

SCIENTIFIC REPORTS



OPEN

Yellowstone Hot Springs are Organic Chemodiversity Hot Spots

Michael Gonsior¹, Norbert Hertkorn², Nancy Hinman³, Sabine E.-M. Dvorski², Mourad Harir², William J. Cooper⁴ & Philippe Schmitt-Kopplin^{2,5}

Yellowstone National Park hydrothermal springs were investigated according to their organic geochemistry with a special focus on the Yellowstone hot spring dissolved organic matter (YDOM) that was solid-phase extracted. Here we show that YDOM has a unique chemodiversity that has not yet been observed anywhere else in aquatic surface environments and that Yellowstone hot springs are organic chemodiversity hot spots. Four main geochemically classified hot spring types (alkaline-chloride, mixed alkaline-chloride, acid-chloride-sulfate and travertine-precipitating) exhibited distinct organic molecular signatures that correlated remarkably well with the known inorganic geochemistry and manifested themselves in excitation emission matrix fluorescence, nuclear magnetic resonance, and ultrahigh resolution mass spectra. YDOM contained thousands of molecular formulas unique to Yellowstone of which 80% contained sulfur, even in low hydrogen sulfide containing alkaline-chloride springs. This unique YDOM reflects the extreme organic geochemistry present in the hydrothermal features of Yellowstone National Park.

Physical, chemical, and biological processes control the evolution of minerals and dissolved organic matter (DOM) over a vast range of mass, length, and time¹. Indeed, mutual interferences and interactions among these processes presumably have defined chemical and biological evolution throughout Earth history. Modern terrestrial biology operates largely within limited temperature, pH, and redox conditions. In contrast, extreme environments, such as hot springs, have wide ranges of physicochemical parameters, generally believed to be similar to environments found on early Earth². Since then, chemical, mineral, and especially biological processes have shaped organic matter diversity on Earth¹. Yet, little is known about the organic composition, specifically the chemical diversity of DOM in geothermal systems, despite some characterization of marine hydrothermal waters^{3,4}. In the case of some Yellowstone hot springs, DOM, such as petroleum products⁵, or microbial metabolites such as lipids⁶ have been previously identified. The incorporation of sulfur into DOM also needs to be considered because of potential reactions between DOM and hydrogen sulfide that is often present in hydrothermal features, and because hydrosulfurization of DOM has been previously suggested in hydrothermal vents³. However, dissolved organic sulfur (DOS) in Yellowstone hot springs has not been characterized to date. The dissolved organic nitrogen (DON) also remains largely uncharacterized at the structural level in aquatic systems, despite the characterization of small biomolecules such as urea⁷ and dissolved combined amino acids (DCAA)⁸, which only account for about 10% of the DON pool. The often used solid-phase extraction of DOM also appears to have a limited extraction efficiency for DON, which makes it difficult to comprehensively evaluate DON. However, DON extraction efficiencies are not known for DON in hydrothermal systems. To the best of our knowledge, no data exists on any DON components in hydrothermal systems.

The Yellowstone Plateau, the youngest in a line of volcanic calderas stretching along the length of the Snake River Plain and dating back 17 million years⁹, exhibits abundant and diverse hydrothermal features that are ideal for studying DOM in extreme environments. The caldera hosts an extensive hydrothermal system, which cools the shallow underlying magma body^{10,11}. The precise details of the magmatic system are still debated, but recent reports suggest that an underlying basaltic magma heats an overlying rhyolitic magma¹¹. The shallower, rhyolitic magma provides the heat that drives hydrothermal circulation. The contributions of this thermal exchange to the inventory of inorganic salts in the hydrothermal system have recently been better established. A caldera-wide

¹University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory, Solomons, Maryland, USA. ²Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany. ³University of Montana, Department of Geosciences, Missoula, USA. ⁴University of California Irvine, Department of Civil and Environmental Engineering, Irvine, USA. ⁵Technische Universität München, Institute for Analytical Food Chemistry, Freising-Weißenstephan, Germany. Correspondence and requests for materials should be addressed to M.G. (email: gonsior@umces.edu) or P.S.-K. (email: Schmitt-kopplin@helmholtz-muenchen.de)

analysis of river composition and discharge¹² suggested that ~40% of the dissolved solids are derived from acid hydrolysis of country rock by acidic magmatic gases. Still, variable concentrations of gases and solids are observed in the numerous thermal features (hot springs, geysers, mud pots, fumaroles, and acid lakes) that are spread throughout the caldera (e.g. Firehole drainage thermal areas) and just outside its margins (e.g. Norris Geyser Basin), suggesting differences in sources of materials and in processes of fluid and gas evolution. The intra-caldera hydrothermal system may serve to enhance adjacent thermal systems at Mammoth Hot Springs, north of the main hydrothermal system¹³. Extra-caldera hydrothermal systems would be expected to have different sources of dissolved constituents.

In 1966, Thomas Brock discovered microorganisms in the boiling hot springs, and their outflow channels, of Yellowstone National Park¹⁴. This remarkable observation raised many new questions. Among these was the question of what carbon and energy sources are available to support heterotrophic extremophiles. Furthermore, *in situ* metabolic processes of the autotrophic and heterotrophic microbial communities in hot springs are largely not understood, despite advances in our understanding of the microbial community composition^{15–17} as well as the inorganic geochemical and mineralogical composition of the thermal waters¹⁰. This is further compounded by our fundamental lack of understanding of the diversity and abundance of organic compounds in these springs.

Hawke *et al.*¹⁸ showed that DOM is largely lost during geothermal heating but did not address changes in organic chemical composition in the remaining material. In Yellowstone hot springs, DOM is leached from deeply buried sedimentary rock⁵, further confounding the sources and composition of DOM. Different subsurface flow paths for heated water and gases define the inorganic geochemical properties of individual springs¹⁰, likely influencing DOM composition as well. Herein, we address the question of chemodiversity of hot spring DOM and approach the possible processes responsible for the specific composition of Yellowstone hot spring DOM (YDOM). A classification of types of hot springs based on the YDOM composition is also evaluated.

Results

Inorganic and organic geochemical comparison of hot spring types. The selected 10 Yellowstone hot springs were analyzed with respect to their inorganic geochemistry and physical properties and clustered into four compositional groups: travertine-precipitating, mixed-alkaline-chloride, alkaline-chloride, and acid-chloride-sulfate springs (Fig. 1 and Supplementary Table. S1). Out of these, four representative hot springs were investigated in detail with respect to DOM composition, namely Narrow Gauge (NG), Mammoth Hot Spring Complex (travertine-precipitating); Rabbit Creek (RC1), Midway Geyser Basin (mixed (terrestrial) alkaline-chloride); Elk Geyser (EG), Norris Geyser Basin (acid-chloride-sulfate); and Octopus Spring (OS), Lower Geyser Basin (alkaline-chloride) (Fig. 1). To describe in detail this DOM, we used electrospray ionization Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry and high field nuclear magnetic resonance (NMR) spectroscopy. It should be noted here that solid-phase extracted material was used for FT-ICR MS and NMR analyses and hence very hydrophilic compounds are not extracted and lost. Furthermore, the analytical window of FT-ICR MS is biased towards the strongest ionizing components in a complex mixture and hence is likely not representative of the whole DOM pool. Nevertheless, these techniques are information-rich methods to accurately depict the molecular diversity of polydisperse and molecularly heterogeneous DOM by identifying the ionizable composition (FT-ICR) and structure (NMR) of solid-phase extractable DOM^{19,20}. Additionally, Excitation Emission Matrix (EEM) fluorescence was used to describe the fluorophores present in these springs.

YDOM composition in hot spring types. These complementary techniques demonstrated unprecedented organic molecular diversity of DOM in individual Yellowstone hot springs (Fig. 2 and Supplementary Figs S1–S8). FT-ICR MS, with its excellent mass resolution and mass accuracy, allows unambiguous assignment of thousands of molecular formulas directly out of complex mixtures, including determination of molecular formulas containing heteroatoms like nitrogen and sulfur²¹. Results from FT-ICR MS showed that organic molecular compositions statistically clustered with the grouping of the 10 springs by inorganic chemistry (Fig. 1) yet also demonstrated distinct DOM compositions in the four individual springs studied in more detail (Fig. 2).

NMR offers quantitative and non-destructive determination of chemical environments for carbon and hydrogen and is particularly useful in depicting aliphatic groups based on sp³-hybridized carbon. NMR results documented that the structure of YDOM is substantially different compared to that of boreal lakes (Fig. 3) and other aquatic systems^{19,22,23}.

EEM fluorescence spectroscopy has been frequently used to describe the chromophoric dissolved organic matter in aquatic systems^{24–27}. Remarkably, each spring type contained different fluorophores that were also not found in other surface waters (Supplementary Fig. S8). Fluorophores are in general indicative of conjugated aromatic π -electron systems that absorb in the ultraviolet and emit light in the visible spectrum. To further describe the unique nature of YDOM, results from each analytical technique are presented in greater detail below.

YDOM characterization using non-target FTICR Mass Spectrometry. Ultrahigh-resolution FT-ICR mass spectra of the four YDOM from the representative springs (NG, EG, RC1, and OS) provided several thousand *m/z* peaks (Supplementary Fig. S1) of which many were assigned to extended molecular formula series containing atomic combinations of carbon, hydrogen, and oxygen (CHO), containing nitrogen (CHNO), containing sulfur (CHOS), and containing both sulfur and nitrogen (CHNOS) (Fig. 2, Supplementary Table. S2). The precisely determined exact molecular formulas (to < 0.0001 amu from theoretical formula) represented a remarkably wide coverage of the compositional space that is shown here in van Krevelen diagrams or elemental plots, where the oxygen to carbon ratio (O/C) is plotted against the hydrogen to carbon ratio (H/C) of each individual molecular formula (Supplementary Figs S2, S3). We define chemodiversity by using the coverage of assigned molecular formulas in van Krevelen space, the molecular weight distribution and numbers of assigned formulas and hence van Krevelen and mass-edited H/C plots can be used to visualize the very high chemodiversity of the YDOM.

