Electronic Supplementary Information: Drivers of Disinfection Byproduct Formation and Speciation in Small, Chlorinated Coastal Groundwater Systems: Relative Roles of Bromide and Organic Matter, and the Need for Improved Source Water Characterization and Monitoring

Text S1. ATP Analyses.

Concentrations of adenosine triphosphate (ATP) – a semi-quantitative indicator of metabolically-active microbial cells – were measured using a commercially-available luminescence assay kit (BacTiter-Glo™; Promega), in accordance with reported procedures.[1] Luminescence measurements were obtained using a Perkin-Elmer Victor3 V microplate reader, with sample reactions conducted in flat-bottomed, opaque white plastic microplates. Total (ATP_{total}), extracellular (ATP_{extra}), and intracellular (ATP_{intra}) ATP concentrations were distinguished by performing complementary measurements of ATP in whole samples (ATP_{total}) and samples filtered through pre-sterilized 0.1 μm PES syringe filters (ATP_{extra}), where $ATP_{intra} = ATP_{total} - ATP_{extra}$, with all measurements corrected for negative control background signals (autoclaved water).

Text S2. PARAFAC Analysis.

The PARAFAC analyses were undertaken on the composite set of Island and San Juan County samples (n = 79) using the DOMFluor toolbox (version 1.7) developed by Stedmon and Bro (available at www.models.life.ku.dk/algorithms).[2] Data was first pre-processed to null data at wavelengths influenced by peak scattering, ensuring removal of Rayleigh and Raman scattering influences from all samples. Five models were fit to the data, with the number of components ranging from 2-7. All components were constrained to be non-negative values. The data set was checked for outliers and to establish a range of components to evaluate when

building the PARAFAC Model. Samples were considered possible outliers if their leverage in the model was high, indicating an extreme sample with unique EEM shape that was altering the model significantly. The following samples were excluded from the final analyses due to high leverage: Q1 IC07 treated, Q2 IC06 raw, and Q4 IC08 raw. The 5 models were then evaluated with the smaller data set (n = 76).

The excitation and emission plots generated for the components identified in each model were visually inspected to determine if components exhibited smooth EEMs with appearances similar to literature-defined natural organic matter components. Models with 6 or more components yielded several component EEMs exhibiting high noise, which did not contribute significantly to the aggregate fluorescence EEMs. The models were also evaluated by visually inspecting the residual EEMs generated by subtraction of sample EEMs from the corresponding modeled EEMs (generated by superposition of all modeled component EEMs for a given sample). For the 2-component model, recurring patterns were identifiable in the residual EEMs for many of the samples, indicating that the model did not appropriately fit to the data, likely due to oversimplification (underdetermination) of the model.

The remaining 3- to 5-component models were evaluated by performing half-split analysis and validation via the DOMFluor toolbox.[2] Briefly, the data were split into quarters and the quarters combined to complete two separate split analyses for each set. The two halves of each dataset were fit to the 3- to 5-component models and compared to each other as a validation step (where for a satisfactory model, the two halves should exhibit similar excitation and emission profiles for each component). The DOMFluor toolbox compares the excitation and emission curves for each model mathematically using Tucker Congruence Coefficients, and reports whether the models are robust and the component EEMs from each half are matching.

The 5-component model did not pass the validation step. Because the sum of the squared error was lower for the 4-component model as compared to the 3-component model, the 4-component model was selected as the most appropriate model for PARAFAC analyses of the study data sets.

The model uses an alternating least squares regression to fit the data, which can be influenced by selection of initial conditions. To validate the least squares model fit, the 4-component model was fit to the entire data set with random initialization parameters. Twenty 4-component models were randomly initialized and modeled. All yielded similar squared error residuals, verifying the robustness and accuracy of the 4-component model.

Following identification of the most appropriate PARAFAC model, F_{max} values (maximum fluorescence intensities, in arbitrary units) were determined for the four components in each sample for comparison with observed DBP formation potentials, DOC levels, and selected DOM properties.

Characteristic excitation and emission maxima are summarized for each component in Table S2, along with descriptions of the character and origin typically ascribed to similar components in the literature, their potential relevance to DBP formation, and identifiers used for similar components by other investigators. Figure S7 includes an example EEM, color contour visualizations of Components 1-4, and F_{max} values for Component 1 (F_{max1}) for all sites over the study period.

Text S3. Additional Source Water Characteristics.

pH. The range of raw water pH measured was 6.8 – 7.8 for IC sources and 7.3 – 8.2 for SJC sources (Table 1). pH increased for IC sources (ranging from 7.1 – 8.3) following full-scale chlorination (in addition to softening and/or iron/manganese removal for IC06, IC07, and IC08).

Total Dissolved Solids and Conductivity. Levels of TDS and conductivity varied from 210 – 730 mg/L and 390 – 1,200 mhos/cm, respectively, across the dataset (Table 1), with generally strong positive correlation ($R^2 = 0.79$) between the two parameters (Figure S2). The upper values observed for each of these parameters are consistent with the relatively high chloride and bromide levels observed for a number of sites, and indicate possible influence of seawater intrusion on their water supplies.

Ammonia Concentrations. Ammonia varied widely across the dataset, with NH_3-N concentrations ranging from below the detection limit (0.005 mg/L) to 3.2 mg/L in the IC samples, whereas lower values (<0.005 – 0.187 mg/L) were measured in SJC samples (Table 1). The relatively high levels of NH_3-N in the IC dataset are likely indicative of reducing conditions in the groundwaters for a number of these sites, consistent with the occurrence of reduced manganese and iron in the IC06, IC07, and IC08 waters,[3] though agricultural or septic field infiltration into recharge zones may also be a contributing factor. Ammonia levels appeared to vary seasonally in several IC systems, with substantially higher levels in Q1 for IC05 Source02 and Source03 and IC08, and during Q1 and Q2 for IC06 and IC07. These trends do not exhibit a general correlation with expected changes in climate (i.e., dry vs. rainy season); hence, this does not appear to be a precipitation-driven trend. Sites with high ammonia concentrations also exhibited high free chlorine demands in 7-day DBP-FP tests, though DBP-FPs did not correlate with NH_3-N, as all formation potential tests were conducted so that free chlorine concentrations added to samples surpassed breakpoint. The presence of high NH_3-N levels could nevertheless play an important role under full-scale treatment conditions at the studied sites if the breakpoint is not exceeded, as formation of combined chlorine could inhibit formation of DBPs (as well as efficacy of disinfection). This warrants careful consideration by utilities in selecting the levels of

chlorine applied to the source waters at the monitored groundwater sites.

