

1 **Molecular signature of organic nitrogen in septic-impacted groundwater**
2

3 William A. Arnold^{†,‡,*}, Krista Longnecker[§], Kevin D. Kroeger^{||}, Elizabeth B. Kujawinski[§]
4

5 [†]Department of Civil, Environmental, and Geo- Engineering, University of Minnesota, 500
6 Pillsbury Dr. SE, Minneapolis, MN 55455, USA

7 [‡]Institute on the Environment, University of Minnesota, 1954 Buford Avenue, Saint Paul, MN
8 55108, USA

9 [§]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution,
10 360 Woods Hole Rd., Woods Hole, MA 02543, USA

11 ^{||}U.S. Geological Survey Woods Hole Coastal & Marine Science Center, 384 Woods Hole Road,
12 Woods Hole, MA 02543 USA

13
14 *Author to whom correspondence should be addressed.

15 Phone: 612-625-8582; Fax: 612-626-7750; email: arnol032@umn.edu
16

17
18 Revised manuscript submitted to *Environmental Science: Processes & Impacts*
19
20
21
22
23
24
25

26 **Abstract**

27

28 Dissolved inorganic and organic nitrogen levels are elevated in aquatic systems due to
29 anthropogenic activities. Dissolved organic nitrogen (DON) arises from various sources, and its
30 impact could be more clearly constrained if specific sources were identified and if the molecular-
31 level composition of DON were better understood. In this work, the pharmaceutical
32 carbamazepine was used to identify septic-impacted groundwater in a coastal watershed. Using
33 ultrahigh resolution mass spectrometry data, the nitrogen-containing features of the dissolved
34 organic matter in septic-impacted and non-impacted samples were compared. The septic-
35 impacted groundwater samples have a larger abundance of nitrogen-containing formulas.
36 Impacted samples have additional DON features in the regions ascribed as ‘protein-like’ and
37 ‘lipid-like’ in van Krevelen space and have more intense nitrogen-containing features in a
38 specific region of a carbon versus mass plot. These features are potential indicators of dissolved
39 organic nitrogen arising from septic effluents, and this work suggests that ultrahigh resolution
40 mass spectrometry is a valuable tool to identify and characterize sources of DON.

41

42 **Introduction**

43 Groundwater is an important source of freshwater to coastal systems.^{1,2} Thus,
44 groundwater is a source of nitrogen to coastal waters, especially if it has been impacted by
45 human activities. Dissolved nitrogen is comprised of both inorganic (nitrate, nitrite, and
46 ammonia) and organic forms. The quantitative balance between inorganic and organic nitrogen
47 will depend on various factors, including sources and biological activity, but evidence suggests
48 that dissolved organic nitrogen (DON) can be quantitatively as important as inorganic forms of
49 nitrogen in groundwater.³ For example, Kroeger et al.³ found that DON comprised 10-93% of
50 total dissolved nitrogen in a set of coastal watersheds. Both the source and environmental
51 conditions influence the bioavailability and processing of DON, with bioavailability ranging
52 from 1-60%.^{4,5} Sources of DON include precipitation, agriculture, natural runoff/infiltration,
53 artificial recharge of wastewater, and septic systems. In populated coastal areas, septic systems
54 may be particularly important sources of both inorganic and organic nitrogen.^{3,6-8} At present,
55 however, it is difficult to link the bulk DON found in impacted aquatic systems to specific
56 sources or processes due to a lack of information about the molecular-level composition of DON.

57 Information regarding the molecular-level composition of DON is important to
58 improving understanding the reactivity of the pool of nitrogen in aquatic systems. The amount
59 of DON present is changed (both increases and decreases have been observed) by human
60 alteration of the landscape or introduction of runoff/effluents.^{9, 10} For example, septic systems are
61 a major source of dissolved nitrogen in low-residential watersheds.¹¹ Compositional changes in
62 the DON pool in human altered systems have received limited attention, and such changes have
63 potentially important ramifications. For example, DON is more recalcitrant than inorganic
64 nitrogen, leading to longer residence times and a larger zone of impact. Additionally, DON
65 concentration and composition will affect water treatment and human exposure to nitrogenous
66 disinfection byproducts.¹²

67 Ultrahigh resolution mass spectrometry techniques, such as Fourier transform ion
68 cyclotron resonance mass spectrometry (FT-ICR-MS), are used to gain insight into the
69 molecular-level composition of dissolved organic matter (DOM).¹³⁻¹⁷ Such instruments provide
70 the opportunity to see differences among samples that may not be apparent through bulk analyses
71 (e.g., measurement of dissolved organic carbon (DOC)) or targeted analysis of specific
72 molecules. FT-ICR-MS has been used to analyze incorporation of nitrogen into organic matter
73 and transformations and sources of the DON pool.¹⁸⁻²⁰ While recent studies have shown
74 compositional differences between effluent organic matter and natural organic matter,^{21, 22} a
75 specific focus on the nitrogen-containing component has yet to be undertaken. That said, past
76 work on septic-impacted groundwater on Cape Cod (MA) noted that a large portion of assigned
77 formulas contained nitrogen, potentially due to microbial activity and only partial mineralization
78 of septic contributions to the DON pool.¹⁷

79 Groundwater impacted by septic system effluent contains numerous human-use
80 compounds, including antibiotics, over-the-counter pharmaceuticals, steroid estrogens,
81 prescription medications, artificial sweeteners, and personal care products.²³⁻²⁷ On Cape Cod,
82 studies have found such compounds in impacted groundwater, which also serves as the drinking
83 water source for the population.^{27, 28} Some of these compounds, including carbamazepine (an
84 anti-convulsant) and artificial sweeteners, appear to be persistent and sorb to soils or sediments
85 only to a limited extent. Thus, they have been proposed as markers for wastewater/septic impact
86 and for contaminant source tracking.^{26, 29-32} In particular, carbamazepine is generally recognized
87 as being persistent in groundwater and an indicator of impact on natural waters by untreated or
88 treated wastewater.^{29, 30} Laboratory studies have shown carbamazepine to be stable in on-site
89 treatment scenarios,³³ and recent results demonstrated the utility of carbamazepine in
90 discriminating nutrient sources in coastal groundwater.⁸

91 On Cape Cod, the vast majority of residents have on-site wastewater treatment,³⁴ and
92 nationwide, 20% of households rely on septic systems. Even when functioning properly, septic
93 systems are a source of nitrogen to groundwater. Because of the prevalence of septic systems on
94 Cape Cod, past detection of pharmaceuticals in groundwater,^{27, 28} known nitrogen pollution
95 issues,^{3, 35, 36} and the fact that the aquifer to which septic systems discharge is also the source of
96 drinking water, this is a logical site to investigate the composition of DON in groundwater with
97 and without potential septic system impact. Our goals were to both discern septic impacted sites
98 and begin to characterize the DON arising from septic tanks. In this work, electrospray ionization
99 (ESI) FT-ICR-MS was used to obtain detailed information about the dissolved organic nitrogen
100 component of DOM, with carbamazepine used as a marker to identify impacted vs. non-
101 impacted waters.

102

103 **Experimental**

104 All solvents were obtained from Fisher and were Optima[®] grade. Trace metal grade
105 hydrochloric acid was also from Fisher. Ultrapure water (18.2 MΩ·cm) was produced by a
106 Millipore Elix 3 and Gradient system. Carbamazepine (Sigma-Aldrich) standards (0.1-1,000
107 ng/mL and a d₁₀-carbamazepine (Cambridge Isotopes) spiking solution in methanol were
108 prepared previously.³⁷ All groundwater samples were collected via a peristaltic pump into
109 prewashed (overnight soak with soap and water, rinsed with ultrapure water, soaked in 10% HCl,
110 rinse with ultrapure water) and autoclaved 2.5 L polycarbonate bottles. The bottles were rinsed
111 three times with ~1 L of collected groundwater and then filled to capacity, with 5 L being
112 collected for each sample. The samples were transported on ice to the laboratory, where they
113 were filtered through 0.2 μm, 47 mm hydrophilic PTFE Omnipore[™] membranes (EMD-
114 Millipore) and acidified to pH 3 using concentrated HCl. If not extracted immediately, the
115 filtered and acidified samples were stored at 4 °C.