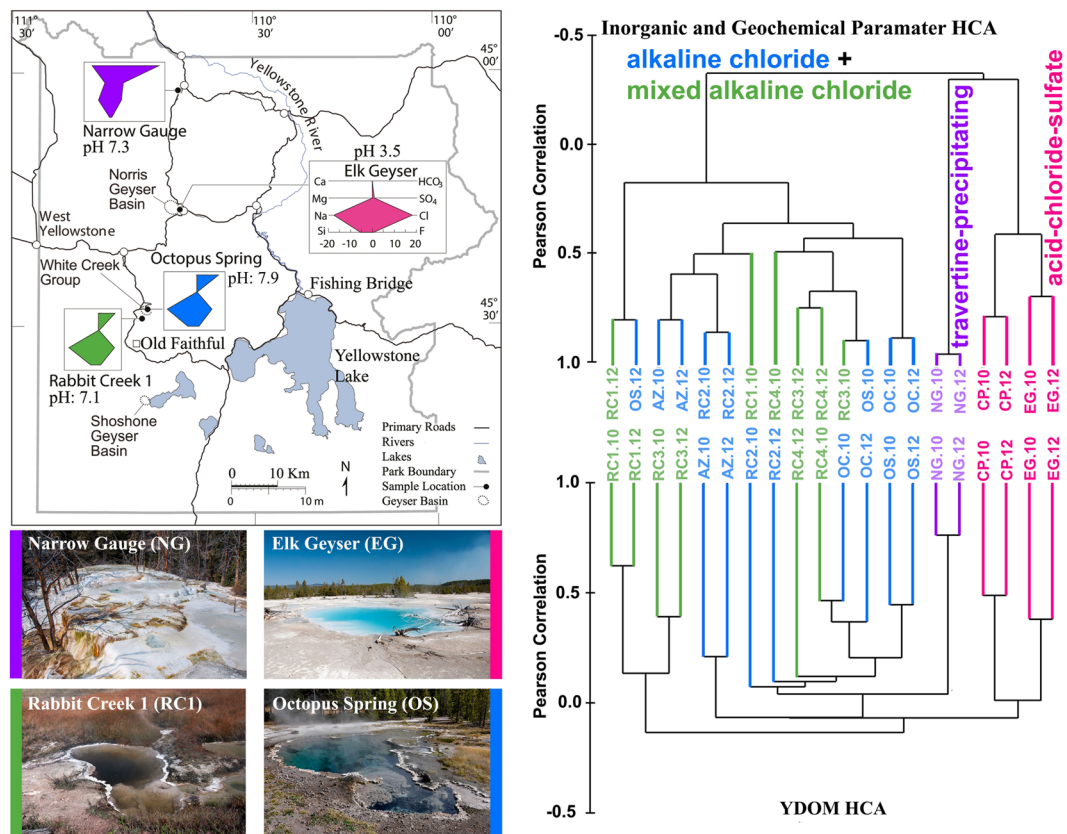


Figure 1. Map of Yellowstone National Park including stiff diagrams of key inorganic ion abundances in hot springs (top, left); photographs of Narrow Gauge Spring (NG), Rabbit Creek (RC1), Elk Geyser (EG) and Octopus Spring (OS) (bottom, left); hierarchical cluster analyses (HCA) of inorganic and geochemical parameters (top, right) (see also Supplementary Table. S1); and YDOM (bottom, right). Note: additional Yellowstone hot springs were used for hierarchical cluster analyses: Azure (AZ), Ojo Caliente (OC), Rabbit Creek 3 and 4 (RC3 and RC4), and Cinder Pool (CP). Details about all springs can be found in the supplementary online material. Replicate samples were collected in 2010 and 2012 (e.g. NG.10 and NG.12). All photographs were taken under Yellowstone Research Permit YELL-2017-SCI-5828.

In essence, mixed-alkaline-chloride and alkaline-chloride springs (RC1, OS) had diverse unsaturated or aromatic ($H/C < 1.5$) molecular ions, but also saturated CHOS molecular formulas up to 800 m/z (Fig. 2 and Supplementary Figs S2–S4), whereas travertine-precipitating hot springs (NG) were enriched in hydrogen-deficient nitrogen and sulfur-containing low molecular weight molecules ($H/C < 1.5$; $m/z < 500$) with limited oxygenation ($O/C < 0.5$). Acid-chloride-sulfate springs (EG) had the widest diversity of unique molecular ions from small to intermediate m/z ($150 < m/z < 650$; Fig. 2 and Supplementary Figs S2–S4). The hot springs NG and EG had the most sulfur-containing molecular formulas with a remarkable coverage of the chemical space shown within the van Krevelen diagrams (Fig. 2 and Supplementary Figs S2–S4). Spring RC1 had the highest number of nitrogen-containing molecular ions, which were mostly unsaturated and/or aromatic in nature ($H/C < 1.5$).

Hierarchical clustering of the FT-ICR MS-derived YDOM chemical compositions (Fig. 1, bottom) showed statistically significant correlation with the previously well-documented inorganic chemistry of geothermal-water types (e.g.¹⁰; Fig. 1 and Supplementary Table. S1). The inorganic constituents derive from the interaction of hydrothermal water with country rock, coupled with retention or release of the water-soluble gases, CO₂ and H₂S, the latter of which oxidizes to form H₂SO₄. The observed congruence of inorganic and organic compositions suggested a decisive influence of physico-chemical and mineral conditions on the synthesis of YDOM, which may be causal in that the inorganic constituents are intimately involved in transformations of YDOM and which might be further mediated by microbiological processes¹⁷. Consequently, the four representative springs contain distinctly different YDOM (Supplementary Figs S2 and S3).

To compare YDOM to the composition of other types of DOM from conventional aquatic systems, we contrasted FT-ICR mass spectra of our representative set of four hot spring YDOM samples with an extensive compendium of lake, river, estuarine, and marine DOM ($n = 114$) (Figs. 3 and 4) that had been isolated by the same solid-phase extraction procedure (SPE-DOM) and analyzed to the same m/z resolution. We found that the consolidated YDOM data set and that of aquatic surface waters, which both showed $>10^4$ assigned molecular formulas, were largely dissimilar. For example, YDOM from the four geochemical systems showed 5,405 unique

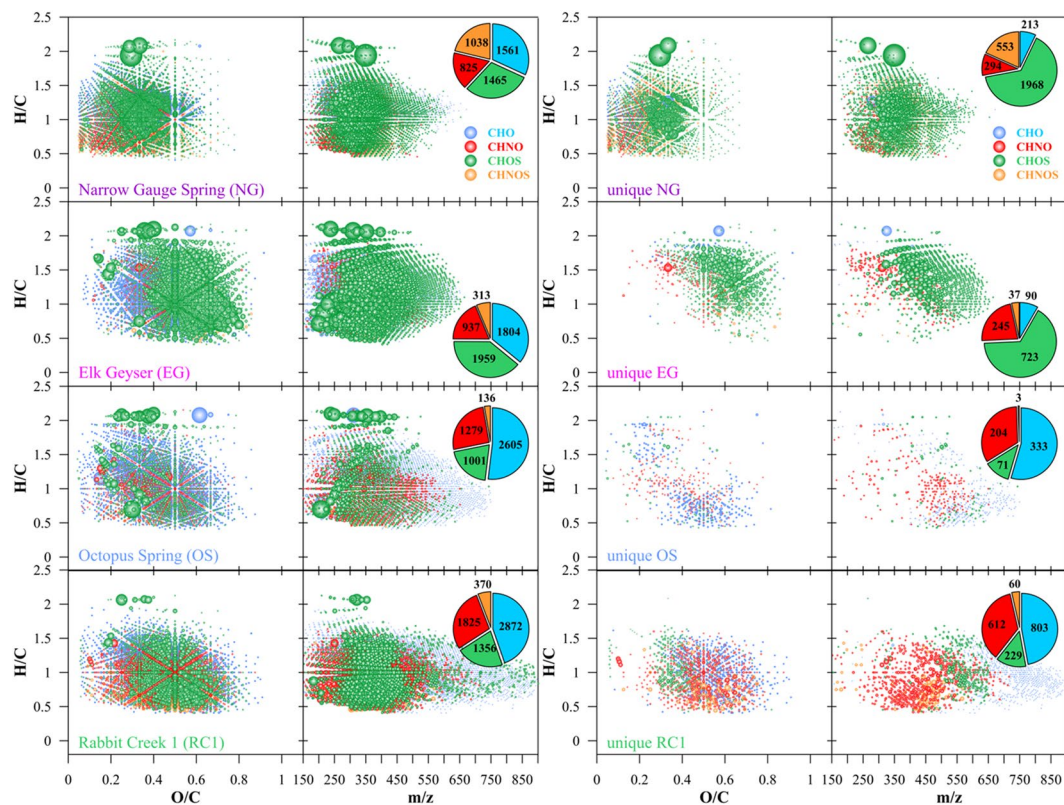


Figure 2. Detailed FT-ICR MS results (van Krevelen diagrams and mass-edited H/C ratio plots) for representative springs of each of the four compositional groups (see also clusters in Fig. 1): Narrow Gauge Spring (NG, travertine-producing), Octopus Spring (OS, alkaline chloride), Elk Geysir (EG, acid-chloride-sulfate) and Rabbit Creek (RC1, mixed alkaline chloride) and their unique molecular signatures, respectively. Note: Bubble area reflects relative abundance of m/z ions; color code of CHO, CHNO, CHOS and CHNOS molecular series according to figures.

molecular formulas that were not present in the consolidated data set for other aquatic SPE-DOM (Figs 3, 4 and Supplementary Table S3). Furthermore, the majority of the unique YDOM molecular formulas contained heteroatoms such as nitrogen and sulfur (~9% CHNO; ~52% CHOS; ~28% CHNOS compounds; Fig. 4).