Heterotrophic Plate Count and ATP. HPC levels were below detection limits (< 1 CFU/mL) for the majority of samples collected during the study, with the exception of IC01, in which HPC values ranged from 4.5 – 60.5 CFU/mL. Consistent with prior observations,[1] HPC measurements exhibited no general correlation with total or intracellular ATP concentrations. Levels of ATP were quite low in nearly all samples ($ATP_{total} < 5\times10^{-11}$ M), with the exception of IC05 Source 2 and IC07 with ATP_{total} of 1.8×10^{-10} M in Q1 and 8×10^{-11} M in Q4, respectively.

Table S1. Site and source characteristics and treatment process descriptions

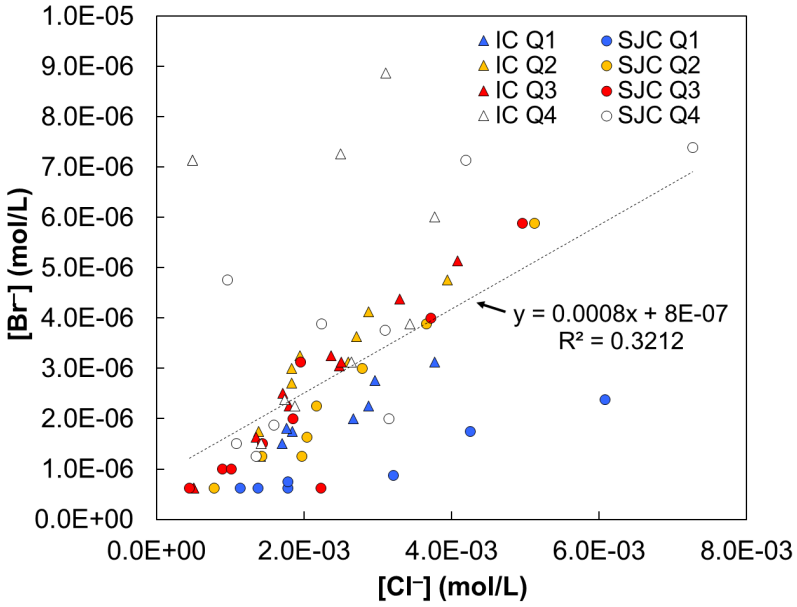
System name	Source Description	Additional Treatment (all sites chlorinate)	Theoretical maximum distribution system residence time (days)[a]	Site Sampling Schedule			
				Q1 (A 15)	Q2 (W 16)	Q3 (Sp 16)	Q4 (Su 16)
IC01	5 sources, one wellfield (Source05), 3 wells (B-Source01,C-Source02,D-Source03) in active use, all combine before treatment		13.7-15.5	9/30/15	12/10/15	3/10/16	6/16/16
IC02	Single source		5.8-13.4	9/24/15	12/10/15	3/10/16	6/16/16
IC03	Single source		4.3-9.5	-	-	3/10/16	6/16/16
IC04	Single source		3.6-6.2	9/24/15	12/10/15	3/10/16	6/16/16
IC05	Two sources at different locations and pump to separate reservoirs, two entry points		1.1-10.3	9/24/15	12/10/15	3/10/16	6/16/16
IC06	3 sources, two wells on together, one treatment plant and one entry point	Fe/Mn removal & corrosion control: free chlorine oxidation, greensand filtration, aeration at reservoir inlets.	2.3-3.4	10/6/15	12/10/15	3/10/16	6/16/16
IC07	Single source	Fe/Mn removal: free chlorine oxidation, greensand filtration.	6.0-10.7	9/24/15	12/10/15	3/10/16	6/16/16
IC08	Two sources pump together, one entry point	Fe/Mn removal & softening: permanganate and free chlorine oxidation, greensand filtration, ion exchange, post-filtration free chlorination.	6.7-9.8	9/24/15	12/10/15	3/10/16	6/16/16
SJC01	Wellfield of 2 sources		7.9-13.4	9/29/15	12/17/15	3/8/16	6/13/16
SJC02	Single source	Aeration tower for TTHM removal	5.1-9.3	9/29/15	12/17/15	3/8/16	6/13/16
SJC03	Two sources at different locations, two entry points		8.4-16.8	9/29/15	12/17/15	3/8/16	6/13/16
SJC04	Two sources at different locations, two entry points		1.2-11.2		12/17/15	3/8/16	6/13/16
SJC05	Single source		4.9-7.6	9/29/15	12/17/15	3/8/16	6/13/16
SJC06	Single source		7.0-10.9	9/29/15	12/17/15	3/8/16	6/13/16
SJC07	Single source		4.3-6.1	9/29/15	12/17/15	3/8/16	6/13/16

[a] Theoretical detention times were calculated by dividing total distribution system storage volume by average water volume treated per day for each month that samples were collected (Sept., Dec., March, and June).

Table S2. Designated components for the 4-component PARAFAC model developed from Q1-4 IC and SJC samples

Component	Ex./Em. maxima		Probable character/origin	DBP relevance	Analogues from previous work
	λ_{ex} (nm)	λ_{em} (nm)			
1	250, 370	460	Terrestrial humic-like[a]	THM and HAA precursor (strongly correlated)	C4 [4] C2 [5] C1 [6] C1 [7] C2 [8]
2	250, 330	400	Terrestrial/anthropogenic humic-like[b]	THM and HAA precursor	C1 [4] C3 [5] C2 [6] C2 [7] C3 [8]
3	260, 410	510	Humic-like[c]	THM and HAA precursor	C4 [4] C2 [5]
4	245	390	Protein/amino acids, tryptophan-like (possibly microbially-derived)[d]	-	C2 [4] C4 [6] C4 [8]

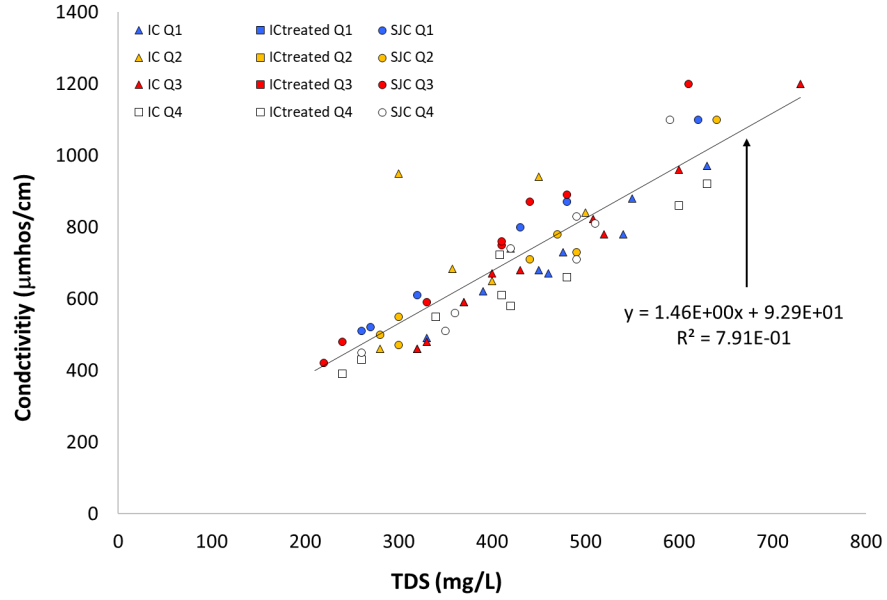
[a]Components likely associated with constituents exhibiting characteristics typical of terrestrial humic substances – i.e., complex macromolecular products resulting from the chemical and/or microbial decay/transformation of vegetative and microbial biomass in terrestrial ecosystems (e.g., grasslands, forests, wetlands), [b]Components likely associated with constituents comprising a mixture of humic substances of natural terrestrial origin and chemically-similar substances originating from anthropogenic sources (e.g., municipal or industrial wastewater, septic leachate), [c]Components likely associated with constituents exhibiting characteristics of broadly-defined humic substances without distinguishing traits of either natural terrestrial or anthropogenic origin, [d]Components likely associated with biomolecules generated during microbial growth (e.g., amino acids, proteins) with fluorescence signatures similar to tryptophan.



