Supplementary Information:

Microbial and environmental processes shape the link between organic matter functional traits and composition

Ang Hu[1,2,#], Kyoung-Soon Jang[3,#], Fanfan Meng[2,4,#], James Stegen[5], Andrew J. Tanentzap[6], Mira Choi[3], Jay T. Lennon[7], Janne Soininen[8], Jianjun Wang[2,4,*]

[1] College of Resources and Environment, Hunan Agricultural University, Changsha 410128, China

[2] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academic of Sciences, Nanjing 210008, China

[3] Bio-Chemical Analysis Team, Korea Basic Science Institute, Cheongju 28119, South Korea

[4] University of Chinese Academy of Sciences, Beijing 100049, China

[5] Pacific Northwest National Laboratory, 902 Battelle Boulevard, P.O. Box 999, Richland, Washington 99352, USA

[6] Ecosystems and Global Change Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[7] Department of Biology, Indiana University, Bloomington IN 47405, USA

[8] Department of Geosciences and Geography, University of Helsinki, Helsinki FIN-00014, Finland

Number of Pages: 30

Number of Figures: 20

Number of Tables: 3

Supplementary Methods

Bacterial community and DOM analyses

Bacterial 16S rRNA genes were sequenced using MiSeq (Illumina, San Diego, USA). We amplified 16S rRNA genes in triplicate using the universal bacterial primers 515F, 5'-GTGCCAGCMGCCGCGGTAA-3' and 806R, 5'-GGACTACHVGGGTWTCTAAT-3', targeting the V4 region. Sample libraries were prepared according to the MiSeqTM Reagent Kit Preparation Guide (Illumina, San Diego, USA). We processed the sequences primarily using the QIIME pipeline (v1.9) [1]. Briefly, overlapped paired-end sequences from MiSeq were assembled using FLASH [2] and poorly overlapped and poor quality sequences were filtered out before de-multiplexing based on barcodes. Then, the sequences were clustered into OTUs at 97% pairwise identity with the seed-based uclust algorithm [3]. After chimeras were removed via Uchime, and representative sequences from each OTU were aligned to the SILVA (v128) reference database [4] using PyNAST [5]. The taxonomic identity of each representative sequence was determined using the RDP Classifier [6] and chloroplasts were removed. The bacterial sequences were rarefied to 20,000 per sample.

Highly accurate mass measurements of DOM within the sediment samples were conducted using a solariX XR 15T ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS, Bruker Daltonics, Billerica, MA). The FT-ICR MS was coupled to an electrospray ionization (ESI) interface, as demonstrated previously [7] with some modifications. DOM was solid-phase extracted for FT-ICR MS measurement [8] with some modifications. Briefly, an aliquot of 0.7 g freeze-dried sediment was sonicated with 30 ml ultrapure water for 2 h, and centrifuged at 5,000 g for 20 min. The extracted water was filtered through the 0.45 μm Millipore filter and further acidified to pH 2 using 1 M HCl. Cartridges were drained, rinsed with ultrapure water and methanol (ULC-MS grade), and conditioned with pH 2 ultrapure water. Calculated volumes of extracts were slowly passed through cartridges based on DOC concentration. Cartridges were rinsed with pH 2 ultrapure water and dried with N_2 gas. Samples were finally eluted with methanol into precombusted amber glass vials, dried with N_2 gas and stored at -20 °C until DOM analysis. The extracts were continuously injected into the standard ESI source with a flow rate of 2 μl min^{-1} and an ESI capillary voltage of 3.5 kV in negative ion mode. One hundred single scans with a transient size of 4 mega word (MW) data points, an ion accumulation time of 0.3 s, and within the mass range of m/z 150-1200, were co-added to a spectrum with absorption mode for phase correction, thereby resulting in a resolving power of 750,000 (FWHM at m/z 400). All FT-ICR mass spectra were internally calibrated using organic matter homologous series separated by 14 Da ($-CH_2$ groups). The mass measurement accuracy was typically within 1 ppm for singly charged ions across a broad m/z range (150-1,200 m/z).

Supplementary Tables

Table S1. Environmental variables and molecular traits of dissolved organic matter (DOM).

Group	Subgroup	Variable	Description
Environmental variables	Contemporary nutrient	TN.sedi	Sediment total nitrogen (TN)
	Contemporary nutrient	TP.sedi	Sediment total phosphorus (TP)
	Contemporary nutrient	NO_x.sedi	Sediment NO_x^-
	Contemporary nutrient	NO_2.sedi	Sediment NO_2^-
	Contemporary nutrient	NH_4.sedi	Sediment NH_4^+
	Contemporary nutrient	PO_4.sedi	Sediment PO_4^{3-}
	Contemporary nutrient	NO_3	Water NO_3^-
	Contemporary nutrient	NO_2	Water NO_2^-
	Contemporary nutrient	NH_4	Water NH_4^+
	Contemporary nutrient	PO_4	Water PO_4^{3-}
	Energy supply	TC.sedi	Sediment total organic carbon (TC)
	Energy supply	DOC.sedi	Sediment dissolved organic carbon (DOC)
	Energy supply	pH	Water pH
	Energy supply	Chla.sedi	Sediment chlorophyll *a* (Chl *a*)
Molecular traits	Molecular weight	Mass	The mass to charge ratio (m/z)
	Molecular weight	C	The number of carbon
	Molecular weight	$kdefect_{CH2}$	Kendrick Defect
	Stoichiometry	O/C	O/C ratio
	Stoichiometry	H/C	H/C ratio
	Stoichiometry	N/C	N/C ratio
	Stoichiometry	P/C	P/C ratio
	Stoichiometry	N/P	N/P ratio
	Stoichiometry	S/C	S/C ratio
	Chemical structure	AI_{Mod}	The modified aromaticity index
	Chemical structure	DBE	Double bond equivalence
	Chemical structure	DBE_O	Double bond equivalence minus oxygen
	Chemical structure	DBE_{AI}	Double bond equivalence minus aromaticity index
	Oxidation state	GFE	Gibbs free energy
	Oxidation state	NOSC	Nominal oxidation state of carbon
	Carbon use efficiency	Y_{met}	Carbon use efficiency

Table S2. Relationships between energy supply and global change drivers that were modeled using multiple ordinary least squares regression. We considered four energy supply variables, namely water pH, sediment chlorophyll *a* (Chl *a*), total organic carbon (TOC) and dissolved organic carbon (DOC). Global change drivers include water temperature (Temp), nutrient enrichment ($ADD.NO_3$) and their interactions ($Temp:ADD.NO_3$).

Response variables	Country	Model R^2	AIC	Explanatory variables and β-weights		
				Temp	$ADD.NO_3$	$Temp:ADD.NO_3$
Water pH	China	0.60	-132.6	[a] 0.65^{***}	0.44^{***}	0.13^{*}
	Norway	0.60	-134.5	0.68^{***}	0.24^{***}	0.30^{***}
Sediment Chl *a*	China	0.25	-38.2	0.43^{***}	0.28^{***}	0.11
	Norway	0.46	-89.5	0.57^{***}	0.25^{***}	0.30^{***}
Sediment TC	China	0.09	-11.6	0.3^{***}	0	0
	Norway	0.2	-32.2	0.17^{*}	0.44^{***}	0
Sediment DOC	China	0.39	-71.1	0.05	-0.61^{***}	0.18^{**}
	Norway	0.39	-73.7	0	-0.63^{***}	0

Note: The best models were identified using Akaike's information criterion (AIC). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. [a] Standardized partial regression coefficients.

Table S3. Peak counts for four fractions of dissolved organic matter (DOM) in China or Norway. In total, 19,538 molecules were parsed into four fractions based on molecular reactivity and activity: labile-active, recalcitrant-active, recalcitrant-inactive, and labile-inactive molecules. 16,101 and 9,449 molecules were the total molecules in China and Norway, respectively.

DOM partitions	H/C ratio	Transformations	China		Norway	
			Counts	Percentage	Counts	Percentage
Labile-active	≥ 1.5	> 10	721	4.48%	610	6.46%
Recalcitrant-active	< 1.5	> 10	1,346	8.36%	1,166	12.34%
Recalcitrant-inactive	< 1.5	≤ 1	2,524	15.68%	895	9.47%
Labile-inactive	≥ 1.5	≤ 1	1,167	7.25%	511	5.41%
Active		> 10	2,067	12.84%	1,776	18.80%
Inactive		≤ 1	3,691	22.92%	1,406	14.88%
Other		2-10	5,758	64.24%	3,182	66.32%
Total			16,101	100%	9,449	100%

Supplementary Figures

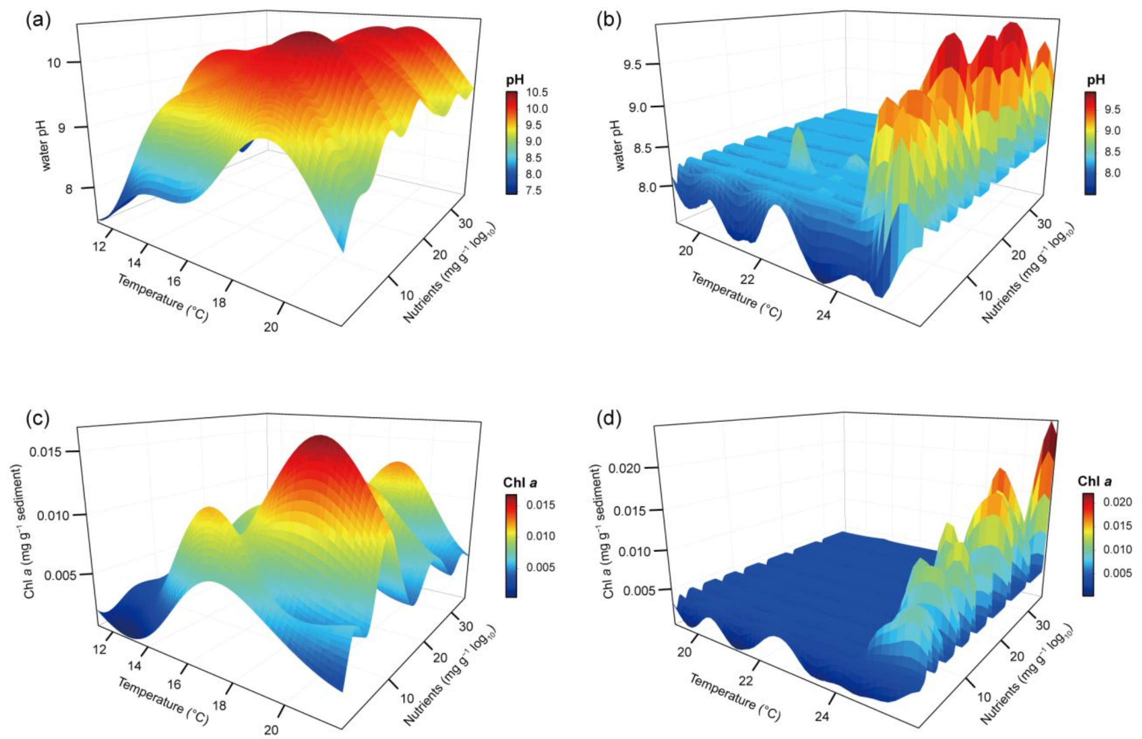


Figure S1. The relationships between energy supply such as water pH (a-b) and sediment chlorophyll *a* (Chl *a*, c-d) and two global change drivers including water temperature and nutrients in China (a, c) and Norway (b, d). Water pH showed strong correlations with sediment Chl *a* at most elevations, and was thus used to represent energy supply (Wang *et al* 2016).

