SUPPORTING INFORMATION: EXTENDED MATERIALS AND METHODS

Representative Pyrogenic Organic Matter (PyOM) Description

To assess the bioavailability of PyOM, we searched primary literature for representative compounds of the PyOM continuum. Specifically, we targeted previously characterized organic compounds from field and laboratory burns of various fuel types representing a range of moisture, temperature, and oxygen conditions (Table S1). The chosen compounds focused on biomass burning alteration products which are often used to characterize PyOM in different environmental media, such as aerosols, soils, and waters. This included compounds such as theoretical BC compounds [defined here as condensed aromatic core structures polysubstituted with O-containing functionalities (Wagner et al., 2017)], anhydrosugars, and polycyclic aromatic hydrocarbons (PAHs). The list also included compounds created and/or transformed from biomass burning, such as those derived from biopolymers like lignin (e.g., methoxyphenols), waxes (e.g., n-alkenes from thermal dehydration of n-alkanols), and resins (e.g., thermally oxidized diterpenoids)(Oros and Simoneit, 2001a, b). In total, our literature search for PyOM chemistries yielded 389 compounds with 207 unique chemical formulae from 12 primary literature sources (Table S1).

Dissolved Organic Matter (DOM) Description

Global surface water and sediment DOM pool composition was measured with Fourier transform ion cyclotron mass spectrometer (FTICR-MS) in order to be able to detected thousands of OM compounds per sample, thereby providing a high-resolution perspective on DOM, and the data used here are described in more detail by Garayburu-Caruso et al. (2020a).

Briefly, the WHONDRS consortium collected surface waters and sediments from 97 river corridors in 8 countries within a 6-week period, from 29 July to 19 September. At each location, collaborators selected sampling sites within 100 m of a station that measured river discharge, height, or pressure. Surface water was collected in triplicate using a 60 mL syringe and then filtered through a 0.22 μm sterivex filter (EMD Millipore) into a 40 mL glass vial (I-Chem amber VOA glass vials; ThermoFisher, pre-acidified with 10 μL of 85% phosphoric acid). Subsequently, 125 mL of surface sediments (1–3 cm depth) were sampled from a ~1 m^2 area at each of three depositional zone with a stainless steel scoop, making sure the sediments were saturated upon collection. All samples were shipped to Pacific Northwest National Laboratory on blue ice within 24 h of collection. Surface water samples were immediately frozen at −20 °C upon receiving. Sediments were sieved to <2 mm, subsampled into proteomic friendly tubes (Genesee Scientific), flash frozen in liquid nitrogen and then stored at -80 until FTICR-MS analysis. Note that in the methods provided by Garayburu-Caruso et al. (2020a) there is an error in the description of the sediment preservation prior to FTICR-MS analysis. Corrected preservation methods are used in this manuscript.

Prior to FTICR-MS analysis, sediment organic matter was extracted in proteomic friendly tubes (Genesee Scientific) with a 1:2 ratio of sediment to water (5 g of sediment to 10 mL of milli-Q water), continuously shaken in the dark at 375 rpm and 21 °C for 2 h. The tubes were centrifuged at 6000 rcf and 21 °C for 5 min. The supernatant was collected and filtered through 0.22 μm polyethersulfone membrane filter (Millipore Sterivex, USA) into borosilicate glass vials. Surface water and sediment extracts were normalized to a standardized NPOC concentration of 1.5 mg C L−1. Diluted samples were acidified to pH 2 with 85% phosphoric acid and extracted with PPL cartridges (Bond Elut), following Dittmar et al.(Dittmar et al., 2008)

A 12 Tesla (12 T) Bruker SolariX FTICR-MS (Bruker, SolariX, Billerica, MA, USA) located at the Environmental Molecular Sciences Laboratory in Richland, WA, was used to collect ultrahigh-resolution mass spectra of surface water and sediment OM pools. Resolution was 220 K at 481.185 m/z. The FTICR-MS was outfitted with a standard electrospray ionization (ESI) source, and data were acquired in negative mode with the voltage set to +4.2 kV. The instrument was externally calibrated weekly to a mass accuracy of <0.1 ppm; in addition, the instrument settings were optimized by tuning on a Suwannee River Fulvic Acid (SRFA) standard. Data were collected with an ion accumulation of 0.05 s for surface water and 0.1 or 0.2 s for sediment from 100–900 m/z at 4 M. One hundred forty-four scans were co-added for each sample and internally calibrated using an OM homologous series separated by 14 Da (–CH2 groups). The mass measurement accuracy was typically within 1 ppm for singly charged ions across a broad m/z range (100 m/z–900 m/z). BrukerDaltonik Data Analysis (version 4.2) was used to convert raw spectra to a list of m/z values by applying the FTMS peak picker module with a signal-to-noise ratio (S/N) threshold set to 7 and absolute intensity threshold to the default value of 100. We aligned peaks (0.5 ppm threshold) and assigned chemical formulas using Formularity.(Tolić et al., 2017) The Compound Identification Algorithm in Formularity was used with the following criteria: S/N > 7 and mass measurement error <0.5 ppm. This algorithm takes into consideration the presence of C, H, O, N, S, and P and excludes other elements.

