UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

A highly resolved molecular-level interrogation of dissolved organic matter (DOM) complexity and composition in a snowmelt-affected system /

Permalink

https://escholarship.org/uc/item/9xt7558d

Author

Ball, Gregory Ian

Publication Date

2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSTIY OF CALIFORNIA, SAN DIEGO

A highly resolved molecular-level interrogation of dissolved organic matter (DOM) complexity and composition in a snowmelt-affected system

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Oceanography

by

Gregory Ian Ball

Committee in charge:

Professor Lihini Aluwihare, Chair
Professor Ellen Druffel
Professor Ralph Keeling
Professor Todd Martz
Professor Mark Thiemens

2013

©

Gregory Ian Ball, 2013

All rights reserved.

The Dissertation of Gregory Ian Ball is approved, and it is acceptable in quality and form for publication on microfilm.

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2013

DEDICATION

This dissertation is dedicated to my family and friends and to all of my past academic mentors who have facilitated, encouraged, and guided me in the pursuit of knowledge.

EPIGRAPH

Dum taxat, rerum magnarum parva potest res
exemplare dare et vestigia notitiai.

So far as it goes, a small thing may give analogy of great things,
and show the tracks of knowledge.

- Titus Lucretious Carus (ca. 99 BC-55 BC), *De Rerum Natura*
- Translation by Martin Ferguson Smith, 1975

The greatest reward for doing is the opportunity to do more.

– Jonas Salk

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Abbreviations viii

List of Symbols ix

Acknowledgements x

Vita xv

Abstract of the Dissertation xvii

Chapter I: Introduction 1

1.1 Dissolved organic carbon (DOC) 2
1.2 Significance of DOC 2
1.3 Characterization of DOC 3
1.4 Organization of the dissertation 6
1.5 References 11

Chapter II: Elucidation of CuO-oxidized dissolved organic matter (DOM) by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOF-MS): a molecular mosaic of thousands of structurally resolved compounds 13

2.1 Abstract 14
2.2 Introduction 15
2.3 Methods 22
2.4 Results 30
2.5 Discussion 40
2.6 Conclusions 46
2.7 References 72

Chapter III: Dissolved organic matter (DOM) composition and molecular diversity in a snowmelt-affected system as elucidated by comprehensive two-dimensional time-of-flight mass spectrometry (GC×GC–TOF–MS) and proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) 76

3.1 Abstract 77
3.2 Introduction 78
3.3 Methods 84
3.4 Results 101
3.5 Discussion 120
3.6 Conclusions 144
3.7 References 188

Chapter IV: Organic geochemistry of late Quaternary sediments from Fallen Leaf Lake, California Sierra Nevada, USA 196

4.1 Abstract 197
4.2 Introduction 198
4.3 Methods 200
4.4 Results 206
4.5 Discussion 212
4.6 Conclusions 217
4.7 References 240

Chapter V: A two-dimensional, heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analyses 243

5.1 Abstract 244
5.2 Introduction 245
5.3 Experimental 249
5.4 Results and discussion 258
5.5 Conclusions 270
5.6 Acknowledgements 271
5.7 References 280

Chapter VI: Conclusions 284

6.1 Original motivations of research 285
6.2 Summary of findings 287
6.3 Future directions 295
6.7 References 301

LIST OF ABBREVIATIONS

^{1}H-NMR	proton nuclear magnetic resonance spectroscopy
^{1}RT	first dimension retention time
^{2}RT	second dimension retention time
^{12}RT	first and second dimension retention time
BPCA	benzene polycarboxylic acids
CSRA	compound-specific radiocarbon analysis
DOC	dissolved organic carbon
DOM	dissolved organic matter
FA	fatty acids
FLL	Fallen Leaf Lake
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
GC	gas chromatography
GC×GC TOF-MS	comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry
LPVI	lignin phenol vegetation index
MS	mass spectrometry
PCGC	preparative capillary gas chromatography
PTV	programmed temperature vaporization
S/N	signal to noise ratio
SPE	solid phase extraction
TIC	total ion chromatogram
TOF	time-of-flight
UAL	Upper Angora Lake
UTR	Upper Truckee River

LIST OF SYMBOLS

$\delta^{13}C$	per mil (‰) deviation of $^{13}C/^{12}C$ from Pee Dee Belemnite
$\delta^{15}N$	per mil (‰) deviation of $^{15}N/^{14}N$ from air
$\Delta^{14}C$	per mil (‰) deviation of ^{14}C from standard (age-corrected to 1950)
C/N	ratio of carbon to nitrogen (by weight)
V	vanillyl lignin phenols
S	syringyl lignin phenols
C	cinnamyl lignin phenols
C/V	ratio of cinnamyl to vanillyl lignin phenols
S/V	ratio of syringyl to vanillyl lignin phenols
Vd/Vl	ratio of vanillic acid to vanillin
Sd/Sl	ratio of syringic acid to syringaldehyde
Ad/Al	acid to aldehyde ratio
Λ_8	carbon normalized sum of eight lignin phenols: mg (100 mg OC)$^{-1}$
Σ_8	volume-normalized (concentration) sum of eight lignin phenols
δ	ppm chemical shift

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Lihini Aluwihare for being an incredibly enthusiastic, helpful, vastly knowledgeable, resourceful, understanding, and extremely accessible academic adviser who encouraged my studies and guided me through six years as her student. Lihini was singularly responsible for introducing me to my dissertation subject of dissolved organic matter (DOM), a subject to which I had very little exposure prior to entering her lab. I quickly contracted Lihini's infectious enthusiasm for everything DOM-related and the more I learned the more interesting it all became. Lihini encouraged me to apply for several fellowships, most of which I was surprised to receive. Lihini also enabled my participation on two research cruises which allowed me to spend over five weeks out at sea, a prospect that had originally motivated me to pursue a graduate program in oceanography. Lihini taught me how to perform isolation of DOM by ultrafiltration and solid-phase extraction, taught me how to collect samples for [TOC] and [DOC] analyses, she taught me how to continuously filter thousands of liters of seawater over a period of several days aboard a moving vessel while avoiding sample contamination with extraneous carbon, and she taught me how to maximize the utility of zip ties, Ziploc bags, and rusted ratchet straps. Lihini enabled my research in the laboratory through her ceaseless and successful grant writing, which included the procurement of state-of-art analytical instrumentation of which I was often the sole user and, ultimately, care-taker. These resources enabled my dissertation research, allowed the pursuit of method-development projects, and provided a technical depth to my graduate education and research which would have otherwise not been possible.

I would also like to thank Devendra Lal. Lal's information-rich and philosophy-imbued lectures were an academic highlight during my time as a graduate student at Scripps. I feel privileged to have met Lal and to have taken his Advanced Isotope Geochemistry class. Ingrained in me are his emphatic exhortations to "think" and to live a life nourished by scientific pursuits, critical thinking, unceasing inquiry, and plenty of puzzles. The graduate fellowship he and his wife, Aruna, very generously endowed to Scripps Institution of Oceanography, supported me for more than a year of graduate school and this support has been deeply appreciated and I feel honored to have received it and been its inaugural recipient.

Three years of my graduate studies and research were enabled by the National Defense Science and Engineering (NDSEG) fellowship (32 CFR 168a), which was sponsored by the Office of Naval Research. I would like to thank Neal Driscoll for inviting me to be a teaching assistant for his SIO35 class on two occasions, for the opportunity to experience teaching, and for the exciting TA meetings that this entailed, whether they were fluffy or bulletproof. I would also like to thank the Scripps Graduate Office for funding me in my first and last years in the program and for providing travel funding to AGU 2012. Becky Burrola and Gayle Aruta are acknowledged for their help in procuring extramural funding and Joshua Reeves and Gilbert Bretado are acknowledged for being helpful resources for academic questions and for facilitating the paperwork involved with my advancement through the program. Laura Schaefer, Jan Hess, Azzure Beale, Monica Bailey, and Anne Cressey of the GRD business office are acknowledged for administering the resources that enabled much of the research in this dissertation.

I would like to thank Bruce Deck and the members of the Aluwihare Lab who were always helpful and available resources, including Susan Lang, Travis Meador, Roman de Jesus, Roberta Hansman, Stuart Goldberg, Nellie Shaul, Jenan Kharbush, Magali Marella, Neal Arakawa, and Brandon Stephens. I also spent several weeks at the National Ocean Sciences Accelerator Mass Spectrometry (NOSAMS) facility in 2010 and I would like to thank the extremely helpful and friendly staff there, including Ann McNichol, Nan Trowbridge, Mark Roberts, Mary Lardie, and Al Gagnon for the help they provided, and especially Li Xu for mentoring me in the radiocarbon prep lab.

I would like to thank the other members of my committee, Profs. Ellen Druffel, Ralph Keeling, Todd Martz, and Mark Thiemens, for agreeing to be on my committee and Ralph for teaching Biogeochemistry and Marine Chemistry. Also important were the cumulative influences and investments that were made towards my education by past academic and research advisers, including Walter Alvarez, Jillian Banfield, Eoin Brodie, Christopher Belnap, Terry Machen, Deirdre Meldrum, Jonathon Ulmer, and Carol Sibley.

Friends helped me strike a balance between life and work and were always reliable sources of support and levity. I would like to thank Katie Blair, Yi Zhang, Trang Nguyen, John Lindsey, my running buddies Tim Arnold, Geir Moholdt, Deanna Nelson, Rebecca Rolph, and Mallory Pickett, with whom I ran up countless mountains, almost weekly, on long, sweaty, and beautiful detours in search of the best burrito in San Diego County, Ryan Rowekamp for being a fantastic friend and room-mate for over five years and Judy Phyo for her support and friendship.

And, most importantly, I would like to thank my mom and my dad for providing me with an upbringing where academics were valued and where curiosities were stoked

and encouraged, my brother Neil, my sister Emma, my extended family, and Jonathan Stine.

Chapter II benefited from the Devendra and Aruna Lal Graduate Fellowship, the Department of Defense (DoD) National Defense Science and Engineering Graduate (NDSEG) Fellowship (32 CFR 168a), the Edna Bailey Sussman Award, and the Scripps Institution of Oceanography's graduate office. Additional funding was provided by ACS-PRF (45608-AC2) and NSF (MRI OCE-072322, OCE05–48275) to Lihini Aluwihare. Software trials and excellent customer support were provided by Allen Zhang, Heidi Neiman, and Todd Richards of LECO Corp. Lake Tahoe lab facilitates were kindly made available by the University of California, Davis Tahoe Environmental Research Center (TERC). Plant, pollen, soil, and sediment samples were generously collected by Paula Noble and plant identification was assisted by Anna Higgins.

Chapter III benefited from support provided by the Devendra and Aruna Lal Graduate Fellowship, the Department of Defense (DoD), National Defense Science and Engineering Graduate (NDSEG) Fellowship (32 CFR 168a), the Edna Bailey Sussman Award, the Scripps Institution of Oceanography's graduate office, and by ACS-PRF (45608-AC2) and NSF (MRI OCE-072322, OCE05–48275) to Lihini Aluwihare, and NSF EAR Postdoctoral Fellowship to Stuart Goldberg. Valuable customer support and software trials were provided by Todd Richards, Allen Zhang, Heidi Neiman of LECO Corp. Lake Tahoe lab facilitates were kindly made available by the University of California, Davis Tahoe Environmental Research Center (TERC).

Chapter IV benefited from the collective contributions of those associated with the BOLLY coring project of Fallen Leaf Lake in November 2010, which was

coordinated by Paula Noble of the University of Nevada, Reno (UNR). LacCore performed the coring, Laurel Stratton (UNR) helped orchestrate the program's logistics, Susan Zimmerman at Lawrence Livermore National Lab contributed AMS data, Simon Poulson (UNR) assisted with the collection of elemental and isotope data and provided mentoring and instruction on instrumental operation, and The Edna Bailey Sussman Foundation provided internship funding to G.I.B. to perform the organic geochemistry work at UNR. Chapter IV also benefited from plant and soil samples provided by Paula Noble (UNR) and species identifications that were assisted by Anna Higgins (UNR), as well as from conversations with Jillian Maloney at Scripps Institution of Oceanography (SIO), Neal Driscoll (SIO), Robert Karlin (UNR), and Briana Johnson (UNR). Brandon Stephens (SIO) helped perform the lignin analyses.

Chapter V is a reprint of, "Ball, G.I., Xu, Li., NcNichol, A.P., Aluwihare, L.I. A multi-dimensional heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analysis. Journal of Chromatography A., 1220 (2012) 122-131." The dissertation author was the primary investigator and author for this paper.

VITA

2006-2007	Research Assistant, Lawrence Berkeley National Lab
2007	Bachelor of Science in Chemical Biology, College of Chemistry, University of California, Berkeley
2007	Bachelor of Arts in Geology (High Honors), College of Letters and Sciences, University of California, Berkeley
2007-2013	Graduate Student Researcher, Scripps Institution of Oceanography
2008-2011	National Defense Science and Engineering Graduate (NDSEG) Fellowship, Office of Naval Research
2010-2013	Teaching Assistant, Scripps Institution of Oceanography
2010	Masters of Science in Earth Science, University of California, San Diego
2010	Guest Student at Woods Hole Oceanographic Institution's National Ocean Sciences Accelerator Mass Spectrometry Facility (NOSAMS)
2011-2012	Devendra and Aruna Lal Graduate Research Fellowship
2012	Professional Intern, Chevron Energy Technology Company
2013	Doctor of Philosophy in Oceanography, Scripps Institution of Oceanography, University of California, San Diego

PUBLICATIONS

Ball, G.I., Xu, Li., NcNichol, A.P., Aluwihare, L.I. A multi-dimensional heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analysis. Journal of Chromatography A., 1220 (2012) 122-131.

Samo, T.J., Pedler, B.E., **Ball, G.I**., Pasulka, A.L., Taylor, A.G., Aluwihare, L.I., Azam, F., Goericke, R., Landry, M.R. Patterns of microbial distribution and activity across a water mass frontal zone in the California Current Ecosystem. Journal of Plankton Research (in review)

Tracy H. Fung, **Gregory I. Ball**, Sarah C. McQuaide, Shih-hui Chao, Alejandro Colman-Lerner, Mark R. Holl, Deirdre R. Meldrum, (2004). Microprinting of on-chip cultures: patterning of yeast cell microarrays using concanavalin-A adhesion. ASME International Mechanical Engineering Congress, IMECE2004-60866.

FIELDS OF STUDY

Major Field: Oceanography

Studies in Organic Geochemistry
Professor Lihini Aluwihare

ABSTRACT OF THE DISSERTATION

A highly resolved molecular-level interrogation of dissolved organic matter (DOM) complexity and composition in a snowmelt-affected system

by

Gregory Ian Ball

Doctor of Philosophy in Oceanography

University of California, San Diego, 2013

Professor Lihini Aluwihare, Chair

To better constrain the chemical composition and molecular diversity of dissolved organic matter (DOM) in natural waters, comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOF-MS) was applied for the first time as a novel technique towards its characterization. This technique chromatographically resolved thousands of peaks per sample and yielded a level of chromatographic resolution that approached the effective mass resolving capabilities of Fourier transform ion cyclotron resonance mass spectrometers (FT-ICR-MS). As with peaks resolved by FT-ICR-MS, the chemical identities of peaks resolved by GC×GC-TOF-MS (> 90 %) remained largely unknown, however GC×GC-TOF-MS provided an

independent means by which peaks could be tied to known compounds, which included (1) electron-impact mass spectra accompanying each peak and (2) the separation of structurally related compounds along predictable trajectories in two-dimensional GC×GC space. The development and implementation of this methodology (Chapter II) identified 451 compounds across diverse CuO-oxidized, TMS-derivatized samples and provided a basis for identifying 313 compounds (6.5 % of peaks) within a solid phase extracted (SPE) DOM sample isolated from the Upper Truckee River, Lake Tahoe, California, USA. This represented the most comprehensive compound-specific characterization of a DOM sample to date and an approximate three-folding of compounds previously identified within this oxidation product mixture, a subject of three decades of extensive study. Novel compounds were discovered, isothermal retention in the second dimension was predicted using molecular mechanics modeling, and evidence for a previously unexplored isomeric diversity within DOM was presented. Thousands (> 90 %) of GC×GC-resolvable peaks, however, resisted facile structural elucidation and motivated the development and implementation of a methodology (Chapter III) by which the chemical distance between specific peak populations (defined by chromatographic localization and peak intensity) of two samples could be quantified. Performed pair-wise across a sample set of Lake Tahoe Basin SPE-DOM samples, distance matrices amenable to linkage analysis were constructed, hierarchical relationships of similarity between peak populations were exposed, and cluster analysis identified sample groupings that were subsequently corroborated by high resolution proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) of CuO-oxidized DOM. Interpretively revealing results were obtained, including the likely identification of peak populations specific to degraded

terrestrial OM. In Chapter IV, a 13,500-year organic geochemical record ($\delta^{13}C$, $\delta^{15}N$, TOC, TN, C:N) spanning the Holocene was created for Fallen Leaf Lake (FLL), California Sierra Nevada, USA. This was complemented by a 4,500-year lignin phenol record which spanned the Pleistocene-Holocene transition. These records were interpreted within the context of Holocene climate and seismicity and a post-glacial history of flora succession was proposed. In Chapter V a newly developed, high resolution, two-dimensional, heart-cutting preparative capillary gas chromatograph (PCGC) was developed, applied, and successfully evaluated as a tool for the enrichment of μg quantities of material destined for natural abundance radiocarbon (^{14}C) analysis by accelerator mass spectrometry (AMS).

Chapter I

Introduction

1.1 Dissolved organic carbon (DOC)

Dissolved organic carbon is operationally defined as the carbon in natural waters that is small enough to pass through a filter with a nominal pore size of about 0.45 μm. Different filter sizes are often used, ranging from 0.1 μm to as high as 1.0 μm. Included in the dissolved phase are viruses, colloids, and macromolecules composed of lipids, sugars, and amino acids. Excluded from the dissolved phase are most bacteria, zooplankton, phytoplankton, pollen, sand, silt, and clay. DOC is a reduced organic carbon mixture; most of its carbon is ultimately derived from the fixation of CO_2 by photosynthesis.