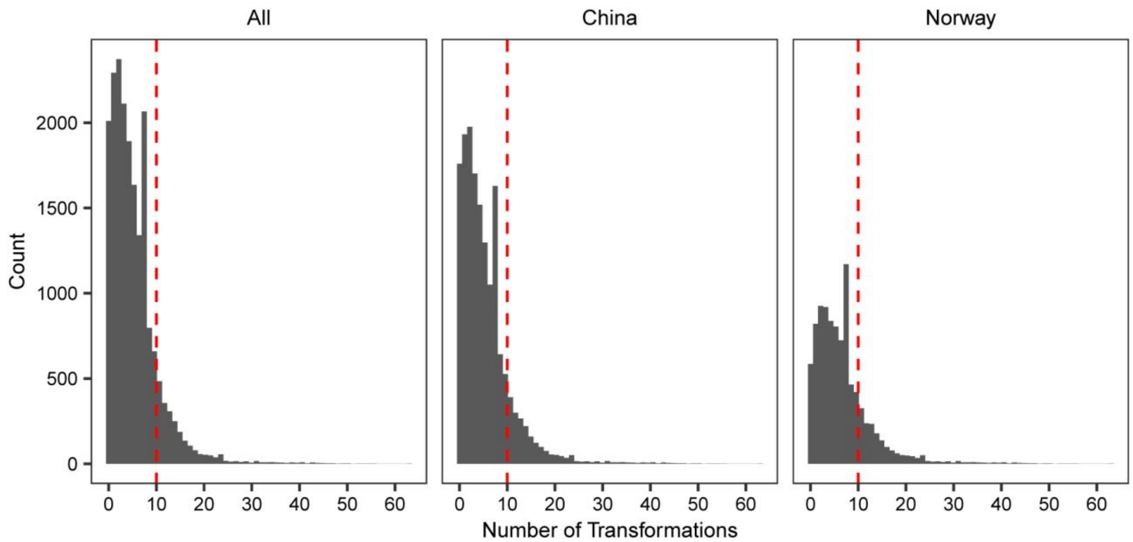


Figure S2. Histogram demonstrating the counts of FT-ICR MS identified peaks with a given number of associated biochemical transformations. The dotted vertical lines indicate the location of the 10 transformation number.

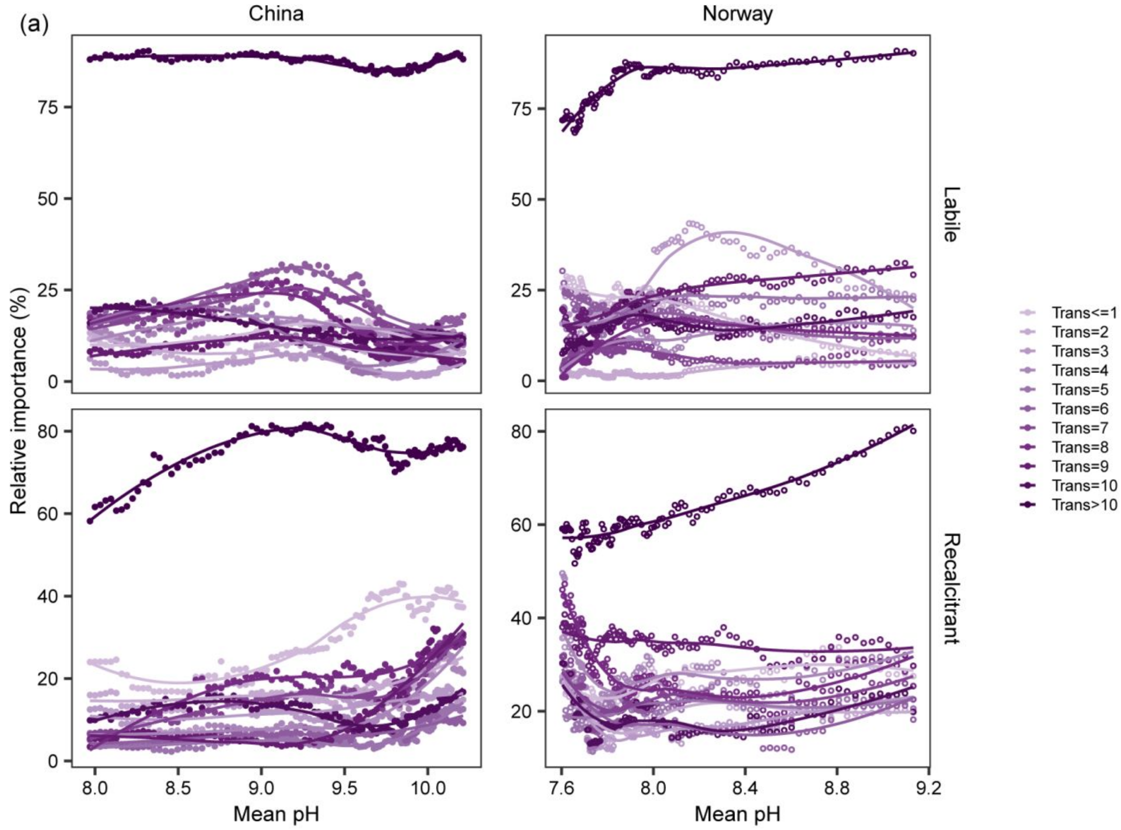
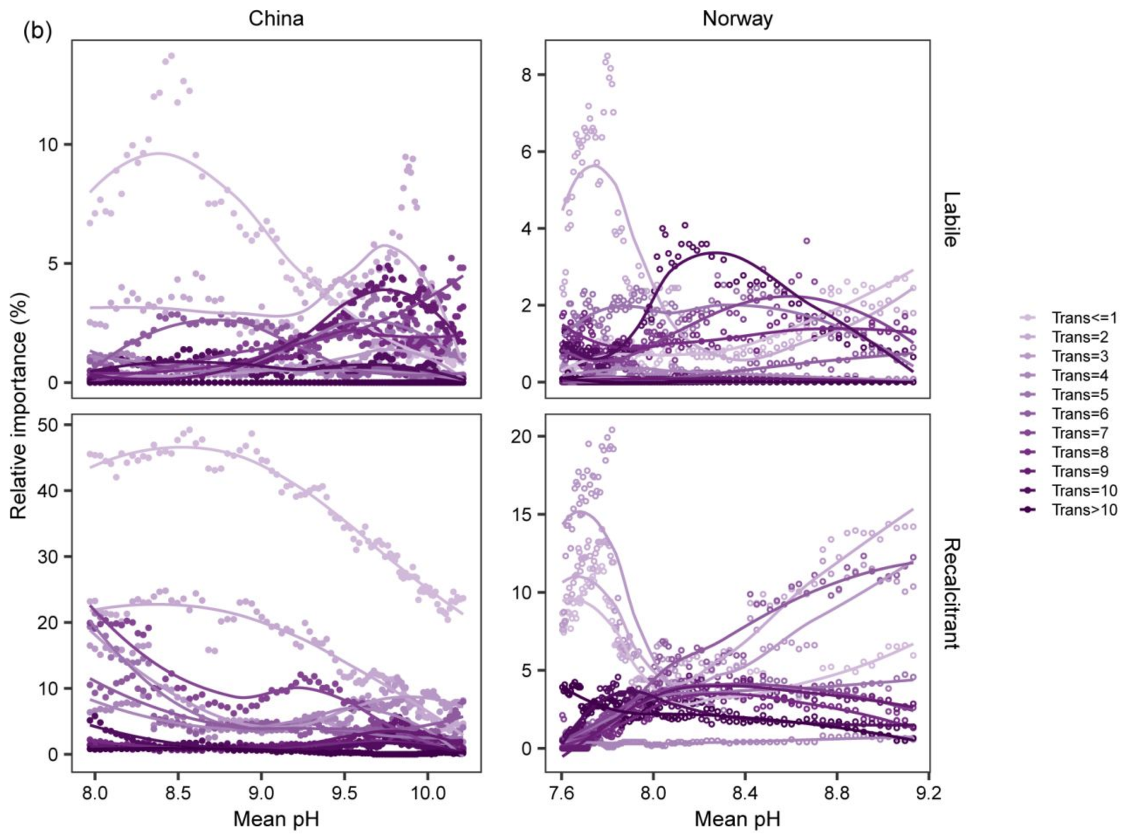


Figure S3. The effects of energy supply (i.e., water pH) on the relative importance of ecological processes underlying molecule assemblages of dissolved organic matter (DOM) with different transformation groups in China and Norway. We plotted the relative importance of variable selection (a), homogeneous selection (b) and stochastic processes (c) against the energy supply (i.e., water pH) gradient using a moving-window approach, and their relationships are visualized with loess regression models. Ten transformation groups are transformations ≤ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and > 10.



108

109 **Figure S3**. Continued.