YDOM characterization using Nuclear Magnetic Resonance (NMR) Spectroscopy. Chemo-diversity of YDOM can also be expressed using quantitative and structural NMR spectroscopy. For comparison with YDOM, we used boreal lake SPE-DOM (collected in central Sweden, Malingsbo region) because boreal lake DOM is largely representative of diverse aquatic surface DOM^{19,22,28,29}. Boreal lake SPE-DOM had a continuous distribution of broad ¹H NMR resonances (Fig. 3) that reflected massive superposition of $>10^6$ individual atomic environments found in common DOM^{19,22,30}. In contrast, YDOM showed better-resolved groups of NMR resonances across the entire chemical shift range and considerable distinction among the four selected hot springs (Supplementary Fig. S4). Hence, ¹H NMR spectra supported the chemical diversity observed in FT-ICR MS and confirmed a remarkable structural diversity within YDOM throughout all classes of chemical bonds (Supplementary Figs S4–S7, Tables S5–S7; see detailed NMR discussion in Supplementary text) along with fundamentally different structural distributions when compared to common aqueous and boreal lake SPE-DOM.

Aliphatic and aromatic structures were observed to varying degrees in YDOM. The four hot springs had significant amounts of aliphatic structures, although the configuration varied between them. YDOM from RC1 and NG was enriched in C_{3–5} branched, open-chain aliphatic compounds, terminated by methyl and carboxylic groups, whereas alicyclic rings such as refractory carboxyl-rich alicyclic molecules (CRAM)²⁸ dominated OS and especially EG (Supplementary Fig. S4), suggesting different formation mechanisms and/or precursor composition. Aromatic functional groups (C_{ar}H) with a chemical shift in δ_H of 9–6.5 ppm in YDOM fell into two groups representing aromatic ethers and ketones (Supplementary Figs S5b, S6b). C_{ar}H shifts were also different from boreal SPE-DOM, reinforcing the unique character of YDOM. EG had a near Gaussian distribution of aromatic NMR resonances, indicating an even abundance of electron-withdrawing, neutral, and electron-donating functional groups (Supplementary Figs S5, S5b, S6b). A ramp-like increase of aromatic hydrogen from lower to higher chemical shift δ_H indicated abundant polyphenolic compounds and/or aromatic ethers decreasing in the order OS \gg RC1 $>$ NG, suggesting variable contributions from terrestrial organic matter and (poly)phenols (Supplementary Figs S4, S5b, S6b).

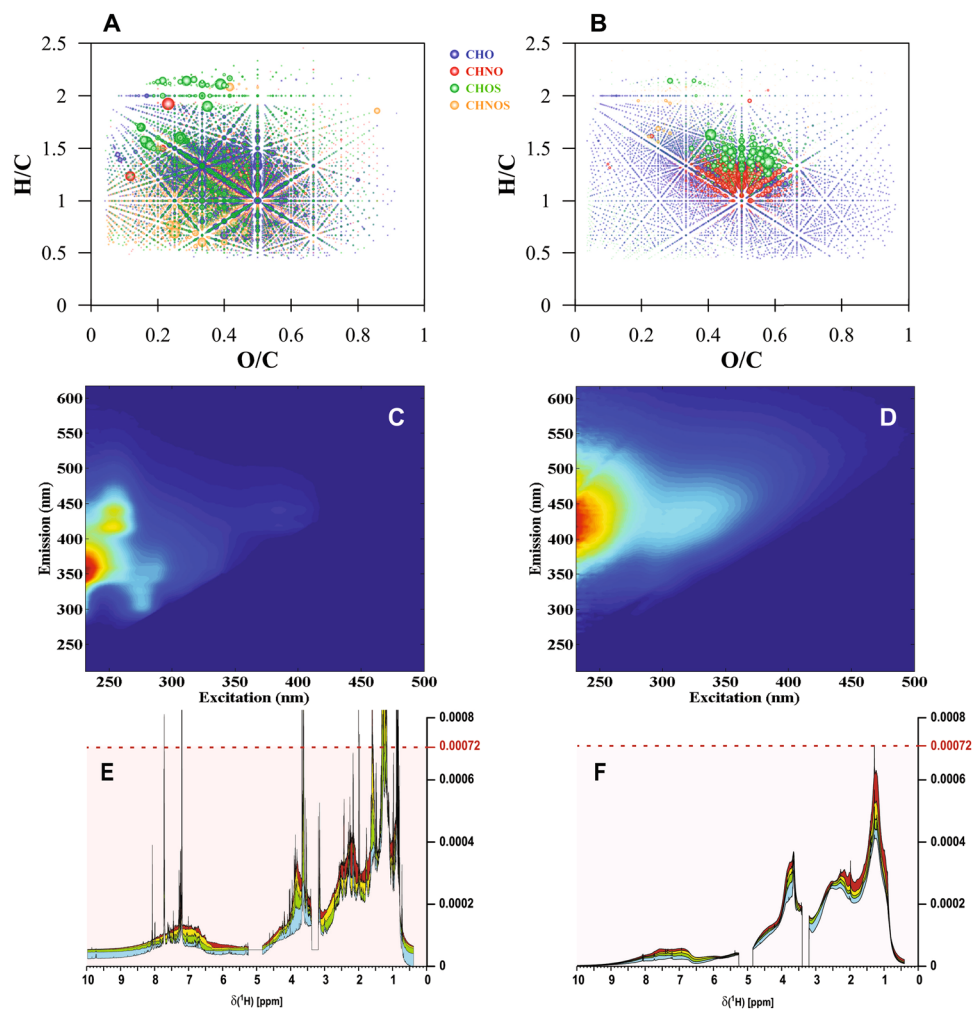


Figure 3. Van Krevelen diagrams of four (NG, EG, RC1, OS) consolidated YDOM samples (A) compared to consolidated solid phase extracted (SPE)-DOM samples from 114 different aquatic systems (B) (see details in supplementary online material). Consolidated YDOM EEM spectrum (C) and ^1H -NMR spectrum in comparison to a consolidated EEM spectrum in ultrapure water and ^1H -NMR data (F) of four Swedish boreal lake SPE-DOM samples. Note: Bubble area in van Krevelen diagrams reflects relative abundance of m/z ions.

In general, aliphatic functional group abundance and diversity ($\delta_{\text{H}} < \sim 3$ ppm) decreased with decreasing pH (Supplementary Figs S5a, S6a, S7). Despite expected hydrolytic degradation of methoxy groups and oxomethylene structures in aliphatic compounds at elevated temperatures, such groups were markedly diverse in YDOM (Supplementary Fig. S8). Overall, the proportion of open-chain, branched aliphatic compounds decreased in the order $\text{RC1} > \text{NG} > \text{OS} > \text{EG}$.

Expected small contributions of heteroatomic functional groups are more difficult to determine in complex ^{13}C and ^1H NMR spectra, however EG also showed a unique set of presumably thiomethyl groups that is highlighted in circle b in Supplementary Fig. S6a.

Overall, NMR results are in very good agreement with results presented by using FT-ICR MS and support the conclusion that YDOM is unique when compared to other surface aquatic systems and that each individual spring type contains distinctly different YDOM.

YDOM characterization using Excitation Emission Matrix (EEM) Fluorescence. The optically active fraction of YDOM also showed unique chemical signatures compared to boreal lake SPE-DOM (Fig. 3 and Supplementary Fig. S8) and other very diverse aquatic systems^{26,31–33}. Maxima of the consolidated YDOM EEM fluorescence peaks of the combined four springs were located at very different excitation/emission couples (Fig. 3). The highest intensity local maximum of the consolidated YDOM dataset was located at an excitation|emission (Ex|Em) couple 230 nm|355 nm, which has not been observed in any aquatic surface SPE-DOM samples. However, to describe different fluorophores, individual EEM spectra need to be used and several distinct fluorescence peaks were observed in individual YDOM springs (Supplementary Fig. S8) (NG: Ex|Em = 240|360 nm, 260|420 nm; EG: Ex|Em = 230|310 nm, 230|410 nm; RC1: Ex|Em = 230|310 nm, OS: Ex|Em = 230|310 nm), representing rather distinct fluorophores of yet unknown provenance, except for peak 260|420 nm in NG, which is similar to the previous defined A peak or humic-like fluorescence³⁴. It appears that