The R package “ftmsRanalysis” (Bramer et al., 2020) was used to (1) remove peaks outside of a high confidence m/z range (200 m/z–900 m/z) and/or with a 13C isotopic signature and (2) to predict chemical class assignments for each DOM molecule using oxygen-to-carbon and hydrogen-to-carbon ratios (i.e., Van Krevelen classes, Kim et al., 2004). To yield a dataset of

globally ubiquitous DOM, surface water and sediment DOM pools were filtered to compounds occurring in 95% of samples.

Although FTICR-MS has the advantage of allowing for the detection of thousands of DOM molecules, one drawback is that formula assignments and follow-on chemical class inferences are computationally assessed rather than referenced against known standards. Because of this, chemical class inferences for DOM molecules in this study are limited to a higher level of molecular taxoDOMy than PyOM molecules that were extracted from primary literature and had known compound identifications. This results in a discrepancy between compound classes presented in the main text figures for DOM and PyOM molecules. We provide finer classifications for PyOM compounds (e.g., phenols), many of which fall into the broader groups depicted for DOM pools (e.g., lipids), to provide the maximum information we are able to infer from each data type.

Substrate-Explicit Model Description

We used a substrate-explicit modelling framework developed by Song et al.(Song et al., 2020) to characterize the bioavailability of each compound and predict its rate of decomposition. The model is compound-specific and environment-agnostic (with excepted specifications of 1 bar atmospheric pressure, pH 7, and 25ºC), meaning that it yields predictions for each input compound (as opposed to as a pool of compounds) and does not consider environmental conditions such as minerology or redox potential. Calculation of thermodynamic functions at pH=7 is important because aqueous species at pH=0 do not necessarily represent the state of biological cells. Correction of Gibbs free energy for any given reaction (*r*) from pH=0 (ΔG_r^0) to 7 (ΔG_r) can be made using the following equation:

$$\Delta G_r = \Delta G_r^0 + RTy_{H^+} \ln\left(10^{-7}\right) \quad \text{(Eq. 1)}$$

where R is the universal gas constant [= 0.008314 kJ/(K · mol)], T is temperature in Kelvin, y_{H^+} is the stoichiometric coefficient of H^+ in a given reaction. With this adjustment, the sign of Gibbs free energy for an electron donor half reaction is often changed from plus to minus.

Three sets of organic molecules were used as model inputs: global (1) surface water DOM and (2) sediment DOM pools (Garayburu-Caruso et al., 2020a); and (3) literature-derived PyOM compounds as described above. Inputs to the model were unique molecular formulae, grouped in subsequent analysis by their corresponding compound classes (Table S1). If one molecular formula was represented by several PyOM compounds (e.g., $C_{10}H_{16}O_2$, which corresponds to the sesquiterpenoid cis-Thujan-10-oic acid and 3-, 4- substituted methylcatechol phenols), we assigned multiple compound classes to that molecular formula.

Briefly, the substrate-explicit model uses the elemental stoichiometry of organic molecules, based on molecular formulae, to predict the number of catabolic reactions that must occur to provide the energy required for the synthesis of a unit carbon mole of biomass. This quantity is described by the parameter lambda (λ) in which lower λ values denote more efficient energetics of catabolism in producing biomass through anabolism. As described above, the model also predicts the Gibbs free energy of C oxidation (ΔG_{Cox}), under standard conditions with a modification to pH 7 adjusted from LaRowe and Van Cappellen (2011) by Song et al. (2020), as well as C use efficiency (CUE) as defined by Saifuddin et al. (2019) Lower ΔG_{Cox} denotes higher thermodynamic favorability in an electron donor half reaction associated with organic matter, and higher CUE reflects more C assimilated into biomass per unit C respired. We also predicted the rate of aerobic metabolism (as oxygen consumed per mol-C biomass produced) under three scenarios commonly observed in aquatic ecosystems: (a) C-limitation, (b) oxygen