116 *Sample sites*

117 Four sites in Falmouth (Cape Cod), MA were sampled in September and October of
118 2013. The shallow groundwater at two of the sites, Head of Waquoit Bay and Sage Lot Pond,
119 with population densities of 190 and 0 people per km² respectively, was expected to have
120 minimal, if any, impact from septic systems.³ The groundwater sample at Head of Waquoit Bay
121 was taken within the Waquoit Bay National Estuarine Research Reserve east of the boathouse at
122 the north point of the transect described by Kroeger and Charette,³⁶ using a small diameter drive
123 point piezometer.³⁸ Groundwater at the Sage Lot Pond site was collected in a forested area from
124 previously installed wells. The other two sites, Green Pond and Site D, were near residences in

125 more densely populated areas (1,000-1,500 people per km²)³ and were expected to be impacted
126 by septic system effluent. Groundwater at Green Pond was collected on the west shore
127 approximately 0.4 kilometers north of the pond outlet into Vineyard Sound. A small diameter
128 drive point piezometer was used approximately 5 m from shore. At Site D, a densely populated
129 residential area west of Waquoit Bay, samples were collected at three depths (6.30, 9.97, and
130 16.11 m below ground surface) during ongoing USGS sampling efforts from a multilevel well.
131 These specific depths for characterization of the DOM and quantification of carbamazepine were
132 chosen based on nitrate and ammonia levels observed in previous sampling efforts to capture
133 potentially different redox zones.

134 *DOC and water quality analysis*

135 For the six samples described above, DOC and total dissolved nitrogen (TDN) were
136 measured with a Shimadzu TOC V_{CSH} with a total nitrogen attachment. Conductivity was
137 measured with a YSI 556MPS handheld meter. Nutrients were measured using an AA3 four-
138 channel segmented flow autoanalyzer (SEAL Analytical). DON was calculated by subtracting
139 the nitrate/nitrite and ammonia concentrations from the TDN. Water quality data from other
140 sampling depths at Site D were provided by the USGS.

141 *Carbamazepine*

142 Duplicate 1-L samples of collected groundwater were spiked with 5 µL of 2 µg/mL d₁₀-
143 carbamazepine for use as a surrogate and extracted following the protocols of ref.³⁷ using Empore
144 SDB-RPS disks (3M) that had been pre-conditioned with acetone, isopropyl alcohol, methanol,
145 and ultrapure water. The disks were stored at -20 °C until extraction. The disks were warmed to
146 room temperature and eluted with 10 mL of methanol and 10 mL of acetone. The combined
147 extract was blown down with nitrogen to a volume of ~5 mL and then reduced to near-dryness

148 via vacuum centrifugation. The extract was then reconstituted with 60 μ L of methanol and 140
149 μ L of ultrapure water. Carbamazepine and d₁₀-carbamazepine were quantified by ultrahigh
150 pressure liquid chromatography tandem mass spectrometry (Thermo TSQ Vantage) with the
151 electrospray ion source in positive mode according to previously described methods.^{37, 39}

152 *ESI-FT-ICR-MS*

153 Samples (1-L) were extracted using Bond Elut PPL cartridges (Agilent) following the
154 protocol of Dittmar et al.⁴⁰ The resulting methanol extracts were analyzed via direct infusion
155 with the electrospray ionization interface of a 7T LTQ FT-ICR-MS (LTQ FT-Ultra, Thermo
156 Scientific). The D2 sample was analyzed twice to test instrument reproducibility. The methanol
157 extracts were diluted with a solution of 70% methanol/30% water (2-fold dilution by volume for
158 Site D samples, 3-fold for Green, and 10-fold for Head of Waquoit and Sage Lot) to achieve a
159 stable spray and analyzed in positive mode, because nitrogen functional groups are generally
160 more amenable to detection under positive ionization. The resolving power was 400,000 at m/z
161 400, and 200 transients were collected per sample. The transients were co-added and processed
162 using the protocols described in Kido Soule et al.⁴¹

163 After correcting the m/z values of the features detected via ESI-FT-ICR-MS for the
164 presence of sodium, a previously developed algorithm^{42, 43} was used to assign elemental formulas
165 with an allowed error of 1 ppm. The algorithm allows formulas including C, H, O, N, S, and P
166 and allows inclusion of ¹³C if the error is reduced by doing so. All data analysis and processing
167 were performed in MATLAB R2013b (MathWorks, Inc.).

168

169 **Results and Discussion**

170 *Dissolved organic carbon, dissolved nitrogen and carbamazepine*

171 Sample collection depths, dissolved organic carbon, total dissolved nitrogen, and
172 carbamazepine data are shown in Table 1. The specific conductance values demonstrate that all
173 samples were collected from fresh water zones, rather than salt water intrusion where specific
174 conductance would be $> 30,000 \mu\text{S}/\text{cm}$.

175 The DOC levels at Head of Waquoit and Sage Lot are higher than those at Green Pond
176 and Site D. The lower DOC concentration for the D1 sample may be due to the deeper vadose
177 zone at Site D, which would allow greater DOC mineralization during transport to the water
178 table. At Site D, the DOC level decreases with depth, consistent with previous observations on
179 Cape Cod,⁴⁴ and with a greater extent of degradation for organic matter that has traveled further
180 or is older (see below).

181 The total dissolved nitrogen and nitrate+nitrite (Table 1) at Green Pond and Site D are
182 substantially higher than the levels at Head of Waquoit and Sage Lot, supporting the
183 expectations regarding the degree of anthropogenic impact at the two sites. The nitrate+nitrite is
184 likely derived from septic discharge and fertilizer application. The DON levels are lowest at the
185 two unimpacted sites (Head of Waquoit and Sage Lot), but at these sites, the DON comprises the
186 largest fraction of TDN ($>65\%$). This indicates decreased processing of the organic nitrogen and
187 less septic impact. At Site D, the DON levels are equal or higher than those at Head of Waquoit
188 or Sage Lot, but DON is only 4-8% of the total TDN. The elevated levels, however, clearly
189 indicate impact from septic systems. While the nitrate levels at Green Pond also indicate septic
190 influence, the calculated DON level is zero. This is in contrast to previous samples from
191 groundwater around Green Pond, which saw levels $\sim 2 \text{ mg}/\text{L}$ as DON.³ This does not mean,
192 however, that one would not expect to observe N-containing formulas via ESI-FT-ICR-MS in
193 our sample. The zero value for DON merely indicates that determining this value as the

194 difference between TDN and nutrients does not have the required precision and that additional
195 analyses are needed to evaluate the role of septic systems as a source of organic nitrogen.

196 For the Head of Waquoit and Sage Lot groundwater samples, average carbamazepine
197 concentrations were below the limit of quantification (1.6 ng/L). Carbamazepine was detected at
198 concentrations ranging from 7.8 to 46.7 ng/L in the samples from Green Pond and Site D. These
199 levels are consistent with those in various wastewater-impacted groundwaters^{25, 30, 32} and a recent
200 analysis of drinking waters drawn from the shallow aquifer on Cape Cod.²⁸

201 It is clear that the samples collected from shallow depths (Green and D1) as well as the
202 deeper samples D2 and D3 are impacted by septic effluent based on the presence of
203 carbamazepine. Yet, the trend of values with depth at Site D merits discussion. Groundwater on
204 Cape Cod flows in layers,^{45, 46} with deeper waters travelling farther distances and being of
205 greater age. Thus, the water from the shallow depth near the water table (D1) is likely derived
206 from proximal sources and influenced by local septic systems. The water at the greatest depth
207 (D3) is likely 10-20 years old^{44, 46, 47} and has travelled from the recharge mound further north.
208 The presence of carbamazepine at this depth suggests that the water is from an area impacted by
209 septic systems or wastewater recharge and supports the notion that carbamazepine is persistent in
210 groundwater systems. The carbamazepine concentration at the intermediate depth is lowest. This
211 may indicate that the region in which this groundwater layer receives recharge is less
212 impacted/populated than the area providing recharge to the deeper groundwater.

213 *Molecular characterization of the organic matter*

214 For the seven samples analyzed via ESI-FT-ICR-MS (including the duplicate runs for
215 sample D2), 12,992 features were detected and 11,434 formulas were assigned. Because of the
216 small number of samples, cluster analysis was performed based on the features found in each

217 sample, and the dendrogram is shown in the Supplementary Information (Figure S1). The two
218 unimpacted sites, Head of Waquoit and Sage Lot, clustered together as did the replicate
219 injections for D2. Overall, the calculated differences were not large (all distances between 0.2
220 and 0.45), and the clustering was not used to guide further analyses.

221 The total formulas and the percentage of formulas in each sample containing only CHO
222 and CHON are given in Table 2. Also given are the percentages of all formulas found by the
223 algorithm that contained N in any allowed combination with other elements. In general, Head of
224 Waquoit and Sage Lot (the two unimpacted sites) had a higher percentage ($\geq 56\%$) of CHO
225 formulas compared to the sites where carbamazepine was detected ($\leq 56\%$). Using a 1-tail t-test
226 with equal variance (as determined by an F-test) to evaluate if the impacted samples have a lower
227 CHO level, gives a p-value of 0.115 (significant with 88% confidence). Similarly, the CHON
228 percentages ($\leq 26\%$) and all N-containing formulas ($\leq 40\%$) at the unimpacted sites are lower than
229 at the impacted sites (28-36% for CHON and 41-56% for all N). In this case, the t-test gives a p-
230 value of 0.063 (significant with 93% confidence). Previous comparisons of effluent organic
231 matter from a wastewater treatment plant with Suwannee River natural organic matter analyzed
232 using negative mode ESI-FT-ICR-MS showed reduced abundance of CHO formulas in the
233 wastewater samples.²¹ The molar ratios of O:C and N:C (Table 2) also indicate that the Head of
234 Waquoit and Sage Lot samples are more oxidized (based on O:C ratio) and that the impacted
235 samples have higher nitrogen content. While these differences are subtle and more data would
236 be needed to derive more robust statistical certainty, these results suggest that septic-impacted
237 groundwater has a greater abundance of N-containing formulas.

