Title: Molecular weight of dissolved organic matter determines its interactions with microbes and its assembly processes in soils

Authors: Pengfa Li [a,b], Meng Wu [b,*], Ting Li [a], Alex J. Dumbrell [c], Muhammad Saleem [d], Lu Kuang [a], Lu Luan [b], Shuang Wang [e], Zhongpei Li [b], Jiandong Jiang [a,*]

Affiliations:

[a] Department of Microbiology, College of Life Sciences, Nanjing Agricultural University, Key Laboratory of Agricultural and Environmental Microbiology, Ministry of Agriculture and Rural Affairs, Nanjing 210095, China

[b] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

[c] School of Life Sciences, University of Essex, Colchester, Essex, United Kingdom

[d] Department of Biological Sciences, Alabama State University, Montgomery, AL 36104, USA

[e] Heilongjiang Academy of Black Soil Conservation & Utilization, Harbin 150086, China

[*]**Corresponding author**: Wu Meng; Jiandong Jiang.

Meng Wu: Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China

Jiandong Jiang: College of Life Sciences, Nanjing Agricultural University, WeiGang Road 1, Nanjing 210095, China

Telephone: +86-025-86881323

Corresponding E-mail: Meng Wu (mwu@issas.ac.cn), Jiandong Jiang (jiang_jjd@njau.edu.cn).

Abstract

Dissolved organic matter (DOM) is involved in numerous biogeochemical processes, and its molecular weight affects many of these processes through its bioavailability and sorptive capacity. However, it remains unknown to what extent the molecular weight of DOM mediates its dynamics, for example, influencing its role in DOM-microbe interactions and the processes determining the compositional assembly of DOM. To address this issue, ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) and high-throughput sequencing were applied to investigate how the molecular weight of DOM was associated with its dynamics in two typical agricultural soils with different fertility. Our results showed that low-molecular-weight DOM had lower biological stability and a higher transformation potential. Analysis of the DOM-microbe co-occurrence network showed that low-molecular-weight DOM displayed tighter interactions with a diversity of microbes, while high-molecular-weight DOM interacted with only a few microbes. Ecological null models revealed that the compositional assembly of low-molecular-weight DOM, but not high-molecular-weight DOM, was more controlled by deterministic processes. Taken together, our results demonstrate the fundamental role the molecular weight of DOM plays in determining biological stability, transformation potential, interactions with microbes, and assembly mechanisms of DOM in agricultural soils. This work provides the foundation for general principles explaining complex dynamics of DOM in natural ecosystems, highlighting that using theories and concepts in metacommunity ecology, such as community diversity and assembly mechanisms, may open a new avenue to understand DOM dynamics from a macro perspective.

Keywords: Dissolved organic matter; Molecular weight; Biological stability; Transformations; DOM-microbe interaction; Assembly process

Introduction

As the most reactive component of organic matter, dissolved organic matter (DOM) plays pivotal roles in biogeochemical cycling of carbon and nitrogen, climate change and human health (Leinemann et al., 2018; Li et al., 2018b; Tanentzap et al., 2019). However, due to the extreme complexity and dynamics of DOM in natural ecosystems, our understanding of the dynamics of DOM is still limited. In metacommunity ecology, over a hundred years of research has enabled us to understand complex ecosystems from a macro perspective. However, most research on DOM has remained in a preliminary stage, describing its composition and relative abundance, or comparing its structural differences across environments (Han et al., 2022). Thus, researchers are increasingly focusing on applying theories and methods from metacommunity ecology, for example, concepts of community diversity and assembly mechanisms, to understand DOM dynamics (Hu et al., 2022a; Li et al., 2022b).

Initial applications of metacommunity theory, demonstrated the interlinkages between the chemodiversity of DOM and both microbial community dynamics (Li et al., 2018a; Li et al., 2019b), and ecosystem functioning (Tanentzap et al., 2019). Recently, concepts and tools developed in metacommunity ecology (i.e. niche vs neutral theories) have also been successfully applied to investigate and understand the compositional assembly of DOM (Danczak et al., 2020). In metacommunity ecology, researchers have attributed changes in the composition and diversity of ecological communities to several eco-evolutionary processes (Li et al., 2019a). There are two divergent but complementary theories that describe community assembly. These are niche theory and neutral theory (Zhou and Ning, 2017). Niche theory posits that deterministic processes such as competition, facilitation, predation, resource differentiation, and other environmental filters

primarily determine the assembly of ecological communities (Dini-Andreote et al., 2015). In contrast, neutral theory suggests that stochastic processes such as colonisation, dispersal, priority effects, and ecological drift regulate the assembly and functioning of ecological communities (Dini-Andreote et al., 2015).

Within natural ecosystems, the composition of DOM represents the collective outcomes of historical processes that have resulted in the gain (e.g. via hydrological activities, and biological and non-biological production), loss (e.g. via hydrological activities, adsorption, and mineralisation), and transformation (including both spontaneous transformation such as spontaneous chemical reactions, and passive transformation such as directed biological transformation) of individual DOM molecules (Fig. 1a). Similar to an ecological community, DOM will also undergo fluctuations in production and degradation rates (analogous to birth and death rates in ecological communities), driven by abiotic or biotic transformations, and controlled by advective hydrologic transport or vector movement (analogous to dispersal in an ecological community) (Graham et al., 2018; Kellerman et al., 2020).

In metacommunity ecology, selection, including via abiotic conditions and various antagonistic or synergistic biotic interactions, is considered a deterministic process (Zhou and Ning, 2017). Similarly, selection pressures derived from both environmental filtering and biotic factors also deterministically shape the compositional assembly of DOM (Boye et al., 2017; Hu et al., 2022b). An ecological community will also be subject to influences from dispersal limitation, mass effects, ecological drift and diversification, which are largely considered to be stochastic processes (Martiny et al., 2006; Urban et al., 2008). To some extent, when compared to these mechanisms structuring ecological communities, the hydrologic transport and vector movement of DOM molecules is

analogous to dispersal, and any unstructured deviations in DOM composition (e.g., uptake by random microorganisms) is analogous to ecological drift (Danczak et al., 2020). However, there are also obvious differences in the processes structing ecological communities and the composition of DOM assemblages. For example, the effects of diversification and competitive exclusion in ecological communities can not be translated to collections of DOM (Vellend, 2010). Nevertheless, given the parallels between the deterministic and stochastic processes acting upon the compositional assembly of both ecological communities and DOM (Fig. 1a), we propose that the application of concepts and tools developed in metacommunity ecology to understanding the compositional assembly of DOM is conceptually reasonable. This combined with a comprehensive understanding of how intrinsic traits of DOM, such as its molecular weight, relate to DOM dynamics, its transformations, its interactions with microbes, and its assembly mechanisms, is likely to provide critical new insights.

A central aim of ecology is to develop general principles that explain ecological phenomenon. For example, how body weight or body size, one of the most important life-history traits (De Bie et al., 2012), determines the metabolic rate of almost all organisms, including animals, plants, and microorganisms (Woo et al., 2018). Therefore, spatiotemporal variation in ecological communities and their underlying dynamics, are well explained by organism body weight (Luan et al., 2020; Li et al., 2022a). Analogous to organism body weight, molecular weight is an intrinsic attribute of organic molecules, and greatly influences the biogeochemical cycling of DOM (Lou and Xie, 2006). Thus, it is likely that generalisable links between molecular weight and DOM dynamics exist. First, DOM molecules of different molecular weights may have different biological stability (Boye et al., 2017; Underwood et al., 2019), which could affect how DOM interacts with microbes. Given that most

low-molecular-weight DOM molecules are rapidly mineralised and that high-molecular-weight DOM molecules are typically refractory (Docherty et al., 2006), we hypothesise that DOM molecules of higher molecular weight are more biologically stable than those of lower molecular weight (**Hypothesis 1**, Fig. 1b). Second, DOM with different molecular weights may be preferentially consumed by different microbes (Wu et al., 2021). For example, high-molecular-weight DOM enrichment has previously selected for methylotrophs in marine systems (Sosa et al., 2015). Moreover, the biological stability of DOM is also an important factor affecting how it interacts with microbes (Logue et al., 2016; Zhang et al., 2021). If an organic molecule is highly biologically stable, only a few specific microbes are involved in its biotransformation. In contrast, a diversity of microorganisms can transform labile compounds (Li et al., 2018b). Therefore, if our prior hypothesis that high-molecular-weight molecules are more biologically stable is true, then we can further hypothesise that any transformation of high-molecular-weight DOM is driven by a few specific microbial groups, while low-molecular-weight DOM is potentially transformed by a greater diversity of microbes (**Hypothesis 2**, Fig. 1c).