(O_2) limitation, and (c) both C and O_2-limitation. For more details of the substrate-explicit modelling approach used, please see Song et al. (2020) Each metric (λ, ΔG_{Cox}, CUE, metabolic rates) denotes a different aspect of bioavailability. Though the relative magnitude of the metrics in comparison to each other will vary based on the specific stoichiometry of a molecule, highly bioavailable compounds are indicated by low λ and ΔG_{Cox} coinciding with high CUE and metabolic rates. More details on Song et al.'s substrate-explicit model are below, and we point the reader to the original publication for the full methodology.

The substrate-explicit model used here leverages two microbial parameters [maximal growth rate (μ_{max}) and harvest volume (V_h) (i.e., the volume that a microbe can access for harvesting energy)] to predict OM-specific oxidative degradation pathways and reaction rates based on the thermodynamic properties of OM pools. The remaining reaction kinetics are formulated from the chemical formula of OM, based on thermodynamic principles. The model is comprised of two major components (1) derivation of stoichiometric equations for catabolic, anabolic, and metabolic reactions by combining a set of standard thermodynamic analyses (Mccarty, 2007; Larowe and Van Cappellen, 2011; Kleerebezem and Van Loosdrecht, 2010) and (2) formulation of kinetic equations for the final oxidative degradation reaction of OC using a relatively recent thermodynamic theory for microbial growth.(Desmond-Le Quéméner and Bouchez, 2014)

For (1), we derived stoichiometric equations following the standard approaches outlined in the literature (Kleerebezem and Van Loosdrecht, 2010; Rittmann and Mccarty, 2001). Step-by-step instructions are available in Song et al. For each OM compound, we derived a stoichiometric equation for oxidative degradation of OC by combining catabolic (i.e., all processes for obtaining energy through substrate oxidation or other means) and anabolic (i.e.,

synthesis of biomass using the energy provided from catabolism) reactions to generate a full metabolic process. To do so, we combined to common approaches to generate metabolic reactions based on stoichiometric equations—the dissipation method (Hoijnen et al., 1992; Heijnen and Van Dijken, 1993) and the thermodynamic electron equivalents model (TEEM)(Mccarty, 2007). The dissipation method provides a basic framework through the determination of the stoichiometric coefficient vector for metabolic reaction by coupling the catabolic and anabolic reactions based on the parameter λ, which in turn was calculated by TEEM along with dissipation energy using information on C source and its conversion into biomass. In all cases, we specified ammonium as the nitrogen source.

To derive kinetic equations in step (2), we use thermodynamic theory by Desmond-Le Quemener and Bouchez (2014) to formulate microbial growth kinetics from stoichiometric equations in step (1). In the case of oxidative degradation of OC, the microbial growth on the *ith* OC (OC_i) can be represented by

$$\mu_i = \mu^{\max} \exp\left(-\frac{|y_{OC,i}|}{V_h\,[\mathrm{OC}_i]}\right) \exp\left(-\frac{|y_{O_2,i}|}{V_h\,[\mathrm{O_2}]}\right) \quad \text{(Eq. 2)}$$

where μ_{max} is the maximal specific growth rate, V_h is the volume that a microbe can access for harvesting energy from the environment (thus termed harvest volume), $y_{OC,i}$ and y_{O_2} are the stoichiometric coefficients of OC and O_2 in the metabolic reaction associated with oxidative degradation of OC_i, and $\left|y_{OC,i}\right|$ and $\left|y_{O_2}\right|$ denote their absolution values.

Importantly, the model utilizes molecular formulae to predict energetic content, metabolic efficiency, and rates of aerobic metabolism, and it does not account for structural components of organic molecules (e.g., double bonds, folding patterns, cross-linkages). It also is

agnostic of environmental parameters known to impact metabolic rates, such as temperature, mineral sorption, and microbial community composition. This simplified approach enables flexibility in application to high-throughput mass spectrometry techniques that yield chemical formulae but not structural information (e.g., Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, FTICR-MS) and supports hypothesis generation regarding in situ molecular dynamics that can be directly measured with targeted laboratory experiments. Despite its limitations, the substrate-explicit model used here has proven useful in linking DOM composition to aerobic metabolism in natural environments (Song et al., 2020; Graham et al., 2017; Garayburu-Caruso et al., 2020b), and its structure is consistent with Harvey et al.(Harvey et al., 2016) who argued for the importance of thermodynamic estimates of PyOM bioavailability that underlie this model. It was chosen to allow for comparison of PyOM to a comprehensive assessment of global aquatic DOM pool composition (Garayburu-Caruso et al., 2020a).

