Supporting Information

Toward Quantitative Understanding of the Bioavailability of Dissolved Organic Matter in Freshwater Lake during Cyanobacteria Blooming

Leilei Bai[†,‡], Chicheng Cao[§], Changhui Wang[†], Huacheng Xu[†], Hui Zhang[§], Vera I Slaveykova[‖], Helong Jiang*[,†]

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

[‡]Graduate University of Chinese Academy of Sciences, Beijing 100049, China

[§]Key Laboratory of Environmental Medicine Engineering of Ministry of Education, School of Public Health, Southeast University, Nanjing, 210009, China

[‖]Department F.-A. Forel for Environmental and Aquatic Sciences, Faculty of Sciences, University of Geneva, Geneva CH-1211, Switzerland.

*Phone/fax: +86 25 8688 2208; e-mail: hljiang@niglas.ac.cn.

Number of pages: 37

Number of tables: 5

Number of figures: 15

22 **Contents**

23

Materials and methods

Bioreactor Set-up and Operation

The laboratory 4-stage plug-flow bioreactor was built following the recommendations of Søndergaard and Worm[1] for the measurement of bioavailable dissolved organic carbon (BDOC) in lake water (Figure S2). Similar bioreactors have been used for BDOC investigation in drinking water, streamwater, and seawater.[2–5]

The bioreactor consisted of 4 darkened glass reactors with various size (Φ3.78×40, 7.32×1.68, 8.01×2.02, and 10.14×4.03 cm, respectively). These reactors were filled with porous glass beads that had a large ratio of surface area to volume. The reactors and glass beads were both precombusted (450 °C, 4 h) to remove organic matter. Then the 4 reactors were linked with sterilized tubes, and peristaltic pumps were also connected to provide a constant flow rate of 0.5 mL min^{-1}. The hydraulic retention times of the 4 reactors were adjusted to 4, 20, 24, and 48 h, respectively. Additionally, an air pump was used to give sufficient sterilized air (filtered by 0.22 μm) for the growth of microorganisms.

To enrich the microorganism community adhered on carriers, the bioreactor was initially incubated with Taihu lake water (filtered by 2.7 μm precombusted Whatman GF/D glass fiber membrane) that contained considerable number of local bacteria. DOC removal in the cultivation was shown in Figure S3. The biodegradation capacity of bioreactor increased with time at the initial cultivation, while the decrease of DOM gradually became stable and effective after 8 months. This incubation time was longer than previous researches,[2,3,5] mainly due to the higher DOC concentrations in CyanoHAB lake water than in drinking water, streamwater and seawater. Moreover, the longer flowing distance of the 4-stage plug-flow bioreactor as compared with the

single-stage bioreactor also resulted in the longer cultivation time. Triplicate samples were collected from the effluent of each reactor (Eff-1, Eff-2, Eff-3, and Eff-4, respectively) for DOM analysis. The interval between each sampling was 4 d.

Batch Bioassays

To verify system-specificity and biodegradation capacity of the 4-stage plug-flow bioreactor, batch bioassays were conducted as follows: triplicates of influent, Eff-1, Eff-2, Eff-3, and Eff-4 samples were inoculated with 5% volume of a local bacteria inoculum. Water for the inoculum was collected at the same site in Meiliang Bay and filtered through 1.2 μm precombusted Whatman GF/D glass fiber membrane to remove grazers, phytoplankton, zooplankton and detritus. Once again, the filtrate was passed through 0.2 μm (PTFE) Teflon filters to retain bacterial cells. Subsequently, the Teflon filters was washed in sterile ASM-1/10 medium, which was then used as the inoculum.[6] Glass bioassays bottles were kept in the dark at 25 ºC and sampled on the 0 (d_0), 7th (d_7), and 56th (d_{56}) day, respectively, for DOC measurement.

Chromophoric DOM Measurement

Absorption spectra were obtained between 200 and 800 nm at 1 nm intervals, using a UV–Vis spectrophotometer (Shimadzu, UV-2550PC) with matching 50 mm quartz cells. The slit width was 1 nm, and the wavelength scan rate was 210 nm/min. Milli-Q water was used in the reference cell. Absorbance measurements at each wavelength (λ) were baseline corrected by subtracting the absorbance at 700 nm. Naperian absorption coefficients (a_λ) was calculated by multiplying the corrected absorbance reading by 2.303/r, where r is the cuvette path length in meters. Concentration

of chromophoric DOM is expressed as absorption coefficient (m^{-1}) at 254 nm (a_{254}). Spectral slope ratio (S_R) was defined as the ratio of the spectral slopes (nm^{-1}) between a_{275} and a_{295} ($S_{275–295}$) and between a_{350} and a_{400} ($S_{350–400}$).[7] Specific UV absorbance ($SUVA_{254}$), a proxy for aromaticity, was calculated by dividing the decadic absorption coefficient 254 by DOC in mg C L^{-1}.[8] The first and second derivative absorption spectra were informative of DOM optical properties and were determined by linear regression over sliding 21 nm intervals.[9]

Fluorescent DOM Measurement

Emission-excitation matrix (EEM) spectra of water samples were measured using a Hitachi F-7000 fluorescence spectrometer (Hitachi High Technologies, Tokyo, Japan). The scan mode was 700–voltage xenon lamp and the spectra was collected with scanning emission (Em) from 250 to 550 nm at 1 nm intervals by varying the excitation (Ex) wavelength from 200 to 450 nm at intervals of 5 nm. The scan speed was set as 2400 nm min^{-1} with Ex and Em silt bandwidths of 5 nm. The blank scans were recorded with Milli-Q water. The collected EEM spectra were corrected for water Raman scatter peaks, inner-filter effects and Rayleigh scattering effects according to previous methods.[10] Daily lamp variations were eliminated by normalizing the corrected and trimmed EEMs to Raman (275 nm) units (RU_{275}). Humification index (HIX) was determined as the ratio of emission scanning areas at the wavelength range of 300–345 nm to 435–480 nm with an excitation wavelength of 254 nm.[11]

Fluorescent DOM (FDOM) components were further identified using a parallel factor analysis (PARAFAC), which separates a data set of EEMs into mathematically and chemically independent components (each representing a single fluorophore or a group of strongly covarying fluorophores).

These components were multiplied by their excitation and emission spectra, representing either pure or combined spectra. PARAFAC modelling was performed using the drEEM toolbox (ver. 0.2.0) for MATLAB (R2012a).[12] A 3-component model was well validated using an extended method of split half analysis, which was developed to assemble 6 different dataset "halves" and produce 3 validation tests "$S_4C_6T_3$" (Splits-4, Combinations-6, Tests-3). The number of components was determined by comparing the residuals from models with gradually increased number of components. To quantify and compare the changes in FDOM components, the concentration of each component was estimated as their maximum fluorescent intensity (F_{max}) in the model.

HPLC-EEM

An Agilent 1200 LC System equipped with fluorescence detector was used in this study. HPLC (Agilent 1200 series) equipped with a reverse phase C18 column (4.6 mm × 150 mm, 5 μm) was applied. The mixture of ammonium acetate (10 mM) and acetonitrile were used as the mobile phase with a flow rate of 1 mL min^{-1}.[13] The gradient elution method is shown in Table S1. Based on the results of EEM, the scan mode was set as Table S2 to obtain the hydrophobicity-distinguished components of FDOM. In RP-HPLC system, the relatively hydrophilic components are generally eluted in shorter retention time, whereas those with longer retention times are hydrophobic.

Molecular Characterization of DOM

Owing to the ESI-FTICR-MS analysis requires a low salt content, the DOM was isolated using solid phase extraction (SPE). The 5 g Bond Elut PPL SPE cartridges (Agilent Technologies) was used and preactivated with LC-MS grade methanol and water. Water samples were adjusted to pH

2 and loaded onto the conditioned PPL SPE cartridges. All DOM loadings were less than the 2 mM C L^{-1} recommend threshold to avoid breakthrough.[14] Salts in the samples were initially eluted by 0.01 M trace metal grade HCl (Fisher Scientific), followed by drying with purity N_2, and then methanol was added to elute the DOM. Ultrapure water was used as blank sample to check for potential contamination. Mean recovery of carbon by the PPL solid phase was similar between DOM sources (65 ± 9%).

The PPL extracted DOM in methanol was diluted by 2 with ultrapure methanol plus 0.1% ammonium hydroxide (pH 8). Additionally, SPE samples were diluted to less than 50 mg C L^{-1} to minimize charge competition during ionization. DOM was continuously infused into an Apollo II ESI Ion source in negative mode of a Bruker Daltonics 7 T Apex Qe FTICR-MS. The injection rate was set at 120 μL h^{-1}. Accumulation of ions in the hexapole ranged from 0.4 to 3 s before being transferred to the ICR cell, where 300 scans, collected with a 4 MWord time domain, were coadded for each sample. The summed free induction decay signal was zero-filled once and Sine-Bell apodized prior to fast Fourier transformation and magnitude calculation using Bruker Daltonics Data Analysis software. Similarly, a 50/50 (v/v) MeOH: H_2O blank spectrum was collected to test for contamination.