1.2 Significance of DOC

DOC is important because it is a large reservoir of reduced carbon that exists at the Earth's surface despite continuous exposure to molecular oxygen. The carbon contained within DOC is almost entirely derived from CO_2 fixation by photosynthesis, although much smaller contributions from chemoautotrophic (Ingalls et al. 2007) and hydrothermal sources are ultimately possible (Lang et al. 2006). Microorganisms also release and utilize DOC as a food source, fueling microbial food webs and enabling the transfer of carbon to higher trophic levels (Azam et al. 1983). The ocean, where the largest inventories of DOC are contained, is estimated to contain 670 Gt C (Hansell and Carlson 1998). By comparison, the carbon contained in Earth's atmosphere is ~760 Gt C while the amount of carbon contained in Earth's living terrestrial biosphere is 800 Gt C (Falkowski et al. 2000). The three carbon reservoirs are similarly scaled in size. If all of the DOC in the ocean were to undergo thermodynamically favorable oxidation to CO_2,

this carbon reservoir would disappear. Yet, this does not happen. In fact, a substantial fraction of DOC in the ocean appears to have accumulated on extended timescales based on its ^{14}C depletion, which reveals ^{14}C ages of several thousand years (Williams and Druffel 1987; Druffel et al. 1992). As such, 97 % of the DOC in the ocean is considered to be resistant to degradation, or 'refractory' (Hansell 2013), and not readily available as a food source to microorganisms. This refractory DOC persists, surviving successive cycles of ocean mixing in its chemically reduced state, effectively eluding re-oxidation to CO_2. While the mechanisms of refractory DOC formation remain unclear, some identified mechanisms include production by bacteria (Ogawa et al. 2001), contributions from black carbon (Dittmar and Koch 2006; Ziolkowski and Druffel 2010), and the formation of carboxylic rich alicyclic molecules (CRAM) which resist degradation (Hertkorn et al. 2006).

The chemical reasons underlying the persistence of this refractory carbon, which also exists in other environmental realms such as lakes, streams, and soil, has therefore been a topic of intense analytical inquiry. However, full deduction of its structure has been elusive. As such, increasingly powerful analytical techniques continue to be applied to its elucidation.

1.3 Characterization of DOC

DOC is most often measured simply in terms of its concentration. DOC concentrations are usually expressed in μM C (in the oceanography literature) or as mg L^{-1} in the estuary and terrestrial literature. DOC is measured either by the high temperature catalytic oxidation (HTCO) method (Sugimura and Suzuki 1988), in which

DOC is oxidized to CO_2 and quantified by non-dispersive infrared (NDIR) detection, or it is measured by wet chemical oxidation followed by CO_2 quantification. This commonly performed measurement, however, reveals nothing about the molecular partitioning of carbon within chemical structures. Other bulk-level measurements that capture all of the carbon present without providing insights into the actual compounds present include total organic carbon analysis, as determined by elemental analysis (EA), and $\delta^{13}C$, as determined by EA-coupled to isotope ratio mass spectrometry (IRMS). Other modes of DOM characterization that provide some measure of information about the types of compounds present include characterization of its cDOM (chromophoric DOM), excitation-emission matrix spectroscopy (EEMS), and sometimes ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectroscopy. While providing some general sense for the types of compounds present, these methods do not identify individual compounds. For individual compounds to be identified, higher resolution analytical methods are required. However, when such higher resolution and compound-specific analytical methods are applied to DOM, the results can often be disappointing. Hydrolyzable amino acids, amino sugars, neutral sugars, sugars, and lignin have all been previously identified in DOM, but each of these analyses must be performed separately and at considerable analytical effort and cost. Additionally, the quantifiable components, which are identified explicitly, only represent a small fraction of the total carbon present. Underscoring this point, in the deep ocean less than 3 % of DOC can be identified on a compound-specific basis while in the surface ocean this number is about 7 % (Benner et al. 1992).

Growing recognition of these analytical limitations, coincident with advances in analytical instrumentation, have led to the wide adoption of Fourier Transform Ion

Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) (Koch et al. 2005; Sleighter and Hatcher 2008; Flerus et al. 2012) to analyze DOM composition and diagenetic alternation on a molecular basis. This technique enables ultra-accurate exact mass determinations for compounds comprising DOM. These exact masses represent unique combinations of C, H, and O, and additional heteroatoms and, as such, afford molecular-level identifications of individual compounds in DOM. The primary limitation of this method, however, is that while compounds are extremely well resolved by mass, they are not resolved by structure. As such, compounds that have identical chemical formulae but different structures will be effectively resolved as identical compounds by FT-ICR-MS, even if they are not the same compound. As a principal focus of this dissertation, a high resolution analytical technique with a chromatographic resolution that approaches the *effective* mass resolution of FT-ICR-MS instrumentation (in terms of the number of resolvable peaks) is applied to the characterization of DOM for the very first time. This newly implemented technique is comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC–TOF-MS) (Fig. 1.1).

For analyses of DOM by FT-ICR-MS, NMR, or by compound-specific quantifications performed by GC or GC×GC-TOF-MS, DOM first needs to be extracted from the aqueous phase. This is performed by solid-phase extraction (SPE) of filtered, acidified water onto a hydrophobic resin. The water sample is filtered to remove particles and the filtrate is then acidified to protonate alcohols and acids. Acidification makes these compounds less polar and more hydrophobic and thus more likely to adsorb to the hydrophobic resin. The resin is then eluted by organic solvent following extraction (Fig. 1.2). The solvent is then removed and a powdered DOM sample is obtained.

1.4 Organization of the dissertation

Chapter II of this dissertation applies GC×GC–TOF-MS for the first time to characterize CuO-oxidized organic matter hydrolysate mixtures *using a compound-specific approach*. Suites of CuO-oxidized plant and soil samples were analyzed to create an electron-impact mass spectra (EI MS) and retention time library for 451 compounds identified from across all CuO-oxidized samples. Identification of these compounds also enabled the comprehensive mapping of the two-dimensional chemical landscape of the CuO hydrolysate as it was painted by this particular GC×GC method. This library of 451 compounds was then used to comprehensively interrogate the molecular composition of CuO-oxidized DOM isolated from the main tributary of Lake Tahoe, the Upper Truckee River (California, USA). More than 300 compounds were identified, several of which were novel structures, and the total number of identifiable compounds in this mixture was expanded almost 3-fold by several hundred compared to the number of compounds that have been previously observed in this mixture during the past 30 years of extensive study. This effort represented one of the most comprehensive compound-specific interrogations of a natural organic matter sample reported to date.

Chapter III then uses the knowledge base provided by Chapter II in order to investigate and quantify relationships of similarity between samples for populations of thousands of physicochemically defined peaks whose identities could not be ascertained. The sample set included 10 SPE-DOM samples isolated from the Upper Truckee River, Lake Tahoe, Fallen Leaf Lake, and Upper Angora Lake, and at multiple time-points. A statistical means of comparing peak populations by retention time alignment and mass spectra matching was evaluated and applied to draw relationships of chemical distance

between populations of peaks that were chemically fractionated as a unique feature of GC×GC. These relationships were then independently corroborated by proton nuclear magnetic resonance (^{1}H-NMR) data and by analysis of fatty acid fingerprint profiles. The relationships of chemical distance that were determined were then interpreted within the context of ancillary geochemical data, providing a unique set of geochemical snapshots that provided insights into DOM dynamics in this high altitude, sub-alpine snowmelt-affected system.

Chapter IV then leaves DOM and GC×GC behind entirely by diving into the sediments of Fallen Leaf Lake (FLL). Three piston cores with independent age control were analyzed from FLL for TOC, TN, C:N, δ^{13}C, and δ^{15}N to create a composite core (ca. -50 to ca. 13.5 kya) that stretched the entire length of the Holocene and into the last glacial period. The elemental and isotopic shifts observed in these cores were examined and interpreted within the context of Holocene climate in the Great Basin and the Sierra Nevada. Further, the core interval that penetrated the Pleistocene was analyzed for its lignin phenol content to examine possible correspondences of this geochemical proxy to isotope excursions that were observed in this interval. Correspondences between these complementary organic geochemical proxies were found and interpreted, and a history of terrestrial vegetation types surrounding FLL was constructed and was suggested to reflect successions of flora colonization following glacial retreat at the end of the Tioga glaciation.

Chapter V then develops a method (Ball et al. 2012) involving a three-oven, two-column, heart-cutting preparative capillary gas chromatograph (PCGC) that is uniquely tailored towards the preparative isolation of individual compounds from complex organic

matter mixtures. The method was demonstrated to be ideally suited for preparative enrichments destined for compound-specific stable isotope measurements and, especially, for natural abundance radiocarbon (^{14}C) analyses by accelerator mass spectrometry (AMS). This chapter appeared in *Journal of Chromatography A* as, "A two-dimensional, heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analyses."

Chapter VI then provides a summary of the dissertation's main conclusions and discusses possible directions for future research.

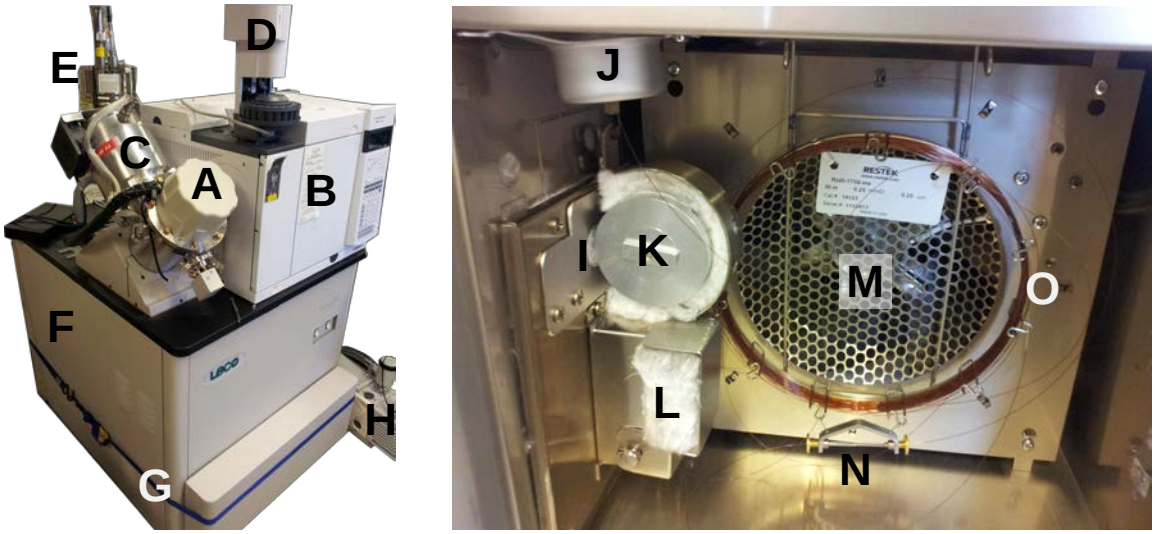


Figure 1.1: The main components of a GC×GC-TOF-MS setup include an ionization source (A), a gas chromatograph (B), source turbo pump (C), an injector (D), a LN_2 dewar to cool the cryogenic modulation gas (E), flight tube and reflector (F), detector (G), rough pump (H), transfer-line (I), injection inlet (J), secondary oven (K), 4-stage hot/cold pulse cryogenic modulator (L), 1˚ GC oven (M), a 1˚ and 2˚ column connector (N), a 1˚ GC column (O).



Figure 1.2: Photograph of solid-phase extracted dissolved organic matter (SPE-DOM) following elution from DAX-8 resin with 3:1 acetonitrile:H_2O. The sample was obtained from the Upper Truckee River, Lake Tahoe, California, USA in 2011.

1.5 References

Azam, F., T. Fenchel, J. Field, J. S. Gray, L. A. Meyer-Reil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. Marine Ecology Progress Series **10**: 257–263.

Ball, G. I., L. Xu, A. P. McNichol, and L. I. Aluwihare. 2012. A two-dimensional, heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analyses. Journal of Chromatography A **1220**: 122–31.

Benner, R., J. Pakulski, M. McCarthy, J. I. Hedges, and P. G. Hatcher. 1992. Bulk chemical characteristics of dissolved organic matter in the ocean. Science **255**: 1561–1564.

Dittmar, T., and B. P. Koch. 2006. Thermogenic organic matter dissolved in the abyssal ocean. Marine Chemistry **102**: 208–217.

Druffel, E. R. M., P. M. Williams, J. E. Bauer, and J. R. Ertel. 1992. Cycling of dissolved and particulate organic matter in the open ocean. Journal of Geophysical Research **97**: 15639–15659.

Falkowski, P., R. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hógberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffen. 2000. The global carbon cycle: a test of our knowledge of earth as a system. Science **290**: 291–296.

Flerus, R., O. J. Lechtenfeld, B. P. Koch, S. L. McCallister, P. Schmitt-Kopplin, R. Benner, K. Kaiser, and G. Kattner. 2012. A molecular perspective on the ageing of marine dissolved organic matter. Biogeosciences **9**: 1935–1955.

Hansell, D. A. 2013. Recalcitrant dissolved organic carbon fractions. Annual review of marine science **5**: 421–45.

Hansell, D. A., and C. A. Carlson. 1998. Deep ocean gradients in the concentration of dissolved organic carbon. Nature **395**: 263–266.

Hertkorn, N., R. Benner, M. Frommberger, P. Schmitt-kopplin, M. Witt, K. Kaiser, A. Kettrup, J. I. Hedges, and P. Schmittkopplin. 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochimica et Cosmochimica Acta **70**: 2990–3010.

Ingalls, A. E., S. R. Shah, R. L. Hansman, L. I. Aluwihare, G. M. Santos, E. R. M. Druffel, and A. Pearson. 2007. Quantifying archaeal community autotrophy in the

mesopelagic ocean using natural radiocarbon. Proceedings of the National Academy of Sciences **103**: 6442–6447.

Koch, B. P., M. Witt, R. Engbrodt, T. Dittmar, and G. Kattner. 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Geochimica et Cosmochimica Acta **69**: 3299–3308.

Lang, S. Q., D. A. Butterfield, M. D. Lilley, H. Paul Johnson, and J. I. Hedges. 2006. Dissolved organic carbon in ridge-axis and ridge-flank hydrothermal systems. Geochimica et Cosmochimica Acta **70**: 3830–3842.

Ogawa, H., Y. Amagai, I. Koike, K. Kaiser, and R. Benner. 2001. Production of refractory dissolved organic matter by bacteria. Science **292**: 917–20.

Sleighter, R. L., and P. G. Hatcher. 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Marine Chemistry **110**: 140–152.

Sugimura, Y., and Y. Suzuki. 1988. A high-temperature catalytic oxidation method for the determination of non-volatile dissolved organic carbon in seawater by direct injection of a liquid sample. Marine Chemistry **24**: 105–131.

Williams, P. M., and E. R. M. Druffel. 1987. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. Nature **330**: 246–248.

Ziolkowski, L. A., and E. R. M. Druffel. 2010. Aged black carbon identified in marine dissolved organic carbon. Geophysical Research Letters **37**.

Chapter II

Elucidation of CuO-oxidized dissolved organic matter (DOM) by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOF-MS): a molecular mosaic of thousands of structurally resolved compounds

2.1 Abstract

Two dimensional comprehensive gas chromatography time-of-flight mass spectrometry (GC×GC-TOF-MS) was applied as a novel technique to interrogate the molecular diversity of CuO-oxidized, trimethylsilyl (TMS) derivatized dissolved organic matter (DOM) isolated from Lake Tahoe's main tributary, the Upper Truckee River (UTR), California, USA. 4,789 chromatographic peaks were resolved at a signal-to-noise (S/N) of 10 or greater and over 300 (> 6 %) were assignable to discrete chemical structures. Comprehensive mapping of the two-dimensional elution time space revealed eight homologous compound series that were defined by successive methylene (–CH_2–) additions. Series of other compounds related to one another by methoxy (–OCH_3) and hydroxy (–OH) substitutions were systematically identified. Elucidation of the retention time (RT) displacements affected by these molecular transformations guided the discovery of novel compounds as well as those that had escaped previous detection within DOM, including a suite of lignin-derived cyclobutane photodimers. Compounds enriched in aldehyde, methoxy, and aromatic groups were preferentially retained in the second dimension (RXI-17) compared to compounds that were poor in these groups or otherwise enriched in aliphatics, alcohols, and carboxylic acids. Analysis of RT shifts among sets of benzene polycarboxylic acid (BPCA) isomers revealed the second chromatographic dimension to strongly retain sterically strained isomers and this observation provided a basis for identifying groups of isomers that were otherwise unidentifiable. The chromatographic resolution of GC×GC-TOF-MS, which approaches the mass resolution of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

(FT-ICR-MS) with respect to the number peaks it effectively resolves, also revealed a substantial and hitherto analytically inaccessible diversity of isomers, which appear to be uniquely resolved by this method, but which would be analytically unresolvable by FT-ICR-MS. Five sets of these isomers are described in detail. Previous investigations over the past 31 years have collectively identified approximately 120 compounds across all varieties of CuO-hydrolyzed organic matter. In this present effort, the identification of over 300 compounds in a *single sample* represents a substantial increase in the amount of chemical information that has been extracted from this commonly used oxidative analytical procedure. This analysis also very likely represents the most comprehensive compound-specific compositional elucidation of any natural organic matter (NOM) sample performed to date.