Statistics and Data Availability

We compared modelling outputs from representative PyOM to outputs of ubiquitous DOM pools to infer relative bioavailability using ANOVA and Tukey HSD statistical tests with R software. All model outputs are available in Tables S2-S4. Code is available at: https://github.com/hyunseobsong/lambda. Data describing DOM pool chemistry are published as a data package (Goldman et al., 2020) (available at: doi:10.15485/1729719) and are discussed in more detail by Garayburu-Caruso et al. (2020a).

SUPPORTING INFORMATION: FIGURES AND TABLES

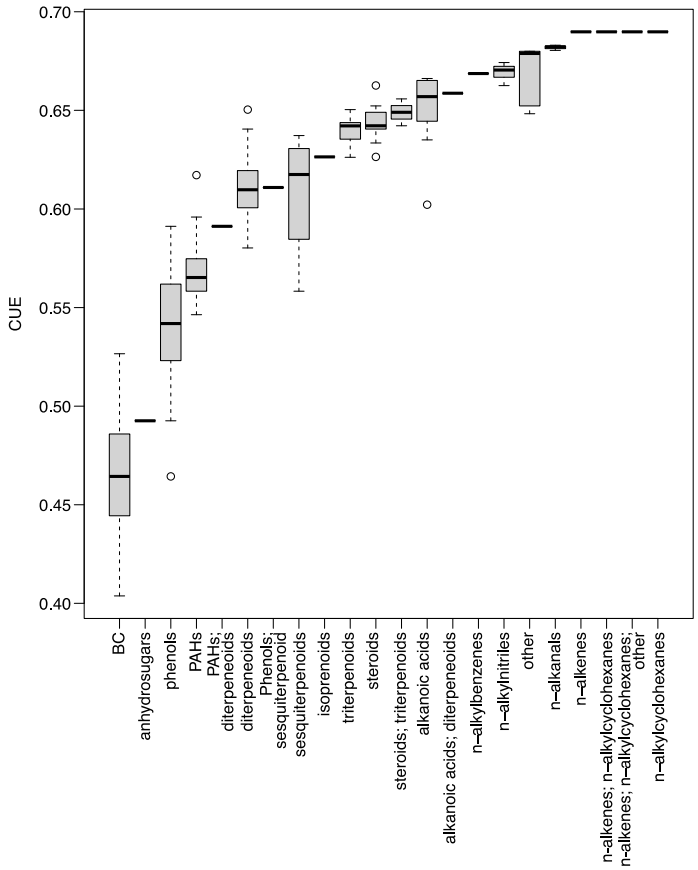

Figure S1. Carbon use efficiency (CUE) of PyOM, grouped by known chemical properties. High CUE values indicate more C incorporated into biomass vs. respired per unit C. Median values are denoted by a bar, the lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the upper and lower whiskers extend from the hinge to the largest/smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles), and data beyond the end of the whiskers are plotted individually.