All m/z lists ranged from 150 m/z to 750 m/z were considered using a signal to noise ratio ≥ 4 and were internally calibrated using data lists of fatty acids and a list of peaks common to all samples.[15] Those peaks detected in the blank spectrum were discarded prior to formulas assignments. Peak detection limits were standardized between samples by adjusting the dynamic range of each sample to that of the sample with the lowest dynamic range (dynamic range = average of the largest 20% of peaks divided by the signal to noise threshold intensity; standardized detection limit =

average of largest 20% of peaks within a sample divided by the lowest dynamic range within the sample set).[16] Peaks below the standardized detection limit were deleted to prevent false negatives for the occurrence of a peak within samples with low dynamic range.

Molecular formulae assignment was performed based on exact masses following the criteria described in Stubbins et al.[17] A molecular formula calculator (Molecular Formula Calc v.1.0 ©NHMFL, 1998) generated formulas using C, H, O, N, and S. The mass accuracy threshed was ≤ ± 1 ppm. Molecular formulae assignment was built on the modified criteria that was more proper for DOM from cyanobacteria[6]. Formulas containing N_2S_2 were removed out from the assignments.

Standardized peak intensities (z) of formulas within a sample were calculated following:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where x is the measured peak intensity, μ is mean peak intensity, and σ is the standard deviation in peak intensity within the sample.

On the bases of presence or absence of the heteroatoms N, P and S in assigned formulae, the molecules were grouped into 4 elemental classes: CHO, CHON, CHOS, and CHONS. Formulae were also grouped into several compound classes based on modified aromaticity index (AI_{mod}), the ratios of H/C and O/C, and the number of N as follows: condensed aromatics ($AI_{mod} > 0.67$), aromatics ($0.5 < AI_{mod} \leq 0.67$), highly unsaturated, high oxygen ($AI_{mod} < 0.5$, $O/C < 0.5$, and $H/C < 1.5$), highly unsaturated, low oxygen ($AI_{mod} < 0.5$, $0.5 \leq O/C < 0.9$, and $H/C < 1.5$), aliphatics ($N = 0$, $O/C < 0.9$, and $1.5 \leq H/C \leq 2$), peptides ($N > 0$, $O/C < 0.9$, and $1.5 \leq H/C \leq 2$), saturated fatty acids ($O/C < 0.9$ and $H/C > 2$), and sugars ($O/C \geq 0.9$).[18] The computed averages values for m/z, AI_{mod}, H, C, N, O, S, and the H/C and O/C ratios were based upon intensity-weighted averages of mass peaks with assigned molecular formulas.

157

158 ***Microbial Community Analysis***

159 Briefly, the glass beads in the bottom 10 cm of each reactor of the bioreactor were extracted
160 and homogenized, respectively, in sterile environment. Then, 8 g of the extracted glass beads were
161 added in sterile polypropylene tubes with 15 mL autoclaved 0.1% tetrasodium pyrophosphate.
162 Tubes were mixed 40 s and sonicated for 5 min with a Bransonic tabletop ultrasonic cleaner model
163 1501 (40 kHz, 70 W, Branson Ultrasonic Corporation, Danbury, CT) to disrupt biofilms and lyse
164 cells. Finally, solutions were filtered by 0.22 μm sterilized polycarbonate filters to obtain the
165 microorganisms and stored in –70 °C before further analysis.

166 DNA was extracted in duplicate for each sample using a PowerSoil kit (MO BIO Laboratories,
167 Carlsbad, CA) according to the manufacturer's directions, and the DNA solutions (approximately
168 100 μL for each extraction) were pooled to reduce sample variability. 16S rRNA was partially
169 amplified using the forward primer 515F (5′-GTGCCAGCMGCCGCGG-3′) and the reverse primer
170 907R (5′-CCGTCAATTCCTTTGAGTTT-3′).[19] The PCR reaction program was as follows: 0.2 μM
171 of each primer and 0.2 mM dNTP, and 1.5 mM $MgCl_2$ and 0.2 units of Taq polymerase. PCR
172 amplification (initial denaturation at 95 °C for 2 min, and 25 cycles at 95 °C for 30 s, 55 °C for 30 s
173 and 72 °C for 30 s, and a final extension at 72 °C for 5 min) and gel purification procedures were
174 conducted and processed by the Meiji Biotechnology Company (Shanghai, China) for high-
175 throughput DNA sequencing with the Illumina MiSeq System (Illumina, San Diego, U.S.A.).[19]

176 All 16S rRNA sequence reads were filtered, denoised and processed by using Trimmomatic
177 (ver. 0.36) (http://www.usadellab.org/cms/?page=trimmomatic) and FLASH
178 (https://ccb.jhu.edu/software/FLASH/) under the following rules: (1) remove the bases with the

179 trailing quality score being under 20, scan the read with a 50-base wide sliding window, do cutting
180 operations when the average quality score per base drops below 20, and drop the reads below 50
181 bases long, (2) merge the reads in pairs into a new one on the basis of their overlap (the minimum
182 length overlapped was 10 bases), (3) remove the merged reads that the mismatch ratio in overlapping
183 regions larger than 0.2, and (4) remove the reads with the mismatch numbers of primer larger than
184 2. The after quality control were analyzed through QIIME (ver. 1.6.0).[20] Sequences with
185 similarities > 97% were clustered into one operational taxonomic unit (OTU) with Usearch (ver. 7.1)
186 (http://drive5.com/uparse). OUT representative sequences were aligned against the Greengenes
187 (Release 13.5 http://greengenes.secondgenome.com/), and chimeric sequences were identified using
188 Mothur (ver. 1.30).[21 22] To compare and perform statistics across samples, the numbers of sequences
189 in each sample were first normalized down to the number in the sample with the fewest sequences.
190 Alpha diversity estimates (Shannon) and Beta diversity were determined in the QIIME program.
191 The raw data has been submitted to NCBI Sequence Read Archive (SRA) with accession number
192 of SRR5252713, SRR5252714, SRR5252715 and SRR5252716.
193

194 ***2D-COS Analysis***

195 The 2D-COS was applied depending on the method of Noda and Ozaki.[23] To obtain the
196 variation of hydrophobicity-distinguished FDOM fractions with HRTs, 2D-COS was performed on
197 the HPLC-EEM spectral data. Prior to 2D correlation spectra analysis, the HPLC-EEM data were
198 normalized to avoid dimension difference, and then the 2D correlation spectra analysis was
199 performed using 2D Shige ver. 1.3 software (Kwansei-Gakuin University, Japan). A set of HRTs-
200 dependent HPLC-EEM data was obtained, and 2D correlation spectra were produced based on the
201 HPLC-EEM data using the HRTs as the external perturbation. The 2D correlation spectra are
202 comprised of synchronous and asynchronous map. In the synchronous 2D correlation spectra, the
203 auto-peak centered at diagonal positions represents the overall extent of intensity variation of a
204 specific spectral variable over the interval. The positive cross-peak at the wavelength pair of λ_1/λ_2
205 suggests the possible existence of a coupled or related origin of the spectral intensity variations
206 measured at λ_1 and λ_2, while a negative cross-peak at the wavelength pair of λ_1/λ_2 suggests that the
207 change occurring at λ_1 and $\lambda_2$was asynchronous or inverse.