238 The van Krevelen plots for all features and for only N-containing features are shown in
239 the Supplementary Information (Figures S2-S7). The four samples collected from shallow depths

240 have a larger number of points with O:C ratios > 0.5 , with the Head of Waquoit and Sage Lot
241 samples having more than the Green Pond and D1 samples. This is true of the plots for all of the
242 detected features as well as those containing N, as quantified in Table 2. The exception is the N-
243 containing compounds for the Head of Waquoit sample, but the count is biased by the fact that
244 this sample has the fewest N-containing features. Overall, these results indicate a more reduced
245 character for the septic-impacted samples.

246 Differences in the N-containing compounds are more readily assessed by identifying
247 features that are common or unique among the samples. The samples were divided into three
248 pairs for this analysis: Head of Waquoit + Sage Lot, Green + D1, and D2 + D3. In each case, the
249 N-containing features present in both samples of one pair and not in both samples of the other
250 pair (and vice versa) were determined. Also found were the features contained in all four of the
251 compared samples.

252 In the van Krevelen diagrams shown in Figure 1a and 1b, the septic impacted samples are
253 compared to the unimpacted samples. It is apparent that the impacted samples have N-containing
254 features that are absent from the unimpacted samples. In comparing Green Pond and D1 with
255 Head of Waquoit and Sage Lot, the impacted samples contain features of H:C > 1.5 across the
256 O:C range that are not in the unimpacted samples (green points in Figure 1a). Similarly, D2 and
257 D3 contain features of H:C > 1.5 and O:C < 0.6 that are not in the unimpacted samples (red
258 points in Figure 1b). The features with H:C > 1.5 overlap with the regions of the diagram
259 generally ascribed to be 'protein-like' and 'lipid-like',⁴⁸ and it may be expected that the lipid and
260 protein content of human waste would be different than that of natural systems. Nitrogen-
261 containing compounds may also arise from household products/pharmaceuticals that are washed
262 down the drain.⁴⁹ A previous study identified the sulfur-containing linear alkyl benzene

263 sulfonates in wastewater effluent using ESI-FT-ICR-MS, supporting the hypothesis that human-
264 use chemicals alter the composition of the dissolved organic matter.²¹

265 In Figure 1c, the differences between the shallow and deep samples impacted by septic
266 effluents are shown. The shallow samples have a cluster of features with higher O:C ratios (0.2-
267 0.8, green points), and the deep samples (red points) have additional features with O:C < 0.3.
268 This could be consistent with reduction of the N-containing features in older, deeper samples in
269 which dissolved oxygen has been consumed.⁴⁸ Also supporting the greater processing of N-
270 containing compounds in the deep, impacted samples are the number of formulas in the ‘protein-
271 like’ area of the van Krevelen diagrams for the individual samples (181 and 387 for D2 and D3,
272 respectively), compared to D1 (528). The reduced numbers in the deep samples may indicate
273 degradation of a portion of the septic-derived N-containing features in these samples.

274 Another means to visualize the differences in nitrogen content are carbon versus mass
275 plots.⁵⁰ In such plots, the number of carbon atoms in a molecule is plotted versus the molecular
276 weight of the molecule, and the size of the point is proportional to the intensity of the ESI-FT-
277 ICR-MS signal. Shown in Figure 2 are such plots for each sample. Note that the four impacted
278 samples have abundant (number of dots) and intense (larger size) features with N:C \geq 0.5 from
279 carbon numbers 10 to 30 with m/z values from 400 to 700 (denoted by pink ovals). These
280 features are less intense and less numerous in the Head of Waquoit and Sage Lot samples,
281 indicating this pattern may be indicative of septic impacts. Interestingly, the four samples from
282 shallow depths all have a greater intensity of points in the upper right portion of the plot (carbon
283 number > 30, m/z > 600; black ovals) with low nitrogen content. The greatest intensity for these
284 points is in the two impacted samples (D1 and Green). While this is not an indicator of organic

285 nitrogen from the septic system effluent, the greater intensity in this region may be an indicator
286 of recent or less-processed septic input.

287 Both the van Krevelen diagrams and carbon versus mass plots show that it is possible to
288 identify features/molecular formulas of DON associated with septic-derived organic matter.
289 Elucidating molecular level composition is the first step to a predictive understanding of the
290 reactivity of the DON pool. Such features can be used to identify the relative contributions of
291 different sources of DON to the total nitrogen pool. This has the potential to allow the
292 assessment of the availability of DON in aquatic systems. With additional structural
293 characterization to identify the abundance of functional groups, it would also be possible to
294 evaluate impacts on water treatment (e.g., the formation of nitrogenous disinfection by products).

295 **Conclusion**

296 This work has provided molecular-level information regarding the DON in groundwaters
297 impacted by septic system effluent. Carbamazepine and nitrate/nitrite were found to be valid
298 indicators of septic influence, but depth, age, and recharge location of groundwater layers need
299 to be carefully considered when comparing these measurements. Overall, there appears to be a
300 greater abundance of N-containing features in DOM impacted by septic effluent, even if the
301 water is older and/or deeper. In van Krevelen space, the impacted samples had N-containing
302 features present at $H:C > 1.5$ that plotted in the ‘lipid-like’ and ‘protein-like’ areas of the
303 diagram. An abundance of features in this area of the diagram may be a useful marker of septic
304 impact in groundwater. The carbon versus mass plots also have abundant, intense features in a
305 specific region of the plot for the impacted samples. Thus, these plots may also be useful in
306 identifying the source of DON. Further work will be needed to identify features in van Krevelen
307 space and carbon versus mass plots that are indicators of other sources of DON in groundwater.

308 This work also demonstrates that unimpacted groundwater samples from the targeted watershed
309 are needed to allow comparisons to be performed to find features indicative of alteration of the
310 organic matter by human activities.

311 This work focused on abundance and intensity of N-containing features. Quantitative
312 sample-to-sample comparison of intensities is limited using ESI-FT-ICR-MS due to issues
313 associated with variations in ion suppression among samples. Thus, it is important to recognize
314 that while the features detected in this work are a first step in using ESI-FT-ICR-MS to develop a
315 molecular level understanding of DON, the overall impact of DON on aquatic systems will be a
316 combination of abundance, concentration, and bioavailability of specific DON-moieties, and
317 ESI-FT-ICR-MS analysis alone will not allow this assessment.

318

319 **Acknowledgements**

320 Thanks to Melissa Kido Soule for performing the carbamazepine and ESI-FT-ICR-MS analyses,
321 Paul Henderson for performing the nutrient analyses, Yina Liu for assistance with MATLAB, the
322 U.S. Geological Survey (Denis LeBlanc and John Colman) for assistance with sampling at Site D
323 and providing additional water quality data, and Matthew Charette for helpful discussions. The
324 Deep Ocean Exploration and Coastal Ocean Institutes at WHOI and the University of Minnesota
325 Institute on the Environment are thanked for financial support.