120

121 **Figure S1.** Bromide concentration versus chloride concentration for all raw water samples
122 collected during the study for each site.

123



124

125 **Figure S2.** Conductivity versus total dissolved solids (TDS) of the source waters for all four
126 quarters.

127

128

129

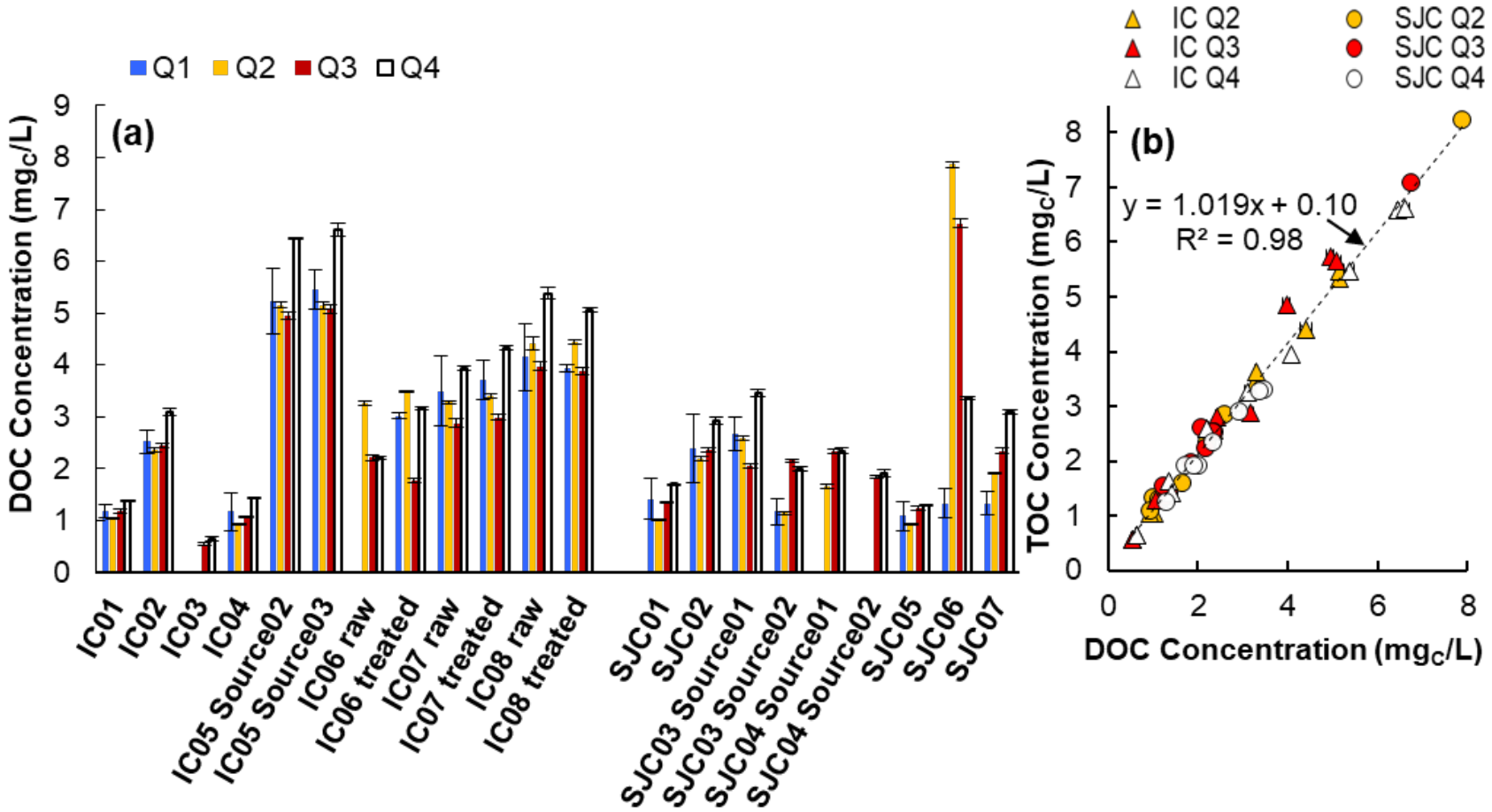


Figure S3. (a) DOC concentrations (Q1–4), and (b) correlation of TOC and DOC concentrations (Q2–4) for IC and SJC samples. Data values and error bars represent means and standard deviations, respectively, obtained from duplicate measurements.