2.2 Introduction

2.2.1 Importance of dissolved organic carbon (DOC)

Dissolved organic carbon (DOC) is a globally important, actively cycling, and chemically complex pool of photosynthetically reduced carbon whose global inventory in the world's oceans, lakes, and streams (~ 670 Gt C) (Hansell and Carlson 1998) rivals that of the carbon content of Earth's atmosphere (~ 760 Gt C) and living terrestrial biosphere (~800 Gt C) (Falkowski et al. 2000). The ultimate origins of this reduced carbon are in the biosynthetic pathways of living organisms, such as plants and autotrophic algae, where it is synthesized into a diverse array of biomolecules, including cellulose, lignin, polysaccharides, lipids, and protein prior to release into the dissolved (<

0.7 μm) phase via dissolution (degradative and non-degradative) and exudation (active and passive). Collectively, these dissolved molecules serve as vital growth substrates for heterotrophic microorganisms, providing the chemical building blocks for cells and the chemical energy that drives the regeneration of ATP that empowers cellular growth and differentiation and that enables the transmission of carbon to higher trophic levels (Azam et al. 1983). Despite the chemical energy it contains, however, only a relatively small fraction of DOC is labile or semi-labile and readily accessible to microorganisms. The remaining non-labile DOC is often referred to as semi-refractory or refractory DOC (RDOC) (Carlson and Ducklow 1995) and can act as a carbon sink and persist in the dissolved phase for extended timescales, which can exceed a millennia in the ocean (Williams and Druffel 1987; Druffel et al. 1992). The compositional and structural reasons underlying its modes of persistence in the environment remain unclear and a comprehensive elucidation of its structure continues to prove elusive. As such, DOM continues to be the subject of intense study and increasingly powerful and high resolution analytical methods continue to be applied to its characterization.

2.2.2 Modes of DOM characterization

The analytical challenges posed by DOC's complexity have resulted in its most commonly measured (and cheapest) parameter to be its concentration. This measurement captures 100 % of the carbon present while providing no information on chemical structure. Other bulk-level techniques that similarly measure all of the carbon present, but which yield no insights into the compound-specific partitioning of this carbon into molecules, include $^{13}C/^{12}C$, which can be a useful discriminator of terrestrial (C_3 plant)

vs. aquatic or C_4-plant carbon, and nuclear magnetic resonance (NMR) spectroscopy, a powerful tool which is uniquely suited for probing the chemical environments of atomic nuclei. Excitation emission matrix spectroscopy (EEMS) has also been used as a relatively high throughput and facile tool by which varying abundances of specific fluorophore types can be used to infer changes in the amounts of humics, tyrosine, and tryptophan across DOM samples (Coble 2007), but again very little, if any, compound-specific information is obtained.

The composition of DOM can also be assessed using compound-specific quantitative methods. In these methods, individual compounds such as lipids, lignin phenols, amino acids, and neutral sugars are individually resolved and quantified following hydrolysis from DOM. Such compound-specific analyses, however, rarely account for a significant fraction of the total carbon present and come only at considerable analytical expense and effort. Emphasizing the limitations of these methods in the deep ocean < 3% of total DOC can be attributable to individual components while in the surface ocean this number is about 7 % (Benner 2002). This low level of characterizable carbon is generally thought to result from the biotic and abiotic diagenetic alternation of recognizable and 'fresh' organic matter (OM) into compounds that become less individually characterizable with time (Benner 2002).

In the past decade, the high-resolution analytical capabilities of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) have been applied to describe the complexity of natural DOC mixtures in unprecedented detail (Kujawinski et al. 2002; Kim et al. 2003; Koch et al. 2005; Dittmar and Koch 2006; Sleighter and Hatcher 2007), especially those compounds that are amenable to electrospray ionization

(ESI), which is the ionization mode most commonly employed in these studies. A particular advantage of DOM characterization by FT-ICR-MS is that no hydrolysis or derivatization is required before analysis and that DOM is analyzed in more or less its natural state if ionization-related molecular cleavages and molecular changes due to sample extraction, pre-treatment, and the pH adjustment (usually performed to assist ionization) are discounted as confounding variables. The high mass resolving power of these instruments, which typically range from 300,000-600,000 full width half max (FWHM) (Sleighter and Hatcher 2007), facilitate the detection of thousands of discrete molecular masses to a level of mass accuracy that enables the accurate assignment of individual elemental formula to thousands of detected mass peaks up to several hundred mass units in size. As only molecular formulae are resolved, however, this high resolution method provides no information as to the chemical structure of the identified molecule, or as to the number of isomers that give rise to any particular mass signal. It is often implicitly, and sometimes explicitly (Koch et al. 2005), assumed in such studies that molecules with identical masses have identical structures and identical origins, but the actual diversity of isomeric structures within DOM appears to be a subject that has foregone explicit examination.

2.2.3 Analysis of CuO-oxidized organic matter by gas chromatography (GC)

The high number of theoretical separation plates in GC has made it an ideal platform with which to resolve complex mixtures. GC's main limitation with respect to its ability to characterize natural organic matter (NOM) stems from NOM's size and polarity, which is too large and condensed to be immediately amenable to separation in the gas

phase. As such, NOM must be first oxidatively or reductively degraded into smaller and more volatile components. In oxidative degradations, however, polar and non-volatile alcohols and acids are produced that are not readily amenable to GC. Therefore, prior to analysis these compounds are converted to their corresponding trimethylsilyl ethers and esters to mask their polarity and make them more volatile.

The most common mode of oxidation relies on the anoxic alkaline hydrolysis of NOM in the presence of CuO at elevated temperature. This mode of oxidation was first described by Pearl in 1942 (Pearl 1942) and was later adapted by Hedges & Parker (Hedges and Parker 1976) to quantify lignin phenol abundances (as trimethylsilyl derivatives) in sediments and, subsequently, in oceanic DOM (Meyers-Schulte and Hedges 1986) for use as an unambiguous tracer of terrestrial organic matter. The exclusive objective of this oxidation technique is often the quantification of a small ensemble of 6–11 lignin phenols. However, the hydrolytic breakdown of NOM by this method, which cleaves oxidatively labile bonds including esters and most ethers (Chang and Allan 1971), yields many more resolvable and identifiable compounds that often escape closer examination. Additional compounds previously identified in this mixture include an expansive set of lignin dimers with diverse linkage structures (Goñi and Hedges 1992), straight chain, branched, α,ω-dioic acids, additional small aromatic acids derived from protein, phenols, polysaccharides, tannins (Goñi and Hedges 1995; Goñi et al. 1998), longer chain α,ω-dioic acids and polyhydroxylated fatty acids derived from plant suberin polyesters and cuticular waxes (Goñi and Hedges 1990), and benzene polycarboxylic acids (Dickens et al. 2007). To date, about 120 compounds have been identified across all types of CuO-hydrolyzed organic matter. This number, however,

represents a small fraction of the thousands of compounds that are ultimately present in the complex mixtures comprising disparate assemblages of NOM at the Earth's surface. Identification and quantification of these additional compounds has hitherto been stymied by their low concentrations and the limited resolution with which they can be chemically resolved by traditional chromatographic methods. With these analytical challenges in mind and motivated by the desire to access the building blocks of DOM's molecular mosaic and the information contained therein, a promising new tool is introduced that offers unique advantages that complement existing high-resolution methods of DOM characterization (FT-ICR-MS), while supplanting the capabilities of traditional 1-D chromatographic analyses in the characterization of CuO-oxidized DOM.

2.2.4 Comprehensive GC×GC-TOF-MS

Comprehensive GC×GC is a powerful technique first implemented by Liu and Phillips (Liu and Phillips 1991) to achieve separations of complex mixtures beyond the existing capabilities of 1-D GC separations. In GC×GC, two orthogonal columns of dissimilar stationary phases are coupled to one another using a cryo-focusing device that periodically samples the effluent of the primary column by freezing it into a narrow band at the head of a second column, which is significantly (~10×) shorter than the first. The narrow band of primary column effluent is then heated near-instantaneously, frozen, and again heated before chromatography begins anew on the second column. Meanwhile, the modulator repeatedly processes a successive interval of primary column eluent every several seconds (the modulation period) for the duration of the chromatographic run. The maximum attainable chromatographic resolution in such systems is the product of the

peak capacities of each respective capillary column and, as such, enormous increases in theoretical peak capacities are achieved. Additionally, cryofocusing leads to dramatically narrower and sharper peaks, which further increases peak capacity while augmenting instrumental sensitivity by increasing the S/N of detected peaks. Finally, a high speed time-of-flight (TOF) detector, which acquires full spectra from m/z 1–1000 at speeds of up to 500 Hz, enables robust spectral deconvolutions that provide a final dimension of separation for peaks that may still remain chromatographically unresolved. By this process, thousands of peaks are individually and discretely resolved, each of which has its own accompanying electron-impact (EI) mass spectra. Each peak is characterized by a primary column retention time (^{1}RT), which is on the order of minutes, and a secondary column retention time (^{2}RT), which is on the order of seconds. The EI mass spectra of each peak can then be interpreted or searched against external and in-house EI mass spectra libraries to enable the linking of detected peaks to discrete chemical structures. Additionally, in GC×GC homologous compound series are separated along predictable trajectories in 2-D retention time space. In combination with mass spectra, these 2-D trends in chromatographic separation, once illuminated, complement MS spectral information and further aid in the elucidation of chemical unknowns. Both GC×GC-TOF-MS and FT-ICR-MS commonly resolve as many as 10,000 peaks for environmental samples. As such, the level of chromatographic resolution of comprehensive GC×GC-TOF-MS setups effectively approach the level of mass resolution afforded by FT-ICR-MS setups. Both types of instrumentation provide similarly large quantities of data and reveal different and complementary features of organic matter complexity and composition. The augmented separation capabilities of comprehensive GC×GC systems

have thus made it an attractive platform with which to characterize complex petroleum mixtures and perform chemical class quantitations (Blomberg et al. 1997; Frysinger and Gaines 1999), elucidate unresolved complex mixtures (UCMs) of hydrocarbons extracted from Archean sediments (Ventura et al. 2008), and to profile molecular transformations associated with dissolution, evaporation, and biodegradation associated with hydrocarbon leakage (Wardlaw et al. 2008). To date, however, this technique has not been applied towards the characterization of more oxygenated NOM mixtures such as DOM. Here, this technique was applied as a novel tool towards the comprehensive chemical characterization of the CuO hydrolysate of a DOM sample isolated from the Upper Truckee River (UTR), South Lake Tahoe, CA, USA.

2.3 Methods

2.3.1 Cupric Oxide (CuO) Oxidation

Microwave-assisted CuO oxidation of OM samples was performed following Goñi & Montgomery's (Goñi and Montgomery 2000) adaptation of Hedges & Ertel (Hedges and Ertel 1982). Briefly, OM samples equivalent to approximately 5 mg OC were mixed with 500 mg CuO, 50 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ and 20 mL 2 *N* NaOH into perfluoroalkoxy (PFA) hydrolysis vessels. The vessels were subjected to 5 purge/evacuation cycles using N_2 (~30 psi) and a vacuum sustained by an oil-less vacuum pump. The mixture was then heated to 150 °C for 90 min following an initial 5 min temperature ramp. The PFA vessels were then cooled to room temperature, internal recovery standards (*trans*-cinnamic acid, ethyl vanillin) were spiked into the mixture and

the mixture was acidified to pH ≈ 2 using 12.1 *N* HCl. The acidic, aqueous solution was extracted (3×) with ethyl acetate that was then dried over Na_2SO_4 and gently blown to complete dryness under N_2 before immediate resuspension of the hydrolysate residue in 400 μL of pyridine. The hydrolysate residue, pyridine, and N,O-BSTFA + 1% TMCS were then combined 1:1:2 in 400 μL GC insert vials. These were heated at 60 °C for 15 min to swap exchangeable hydrogens in the hydrolysate with TMS groups, which increase the volatility of derivatized compounds. Samples were analyzed by both GC-MS and GC×GC-TOF-MS immediately following derivatization.

2.3.2 Comprehensive GC×GC-TOF-MS

The GC×GC-TOF-MS comprises an Agilent 7890A gas chromatograph outfitted with a 7893A Automated Liquid Sampler (ALS). The GC was interfaced to a LECO (LECO Corp's, St. Joseph, MI) Pegasus 4D time-of-flight (TOF) mass spectrometer. The primary (1°) column was 10.0 m, 0.18 mm i.d., 0.2 μm df with a RTX-5MS stationary phase. The secondary (2°) column was a 1.20 m, 0.10 mm i.d., 0.1 μm df and had an RXI-17 stationary phase. More polar compounds elute with greater ^{2}RTs in this method because an intermediate polarity (RXI-17) stationary phase was used in the second column. A Siltek-coated Vu2 Union® (Restek) joined the two columns inside the primary GC oven. The Agilent multimode inlet was kept at 300 ˚C and was operated in splitless mode with 100 mL min^{-1} purge flow at 60 s following injection. The inlet liner was a Restek gooseneck splitless deactivated glass liner containing deactivated glass wool. The instrument was operated in constant flow mode at 0.8 ml min^{-1} and He was used as the carrier gas. The system contains two separate ovens which were programmed

separately. Both primary (100–300 ˚C) and secondary (105–305 ˚C) ovens were programmed with 0.20 min initial hold times, and both were then heated at 6.0 ˚C min^{-1} until their final temperatures were reached at which point this temperature was held for 1 min. The GC runtime was only 34.5 minutes. These conditions were chosen after extensive method optimization that sought to maximize the spread of compounds across 2-D retention time space. As a result of these optimizations, the selected modulation period used was 6.0 seconds with a hot pulse time of 1.32 s and a cool time between stages of 1.68 s. The transfer line temperature was maintained at 280 ˚C and the modulator temperature offset was maintained such that it was always 20 ˚C above that of the primary oven. This offset is necessary to ensure complete and rapid volatilization of compounds after cryogenic LN_2 focusing during modulation. The injection volume was 1 µL. Larger injections were used when higher quality mass spectra of low abundance compounds were desired.

A solvent delay of 4.0 min was used, the mode of ionization was electron impact (EI) at 70 eV, the source temperature was 250 ˚C, and the acquisition rate and m/z range were 100 spectra $second^{-1}$ and 5-1000, respectively. The detector voltage was 1600 V. System optimization, source focusing, ion optic focusing, mass calibrations, tune, and leak checks were all performed immediately prior to sample analysis. These acquisition parameters yielded manageable file sizes of 357 MB per sample. File size scales proportionally with acquisition rate and run time. For instance, a doubling of both GC runtime and acquisition rate would increase file sizes roughly 4-fold, resulting in unwieldy files sizes exceeding 1 GB each. The injector, modulator, LN_2 dewar, injector,

TOF detector, and the GC and their two ovens were driven by ChromaTOF, proprietary software of LECO Corp.

2.3.3 GC×GC-TOF-MS Data Processing and Peak Detection

Data processing was performed within ChromaTOF. Here, raw data is processed and peaks are identified based on the occurrence of unique (U^{+}) apexing m/z values at discrete two-dimensional retention times (^{12}RT) that exceed a user-defined S/N threshold. Specifying a lower S/N threshold for peak detection invariably results in the detection of more peaks while specifying a higher S/N threshold results in the detection of fewer peaks, although the mass spectra of these peaks is generally of higher quality. If two peaks co-elute, ChromaTOF identifies unique apexing masses for each compound and deconvolutes the individual mass spectra for each compound from the TIC signal. This deconvolution is robust and is greatly assisted by (1) rapid (full) spectra acquisition rates of up to 500 Hz that enable well-determined deconvolutions and by (2) the absence of spectral skewing in TOF detectors (there is no scanning), which results in the Gaussian-shaped peaks ideally suited for standard deconvolutions. Full scan speeds for quadrupole-MS instruments amount to 1-3 full (50-1000) spectra second^{-1}, an acquisition rate that is more than two orders of magnitude lower than that which is possible with high speed TOF detectors.

The rapid acquisition speeds of TOF detectors are required to resolve the narrow peak widths (100-300 ms) that arise as a result of cryofocusing during GC×GC modulation. Generally, a chromatographic peak must be sampled 10 times across its width to be accurately quantified (Novak 1988) while for a deconvolution of two co-

eluting peaks to be robust, a peak must be sampled 18-20 times across its width. The efficacy of ChromaTOF's deconvolution algorithm and its performance compared to other chromatographic deconvolution algorithms has been evaluated elsewhere (Lu et al. 2008).

2.3.4 Oxidation of potential DOM source materials

A unique feature of GC×GC is that chemically related compounds are chromatographically separated by predictable shifts in two-dimensional retention times. Elucidation of these shifts is enabled by the examination of the elution patterns of known compounds related to one another by regular chemical permutations or transformations. Once illuminated, these patterns guide the identification of unknowns. Such a diverse array of compounds, however, is not readily available on a commercial basis at accessible costs. Therefore, CuO-oxidized hydrolysate mixtures derived from a variety of environmental sources were analyzed. Analysis and identification of compounds in these samples enabled the creation of an electron impact (EI) EI-MS TOF mass spectral library that assisted with the identification of lower-abundance compounds that were observed in other samples, and in the UTR DOM sample, but which had lower quality mass spectra that would have made their *de novo* identification in these matrices more tenuous.

The analyzed samples included an array of woody and nonwoody plant tissues of gymnosperm and angiosperm plants, including bark and cones from *Pinus jeffreyi* (Jeffrey pine), *Abies concolor* (white fir, needles, cone, bark), browned leaves from *Arctostaphylos patula* (greenleaf manzanita, leaf, twig), green leaves from *Quercus vacciniifolia* (huckleberry oak, leaf, twig), pollen skimmed from the surface of Fallen

Leaf Lake, California, USA, and a core-top sediment sample from this lake. Forest litter from nearby Lily Lake and Glen Alpine Creek were also collected and analyzed. The samples were collected June 9, 2012 to be broadly reflective of the different types of organic matter that surrounded UTR. All samples were dried and homogenized prior to CuO-oxidation following the aforementioned procedure.

A large volume (838 L) DOM sample was isolated by solid phase extraction (SPE) from the Upper Truckee River onto 1 kg of DAX-8 resin. Before use, the resin was thoroughly cleaned following the procedures outlined by Aiken *et al.* (Aiken et al. 1992). The cleaning procedure included Soxhlet extraction for ≥ 50 cycles of methanol followed by ≥ 50 cycles of acetonitrile. The resin was then poured into the extraction column and again cleaned with 20 L of 0.1 *N* NaOH. The base was then diluted with 1-2 L of milliQ water and the resin was again cleaned with 20 L of 0.1 *N* HCl and the poured column was stored in 0.1 *N* HCl prior to extraction.