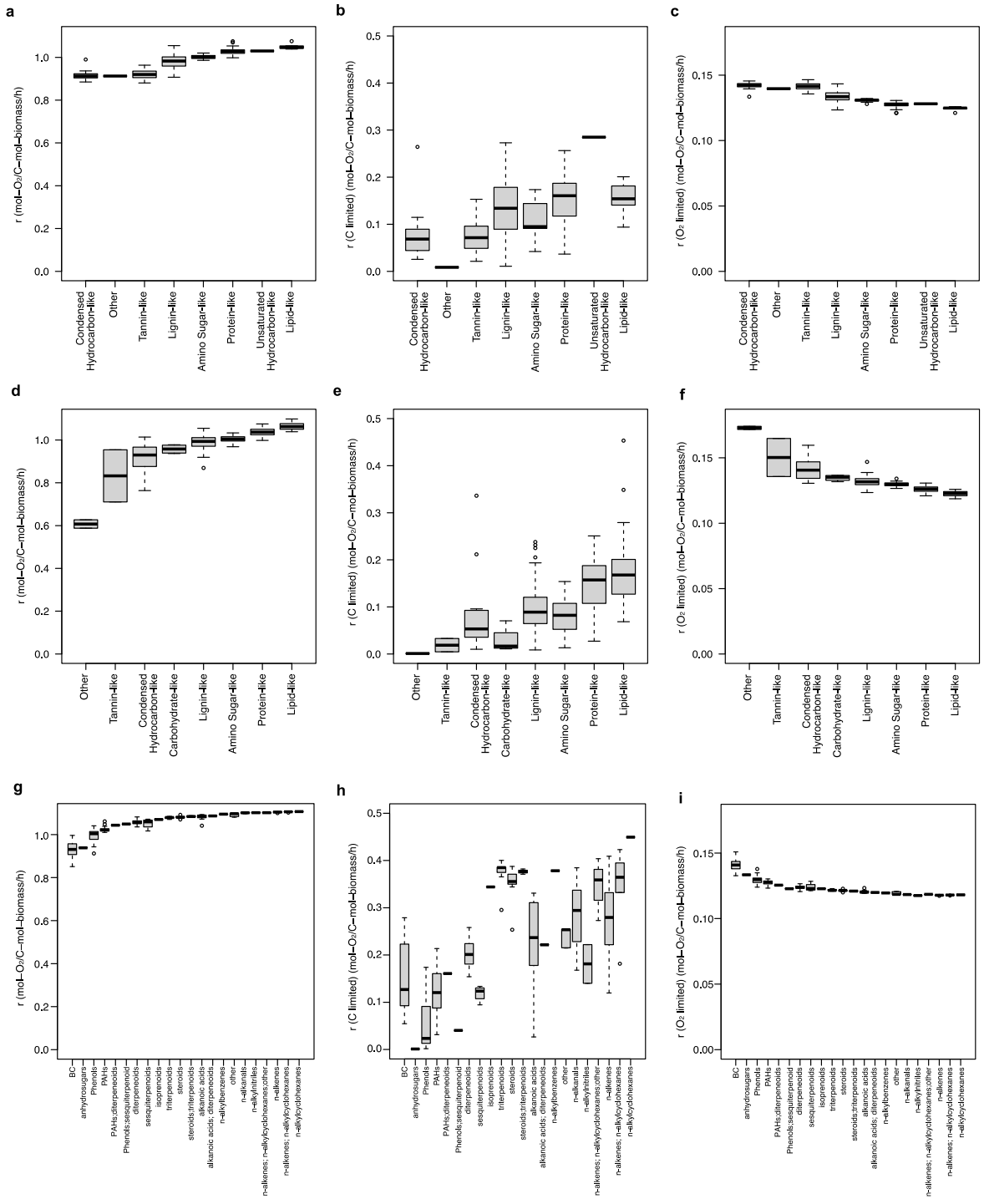

Figure S2. Predicted metabolism of ubiquitous (a–c) surface water DOM, (d–f) sediment DOM, and (g–i) PyOM molecules, grouped by inferred chemical class or known chemical properties. The first column shows predicted rates of aerobic metabolism without any elemental limitations.

C-limited and O_2-limited scenarios are show in the second and third columns, respectively. Median values are denoted by a bar, the lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the upper and lower whiskers extend from the hinge to the largest/smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles), and data beyond the end of the whiskers are plotted individually.

Table S1. PyOM molecules and chemical properties.

Table S2. Substrate-explicit model outputs for each PyOM compound.

Table S3. Substrate-explicit model outputs for each sediment DOM compound.

Table S4. Substrate-explicit model outputs for each surface water DOM compound.

References.