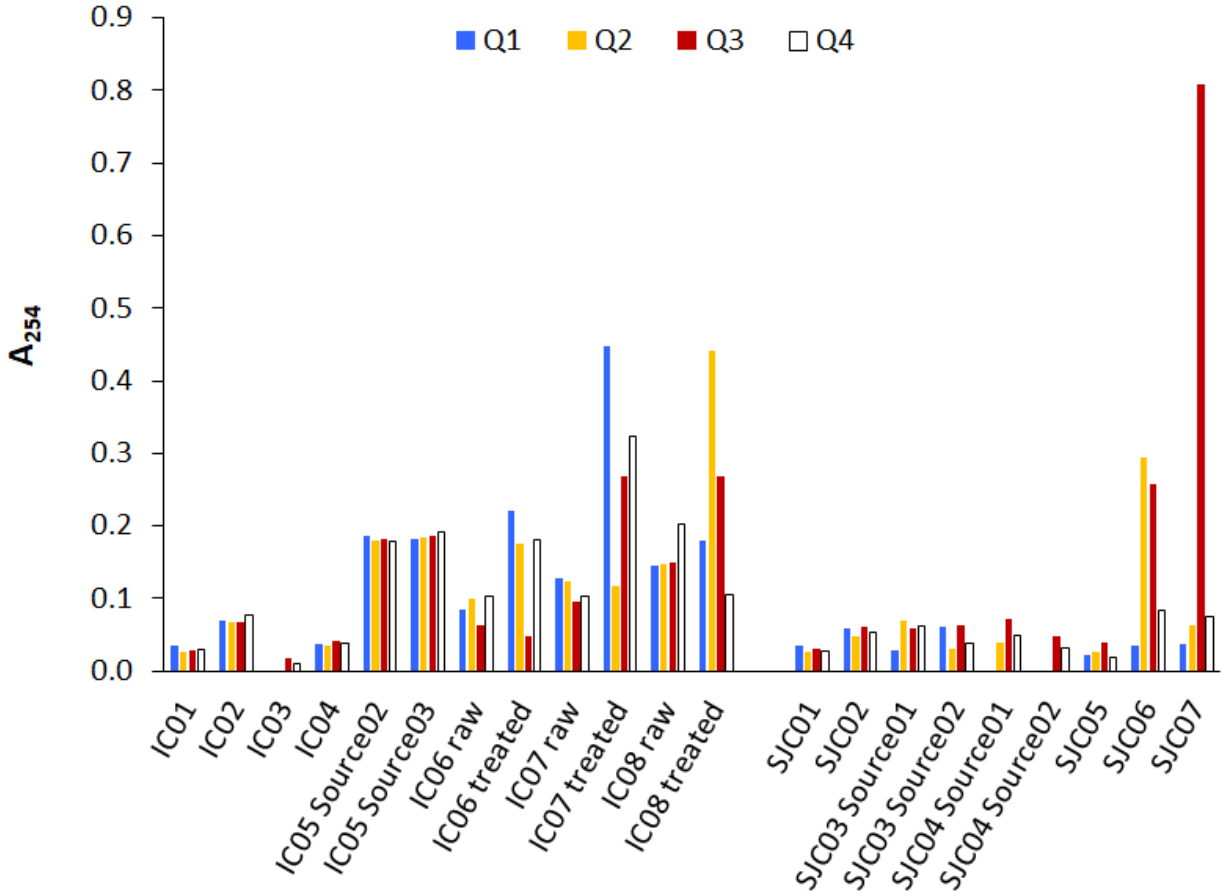


Figure S4. A_{254} measurements (Q1–4) for IC and SJC samples.

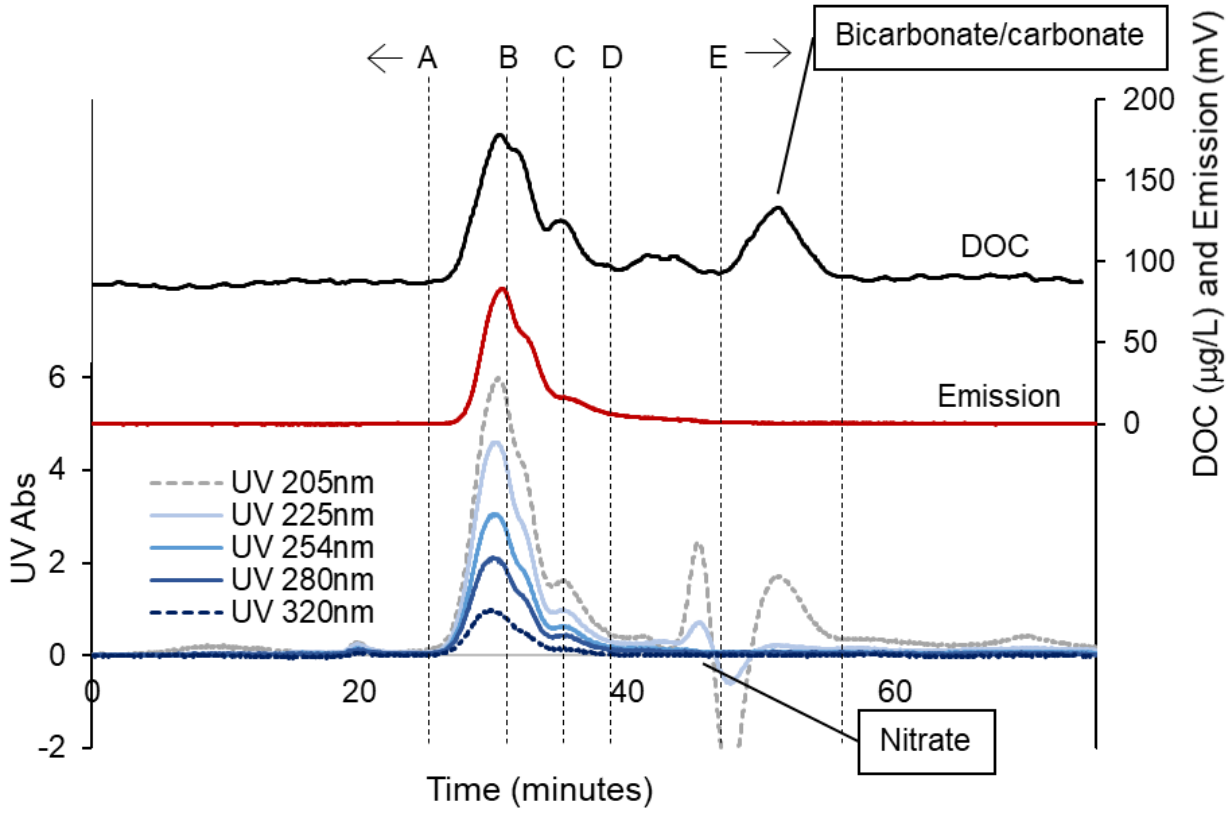
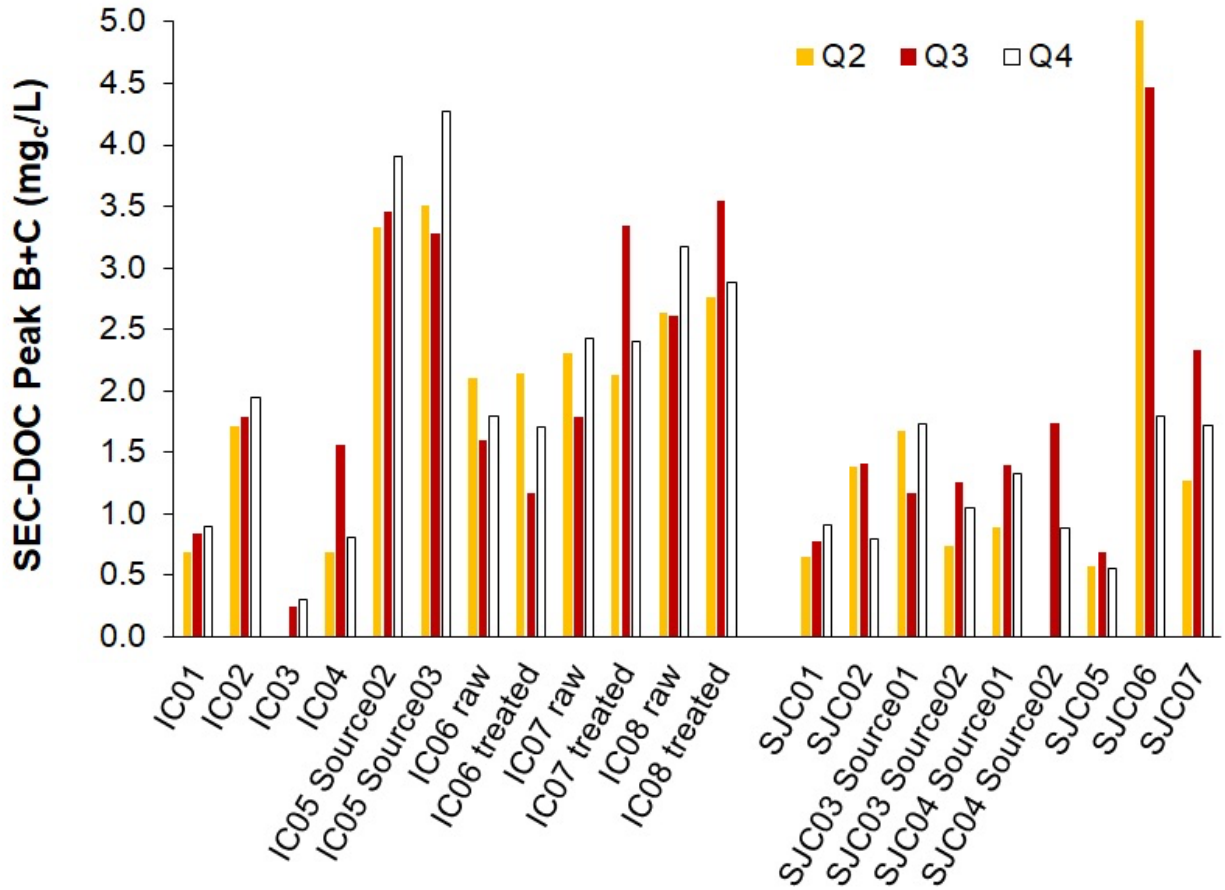