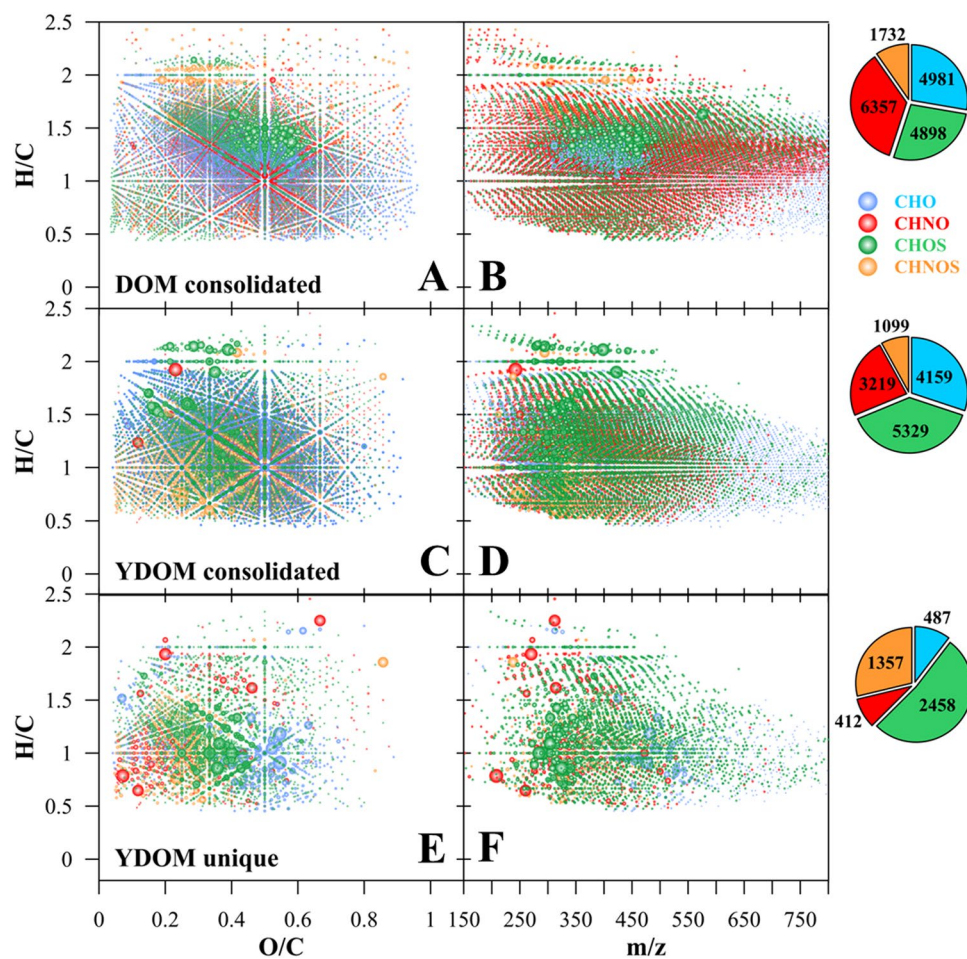


Figure 4. Comparison of consolidated SPE-DOM molecular signatures (A,B) analyzed by non-targeted FT-ICR MS and collected on a large spatial scale globally and consolidated YDOM signatures (C,D), including the unique signatures isolated from this comparison (E,F). Left column (A,C,E): van Krevelen diagrams; right column (B,D,F): mass-edited H/C ratio plot; pie diagrams: counts of assigned elemental compositions. Note: Bubble area reflects relative abundance of m/z ions.

there might be similar fluorophores present in EG, RC1 and OS with additional fluorophores at higher emission wavelengths in EG. NG appeared to be very different from all other springs, which may reflect the different source of the hydrothermal water. In agreement with NMR and FT-ICR mass spectra, the EEM fluorescence spectra of the four individual hot springs also showed clear distinction between these representative hot springs (Supplementary Fig. S8) and clear differences to SPE-DOM collected from diverse surface waters.

Unique Chemodiversity of YDOM. Combined results from these complementary analytical techniques demonstrated that the extensive compositional diversity as well as the structural specificity of YDOM make Yellowstone National Park hot springs DOM chemodiversity hotspots and further, that their DOM is largely indigenous to the geochemical types of hot springs. DOM arising from thermophiles will contribute to the observed chemodiversity. However, the altogether limited microbial abundance, and the all in all restricted microbial biodiversity observed in these extreme environments (Supplementary Table S4) likely cannot account for the observed molecular complexity of YDOM. This is also supported by the EEM fluorescence data, because only very weak signals that indicated the presence of proteins were found, which typically show a fluorescence signal at higher excitation wavelengths of either $Ex|Em = 275|300$ nm (tyrosine) or $Ex|Em = 280|345$ nm (tryptophan)³⁵. Therefore, alternative sources of DOM must contribute to the diversity of organic molecules. While atmospheric deposition into small surface area springs in the pristine environments of Yellowstone National Park will likely be negligible, contributions from small molecules from deep thermal waters (>350 °C) are clearly conceivable. At these elevated temperatures and pressures, complex DOM is likely to decompose into small, and often oxygenated, molecules¹⁸. A proportion of this thermally transformed DOM is expected to be non-extractable¹⁸ and hence would escape our analysis in this study. Surface water runoff and groundwater are possible additional sources of DOM to these springs (Fig. 5).

While surface water DOM would resemble common SPE-DOM of rivers and lakes, groundwater may become severely transformed before entering the spring aquifers, depending on underground exposure to temperature, pressure, gases (e.g. HCl, H₂S, SO₂, NH₃) and minerals (e.g. clays,

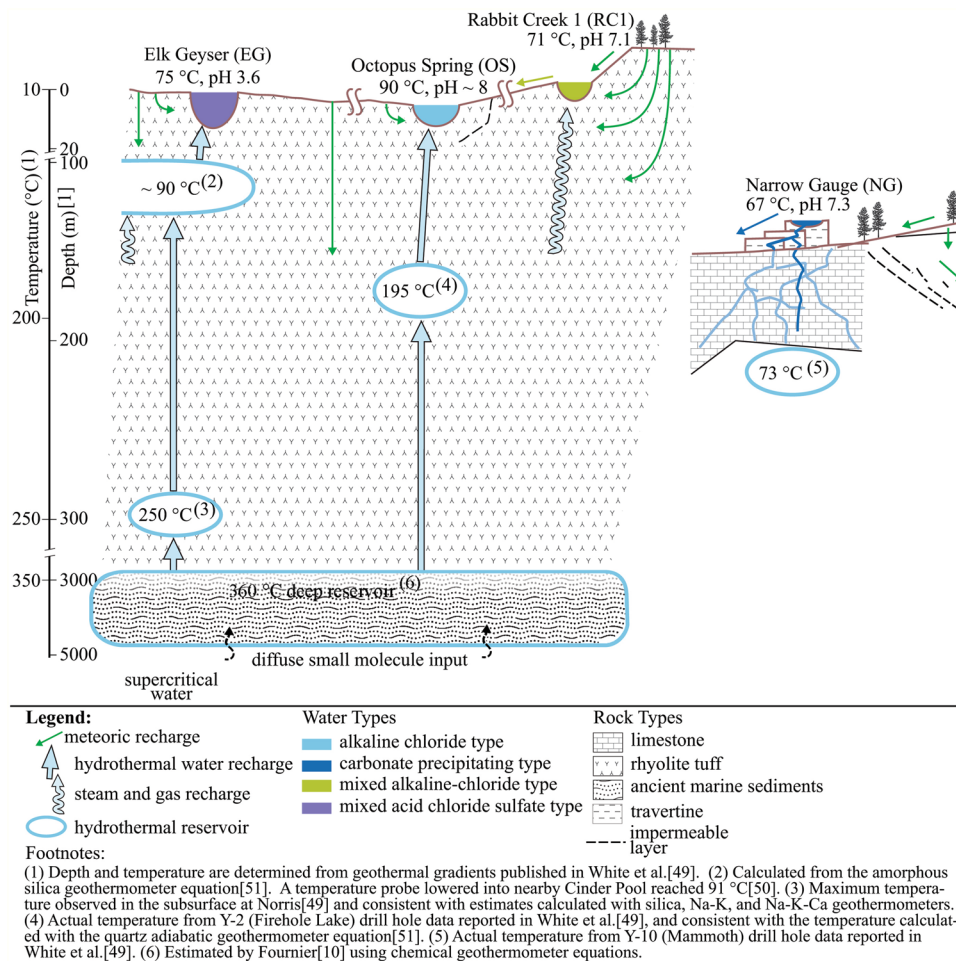


Figure 5. Conceptual diagram of YDOM transformation in Yellowstone hot springs as derived from drill core studies.

minerals³⁶, and redox-active transition-metal compounds³⁷) will become quite reactive reaction partners and catalysts³⁸ for transformation of aqueous DOM at shallow depths, which is feasible for EG, considering that groundwater does not enter in Norris Geyser Basin deeper than 100 m³⁹. Commonplace surface runoff DOM entering the springs will be processed at these near-boiling conditions, but it is not clear whether or not this terrestrially derived DOM can be so severely transformed in individual hot springs as to create the observed unique YDOM, especially when residence times of water in hot springs are low which is certainly the case for NG with no exposed pool and OS with a rather large discharge of an estimated 30 (base discharge) to 192 (surge discharge) L s⁻¹⁴⁰.

The molecular complexity of YDOM is likely to be influenced by several factors, once superheated fluids reach the surface. These factors may include both abiotic (e.g. redox- and photochemistry) and biotic (e.g. transformation of plant-derived DOM) processes that jointly define the environments. To further assess likely contributing sources of DOM, we evaluated each spring for possible contributions from marshes (runoff) and groundwater and put these contributions into context with direct hydrothermal sources to each spring (alkaline-chloride, acid-sulfate, and steam-condensates) (Table 1).

White Creek (locus of OS) and Norris Geyser Basin (locus of EG) are among the few areas for which groundwater – surface water interactions have been studied (White Creek⁴¹; Norris Geyser Basin^{39,42}). Gibson and Hinman⁴¹ concluded that there is some potential for episodic exchange across a leaky sinter barrier between the adjacent marsh and Octopus Spring but did not provide an estimate. Gardner *et al.*³⁹ concluded, based on tritium and CFC analyses, that 10% of “cold, shallow, young” water mixes with hydrothermal water at Norris Geyser Basin. It is worth noting that they comment on the complexity of the groundwater system at Norris Geyser Basin and therefore, the amount of potential mixing at EG might vary significantly from this value.