110

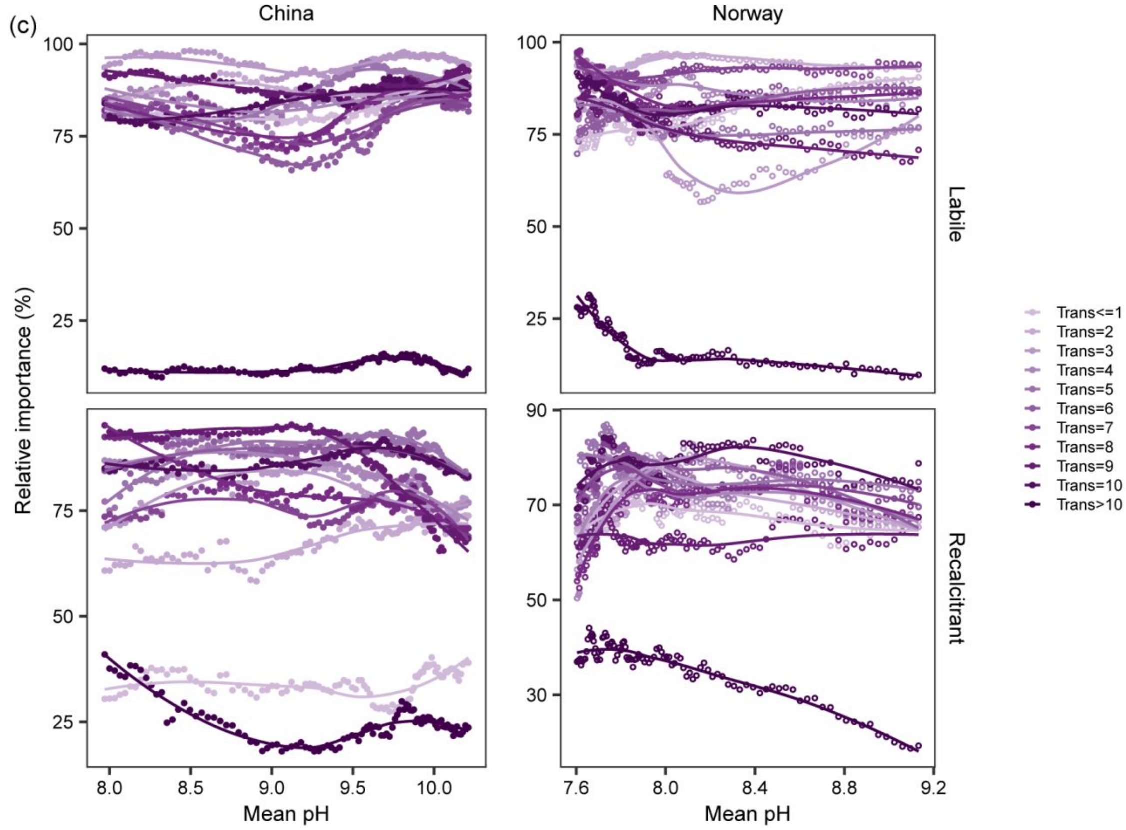


Figure S3. Continued.

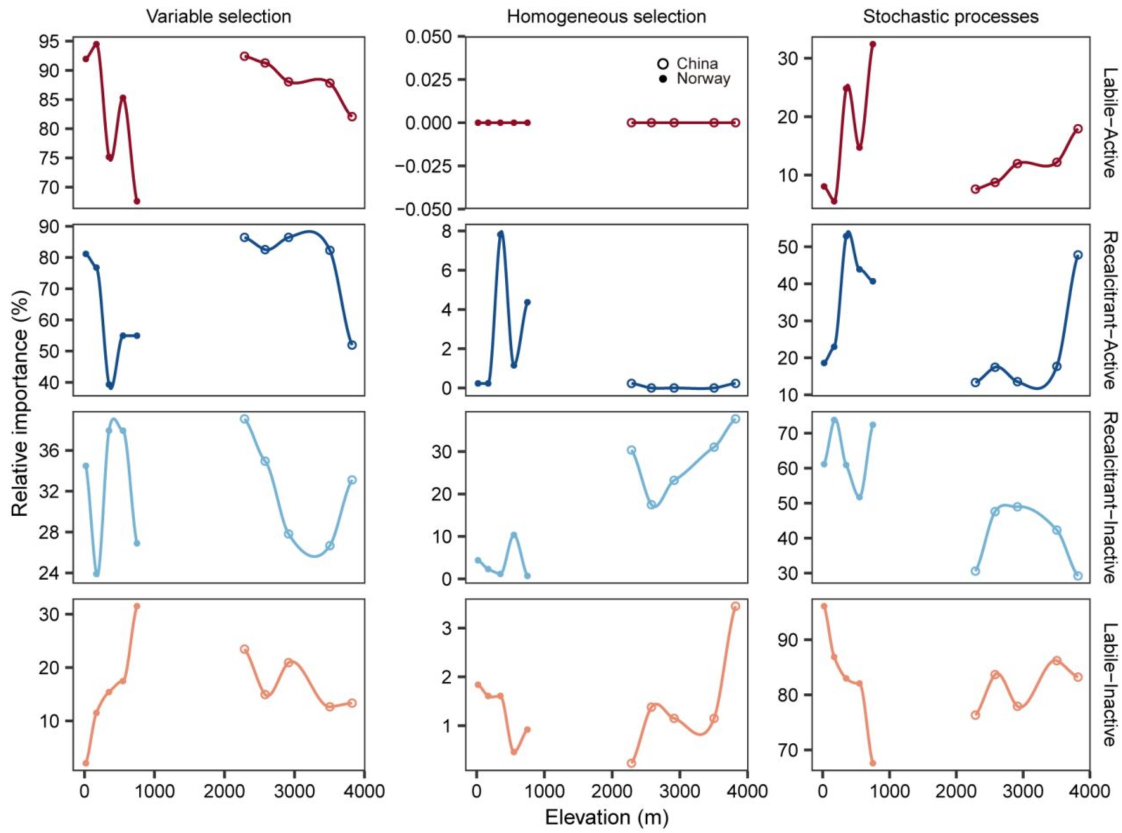


Figure S4. The effects of elevation on the relative importance of ecological processes underlying molecule assemblages of four fractions of dissolved organic matter (DOM) in China (hollow points) and Norway (solid points). We plotted the relative importance of variable selection (left panels), homogeneous selection (middle panels) and stochastic processes (right panels) against elevational gradient, and their relationships are visualized with loess regression models. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

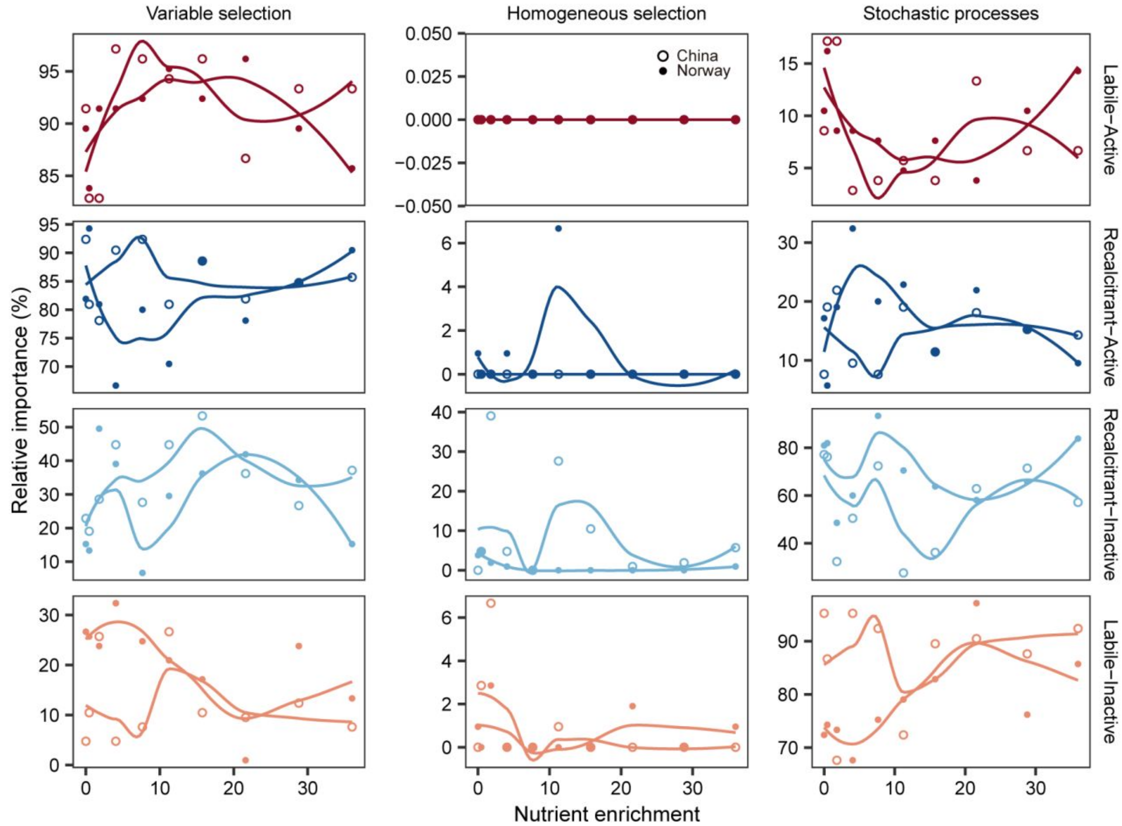


Figure S5. The effects of nutrient enrichment on the relative importance of ecological processes underlying molecule assemblages of four fractions of dissolved organic matter (DOM) in China (hollow points) and Norway (solid points). We plotted the relative importance of variable selection (left panels), homogeneous selection (middle panels) and stochastic processes (right panels) against nutrient gradient, and their relationships are visualized with loess regression models. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