DOM transformations in natural ecosystem can be either spontaneous or passive, and microbes play a fundamental role in these transformations (Huygens et al., 2016; Osterholz et al., 2016). These microbial-driven transformations can be treated as a selection pressure acting upon DOM, and DOM-associated microbes are a biotic factor deterministically regulating the compositions of DOM assemblages (Graham et al., 2018). Consequently, the composition of a DOM assemblage may be more highly controlled by deterministic processes if it interacts with more microbes. Therefore, if our previous hypothesis that low-molecular-weight DOM is potentially transformed by a diversity of microbes is supported, we can further hypothesise that the

compositional assembly of low-molecular-weight DOM is governed by a higher proportion of deterministic processes due to its potentially tighter interactions with microbes (**Hypothesis 3**, Fig. 1d).

Here, we seek to investigate the extent to which molecular weight mediates the dynamics of DOM, especially DOM-microbe interactions and processes governing the compositional assembly of DOM. We conducted field sampling to answer the following questions: (i) does DOM with different molecular weights have different biological stability? (ii) does DOM with different molecular weights have different DOM-microbe interactions? (iii) what are the relative influences of stochastic and deterministic processes mediating the assembly processes of DOM with different molecular weights? To address these questions, we collected two typical agricultural soils with distinct fertility levels (Low fertility: 75 red soil samples; High fertility: 59 black soil samples) in China, and then used high-throughput sequencing and FT-ICR-MS techniques, to quantify soil microbial communities and DOM pools. We then applied ecological modelling approaches to investigate how the molecular weight of DOM is associated with its multidimensional dynamics.

Materials and methods

Soil sample collection, FT-ICR-MS sample preparation and data preprocessing

To test our hypotheses, two typical agricultural soils in China (red paddy soil and black dryland soil), were collected. These soils are distributed in different climatic regions, developing from distinct parent materials, cropping patterns and fertility levels. Red and black soils represent typical low-fertility and high-fertility soils, respectively (Liang et al., 2019). The soil organic matter (SOM) content in our collected red paddy soil ranged from 0.7 to 2.1%, and in black dryland soil ranged

from 2.1 to 3.6%. Red soil samples were collected from rice paddy fields in subtropical China (Yujiang, Jiangxi Province, China; 116°55′ E, 28°15′ N). This region has subtropical monsoon climates, with abundant sunshine and rainfall (mean annual sunshine hours, 1,739.4 h; mean annual temperature, 17.6 ℃; mean annual precipitation, 1,750 mm). A total of 75 paddy soil samples were collected from this area after the rice harvest. Black soil samples were collected from dryland fields in Northeast China (Harbin City, Heilongjiang Province, China; 126°35′ E, 45°40′ N). This region has a temperate continental monsoon climate, with a frost-free period of 135 days and an altitude of 151 m (mean annual temperature, 3.5 ℃; mean annual precipitation, 533 mm). A total of 59 black dryland soils were collected from this area after the soybean harvest. Across sites, within each field, five 20-cm-depth soil cores (6-cm diameter, free from plant roots) were collected and combined into one sample. The soil samples were homogenized and subsampled for further analyses. Subsamples for microbial properties were stored at -40 ℃. Subsamples for DOM extraction were air dried, ground, and sieved through 2-mm mesh.

Soil DOM was extracted from a 10 g sample using ultrapure water with a 1:10 soil:water ratio, and then shaken for 12 h in a horizontal shaker at the room temperature (~ 25 ℃). The soil suspensions were centrifuged at 2800 × g for 20 min, and then filtered through a 0.45-μm membrane filter. The PPL cartridges (Agilent Technologies, Santa Clara, CA, USA) were cleaned with 10 mL HPLC grade methanol (HPLC grade; Merck, Germany) and acidified ultrapure water (10 mL, pH = 2) before DOM analysis. Then, the DOM solution was loaded onto PPL cartridges. After that, the DOM was collected from PPL cartridges using 10 mL of HPLC grade methanol. The DOM elutes were kept at -20 ℃ in the dark prior to measurement via electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR-MS).

We added deuterated octadecanoic acid as an internal standard with a dose of 15 μL (5×10^{-7} mol/L) per milliliter of the DOM sample. The ESI FT-ICR-MS (Bruker, Billerica, MA, USA) was equipped with a 9.4 T superconducting magnet interfaced with negative-ion mode electrospray ionization. We injected each sample into the ESI source at a speed of 180 μL h^{-1} using a syringe pump. The polarization voltage was 4.0 kV. The capillary column introduction and outlet voltages, were 4.5 kV and 320 V, respectively. The ions accumulated in the hexapole for 0.001s before being transferred to the ICR cell. The m/z range was 150–800 Da. A 4 M word size was selected for the time domain signal acquisition. The signal to noise ratio and dynamic range was enhanced through accumulating 128 time domain FT-ICR transients.

Data analysis software (Bruker Daltonics version 4.2) was used to convert raw spectra to final values (m/z) using the FTMS peak picker (S/N threshold of 6; absolute intensity threshold of 100). To reduce cumulative errors, all peaks from the entire dataset were aligned to each other to check the mass shift. The molecular formulae of peaks were calculated using custom software. Only peaks that were observed in at least two samples were selected for further analysis in order to minimize detection error. Furthermore, only the molecules that were successfully assigned to molecular formula were considered in the downstream analysis.

Determination of biological stability and transformation potential of DOM molecules

The Gibbs free energy for the half-reaction of carbon oxidation ΔG°_{Cox} was calculated for DOM molecules to infer their thermodynamic quality, which reflects their biological stability (LaRowe and Van Cappellen, 2011). ΔG°_{Cox} is estimated from the following empirical equation.

$$\Delta G^{\circ}_{Cox} = 60.3 - 28.5 \times NOSC$$

where NOSC is the nominal oxidation state of carbon, which is estimated using the following

196 equation.

197 $$NOSC = 4-[(-Z+4C+H-3N-2O+5P-2S)/C]$$

198 Here, C, H, N, O, P, and S represent the number of atoms of elements C, H, N, O, P, and S
199 (respectively) in a given organic carbon compound, and Z is the corresponding net charge (we
200 assume a neutral charge per molecule). A higher ΔG°_{Cox} for a given organic compound indicates
201 reduced thermodynamic quality and *vice versa* (Stegen et al., 2018). The molecule with a high
202 ΔG°_{Cox} indicates that more energy is required to oxidize this molecule. Therefore, a higher ΔG°_{Cox}
203 reflects a relatively higher biological stability of DOM molecules; and low ΔG°_{Cox} DOM is considered
204 a readily accessible resources for microbes (Zhang et al., 2021).

205 The potential transformation between DOM molecules was determined using the pairwise
206 mass differences between identified molecules following a previously described pipeline with some
207 modifications (Fig. S1) (Danczak et al., 2020). For example, if the mass difference between two DOM
208 molecules was 18.010565, that would putatively indicate a loss or gain of a molecule of water, while
209 a mass difference of 163.063329 would putatively indicate loss or gain of a tyrosine. If the mass
210 difference cannot match a molecular weight of a known molecule, we consider this an invalid
211 transformation. We also checked whether the elements before and after transformation are
212 conservative, and the nonconservative transformations were eliminated (Wu et al., 2022). Based on
213 these, a transformation of DOM molecules was constructed. Using the transformation network, the
214 correlations between DOM molecules were calculated by picking the largest cluster of
215 interconnected nodes (every node denotes an individual DOM molecule). Then, we measured the
216 stepwise distance between each pair of the DOM molecules. Afterwards, the transformation-based
217 dendrogram of DOM molecules was constructed using the UPGMA method (Danczak et al., 2020),

based on the standardized Euclidean distances (Fig. S2).