A conceptual diagram of the Yellowstone hydrothermal system illustrates how the surface and subsurface water flow paths connect water and vapor with the four springs in this study (Fig. 5) and highlights the potential for transformation of organic matter at different depths and corresponding temperatures⁴³. The general circulation model for the Yellowstone caldera posits recharge from meteoric water at high elevations or through fractures within and across the rim of the caldera, as well as recirculation within the local hydrothermal systems^{10,12,42,44–47}. Some of the DOM therefore likely originates through downward leaching of the regionally thin soils⁴⁸ and organic-poor volcanic rocks. Most of the DOM in hot springs directly connected to deep sources (intracaldera, OS and Mammoth Hot Springs, NG) must be generated through internal reactions of DOM of

spring name	surface runoff	ground-water	alkaline-chloride	acid-sulfate	steam-heated
RC1 ^a	y**	n	y*	n	y*
EG ^b	n	y*	y**	y**	n
OS ^c	y*	n	y**	n	n
NG ^d	n	y**	n	n	n

Table 1. Major and minor sources that contribute to the organic complexity in different types of hot springs in Yellowstone (major source (y**); minor source (y*); no contribution (n)). Note: ^aBased on visual observations of surface water input; [Si], [Cl⁻], and [Na⁺], which are traditional alkaline-chloride water components¹⁰; and slightly lower pH in the absence of significant [SO₄²⁻], which is an indication of steam input¹⁰. ^bBased on [Si], [Cl⁻], and [Na⁺], which are traditional alkaline-chloride water components¹⁰, low pH and high [SO₄²⁻], and reported contributions of ~10% “cold, shallow, young” water³⁹. ^cBased on topographic observations (OS is at the base of a hill); [Si], [Cl⁻], and [Na⁺], which are traditional alkaline-chloride water components¹⁰; and reports of possible groundwater exchange⁴¹. ^dBased on topographic observations (NG is at relatively high elevation above forest floor) and previous hydrological investigations⁶⁶.

these three sources: deep hydrothermal water, intra-caldera recirculated water, and infiltrating meteoric water, whereas DOM in the other springs (RC1 and EG) is acquired at shallower depths and lower temperatures. Each spring, therefore, bears an organic geochemical signature dictated in part by the source of the water and partly from reactions that occur in all hydrothermal systems.

OS receives almost all of its water and solutes directly from a deep hydrothermal source (~195 °C and ~150 m depth⁴⁹ (Fig. 5)), most closely representing an endmember among our sites for both organic and inorganic components in the hydrothermal system. In addition to deep hydrothermal water, EG receives groundwater from DOM-containing surface recharge, acidic gases, and steam from the underlying deep aquifer (below 23 m depth^{49,50}). This acidified shallow groundwater also leaches solutes from bedrock. In contrast, RC1 is disconnected from the deep aquifer that feeds OS. Instead, water and heat come from steam condensate that substantially interacts with bedrock and perhaps shallow groundwater, as well as DOM-containing water from surface runoff. NG is directly connected to a separate limestone aquifer source, which is absent from the caldera, with a theoretical temperature of ~110 °C based on a quartz adiabatic geothermometer calculation⁵¹. However, the actual temperature from the nearby Mammoth drill core Y10 maintained a constant temperature of 70 °C from ~15 m below the surface to the maximum depth of the drill hole at ~135 m⁴⁹.

Temperature estimates from drill hole geothermal gradients⁴⁹ revealed variations of depth at which a temperature of 350 °C, the proposed temperature of the parent thermal water¹⁰ and near supercritical conditions (374 °C), would be reached, ranging from 690 m (NG; drill hole Y10) > 570 m (EG; drill hole Y12) ≈ 560 m (RC1; drill hole Y5) > 470 m (OS; drill hole Y2). The depth at which supercritical conditions for aqueous fluids would be reached is too shallow to allow contributions of organic molecules from the marine rocks deeply buried (~3 km) under the caldera except for very small molecules capable of surviving these extreme conditions¹⁸. Circulating upwards, this deep hydrothermal water then accumulates in shallower aquifers (~150 m depth) where subsequent modifications of DOM may occur. The reactions that occur in the shallower aquifers are the ones that substantially modify the composition of hot spring water in individual springs.

Additional reactions, though, contribute to the diversity of organic molecules, where temperature, dissolved gases, and water-property changes play a critical role^{52,53}. For example, the high number of unique sulfur-containing molecular ions in YDOM and specifically in the NG and EG samples is likely a result of reactions of DOM with hydrogen sulfide⁵⁴. However, a distinct precursor pool of organic molecules must exist to create these unique molecular signatures. Indeed, for each YDOM sulfur-containing formula (CHOS), there is a feasible precursor CHO formula. These molecular signatures are also very different from DOM hydro-sulfurized under anaerobic conditions in sediments or soils (Supplementary Fig. S9). A comparison of YDOM sulfur-containing formulas with those of a sulfate-rich anaerobic-sediment pore-water DOM collected in the Chesapeake Bay (38°20′54.19″N, 76°18′47.34″W) revealed only a small overlap [94 (~4%) out of 2,458 CHOS formulas] despite a very high diversity of CHOS molecules present in the sediment pore water (1,703 CHOS formulas), further highlighting the unusual and largely unique nature of YDOM (Supplementary Fig. S9).

Geochemically formed organic molecules likely are produced from supercritical CO₂ in deep-reaching tectonic faults^{55,56} and in alkaline hydrothermal vents⁵⁷. Laboratory-based experiments confirmed extensive transformation reactions of hydrocarbons below 300 °C and 100 MPa pressure at hydrothermal conditions and document that complex organic molecules were indeed produced^{138,52}. Remarkably, the dielectric constant of water decreases from 80.1 (20 °C) down to 19.7 (300 °C) making its solvent properties at ~300 °C roughly comparable with that of acetone at 25 °C⁵⁸ whereas its dissociation constant increases dramatically, leading to more acidic conditions at equivalent pH values. All this facilitates solubilization of organic matter coupled with transformation by ionic condensation, cleavage, and hydrolysis. Under these conditions, water can act as a powerful acid-base catalyst³⁸, and ionic reaction pathways could be further facilitated by acidic and basic minerals and dissolved salts. These possible reaction pathways may result in the formation of complex YDOM.

Additionally, above the critical temperature ($T_c = 374$ °C), thermally induced free-radical reactions are likely to become more significant and may even dominate as temperature increases³⁸. Water may contribute hydrogen and oxygen for formation of hydrocarbons and oxygenated alteration products⁴³. Deep hydrothermal geochemistry near and above supercritical conditions drastically affects hydrogen bonding in aqueous fluids⁵⁹ and might also enable the formation of specific YDOM constituents that are improbable under (near) surface conditions.

Possible geochemical conditions that may be responsible for such a tremendous chemodiversity of YDOM include pressurized hot alkaline-chloride solutions at depth (**RC1**, **OS**), heated H_2S and H_2SO_4 (**EG**), and deep circulating crustal fluids in presence of CaCO_3 (metamorphosed marine sediments) and H_2S (**NG**). Thermochemical sulfate reduction processes operating above $\sim 140^\circ\text{C}$ ⁴³ further contribute to the observed diversity of CHOS and CHNOS compounds in YDOM.

We conclude (A) that the uniqueness of YDOM was confirmed by compositional comparison with globally collected surface water SPE-DOM samples from rivers, lakes, estuaries, and oceans and by structural and fluorometric comparison with boreal lake SPE-DOM; (B) YDOM complexity cannot be explained by limited biodiversity and expected metabolites, and therefore it must result from its exposure to the inorganic geochemical environment in each spring; (C) YDOM molecular complexity offers a new dimension in defining extreme environments; and (D) YDOM organic chemodiversity can be used to classify springs analogously to the established inorganic geochemical classification currently employed.

Methods

Description of the Sampled Hot Springs. The springs reported here are examples of each of these three physiographic types (intra-caldera, caldera-rim, and extra-caldera) and are of different compositional types (alkaline-chloride, mixed-alkaline chloride, acid-chloride-sulfate, and travertine-precipitating), presenting different degrees of connectedness with primary thermal waters, shallow meteoric aquifers, and other crustal fluids. In general, hot springs in the Yellowstone hydrothermal system are considered to show long-term stability of thermal water chemistry⁶⁰, which is also reflected in the reproducible data presented in the HCA in Fig. 1. Figure 5 includes a schematic diagram of the proposed water sources and relative placement of springs within the Yellowstone hydrologic and volcanic system.

The first detailed site, Octopus Spring (**OS**), located within the White Creek Group, Lower Geyser Basin, along with Rabbit Creek 4, Azure Spring, and Ojo Caliente, represents the intra-caldera, high-silica alkaline-chloride hot spring water from the primary hot-water source stored within the rhyolites of the first and third eruption cycles of the Yellowstone caldera. The temperature of the source water for **OS** is $\sim 195^\circ\text{C}$, as calculated with the adiabatic quartz geothermometry equation developed by Fournier⁵¹. This translates to a depth of approximately 130–150 m, following the calculated boiling point curve and by comparison with actual downhole temperatures of the closest drill hole, Y2, located about 1.5 kilometers away⁴⁹. Rhyolites of Central Plateau flows of the first and third eruption cycles host the deep hydrothermal aquifer in this area⁴⁹. The White Creek Group, along with other groups in the Firehole Lake area, comprises several hot springs that are influenced by small amounts of steam condensate, formed when steam, released during subsurface boiling, reaches an impermeable rock layer after which the water condenses. This hot, slightly acidic water reacts with the country rock, leaching trace elements from the minerals. The alkaline-chloride-type water is the same type as that of better-known thermal features, such as Morning Glory Pool, Grand Prismatic Spring, and Old Faithful Geyser¹⁰. Recent work modeling the regional hydrograph suggests much of the water in the Yellowstone Plateau is recirculated through the hydrothermal system⁴². But some variability in the chemistry of Octopus Spring can be attributed to episodic influx of small amounts of water from an adjacent cold-water marsh, as previously reported⁴¹; see also supporting online text. **OS** has been the focus of extensive microbiological^{14,61}, geochemical^{140,62}, and hydrogeological^{141,62} studies.