The water sample was collected at the United States Geological Survey's (USGS) water monitoring station (#10336610; 38.922879 N, -119.991657 W) at South Lake Tahoe over four days from June 28 to July 1, 2011, a period that captured peak snowmelt. During this time river discharge averaged 867 ft^3 sec^{-1}, gauge height averaged 8.53', temperature averaged 10.5 °C, and TOC and DOC concentrations averaged 254 and 252 μM, respectively.

Water was collected into 20 L carboys and was transported to the Tahoe Environmental Research Center (Incline, Village), NV. Here, the water was deposited into a 55-gallon barrel lined with a Teflon bag (Welch Fluorocarbon). The water was then 0.2 μm-filtered into another Teflon-lined 55-gallon barrel before it was acidified to a pH

of 2 using 12.1 *N* HCl. The filtered, acidified water was then extracted onto the DAX-8 resin at a flow rate of 10 bed volumes or, equivalently, 10 L $hour^{-1}$. SPE carbon recoveries were monitored periodically and were calculated as $(1- [DOC]_{out}/DOC_{in}) \times 100$ % using a Shimadzu TOC-V CSN high temperature combustion analyzer and 7-point standard calibration curves ($R^2 \geq 0.99$). DOC recovery began at 53 % at the start of SPE extraction and ended at 38 % as the final liters of water were extracted, averaging 46 % over the course of the whole extraction. The observed decline in recoveries was consistent with increasing column saturation, which was visibly apparent. Following extraction, the resin was gravity eluted using 4 L of a 3:1 acetonitrile:water solution. The acetonitrile was removed by rotary evaporation and the remaining water was removed by lyophilization, yielding 1.9156 g of SPE-DOM with the following elemental and isotopic characteristics, expressed as the average (± 1σ) of triplicates: 55.5 ± 0.4 wt% C, 1.3 ± 0.7 wt% N, $\delta^{13}C$ = -26.3 ± 0.1 ‰, $\delta^{15}N$ = 2.8 ± 0.3 ‰. These values were measured following established procedures at the Scripps Institution of Oceanography's Analytical Facility. Integrated carbon recovery calculated on the alternative basis of the weight of recovered DOM (1.9156 g) and its wt% C was similar, but slightly lower at 42.0 %. This small discrepancy could arise due to sample losses during dry-down, lyophilization, and/or incomplete elution, or due to a slight over-estimation in the average recovery, perhaps caused by noncontinuous sampling (n=5) of the column effluent and influent streams that were the basis for this measurement.

2.3.5 Compound Identification

Mass spectra from peaks identified by GC×GC-TOF-MS of CuO hydrolysate mixtures were linked to known compounds by comparison of their spectra to (1) mass spectra obtained from commercially available standards, (2) to mass spectra contained within the NIST 2011 EI-MS main and replicate libraries, and (3) to those published elsewhere (Goñi and Hedges 1990, 1992; Hartley et al. 1990; Goñi 1992; Kallio et al. 2006; Dickens et al. 2007). Compounds were also identified by locating peaks that were separated by retention time shifts that were indicative of the continuation of homologous compound series, or by specific molecular substitutions. If a compound was identified in this manner, its presence had to be confirmed by MS interpretation and by the presence of incremented m/z fragments characteristic of the predicted molecular substitutions. Compounds identified on the basis of comparison to NIST library spectra were treated with *extreme skepticism*. If good agreement between library and experimental spectra existed, for instance, a plausible environmental origin of the identified compound was also required, and the presence of any particular compound had to be consistent with the manner in which the sample was prepared. As such, peaks that had high MS similarity scores to compounds in the NIST database, but which contained alkyl-ester linkages (which should be cleaved during oxidation), or that contained derivatization groups other than TMS, or that contained –OH or –COOH groups that should have been derivatized with TMS, were considered to be erroneous assignments. Furthermore, the peak's position in 2-D retention time (RT) space had to be consistent with the polarity of the functional groups of the matching candidate compound. For instance, if a particular compound contained a preponderance of polar functional groups (e.g. methoxy or

aldehyde groups) it needed to have an elevated second dimension retention time (^{2}RT) to be considered a positive identification. Lastly, if a published relative retention time (RRT) was available for a tentatively identified compound, this published RRT was translated into the present method and had to agree with the peak's observed retention time. Published mass spectra and RRTs beyond the NIST library were also important resources for identifying peaks corresponding to lignin dimers (Hartley et al. 1990; Goñi and Hedges 1992), cutin acids (Goñi and Hedges 1990; Kallio et al. 2006), and benzene polycarboxylic acids (BPCAs) (Dickens et al. 2007) which were observed in the CuO-oxidized OM.

2.4 Results

2.4.1 Comparison of comprehensive GC×GC-TOF-MS to 1-D GC

GC×GC of the TMS-derivatized CuO hydrolysate of SPE-DOM from the Upper Truckee River (UTR) yielded a detail-rich two-dimensional chromatogram (Fig. 2.1B) that contained 4,789 discretely resolvable peaks. Meanwhile, analysis of the same SPE-DOM sample by traditional 1-D GC, using a column and runtime that were 6× and 2× longer, respectively, yielded approximately 200 peaks, most of which were poorly or indistinctly resolved (Fig. 2.1A). At a *minimum*, therefore, the GC×GC method provided over an order of magnitude (20×) greater resolution than was attainable with a similarly optimized 1-D method. Moreover, the peaks obtained by GC×GC were narrower, better resolved, and each had a clean, m/z 1-1000 mass spectra obtained without mass scanning. This was in comparison to the 1-D GC whose chromatogram displays closely-eluting and

poorly-resolvable peaks that produce indistinct hump-like regions (especially visible after 35 min in Fig. 2.1A), akin to UCMs. The resolution of these UCM-like regions were significantly enhanced using GC×GC.

2.4.2 Linking GC×GC peaks to chemical structures

The oxidation of a diverse range of OM materials, including plant tissues, soil samples, and lake sediments enabled the production of high quality mass spectra for compounds that were abundant in these materials. Following the criteria for compound identification outlined in §2.3.5, 451 compounds were ultimately identified across all CuO-hydrolyzed OM samples. This, in turn, enabled the creation of a local EI-MS library that facilitated, in addition to retention time homology, the identification of these same compounds which were present at lower-concentrations in other OM matrices (such as the SPE-DOM of UTR) where the spectral qualities of these same compounds were too poor to definitively interpret. As peaks were linked to known structures, patterns in the retention times of chemically related compounds became apparent, and these patterns then assisted with further linking peaks to discrete chemical structures.

2.4.3 Observation of eight homologous compound series

Across all samples, a total of eight homologous compound series were observed to exist in the GC×GC space of CuO-hydrolyzed DOM samples. These compound series included (1) 1-alkanols, (2) alkanediols (generally α,ω), (3) 3-hydroxy fatty acids, (4) 2-hydroxy fatty acids, (5) normal fatty acids, (6) α,ω-dioic acids, (7) 2-hydroxy α,ω-dioic acids, and (8) (ω-1)-oxo fatty acids and these homologous compound series are shown in

Fig. 2.2. In these series, even-numbered carbon chains predominated over odd-numbered chains. The 1-alkanol series contained only four members, C_{11}, C_{12}, C_{16}, and C_{18}. The alkanediol series spanned carbon numbers from $C_{6,8\text{-}16}$. The 3-hydroxy fatty acids spanned $C_{6\text{-}16}$ and seven branched isomers were observed ($C_{11b,13b,14b,(15b\text{-}16b)\times2}$). The 2-hydroxy fatty acids were present from $C_{6,7,9\text{-}16}$ and four branched isomers ($C_{12b\text{-}15b}$) were observed while C_8 was absent from all samples. Normal fatty acids were present continuously from $C_{8\text{-}18}$ and for $C_{20,22,24}$ and five branched isomers were observed ($C_{11b,13b,(14b\text{-}15b)\times2}$). In all instances where branching of aliphatics was observed, ^{2}RTs were only affected slightly, but ^{1}RTs decreased relative to that observed for their straight chain equivalents (Fig. 2.2). The α,ω-dioic acids spanned C_3 (malonic acid) to C_{22}, continuously. The 2-hydroxy α,ω-dioic acids spanned $C_{4\text{-}22}$ continuously, and the (ω-1)-oxo fatty acids spanned $C_{6\text{-}13}$ continuously. Of the observed homologous compound series, the (ω-1)-oxo fatty acids had the greatest ^{2}RT, followed by the α,ω-dioic acids. The ^{2}RTs of the n-fatty acids and the 2-hydroxy α,ω-dioic acids were very similar, but the compounds in these two series were nevertheless well resolved from one another, yielding clean mass spectra, even for the longer chain members of these series whose elution times appear to be superimposed on one another in Fig. 2.2. The 2-hydroxy and 3-hydroxy fatty acids had similar ^{2}RTs, but the ^{2}RTs of the 2-hydroxy series was slightly greater. The alkanediols had the shortest ^{2}RTs of the observed compound series and the 1-alkanols had the second shortest ^{2}RTs. In these series, methylene ($–CH_2–$) insertions always increased ^{1}RTs while having little effect on ^{2}RTs, except for at the beginning and end of the chromatogram.

2.4.4 Characteristic retention time shifts caused by functional group permutations to core molecular structures

No proper homologous compound series were observed for aromatic compounds. However, among the compounds that were examined, several characteristic retention time shifts were observed and these were associated with defined chemical permutations to a core molecular structure that otherwise remained unchanged. These transformations included the replacement of aldehydic *hydrogens* with either $–CH_3$ or –OH groups and the replacement of 3- and 5-hydrogens of *p*-hydroxyphenyl aromatics with methoxy ($–OCH_3$) groups. The effect of these additions is shown in Fig. 2.2 and 2.3. Addition of $–OCH_3$ groups led to increases in both ^{1}RT and ^{2}RT and an M^+ increase of 30. Replacement of an aldehydic hydrogen with a CH_3 group ($M^+ + 14$) led to an increase in ^{1}RT while ^{2}RT remained largely unaffected. Replacement of an aldehydic hydrogen with an alcohol ($M^+ + 88$) led to an increase in ^{1}RT and a sharp *decrease* in ^{2}RT due to the polarity-shielding effect of the TMS group.

The effects that these functional group replacements had on two-dimensional retention times are clearly seen in Fig. 2.1 for the six lignin phenols, vanillin (Vl), acetovanillone (Vn), vanillic acid (Vd), syringaldehyde (Sl), acetosyringone (Sn), and syringic acid (Sd). Sl, Sn, and Sd are all related to Vl, Vn, and Vd, respectively, by the addition of a 3-methoxy group, as shown for these compounds in Fig. 2.2. Meanwhile, Vl is related to Vn as Sl is related to Sn by replacement of an aldehydic hydrogen with a methyl group which leads to an increased ^{1}RTs, but to only a very small change in ^{2}RT, as shown by the black lines in Fig. 2.2. Meanwhile, Vl and Vd as well as Sl and Sd are

each related to one another by replacement of an aldehydic proton with an –OH (–OTMS) group and this replacement leads to increased ^{1}RTs and decreased ^{2}RTs in both cases, as shown by the blue lines in Fig. 2.2. The effect of methoxy replacement of aromatic hydrogen is even more apparent for the α,1-monoketone and β,1-diketone lignin dimers (Fig. 2.3). In the case of the α,1-monoketone dimers (Fig. 2.3), for instance, zero to four levels of methoxy additions are observed, resulting in predictable increments of ^{1}RT and ^{2}RT that propel this compound series along a defined upward trajectory in two-dimensional retention time space. Similar relationships were observed for lignin dimers of other linkage types and these are shown in Fig. 2.3.

2.4.5 Identification of new compounds

2.4.5.1 Novel lignin dimers

Tying retention time shifts to specific functional group permutations enabled the discovery of eight new lignin dimers that appear to have hitherto gone undetected in the literature. These dimers are indicated (ǂ) in Fig. 2.3 and are listed in Table 2.1, except for α,5-(4-hydroxyacetophenone-vanillone), which was observed in the bark tissue of *Pinus jeffreyi*, but not in the UTR sample. These new dimers represent additional functional group permutations beyond those previously identified in CuO-oxidized DOM (Goñi and Hedges 1992). An additional set of nine lignin cyclobutane dimers (Fig. 2.3, 2.4) were newly identified in CuO oxidized DOM. These cyclobutane dimers appear to arise as a result of photo-dimerization of *p*-coumaric acid (pCd) and ferulic acid (Fd) (Fig. 2.4) (Ford and Hartley 1989; Hartley et al. 1990). These cyclobutane photodimers were

observed in CuO-oxidized pollen, nonwoody plant tissues, FLL sediments, and in UTR SPE-DOM. The following dimer types were observed: pCd-pCd, pCd-Fd, and Fd-Fd. Both asymmetrical "head-to-head" truxinic acid-like and symmetrical 'head-to-tail' truxillic acid-like dimers of the pCd-pCd dimer type were observed, although the former was only present for the pCd-pCd dimer type, consistent with previous observations (Hartley et al. 1990). These two dimer orientations were distinguishable by MS because of the symmetrical fragmentation of the "head-to-tail" dimers compared to the asymmetrical fragmentation of the "head-to-head" dimer (Fig. 2.4). While the individual stereoisomers of the "head-to-tail" dimers were indistinguishable by their mass spectra, they were nevertheless chromatographically resolvable, due to the large conformational differences, (accentuated by TMS derivatization) that arise from the differential orientation of the p-hydroxyphenyl and carboxylic acid substituents above or below the plane of the cyclobutane ring. In total, 17 new dimers were observed.

2.4.5.2 Structural elucidation of hydroxy benzene polycarboxylic acid isomers

Investigation of the retention time shifts within and between isomer sets of the benzene polycarboxylic acid (BPCA) compound series (Fig. 2.5) revealed additional patterns of separation that contrasted sharply with the simpler retention time trends that were observed for $–CH_2–$ insertion, or from replacement of aldehydic hydrogens with $–CH_3$ or –OH groups. Structural isomers of benzene dicarboxylic acids (2-BPCAs) and of benzene tricarboxylic acids (3-BPCAs) were better separated by ^{2}RT than by ^{1}RT (Fig.

2.5). The addition of –OH (–OTMS) groups to the 2-BPCA and 3-BPCA series led to consistent, but variable, increases in ^{1}RT and these increases were larger for the 2-BPCAs series, perhaps due to their smaller size relative to the added –OTMS substituent. However, the addition of –OH (–OTMS) to the different positional isomers of the 2- and 3-BPCA series led to inconsistent ^{2}RT shifts whereby –OH (–OTMS) substitution either increased or decreased the ^{2}RT of the resulting compound. Within each BPCA isomer set the isomer with the most crowded substitution pattern appeared to have the greatest ^{2}RTs, suggesting that steric strain influenced separation of these isomers in the second dimension. ^{2}RT times of individual isomers within each isomer set also scaled inversely with the molecular surface area computed for each isomer in its minimum (global) energy conformation, as described in Fig. 2.5. As such, examination of the separation of 2-, 3-BPCAs, and 2-BPCA-OH isomers in ^{2}RT-space, for which standards were analyzed, enabled the assignment of structures to isomers of the 3-BPCA-OH set, which has not been previously achieved (Dickens et al. 2007) due to the incomplete resolution of these isomers by regular GC methods and due the lack of currently available standards for 3-BPCA-OH compounds.

2. 4.6 Compounds identified in SPE-DOM from the Upper Truckee River (UTR)

The identification of 451 compounds across all CuO-oxidized OM samples provided a template for identifying unknowns within the CuO hydrolysate of the SPE-DOM extracted from UTR, a complex sample mixture that potentially integrates contributions from many of the OM sources analyzed. This procedure ultimately enabled the identification of 313 compounds in the CuO hydrolysate of UTR's SPE-DOM, which

represented 6.5 % of the 4,789 peaks that were discretely resolvable by GC×GC-TOF-MS at a S/N of 10 or higher. These 313 compounds are listed in Table 2.1 by their ^{1}RT and ^{2}RTs along with their respective CAS numbers and molecular formulae (of the derivatized TMS esters/ethers), most abundant unique ion (U+), and the area (or volume) derived from integration of the unique ion trace. These compounds are also shown as red dots in Fig. 2.6 overlain on top of the 4,789 identified peaks, which are shown as small black dots. The vast majority of compounds, however, still remain unidentified. Identified and unidentified peaks in Fig. 2.6 are superimposed on a higher-level classification scheme that was developed from analysis of the retention times of chemically related compounds identified in Fig. 2.2. Development of this classification scheme then enabled the comprehensive and visual assessment of the chemical composition, or molecular fingerprint, of DOM isolated from other aquatic settings, such as the Upper Truckee River which was examined in detail here. For example, inspection of Fig. 2.6 quickly revealed that the sample was lignin rich, containing a full complement of aldehydic and carboxy lignin monomers as well as a full suite of lignin dimers (except one, as previously noted). The full complement of BPCA and BPCA-OH compounds, thought to derive from tannin-like material during CuO oxidation and previously shown to be indicators of OM degradation (Otto and Simpson 2006), were also present. Cutin-derived longer chain (primarily $C_{\geq 16}$) hydroxylated aliphatics are largely absent, except for 9,16 and 10,16-dihydroxy hexadecanoic acids, which are the most abundant C_{16} cutin acids present in plant tissues containing cutin (Goñi and Hedges 1990; Kallio et al. 2006). Cutin-derived acids, including *trans*-9-octadecenoic acid, octadecanoic acid, and hexadecanoic acid were also present, although these are not unique to cutin. n-fatty acids

were observed continuously up to C_{18}, α,ω-dioic acids and 2-hydroxy α,ω-dioic acids were both observed continuously up to C_{13}, 2-hydroxy fatty acids were observed for $C_{6-7,9-12,14-16}$. 3-hydroxy fatty acids were observed for $C_{6-10,12,14,16}$. The only branched fatty acids observed were the 2- and 3-monohydroxy branched C_{15} fatty acids. (ω-1)-oxo-acids were observed for C_{7-10}, continuously. 1-alkanols were observed for $C_{11,12,18}$ and were of very low abundance and aliphatic diols were observed for $C_{6,8,8i}$ and the C_6 α,ω-diol was the most abundant of these.