Soil DNA extraction, sequencing, and processing

Genomic DNA was extracted from 0.5 g of soil using the FastDNA SPIN Kit for soil (MP Biomedicals, Santa Ana, CA). The *16S* rRNA gene primers, 519F and 907R, were used for PCR amplification of the bacterial community (Table S1). We performed high-throughput sequencing using the Illumina MiSeq platform (Illumina Inc., CA, USA). Raw sequence data were demultiplexed and quality filtered using the q2-demux plugin followed by denoising with DADA2 (via q2-dada2) (Callahan et al., 2016), and any sequences not present in at least two samples were filtered out. After quality filtering and the removal of chimaeras, sequences in red and black soils were clustered into 11,215 and 4160 amplicon sequence variants (ASVs), respectively. After rarefying (based on the sample with the minimum numbers of reads) (McKnight et al., 2019), 25,162 sequences per red soil sample and 24,043 sequences per black soil sample were retained. The taxonomic assignment of representative sequences was performed using RDP classifier (http://rdp.cme.msu.edu/classifier/) with 80% confidence threshold (Wang et al., 2007).

Statistical analysis

Analysis of the DOM-microbe interaction network was conducted to investigate potential co-occurrences between microbes and DOM using CoNet in Cytoscape 3.5.1 (Shannon et al., 2003). Significant correlations between DOM and ASVs were determined by the Spearman's correlation coefficient with false discovery rate (FDR) adjustment (P values < 0.05). We constructed correlation networks in which each node represented an ASV or a DOM molecule, and each edge represented a significant and strong correlation between nodes. After removing the microbe-microbe links and

DOM-DOM links, a DOM-microbe co-occurrence network was produced.

We used two approaches, namely, modified normalised stochasticity ratio (NST) (Ning et al., 2019) and an ecological assembly model (betaNTI: beta nearest taxon index) (Dini-Andreote et al., 2015), to investigate the assembly processes governing the composition of soil microbial communities and metabolites. These models are frequently used in metacommunity ecology to disentangle the determinism-stochasticity balance, and were chosen in this study due to their applicability to occurrence only (presence/absence) data. Moreover, betaNTI considers the transformation between DOM molecules (based on a transformation dendrogram) while NST is a dendrogram-free model; thus, we can investigate the determinism-stochasticity balance of the DOM collection with or without considering transformations. Moreover, we choose NST to indicate assembly processes because our research met the requirements of this method: (i) local/landscape scale sampling as opposed to global scale; (ii) $n > 6$. The NST is an index with a threshold of 50% as the boundary point between more deterministic (< 50%) and more stochastic (> 50%) assembly (Ning et al., 2019). For instance, community assembly will be more deterministic and stochastic if MST is < 50% and > 50%, respectively. This analysis was performed in the R package 'NST'. The parameters were set as follows: "dist.method" of "jaccard," "abundance.weighted" of "TRUE", and "rand" of "1000" (Li et al., 2020). For betaNTI analysis, the dendrogram and abundance table were used to determine the betaMNTD (beta mean nearest taxon distance) and NTI (nearest taxon index). The betaMNTD is the mean phylogenetic distance to the closest relative between pairs of communities, and the betaNTI is the between-assemblage analog of the NTI. The community is considered to be deterministically assembled if the absolute betaNTI > 2, and the community is considered to be stochastically assembled if the absolute betaNTI < 2 (Dini-Andreote et al., 2015).

The betaNTI matrix was calculated using the "picante" package in R. Briefly, the betaNTI was calculated as follows:

$$\mathrm{betaNTI} = -1\left(\frac{\mathrm{betaMNTD_{obs}} - \overline{\mathrm{betaMNTD_{null}}}}{\mathrm{SD(betaMNTD_{null})}}\right)$$

where the $\mathrm{betaMNTD_{obs}}$ is the observed betaMNTD for the observed assemblages, while the $\overline{\mathrm{betaMNTD_{null}}}$ is the average betaMNTD for the null communities. The $\mathrm{SD(betaMNTD_{null})}$ is the standard deviation of the $\mathrm{betaMNTD_{null}}$ values.

Results

Molecular composition of DOM with different molecular weights

In red soils, the FT-ICR-MS analysis identified 12,927 DOM molecules across all samples, while each sample had 6,125 DOM molecules on average (Fig. 2a). The molecular weight of the detected DOM molecules ranged from 151 to 759 Da, with an average molecular weight of 408 Da. According to the oxygen/carbon ratio (C/N) and hydrogen/carbon ratio (H/C) information, lignins comprised the highest proportion present (44.98%), followed by other compounds (21.78%), lipids (16.53%), protein/amino sugars (6.74%), tannins (3.38%), condensed aromatics (3.06%), carbohydrates (2.35%), and unsaturated hydrocarbons (1.17%, Fig. 2a). Dissolved carbohydrates (CHO) had highest proportion (33.21%), followed by CHNO (30.20%) and CHOS (20.44%), while CHNOS (10.57%) and CHOP (5.59%) had the lowest (Fig. S3). In black soils, FT-ICR-MS analysis identified 10,910 DOM molecules across all samples, while each sample had 5,251 DOM molecules on average. Although black and red soils have distinct fertility levels, the molecular composition of DOM in black soils was similar to that in red soils (Fig. 2a, Fig. S3).

We divided the DOM molecules from red and black soils into 25 groups and 21 groups

282 respectively, according to their molecular weight (each group has equal number of DOM
283 molecules). With increased molecular weight, the proportions of lignins, lipids and other DOM
284 compounds gradually increased, while the proportions of condensed aromatics, carbohydrates,
285 protein/amino sugars, tannins and unsaturated hydrocarbons gradually declined (Fig. 2b-c). In both
286 red and black soils, our results showed a clear decrease of nitrogen-containing DOM (CHNO and
287 CHNOS) with increased molecular weight. On the contrary, the proportion of nitrogen-free (CHO,
288 CHOP and CHOS) DOM increased with increased molecular weight (Fig. S3). Correlation analysis
289 showed that the proportion of nitrogen-containing DOM and nitrogen-free DOM was significantly
290 negatively and positively correlated to DOM molecular weight, respectively (Fig. S3).

291 We determined the correlation between DOM molecular weight and their alpha- and beta-
292 diversity indices. Results showed that DOM molecular weight was significantly negatively correlated
293 with their alpha-diversity (Red soil: r = -0.884, P < 0.001; Black soil: r = -0.900, P < 0.001; Fig. S4),
294 but significantly positively correlated with their beta-diversity (Red soil: r = 0.924, P < 0.001; Black
295 soil: r = 0.894, P < 0.001; Fig. S4). These results indicated that the higher-molecular-weight DOM
296 had lower occurrence, and was more compositionally distinct across different samples.

297 **Biological stability and transformation potential of DOM with different molecular weights**

298 The Gibbs free energy ($\Delta G°$) of each DOM molecule was estimated to indicate biological stability
299 (Zhang et al., 2021; Wu et al., 2022). Our results showed that $\Delta G°$ was significantly positively
300 correlated with the molecular weight of DOM (Red soil: r = 0.340, P < 0.001; Black soil: r = 0.323, P
301 < 0.001; Fig. 3a). For different categories of DOM molecules, the significant positive correlation
302 between their molecular weight and $\triangle G°$ was also consistently observed (P < 0.001, Fig. S5).
303 Specifically, the correlation coefficients between molecular weight and $\triangle G°$ for CHNO, CHNOS, CHO,

CHOP, and CHOS molecules in red soils were 0.296, 0.533, 0.135, 0.278, and 0.469, respectively. Similarly, the correlation coefficients for CHNO, CHNOS, CHO, CHOP, and CHOS molecules in black soils were 0.231, 0.612, 0.131, 0.396, and 0.523, respectively (Fig. S5).

After determining the mass differences among all DOM molecules, and excluding false-positive transformations (potential transformation of element non-conservation), a DOM transformation network was constructed (Fig. S6). The transformation network showed that the DOM with higher degree values consistently had lower molecular weight in both red and black soils (Fig. S6). A correlation analysis showed that the number of potential transformations was significantly negatively correlated with DOM molecular weight (Red soil: r = -0.468, P < 0.001; Black soil: r = -0.513, P < 0.001; Fig. 3b). This negative correlation was also observed across different categories of DOM. In red soils, the correlation coefficients between molecular weight and potential transformations for CHNO, CHNOS, CHO, CHOP, and CHOS molecules were -0.430, -0.242, -0.603, -0.360, and -0.567, respectively. In black soils, the correlation coefficients for CHNO, CHNOS, CHO, CHOP, and CHOS molecules were -0.482, -0.250, -0.642, -0.054, and -0.649, respectively (P < 0.001, Fig. S7).