326

327 **References**

- 328 1. W. S. Moore, *Annu. Rev. Mar. Sci.*, 2010, **2**, 59-88.
- 329 2. A. E. Mulligan and M. A. Charette, *J. Hydrol.*, 2006, **327**, 411-425.
- 330 3. K. D. Kroeger, M. L. Cole and I. Valiela, *Limnol. Oceanogr.*, 2006, **51**, 2248-2261.
- 331 4. D. A. Bronk, Q. N. Roberts, M. P. Sanderson, E. A. Canuel, P. G. Hatcher, R. Mesfioui,
- 332 K. C. Filippino, M. R. Mulholland and N. G. Love, *Environ. Sci. Technol.*, 2010, **44**,
- 333 5830-5835.
- 334 5. S. P. Seitzinger, R. W. Sanders and R. Styles, *Limnol. Oceanogr.*, 2002, **47**, 353-366.
- 335 6. M. Badruzzaman, J. Pinzon, J. Oppenheimer and J. G. Jacangelo, *J. Environ. Manage.*,
- 336 2012, **109**, 80-99.
- 337 7. C. Meile, W. P. Porubsky, R. L. Walker and K. Payne, *Water Res.*, 2010, **44**, 1399-1408.
- 338 8. S. Zhao, P. Zhang, J. Crusius, K. D. Kroeger and J. F. Bratton, *J. Environ. Monit.*, 2011,
- 339 **13**, 1337-1343.
- 340 9. B. A. Pellerin, S. S. Kaushal and W. H. McDowell, *Ecosystems*, 2006, **9**, 852-864.
- 341 10. E. H. Stanley and J. T. Maxted, *Ecol. Appl.*, 2008, **18**, 1579-1590.
- 342 11. S. S. Kaushal, P. M. Groffman, L. E. Band, E. M. Elliott, C. A. Shields and C. Kendall,
- 343 *Environ. Sci. Technol.*, 2011, **45**, 8225-8232.
- 344 12. A. D. Shah and W. A. Mitch, *Environ. Sci. Technol.*, 2011, **46**, 119-131.
- 345 13. E. B. Kujawinski, P. G. Hatcher and M. A. Freitas, *Anal. Chem.*, 2002, **74**, 413-419.
- 346 14. C. E. Rostad and J. A. Leenheer, *Anal. Chim. Acta*, 2004, **523**, 269-278.
- 347 15. A. C. Stenson, A. G. Marshall and W. T. Cooper, *Anal. Chem.*, 2003, **75**, 1275-1284.
- 348 16. B. P. Koch, M. Witt, R. Engbrodt, T. Dittmar and G. Kattner, *Geochim. Cosmochim.*
- 349 *Acta*, 2005, **69**, 3299-3308.
- 350 17. K. Longnecker and E. B. Kujawinski, *Geochim. Cosmochim. Acta*, 2011, **75**, 2752-2761.
- 351 18. G. McKee and P. G. Hatcher, *Geochim. Cosmochim. Acta*, 2010, **74**, 6436-6450.
- 352 19. F. Schmidt, M. Elvert, B. P. Koch, M. Witt and K.-U. Hinrichs, *Geochim. Cosmochim.*
- 353 *Acta*, 2009, **73**, 3337-3358.
- 354 20. F. Schmidt, B. P. Koch, M. Elvert, G. Schmidt, M. Witt and K.-U. Hinrichs, *Environ. Sci.*
- 355 *Technol.*, 2011, **45**, 5223-5229.
- 356 21. M. Gonsior, M. Zwartjes, W. J. Cooper, W. Song, K. P. Ishida, L. Y. Tseng, M. K. Jeung,
- 357 D. Rosso, N. Hertkorn and P. Schmitt-Kopplin, *Water Res.*, 2011, **45**, 2943-2953.
- 358 22. P. F. Greenwood, L. J. Berwick and J.-P. Croue, *Chemosphere*, 2012, **87**.
- 359 23. J. Carrera, C. J. Ptacek, W. D. Robertson, D. W. Blowes, M. C. Moncur, E. Sverko and S.
- 360 Backus, *Environ. Sci. Technol.*, 2008, **42**, 2805-2811.
- 361 24. E. Godfrey, W. W. Woesnner and M. J. Benoit, *Ground Water*, 2007, **45**, 263-271.
- 362 25. D. J. Lapworth, N. Baran, M. E. Stuart and R. S. Ward, *Environ. Poll.*, 2012, **163**, 287-
- 363 303.
- 364 26. W. D. Robertson, D. R. Van Stempvoort, D. K. Solomon, J. Homewood, S. J. Brown, J.
- 365 Spoelstra and S. L. Schiff, *J. Hydrol.*, 2013, **477**, 43-54.
- 366 27. C. H. Swartz, S. Reddy, M. J. Benoit, H. Yin, L. B. Barber, B. J. Brownawell and R. A.
- 367 Rudel, *Environ. Sci. Technol.*, 2006, **40**, 4894-4902.
- 368 28. L. A. Schaidler, R. A. Rudel, J. M. Ackerman, S. C. Dunagan and J. G. Brody, *Sci. Total*
- 369 *Environ.*, 2014, **268-269**, 384-393.
- 370 29. G. Gasser, M. Rona, A. Voloshenko, R. Shelkov, O. Lev, s. Elhanany, F. T. Lange, M.
- 371 Scheure and I. Pankratov, *Desalination*, 2011, **273**, 398-404.

- 372 30. A. Jurado, R. Lopez-Serna, E. Vazquez-Sune, J. Carrera, E. Pujades, M. Petrovic and D.
373 Barcelo, *Chemosphere*, 2014, in press.
374 <http://dx.doi.org/10.1016/j.chemosphere.2014.01.014>.
- 375 31. M. E. Stuart, D. J. Lapworth, J. Thomas and L. Edwards, *Sci. Total Environ.*, 2014, **468-**
376 **469**, 564-577.
- 377 32. D. R. Van Stempvoort, J. W. Roy, J. Grabuski, S. J. Brown, G. Bickerton and E. Sverko,
378 *Sci. Total Environ.*, 2013, **461-462**, 348-359.
- 379 33. J. Teerlink, V. Martinez-Hernandez, C. P. Higgins and J. E. Drewes, *Water Res.*, 2012,
380 **46**, 5174-5184.
- 381 34. *Massachusetts Executive Office of Environmental Affairs, Cape Cod Watershed*
382 *Assessment and 5-Year Action Plan*, Boston, MA, 2004.
- 383 35. J. R. Barbaro, D. A. Walter and D. R. LeBlanc, *U.S. Geological Survey Scientific*
384 *Investigations Report 2013-5061*, 2013, 37 p. <http://pubs.usgs.gov/sir/2013/5061/>.
- 385 36. K. D. Kroeger and M. A. Charette, *Limnol. Oceanogr.*, 2008, **53**, 1025-1039.
- 386 37. D. R. Griffith, *Joint Program in Oceanography/Applied Ocean Science and Engineering*
387 *and Department of Civil and Environmental Engineering, Massachusetts Institute of*
388 *Technology/Woods Hole Oceanographic Institution, Ph.D. Thesis. Available at:*
389 <http://dspace.mit.edu/handle/1721.1/85824>, 2013, [Accessed May 20, 2014].
- 390 38. M. A. Charette and M. C. Allen, *Ground Water Monit. Remed.*, 2006, **26**, 87-93.
- 391 39. X.-S. Miao and C. S. Metcalfe, *Anal. Chem.*, 2003, **75**, 3731-3738.
- 392 40. T. Dittmar, B. Koch, N. Hertkorn and G. Kattner, *Limnol. Oceanogr.: Methods*, 2008, **6**,
393 230-235.
- 394 41. M. C. Kido Soule, K. Longnecker, S. J. Giovannoni and E. B. Kujawinski, *Org.*
395 *Geochem.*, 2010, **41**, 725-733.
- 396 42. E. B. Kujawinski and M. D. Behn, *Anal. Chem.*, 2006, **78** 4363-4373.
- 397 43. E. B. Kujawinski, K. Longnecker, N. V. Blough, R. Del Vecchio, L. Finlay, J. B. Kitner
398 and S. J. Giovannoni, *Geochim. Cosmochim. Acta*, 2009, **73**, 4384-4399.
- 399 44. W. J. Pabich, I. Valiela and H. F. Hemond, *Biogeochemistry*, 2001, **55**, 247-268.
- 400 45. T. C. Cambareri and E. M. Eichner, *Ground Water*, 1998, **36**, 626-634.
- 401 46. O. Portniaguine and D. K. Solomon, *Water Resour. Res.*, 1998, **34**, 637-645.
- 402 47. D. K. Solomon, R. J. Poreda, P. G. Cook and A. Hunt, *Ground Water*, 1995, **33**, 988-996.
- 403 48. S. Kim, R. W. Kramer and P. G. Hatcher, *Anal. Chem.*, 2003, **75**, 5336-5344.
- 404 49. X. Li and B. J. Brownawell, *Environ. Sci. Technol.*, 2010, **44**, 7561-7568.
- 405 50. T. Reemtsma, A. These, M. Linscheid, J. Leenheer and A. Spitzzy, *Environ. Sci. Technol.*,
406 2008, **42**, 1430-1437.
407
408
409

410

411

412 **Table 1.** Specific conductance, nutrient, dissolved organic carbon, and carbamazepine levels in the sampled groundwaters.