Of the 313 compounds identified in UTR SPE-DOM, 70 % were mono-acids, 28 % were di-acids, 5% were tri-acids, and 1 % were tetra-acids. 63 % of the compounds contained at least one aromatic ring and 17 % of the compounds contained two aromatic rings. 10 % of the identified compounds were mono-aldehydes and 2 % were di-aldehydes. 68 % of the compounds contained at least one –OH group, 23 % contained two, and 1 % of the identified compounds contained three. 97 , 72, 26, and 7 % of compounds contained one, two, three, or four TMS groups, respectively. And 26, 13, 4, and 2 % of the identified compounds contained one, two, three, or four methoxy groups, respectively. The average number of TMS, aldehyde, methoxy, aromatic, acid, and hydroxy functional groups across the set of identified compounds was 2.0, 0.1, 0.5, 0.8, and 0.9, respectively. The highest molecular weight compounds identified were the stereoisomers of the Fd-Fd cyclobutane lignin dimers (Fig. 2.3, 2.4), which had molecular weights of 676 g mol^{-1}. Eight (2.5 %) of the identified compounds contained at least 1 nitrogen. 94 (30 %) contained at least one aromatic ring and one carboxylic acid. It is likely that the most frequently observed functional groups (i.e. acids) are products of the CuO oxidation of a disparate range of biopolymers, including protein, cellulose,

xylan, rubisco, lipids, and lignin, all of which produce small acids upon CuO oxidation, although only a small suite of these acids has been previously characterized (Goñi and Hedges 1995) . The CuO oxidation is not believed to lead to the hydroxylation of aromatic rings. The observed aldehydic groups are near-exclusively (29 of 31) lignin-derived and are formed concomitantly with lignin acetophenones and lignin acids as a result of the oxidation of lignin, principally aryl-O-ether linkages (Chang and Allan 1971). No evidence exists for significant formation of aromatics by CuO oxidation.

2.4.7 Additional structural isomer sets resolvable by GC×GC-TOF-MS

While over 300 compounds were identified within the CuO-oxidized SPE-DOM sample examined, the structures of the vast majority (93.5 %) of the remaining peaks resisted immediate structural elucidation. Manual inspection of mass spectra of subsets of these peaks, however, revealed many of them to be near-identical to one another, suggesting the resolution of isomeric structures at discrete retention times. In total, five additional isomer sets were readily apparent beyond those for the BPCA isomer sets discussed in §2.4.5.2 and shown in Fig. 2.5. These additional sets are labeled I–V in Fig. 2.7 and a blowup of set I is provided in Fig. 2.S1. Members of individual isomer sets were identified by their mass spectra (thin black arrows), all of which showed high similarity to one another based on reverse library searches. In isomer sets I, II, III, and V there was at least one isomer (thick black arrows) for which published relative retention time and mass spectra data were available from NIST databases. In no instances did the number of observed isomers exceed the number of possible positional ring isomers (PRI). Also identified by color-coded circles in Fig. 2.7 are groups of peaks with identical M^+

ions, resolved to 1 mass unit, which invariably represent different compounds of the same molecular weight. However, for compounds with M+'s of 222 (I and II), 296 (III), 250 (IV), and 324 (V), several of the encircled peaks with identical M^+ did, in fact, represent the same compound (as indicated) based on high mass spectra similarity.

2.5 Discussion

2.5.1 Resolution of CuO-hydrolyzed DOM by GC×GC-TOF-MS

It was apparent from comparison of 1-D GC and GC×GC chromatograms that GC×GC revealed a greater level of the CuO hydrolysate's molecular diversity than did 1-D GC. This was also shown quantitatively. At *least* 20× more peaks were identified by GC×GC than by 1-D GC.

Past studies characterizing CuO-oxidized organic matter have collectively identified about 120 compounds across these oxidation product mixtures (Hedges and Ertel 1982; Goñi 1992; Goñi and Hedges 1992, 1995; Goñi et al. 1998; Dickens et al. 2007). The present effort, which led to the identification of 451 compounds across the aforementioned set of CuO-oxidized OM samples, represents over a 3.7-fold increase in the number of characterized compounds in CuO-hydrolysate mixtures. This high level of compound identification was uniquely enabled by the enhanced chromatographic and ultra-fast, full spectra mass detection capabilities of GC×GC-TOF-MS which revealed 4,789 peaks in a single CuO-oxidized SPE-DOM sample. The average FWHM of all of these peaks was 113 milliseconds. On this basis, a maximum of 53 peaks could

theoretically be resolved in a 6 s modulation window without the aid of deconvolution. With deconvolution, this number is higher and dependent on the acquisition rate.

High quality mass spectra for each peak, combined with the separation of chemically related compounds by predictable 2-D retention time displacements, assisted with linking unknown peaks to discrete chemical structures. Following this methodology, previously unreported compounds were identified within CuO hydrolysate mixtures, examples of which were provided for the lignin dimers in Figs. 3-4 where 17 lignin dimer types new to CuO-hydrolyzed DOM were shown, eight of which appear to be novel structures while the other nine photodimers have been previously observed in directed studies of plant tissues (Ford and Hartley 1989; Hartley et al. 1990). The method provided particularly valuable at capturing lignin diversity, essentially binning all of the lignin dimers into the upper right quadrant of the chromatogram where they were exceptionally well resolved. Other newly observed compound series included the 2- and 3-hydroxy fatty acids, an extended range of α,ω-dioic acids, and the 2-hydroxy α,ω-dioic acid series.

The task of assigning structures to thousands of peaks is primarily a product of effort expended (Fig. 2.S2). This process starts with making educated guesses concerning the types of compounds that could conceivably be present within different CuO-hydrolyzed mixtures and then searching the NIST library and the published literature to see what information, if any, is available for these compounds in the way of RRT and mass spectra data. If published spectra are available, then reverse library searches can assist with identifying predicted compounds from the peak data provided that the targeted compounds are indeed present. These assignments can then be corroborated by RRT

alignments if RRT data are available. If an envisioned compound, however, does not have mass spectra coverage in the NIST database, fragmentation patterns must be predicted manually. These predicted fragment ions can then be used to search the peak data for these compounds.

As an example, such a procedure was used unsuccessfully to see if degradation products of the lignin cyclobutane photodimers could be identified within CuO hydrolyzed SPE-DOM. Possible degradation pathways (e.g. demethylation, ring opening, etc) were used to predict degraded forms for these structures which, if present, would represent terrestrial organic carbon in hitherto unrecognized structures. Fragment ions characteristic of these compounds were then predicted and used to search the peak data. This procedure was also used unsuccessfully to attempt the identification of photo-oxidation products of α-pinene, a volatile monoterpene released by conifers, which are of great abundance surrounding Lake Tahoe and UTR. Pursuits such as these, however, rarely yielded success and were very time intensive, and underscored the importance of having standard reference spectra available, whether they were authentic pure compounds, NIST library matches, or "natural" standards enriched in specific compounds that would otherwise be too challenging to identify *de novo* at lower abundances in more complex mixtures such as DOM.

2.5.2 Compounds present in UTR DOM

A large number of compounds were observed in the CuO hydrolysate of UTR's DOM. The vast majority of these compounds are produced as a result of the oxidation of a diverse range of biopolymers while other compounds survive through the oxidation

with minimal alteration. For instance, the oxidation of the amino acid phenylalanine produces a suite of small acids, including benzoic acid, benzeneacetic acid, butanedioic acid, benzene glyoxalic acid, and α-hydroxy benzeneacetic acid in yields of < 1 % each (Goñi and Hedges 1995). All of these oxidation products are observed in UTR DOM, but they have additional potential sources as well. Different types of biopolymers also produce similar sets of small 4-hydroxy aromatic acids and α,ω-diacids (up to C_5), but lipid-like compounds like tristearin (that yields only octadecanoic acid) yield oxidation products more reflective of their source due to their limited number of oxidatively labile bonds (Goñi and Hedges 1995). It has also been observed (Goñi and Hedges 1990) that some primary alcohols can be oxidized to acids in yields of up to (40 %). This could explain, along with the absence of abundant sources for these compounds, the discontinuousness of the homologous series of aliphatic alcohols observed in this study. The observation of a small number of unsaturated aliphatics among the identified oxidation products indicates that double bonds must survive the oxidation intact. Consistent with alkaline oxidation with CuO, no esters or ethers (beyond methoxy) were observed in the compounds identified. Hydroquinone and resorcinol were both observed, but catechol was not, and the yield of 3,5 dihydroxybenzoic acid was >> than that of 3,4-dihydroxybenzoic acid, and gallic acid was not observed, suggesting that vicinal aromatic diols are labile to the CuO oxidation.

The most abundant compounds identified by GC×GC-TOF-MS of the UTR DOM sample tended to be those compounds that have been previously observed in CuO oxidation mixtures. Thirty-seven (12 %) compounds had abundances > 100 ($\times 10^4$) and these included many of the BPCAs, lignin monomers, including the carboxy and formyl

lignin monomers, dodecanoic acid (C_{12}-FA) and 5,5'-dehydrovanillinvanillic acid. Some compounds at this concentration level that appear to have escaped previous detection in CuO oxidation studies included glutaconic acid, vanillylpropionic acid, an isomer of 2-hydroxy, 5-ethanone benzoic acid, and a newly identified di-carboxylic, mono-formyl guiacol compound that is probably lignin derived. Ninety (29 %) compounds had abundances between 10–100 ($\times 10^4$), 111 (35 %) compounds had abundances between 1–10 ($\times 10^4$), and 75 (24 %) had abundances below 10^4. The expanded diversity of the CuO hydrolysate that was revealed by GC×GC-TOF-MS was clearly driven by the enhanced resolution and detection of lower abundance compounds.

The majority of the compounds (~ 62 %) identified within the CuO-oxidized terrestrial SPE-DOM samples in this study have not been observed in previous CuO oxidation studies. Some of these compounds have, however, been observed in previous pyrolysis mass spectrometry (Py-MS) studies of dissolved humic and fulvic acids (Schulten 1999). Some of the compounds that fell within the analytical window of both CuO-oxidation-GC×GC-TOF-MS and Py-MS methods included a variety of substituted phenols as shown in Fig. 2.8.

2.5.3 Complementarity of GC×GC-TOF-MS data to FT-ICR-MS data

Recent advances in DOM characterization have principally been made by FT-ICR-MS whose high mass resolutions resolve thousands of mass peaks per sample. Here, GC×GC-TOF-MS similarly resolved thousands of peaks per sample, but on the basis of chromatography as opposed to mass. Both high resolution techniques resolve similar numbers of peaks. In FT-ICR-MS, structural isomers cannot be resolved and, instead, all

isomers contribute to one discrete mass intensity. Meanwhile, in GC×GC-TOF-MS compounds with indistinguishable retention times (or apexing masses) cannot be resolved, but structural isomers of the same exact mass *can* be resolved. This was described for five specific isomer sets in §2.4.7 and was shown in Fig. 2.6. It was also suggested that additional, higher molecular weight isomer sets likely exist beyond the five that were specifically identified. While the specific isomeric configurations of these five compounds, which have not been previously identified in CuO-oxidized DOM, were not determined, a methodology for assigning structures to individually resolved isomers was successfully tested for sets of BPCA isomers examined (Fig. 2.5). It was found that the systematic retention time shifts caused by functional group additions and removals (i.e. ± –CH2–,± $–OCH_3$, ± –OH) broke down in highly strained molecules where higher levels of strain led to lower molecular surface areas which, in turn, corresponded with longer ^{2}RTs. This evidenced a unique steric-selectivity for this setup's second dimension separation. One explanation for the longer ^{2}RTs of sterically hindered molecules could derive from the reduced ability of the TMS groups to mask the polarity of oxygen and aromatic groups in these strained molecules, leading to greater interaction with the diphenyl stationary phase and further retention in secondary column. Further investigation of the isomeric diversity of DOM by GC×GC-TOF-MS could begin to inform the prevalence of different substitution patterns within and between DOM samples as well as provide constraints on the actual levels of isomer diversity that exist with DOM mixtures. If the entropy-driven evolution of molecular diversity in NOM does in fact play a crucial role in the formation of refractory DOM as has been suggested by

Hertkorn *et al.* (Hertkorn et al. 2007), then equal abundances of isomeric structures might well be encountered in RDOM.

2.6 Conclusions

DOC is ultimately entwined within discrete molecules of DOM whose structures can reveal the source, reactivity, and fate of this dissolved carbon in the environment. In this study, the groundwork was laid for using GC×GC-TOF-MS as a tool for the comprehensive characterization of CuO hydrolysate mixtures. The high chromatographic resolution of this tool, which approaches the mass resolution afforded by FT-ICR-MS instruments in terms of the number of peaks that are effectively resolved, enabled the resolution of thousands of peaks per sample and 6.5 % of these peaks were ultimately tied to individual compounds in the SPE-DOM sample that was analyzed. The separation of these peaks by regular and predictable shifts in retention times caused by methylene ($–CH_2–$), methoxy ($–OCH_3$), alcohol (–OTMS), and carboxy ($–CO_2TMS$) replacements guided the identification of novel unknowns and of compounds that had not been previously observed within CuO-oxidized OM, such as the cyclobutane lignin dimers. The ^{12}RT shifts resulting from these molecular transformations were, however, observed to break down in the case of the BPCA compound series, however closer examination revealed this deviation to be analytically advantageous as compounds that were identical by formulae and by functional group composition were nevertheless well-resolved in the second chromatographic dimension. It was also suggested that separation of these compounds in the second dimension is most likely attributable to crowding of substituents and this observation was corroborated by molecular surface area calculations

and was used to tentatively assign structures to the 3-BPCA-OH compound series. A rich diversity of isomeric structures was also uniquely revealed by this method whereas isomeric structures are incapable of being resolved by FT-ICR MS.

Moreover, the analysis of a diverse array of OM mixtures facilitated the creation of a classification template that enabled the rapid, visual inspection of a CuO hydrolysate's chemical composition. It was shown, for example, that the cutin-derived acids, lignin dimers, aldehydic lignin monomers, carboxylated lignin monomers, and hydroxylated fatty acids and, to some extent, BPCAs, all elute within discrete areas of the 2-D chromatogram. This classification template can quickly be overlain on chromatograms of other CuO-oxidized samples to provide a rapid assessment of the CuO hydrolysate's chemical composition and the types of organic matter it contains, which will ultimately be reflective of its origins and diagenetic state in the environment. While over 300 compounds were identified in this mixture and this likely represents one of the most comprehensive compound-specific characterizations of a natural organic mixture to date, over 90 % of the peaks in this sample still remain unidentified. Further illumination of this population of peaks and comparison of peak populations between contrasting environmental samples may gradually reveal the identity of individual compounds that constitute the majority of refractory DOM.

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM. The 313 compounds and their unique ion (U^+) abundances identified by GC×GC-TOF-MS of CuO-oxidized SPE-DOM from the Upper Truckee River (UTR), ordered by retention time. Chemical Abstracts Service (CAS) Registry Numbers (where available), chemical formulae, and molecular weights (MW) refer to the TMS derivatives of the identified compounds. Areas are computed using the ion counts of the U+, which is identified automatically by ChromaTOF during data processing. Areas for compounds that were sliced by the modulator, but which were subsequently recombined during data processing, effectively represent volumes. Compound identifications were made, as indicated, based on information derived from: **a** = NIST 2011 mass spectra library and replicate library matches; **b** = authentic standards; **c** = (Goñi and Hedges 1992); **d** = retention time (RT) extrapolation based on the apparent presence of homologous compound series and corroborated by relevant m/z ions; **e** = interpretation of mass spectra; **f** = retention time (RT) agreement based on translation of published relative retention times (RRTs) to the present method from those contained in NIST 2011 and elsewhere; **g** = (Dickens et al. 2007); **h** = (Goñi and Hedges 1990); **i** = (Goñi and Hedges 1995); **j** = (Goñi 1992); **k** = (Kallio et al. 2006); **m** = (Hartley et al. 1990); **(I)** = positional isomer; **(SI)** stereoisomer; **(HH)** = "head-to-head" photodimer; **(HT)** = "head to tail" photodimer; mk = monoketone; dk = diketone; **ǂ =** newly identified lignin monomer or dimer in CuO-hydrolyzed OM; **8** unambiguously lignin-derived.