Figure S5. Example SEC chromatogram of IC05 Source03 from quarter 3 (note that the DOC and fluorescence emission chromatograms provided are plotted against the secondary y-axis, right-side).



140

141 **Figure S6.** SEC-DOC Peak B+C concentration measurements (Q2–4) for IC and SJC samples.

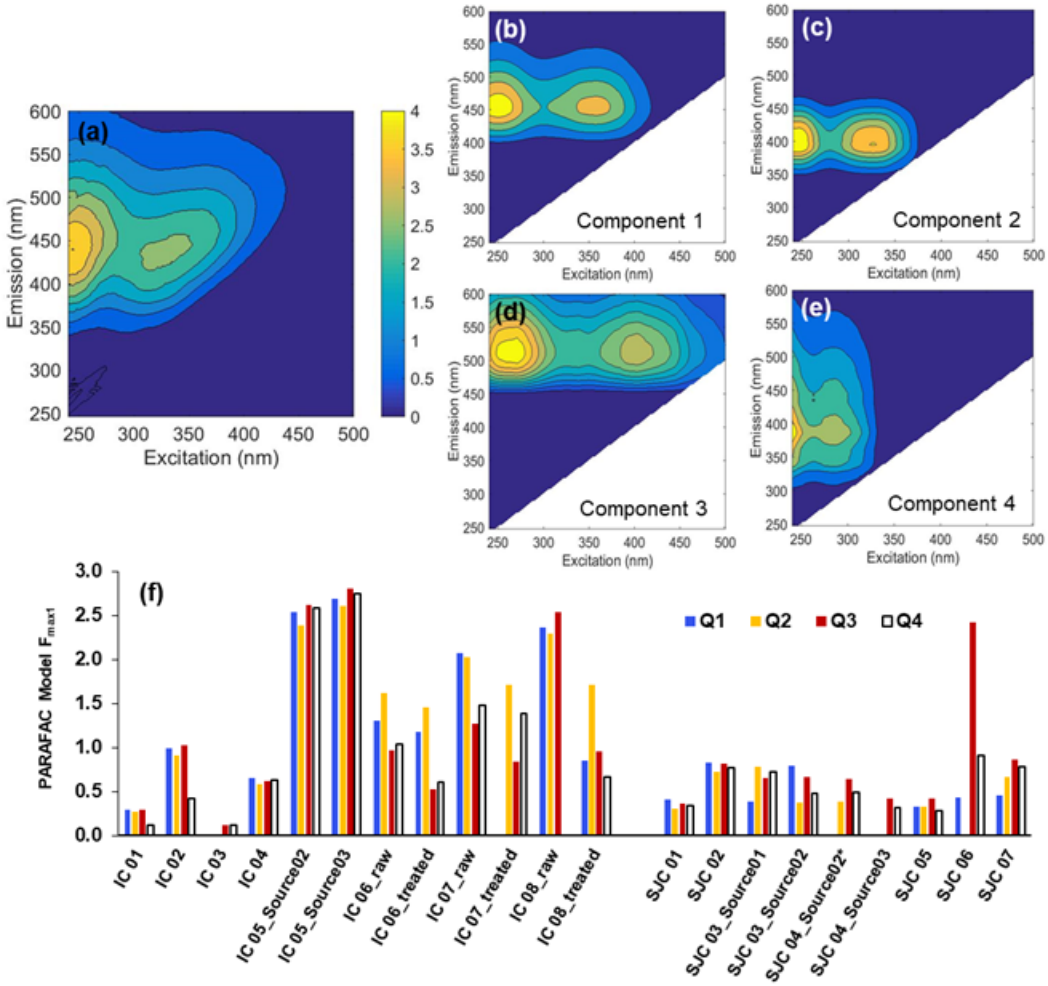
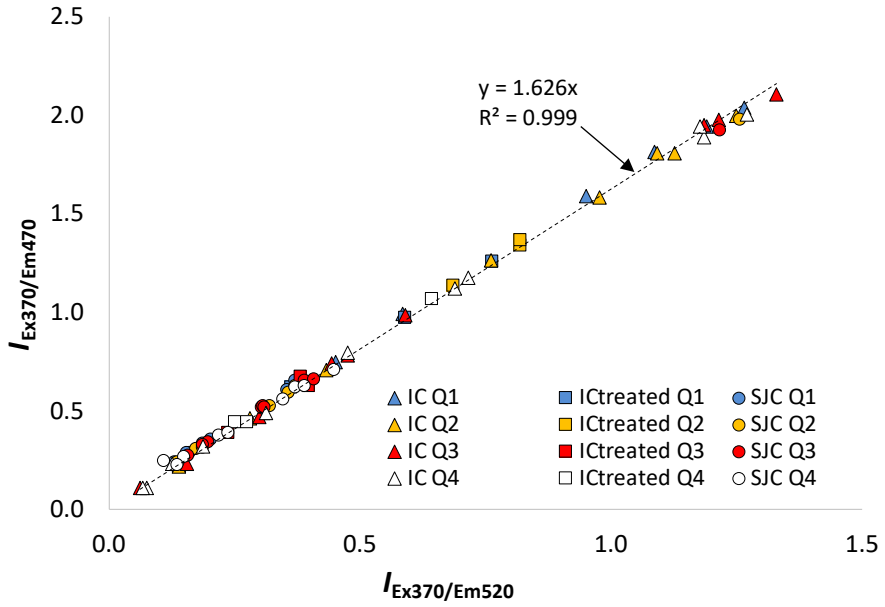