Site ^a	Depth below surface (m)	SC (μS/cm)	DOC (mg C/L)	[NO ₃ ⁻ + NO ₂ ⁻] (mg/L as N)	[NH ₄ ⁺] (mg/L as N)	TDN (mg/L as N)	DON (mg/L as N)	Carbamazepine (ng/L; average of n = 2)
Head of Waquoit	1.17	680	4.6	0.05	0.03	0.3	0.3	<LOQ ^c
Sage Lot	2 – 4 ^b	-	8.6	0.01	0.03	0.2	0.2	<LOQ
Green	0.86	531	1.7	3.68	0.02	3.7	0	12.1
Site D1	6.30 ^c	171	1.8	7.33	0.14	7.8	0.3	46.7
	7.53	181	1.1	1.67	0.57	2.4	0.11	
	8.75	203	1.2	2.34	0.69	3.2	0.17	
Site D2	9.97	236	0.4	12.09	0.51	13.7	1.1	7.8
	11.19	225	0.3	15.7	<0.02	14.9	- ^d	
	12.41	209	0.5	12.9	<0.02	12.3	- ^d	
	13.66	219	0.5	14.5	<0.02	14.9	0.4	
	14.88	255	0.4	19.4	<0.02	18.7	- ^d	
Site D3	16.11	321	0.5	24.39	0.02	26.3	1.9	30.6
	17.33	1081	0.6	23.1	<0.02	16.1	- ^d	

413 ^aFor unlabeled rows at Site D, only water quality data were collected by the USGS (provided courtesy of Denis LeBlanc and John Colman), and no quantification of
414 carbamazepine or analysis by ESI-FT-ICR-MS was performed.415 ^bExact depth between these depths. Water table at ~2 m below ground surface.416 ^cThe interval at 5.1 m was dry, indicating the water table is between 5.1 and 6.3 m in depth417 ^dSubtracting the inorganic components from total nitrogen results in a negative value.418 ^eOne replicate was above the LOQ of 1.6 ng/L and one below, with the average value below the LOQ.

419

420 **Table 2.** The total number of formulas and percentages (by number) of the types of formulas assigned to the groundwater samples,
 421 weighted averages of O:C and N:C molar ratios, and number of formulas with O:C ratio > 0.5.

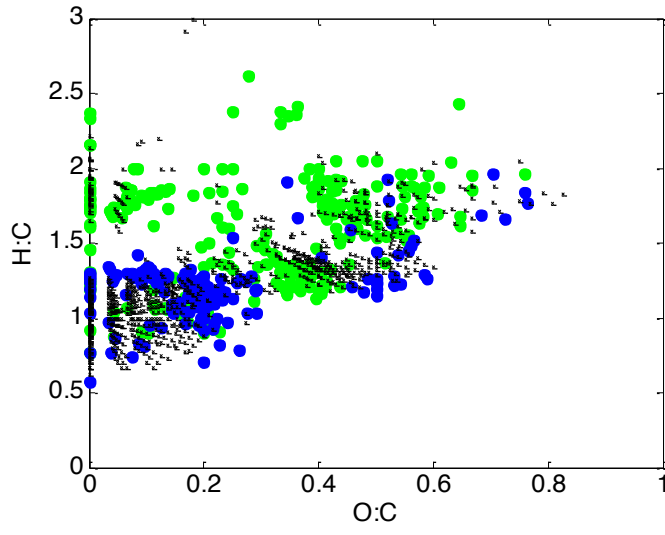
Site	Total Formulas	%CHO	% CHON	% containing N ^b	(O:C) _{wa}	(N:C) _{wa}	O:C > 0.5 all formulas	O:C > 0.5 N-containing formulas
Head of Waquoit	5,667	61	25	37	0.34	0.07	1193	278
Sage Lot	6,638	56	26	40	0.35	0.07	1337	522
Green	5,731	54	28	43	0.30	0.09	947	374
Site D1	5,803	56	29	41	0.30	0.07	661	375
Site D2 ^a	4,413	41	36	56	0.20	0.17	425	177
Site D3	5,368	53	30	44	0.27	0.09	505	262

422 ^aAverage of duplicate injections.

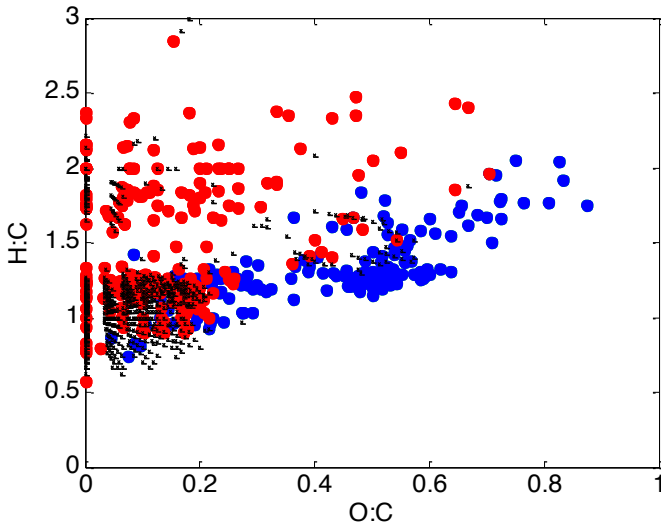
423 ^bIncludes all possible formulas found by the algorithm that contain N in any allowed combination with C,H,O,S, and/or P.