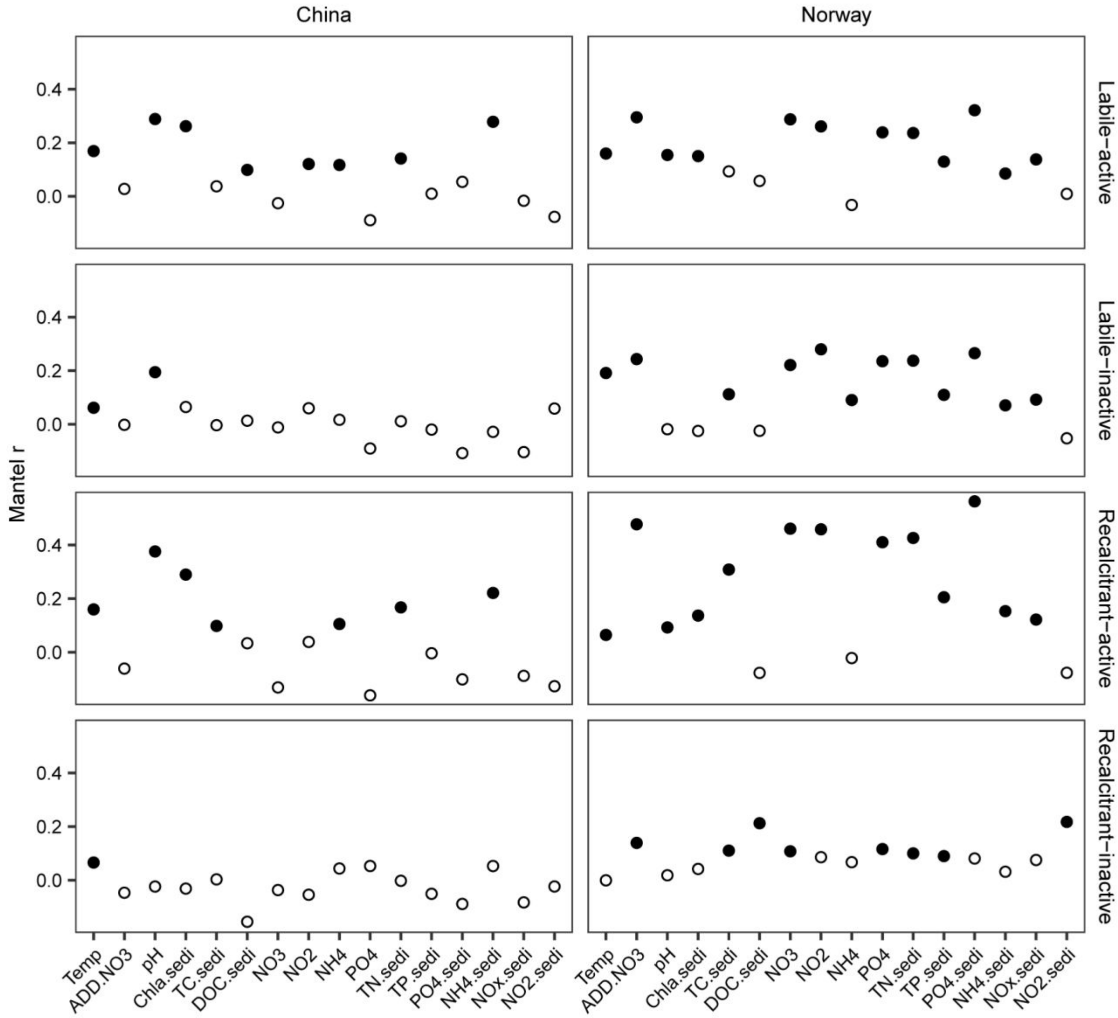


Figure S6. The relative influence of explanatory variables on β-nearest taxon index (βNTI) across four fractions of dissolved organic matter (DOM) using Mantel test. Each circle is the values of Mantel r for individual explanatory variable in China and Norway, respectively. Solid and open circles indicate the significant ($P \leq 0.05$) and non-significant ($P > 0.05$) Mantel r, respectively. The details of abbreviations of explanatory variables are available in Table S1. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

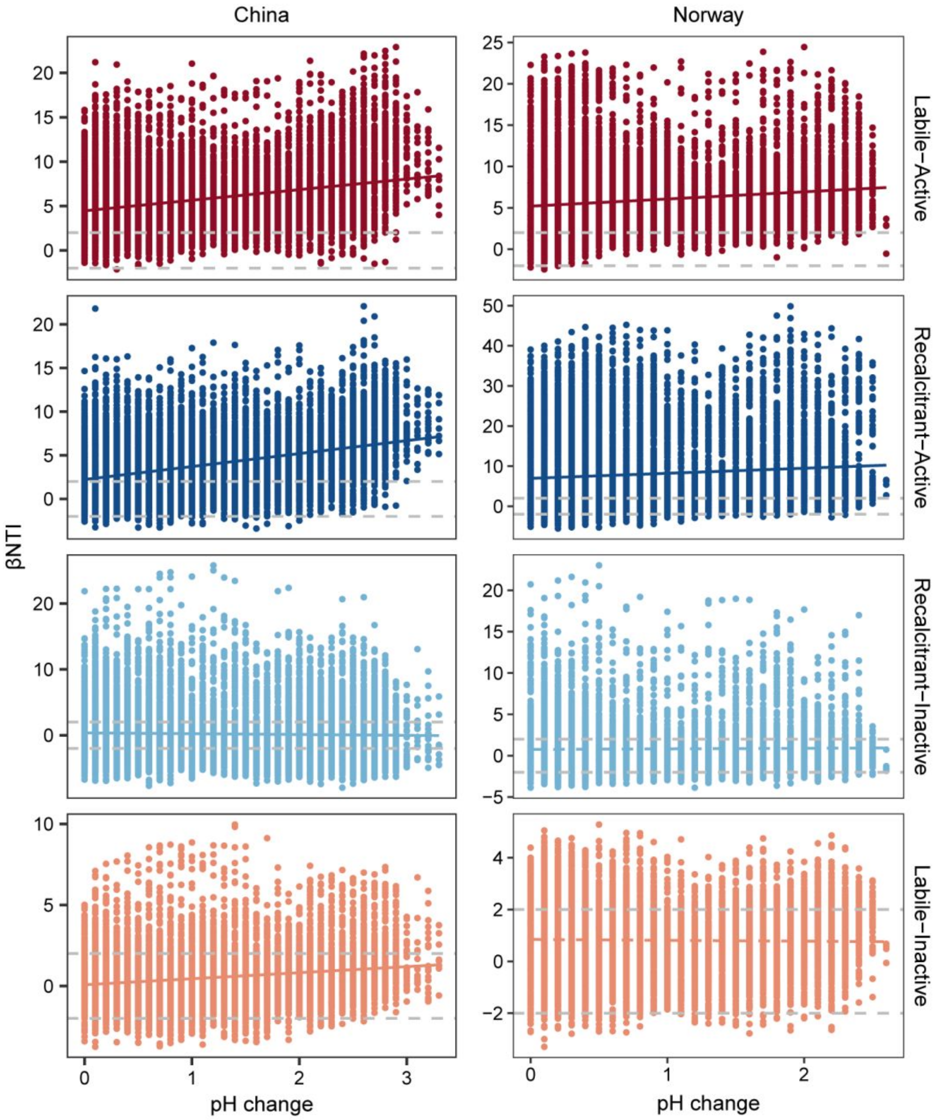


Figure S7. β-nearest taxon index (βNTI) for all pairwise assemblage comparisons of four fractions of dissolved organic matter (DOM) as a function of the change in water pH between samples in China and Norway. Horizontal dashed lines indicate the upper (+2) and lower (-2) significance thresholds. The relationships between βNTI and changes in pH are indicated by solid ($P \leq 0.05$) and dotted ($P > 0.05$) lines estimated using linear models. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

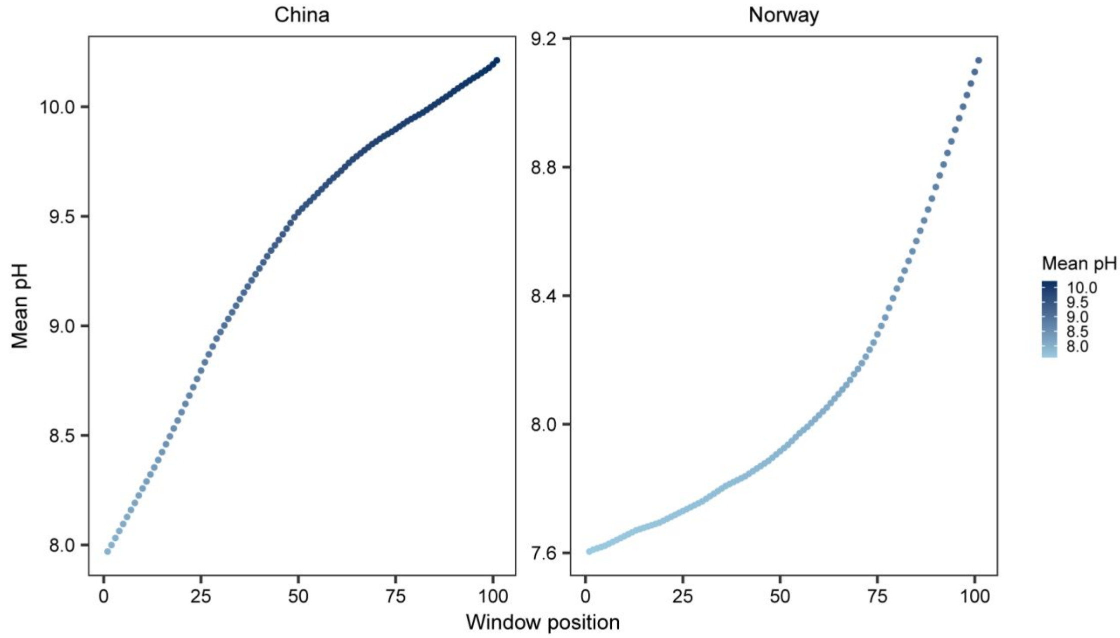


Figure S8. Mean water pH for each pH-based window. We sorted the samples along the water pH gradient, from minimum to maximum water pH, separately for China and Norway, resulting in 101 windows with a fixed size of 50 samples.

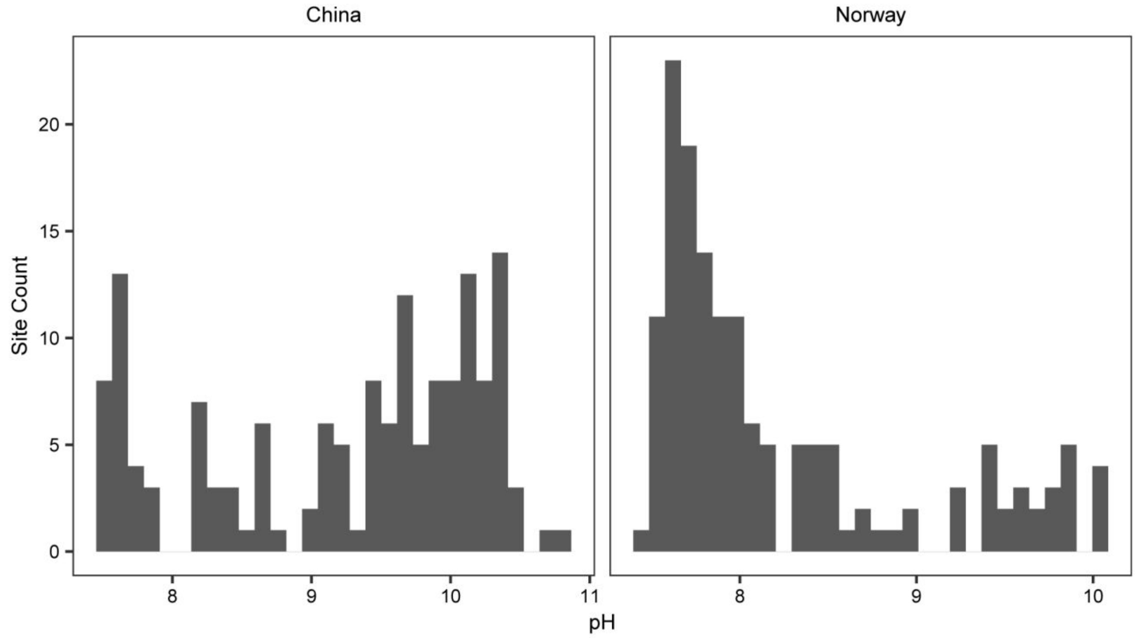


Figure S9. Histogram demonstrating the counts of samples in China or Norway with a given water pH.