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^+ m/z	U^+ Area ($\times 10^4$)	Basis for ID
1	4.1	1.54	2-pentenoic acid, 2-hydroxy, 4-methyl	72088-30-3	C12H26O3Si2	159	5.2	a, f
2	4.2	1.77	phenol, 3,5-dimethyl	17994-05-7	C11H18OSi	105	45.6	a, f
3	4.2	1.90	phenol, 2-isopropyl	-	C12H20OSi	193	1.2	a, f
4	4.2	2.42	furan, 2-carboxylic acid, 3-methyl	-	C9H14O3Si	139	6.4	a
5	4.3	1.84	phenol, 2,6-dimethyl	16286-54-7	C11H18OSi	179	89.8	a, f
6	4.4	1.17	2-butenoic acid, 3-methyl, 2-hydroxy	55044-79-6	C11H24O3Si2	66	1.9	a, f
7	4.4	1.73	octanoic acid	55494-06-9	C11H24O2Si	201	2.2	a, f
8	4.4	2.13	phenol, 2-methoxy	6689-37-8	C10H16O2Si	196	18.7	a, f
9	4.5	1.53	pentanoic acid, 4-methyl, 2-hydroxy	54890-08-3	C12H28O3Si2	159	23.2	a, f
10	4.5	2.03	phenol, 3-propyl	-	C12H20OSi	193	6.0	e
11	4.5	2.25	hexanoic acid, 2-oxo	74367-72-9	C9H18O3Si	145	96.8	a
12	4.6	1.68	hexanoic acid, 2-hydroxy	54890-07-2	C12H28O3Si2	159	11.6	a, f
13	4.6	2.00	phenol, 2,3-dimethyl	-	C11H18OSi	149	17.8	a, f
14	4.6	2.23	benzoic acid	2078-12-8	C10H14O2Si	105	248.1	a, b, f
15	4.6	2.38	pyridine, 3-carboxylic acid	25436-37-7	C9H13NO2Si	180	0.5	a, f
16	4.7	2.43	2-thiophenecarboxylic acid	55557-11-4	C8H12O2SSi	186	0.2	a
17	4.8	1.45	glycerol	6787-10-6	C12H32O3Si3	205	0.5	a, f
18	4.8	2.09	phenol, 3-isopropyl	-	C12H20OSi	193	17.6	a
19	4.8	2.43	phenol, 4-methoxy	6689-38-9	C10H16O2Si	181	9.6	a, f
20	5.0	2.04	phenol, 4-isopropyl	-	C12H20OSi	193	28.6	a, f
21	5.0	2.31	butenedioic acid (cis)	23508-82-9	C10H20O4Si2	147	7.7	a, f
22	5.0	2.36	phenol, 3-bromo	36601-47-5	C9H13BrOSi	137	0.6	a

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
23	5.0	2.55	benzeneacetic acid	2078-18-4	C11H16O2Si	65	16.9	a, f
24	5.1	2.11	butanedioic acid	40309-57-7	C10H22O4Si2	14	16.3	a, f
25	5.1	2.88	1-cyclohexene-1-ol	6651-36-1	C9H18OSi	155	0.1	a
26	5.2	1.68	hexanoic acid, 3-hydroxy	136788-82-4	C12H28O3Si2	219	0.5	a, f
27	5.2	1.95	butanedioic acid, 2,2-dimethyl	55530-54-6	C12H26O4Si2	69	37.7	a, f
28	5.3	2.01	butanedioic acid, 2-methyl	55557-26-1	C11H24O4Si2	147	843.6	a, f
29	5.4	2.09	phenol, 2-methyl, 4-isopropyl	55012-80-1	C13H22OSi	207	3.1	a, f
30	5.4	2.40	phenol, 4-bromo	36971-28-5	C9H13BrOSi	231	0.8	a, f
31	5.4	2.48	benzoic acid, 2-methyl	55557-15-8	C11H16O2Si	119	16.0	a
32	5.5	3.17	2-furancarboxylic acid, tetrahydro-2-methyl-5-oxo	-	C9H16O4Si	173	11.1	a
33	5.5	3.33	pyridine, 2-carboxylic acid	17881-49-1	C9H13NO2Si	136	2.7	a
34	5.5	3.37	cyclohexene, n-heptyl	15232-93-6	C13H24	81	0.5	a, f
35	5.6	1.72	propanoic acid, 2,3-dihydroxy	38191-87-6	C12H30O4Si3	189	13.7	a, f
36	5.6	1.91	butenedioic acid (trans)	17962-03-7	C10H20O4Si2	147	647.2	a, f
37	5.6	2.19	phenol, -methyl, -isopropyl (I)	-	C13H22OSi	207	0.5	a
38	5.6	2.24	butanedioic acid, methylene	55494-04-7	C11H22O4Si2	147	54.5	a, f
39	5.6	2.69	benzaldehyde, 2-hydroxy	1078-31-5	C10H14O2Si	179	24.1	a
40	5.6	4.22	phthalic anhydride	85-44-9	C8H4O3	104	0.9	a
41	5.7	1.67	hexane, 1,6-diol	6222-22-6	C12H30O2Si2	82	0.5	a
42	5.7	2.93	exo-norborneol	37555-26-3	C10H20OSi	184	2.3	a
43	5.8	1.92	nonanoic acid	82326-11-2	C12H26O2Si	215	3.2	a, d, f
44	5.8	2.30	pyrrole-2-carboxylic acid	-	C11H21NO2Si2	166	75.2	a
45	5.8	2.39	butenedioic acid, methyl	77220-12-3	C11H22O4Si2	259	7.9	a, f
46	5.8	2.54	benzoic acid, 3-methyl	-	C11H16O2Si	119	57.0	a, f
47	5.8	2.68	heptanoic acid, 6-oxo	-	C10H20O3Si	198	15.7	a, d
48	5.9	1.80	heptanoic acid, 2-hydroxy	-	C13H30O3Si2	173	7.8	a
49	5.9	1.99	butanedioic acid, 2,3-dimethyl	55557-29-4	C12H26O4Si2	69	32.3	a
50	5.9	2.62	benzoic acid, 4-methyl	-	C11H16O2Si	119	31.3	a
51	6.0	2.33	phenol, -methyl, -isopropyl (I)	-	C13H22OSi	207	0.3	a
52	6.0	3.94	acetophenone, 4-methoxy	100-06-1	C9H10O2	135	0.8	a, b, f
53	6.1	2.88	benzaldehyde, 4-hydroxy	1012-12-0	C10H14O2Si	179	270.6	a

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
54	6.2	2.18	benzene, 1,3-dihydroxy	4520-29-0	C12H22O2Si2	239	23.1	a, f
55	6.2	2.35	phenol, -methyl, -isopropyl (I)	-	C13H22OSi	207	2.0	a
56	6.2	2.92	propiophenone, 2-hydroxy	33342-87-9	C12H18O2Si	193	16.8	a
57	6.3	1.82	heptanoic acid, 3-hydroxy	-	C13H30O3Si2	159	0.3	d, e
58	6.4	2.36	phenol, -methyl, -isopropyl (I)	-	C13H22OSi	133	0.4	a
59	6.5	1.52	1,1,1-tris(hydroxymethyl)propane	-	C15H38O3Si3	191	0.4	a, f
60	6.5	2.11	benzene, 1,4-dihydroxy	2117-24-0	C12H22O2Si2	239	2.5	a, f
61	6.5	2.23	pentanedioic acid	55494-07-0	C11H24O4Si2	55	235.1	a, d, f
62	6.5	2.68	benzoic acid, 3-chloro	114521-53-8	C10H13ClO2Si	213	0.7	a, f
63	6.5	3.21	phenol, 2,6-dimethoxy	-	C11H18O3Si	196	34.8	a, f
64	6.7	1.87	butanoic acid, 2,4-dihydroxy	55191-52-1	C13H32O4Si3	103	17.6	a, f
65	6.7	1.87	2-propanone, 1,3-dihydroxy	17877-42-8	C9H22O3Si2	103	17.6	a
66	6.7	2.16	pentanedioic acid, 2-methyl	55530-53-5	C12H26O4Si2	55	82.0	a, f
67	6.7	2.85	benzenepropanoic acid	21273-15-4	C12H18O2Si	104	9.0	a, f
68	6.8	3.01	acetophenone, 2-hydroxy	33342-85-7	C11H16O2Si	208	4.5	a
69	6.8	4.47	3-methyl phthalic anhydride	4792-30-7	C9H6O3	90	2.2	a
70	6.9	2.13	pentanedioic acid, 3-methyl	55517-41-4	C12H26O4Si2	31	0.5	a, f
71	7.0	2.28	propanedioic acid	56051-48-0	C12H28O4Si3	179	12.3	a
72	7.0	2.70	benzoic acid, dimethyl (I)	-	C12H18O2Si	207	1.2	a
73	7.1	2.14	pentanedioic acid, 3-ethyl	-	C13H28O4Si2	204	3.3	a
74	7.2	2.32	benzene, 1-methyl, 3,5-dihydroxy	89267-67-4	C13H24O2Si2	268	22.0	a, f
75	7.2	2.42	pentenedioic acid, 3-methyl	55887-63-3	C12H24O4Si2	82	6.9	a
76	7.2	2.76	benzoic acid, dimethyl (I)	-	C12H18O2Si	207	1.6	a
77	7.2	2.94	octanoic acid, 7-oxo	-	C11H22O3Si	117	15.8	a, d
78	7.2	3.42	benzeneglyoxalic acid	55517-36-7	C11H14O3Si	105	75.5	a, f
79	7.2	4.75	2,2'-bipyridine	366-18-7	C10H8N2	156	23.4	a, f
80	7.3	2.05	decanoic acid	55494-15-0	C13H28O2Si	229	1.0	a, d, f
81	7.3	2.45	pentenedioic acid	55255-43-1	C11H22O4Si2	68	35.5	a, f
82	7.3	2.89	benzoic acid, dimethyl (I)	-	C12H18O2Si	132	1.2	a
83	7.3	4.44	benzofuran (7,7a-dimethyl-3a,4,5,7a-tetrahydro-3H-benzofuran-2-one)	-	C10H14O2	123	5.8	a
84	7.5	1.82	1-undecanol	17957-64-1	C14H32OSi	229	0.1	a, f
85	7.6	2.39	benzyl alcohol, hydroxy (I)	-	C13H24O2Si2	179	2.8	0

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
86	7.6	4.50	2,4-2H-benzo[c]furanone, 3,3,4-trimethyl-	146950-80-3	C11H12O2	161	5.2	a
87	7.7	1.92	octanoic acid, 3-hydroxy	141-42-4	C14H32O3Si2	173	0.5	a, f
88	7.7	2.04	butanedioic acid, 2-methyl, 2-hydroxy	-	C14H32O5Si3	247	33.1	a, f
89	7.7	3.15	acetophenone, 4-hydroxy	18803-29-7	C11H16O2Si	193	296.1	a, b
90	7.8	2.67	benzeneacetic acid, alpha-hydroxy	2078-19-5	C14H24O3Si2	179	44.7	a, f
91	7.9	2.46	pentenedioic acid, 2-methyl	55887-63-3	C12H24O4Si2	82	3.0	a, f
92	8.0	2.44	benzyl alcohol, 4-hydroxy	18401-58-6	C13H24O2Si2	179	1.4	a, f
93	8.0	2.89	benzoic acid, 2,4-dimethyl	-	C12H18O2Si	133	11.1	a, f
94	8.1	2.04	butanedioic acid, 2-hydroxy	38166-11-9	C13H30O5Si3	233	5.9	a
95	8.2	2.38	hexanedioic acid	18105-31-2	C12H26O4Si2	111	39.4	a, d, f
96	8.2	2.56	propane, 2-phenyl, 1-2-dihydroxy	294847-15-7	C15H28O2Si2	193	16.8	a
97	8.3	2.74	benzoic acid, 2-hydroxy	3789-85-3	C13H22O3Si2	267	105.9	a, f
98	8.3	3.53	2-(4-Isopropylphenyl)-2-propanol	-	C15H26OSi	235	4.4	a
99	8.4	2.53	2-hexenedioic acid	55494-10-5	C12H24O4Si2	82	5.8	a
100	8.4	2.53	*trans*-3-hexenedioic acid	-	C12H24O4Si2	82	5.8	a, f
101	8.5	2.09	octane, 1,n-diol (I)	-	C14H34O2Si2	83	3.0	-
102	8.5	3.28	4-methoxy benzoic acid	2078-14-0	C11H16O3Si	135	15.0	a, f
103	8.7	1.87	octane, 1,8-diol	16654-42-5	C14H34O2Si2	215	0.2	a
104	8.7	3.89	vanillin ☙	6689-43-6	C11H16O3Si	194	1980.0	a, b, f
105	8.8	4.78	2(4H)-benzofuranone, 5,6,7,7a-tetrahydro-4,4,7a-trimethyl- (mint furanone)	15356-74-8	C11H16O2	111	6.7	a, f
106	8.9	1.95	nonanoic acid, 2-hydroxy	-	C15H34O3Si2	201	0.6	d, e
107	8.9	3.05	nonanoic acid, 8-oxo	-	C12H24O3Si	187	0.8	a, d
108	9.0	2.12	undecanoic acid	-	C14H30O2Si	117	1.6	a, d, f
109	9.0	2.84	2-furancarboxylic acid, 5-hydroxymethyl	55517-40-3	C12H22O4Si2	271	1.0	a, f
110	9.0	3.66	2-(4-Isopropylphenyl)-2-propanol (I)	-	C15H26OSi	235	7.8	a
111	9.1	2.86	2-furanacetic acid, α-hydroxy	55517-39-0	C12H22O4Si2	169	3.6	a
112	9.2	1.88	1-dodecanol	6221-88-1	C15H34OSi	243	0.3	a, f
113	9.2	1.98	nonanoic acid, 3-hydroxy	-	C15H34O3Si2	233	0.1	d, e
114	9.2	2.18	2-hydroxybut-2-enedioic acid	55517-46-9	C13H28O5Si3	171	0.8	a
115	9.2	2.65	benzoic acid, 3-hydroxy	3782-84-1	C13H22O3Si2	267	181.6	a, f
116	9.2	3.44	2-napthanol	18081-08-8	C13H16OSi	201	4.5	a, f
117	9.2	3.62	2-(4-Isopropylphenyl)-2-propanol (I)	-	C15H26OSi	235	2.0	a

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
118	9.2	4.09	phthalimide	10416-67-8	C11H13NO2Si	204	0.4	a
119	9.3	2.85	benzeneacetic acid, 2-hydroxy	27750-52-3	C14H24O3Si2	164	9.1	a, f
120	9.3	5.39	acetophenone, 2,5-dimethoxy	829-20-9	C10H12O3	165	0.2	a, b, f
121	9.5	2.18	pentanedioic acid, 2-hydroxy	55530-62-6	C14H32O5Si3	129	40.0	a, f
122	9.6	2.06	pentanedioic acid, 2-methyl-2-hydroxy	-	C15H34O5Si3	99	14.7	a
123	9.6	2.63	α-hydroxy-3-phenylpropionic acid	27750-45-4	C15H26O3Si2	193	3.5	a, f
124	9.6	2.81	benzoic acid, 3-methyl, 2-hydroxy	-	C14H24O3Si2	255	0.5	a
125	9.6	3.53	2-propenoic acid, 3-(p-tolyl)	-	C13H18O2Si	145	8.3	a
126	9.7	2.50	1-(1-hydroxyethenyl)-4-hydroxy-benzene	60068-18-0	C14H24O2Si2	279	6.1	a
127	9.8	2.79	benzoic acid, 4-methyl, 2-hydroxy	-	C14H24O3Si2	281	27.5	a
128	9.9	2.45	heptanedioic acid	55530-58-0	C13H28O4Si2	55	34.9	a, d, f
129	9.9	2.48	hexanedioic acid, 3-methyl	55520-93-9	C13H28O4Si2	155	22.4	a, f
130	10.0	2.09	pentanedioic acid, 3-methyl, 3-hydroxy	55590-95-9	C15H34O5Si3	109	1.2	a, f
131	10.1	2.34	2-pentene dioic acid, 2-hydroxy	55590-97-1	C14H30O5Si3	128	3.2	a, f
132	10.1	2.81	benzeneacetic acid, 3-hydroxy	27750-55-6	C14H24O3Si2	296	8.4	a, f
133	10.1	2.92	benzoic acid, 2-amino	18406-07-0	C13H23NO2Si2	266	25.1	a, f
134	10.2	2.29	2,4-hexadienedioic acid	55517-43-6	C12H22O4Si2	271	6.2	a
135	10.2	2.83	benzoic acid, 3-methyl, 4-hydroxy (I)	-	C14H24O3Si2	215	0.3	a
136	10.2	3.59	benzoic acid, trimethyl (I)		C13H20O2Si	221	40.0	a
137	10.2	3.96	acetovanillone ȣ	-	C12H18O3Si	193	609.2	a, b, f
138	10.3	2.61	benzoic acid, 4-hydroxy	2078-13-9	C13H22O3Si2	193	408.2	a, b, f
139	10.3	4.33	2-(4-Isopropylphenyl)-2-propanol (I)	-	C15H26OSi	161	16.2	a
140	10.5	1.99	decanoic acid, 2-hydroxy	-	C16H36O3Si2	215	0.3	d, e, f
141	10.5	2.89	benzeneacetic acid, 4-hydroxy	27750-57-8	C14H24O3Si2	252	6.4	a, f
142	10.6	2.22	dodecanoic acid	55520-95-1	C15H32O2Si	117	258.3	a, d, f
143	10.6	2.73	benzoic acid, 4-hydroxy-3-methyl	-	C14H24O3Si2	237	50.6	a
144	10.6	3.12	decanoic acid, 9-oxo	-	C13H26O3Si	201	1.3	a, d
145	10.6	4.02	2-(4-Isopropylphenyl)-2-propanol (I)	-	C15H26OSi	235	37.1	a
146	10.7	2.37	butanedioic acid, 2-carboxy, 3-methyl	61713-71-1	C15H32O6Si3	259	6.5	a
147	10.7	2.98	2,5-furandicarboxylic acid	55494-09-2	C12H20O5Si2	285	3.5	a, f
148	10.8	2.00	decanoic acid, 3-hydroxy	-	C16H36O3Si2	233	0.6	a
149	11.0	2.48	pentenedioic acid, 2-hydroxy	-	C14H30O4Si3	156	11.2	a
150	11.1	2.64	benzoic acid, 3-methyl, 4-hydroxy (I)	-	C14H24O3Si2	281	4.3	a