424

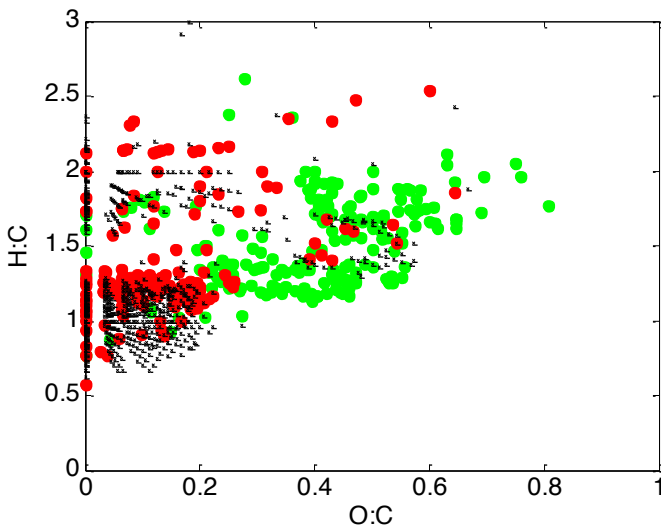
425



426

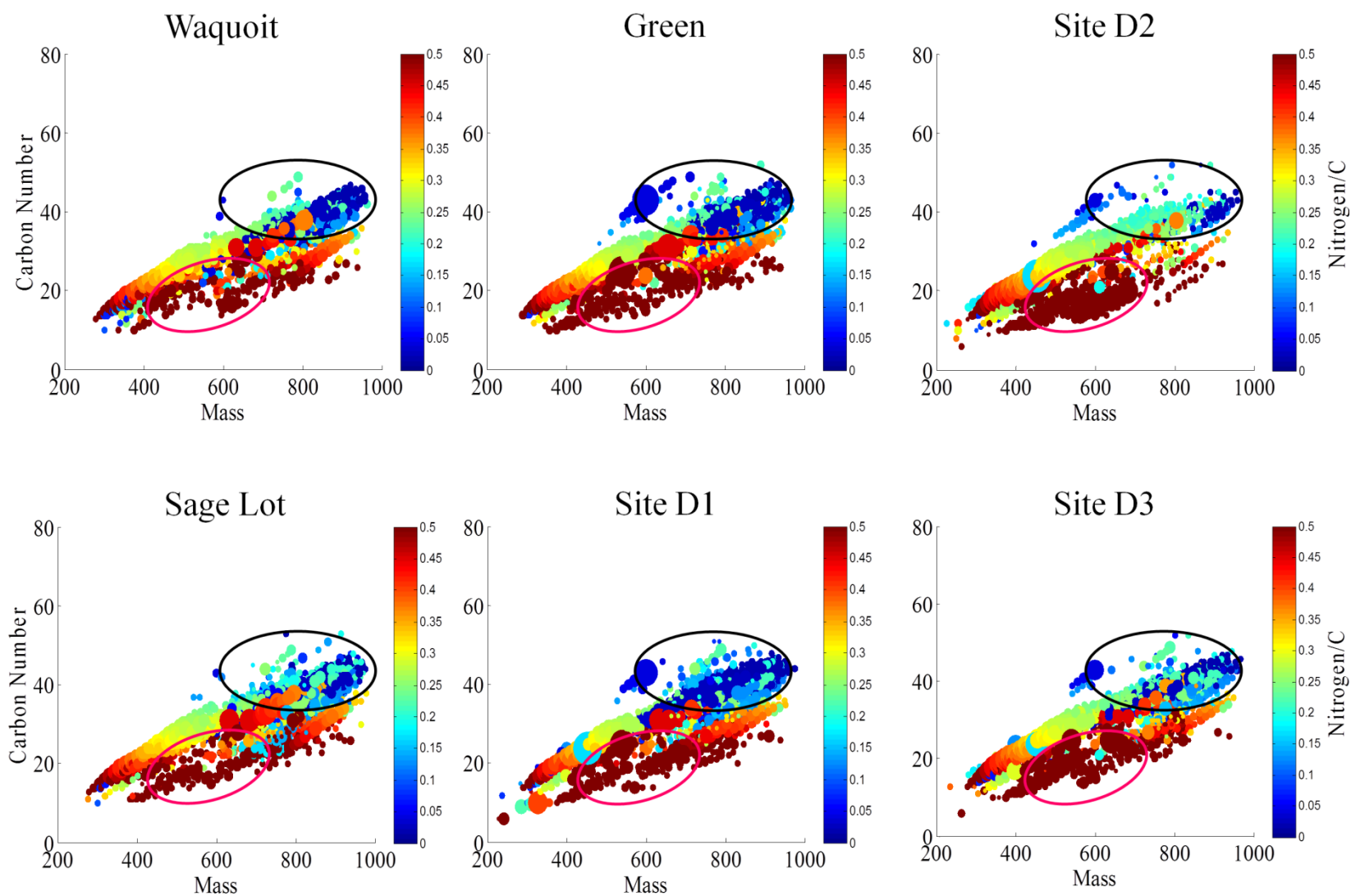


427



428

429 **Figure 1.** Comparisons of N-containing features in selected samples. (a) The green symbols are
430 features present in the shallow, impacted samples (Green and D1) but not in the unimpacted sites
431 (Head of Waquoit and Sage Lot). The blue are the features in Head of Waquoit and Sage Lot, but
432 not the other two samples. Black symbols are present in all four samples. (b) The blue symbols
433 are features present in the shallow, unimpacted samples (Head of Waquoit and Sage Lot) but not
434 in D2 and D3. The red are the features in D2 and D3, but not the other two samples. Black
435 symbols are present in all four samples. (c) The green symbols are features present in the
436 shallow, impacted samples (Green and D1) but not in D2 and D3. The red are the features in D2
437 and D3, but not the other two samples. Black symbols are present in all four samples
438



439

440 **Figure 2.** Carbon versus mass plots for the nitrogen containing features detected via ESI-FT-ICR-MS in the groundwater samples.
 441 Larger diameter denotes greater intensity and color indicates the N:C ratio. The pink oval marks the region of the plot that is

442 comprised of high nitrogen ratio features that are more intense in the septic -impacted samples. The black oval indicates low nitrogen
443 features that are more abundant/intense in shallow water samples.

Supplementary Information for

Molecular signature of organic nitrogen in septic-impacted groundwater

William A. Arnold^{†,‡,*}, Krista Longnecker[§], Kevin D. Kroeger^{||}, Elizabeth B. Kujawinski[§]

[†]Department of Civil, Environmental, and Geo- Engineering, University of Minnesota, 500 Pillsbury Dr. SE, Minneapolis, MN 55455, USA

[‡]Institute on the Environment, University of Minnesota, 1954 Buford Avenue, Saint Paul, MN 55108, USA

[§]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, 360 Woods Hole Rd., Woods Hole, MA 02543, USA

^{||}U.S. Geological Survey Woods Hole Coastal & Marine Science Center, 384 Woods Hole Road, Woods Hole, MA 02543 USA

*Author to whom correspondence should be addressed.

Phone: 612-625-8582; Fax: 612-626-7750; email: arno1032@umn.edu

Environmental Science: Processes & Impacts

Dendrogram from Cluster Analysis (Figure S1)

p. S2

Van Krevelen diagrams (Figures S2-S7)

p. S3

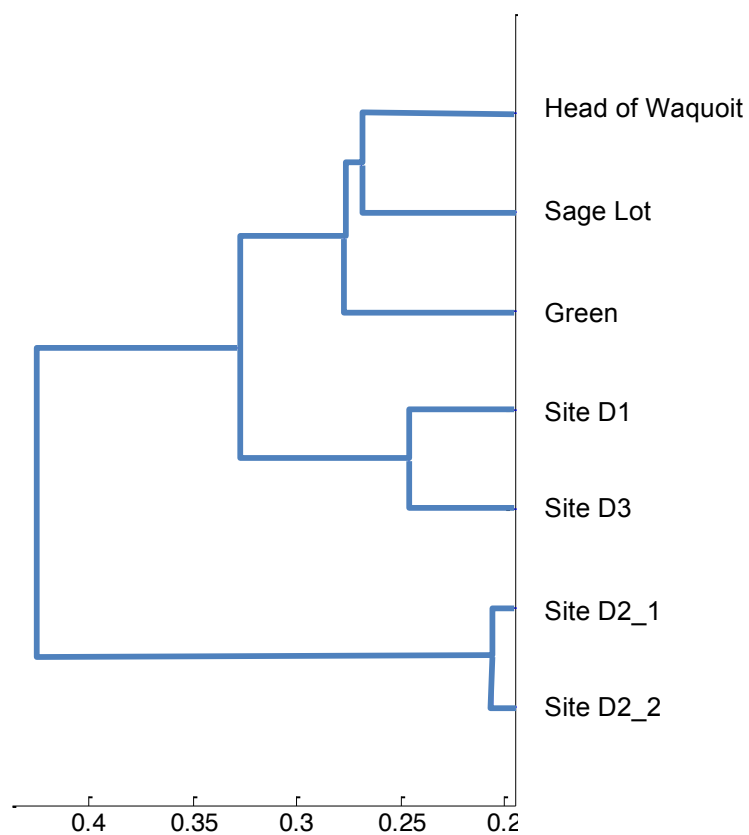
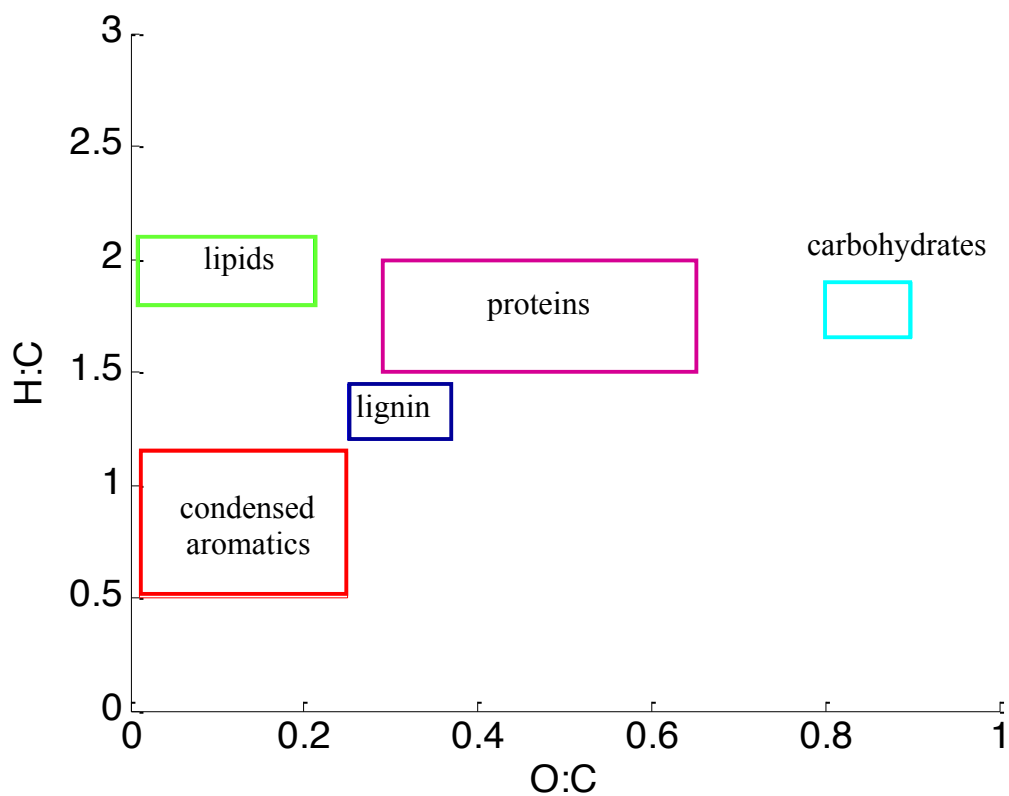


Figure S1. Dendrogram derived from cluster analysis determined from the FTICR-MS data from the six samples. The sample for Site D2 was analyzed twice, and the close association of the replicates indicates limited instrument variability.

Van Krevelen diagrams

In the following van Krevelen diagrams, the left hand plot contains all of the identified features for which formulas were assigned. The plot on the right hand side contains only N-containing features. The rectangles indicate regions associated with condensed hydrocarbons (red), lignin (dark blue), lipids (green), proteins (pink), and carbohydrates (light blue), as shown below.