Figure S7. (a) Representative measured fluorescence EEM, obtained for the Q3 IC05 Source03 raw water sample; (b)-(e) EEM fluorescence spectra for Components 1-4 as defined in the PARAFAC model; and (f) F_{max1} values corresponding to Component 1, generated for each sample by the PARAFAC model.

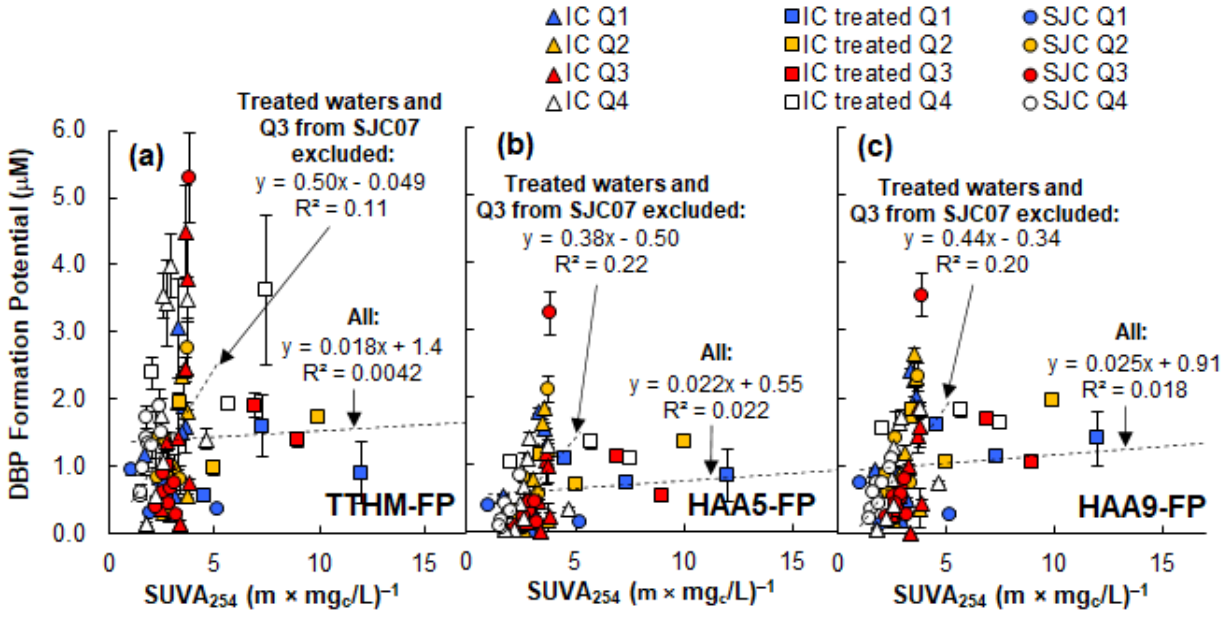


147

148 **Figure S8.** Magnitude of fluorescence emission intensity at λ_{em} = 470 nm versus emission
149 intensity at λ_{em} = 520 nm (both at an excitation wavelength, λ_{ex} = 370 nm) for each sample. The
150 average fluorescence index across the sample set (i.e., the ratio of the former to the latter[9, 10]),
151 was calculated from the slope of the line resulting from least squares regression of the dataset.

152

153



154

155 **Figure S9.** (a) TTHM-FP; (b) HAA5-FP and; (c) HAA9-FP plotted versus $SUVA_{254}$. Data points
156 and error bars represent means and standard deviations, respectively, obtained from at least
157 duplicate (and typically triplicate) experimentally-independent measurements.

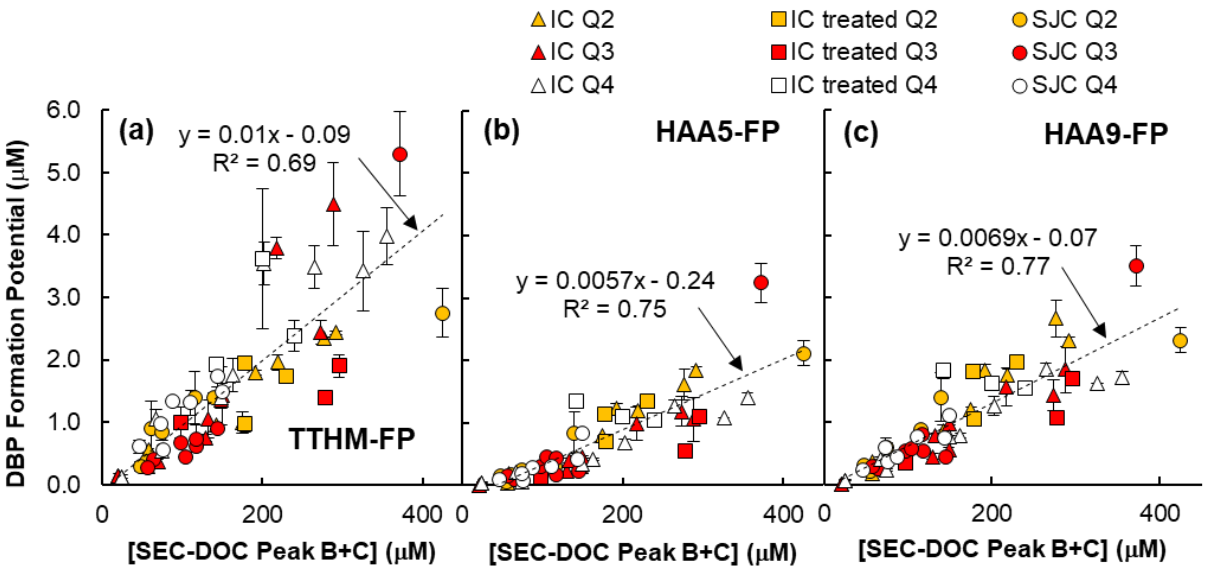


Figure S10. (a) TTHM-FP; (b) HAA5-FP and; (c) HAA9-FP plotted versus [SEC-DOC Peak B+C]. Data points and error bars represent means and standard deviations, respectively, obtained from at least duplicate (and typically triplicate) experimentally-independent measurements.