208 As to the asynchronous 2D correlation spectra, if the change of λ_1 and λ_2 was concurrent, the
209 positive cross-peak at the wavelength pair of λ_1/λ_2 suggests that the change occurring at the
210 wavelength λ_1 is quicker than that occurring at the wavelength λ_2, whereas a negative cross-peak at
211 the wavelength pair of λ_1/λ_2 suggests that the change occurring at the wavelength λ_1 is slower than
212 that occurring at the wavelength λ_2. If the change of λ_1 and λ_2 was asynchronous or inverse, the
213 same conditions of peaks indicate the opposite direction of the change.[24]
214

215 **Table S1.** HPLC Gradient Elution Scheme

Time (min)	Acetonitrile (%)	Ammonium acetate (%)	Flow rate (mL min^{-1})
0.00	20.0	80.0	1.0
4.00	70.0	30.0	1.0
5.00	90.0	10.0	1.0
10.00	95.0	5.0	1.0
15.00	95.0	5.0	1.0

216

Table S2. FLD Parameters Setup

Scan mode	Ex (nm)	Em (nm)
Multi-excitation scan	230-290	324
	220-380	425
Multi-emission scan	235	300-500

Table S3. Changes of DOC Concentration in the 4-Stage Plug-Flow Bioreactor and Batch Inoculation (mg L^{-1})

	4-stage plug-flow bioreactor	Bioassay	
		7 days inoculation	56 days inoculation
Influent	35.77 ± 6.19	19.61 ± 3.68	11.74 ± 0.92
Eff-1	21.60 ± 0.88	15.07 ± 1.29	9.94 ± 1.76
Eff-2	10.68 ± 0.34	9.51 ± 1.33	8.87 ± 1.25
Eff-3	8.62 ± 1.35	9.20 ± 0.38	8.95 ± 0.11
Eff-4	7.44 ± 0.51	9.18 ± 0.43	8.64 ± 0.92

222 **Table S4.** DOM Molecular Characterization after Having Passed through the 4-Stage Plug-Flow
223 Bioreactor

	Influent	Eff-1	Eff-2	Eff-3	Eff-4
Total number	1065	1011	1032	917	857
m/z_w[a]	334	478	373	385	396
AI_{mod}[a]	0.12	0.22	0.17	0.21	0.18
C_w[a]	15.32	24.40	17.82	19.16	20.46
H_w[a]	21.54	31.71	23.80	25.00	29.75
O_w[a]	6.76	6.42	6.70	6.37	6.09
N_w[a]	0.78	1.46	0.81	0.86	0.73
S_w[a]	0.34	0.99	0.54	0.56	0.45
O/C_w[a]	0.49	0.32	0.44	0.40	0.34
H/C_w[a]	1.42	1.32	1.37	1.33	1.45
N/C_w[a]	0.06	0.06	0.05	0.05	0.04
S/C_w[a]	0.02	0.04	0.03	0.03	0.02
%CHO[b]	39%	25%	41%	43%	50%
%CHON[b]	34%	21%	23%	24%	23%
%CHOS[b]	17%	17%	16%	14%	15%
%CHONS[b]	10%	36%	20%	19%	12%
%Condensed aromatics[b]	5%	12%	7%	9%	4%
%Aromatics[b]	6%	12%	8%	10%	5%
%Highly unsaturated, low oxygen[b]	22%	26%	21%	24%	28%
%Highly unsaturated, high oxygen[b]	27%	10%	24%	22%	17%
%Aliphatics[b]	20%	13%	19%	16%	26%
%Peptide-like[b]	16%	21%	18%	15%	16%
%Saturated fatty acids[b]	2%	5%	2%	3%	4%
%Sugars[b]	2%	1%	1%	1%	0%

[a]Intensity-weighted average. [b]The composition is based on the total intensity-weighted proportion of molecules

224
225

226 **Table S5.** Molecular Characterization of the Labile DOM, Semilabile DOM, Refractory DOM and
227 Bioproduced DOM

	Labile	Semilabile	Refractory	Bioproduced
m/z[a]	377	309	303	479
AI_{mod}[a]	0.09	0.01	0.20	0.20
O/C[a]	0.46	0.57	0.50	0.27
H/C[a]	1.42	1.48	1.33	1.47
N/C[a]	0.07	0.07	0.03	0.06
S/C[a]	0.03	0.03	0.01	0.04
%CHO[b]	21	22	65	24
%CHON[b]	41	49	29	28
%CHOS[b]	24	25	6	20
%CHONS[b]	14	4	0	28
%Condensed aromatics[b]	7	2	0	10
%Aromatics[b]	12	5	4	8
%Highly unsaturated, low oxygen[b]	19	17	28	23
%Highly unsaturated, high oxygen[b]	16	28	41	5
%Aliphatics[b]	17	25	18	22
%Peptide-like[b]	23	17	9	24
%Saturated fatty acids[b]	3	2	0	8
%Sugars[b]	3	4	0	0

[a]Average values of all molecules. [b]This percentage was calculated based on number of molecules

228

229 **Figure S1.** Location of Meiliang Bay and Taihu sampling site (coordinates: 31°28'46.17"N,

230 120°11'19.70"S).

231

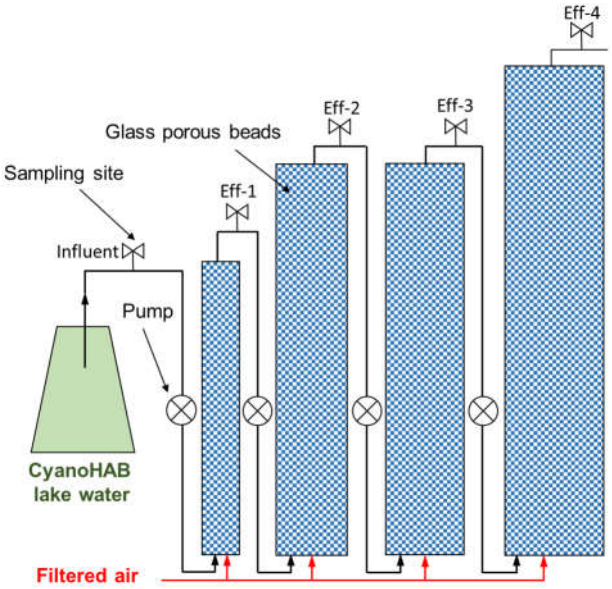
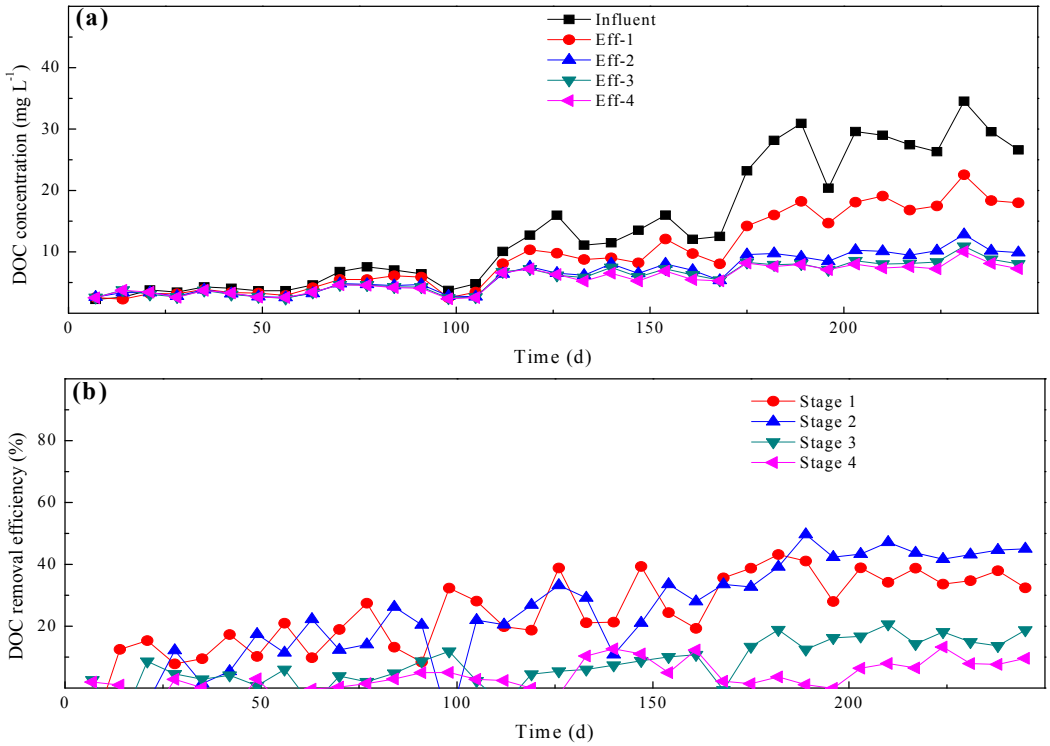


Figure S2. Setup and operation schemes of the 4-stage plug-flow bioreactor.

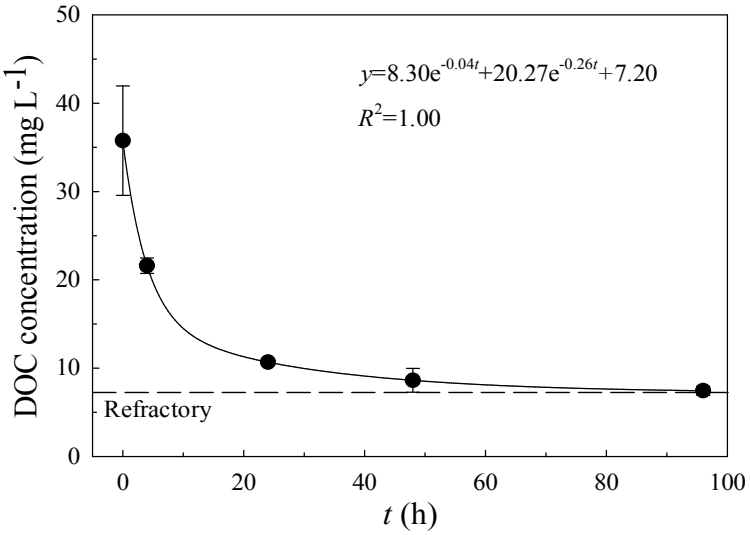


235 **Figure S3.** Concentrations and removal efficiencies of DOC in the 4-stage plug-flow bioreactor

236 during incubation. (a) DOC concentrations, and (b) DOC removal efficiencies in each stage of the

237 bioreactor.