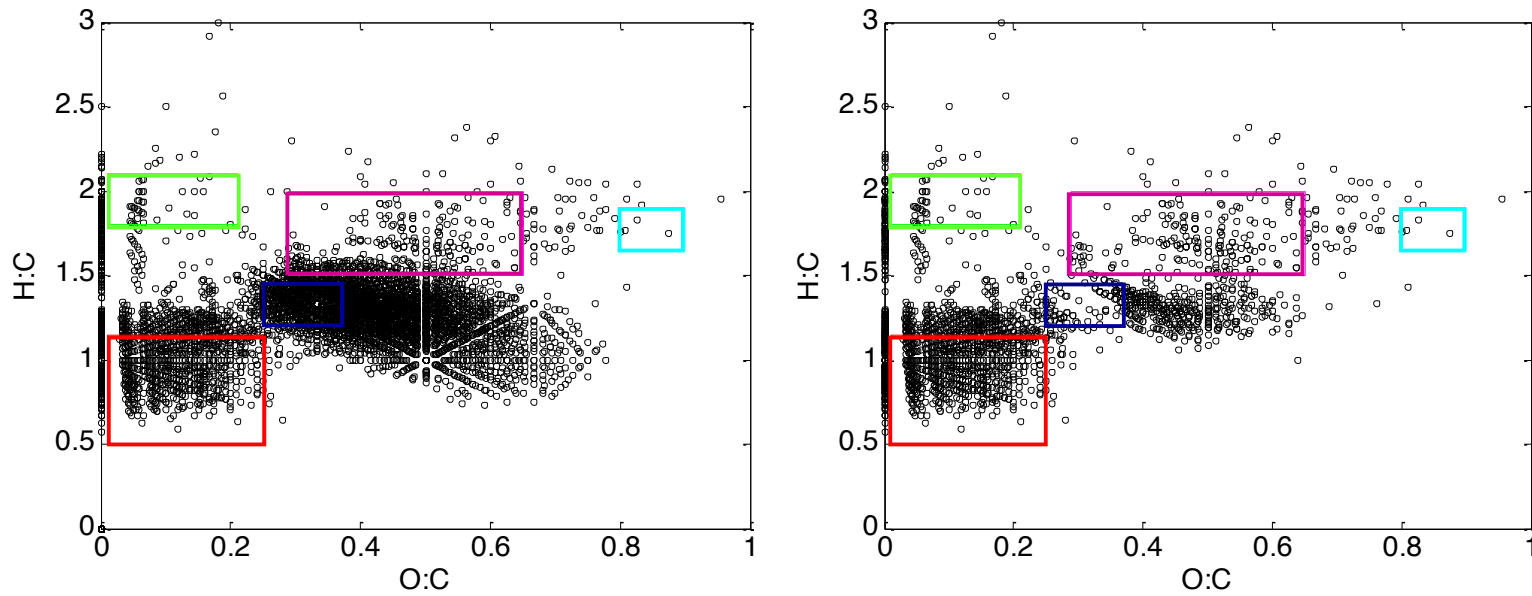


Figure S2. Van Krevelen diagrams for the Waquoit ground water samples.

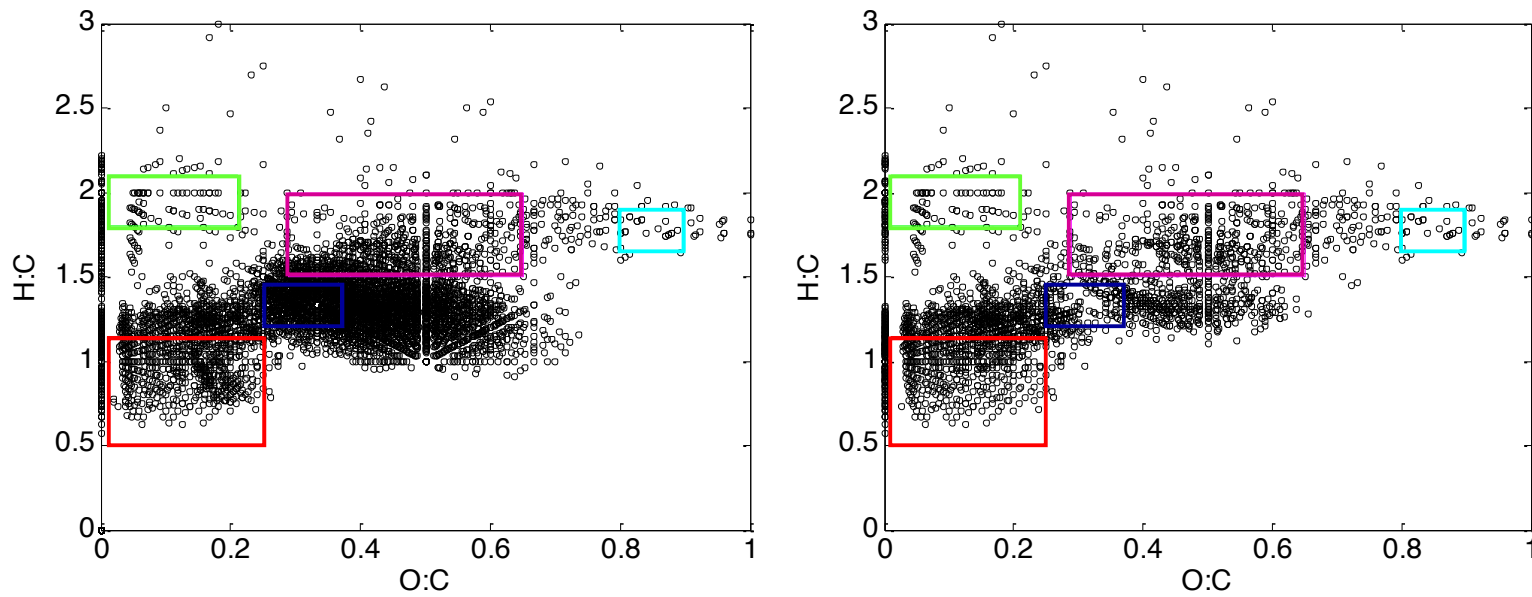


Figure S3. Van Krevelen diagrams for the Sage Lot ground water samples.

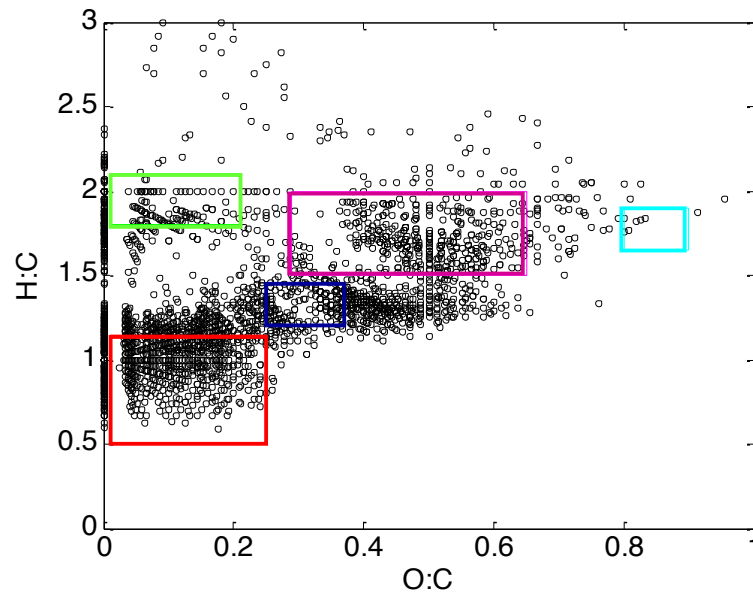
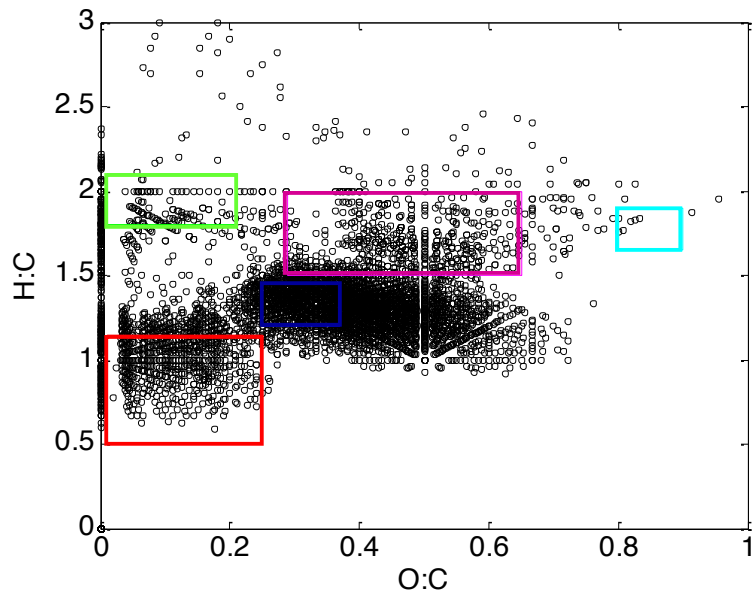


Figure S4. Van Krevelen diagrams for the Green Pond ground water samples.