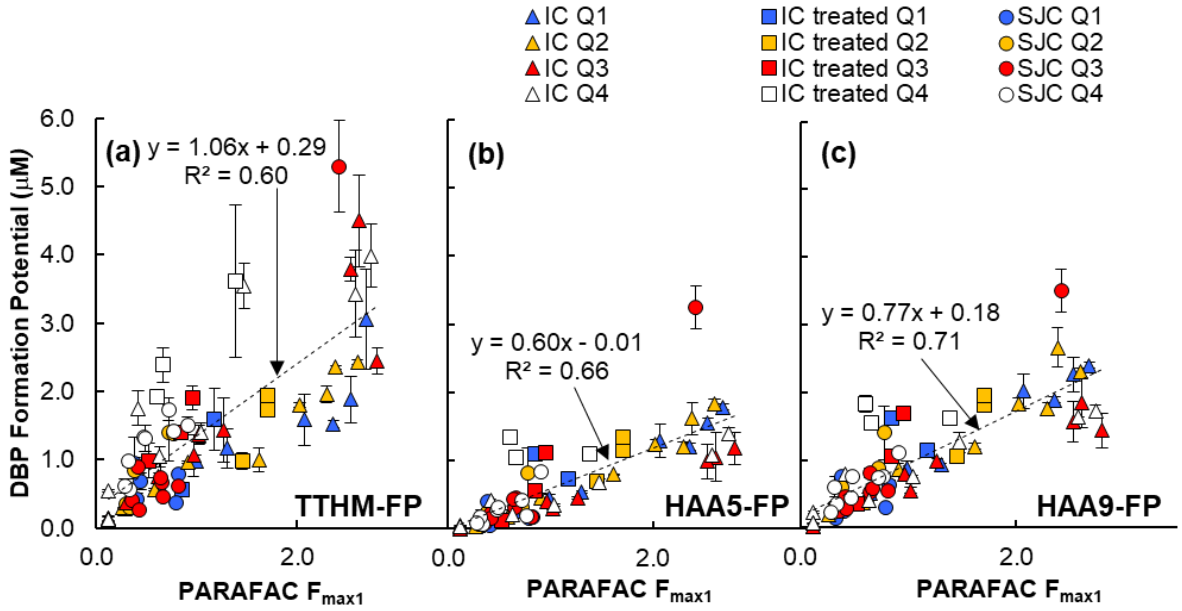
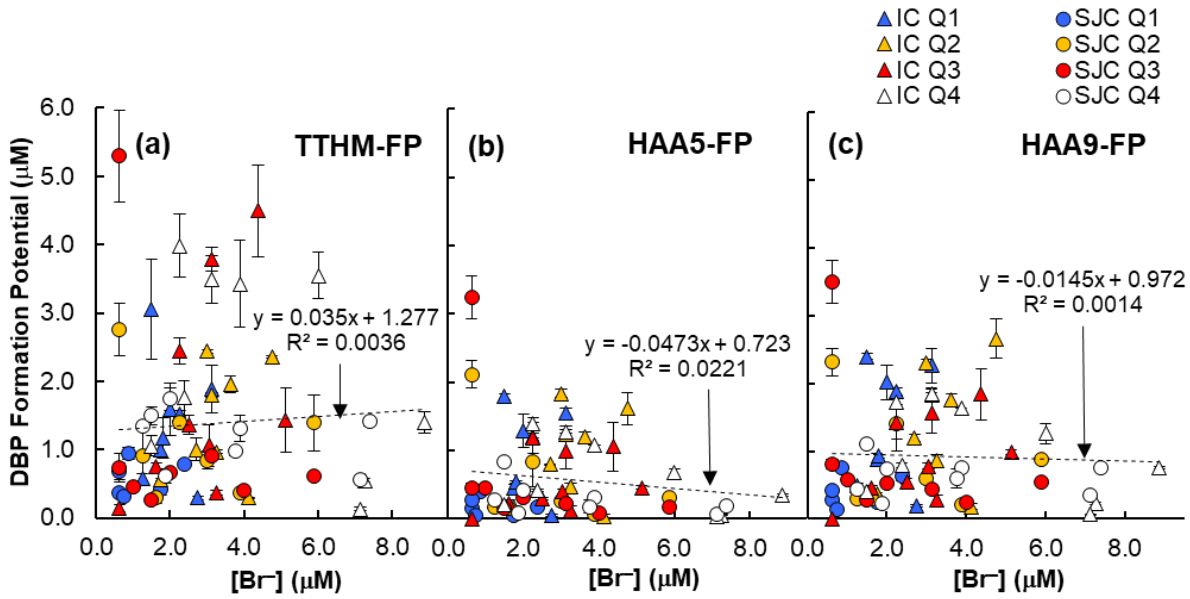


Figure S11. (a) TTHM-FP; (b) HAA5-FP and; (c) HAA9-FP plotted versus F_{max1} values generated by the PARAFAC model. Measurements of F_{max1} were obtained at native sample pH. Data points and error bars represent means and standard deviations, respectively, obtained from at least duplicate (and typically triplicate) experimentally-independent measurements.