238



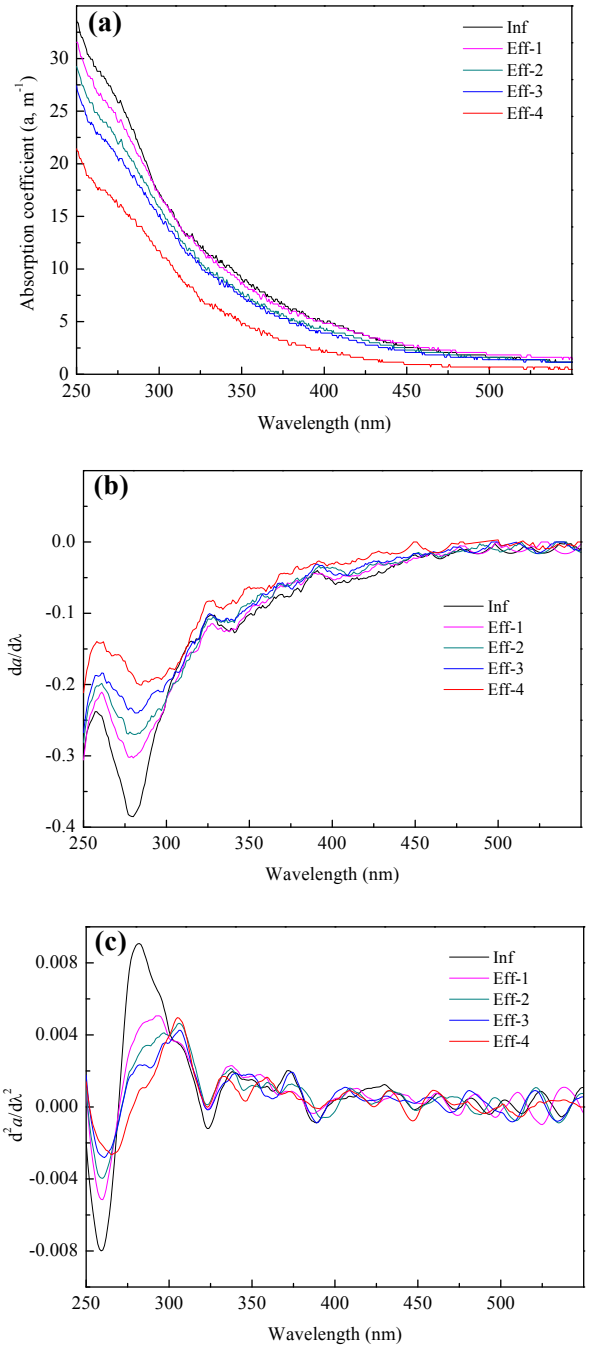
239

240 **Figure S4.** *G* model fitting results of DOC concentrations in the 4-stage plug-flow bioreactor. *G*

241 model: $DOC=C_1e^{-k_1t}+C_2e^{-k_2t}+C_3e^0$, where C_1, C_2 and C_3 were the concentrations of labile DOC,

242 semilabile DOC and refractory DOC, respectively.

243



244 **Figure S5.** (a) CDOM absorption coefficient spectra (a, m^{-1}), (b) first derivative absorption spectra,

245 and (c) second derivative absorption spectra of the DOM as passed through the 4-stage plug-flow

246 bioreactor.

247

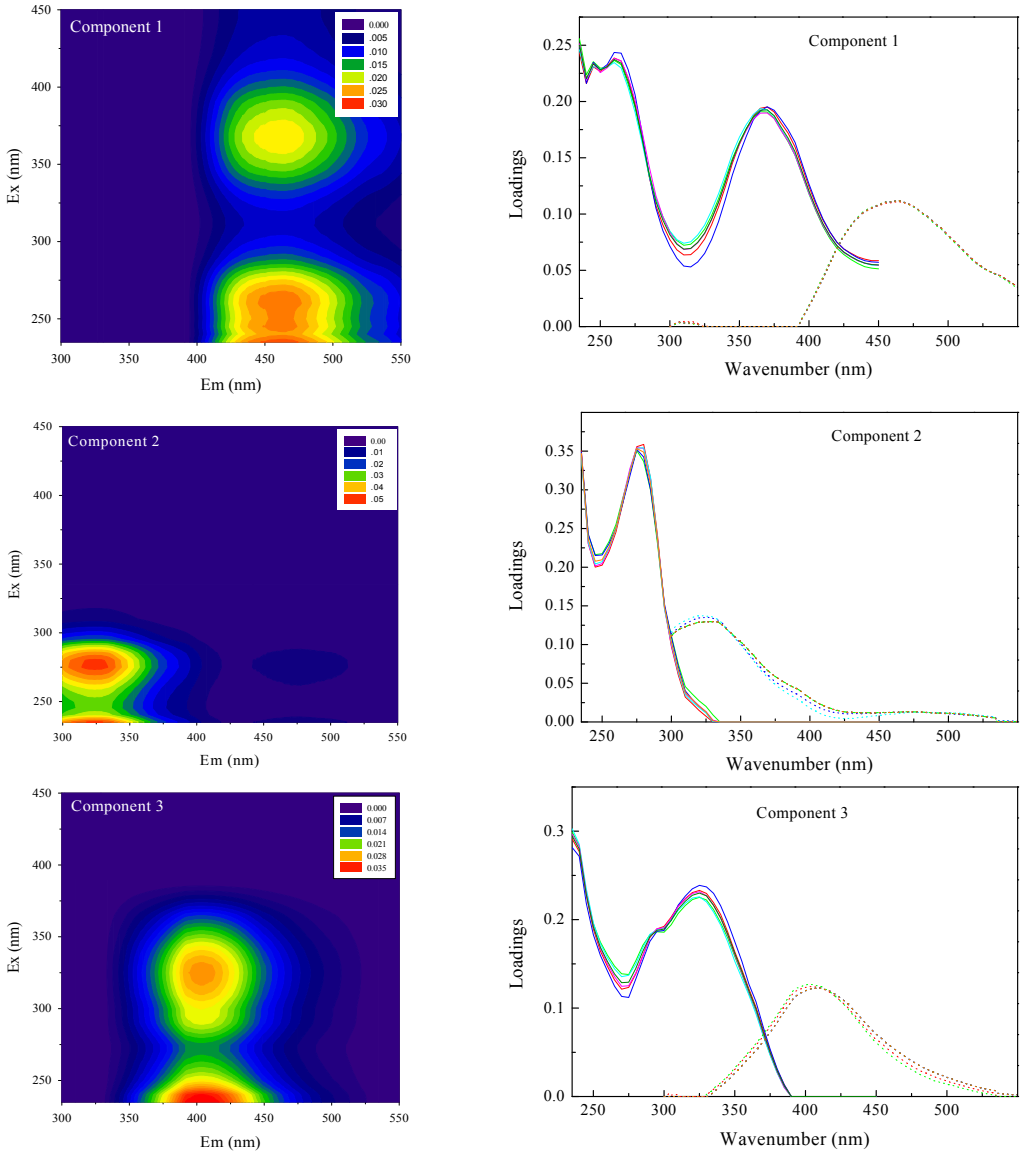
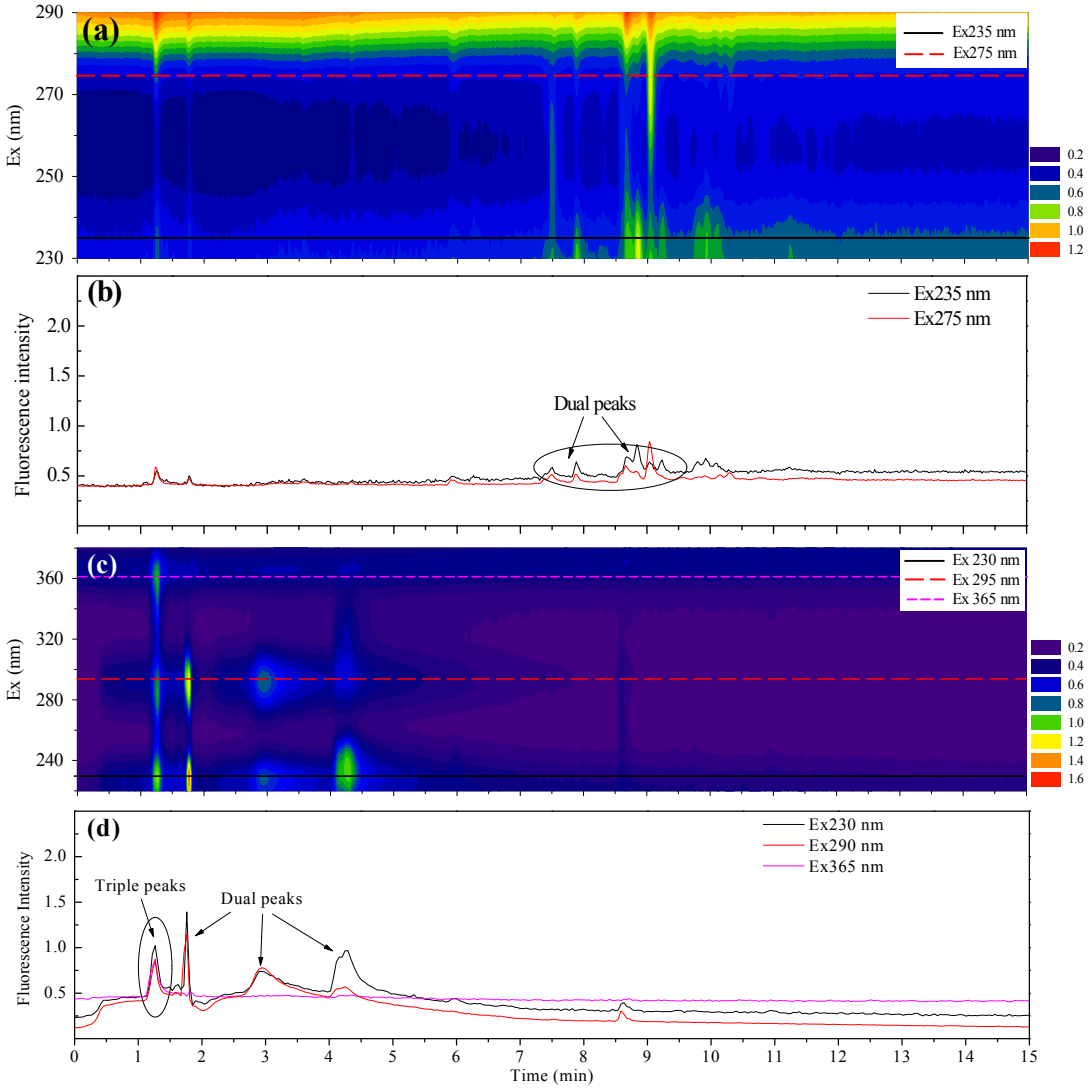