Interactions between microbes and DOM molecules

After excluding the DOM-DOM links and microbe-microbe links, we constructed a DOM-microbe co-occurrence network to investigate the potential interactions between DOM and microbes (Fig. 4a). In red soils, the DOM-microbe network had a total of 929,426 links, 736,615 of which were positive (79.25%) and 192,811 were negative (20.75%). Of all 11,215 microbial ASVs, only 261 ASVs entered the DOM-microbe network. While these 261 microbial ASVs only accounted for 1.57% of all ASVs, they accounted for 56.66% of total sequences. In black soils, the DOM-microbe network

had a total of 991,895 links, 396,462 of which were positive (39.97%) and 595,433 were negative (60.03%). After extracting the microbial nodes, 753 out of 4160 microbial ASVs (18.10%), representing 74.09% of total sequences, were observed in the DOM-microbe network. The DOM-associated ASVs in red soils consisted of 10 microbial phyla, with major constituents being *Firmicutes* (99 ASVs), *Chloroflexi* (76 ASVs), *Proteobacteria* (24 ASVs), *Myxococcota* (18 ASVs), and *Acidobacteriota* (14 ASVs) (Fig. 4b). In black soils, the DOM-associated ASVs consisted of 18 microbial phyla, with major constituents being *Actinobacteriota* (237 ASVs), *Proteobacteria* (192 ASVs), *Chloroflexi* (103 ASVs), *Acidobacteriota* (88 ASVs), and *Gemmatimonadota* (39 ASVs) (Fig. 4b).

We calculated the number of microbes that were linked to each DOM molecule, and compared this to its molecular weight. The results showed that DOM molecular weight was consistently negatively correlated to DOM-microbe co-occurrences (Red soil: r = -0.238, P < 0.001; Black soil: r = -0.277, P < 0.001; Fig. 4c). After grouping the DOM molecules by element ratio information, these negative relationships remained, with correlation coefficients ranging from -0.110 (Carbohydrate in black soil, P = 0.219) to -0.398 (Lignin in red soil, P < 0.001, Table S2). Moreover, for different categories of DOM molecules, the negative correlation between DOM molecular weight and DOM-microbe co-occurrences can also be consistently observed, with the exception of CHOP in red soils (Table S2).

Assembly processes of DOM with different molecular weights

Two ecological null models were used to investigate the compositional assembly of DOM with different molecular weights. First, a DOM dendrogram was constructed after standardizing the links from our DOM transformation network, and a dendrogram-informed ecological model was applied.

348 The model showed that the assembly of low-molecular-weight DOM was highly governed by
349 deterministic processes in both red (Fig. 5a) and black soils (Fig. 5b). For example, in red soils, more
350 than 98% of paired betaNTI indices of the DOM in group 1, whose average molecular weight was
351 177.9 Da, were > 2 (deterministic process). The assembly of high-molecular-weight DOM was highly
352 influenced by stochastic processes. For example, 76% of paired betaNTI indices of the DOM in
353 group 25, whose average molecular weight was 652.6 Da, were < 2 (stochastic process; Fig. 5a). We
354 also investigated the assembly processes of DOM with different molecular weights using the
355 dendrogram-free model, modified normalised stochasticity ratio (NST). The results showed that
356 most NST indices of the low-molecular-weight DOM were close to 0.5 or < 0.5 (more deterministic),
357 while most NST indices of high-molecular-weight DOM were > 0.5 (more stochastic; Fig. S8).
358 We then compared the average betaNTI indices and average NST indices of DOM against its
359 average molecular weight. This yielded significant negative correlations between average betaNTI
360 indices and DOM molecular weight (Red soil: r = -0.781, P < 0.001, Fig. 5c; Black soil: r = -0.537, P
361 = 0.012, Fig. 5d), but significant positive correlations between average NST indices and DOM
362 molecular weight (Red soil: r = 0.933, P < 0.001; Black soil: r = 0.846, P < 0.001; Fig. S8).

363 **Discussion**

364 Our study aimed to develop a complex ecological model of DOM in natural ecosystems. By
365 applying the concepts and models that underpin metacommunity ecology, our results revealed the
366 fundamental role of an intrinsic trait of DOM (i.e., its molecular weight) in determining the
367 biological stability, transformation potential, interaction with microbes, and the compositional
368 assembly of DOM in agricultural soils.

We observed consistent positive correlations between DOM molecular weight and ΔG° (Fig. 3a), an index indicating biological stability of organic molecules. This supports our **Hypothesis 1**, that high-molecular-weight DOM molecules are more biologically stable than low-molecular-weight DOM molecules, and suggests that low-molecular-weight rather than high-molecular-weight compounds are more thermodynamically available to microbes (Zhang et al., 2021). The labile low-molecular-weight compounds can be rapidly mineralised (Docherty et al., 2006), or converted into refractory molecules by microbes (Ogawa et al., 2001). As a consequence, the low-molecular-weight DOM had a higher transformation potential, while the high-molecular-weight DOM had a much lower transformation potential (Fig. 3b).

Using network analysis, our results showed more complex DOM-microbe co-occurrences in black soils than in red soils (Fig. 4), suggesting that higher soil fertility may promote DOM-microbe interactions. More importantly, the DOM molecular weight was consistently negatively correlated to DOM-microbe co-occurrences (Fig. 4c). This implied that the low-molecular-weight DOM interacted with a greater diversity of microbial taxa, while the high-molecular-weight DOM was associated with fewer microbes. These results supported our **Hypothesis 2**, that the transformation of high-molecular-weight DOM is driven by a few specific microbial groups, while low-molecular-weight DOM is potentially transformed by a greater diversity of microbes. One possible explanation for the negative relationship between DOM molecular weight and DOM-microbe co-occurrences is that the low-molecular-weight DOM has lower biological stability than high-molecular-weight DOM (Fig. 2b). The low-molecular-weight DOM molecules with lower biological stability are mainly composed of readily accessible resources for microbes (Stegen et al., 2018). As a consequence, these low-molecular-weight DOM molecules can readily supply a greater diversity of microbes with

391 resources, resulting in numerous and tighter DOM-microbe co-occurrences. Another possible
392 explanation for such a negative relationship is the distinct composition of DOM that have different
393 molecular weights. For example, condensed aromatics and carbohydrates, which can be easily
394 utilized by many microbial taxa, always have relatively lower molecular weights (Fig. 2). On the
395 contrary, lignins have complex chemical structures and therefore can only be degraded by specialist
396 taxa such as *Sterptomyces*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, white rot fungi and brown rot
397 fungi (De Filippo et al., 2010). As a consequence, the high proportion of lignins in high-molecular-
398 weight DOM is unlikely to shape complex DOM-microbe co-occurrences.

399 The ecological null models showed that the compositional assembly of DOM in red soils was
400 more controlled by stochastic processes than in black soils, possibly due to different field
401 management practices. The red soils in the current study are flooded for a long time throughout
402 the year, which may improve the dispersal chance of DOM molecules between different fields, and
403 consequently lead to a relatively higher proportion of stochastic processes. Moreover, the
404 difference in soil fertility, soil texture, local climates and cover plants may also affect the assembly
405 processes of DOM, and the mechanism behind this needs to be disentangled using finely-designed
406 experiments in the future. We observed a significant negative correlation between average betaNTI
407 indices and DOM molecular weight (Fig. 5d), and a significant positive correlation between average
408 NST indices and DOM molecular weight (Fig. S8). These results consistently supported our
409 **Hypothesis 3,** that the compositional assembly of low-molecular-weight DOM was mainly
410 controlled by deterministic processes, whereas the assembly of high-molecular-weight DOM was
411 mainly governed by stochastic processes. In metacommunity ecology, an ecological community
412 may be more deterministically assembled as a consequence of strong environmental selection