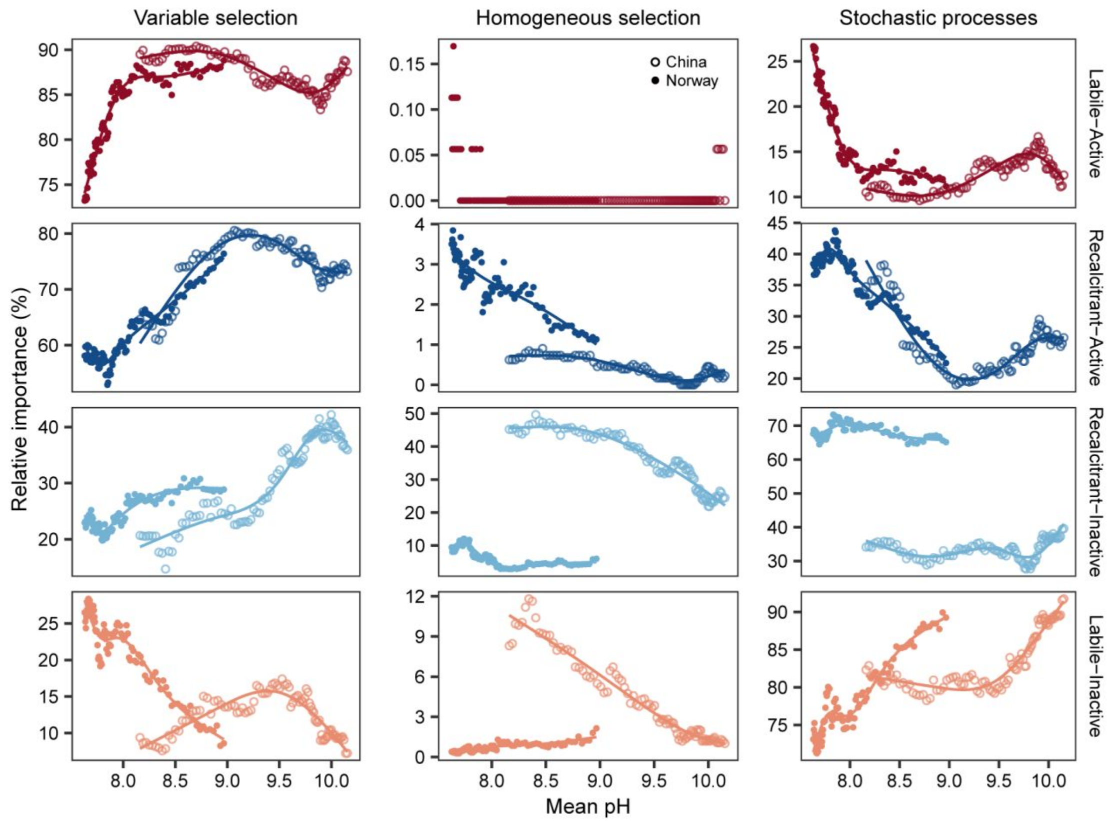


Figure S10. The effects of energy supply (i.e., water pH) on the relative importance of ecological processes underlying molecule assemblages of four fractions of dissolved organic matter (DOM) in China (hollow points) and Norway (solid points). We plotted the relative importance of variable selection (left panels), homogeneous selection (middle panels) and stochastic processes (right panels) against the energy supply (i.e., water pH) gradient using a moving-window approach with a fixed size of 60 samples, and their relationships are visualized with generalized additive models with k of 5. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

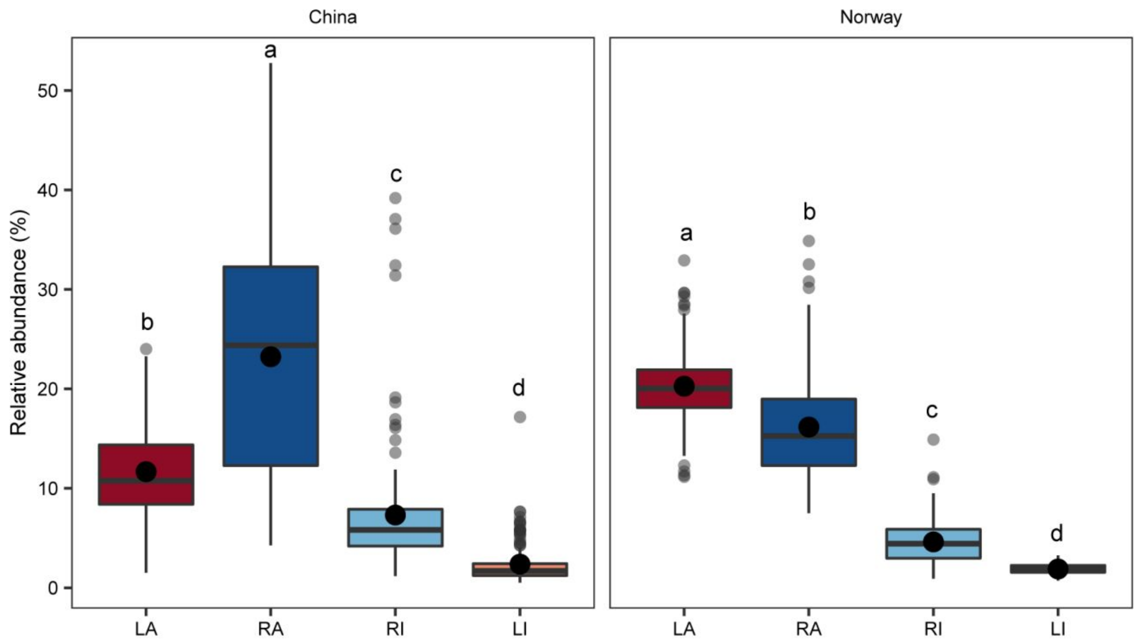


Figure S11. Box plots of the relative abundance of molecular formulae of four fractions of dissolved organic matter (DOM) in China and Norway. Large dots are the mean values of relative abundance for each fraction in China or Norway. Different letters (a, b, c, d) indicate significant differences between four fractions ($P \leq 0.05$) by Wilcoxon test. Four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI) and labile-inactive (LI) based on two dimensions of molecular reactivity and activity.

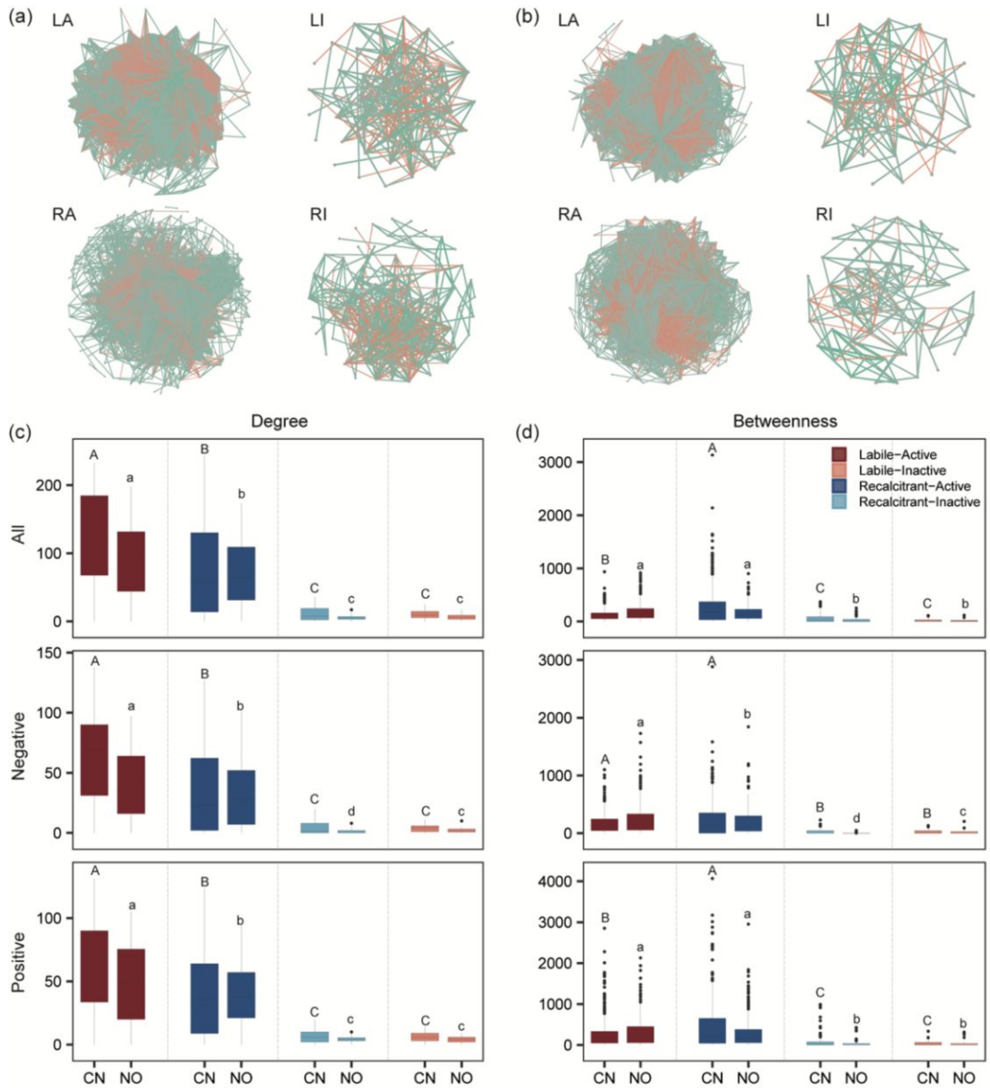


Figure S12. Networks of co-occurring molecules based on SparCC correlation analysis for four fractions of dissolved organic matter (DOM) in China (a) and Norway (b). Blue and red lines indicate positive and negative interactions, respectively. (c-d) Network properties such as degree (c) and betweenness centrality (d) for all, negative and positive interactions. Betweenness centrality measures the extent to which a node lies on paths between other nodes. Molecules with a higher betweenness centrality communicate more with other molecules within an assemblage. Different letters (a, b, c, d) indicate significant differences between four fractions ($P \leq 0.05$) by Wilcoxon test. Four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI) and labile-inactive (LI) based on two dimensions of molecular reactivity and activity. CN: China. NO: Norway.