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
151	11.2	2.22	hexanedioic acid, 2-hydroxy	-	C15H34O5Si3	129	14.1	a, f
152	11.2	2.94	benzaldehyde, 2,4-dihydroxy	33617-38-8	C13H22O3Si2	267	95.9	a
153	11.4	2.66	benzenepropanoic acid, 2-hydroxy	-	C15H26O3Si2	251	0.6	a
154	11.4	3.16	pyridine, -hydroxy, -carboxylic acid (I)	36972-82-4	C12H21NO3Si2	268	1.4	a, f
155	11.5	2.52	octanedioic acid	43199-48-0	C14H30O4Si2	55	29.6	a, d, f
156	11.5	3.24	benzoic acid, 2-hydroxy-3-methoxy	-	C14H24O4Si2	297	17.1	a, f
157	11.5	3.38	o-benzenedicarboxylic acid	2078-22-0	C14H22O4Si2	147	104.0	a, b, f
158	11.5	3.76	benzoic acid, 3,4-dimethoxy	2078-16-2	C12H18O4Si	239	3.0	a, f
159	11.6	2.66	benzoic acid, 3-methyl, 4-hydroxy (I)	-	C14H24O3Si2	237	17.2	a
160	11.6	4.34	syringaldehyde ϐ	6651-62-3	C12H18O4Si	224	480.3	a, b
161	11.8	2.73	benzenepropanoic acid, 3-hydroxy	55887-87-1	C15H26O3Si2	220	16.8	a, f
162	11.8	2.98	acetophenone, 2,4-dihydroxy	-	C14H24O3Si2	281	29.4	a
163	11.9	2.19	pent-2-enedioic acid, 2-hydroxy	72361-20-7	C14H30O5Si3	347	1.0	a
164	11.9	3.10	benzoic acid, 3-hydroxy, 4-methoxy	-	C14H24O4Si2	251	2.9	a, f
165	12.0	2.04	undecanoic acid, 2-hydroxy	-	C17H38O3Si2	83	0.3	a
166	12.0	2.80	4-(1-hydroxyvinyl)-2-methoxyphenol	54619-52-2	C15H26O3Si2	279	15.4	a
167	12.2	2.42	pentanedioic acid, 3-carboxy	-	C15H32O6Si3	185	10.5	a, f
168	12.2	2.95	benzoic acid, hydroxy, -methoxy (I)	-	C14H24O4Si2	297	5.2	a, f
169	12.3	2.59	glutaconic acid	-	C14H30O4Si3	156	167.1	a
170	12.3	2.59	pentenedioic acid, ?-hydroxy (I)	-	C14H30O4Si3	156	42.7	a
171	12.5	2.46	prop-1-ene-1,2,3-tricarboxylic acid	55530-72-8	C15H30O6Si3	147	83.6	a, f
172	12.5	2.86	m-benzenedicarboxylic acid	106450-31-1	C14H22O4Si2	295	27.6	a, b
173	12.5	2.90	benzenepropanoic acid, 4-hydroxy	27750-62-5	C15H26O3Si2	219	2.0	a, f
174	12.6	3.18	vanillic acid ϐ	2078-15-1	C14H24O4Si2	223	662.4	a, b, f
175	12.8	2.26	heptanedioic acid, 2-hydroxy	-	C16H36O5Si3	185	1.5	d, e
176	12.8	3.06	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	279	1.4	a
177	12.8	4.30	acetosyringone ϐ	-	C13H20O4Si	253	98.6	b
178	12.9	3.84	1-naphthalenecarboxylic acid	-	C14H16O2Si	127	0.4	a
179	13.0	2.62	p-benzenedicarboxylic acid	4147-84-6	C14H22O4Si2	103	502.4	a, b
180	13.0	3.81	1,3-dioxoisobenzofuran-5-carboxylic acid	-	C12H12O5Si	249	2.1	a, f
181	13.1	2.55	nonanedioic	17906-08-0	C15H32O4Si2	55	40.2	a, d, f
182	13.2	4.19	5-formylvanillin ϐ	-	C12H16O4Si	237	34.8	c, f
183	13.3	3.09	cinnamic acid, 2-hydroxy	32426-62-3	C15H24O3Si2	125	1.0	a, f

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
184	13.3	3.44	phenol, 2-ethanone, 4-methyl	97389-70-3	C12H18O2Si	207	29.9	a
185	13.5	2.56	benzoic acid, 3,4-dihydroxy	2347-40-2	C16H30O4Si3	372	9.0	a, b, f
186	13.5	2.83	3-vanilpropanol ☋	-	C16H30O3Si2	206	3.6	a
187	13.6	2.05	dodecanoic acid, 2-hydroxy	-	C18H40O3Si2	83	0.2	d, e
188	13.6	2.98	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	293	6.0	a
189	13.7	2.52	benzoic acid, 3,5-dihydroxy	79314-27-5	C16H30O4Si3	170	80.4	a, b, f
190	13.8	2.08	dodecanoic acid, 3-hydroxy	-	C18H40O3Si2	233	0.5	a
191	13.8	3.06	benzeneglyoxalic acid, 4-hydroxy	-	C14H22O4Si2	193	321.4	i
192	13.8	3.06	benzoic acid, 2,5-dihydroxy	3618-20-0	C16H30O4Si3	193	321.4	a, f
193	13.8	4.42	5-formylacetovanillone ☋	-	C13H18O4Si	251	102.7	c, f
194	13.9	2.25	tetradecanoic acid	18603-17-3	C17H36O2Si	104	0.1	a, d, f
195	13.9	3.30	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	275	1.2	a
196	14.0	2.88	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	293	4.4	a
197	14.0	3.53	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	309	151.3	a
198	14.2	3.13	4-(1-hydroxyvinyl)-2,3-dimethoxyphenol	-	C16H28O4Si2	340	1.4	d, e
199	14.3	2.27	octanedioic acid, 2-hydroxy	-	C17H38O5Si3	149	1.5	d, e
200	14.3	2.80	cinnamic acid, 3-hydroxy	32342-01-1	C15H24O3Si2	203	4.5	a, f
201	14.3	3.44	benzoic acid, 2-hydroxy, 5-ethanone	-	C15H24O4Si2	275	2.7	a, f
202	14.5	3.13	naphthoresorcinol	-	C16H24O2Si2	304	2.9	a
203	14.7	2.53	decanedioic acid	18408-42-9	C16H34O4Si2	129	2.2	a, d, f
204	14.7	3.24	vanillylpropionic acid ☋	56051-49-1	C16H28O4Si2	309	163.8	a, f
205	14.8	3.46	syringic acid ☋	10517-29-0	C15H26O5Si2	141	157.2	a, b
206	14.9	3.09	benzoic acid, 2-hydroxy, 5-ethanone (I)	-	C15H24O4Si2	309	178.8	a
207	15.4	2.95	cinnamic acid, 4-hydroxy ☋	10517-30-3	C15H24O3Si2	293	38.9	a, b, f
208	15.5	3.20	3-hydroxy-1,2-benzenedicarboxylic acid	-	C17H30O5Si3	221	58.8	g, f
209	15.6	2.88	2-hydroxy-1,3-benzenedicarboxylic acid	-	C17H30O5Si3	221	20.9	g, f
210	15.7	3.70	5-formylvanillic acid ☋	-	C15H24O5Si2	325	346.3	c, f
211	15.8	2.29	nonanedioic acid, 2-hydroxy	-	C18H40O5Si3	95	3.8	d, e
212	15.8	3.13	5-carboxyvanillin ☋	-	C15H24O5Si2	251	21.2	c, f
213	16.0	3.43	vanillylglyoxalic acid ☋	-	C15H24O5Si2	223	428.1	c, f
214	16.1	2.35	4-pyridoxic acid	-	C17H33NO4Si3	281	2.5	a, f
215	16.1	2.56	2-hydroxy-1,4-benzenedicarboxylic acid	-	C17H30O5Si3	221	128.7	b, g, f
216	16.2	2.53	undecanedioic acid	106450-25-3	C17H36O4Si2	129	1.0	a, d, f

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
217	16.2	2.85	4-hydroxy-1,2-benzenedicarboxylic acid	-	C17H30O5Si3	147	196.7	b, g, f
218	16.2	3.41	6-carboxyvanillin ☋	-	C15H24O5Si2	325	8.4	c, f
219	16.3	2.76	4-hydroxy-1,3-benzenedicarboxylic acid	-	C17H30O5Si3	221	690.6	b, g, f
220	16.5	2.09	tetradecanoic acid, 2-hydroxy	-	C20H44O3Si2	83	0.3	d, e
221	16.5	2.52	5-hydroxy-1,3-benzenedicarboxylic acid	-	C17H30O5Si3	91	105.5	b, g, f
222	16.7	2.11	tetradecanoic acid, 3-hydroxy	-	C20H44O3Si2	233	0.3	a
223	16.7	3.73	6-carboxyacetovanillone ☋	-	C16H26O5Si2	339	223.8	c, f
224	16.8	3.36	2-carboxyvanillin ☋	-	C15H24O5Si2	325	178.8	c, f
225	16.9	2.28	hexadecanoic acid	55520-89-3	C19H40O2Si	117	51.9	a, d, f
226	17.0	3.49	5-carboxyacetovanillone ☋	-	C16H26O5Si2	339	302.6	c, f
227	17.2	2.30	decanedioic acid, 2-hydroxy	-	C19H42O5Si3	109	3.3	a, f
228	17.4	2.08	pentadecanoic acid, 2-hydroxy	-	C21H46O3Si2	57	0.2	d, e
229	17.4	2.10	pentadecanoic acid (branched), 2-hydroxy	-	C21H46O3Si2	285	0.1	d, e
230	17.4	2.30	decanedioic acid, 3-hydroxy	76735-13-2	C19H42O5Si3	95	0.2	a
231	17.7	2.17	pentadecanoic acid (branched), 3-hydroxy	-	C21H46O3Si2	233	0.2	d, e
232	17.7	2.52	dodecanedioic acid	22396-19-6	C18H38O4Si2	336	0.1	a, d, f
233	17.7	3.26	cinnamic acid, 3-methoxy, 4-hydroxy ☋	32342-04-4	C16H26O4Si2	338	38.8	a, f
234	18.0	3.43	2-carboxyacetovanillone ☋	-	C16H26O5Si2	339	0.7	c, f
235	18.0	3.65	syringylglyoxalic acid ☋	-	C16H26O6Si2	253	7.5	c, f
236	18.3	3.24	6-carboxyvanillic acid ☋	-	C18H32O6Si3	413	12.1	c, f
237	18.3	3.40	1,2,3-benzenetricarboxylic acid	-	C18H30O6Si3	175	6.2	b, g, f
238	18.3	3.78	2-carboxysyringaldehyde ☋	-	C18H32O6Si3	136	15.8	c, f
239	18.5	2.05	1-octadecanol	18748-98-6	C21H46OSi	57	0.1	a, f
240	18.5	3.04	5-carboxyvanillic acid ☋	-	C18H32O6Si3	251	891.5	c, f
241	18.6	2.31	undecanedioic acid, 2-hydroxy	-	C20H44O5Si3	81	0.5	d, e
242	18.7	2.73	1,2,4-benzenetricarboxylic acid	-	C18H30O6Si3	147	409.4	b, g, f
243	19.0	2.54	tridecanedioic acid	-	C19H40O4Si2	129	1.1	d, e
244	19.3	2.11	hexadecanoic acid, 2-hydroxy	-	C22H48O3Si2	57	0.3	d, e
245	19.4	2.39	1,3,5-benzenetricarboxylic acid	-	C18H30O6Si3	337	1.3	b, g, f
246	19.5	2.12	hexadecanoic acid, 3-hydroxy	-	C22H48O3Si2	233	0.1	a
247	19.5	2.38	trans-9-octadecenoic acid	96851-47-7	C21H42O2Si	337	1.5	a, f
248	19.7	2.31	octadecanoic acid	18748-91-9	C21H44O2Si	117	14.3	a, d, f
249	19.9	3.48	sinapic acid ☋	27750-80-7	C17H28O5Si2	338	1.4	a, f

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	Basis for ID
250	20.0	2.32	dodecanedioic acid, 2-hydroxy	-	C21H46O5Si3	345	0.0	d, e
251	20.7	2.66	1,2,4-benzenetricarboxylic acid, 6-hydroxy	-	C21H38O7Si4	499	27.6	d, e
252	20.9	2.65	1,2,4-benzenetricarboxylic acid, 5-hydroxy	-	C21H38O7Si4	395	1.0	d, e
253	20.9	2.80	1,2,4-benzenetricarboxylic acid, 3-hydroxy	-	C21H38O7Si4	247	18.2	d, e
254	20.9	3.31	ǂ guiacol, -dicarboxy, -formyl (I) ȣ	-	C19H32O7Si3	339	100.5	0
255	21.1	2.54	1,3,5-benzenetricarboxylic acid, 2-hydroxy	-	C21H38O7Si4	337	46.8	d, e
256	21.2	2.39	tridecanedioic acid, 2-hydroxy	-	C22H48O5Si3	18	4.6	d, e
257	21.2	3.21	ǂ guiacol, -dicarboxy, -formyl (I) ȣ	-	C19H32O7Si3	339	19.8	0
258	21.3	3.09	1,2,3-benzenetricarboxylic acid, 5-hydroxy	-	C21H38O7Si4	337	6.7	d, e
259	21.6	3.36	1,2,3-benzenetricarboxylic acid, 4-hydroxy	-	C21H38O7Si4	499	1.4	d, e
260	21.7	4.03	dehydroabietic acid	21414-49-3	C23H36O2Si	239	21.0	a, f
261	22.3	3.66	ǂ p-hydroxyphenyl-p-hydroxyacetophenone (α,1-mk) ȣ	-	C19H26O3Si2	193	7.2	d, e
262	22.8	2.58	1,2,4,5-benzenetetracarboxylic acid	-	C22H38O8Si4	527	1.6	d, e
263	22.9	2.95	1,3,4,5-benzenetetracarboxylic acid	-	C22H38O8Si4	527	5.1	d, e
264	23.2	2.78	1,2,3,4-benzenetetracarboxylic acid	-	C22H38O8Si4	528	2.9	d, e
265	23.4	4.44	5,5'-dehydrovanillin-p-hydroxybenzaldehyde ȣ	-	C21H28O5Si2	401	0.4	c, f
266	24.0	2.19	hexadecanoic acid, 9 and 10,16-dihydroxy	-	C25H56O4Si3	129	0.1	j, k, f
267	24.0	4.41	ǂ 5,5'-dehydroacetovanillone-p-hydroxybenzaldehyde ȣ	-	C22H30O5Si2	415	0.1	d, e
268	24.1	4.06	p-hydroxyphenylvanillone (α,1-mk) ȣ	-	C20H28O4Si2	358	0.5	c, f
269	24.2	4.44	5,5'-dehydrovanillin-p-hydroxyacetophenoneȣ	-	C22H30O5Si2	415	0.2	c, f
270	24.8	3.74	ǂ 5,5'-dehydrovanillic-p-hydroxybenzaldehydeȣ	-	C24H36O6Si3	327	1.8	d, e
271	24.8	3.81	ǂ p,p'-dihydroxybenzil (β1-dk) ȣ	-	C20H26O4Si2	194	0.4	d, e
272	25.1	3.70	5,5'-dehydrovanillin-p-hydroxybenzoic acid ȣ	-	C24H36O6Si3	327	6.0	c, f
273	25.1	4.84	5,5'-dehydrodivanillin ȣ	-	C22H30O6Si2	431	31.8	c, f
274	25.4	4.58	ǂ α,5-(4-hydroxyacetophenone-vanillin) ȣ	-	C21H28O5Si2	401	0.3	d, e
275	25.6	4.87	5,5'-dehydrovanillinacetovanillone ȣ	-	C23H32O6Si2	445	25.4	c, f
276	25.7	3.14	1,2,3,4,5-benzenepentacarboxylic acid	-	C26H46O10Si5	643	0.1	d, e
277	25.7	4.32	vanillovanillone (α,1-mk) ȣ	-	C21H30O5Si2	165	1.5	c, f
278	26.0	4.94	5,5'-dehydrodiacetovanillone ȣ	-	C24H34O6Si2	459	4.1	c, f
279	26.2	3.17	ǂ 5,5'-dehydrovanillic-p-hydroxybenzoicacid ȣ	-	C27H44O7Si4	415	6.5	d, e
280	26.3	4.18	5,5'-dehydrovanillinvanillic acid ȣ	-	C25H38O7Si3	357	140.8	c, f
281	26.4	4.10	ǂ p-hydroxybenzovanillil (β1-dk) ȣ	-	C21H28O5Si2	223	0.3	d, e

Table 2.1: List of 313 compounds identified in CuO-oxidized riverine DOM, continued . . .