The second spring, Rabbit Creek 1 (**RC1**) (MRCHSG032, Rabbit Creek Group, Midway Geyser Basin) along with Rabbit Creek 3, discharges water from steam condensate generated by the primary hot-water source mixed with terrestrial input from a small marsh uphill from **RC1** (mixed alkaline-chloride). Organic matter would be exposed to hot mildly acidic conditions, creating background DOM from the milieu of primary sources: terrestrial inputs, microbial inputs, leached soils, and atmospheric inputs. The discharge from the spring is visibly greater than the surface inputs, although measurements of recharge and discharge were not made.

The third spring Elk Geyser (**EG**), along with Cinder Pool, is an example of an acid-chloride-sulfate spring in Norris Geyser Basin (Type III of⁵⁰), which is located outside the Yellowstone caldera at approximately the intersection of the caldera-rim fracture system and the north-south-trending Norris Mammoth corridor. **EG** was not present in Norris Geyser Basin at the time the White *et al.*⁵⁰ work was conducted. Two major types of hydrothermal water are present at Norris Geyser Basin: alkaline-chloride (Type I of⁵⁰) and acid-sulfate (Type IV of⁵⁰). Alkaline-chloride waters are apparently cooled by non-adiabatic processes⁴². Gardner's *et al.*⁴² comprehensive work does not address the origin of heat and solutes in acid-sulfate or in acid-chloride-sulfate springs. In the classical model, acid-sulfate springs form from the interaction of gases and water vapor, released by subsurface boiling from the underlying magma chamber, with the local meteoric aquifer^{10,11}. Carbon dioxide (CO_2) and hydrogen sulfide (H_2S) are the dominant gases in such springs. The latter oxidizes to form sulfuric acid, which accounts for the acidity along with carbonic acid from dissolved CO_2 . In the case of acid-sulfate-chloride waters, recirculated alkaline-chloride waters mix with acid-chloride-sulfate waters or are infused with acid gases at depth (hydrothermal with subsurface boiling and hot gas infusion)⁶³. Many springs (e.g., Cinder Pool) have detectable quantities of H_2S along with microbial populations for its oxidation⁶⁴. Other springs, like **EG**, show no evidence of H_2S , suggesting oxidation takes place in the subsurface. The water in **EG** is isolated from the deep hydrothermal aquifer of the intra-caldera zone by physiographic and hydrological barriers. **EG** is a mixture of extensively acidified, steam-heated shallow aquifer and the more deeply circulating alkaline chloride waters of Norris Geyser Basin⁵⁰. As such, it would have the alkaline-chloride background source of DOM modified through acidic steam extracts.

The fourth spring, Narrow Gauge (**NG**), is in the Mammoth Hot Springs area, north of the Yellowstone Plateau. It is a travertine-precipitating spring. Isotopic evidence suggests the Mammoth system may issue from a separate hydrothermal source from the springs of the Yellowstone Plateau¹³. Source waters for this spring come from deeply circulating crustal fluids that encounter limestone at depth, dissolve CaCO_3 , and become enriched in CO_2 gas. Such springs are common throughout the northern Rocky Mountains⁶⁵, although none are as extensive

as the Mammoth Hot Spring. Upon reaching the surface, CO₂ exsolves, leading to rapid precipitation of CaCO₃ in massive travertine terraces. The fluids carry dissolved H₂S, probably leached from Mesozoic marine sediments. H₂S is not oxidized in the subsurface and emanates with the carbonate-precipitating fluids. Background organic matter derives from the metamorphosed marine sediments, similar to DOM from alkaline-chloride springs, but recirculation and modification of organic matter derived from this source and modified by contributions from terrestrial and microbial sources appears unlikely due to the relief of the travertine terraces. In fact, most discharge from the Mammoth terraces is fault-controlled and thought to flow from Boiling River⁶⁶, although complete mass balance measurements are not possible. So the DOM associated with Mammoth Hot Springs may be somewhat different from DOM found in acid-chloride-sulfate springs and it appears not to have been exposed to such high temperatures as those achieved by acid-chloride-sulfate springs.

Sample Collection and Isolation of DOM. Hot spring water samples were collected from selected hot springs in 2010 and 2012. The sampling was undertaken by submerging 2.5 L Pyrex glass bottles into the spring. The 2.5 L glass bottle containing the sample was then allowed to cool below 60 °C before the water was transferred to 20 L glass bottles. This process was repeated until 20 L were collected. The Narrow Gauge spring was sampled using Teflon tubing and spring water was siphoned directly into the 20 L glass containers. 20 L was a minimal requirement to extract sufficient DOM to allow for NMR analyses. All samples were transported to the laboratory located in Yellowstone National Park and then filtered through Whatman GF/F glass fiber filters. A previously described solid-phase extraction procedure⁶⁷ was used to isolate DOM from the water samples using highly efficient (carbon extraction efficiency is about 60%) Agilent Bond Elut PPL solid-phase extraction (SPE) cartridges filled with 1 g of a functionalized styrene-divinylbenzene polymer (PPL) resin. Briefly, after conditioning the Agilent PPL cartridges with 2 cartridge volumes of high purity methanol and rinsing with 1 cartridge volume of acidified ultrapure water (acidified to pH 2), the filtered water samples were acidified (pH 2 with 32% HCl) and gravity-fed to the cartridge. The adsorbed DOM was eluted off the cartridge by using 10 mL high purity methanol and the isolate stored at -20 °C in the freezer prior to FTICR mass spectrometry, NMR spectroscopy and EEM fluorescence analyses.

The following springs were sampled for this study: Octopus Spring, Narrow Gauge, Rabbit Creek (three springs at this location), Elk Spring, Cinder Pool, Ojo Caliente, Azure Spring, and Cinder Pool. To demonstrate the drastic differences of DOM present in a wide variety of common, non-hydrothermal aquatic systems with Yellowstone hot spring DOM, we also collected 1 L freshwater or 10 L seawater samples from 114 sites from diverse aquatic systems: aquatic biomes in New Zealand; the Suwannee River, Georgia, USA (also an IHSS standard reference material); boreal lakes in the Malingsbo area, Arctic lakes in the Abisko area and samples from the Baltic Sea in Sweden; underneath the sea ice in Antarctica, Ross Island; Amazonian rivers and saltwater lagoons in Brazil; the North Pacific and Atlantic Ocean, including Sargasso Sea. DOM from all aquatic systems mentioned above were solid-phase extracted according to the same procedure given above. More details about sampling locations are given in the online supplementary material.

Characterization of DOM. FT-ICR MS and NMR analyses were undertaken at the Helmholtz Center, Munich, Germany. Mass spectrometric molecular formula assignments and NMR data processing was carried out in a similar fashion as described in previous studies^{19,68}. NMR analysis needed to be restricted to a selected set of samples because of the time constraints of analysis. EEM spectra were recorded using a Horiba Aqualog fluorometer at the University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory. Details about each analytical technique are given below.

FT-ICR MS. Mass spectra were obtained in negative mode electrospray ionization (ESI) using a Bruker Solarix 12 Tesla FT-ICR mass spectrometer. All methanolic SPE-samples were directly injected into the ionization source at a flow rate of 120 µL min⁻¹ and a voltage of 3,600 V. Five hundred transient spectra were averaged at a 4 mega word time domain to yield very accurate and highly resolved *m/z* molecular ions. The high magnetic field and its resulting ultrahigh resolution allowed assigning precise molecular formulas to the majority of observed *m/z* ions. The calculated error between measured and actual mass of assigned formulas was always better than 0.2 ppm. More details about this specific instrument and the associated data analysis have been previously published^{19,26}. Double bond equivalency was calculated according to Bae *et al.*⁶⁹.

Excitation emission matrix (EEM) fluorescence. One mL of methanolic YDOM and aquatic SPE-DOM samples were dried under ultrapure nitrogen and then re-dissolved in 5 mL ultrapure MilliQ water. EEM spectra were then measured on the aqueous samples using a Horiba Jobin Yvon Aqualog fluorometer at excitation wavelengths ranging from 230–500 nm and emission wavelengths between 200–600 nm. The recorded EEM spectra were then corrected for Raleigh and Raman scattering, inner filtering effect, and normalized to a 1 ppm quinine sulfate standard and expressed in quinine sulfate units (QSU).

Nuclear magnetic resonance spectroscopy (NMR). ¹H NMR spectra of methanolic YDOM extracts were acquired with a Bruker Avance NMR spectrometer at 800.13 MHz (B₀ = 18.7 T) at 283 K from a few mg of solid obtained by evaporation of original methanol solution, dissolved in approx. 130 µL CD₃OD (Merck, 99.95% ²H) solution with a 5 mm z-gradient ¹H/¹³C/¹⁵N/³¹P QCI cryogenic probe (90° excitation pulses: ¹³C-¹H~10 µs) in sealed 2.5 mm Bruker MATCH tubes. 1D ¹H NMR spectra were recorded with a spin-echo sequence (10 µs delay) to allow for high-Q probe ringdown, and classical presaturation to attenuate residual water present “*noesypr1d*” (5 s acquisition time, 5 s relaxation delay, 1 ms mixing time; 1 Hz exponential line broadening). A phase sensitive, gradient enhanced echo-antiecho TOCSY NMR spectrum with solvent suppression (*dipsi2etgpsi19*) was acquired with an acquisition time of 1 s, a mixing time of 70 ms, and a relaxation delay of 1 s (spectral width of 9615.4 Hz, computed

to a 16384 × 2048 matrix. The one bond coupling constant $^1J(\text{CH})$ used in 2D ^1H , ^{13}C DEPT-HSQC spectra (*hsqc-detgpsisp2.2*) was set to 145 Hz; other conditions: ^{13}C 90 degree decoupling pulse, GARP (70 μs); 50 kHz WURST 180 degree ^{13}C inversion pulse (Wideband, Uniform, Rate, and Smooth Truncation; 1.2 ms); F2 (^1H): spectral width of 9572.2 Hz (11.96 ppm); 1.25 s relaxation delay; F1 (^{13}C): SW = 40252 Hz (200 ppm); 36224.9 Hz (180 ppm, for OS and RC1). HSQC-derived NMR spectra were computed to a 4096 × 512 matrix. Gradient (1 ms length, 450 μs recovery) and sensitivity enhanced sequences were used for all 2D NMR spectra. Similarity of ^1H NMR spectra (Supplementary Fig. S4) was computed from 0.001 ppm section integrals in the range $\delta_{\text{H}} = 0.5\text{--}9.5$ ppm, with exclusion of methanol and residual water (Bruker AMIX software, version 3.9.4.) with Hierarchical Cluster Explorer (HCE); similarity versus distance metrics used Pearson correlation coefficients. Other NMR acquisition conditions are given in Supplementary Table. S5.