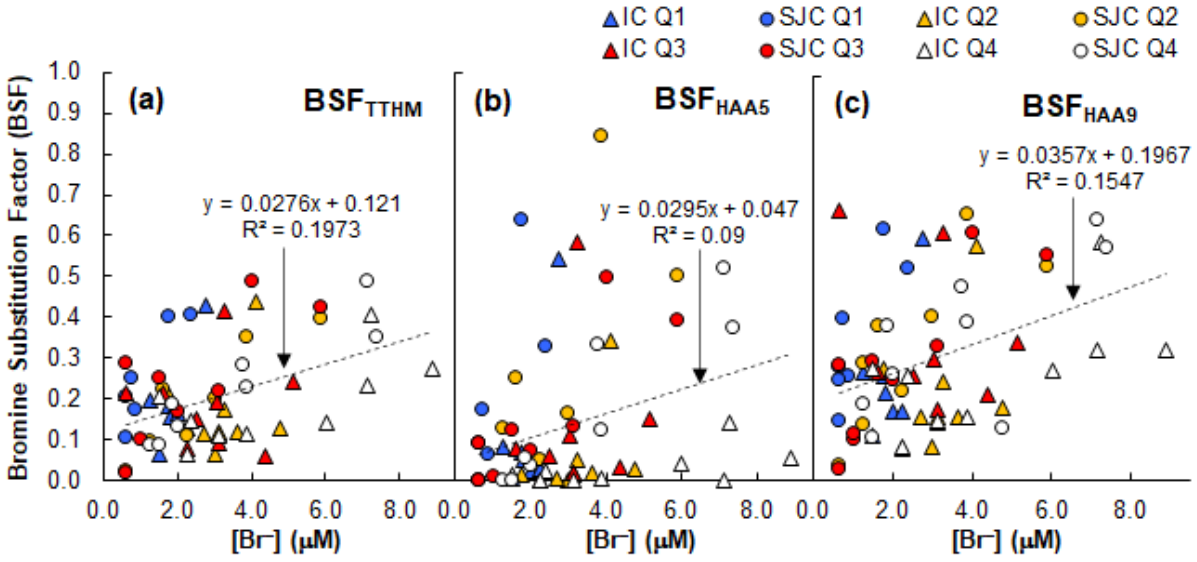


170

171 **Figure S12.** (a) TTHM-FP; (b) HAA5-FP; and (c) HAA9-FP plotted versus [Br^-] in raw water
172 samples.

173

174

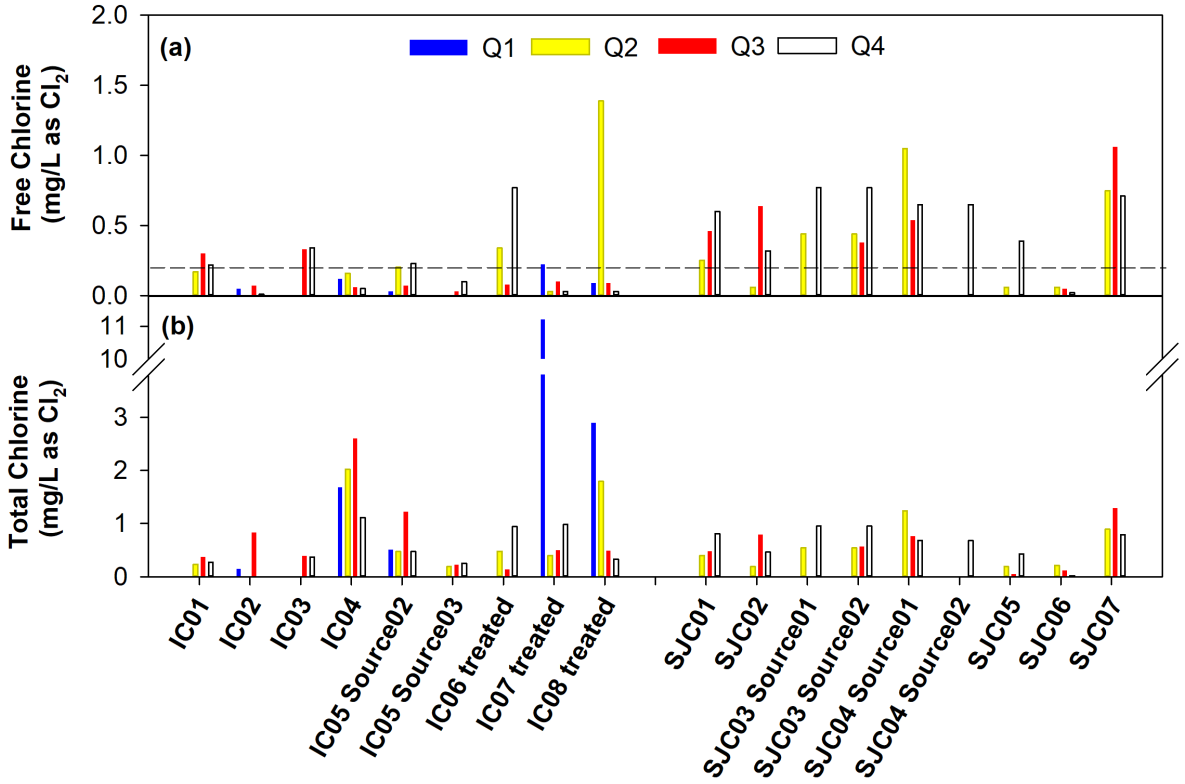


175

176 **Figure S13.** (a) BSF_{TTHM}; (b) BSF_{HAA5}; (c) BSF_{HAA9} plotted versus [Br^-] in raw water samples.

177

178



179

180 **Figure S14.** (a) Average free chlorine residuals (black dashed line is 0.2 mg/L as Cl_2); and (b)
181 average total chlorine residuals following full-scale chlorination at each site. Note the break in
182 the y-axis in panel (b), which only applies to Q1 for source IC06.

183

References

1. F. Hammes, F. Goldschmidt, M. Vital, Y. Y. Wang, T. Egli, Measurement and interpretation of microbial adenosine tri-phosphate (ATP) in aquatic environments, *Water Res.*, 2010, 44(13), 3915-23.
2. C. A. Stedmon, R. Bro, Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial, *Limnol. Oceanogr.: Methods*, 2008, 6(572-9.
3. Sentry Internet - Washington State Water System Data [Internet]. Available from: https://fortress.wa.gov/doh/eh/portal/odw/si/Intro.aspx.
4. A. D. Pifer, J. L. Fairey, Improving on SUVA 254 using fluorescence-PARAFAC analysis and asymmetric flow-field flow fractionation for assessing disinfection byproduct formation and control, *Water Res.*, 2012, 46(9), 2927-36.
5. S. K. Ishii, T. H. Boyer, Behavior of reoccurring PARAFAC components in fluorescent dissolved organic matter in natural and engineered systems: a critical review, *Environ. Sci. Technol.*, 2012, 46(4), 2006-17.
6. S. Baghoth, S. Sharma, G. Amy, Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC, *Water Res.*, 2011, 45(2), 797-809.
7. D. W. Johnstone, N. P. Sanchez, C. M. Miller, Parallel factor analysis of excitation–emission matrices to assess drinking water disinfection byproduct formation during a peak formation period, *Environmental Engineering Science*, 2009, 26(10), 1551-9.
8. K. M. Beggs, R. S. Summers, Character and chlorine reactivity of dissolved organic matter from a mountain pine beetle impacted watershed, *Environ. Sci. Technol.*, 2011, 45(13), 5717-24.

9. R. M. Cory, M. P. Miller, D. M. McKnight, J. J. Guerard, P. L. Miller, Effect of instrument-specific response on the analysis of fulvic acid fluorescence spectra, *Limnol. Oceanogr.: Methods*, 2010, 8(2), 67-78.

10. D. M. McKnight, E. W. Boyer, P. K. Westerhoff, P. T. Doran, T. Kulbe, et al., Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity, *Limnol. Oceanogr.*, 2001, 46(1), 38-48.