(Zhou and Ning, 2017). For an assemblage of DOM, microbial transformations can be considered as a selection pressure. The low-molecular-weight DOM was potentially transformed by a diversity of microbes (Fig. 4), and thus these low-molecular-weight DOM molecules experience stronger selection pressure by microbes than high-molecular-weight DOM. As a consequence, low-molecular-weight DOM molecules were more deterministically assembled than high-molecular-weight DOM molecules. It is worth noting that not all transformations should be considered deterministic. For example, the random binding and substitution of functional groups in complex mixed systems may lead to undirected transformations, which may be considered stochastic, and may have an increased chance of happening for high-molecular-weight DOM due to its greater number of functional groups. In metacommunity ecology, an ecological community may be more stochastically assembled as a consequence of ecological drift, which is primarily caused by random fluctuations in birth and death rates (Stegen et al., 2013). In the absence of temporally or spatially consistent factors determining variation in DOM molecule production or degradation rates, a dynamic similar to ecological drift could emerge. This 'DOM drift' would comprise unstructured compositional deviations (e.g., uptake by random microorganisms; driven by random abiotic processes operating on molecules such as adsorption, reduction, and changing environment such as moisture, temperature, pH, and redox; and any process randomly generating a free-radical, or random scission in large DOM molecules) and could occur in DOM assemblages (Danczak et al., 2020). Generally, an ecological community with lower diversity may be more influenced by ecological drift (Martiny et al., 2006). Our results showed that the high-molecular-weight DOM had much lower diversity than low-molecular-weight DOM (Fig. S4), which may lead to stronger 'DOM drift' for high-molecular-weight DOM and thus a greater influence of stochastic processes acting

upon its assembly.

Conclusion

Overall, this study revealed that the molecular weight, an intrinsic trait of DOM, influences the multifaceted dynamics of DOM in soils. Using ecological modelling approaches, we demonstrated that the molecular weight of DOM is positively correlated with its biological stability, and is negatively correlated with its transformation potential and interaction with microbes. Moreover, our results showed that the compositional assembly of low-molecular-weight DOM is generally controlled by deterministic processes, whereas the compositional assembly of high-molecular-weight DOM is generally governed by stochastic processes. Although the relationships between molecular weight and multiple DOM dynamics are generally consistent across the two investigated soils with different fertility levels, soils with distinct physical texture or biological properties may retain different DOM molecules and hence different relationships with DOM molecular weight underpinning its dynamics. It is important to acknowledge that the relationships between molecular weight and DOM dynamics may vary across different biomes and geographical scales, since advective hydrologic transport or vector movement of DOM molecules may differ greatly across biomes (Kellerman et al., 2014), which would also significantly affect the determinism-stochasticity balance behind DOM assembly. It is also worth noting that this study aimed to understand the dynamics of DOM from a macro perspective, and therefore, our findings may not be fully applicable to some specific DOM molecules because all general ecological principles have exceptions. For example, while benzene has a small molecular weight, it has a very stable structure in the absence of molecular oxygen and thus is hard to transform by microbes (van der Waals et

al., 2017). Moreover, natural ecosystems contain many DOM molecules with extremely large molecular weights, which are out of the detection scope of FT-ICR-MS (Chuang et al., 2020; Li et al., 2022c). The applicability of our conclusions may need to be refined after further extending analysis to include these extremely large DOM molecules, which may be more bioactive than low-molecular-weight DOM in, for example, aquatic systems (Amon and Benner, 1994). This study paves the way for future research on DOM dynamics, especially on DOM transformations and DOM-microbe interactions in natural ecosystems, using theories, concepts and modelling approaches that were previously established in community ecology.

Acknowledgments

We are grateful to two anonymous reviewers for their valuable suggestions, recommendations, and corrections that improved the quality of this manuscript. This study was supported by the Natural Science Foundation of Jiangsu Province (BK20221005), National Natural Science Foundation of China (42207349, 42077021), Fundamental Research Funds for the Central Universities (KYQN2023027), National Key R&D Program of China (2021YFD1500300), China Postdoctoral Science Foundation (2022M711653), and Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB331).

Competing interests

The authors declare that they have no conflict of interest.

Author contributions

MW and JJ designed the framework. PL, WM and SW performed the experiment. PL, MW, TL and LK did the data analysis. PL, AJD, MS, LL, ZL and JJ wrote the paper. All authors discussed the results and commented on the manuscript.

Data accessibility statement

The *16S* rRNA gene sequencing data we used and other source data generated in the current study are publicly available in Figshare (https://doi.org/10.6084/m9.figshare.21800850).

References

Amon, R.M.W., Benner, R., 1994. Rapid-cycling of high-molecular-weight dissolved organic-matter in the ocean. Nature 369, 549-552.

Boye, K., Noel, V., Tfaily, M.M., Bone, S.E., Williams, K.H., Bargar, J.R., Fendorf, S., 2017. Thermodynamically controlled preservation of organic carbon in floodplains. Nature Geoscience 10, 415-419.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13, 581-583.

Chuang, C.W., Hsu, L.F., Tsai, H.C., Liu, Y.Y., Huang, W.S., Chen, T.C., 2020. Nickel binding affinity with size-fractioned sediment dissolved and particulate organic matter and correlation with optical indicators. Applied Sciences-Basel 10, 8995.

Danczak, R.E., Chu, R.K., Fansler, S.J., Goldman, A.E., Graham, E.B., Tfaily, M.M., Toyoda, J., Stegen, J.C., 2020. Using metacommunity ecology to understand environmental metabolomes. Nature Communications 11, 6369.

De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, H., Denys, L., Vanhecke, L., Van der Gucht, K., Van Wichelen, J., Vyverman, W., Declerck, S.A.J., 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. Ecology Letters 15, 740-747.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini,

500 G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a comparative study
501 in children from Europe and rural Africa. Proceedings of The National Academy of Sciences of
502 the United States of America 107, 14691-14696.
503 Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., Salles, J.F., 2015. Disentangling mechanisms that
504 mediate the balance between stochastic and deterministic processes in microbial succession.
505 Proceedings of the National Academy of Sciences of the United States of America 112, E1326-
506 E1332.
507 Docherty, K.M., Young, K.C., Maurice, P.A., Bridgham, S.D., 2006. Dissolved organic matter
508 concentration and quality influences upon structure and function of freshwater microbial
509 communities. Microbial Ecology 52, 378-388.
510 Graham, E.B., Crump, A.R., Kennedy, D.W., Arntzen, E., Fansler, S., Purvine, S.O., Nicora, C.D., Nelson,
511 W., Tfaily, M.M., Stege, J.C., 2018. Multi 'omics comparison reveals metabolome biochemistry,
512 not microbiome composition or gene expression, corresponds to elevated biogeochemical
513 function in the hyporheic zone. Science of the Total Environment 642, 742-753.
514 Han, H.X., Feng, Y.J., Chen, J., Xie, Q.R., Chen, S., Sheng, M., Zhong, S.J., Wei, W., Su, S.H., Fu, P.Q.,
515 2022. Acidification impacts on the molecular composition of dissolved organic matter revealed
516 by FT-ICR MS. Science of the Total Environment 805, 150284.
517 Hu, A., Choi, M., Tanentzap, A.J., Liu, J.F., Jang, K.S., Lennon, J.T., Liu, Y.Q., Soininen, J., Lu, X.C., Zhang,
518 Y.L., Shen, J., Wang, J.J., 2022a. Ecological networks of dissolved organic matter and
519 microorganisms under global change. Nature Communications 13, 3600.
520 Hu, A., Jang, K.S., Meng, F.F., Stegen, J., Tanentzap, A.J., Choi, M., Lennon, J.T., Soininen, J., Wang, J.J.,
521 2022b. Microbial and environmental processes shape the link between organic matter

functional traits and composition. Environmental Science & Technology 56, 10504-10516.

Huygens, D., Diaz, S., Urcelay, C., Boeckx, P., 2016. Microbial recycling of dissolved organic matter confines plant nitrogen uptake to inorganic forms in a semi-arid ecosystem. Soil Biology & Biochemistry 101, 142-151.

Kellerman, A.M., Arellano, A., Podgorski, D.C., Martin, E.E., Martin, J.B., Deuerling, K.M., Bianchi, T.S., Spencer, R.G.M., 2020. Fundamental drivers of dissolved organic matter composition across an Arctic effective precipitation gradient. Limnology and Oceanography 65, 1217-1234.

Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J., 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nature Communications 5, 3804.

LaRowe, D.E., Van Cappellen, P., 2011. Degradation of natural organic matter: A thermodynamic analysis. Geochimica Et Cosmochimica Acta 75, 2030-2042.