Figure S6. EEM spectrum of the 4 components identified by the PARAFAC analysis for the DOM samples and the highly overlaid excitation and emission spectra estimated using the split-half validation procedure.



252 **Figure S7.** HPLC fluorescence Excitation-Time-Maps of CyanoHAB-DOM. (a) and (b) Excitation-

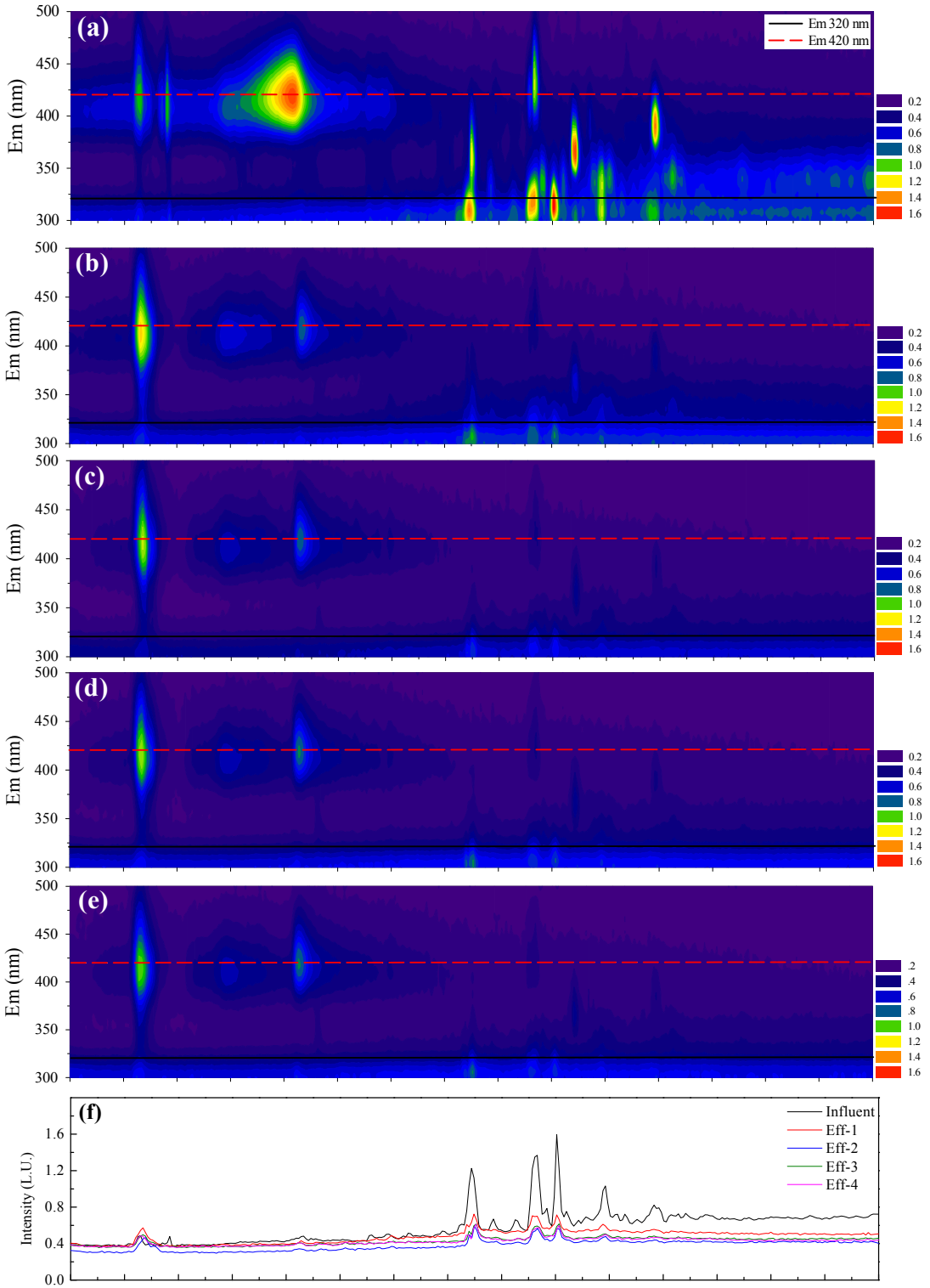
253 Time-Map at Em 324 nm and the extracted chromatograms at Em 235 and 275 nm, respectively,

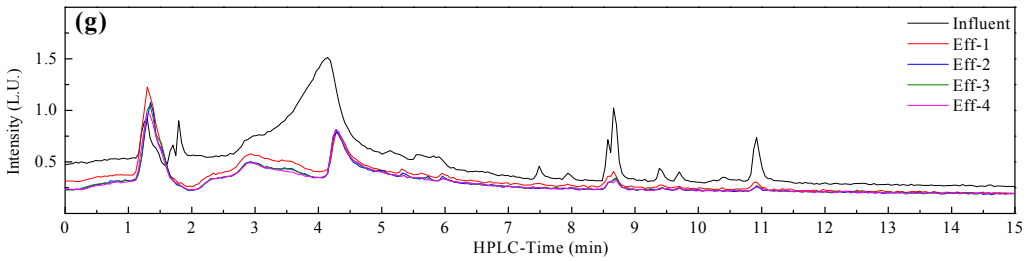
254 and (c) and (d) Excitation-Time-Map at Em 425 nm and the extracted chromatograms at Em 230,

255 290 and 365 nm, respectively.