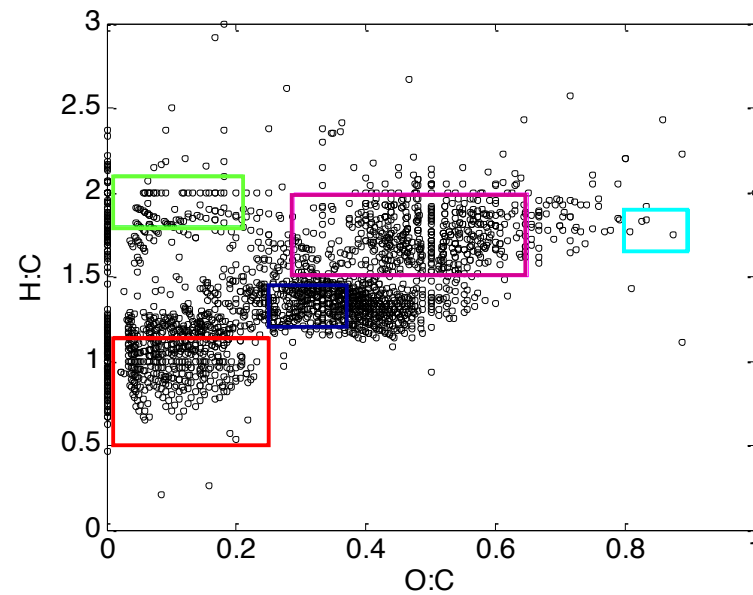
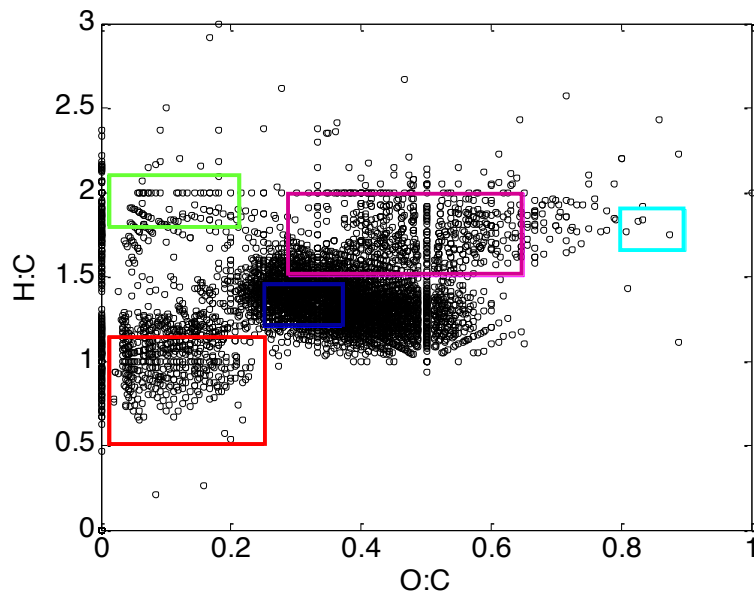


Figure S5. Van Krevelen diagrams for the Site D1 ground water samples.

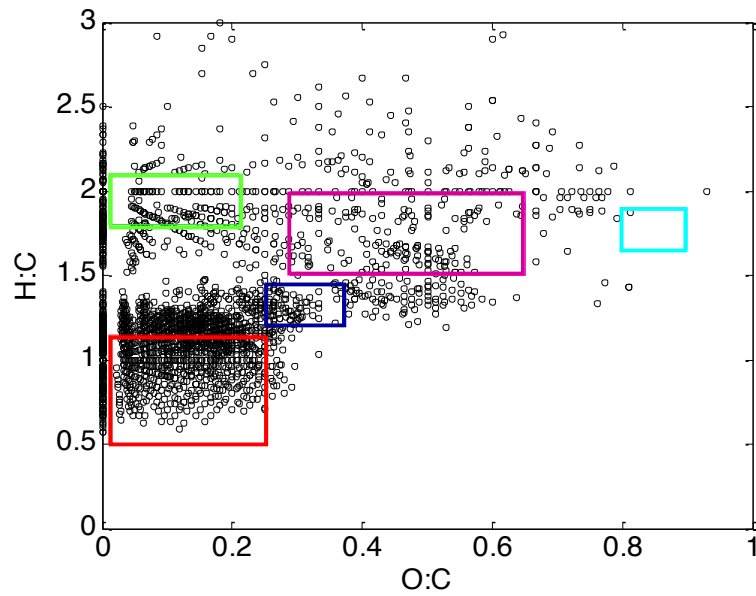
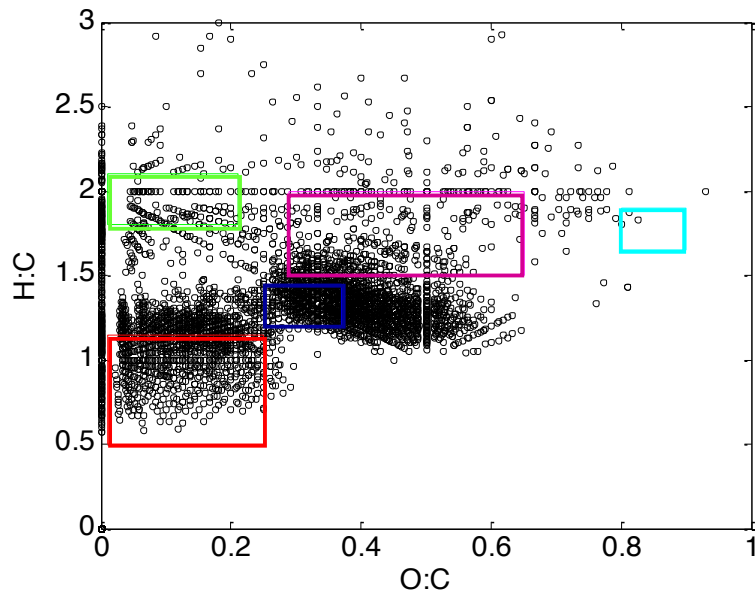


Figure S6. Van Krevelen diagrams for the Site D2 ground water samples.

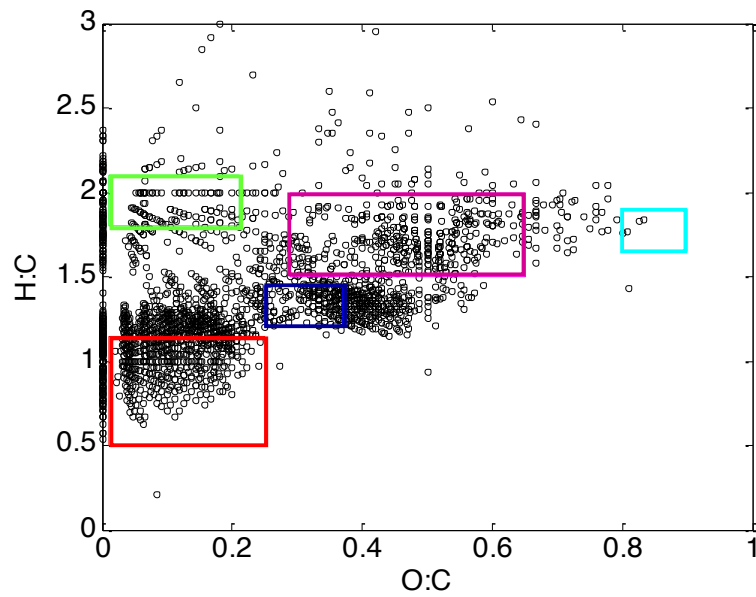
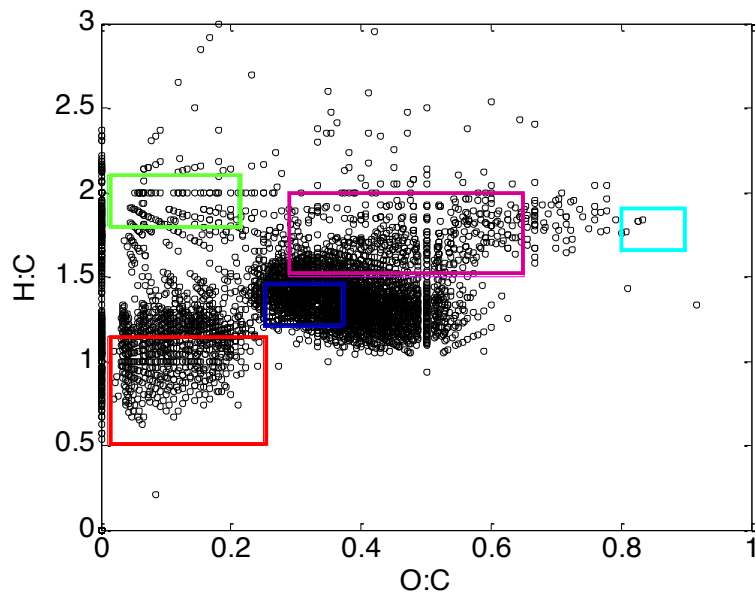


Figure S7. Van Krevelen diagrams for the Site D3 ground water samples.