Leinemann, T., Preusser, S., Mikutta, R., Kalbitz, K., Cerli, C., Hoschen, C., Mueller, C.W., Kandeler, E., Guggenberger, G., 2018. Multiple exchange processes on mineral surfaces control the transport of dissolved organic matter through soil profiles. Soil Biology & Biochemistry 118, 79-90.

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 with microbial community at continental scales. Microbiome 6, 169.

Li, P., Dumbrell, A.J., Saleem, M., Kuang, L., Li, T., Luan, L., Li, W., Li, G., Wu, M., Wang, B., Jiang, J., Liu, M., Li, Z., 2022a. Linking microbial body size to community co-occurrences and stability at multiple geographical scales in agricultural soils. Advances in Ecological Research 67, 1-26.

Li, P.F., Li, W.T., Dumbrell, A.J., Liu, M., Li, G.L., Wu, M., Jiang, C.Y., Li, Z.P., 2020. Spatial variation in soil fungal communities across paddy fields in subtropical China. mSystems 5, e00704-19.

Li, P.F., Liu, J., Jiang, C.Y., Wu, M., Liu, M., Li, Z.P., 2019a. Distinct successions of common and rare bacteria in soil under humic acid amendment - A microcosm study. Frontiers in Microbiology 10, 2271.

Li, P.F., Liu, J., Saleem, M., Li, G.L., Luan, L., Wu, M., Li, Z.P., 2022b. Reduced chemodiversity suppresses rhizosphere microbiome functioning in the mono-cropped agroecosystems. Microbiome 10, 108.

Li, X.M., Chen, Q.L., He, C., Shi, Q., Chen, S.C., Reid, B.J., Zhu, Y.G., Sun, G.X., 2019b. Organic carbon amendments affect the chemodiversity of soil dissolved organic matter and its associations with soil microbial communities. Environmental Science & Technology 53, 50-59.

Li, X.M., Sun, G.X., Chen, S.C., Fang, Z., Yuan, H.Y., Shi, Q., Zhu, Y.G., 2018b. Molecular Chemodiversity of Dissolved Organic Matter in Paddy Soils. Environmental Science & Technology 52, 963-971.

Li, Y., Chen, Z.M., Chen, J., Castellano, M.J., Ye, C.L., Zhang, N., Miao, Y.C., Zheng, H.J., Li, J.J., Ding, W.X., 2022c. Oxygen availability regulates the quality of soil dissolved organic matter by mediating microbial metabolism and iron oxidation. Global Change Biology 28, 7410-7427.

Liang, F., Li, J.W., Zhang, S.Q., Gao, H.J., Wang, B.R., Shi, X.J., Huang, S.M., Xu, M.G., 2019. Two-decade long fertilization induced changes in subsurface soil organic carbon stock vary with indigenous site characteristics. Geoderma 337, 853-862.

Logue, J.B., Stedmon, C.A., Kellerman, A.M., Nielsen, N.J., Andersson, A.F., Laudon, H., Lindstrom, E.S., Kritzberg, E.S., 2016. Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. ISME Journal 10, 533-545.

566 Lou, T., Xie, H.X., 2006. Photochemical alteration of the molecular weight of dissolved organic matter.
567 Chemosphere 65, 2333-2342.
568 Luan, L., Jiang, Y.J., Cheng, M.H., Dini-Andreote, F., Sui, Y.Y., Xu, Q.S., Geisen, S., Sun, B., 2020.
569 Organism body size structures the soil microbial and nematode community assembly at a
570 continental and global scale. Nature Communications 11,6406.
571 Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine,
572 M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreas, L., Reysenbach, A.L.,
573 Smith, V.H., Staley, J.T., 2006. Microbial biogeography: putting microorganisms on the map.
574 Nature Reviews Microbiology 4, 102-112.
575 McKnight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., Zenger, K.R., 2019. Methods
576 for normalizing microbiome data: An ecological perspective. Methods in Ecology and
577 Evolution 10, 389-400.
578 Ning, D.L., Deng, Y., Tiedje, J.M., Zhou, J.Z., 2019. A general framework for quantitatively assessing
579 ecological stochasticity. Proceedings of the National Academy of Sciences of the United States
580 of America 116, 16892-16898.
581 Ogawa, H., Amagai, Y., Koike, I., Kaiser, K., Benner, R., 2001. Production of refractory dissolved
582 organic matter by bacteria. Science 292, 917-920.
583 Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J., Dittmar, T., 2016.
584 Deciphering associations between dissolved organic molecules and bacterial communities in
585 a pelagic marine system. ISME Journal 10, 1717-1730.
586 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B.,
587 Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular

588 interaction networks. Genome Research 13, 2498-2504.
589 Sosa, O.A., Gifford, S.M., Repeta, D.J., DeLong, E.F., 2015. High molecular weight dissolved organic
590 matter enrichment selects for methylotrophs in dilution to extinction cultures. ISME Journal 9,
591 2725-2739.
592 Stegen, J.C., Johnson, T., Fredrickson, J.K., Wilkins, M.J., Konopka, A.E., Nelson, W.C., Arntzen, E.V.,
593 Chrisler, W.B., Chu, R.K., Fansler, S.J., Graham, E.B., Kennedy, D.W., Resch, C.T., Tfaily, M., Zachara,
594 J., 2018. Influences of organic carbon speciation on hyporheic corridor biogeochemistry and
595 microbial ecology. Nature Communications 9, 1034.
596 Stegen, J.C., Lin, X.J., Fredrickson, J.K., Chen, X.Y., Kennedy, D.W., Murray, C.J., Rockhold, M.L.,
597 Konopka, A., 2013. Quantifying community assembly processes and identifying features that
598 impose them. ISME Journal 7, 2069-2079.
599 Tanentzap, A.J., Fitch, A., Orland, C., Emilson, E.J.S., Yakimovich, K.M., Osterholz, H., Dittmar, T., 2019.
600 Chemical and microbial diversity covary in fresh water to influence ecosystem functioning.
601 Proceedings of The National Academy of Sciences of the United States of America 116, 24689-
602 24695.
603 Underwood, G.J.C., Michel, C., Meisterhans, G., Niemi, A., Belzile, C., Witt, M., Dumbrell, A.J., Koch,
604 B.P., 2019. Organic matter from Arctic sea-ice loss alters bacterial community structure and
605 function. Nature Climate Change 9, 170-176.
606 Urban, M.C., Leibold, M.A., Amarasekare, P., De Meester, L., Gomulkiewicz, R., Hochberg, M.E.,
607 Klausmeier, C.A., Loeuille, N., de Mazancourt, C., Norberg, J., Pantel, J.H., Strauss, S.Y., Vellend,
608 M., Wade, M.J., 2008. The evolutionary ecology of metacommunities. Trends in Ecology &
609 Evolution 23, 311-317.

van der Waals, M.J., Atashgahi, S., da Rocha, U.N., van der Zaan, B.M., Smidt, H., Gerritse, J., 2017. Benzene degradation in a denitrifying biofilm reactor: activity and microbial community composition. Applied Microbiology and Biotechnology 101, 5175-5188.

Vellend, M., 2010. Conceptual synthesis in community ecology. Quarterly Review of Biology 85, 183-206.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73, 5261-5267.

Woo, C., An, C., Xu, S., Yi, S.M., Yamamoto, N., 2018. Taxonomic diversity of fungi deposited from the atmosphere. ISME Journal 12, 2051-2060.

Wu, M., Li, P.F., Li, G.L., Liu, K., Gao, G.Z., Ma, S.Y., Qiu, C.P., Li, Z.P., 2022. Using potential molecular transformation to understand the molecular trade-offs in soil dissolved organic matter. Environmental Science & Technology 56, 11827-11834.

Wu, M., Li, P.F., Li, G.L., Petropoulos, E., Feng, Y.Z., Li, Z.P., 2021. The chemodiversity of paddy soil dissolved organic matter is shaped and homogenized by bacterial communities that are orchestrated by geographic distance and fertilizations. Soil Biology & Biochemistry 161, 108374.

Zhang, J.W., Feng, Y.Z., Wu, M., Chen, R.R., Li, Z.P., Lin, X.G., Zhu, Y.G., Delgado-Baquerizo, M., 2021. Evaluation of microbe-driven soil organic matter quantity and quality by thermodynamic theory. mBio 12, e03252-20.