Statistical Analyses. Stiff diagrams were used to visualize the inorganic composition of selected springs. Inorganic parameters used for Hierarchical Cluster Analysis (HCA) are summarized in Supplementary Table S1. All HCA were undertaken on auto-scaled data⁷⁰ using Pearson correlations and average linkage. Simple absence presence analyses were used to isolate unique molecular formulas for the YDOM and SPE-DOM sample sets.

References

- Hazen, R. M., Jones, A. P. & Baross, J. A. *Carbon in Earth*. Vol. 75 (the Mineralogical Society of America, 2013).
- Nisbet, E. G. & Sleep, N. H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).
- Gomez-Saez, G. V. *et al.* Molecular evidence for abiotic sulfuration of dissolved organic matter in marine shallow hydrothermal systems. *Geochimica et Cosmochimica Acta* **190**, 35–52, <https://doi.org/10.1016/j.gca.2016.06.027> (2016).
- Rossel, P. E., Stubbins, A., Hach, P. F. & Dittmar, T. Bioavailability and molecular composition of dissolved organic matter from a diffuse hydrothermal system. *Marine Chemistry* **177**, 257–266, <https://doi.org/10.1016/j.marchem.2015.07.002> (2015).
- Clifton, C. G., Walters, C. C. & Simoneit, B. R. T. Hydrothermal petroleum from Yellowstone National Park, Wyoming, USA. *Applied Geochemistry* **5**, 169–191, [https://doi.org/10.1016/0883-2927\(90\)90047-9](https://doi.org/10.1016/0883-2927(90)90047-9) (1990).
- Schouten, S. *et al.* Archaeal and Bacterial Glycerol Dialkyl Glycerol Tetraether Lipids in Hot Springs of Yellowstone National Park. *Applied and Environmental Microbiology* **73**, 6181–6191, <https://doi.org/10.1128/aem.00630-07> (2007).
- Bogard, M. J., Donald, D. B., Finlay, K. & Leavitt, P. R. Distribution and regulation of urea in lakes of central North America. *Freshwater Biology* **57**, 1277–1292, <https://doi.org/10.1111/j.1365-2427.2012.02775.x> (2012).
- Duan, S. & Bianchi, T. S. Particulate and dissolved amino acids in the lower Mississippi and Pearl Rivers (USA). *Marine Chemistry* **107**, 214–229, <https://doi.org/10.1016/j.marchem.2007.07.003> (2007).
- Pierce, K. L. & Morgan, L. A. Is the track of the Yellowstone hotspot driven by a deep mantle plume? — Review of volcanism, faulting, and uplift in light of new data. *Journal of Volcanology and Geothermal Research* **188**, 1–25, <https://doi.org/10.1016/j.jvolgeores.2009.07.009> (2009).
- Fournier, R. O. Geochemistry and Dynamics of the Yellowstone National Park Hydrothermal System. *Annual Review of Earth and Planetary Sciences* **17**, 13–53, <https://doi.org/10.1146/annurev.ea.17.050189.000305> (1989).
- Lowenstern, J. B. & Hurwitz, S. Monitoring a supervolcano in repose: Heat and volatile flux at the yellowstone caldera. *Elements* **4**, 35–40, <https://doi.org/10.2113/gselements.4.1.35> (2008).
- Hurwitz, S., Evans, W. C. & Lowenstern, J. B. River solute fluxes reflecting active hydrothermal chemical weathering of the Yellowstone Plateau Volcanic Field, USA. *Chem Geol* **276**, 331–343, <https://doi.org/10.1016/j.chemgeo.2010.07.001> (2010).
- Kharaka, Y. K., Sorey, M. L. & Thordsen, J. J. Large-scale hydrothermal fluid discharges in the Norris-Mammoth corridor, Yellowstone National Park, USA. *Journal of Geochemical Exploration* **69**, 201–205, [https://doi.org/10.1016/s0375-6742\(00\)00025-x](https://doi.org/10.1016/s0375-6742(00)00025-x) (2000).
- Brock, T. D. Life at High Temperatures: Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed. *Science* **158**, 1012–1019 (1967).
- Stahl, D. A., Lane, D. J., Olsen, G. J. & Pace, N. R. Characterization of a Yellowstone Hot-Spring Microbial Community by 5S Ribosomal-RNA Sequences. *Applied and Environmental Microbiology* **49**, 1379–1384 (1985).
- Meyer-Dombard, D. R., Shock, E. L. & Amend, J. P. Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**, 211–227, <https://doi.org/10.1111/j.1472-4669.2005.00052.x> (2005).
- Inskeep, W. P. *et al.* The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Frontiers in Microbiology* **4**, <https://doi.org/10.3389/fmicb.2013.00067> (2013).
- Hawkes, J. A. *et al.* Efficient removal of recalcitrant deep-ocean dissolved organic matter during hydrothermal circulation. *Nature Geosci* **8**, 856–860, <https://doi.org/10.1038/ngeo2543> (2015).
- Hertkorn, N., Harir, M., Koch, B. P., Michalke, B. & Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **10**, 1583–1624, <https://doi.org/10.5194/bg-10-1583-2013> (2013).
- Hertkorn, N. *et al.* High-precision frequency measurements: Indispensable tools at the core of the molecular-level analysis of complex systems. *Anal. Bioanal. Chem.* **389**, 1311–1327, <https://doi.org/10.1007/s00216-007-1577-4> (2007).
- Schmitt-Kopplin, P. *et al.* High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *PNAS* **106**, 2763–2768, <https://doi.org/10.1073/pnas.0912157107> (2010).
- Hertkorn, N., Harir, M., Cawley, K. M., Schmitt-Kopplin, P. & Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **13**, 2257–2277, <https://doi.org/10.5194/bg-13-2257-2016> (2016).
- Lam, B. *et al.* Major Structural Components in Freshwater Dissolved Organic Matter. *Environ. Sci. Technol.* **41**, 8240–8247 (2007).
- Stedmon, C. A. & Nelson, N. B. In *Biogeochemistry of Marine Dissolved Organic Matter, 2nd Edition* (eds D. A. Hansell & C. A. Carlson) 481–503 (Academic Press - Elsevier, 2015).
- Jørgensen, L. *et al.* Global trends in the fluorescence characteristics and distribution of marine dissolved organic matter. *Marine Chemistry* **126**, 139–148, <https://doi.org/10.1016/j.marchem.2011.05.002> (2011).
- Gonsior, M. *et al.* Chemodiversity of dissolved organic matter in the Amazon Basin. *Biogeosciences* **13**, 4279–4290, <https://doi.org/10.5194/bg-13-4279-2016> (2016).
- Kowalczyk, P. *et al.* Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model. Relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. *Mar. Chem.* **118**, 22–36, <https://doi.org/10.1016/j.marchem.2009.10.002> (2010).
- Hertkorn, N. *et al.* Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010 (2006).
- Lam, B. & Simpson, A. J. Direct ^1H NMR spectroscopy of dissolved organic matter in natural waters. *Analyst* **133**, 263–269 (2008).
- Hertkorn, N. *et al.* Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Analytical Chemistry* **80**, 8908–8919, <https://doi.org/10.1021/ac800464g> (2008).