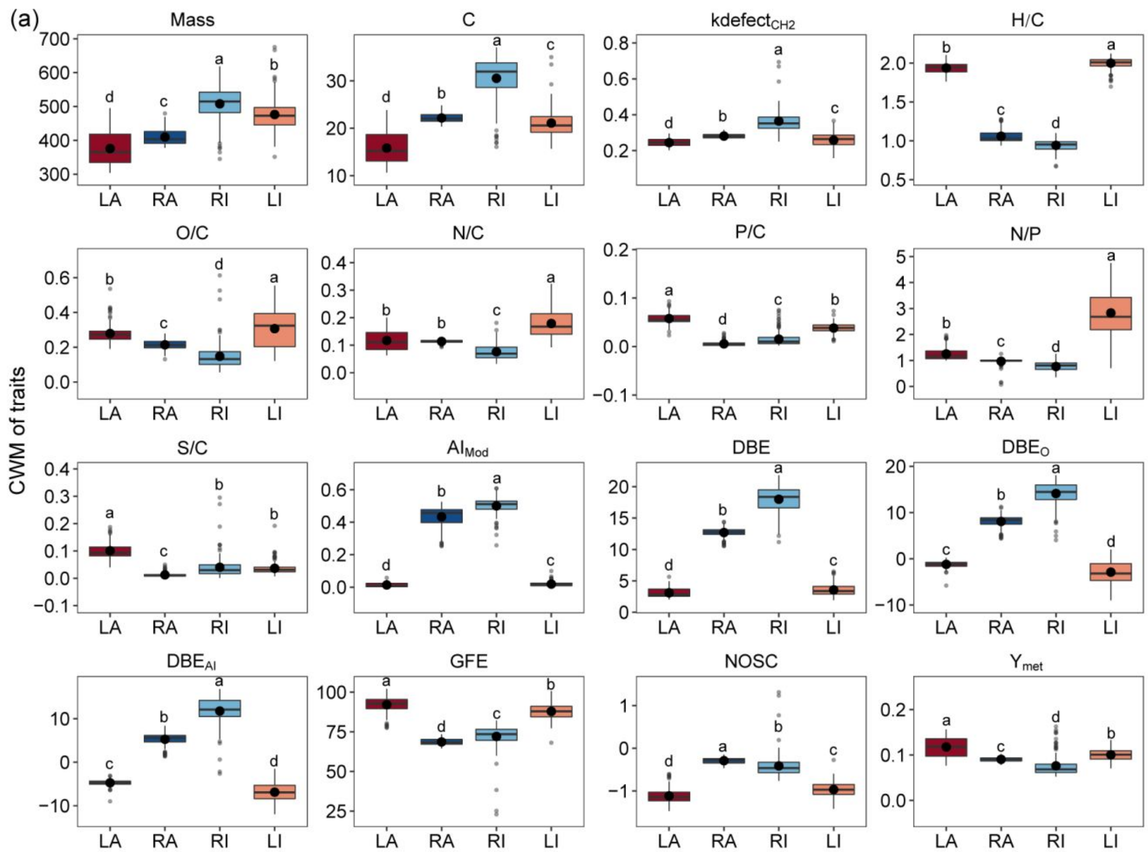


Figure S13. Box plots of the molecular traits of four fractions of dissolved organic matter (DOM) in China and Norway. Large dots are the mean values of weighted means (WM) of molecular traits for each fraction in China (a) or Norway (b). Different letters (a, b, c, d) indicate significant differences between four fractions ($P \leq 0.05$) by Wilcoxon test. Four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI) and labile-inactive (LI) based on two dimensions of molecular reactivity and activity. The details of abbreviations of molecular traits are available in Table S1.

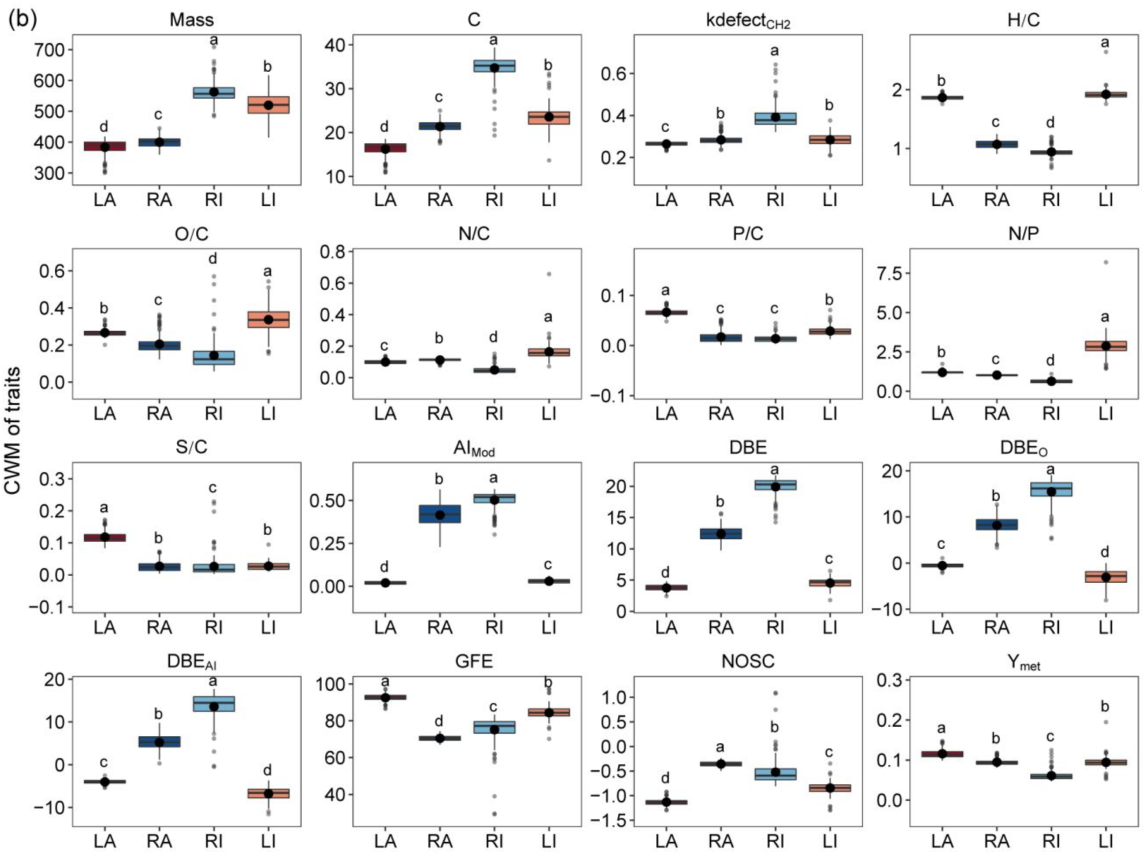


Figure S13. Continued.

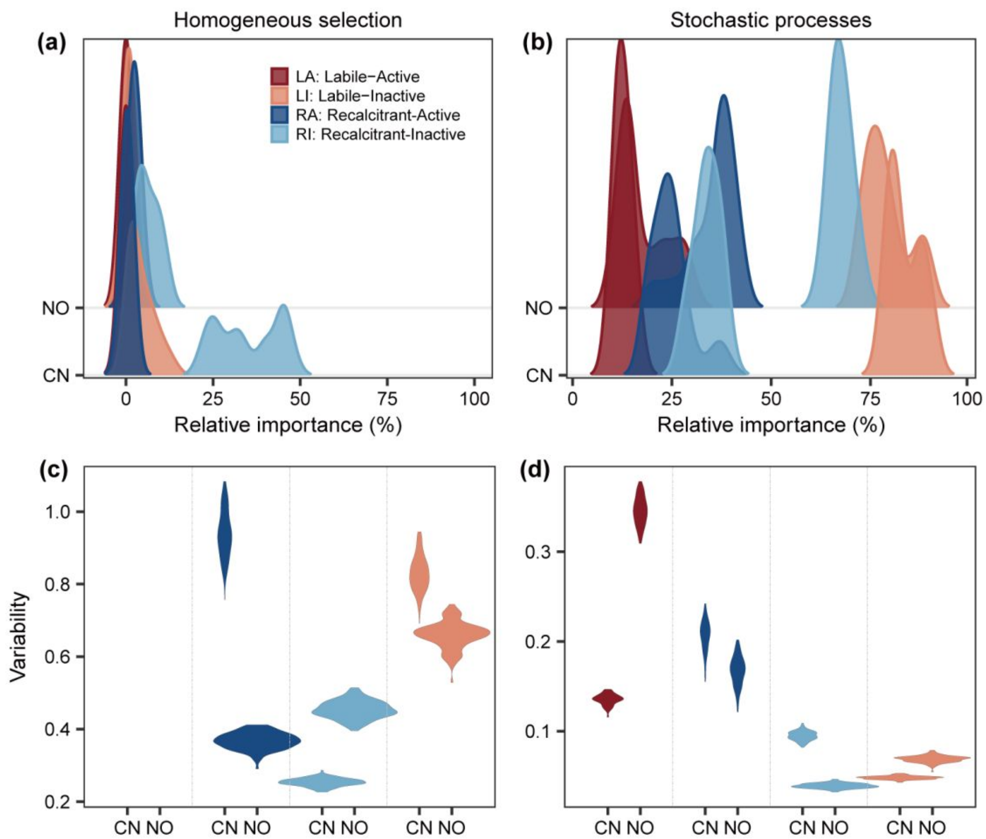


Figure S14. **Ecological processes structuring dissolved organic matter (DOM) fractions under global change.** (a-b) The distribution of relative importance of ecological processes across the pH gradients for four DOM fractions in China (CN) and Norway (NO). We considered three ecological processes, that is variable selection (shown in the main text), homogeneous selection (a) and stochastic processes (b). The four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI), and labile-inactive (LI) based on two dimensions of molecular reactivity and activity. (c-d) Violin plots of variability of relative importance of homogeneous selection (c) and stochastic processes (d) across the pH gradients for four DOM fractions in China and Norway. Variability was calculated as the ratio of the standardized deviation to mean of the relative importance of each ecological process calculated across the pH gradients.