Zhou, J.Z., Ning, D.L., 2017. Stochastic community assembly: Does it matter in microbial ecology? Microbiology and Molecular Biology Reviews 81, e00002-17.

632

633 **Figure legends**

634 **Figure 1. The conceptual framework for the current study.** (**a**) The compositional dynamics of
635 DOM, and the parallels between the deterministic and stochastic processes acting upon the
636 assembly of ecological communities and DOM. (**b**) The hypothesised relationship between
637 molecular weight and biological stability of DOM (**Hypothesis 1**). Here we hypothesise that
638 biological stability will increase with increased molecular weight of DOM molecules. (**c**) The
639 hypothesised relationship between molecular weight and DOM-microbe interactions (**Hypothesis**
640 **2**). Here we hypothesise that the DOM-microbe interactions will decline with increased molecular
641 weight of DOM molecules. (**d**) The hypothesised relationship between molecular weight and
642 assembly processes of DOM (**Hypothesis 3**). Based on our previous hypothesis that low-molecular-
643 weight DOM is potentially transformed by a greater diversity of microbes, we further hypothesise
644 that low-molecular-weight DOM is governed by a higher proportion of deterministic process due
645 to potentially tighter interactions with microbes.

646

647 **Figure 2. Molecular composition of DOM in red and black soils.** (**a**) Kernal-density estimation
648 of DOM molecules. (**b**) Molecular composition of DOM with different molecular weights. (**c**) The
649 correlations between molecular weight and relative frequency of different categories of DOM
650 molecules. *, **, and *** indicates a significant correlation at $P < 0.05$, 0.01, and 0.001, respectively.

651

652 **Figure 3. Biological stability and transformation potential of DOM with different molecular**
653 **weights.** (**a**) The correlation between molecular weight and Gibbs free energy of DOM molecules.
654 (**b**) The correlation between DOM molecular weight and number of potential transformations. Lines

represent the least squares regression fits and shaded areas represent 95% confidence intervals. We applied one-side *F* and two-side *t* tests, and calculated *P* values as shown.

Figure 4. Interactions between microbes and DOM. (**a**) The DOM-microbe co-occurrence network. (**b**) Species composition of DOM-associated microbes. (**c**) The correlations between DOM molecular weight and number of DOM-microbe co-occurrences. Lines represent the least squares regression fits and shaded areas represent the 95% confidence intervals. We applied one-side *F* and two-side *t* tests, and calculated *P* values as shown.

Figure 5. Assembly processes of DOM with different molecular weights. The distributions of betaNTI indices of DOM with different molecular weights in (**a**) red soil (**b**) and black soil. The correlations between DOM molecular weight and average betaNTI indices in (**c**) red soil (**d**) and black soil. Lines represent the least squares regression fits and shaded areas represent the 95% confidence intervals. We applied one-side *F* and two-side *t* tests, and then calculated *P* values as shown.