256

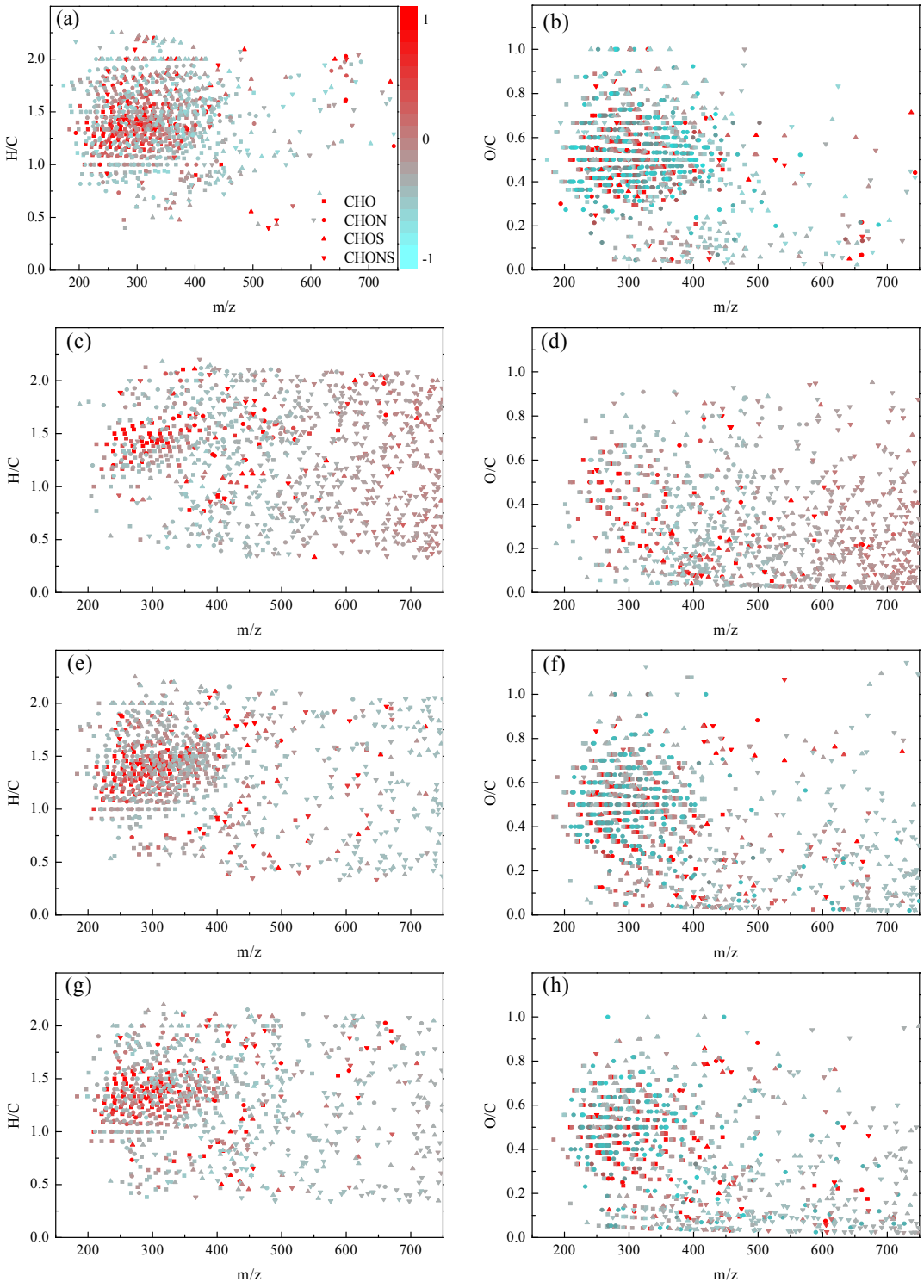
257

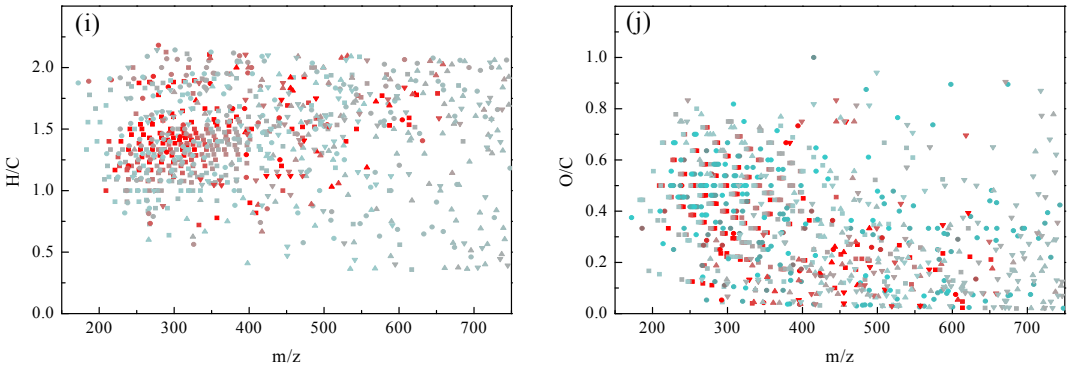

(a)
Em 320 nm
Em 420 nm
(b)
(c)
(d)
(e)
Em (nm)
500
450
400
350
300
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
(f)
Influent
Eff-1
Eff-2
Eff-3
Eff-4
Intensity (L.U.)
1.6
1.2
0.8
0.4
0.0



258 **Figure S8.** HPLC fluorescence emission-time-maps at Ex 235 nm for the DOM as passed through
259 the 4-stage plug-flow bioreactor. (a)–(e) Emission-time-maps of influent, Eff-1, Eff-2, Eff-3, Eff-4,
260 respectively, and (f) and (g) the extracted chromatograms at Em 320 nm and Em 420 nm,
261 respectively.

262


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
H/C
O/C
m/z
CHO
CHON
CHOS
CHONS
1
0
-1



263 **Figure S9.** H/C versus mass and O/C versus mass of DOM samples. (a) and (b) Influent DOM, (c)

264 and (d) Eff-1 DOM, (e) and (f) Eff-2 DOM, (g) and (h) Eff-3 DOM, and (i) and (j) Eff-4 DOM.

265

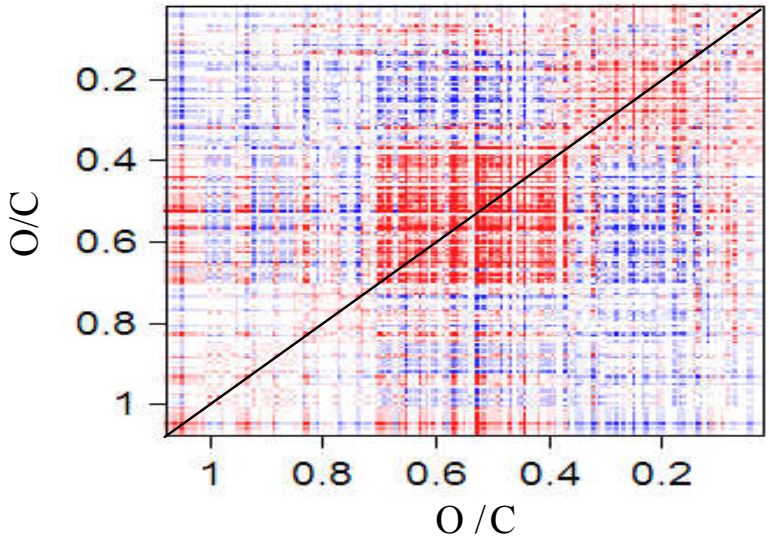


Figure S10. The synchronous plots of generated by applying 2D-COS on the O/C of CHO compounds. The correlation was based on the presence/absence of each individual formula as DOM passed through the 4-stage plug-flow bioreactor. Red color indicates a positive correlation, while blue color indicates a negative correlation.

There were three distinct areas with strong signals along the diagonal line. Formulas at high O/C of 0.4–0.7 showed the greatest changes in the number of CHO formulas with biodegradation, followed by formulas with O/C of 0.1–0.4, whereas the formulas at O/C of 0.7–1.1 exhibited the smallest changes. By reexamining the CHO formulas with the O/C of 0.4–0.7, it was clear that the number of formulas in this region decreased as passed through the 4-stage bioreactor. The cross correlation (off-diagonal) signals indicated that the formulas that fell in the region of O/C 0.4–0.7 had positive correlations with formulas at O/C ratios of 0.7–1.1, suggesting that the number of these 2 groups of formulas decreased together. However, formulas with O/C of 0.1–0.4 were negatively correlated with those at 0.4–0.7, implying that new compounds at O/C of 0.1–0.4 were added to the DOM pool as the number of formulas at 0.4–0.7 and 0.7–1.1 decreased with biodegradation. These results confirmed that the formulas with higher O/C ratios were most labile and supposed to be

metabolized quickly, whereas some low oxygenated compounds were produced. Although some observation can be made from the O/C versus mass plots visually, 2D-COS can provide towards insights when trying to identify the different regions with regard to their bioreactivity and those changes in the number of formulas in each region.