Bramer, L. M., White, A. M., Stratton, K. G., Thompson, A. M., Claborne, D., Hofmockel, K., and McCue, L. A.: ftmsRanalysis: An R package for exploratory data analysis and interactive visualization of FT-MS data, PLoS computational biology, 16, e1007654, 2020.
Desmond-Le Quéméner, E. and Bouchez, T.: A thermodynamic theory of microbial growth, The ISME journal, 8, 1747-1751, 2014.
Dittmar, T., Koch, B., Hertkorn, N., and Kattner, G.: A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater, Limnology and Oceanography: Methods, 6, 230-235, 2008.
Garayburu-Caruso, V. A., Danczak, R. E., Stegen, J. C., Renteria, L., Mccall, M., Goldman, A. E., Chu, R. K., Toyoda, J., Resch, C. T., and Torgeson, J. M.: Using Community Science to Reveal the Global Chemogeography of River Metabolomes, Metabolites, 10, 518, 2020a.
Garayburu-Caruso, V. A., Stegen, J. C., Song, H.-S., Renteria, L., Wells, J., Garcia, W., Resch, C. T., Goldman, A. E., Chu, R. K., Toyoda, J., and Graham, E. B.: Carbon limitation leads to thermodynamic regulation of aerobic metabolism, Environmental Science & Technology Letters, 7, 517-524, 2020b.
Goldman, A. E., Chu, R. K., Danczak, R. E., Daly, R. A., Fansler, S., Garayburu-Caruso, V. A., Graham, E. B., McCall, M. L., Ren, H., and Renteria, L.: WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Sediment FTICR-MS, NPOC, and Aerobic Respiration, Environmental System Science Data Infrastructure for a Virtual Ecosystem …, 2020.
Graham, E. B., Tfaily, M. M., Crump, A. R., Goldman, A. E., Bramer, L. M., Arntzen, E., Romero, E., Resch, C. T., Kennedy, D. W., and Stegen, J. C.: Carbon inputs from riparian vegetation limit oxidation of physically bound organic carbon via biochemical and thermodynamic processes, Journal of Geophysical Research: Biogeosciences, 122, 3188-3205, 2017.
Harvey, O. R., Myers-Pigg, A. N., Kuo, L.-J., Singh, B. P., Kuehn, K. A., and Louchouarn, P.: Discrimination in degradability of soil pyrogenic organic matter follows a return-on-energy-investment principle, Environmental science & technology, 50, 8578-8585, 2016.
Heijnen, J. and Van Dijken, J.: Response to comments on “in search of a thermodynamic description of biomass yields for the chemotropic growth of microorganisms”, Biotechnology and Bioengineering, 42, 1127-1130, 1993.
Hoijnen, J., van Loosdrecht, M. C., and Tijhuis, L.: A black box mathematical model to calculate auto-and heterotrophic biomass yields based on Gibbs energy dissipation, Biotechnology and bioengineering, 40, 1139-1154, 1992.
Kim, S., Kaplan, L. A., Benner, R., and Hatcher, P. G.: Hydrogen-deficient molecules in natural riverine water samples—evidence for the existence of black carbon in DOM, Marine Chemistry, 92, 225-234, 2004.
Kleerebezem, R. and Van Loosdrecht, M. C.: A generalized method for thermodynamic state analysis of environmental systems, Critical Reviews in Environmental Science and Technology, 40, 1-54, 2010.
LaRowe, D. E. and Van Cappellen, P.: Degradation of natural organic matter: a thermodynamic analysis, Geochimica et Cosmochimica Acta, 75, 2030-2042, 2011.

McCarty, P. L.: Thermodynamic electron equivalents model for bacterial yield prediction: modifications and comparative evaluations, Biotechnology and bioengineering, 97, 377-388, 2007.
Oros, D. R. and Simoneit, B. R.: Identification and emission factors of molecular tracers in organic aerosols from biomass burning Part 1. Temperate climate conifers, Applied Geochemistry, 16, 1513-1544, 2001a.
Oros, D. R. and Simoneit, B. R.: Identification and emission factors of molecular tracers in organic aerosols from biomass burning Part 2. Deciduous trees, Applied Geochemistry, 16, 1545-1565, 2001b.
Rittmann, B. E. and McCarty, P. L.: Environmental biotechnology: principles and applications, McGraw-Hill Education2001.
Saifuddin, M., Bhatnagar, J. M., Segrè, D., and Finzi, A. C.: Microbial carbon use efficiency predicted from genome-scale metabolic models, Nature communications, 10, 1-10, 2019.
Song, H.-S., Stegen, J. C., Graham, E. B., Lee, J.-Y., Garayburu-Caruso, V., Nelson, W. C., Chen, X., Moulton, J. D., and Scheibe, T. D.: Representing Organic Matter Thermodynamics in Biogeochemical Reactions via Substrate-Explicit Modeling, Frontiers in Microbiology, https://doi.org/10.3389/fmicb.2020.531756, 2020.
Tolić, N., Liu, Y., Liyu, A., Shen, Y., Tfaily, M. M., Kujawinski, E. B., Longnecker, K., Kuo, L.-J., Robinson, E. W., and Paša-Tolić, L.: Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra, Analytical chemistry, 89, 12659-12665, 2017.
Wagner, S., Ding, Y., and Jaffé, R.: A new perspective on the apparent solubility of dissolved black carbon, Frontiers in Earth Science, 5, 75, 2017.