Figure 1

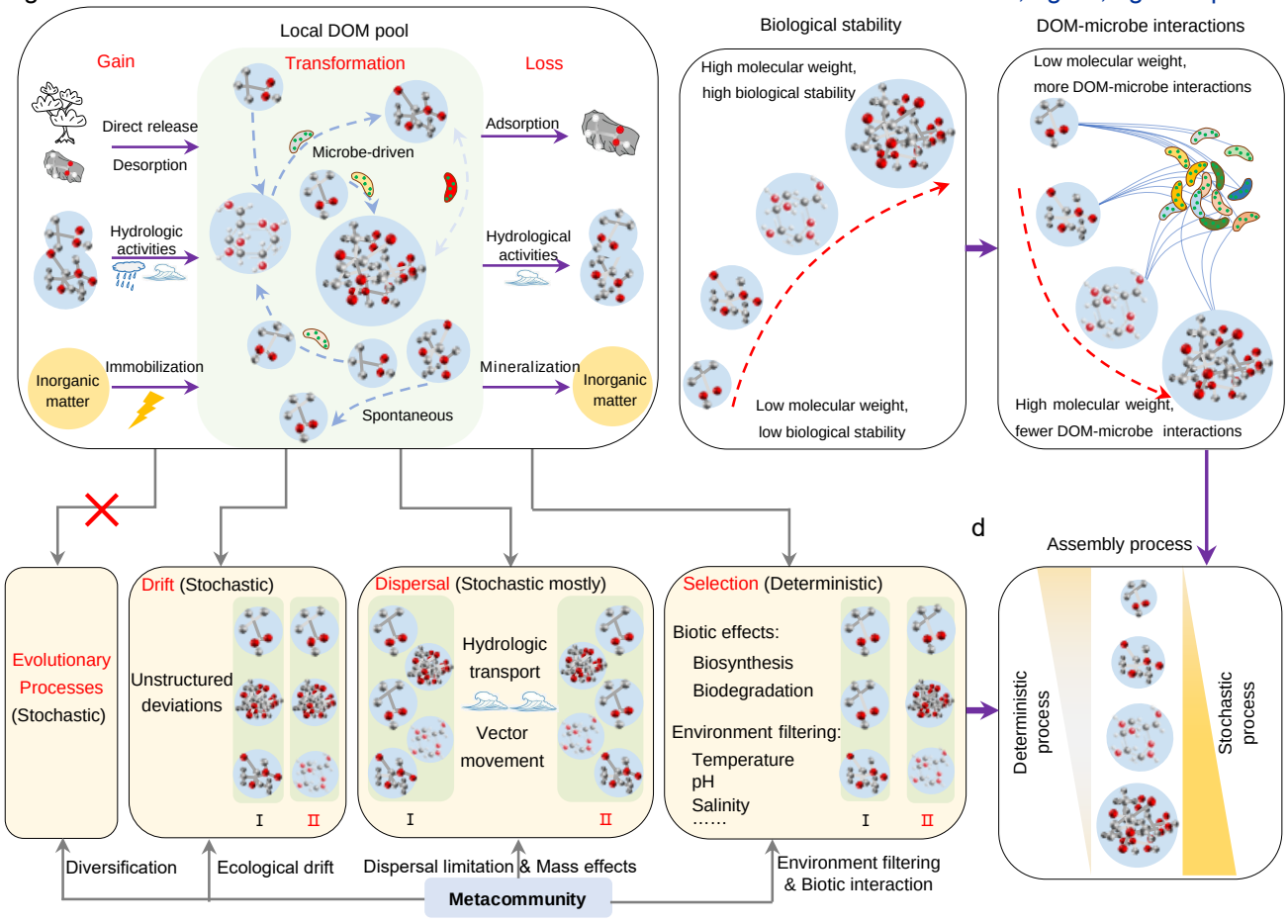

Figure 2

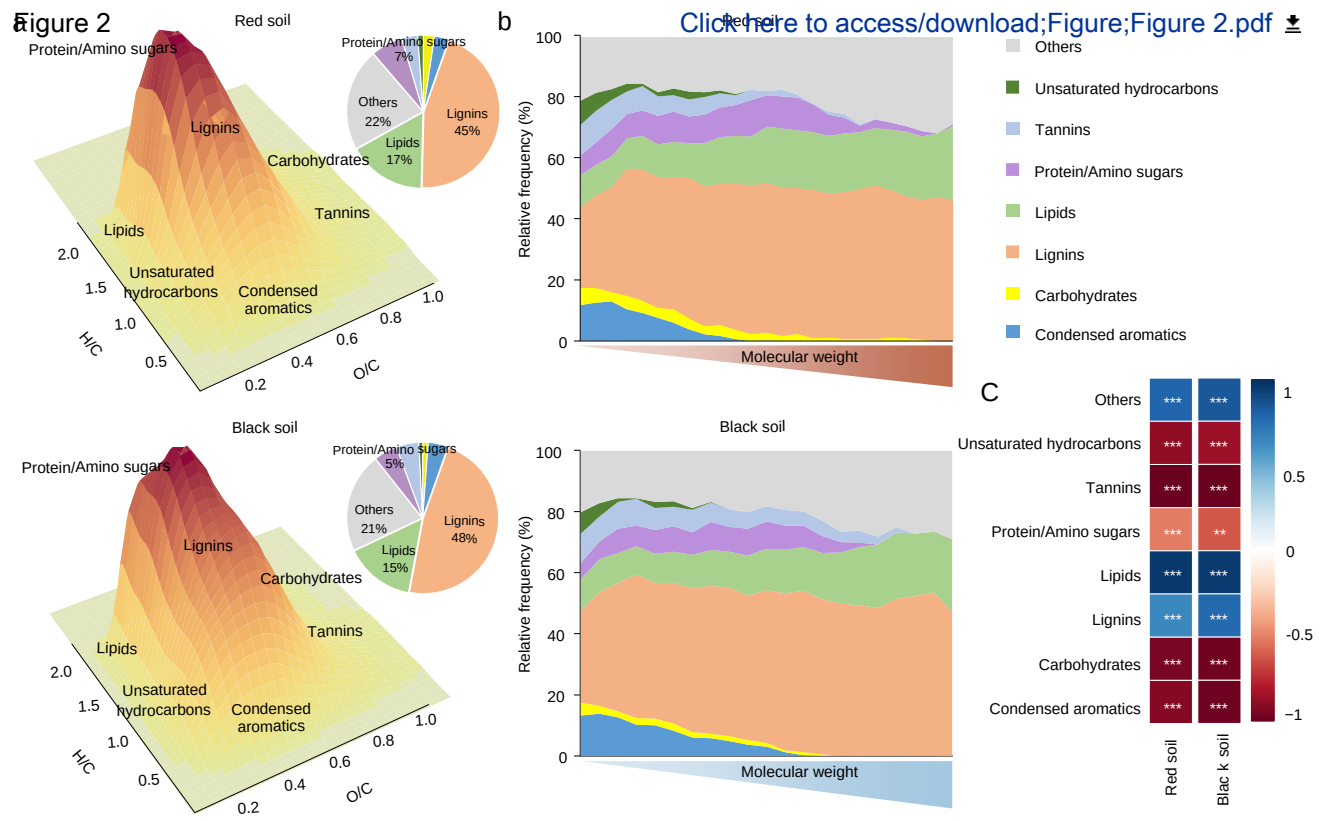

Figure 3

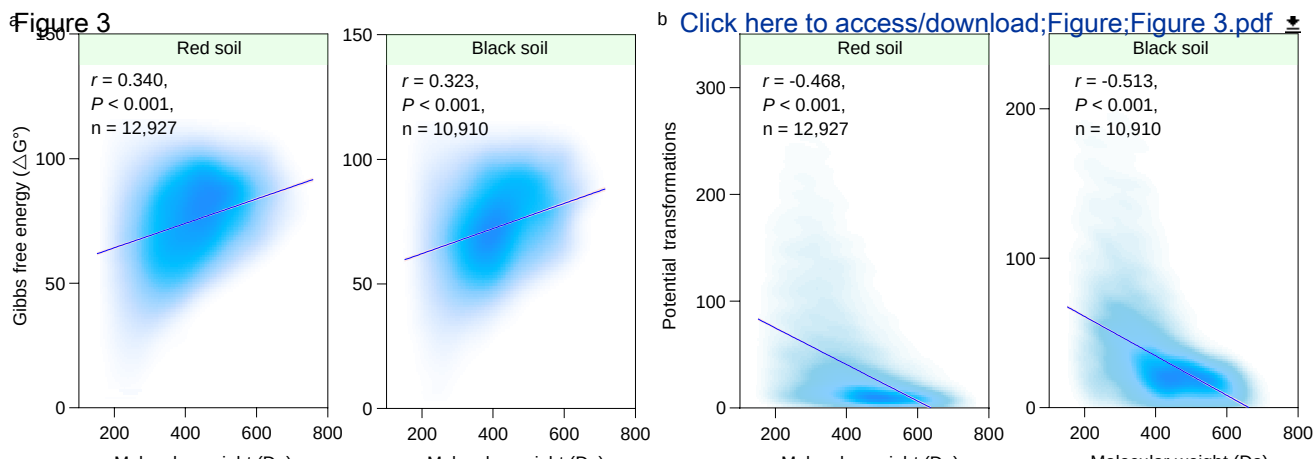
a
Red soil
r = 0.340,
P < 0.001,
n = 12,927
Black soil
r = 0.323,
P < 0.001,
n = 10,910
Gibbs free energy (△G°)
Molecular weight (Da)
Molecular weight (Da)
b
Click here to access/download;Figure;Figure 3.pdf
Red soil
r = -0.468,
P < 0.001,
n = 12,927
Black soil
r = -0.513,
P < 0.001,
n = 10,910
Potential transformations
Molecular weight (Da)
Molecular weight (Da)

Figure 4

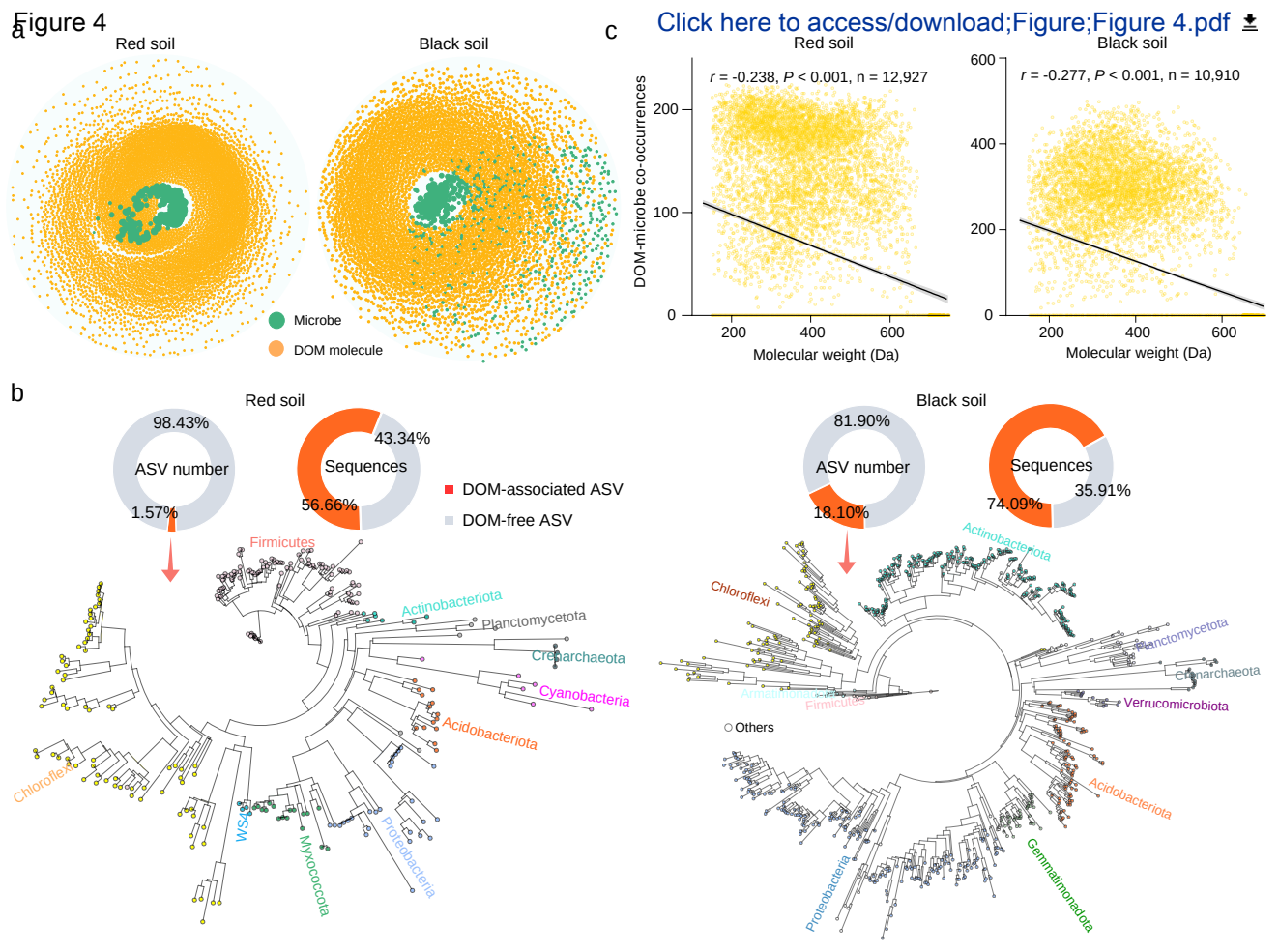

Figure 5