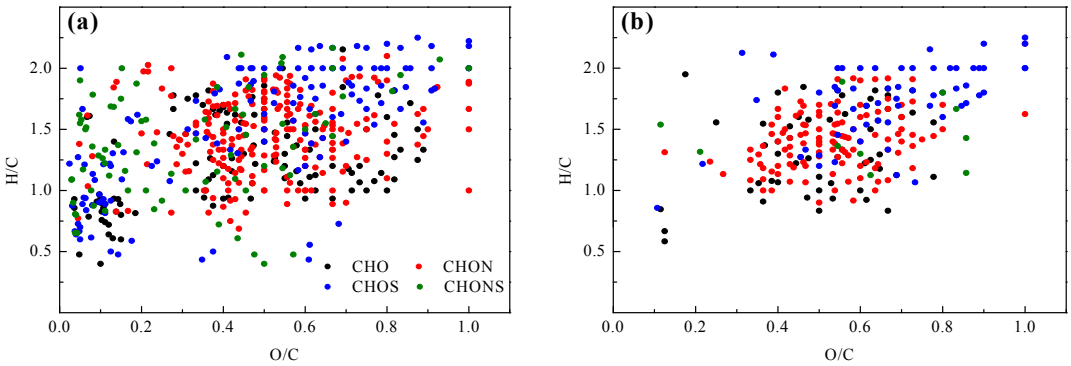
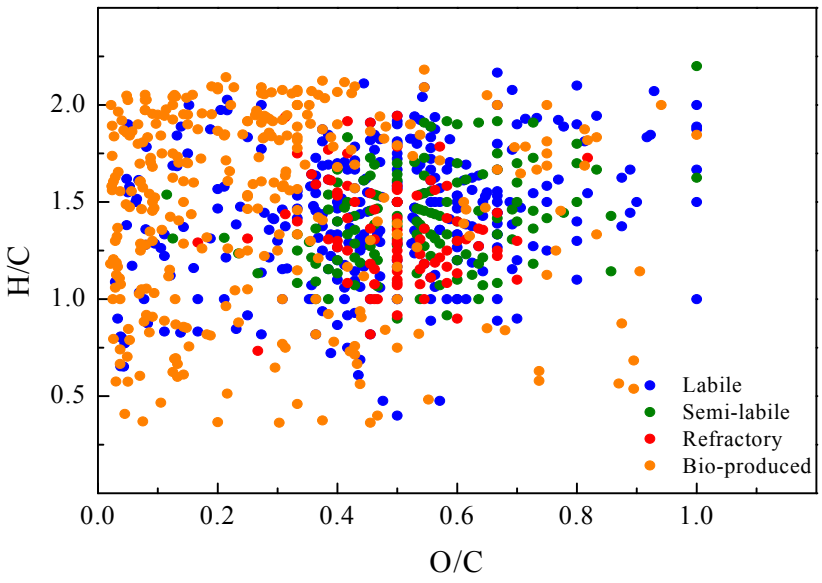


Figure S11. van-Krevelen diagrams for labile DOM molecules. (a) Labile molecules in influent, and (b) semilabile molecules in influent.

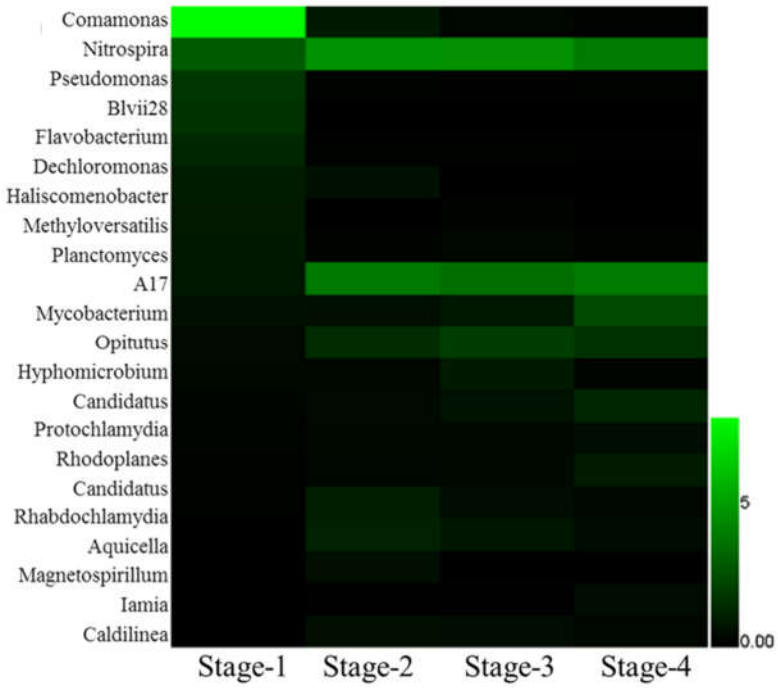


293

294 **Figure S12.** van-Krevelen diagram for nitrogen-containing molecules (CHON and CHONS) in

295 different DOM bioreactivity classification.

296

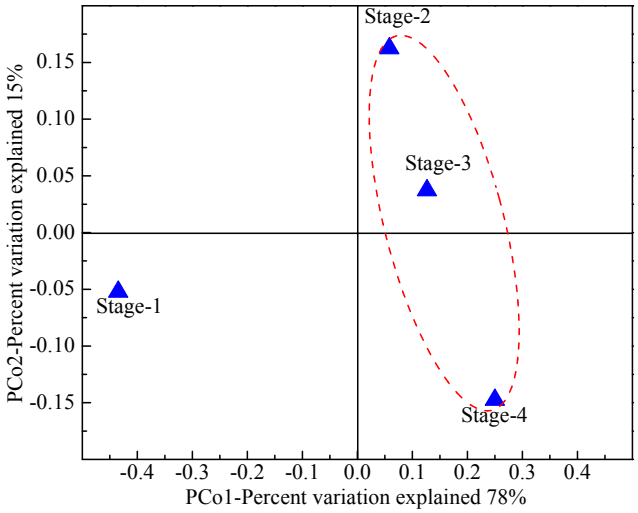


297

298 **Figure S13.** Heat map of microbial groups at the genus level at 10 cm depth of each stage of the

299 bioreactor.

300



301

302 **Figure S14.** Principal coordinate analysis of microbial structures in the 4-stage plug-flow bioreactor.

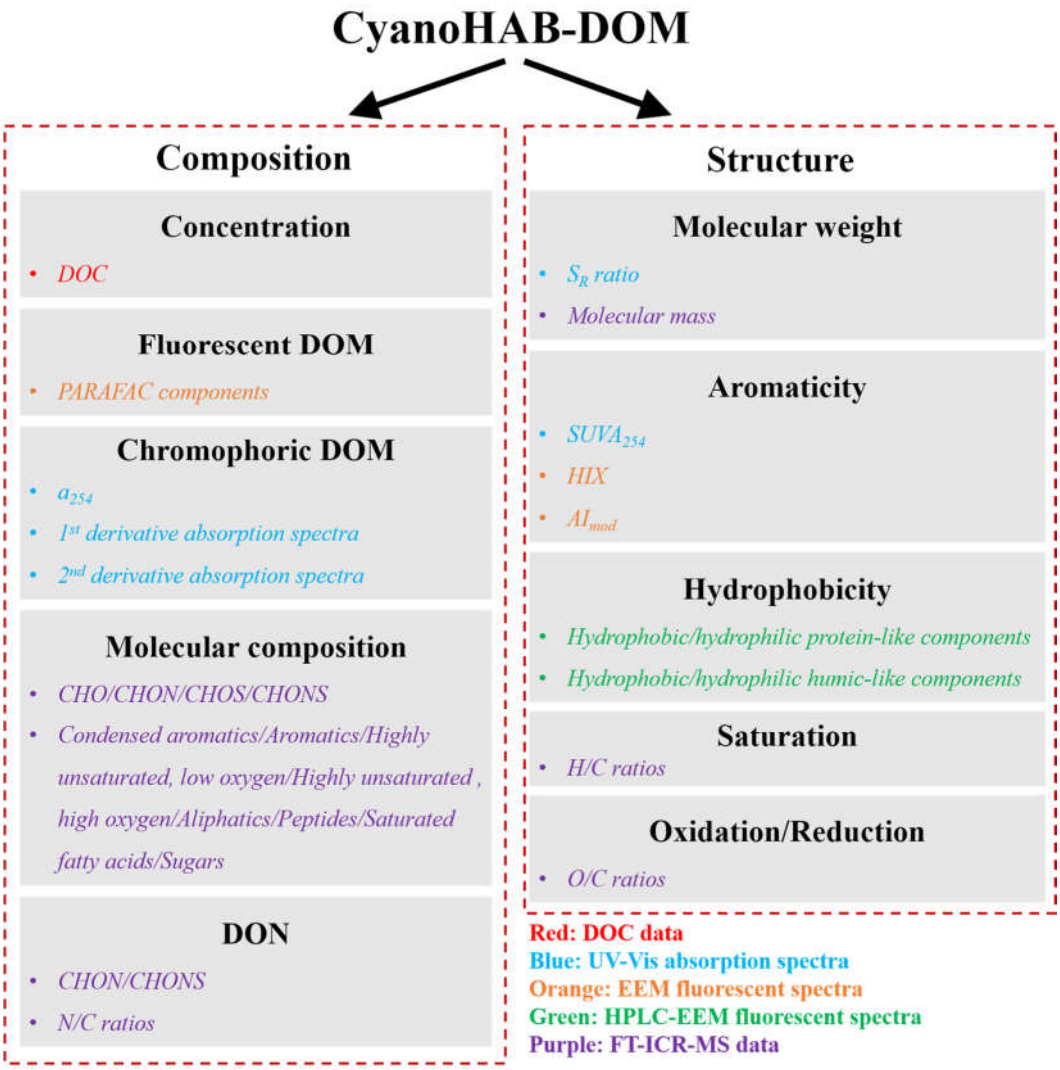


Figure S15. A succinct synthesizing statement of the multiple analytical tools used in this study.