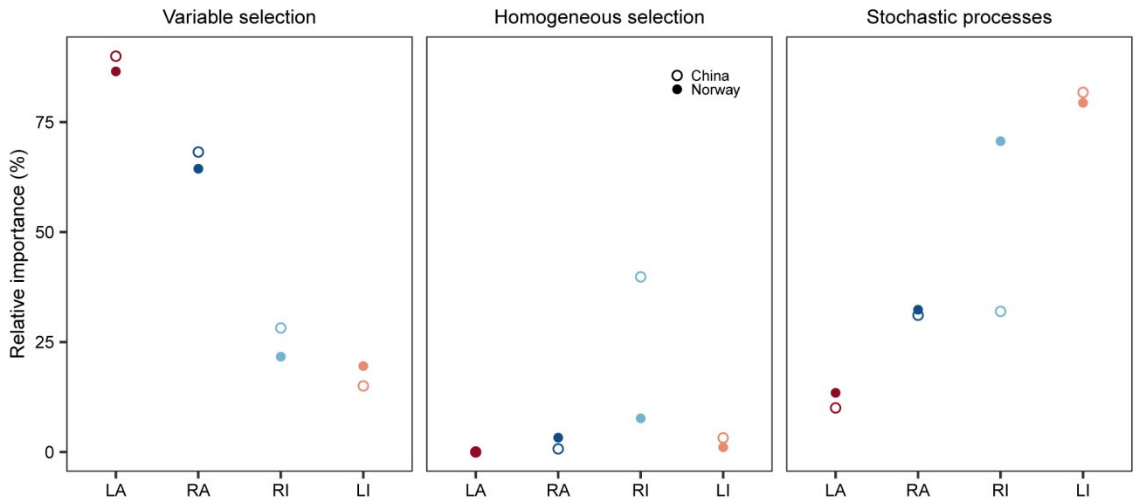


Figure S15. Relative importance of ecological processes underlying molecule assemblages of four fractions of dissolved organic matter (DOM) in China (hollow points) and Norway (solid points). We considered three ecological processes including variable selection (left panels), homogeneous selection (middle panels) and stochastic processes (right panels). Four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI) and labile-inactive (LI) based on two dimensions of molecular reactivity and activity.

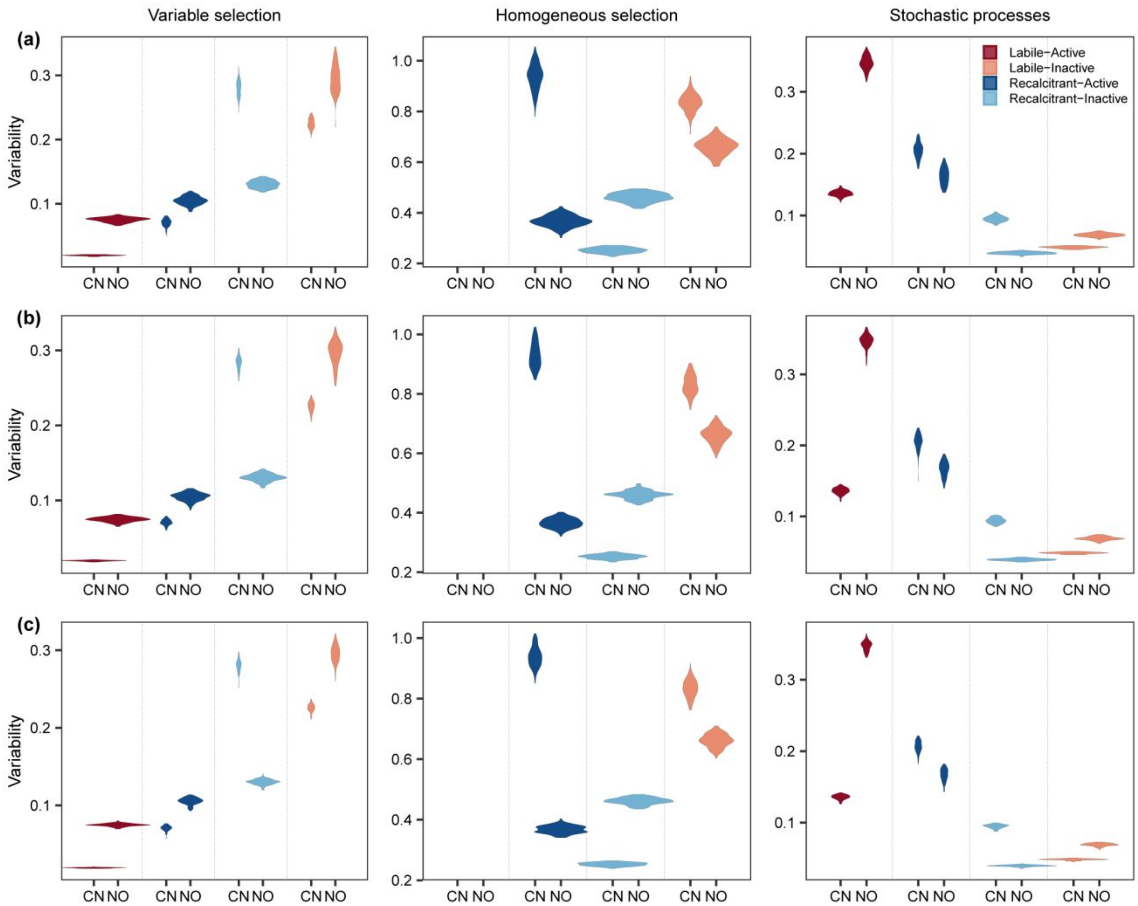


Figure S16. Boxplots of variability of the relative importance of variable selection across the energy supply gradients for four fractions of dissolved organic matter (DOM) in China (CN) and Norway (NO). The variability was calculated as the ratio of standardized deviation and mean of the relative importance of variable selection by randomly using 60 (a), 70 (b) or 80 (c) windows (100 bootstraps) along the energy supply gradient for each DOM fraction in China or Norway. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

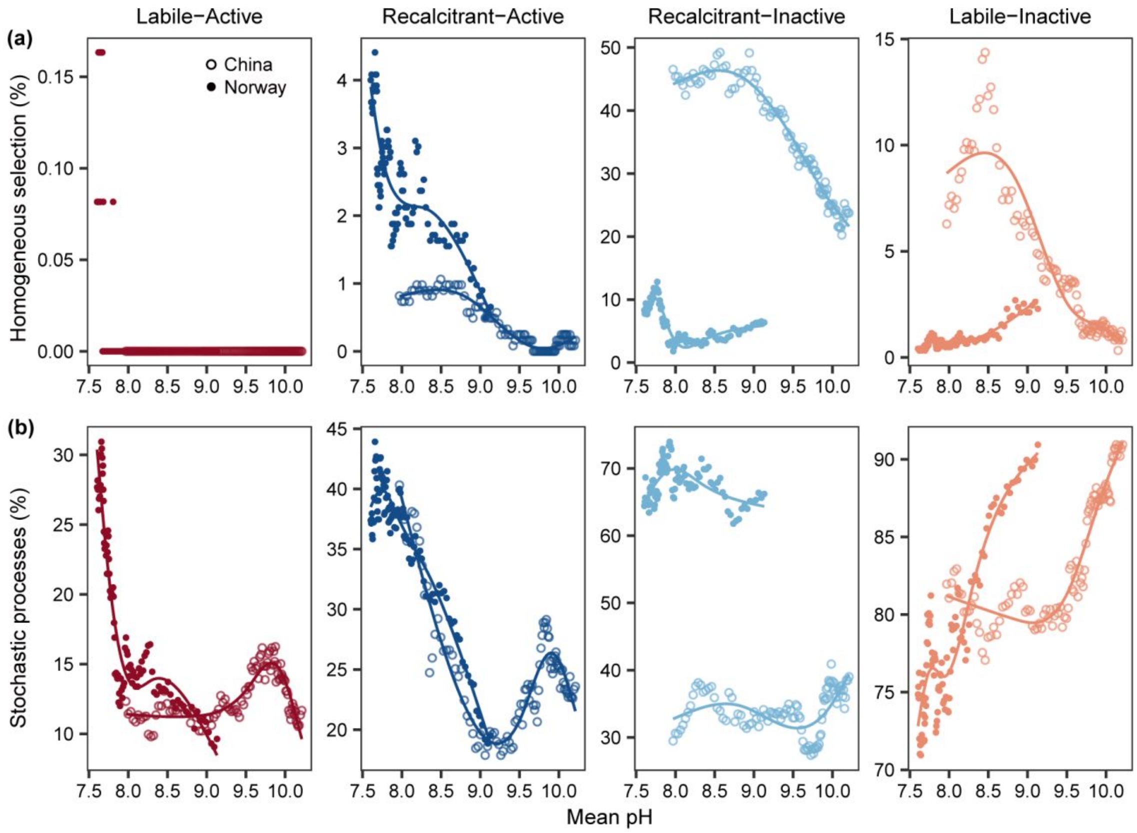


Figure S17. Effects of energy supply on relative importance of ecological processes structuring dissolved organic matter (DOM) fractions. We plotted the relative importance of homogeneous selection (a) and stochastic processes (b) against pH for four DOM fractions in China (hollow points) and Norway (solid points). The other ecological processes, that is variable selection, were shown in Fig. 4. Relationships are visualized with generalized additive models with k of 5. The four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive molecules and were based on two dimensions of molecular reactivity and activity.