31. Gonsior, M., Luek, J., Schmitt-Kopplin, P., Grebmeier, J. M. & Cooper, L. W. Optical Properties and Molecular Diversity of Dissolved Organic Matter in the Bering Strait and Chukchi Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* (2017).
32. Maie, N., Yamashita, Y., Cory, R. M., Boyer, J. N. & Jaffe, R. Application of excitation emission matrix fluorescence monitoring in the assessment of spatial and seasonal drivers of dissolved organic matter composition: Sources and physical disturbance controls. *Applied Geochemistry* **27**, 917–929, <https://doi.org/10.1016/j.apgeochem.2011.12.021> (2012).
33. Guéguen, C., Granskog, M. A., McCullough, G. & Barber, D. G. Characterisation of colored dissolved organic matter in Hudson Bay and Hudson Strait using parallel factor analysis. *Journal of Marine Systems* **88**, 423–433, <https://doi.org/10.1016/j.jmarsys.2010.12.001> (2011).
34. Coble, P. G. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* **51**, 325–346 (1996).
35. Yamashita, Y. & Tanoue, E. Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. *Marine Chemistry* **82**, 255–271, [https://doi.org/10.1016/S0304-4203\(03\)00073-2](https://doi.org/10.1016/S0304-4203(03)00073-2) (2003).
36. Shipp, J. A., Gould, I. R., Shock, E. L., Williams, L. B. & Hartnett, H. E. Sphalerite is a geochemical catalyst for carbon–hydrogen bond activation. *Proceedings of the National Academy of Sciences* **111**, 11642–11645, <https://doi.org/10.1073/pnas.1324222111> (2014).
37. Wang, J., Liu, L., Zhang, L. & Li, Z. Aquathermolysis of Heavy Crude Oil with Amphiphilic Nickel and Iron Catalysts. *Energy & Fuels* **28**, 7440–7447, <https://doi.org/10.1021/ef502134p> (2014).
38. Siskin, M. & Katritzky, A. R. Reactivity of Organic Compounds in Superheated Water: General Background. *Chemical Reviews* **101**, 825–836, <https://doi.org/10.1021/cr000088z> (2001).
39. Gardner, W. P., Susong, D. D., Solomon, D. K. & Heasler, H. P. A multitracer approach for characterizing interactions between shallow groundwater and the hydrothermal system in the Norris Geyser Basin area, Yellowstone National Park. *Geochemistry, Geophysics, Geosystems* **12**, n/a–n/a, <https://doi.org/10.1029/2010gc003353> (2011).
40. Hinman, N. W. & Lindstrom, R. F. Seasonal changes in silica deposition in hot spring systems. *Chem Geol* **132**, 237–246, [https://doi.org/10.1016/S0009-2541\(96\)00060-5](https://doi.org/10.1016/S0009-2541(96)00060-5) (1996).
41. Gibson, M. L. & Hinman, N. W. Mixing of hydrothermal water and groundwater near hot springs, Yellowstone National Park (USA): hydrology and geochemistry. *Hydrogeology Journal* **21**, 919–933, <https://doi.org/10.1007/s10040-013-0965-4> (2013).
42. Gardner, W. P., Susong, D. D., Solomon, D. K. & Heasler, H. Snowmelt hydrograph interpretation: Revealing watershed scale hydrologic characteristics of the Yellowstone volcanic plateau. *Journal of Hydrology* **383**, 209–222, <https://doi.org/10.1016/j.jhydrol.2009.12.037> (2010).
43. Seewald, J. S. Organic–inorganic interactions in petroleum-producing sedimentary basins. *Nature* **426**, 327–333 (2003).
44. Lowenstern, J. B., Bergfeld, D., Evans, W. C. & Hurwitz, S. Generation and evolution of hydrothermal fluids at Yellowstone: Insights from the Heart Lake Geyser Basin. *Geochem. Geophys. Geosyst.* **13**, Q01017, <https://doi.org/10.1029/2011gc003835> (2012).
45. Sorey, M. L. Effects of potential geothermal development in the Corwin Springs Known Geothermal Resources Area, Montana, on the thermal features of Yellowstone National Park. Report No. 91–4052 (1991).
46. Rye, R. O. & Truesdell, A. H. In *Integrated geoscience studies in the greater Yellowstone area—Volcanic, tectonic, and hydrothermal processes in the Yellowstone geocosystem: U.S. Geological Survey Professional Paper 1717* (ed. L. A. Morgan) Ch. H, (U.S. Geological Survey, 2007).
47. Kharaka, Y. K., Thordsen, J. J. & White, L. D. Isotope and chemical compositions of meteoric and thermal waters and snow from the greater Yellowstone National Park region. Report No. 2002–194 (2002).
48. Rodman, A. W., Shovic, H. F. & Thoma, D. P. *Soils of Yellowstone National Park*. (Yellowstone Center for Resources, National Park Service, 1996).
49. White, D. E., Fournier, R. O., Muffler, L. J. P. & Truesdell, A. H. Physical results of research drilling in thermal areas of Yellowstone National Park, Wyoming. Report No. 892, (USGS, 1975).
50. White, D. E., Hutchinson, R. A. & Keith, T. E. C. Geology and remarkable thermal activity of Norris Geyser Basin, Yellowstone National Park, Wyoming. *U.S. Geol. Surv., Prof. Pap.* **75**, 1–84 (1988).
51. Fournier, R. O. Chemical geothermometers and mixing models for geothermal systems. *Geothermics* **5**, 41–50, [https://doi.org/10.1016/0375-6505\(77\)90007-4](https://doi.org/10.1016/0375-6505(77)90007-4) (1977).
52. Shipp, J. *et al.* Organic functional group transformations in water at elevated temperature and pressure: Reversibility, reactivity, and mechanisms. *Geochimica et Cosmochimica Acta* **104**, 194–209, <https://doi.org/10.1016/j.gca.2012.11.014> (2013).
53. Watanabe, M. *et al.* Chemical Reactions of C1 Compounds in Near-Critical and Supercritical Water. *Chemical Reviews* **104**, 5803–5822, <https://doi.org/10.1021/cr020415y> (2004).
54. Hebling, Y. *et al.* Biomarker Evidence for a Major Preservation Pathway of Sedimentary Organic Carbon. *Science* **312**, 1627–1631, <https://doi.org/10.1126/science.1126372> (2006).
55. Schreiber, U., Locker-Gruetjen, O. & Mayer, C. Hypothesis: Origin of Life in Deep-Reaching Tectonic Faults. *Origins of Life and Evolution of Biospheres* **42**, 47–54, <https://doi.org/10.1007/s11084-012-9267-4> (2012).
56. Mayer, C., Schreiber, U. & Davila, M. J. Periodic Vesicle Formation in Tectonic Fault Zones—an Ideal Scenario for Molecular Evolution. *Origins of Life and Evolution of Biospheres* **45**, 139–148, <https://doi.org/10.1007/s11084-015-9411-z> (2015).
57. Martin, W. & Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B-Biological Sciences* **362**, 1887–1925, <https://doi.org/10.1098/rstb.2006.1881> (2007).
58. Siskin, M., Katritzky, A. R. & Balasubramanian, M. Aqueous organic chemistry. 4. *Cleavage of diaryl ethers*. *Energy & Fuels* **5**, 770–771, <https://doi.org/10.1021/ef00029a028> (1991).
59. Brunner, G. Near critical and supercritical water. Part I. Hydrolytic and hydrothermal processes. *The Journal of Supercritical Fluids* **47**, 373–381, <https://doi.org/10.1016/j.supflu.2008.09.002> (2009).
60. Hurwitz, S. & Lowenstern, J. B. Dynamics of the Yellowstone hydrothermal system. *Reviews of Geophysics* **52**, 375–411, <https://doi.org/10.1002/2014rg000452> (2014).
61. Brock, T. D. The Road to Yellowstone - and beyond. *Annual Review of Microbiology* **49**, 1–28 (1995).
62. Stauffer, R. E., Jenne, E. A. & Ball, J. W. Chemical Studies of Selected Trace Elements in Hot-Spring Drainages of Yellowstone National Park. (U.S. Geological Survey, Washington D.C., 1980).
63. Nordstrom, D. K., McCleskey, R. B. & Ball, J. W. Sulfur geochemistry of hydrothermal waters in Yellowstone National Park: IV Acid-sulfate waters. *Applied Geochemistry* **24**, 191–207, <https://doi.org/10.1016/j.apgeochem.2008.11.019> (2009).
64. Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A. & Fischer, J. Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? *Applied and Environmental Microbiology* **67**, 2873–2882, <https://doi.org/10.1128/aem.67.7.2873-2882.2001> (2001).
65. Chadwick, R. A. & Leonard, R. B. Structural Controls of Hot-Spring Systems in Southern Montana. (United States Department of the Interior Geological Survey, Helena, Montana, 1979).
66. Sorey, M. L. & Colvard, E. M. Hydrologic investigations in the Mammoth Corridor, Yellowstone National Park and vicinity. *USA Geothermics* **26**, 221–249, [https://doi.org/10.1016/S0375-6505\(96\)00041-7](https://doi.org/10.1016/S0375-6505(96)00041-7) (1997).
67. 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* **6**, 230–235 (2008).
68. Gonsior, M. *et al.* Characterization of dissolved organic matter across the Subtropical Convergence off the South Island, New Zealand. *Marine Chemistry* **123**, 99–110, <https://doi.org/10.1016/j.marchem.2010.10.004> (2011).
69. Bae, E. J. *et al.* Study of Double Bond Equivalents and the Numbers of Carbon and Oxygen Atom Distribution of Dissolved Organic Matter with Negative-Mode FT-ICR MS. *Anal. Chem.* **83**, 4193–4199, <https://doi.org/10.1021/ac200464q> (2011).
70. Bro, R. & Smilde, A. K. Principal component analysis. *Analytical Methods* **6**, 2812–2831, <https://doi.org/10.1039/c3ay41907j> (2014).

Acknowledgements

The collection of the presented diversity of aquatic DOM samples would not have been possible without the help of a large group of people and we would like to thank: Ditte M. Mikkelsen and Søren Rysgaard in assisting to collect DOM samples in Greenland, David Bastviken in assisting to collect DOM samples in Sweden, Alex Enrich-Prast in assisting to collect samples in Brazil, Miles Lamare in assisting to collect samples in Antarctica, Michael Perdue in assisting to collect samples in the Suwannee River, Florida, Maureen Conte in assisting to collect marine samples in the Sargasso Sea, Wesley Kiggins for helpful discussions and the Yellowstone Center for Resources (Yellowstone National Park, WY) for permission and access to collect hot spring samples (Permit No. YELL-2009-SCI-5828 YELL-2010-SCI-5828, YELL-2010-SCI-1660, YELL-2012-SCI-5828, and YELL-2012-SCI-1660). The University of Montana provided support for Nancy Hinman. This is contribution 5518 of the University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory.

Author Contributions

M.G. designed research and analyzed EEM and FT-ICR MS data of YDOM and SPE-DOM, No.H. performed all NMR analyses, Na.H. organized sampling and analyzed all inorganic parameters, S.D. assisted in data mining of FT-ICR MS. M.H. measured all FT-ICR MS spectra, B.C. coordinated sampling and P.S. designed research together with M.G. All authors contributed to write the article.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32593-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018