Multiple analytical tools including DOC concentration, molecular composition, absorption and fluorescence spectra, and hydrophobicity-distinguished components were used to characterize the temporary evolution in composition and structure of CyanoHAB-DOM in the 4-stage plug-flow bioreactor (Figure S15). Through *G* model fitting, the bulk DOC was separated into labile, semilabile, and refractory fractions. This attenuation profiles provided an excellent way to generally distinguish the DOM compounds with various bioavailability. Concentrations of chromophoric and fluorescent DOM were represented as a_{254} and F_{max} of PARAFAC components, respectively. With the use of FTICR-MS, 4 kinds of molecules, including CHO, CHON, CHOS and CHONS, were

assigned and they were further classified into different compounds, e.g. aromatics, unsaturated compounds, peptides and so on. Additionally, changes in DON were investigated with nitrogen-bearing molecules.

Results also provided information on the structure of CyanoHAB-DOM, including molecular weight, aromaticity, saturation, oxidation/reduction and hydrophobicity. The S_R ratio and molecular mass (m/z) reflected the changes in molecular weight of DOM. The aromaticity of DOM was related to the values of $SUVA_{254}$, HIX and AI_{mod}. H/C and O/C ratios, which were obtained by FTICR-MS, indicated the saturation and oxidation of DOM molecules, respectively. In combination of 2D-COS, HPLC-EEM further revealed the evolution of hydrophobicity-distinguished components in a successive biodegradation. Overall, the multiple analytical data provide a deeper insight on the relationship between the composition and structure of CyanoHAB-DOM and its biodegradation.

REFERENCES

1. Søndergaard, M.; Worm, J., Measurement of biodegradable dissolved organic carbon (BDOC) in lake water with a bioreactor. *Water Res.* **2001,** *35*, (10), 2505-2513.
2. Ribas, F.; Frias, J.; Lucena, F., A new dynamic method for the rapid-determination of the biodegradable dissolved organic-carbon in drinking-water. *J. Appl. Bacteriol.* **1991,** *71*, (4), 371-378.
3. Kaplan, L. A.; Newbold, J. D., Measurement of streamwater biodegradable dissolved organic carbon with a plug-flow bioreactor. *Water Res.* **1995,** *29*, (12), 2696-2706.
4. Sleighter, R. L.; Cory, R. M.; Kaplan, L. A.; Abdulla, H. A. N.; Hatcher, P. G., A coupled geochemical and biogeochemical approach to characterize the bioreactivity of dissolved organic matter from a headwater stream. *J. Geophys. Res.: Biogeosci.* **2014,** *119*, (8), 1520-1537.
5. De Vittor, C.; Larato, C.; Umani, S. F., The application of a plug-flow reactor to measure the biodegradable dissolved organic carbon (BDOC) in seawater. *Bioresource Technol.* **2009,** *100*, (23), 5721-5728.
6. Bittar, T. B.; Vieira, A. A. H.; Stubbins, A.; Mopper, K., Competition between photochemical and biological degradation of dissolved organic matter from the cyanobacteria Microcystis aeruginosa. *Limnol. Oceanogr.* **2015,** *60*, (4), 1172-1194.
7. Helms, J. R.; Stubbins, A.; Ritchie, J. D.; Minor, E. C.; Kieber, D. J.; Mopper, K., Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* **2008,** *53*, (3), 955-969.
8. Weishaar, J. L.; Aiken, G. R.; Bergamaschi, B. A.; Fram, M. S.; Fujii, R.; Mopper, K., Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* **2003,** *37*, (20), 4702-4708.
9. Loiselle, S. A.; Bracchini, L.; Dattilo, A. M.; Ricci, M.; Tognazzi, A.; Cózar, A.; Rossi, C., The optical characterization of chromophoric dissolved organic matter using wavelength distribution of absorption spectral slopes. *Limnol. Oceanogr.* **2009,** *54*, (2), 590-597.
10. Zhou, Y. Q.; Zhang, Y. L.; Jeppesen, E.; Murphy, K. R.; Shi, K.; Liu, M. L.; Liu, X. H.; Zhu, G. W., Inflow rate-driven changes in the composition and dynamics of chromophoric dissolved organic matter in a large drinking water lake. *Water Res.* **2016,** *100*, 211-221.
11. Huguet, A.; Vacher, L.; Relexans, S.; Saubusse, S.; Froidefond, J. M.; Parlanti, E., Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* **2009,** *40*, (6), 706-719.
12. Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R., Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Anal. Methods* **2013,** *5*, (23), 6557-6566.
13. Li, W. T.; Xu, Z. X.; Li, A. M.; Wu, W.; Zhou, Q.; Wang, J. N., HPLC/HPSEC-FLD with multi-excitation/emission scan for EEM interpretation and dissolved organic matter analysis. *Water Res.* **2013,** *47*, (3), 1246-1256.
14. 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.: Methods* **2008,** *6*, 230-235.
15. Sleighter, R. L.; McKee, G. A.; Liu, Z.; Hatcher, P. G., Naturally present fatty acids as internal calibrants for Fourier transform mass spectra of dissolved organic matter. *Limnol. Oceanogr.: Methods* **2008,** *6*, (6), 246-253.
16. Spencer, R. G. M.; Guo, W. D.; Raymond, P. A.; Dittmar, T.; Hood, E.; Fellman, J.; Stubbins, A., Source and biolability of ancient dissolved organic matter in glacier and lake ecosystems on the Tibetan Plateau. *Geochim. Cosmochim. Ac.* **2014,** *142*, 64-74.
17. Stubbins, A.; Spencer, R. G. M.; Chen, H. M.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J., Illuminated darkness: Molecular signatures of Congo River

368 dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry.
369 *Limnol. Oceanogr.* **2010,** *55*, (4), 1467-1477.
370 18. Stubbins, A.; Lapierre, J. F.; Berggren, M.; Prairie, Y. T.; Dittmar, T.; del Giorgio, P. A., What's in an EEM?
371 Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. *Environ. Sci. Technol.*
372 **2014,** *48*, (18), 10598-10606.
373 19. Bernard, L.; Chapuis-Lardy, L.; Razafimbelo, T.; Razafindrakoto, M.; Pablo, A. L.; Legname, E.; Poulain, J.;
374 Bruls, T.; O'Donohue, M.; Brauman, A.; Chotte, J. L.; Blanchart, E., Endogeic earthworms shape bacterial functional
375 communities and affect organic matter mineralization in a tropical soil. *ISME J.* **2012,** *6*, (1), 213-222.
376 20. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena,
377 A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone,
378 C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.;
379 Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community
380 sequencing data. *Nat. Methods* **2010,** *7*, (5), 335-336.
381 21. Schloss, P. D.; Westcott, S. L., Assessing and improving methods used in operational taxonomic unit-based
382 approaches for 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.* **2011,** *77*, (10), 3219-3226.
383 22. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O., The SILVA
384 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **2013,** *41*,
385 (D1), D590-D596.
386 23. Noda, I.; Ozaki, Y., *Two-dimensional correlation spectroscopy: applications in vibrational and optical*
387 *spectroscopy*. John Wiley and Sons Inc. London: 2005.
388 24. He, X. S.; Xi, B. D.; Li, W. T.; Gao, R. T.; Zhang, H.; Tan, W. B.; Huang, C. H., Insight into the composition
389 and evolution of compost-derived dissolved organic matter using high-performance liquid chromatography
390 combined with Fourier transform infrared and nuclear magnetic resonance spectra. *J. Chromatogr. A* **2015,** *1420*,
391 83-91.