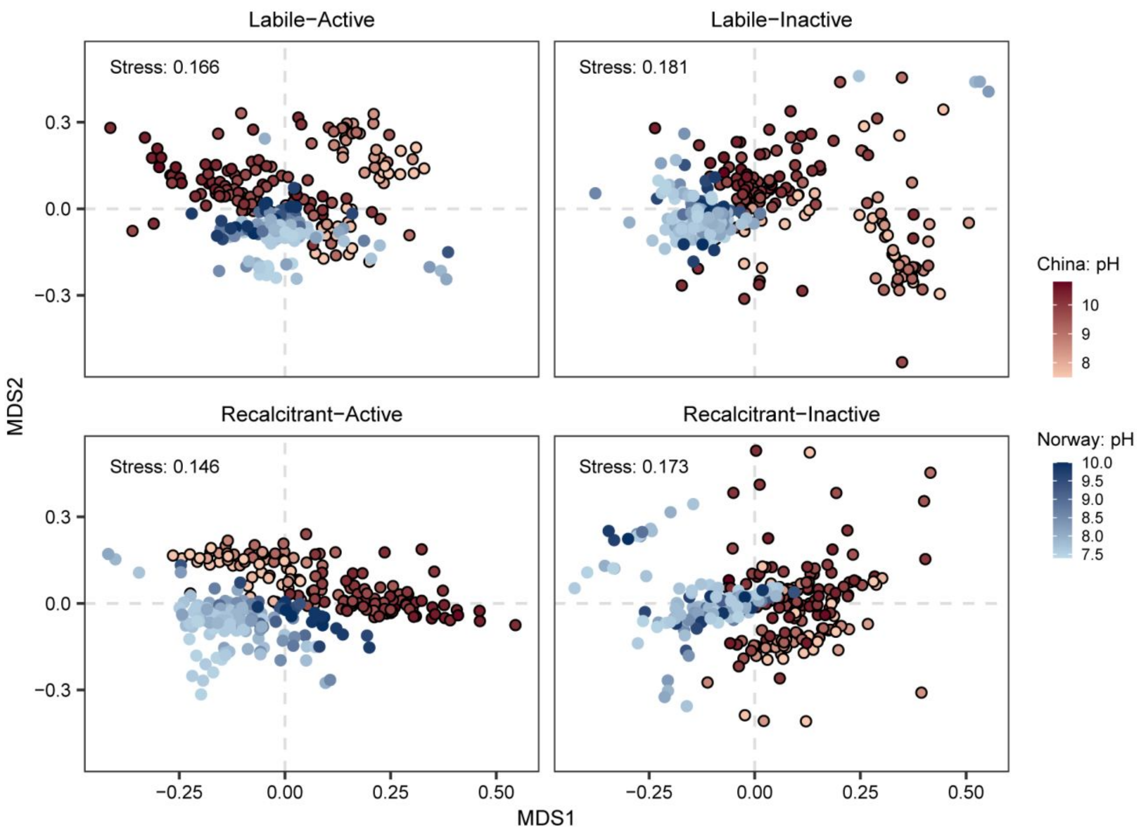


Figure S18. Non-metric multidimensional scaling (NMDS) of molecular compositions of four fractions of dissolved organic matter (DOM) across energy supply (i.e., water pH) gradients in China and Norway. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity.

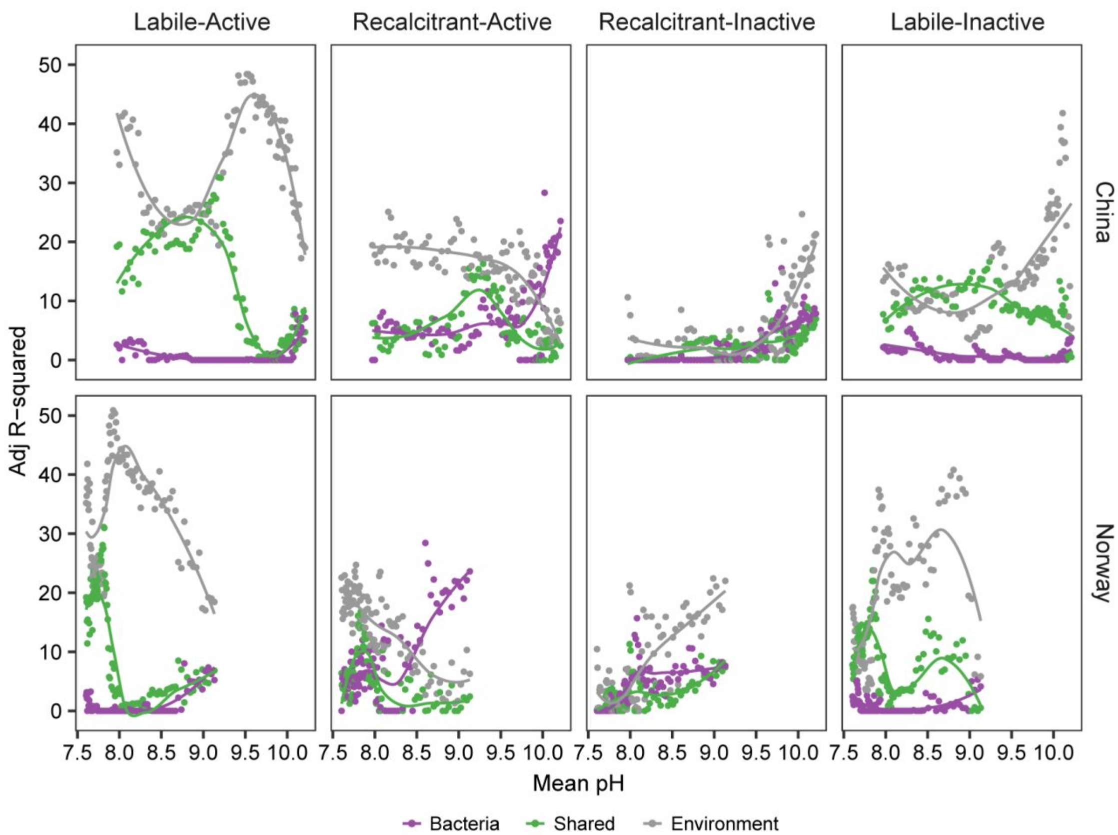


Figure S19. The effects of energy supply (i.e., water pH) on the pure effects of bacterial communities and environmental conditions, and their shared effects across four fractions of dissolved organic matter (DOM) in China and Norway. These effects of explanatory variables on the four DOM fractions were determined by variation partitioning analysis. We plotted these effects against the energy supply (i.e., water pH) gradient across four fractions in China (Upper panel) and Norway (lower pannel), and their relationships are visualized with loess regression models. The adjusted R-squared indicates the explained variations by the explanatory variables of bacterial communities and/or environments. Four DOM fractions are labile-active (LA), recalcitrant-active (RA), recalcitrant-inactive (RI) and labile-inactive (LI) based on two dimensions of molecular reactivity and activity.

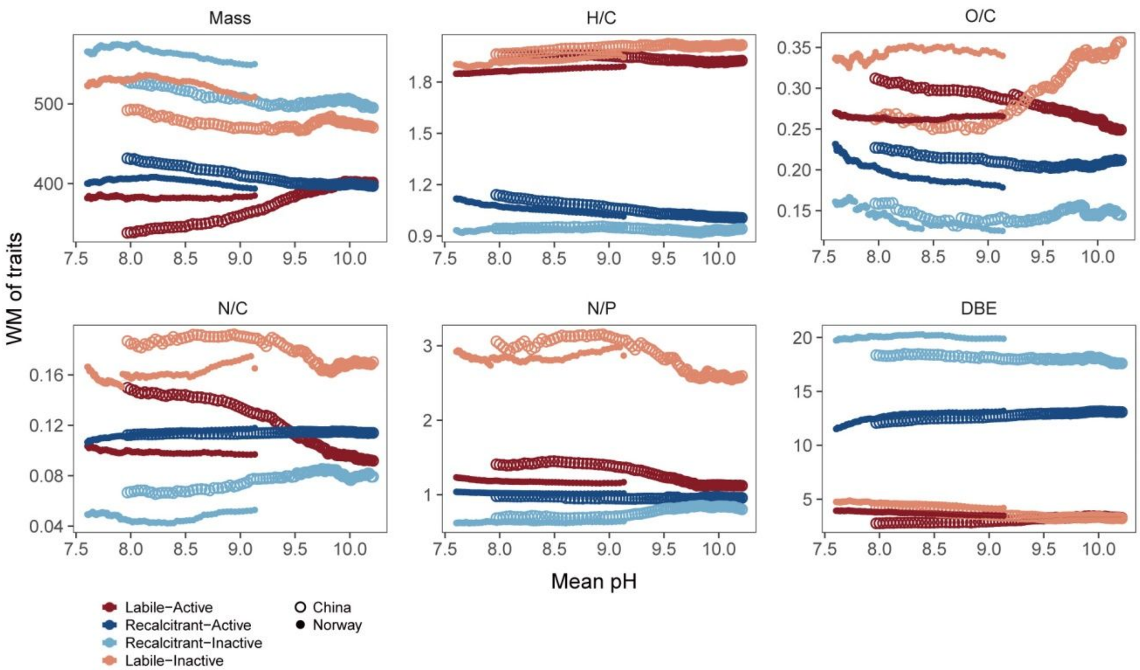


Figure S20. The effects of energy supply (i.e., water pH) on molecular traits of four fractions of dissolved organic matter (DOM) in China (hollow points) and Norway (solid points). The relationships between weighted mean (WM) of DOM traits and energy supply are visualized with loess regression models. Four DOM fractions are labile-active, recalcitrant-active, recalcitrant-inactive and labile-inactive based on two dimensions of molecular reactivity and activity. The details of abbreviations of molecular traits are available in Table S1.

References

1. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010,** *7*, (5), 335-336.
2. Magoč, T.; Salzberg, S. L., FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* **2011,** *27*, (21), 2957-2963.
3. Edgar, R. C., UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **2013,** *10*, (10), 996-998.
4. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **2013,** *41*, (Database issue), D590-6.
5. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010,** *26*, (2), 266-267.
6. Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R., Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **2007,** *73*, (16), 5261-5267.
7. Choi, J. H.; Jang, E.; Yoon, Y. J.; Park, J. Y.; Kim, T. W.; Becagli, S.; Caiazzo, L.; Cappelletti, D.; Krejci, R.; Eleftheriadis, K.; Park, K. T.; Jang, K. S., Influence of Biogenic Organics on the Chemical Composition of Arctic Aerosols. *Glob Biogeochem Cycle* **2019,** *33*, (10), 1238-50.
8. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G., A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Meth.* **2008,** *6*, (6), 230-235.