#	^{1}RT (min)	^{2}RT (s)	Compound Identified via GC×GC-TOF-MS	CAS #	Formula	U^{+} m/z	U^{+} Area ($\times 10^{4}$)	
282	26.5	4.90	α,5-vanillovanillin ☋	-	C22H30O6Si2	432	2.2	c, f
283	26.7	4.19	5,5'-dehydroacetovanillonevanillic acid ☋	-	C26H40O7Si3	371	47.6	c, f
284	26.8	4.65	2-vanillylsyringaldehyde (α,2-m) ☋	-	C23H34O6Si2	447	4.7	c, f
285	27.1	3.77	ǂ α,5-(4-hydroxyacetophenone-vanillic acid) ☋	-	C24H36O6Si3	489	0.6	d, e
286	27.1	4.72	vanillosyringone (α,1-mk) ☋	-	C22H32O6Si2	59	0.5	c, f
287	27.2	3.58	5,5'-dehydrodivanillic acid ☋	-	C28H46O8Si4	445	87.8	c, f
288	27.2	5.04	α,5-vanilloacetovanillone ☋	-	C23H32O6Si2	445	2.5	c, f
289	27.3	4.73	2-vanillylacetosyringone (α,2-m) ☋	-	C24H36O6Si2	461	2.0	c, f
290	27.5	5.41	α,5-syringovanillin ☋	-	C23H32O7Si2	461	0.9	c, f
291	27.6	5.20	2-syringylsyringaldehyde (α,2-m) ☋	-	C24H36O7Si2	477	11.0	c, f
292	27.7	4.58	vanillil (β,1-dk) ☋	-	C22H30O6Si2	223	23.1	c, f
293	27.8	3.96	2-vanillylsyringic acid (α,2-m) ☋	-	C26H42O7Si3	223	4.2	c, f
294	27.8	5.36	2-syringylacetosyringone (α,2-m) ☋	-	C25H38O7Si2	73	15.1	c, f
295	28.0	4.02	α,5-vanillovanillic acid ☋	-	C25H38O7Si3	519	5.3	c, f
296	28.0	4.59	2-syringylsyringic acid (α,2-m) ☋	-	C27H44O8Si3	565	1.2	c, f
297	28.1	5.48	α,5-syringoacetovanillone ☋	-	C24H34O7Si2	475	0.9	c, f
298	28.4	4.40	5,5'-dehydrovanillinvanillylglyoxalic acid ☋	-	C26H38O8Si3	445	2.3	c, f
299	28.5	3.01	ǂ truxinic acid, 4,4'-dihydroxy, (HH) ☋	-	C30H48O6Si4	308	2.8	m, f
300	28.5	5.20	syringosyringone (α,1-mk) ☋	-	C23H34O7Si2	478	0.4	c, f
301	28.7	2.91	ǂ truxillic acid, 4,4'-dihydroxy (HT) (SI) ☋	-	C30H48O6Si4	308	1.6	m, f
302	28.7	4.51	5,5'-dehydroacetovanillonevanillylglyoxalic acid ☋	-	C27H40O8Si3	459	0.5	c, f
303	28.9	5.00	vanillosyringil (β,1-dk) ☋	-	C23H32O7Si2	253	0.5	c, f
304	29.0	3.68	α,5-syringovanillic acid ☋	-	C26H40O8Si3	550	2.0	c, f
305	29.1	3.81	5,5'-dehydrovanillicvanillylglyoxalic acid ☋	-	C29H46O9Si4	533	1.0	c, f
306	29.4	3.14	ǂ truxillic acid, 4,4'-dihydroxy (HT) (SI) ☋	-	C30H48O6Si4	308	2.8	m, f
307	29.5	3.18	ǂ truxillic acid, 4,4'-dihydroxy (HT) (SI) ☋	-	C30H48O6Si4	308	10.1	m, f
308	29.6	3.14	ǂ truxillic acid, 4,4'-dihydroxy, 3-methoxy (HT) ☋	-	C31H50O7Si4	338	0.5	m, f
309	29.9	5.55	syringil (β,1-dk) ☋	-	C24H34O8Si2	253	0.8	c, f
310	30.3	3.42	ǂ truxillic acid, 4,4'-dihydroxy, 3,3'-dimethoxy (HT) (SI) ☋	-	C32H52O8Si4	338	4.7	m, f
311	30.5	3.50	ǂ truxillic acid, 4,4'-dihydroxy, 3,3'-dimethoxy (HT) (SI) ☋	-	C32H52O8Si4	338	0.9	m, f
312	30.8	3.52	ǂ truxillic acid, 4,4'-dihydroxy, 3,3'-dimethoxy (HT) (SI) ☋	-	C32H52O8Si4	471	0.4	m, f
313	31.1	3.66	ǂ truxillic acid, 4,4'-dihydroxy, 3,3'-dimethoxy (HT) (SI) ☋	-	C32H52O8Si4	338	1.3	m, f

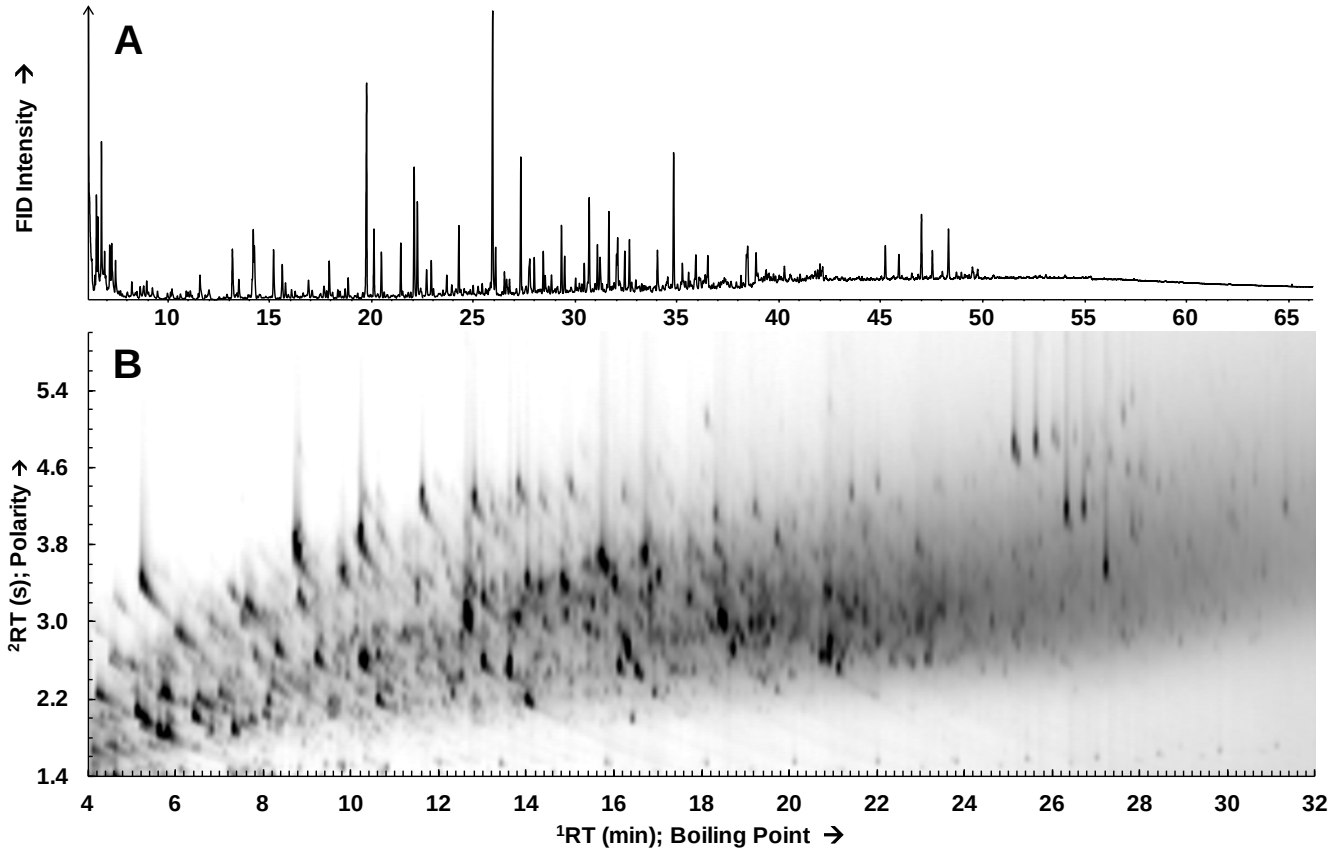


Figure 2.1: The TMS-derivatized CuO hydrolysate of Upper Truckee River SPE-DOM visualized using a traditional 1-D GC-FID setup (A) and a GC×GC-TOF-MS setup (B). The 1-D GC-FID chromatogram was acquired on a DB-5 capillary column (60 m, 0.25 mm i.d., 0.5μm df) using a 66 min oven program of 100–320 °C at 4.0 °C min^{-1}, with initial and final holds of 1 and 10 min, respectively. The GC×GC-TOF-MS method is described in §2.2.4 and was 34.5 min in duration. The 2-D chromatogram shown is constructed from the sum of the m/z values indicated in Fig. 2.2. 4,789 discretely resolvable peaks were detected at a S/N of 10 or greater by GC×GC-TOF-MS whereas about 200 peaks were counted in the 1-D trace, however the majority of these peaks were poorly resolvable or co-eluting. As such, a *minimum* of 24× more peaks were identified by GC×GC-TOF-MS than were identified by 1-D GC, despite the 1-D GC's reliance on a column and runtime that were 6× and 2× longer, respectively.

Figure 2.2: The two-dimensional GC×GC retention time space spanned by 451 compounds identified to date in the hydrolysate mixtures of CuO-oxidized NOM derived from a variety of source materials, illustrating the analytical window and patterns of separation produced by this particular method. Eight aliphatic homologous compound series defined by successive methylene (–CH_2–) additions were observed. These series were also separated in ^{2}RT-space according to their proportion of TMS-bonded oxygens (which led to lower ^{2}RTs) to C=O oxygens (which led to higher ^{2}RTs). In these series, even-numbered carbon chains predominated over odd-numbered carbon chains. These eight observed aliphatic homologous compound series included: (1) 1-alkanols, (2) alkanediols (generally α,ω), (3) 3-hydroxy fatty acids, (4) 2-hydroxy fatty acids, (5) normal fatty acids, (6) α,ω-dioic acids, (7) 2-hydroxy, α,ω-dioic acids, and (8) (ω-1)-oxo fatty acids. The addition of –CH_2–, –OCH_3, and –OH groups to compounds led to predictable shifts in ^{1}RT and ^{2}RT that propelled compounds along definable and consistent trajectories, as indicated. Addition of methylene (–CH_2–) groups (black lines) effected little change in ^{2}RTs while the addition of methoxy (–OCH_3) groups (dotted red lines) significantly increased ^{2}RTs and the addition of derivatizable hydroxy (–OTMS) groups (blue lines) and carboxy (–CO_2TMS) (not shown) groups significantly decreased the ^{2}RTs of compounds containing these groups, such as for the annotated hydroxy- ,carboxy-, and methyl-functionalized α,ω-dioic acids, as indicated. All functional group additions increased ^{1}RT. Color-coded annotations refer to suites of identified compounds which included lignin dimers (c.f. Fig. 2.3), lignin monomers, benzenepolycarboxylic acids (c.f. Fig. 2.5), and lignin carboxy and formyl monomers, and two new guiacol-type monomers, as indicated (ǂ) and listed in Table 2.1. Labels are as follows: Pl = benzaldehyde, 4-hydroxy; Pn = acetophenone, 4-hydroxy; Pd = benzoic acid, 4-hydroxy; Vl = vanillin; Vn = acetovanillone; Vd = vanillic acid; Sl = syringaldehyde; Sn = acetosyringone; Sd = syringic acid. The background image is the summed selected ion chromatogram of the TMS-derivatized CuO hydrolysate of UTR SPE-DOM. The following masses, which were broadly inclusive of the 451 compounds' characteristic fragment masses, were summed to produce this chromatogram: 29 43 57 73 89 91 103 105 117 147 177 191 193 204 217 221 223 233 251 253 293 295 308 309 310 318 327 337 357 367 371 383 388 395 397 398 401 411 415 418 425 426 431 439 445 447 448 459 453 461 475 477 478 491 499 514 519 533 549 565.

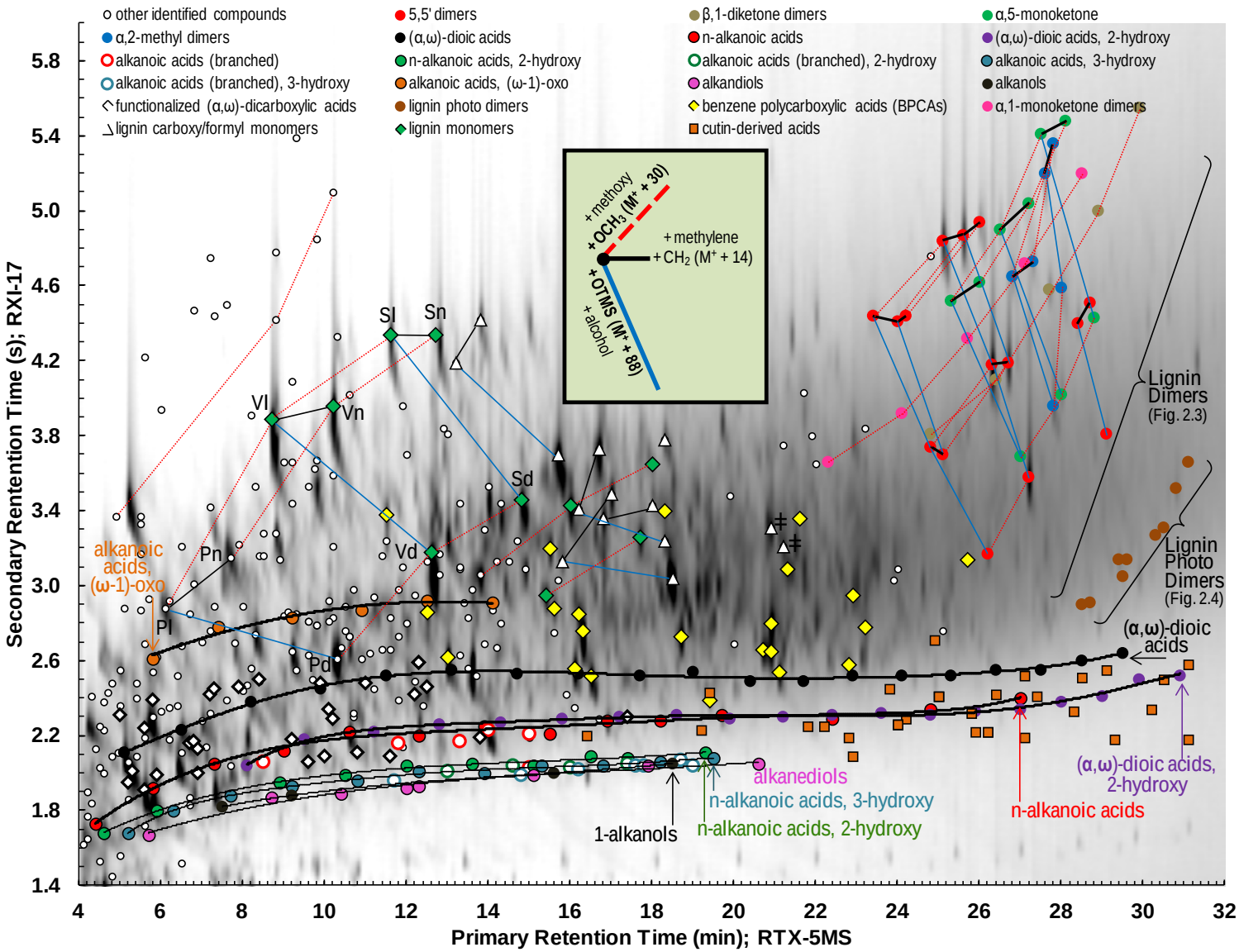
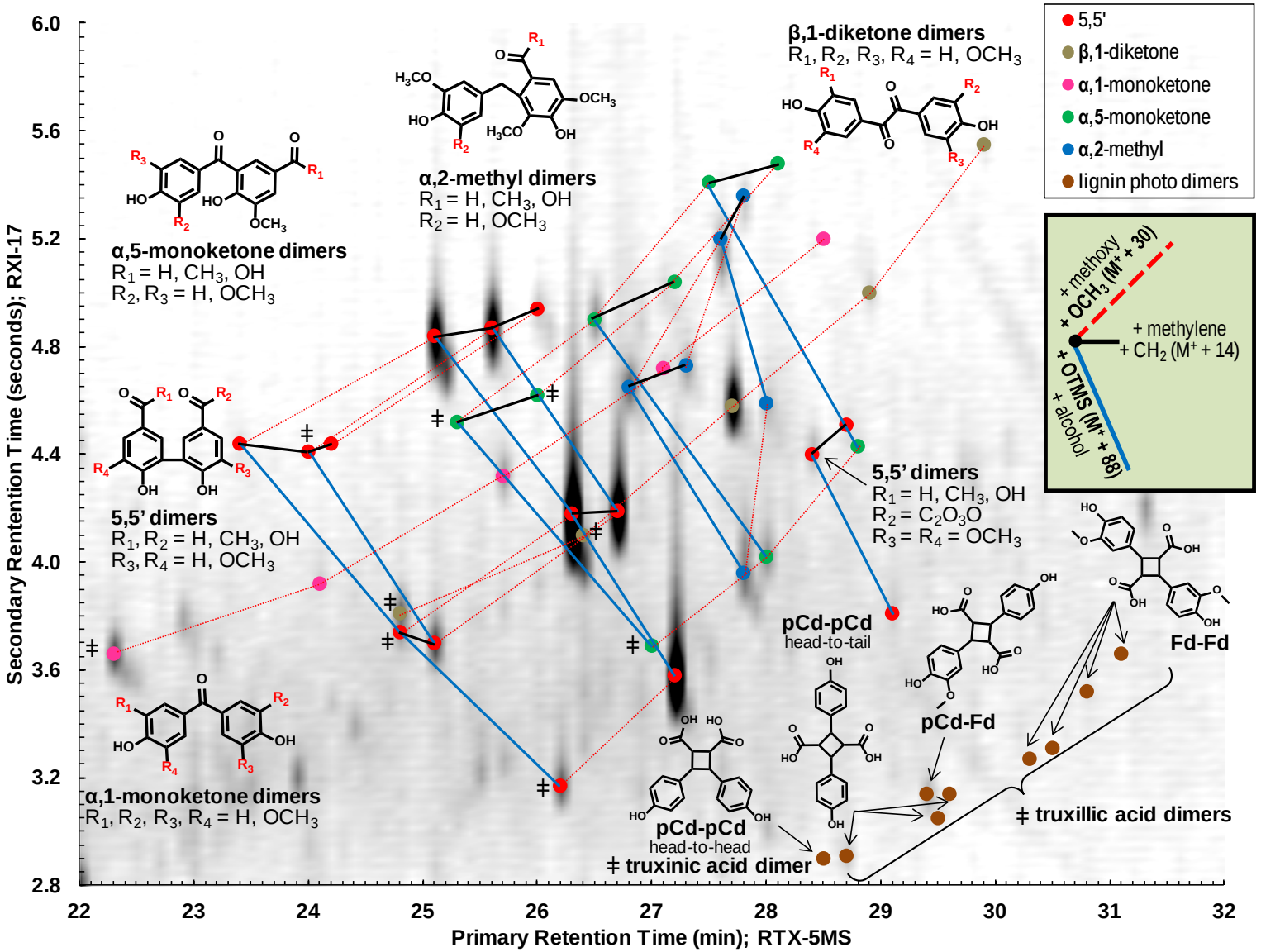

o other identified compounds
5,5' dimers
β,1-diketone dimers
α,5-monoketone
α,2-methyl dimers
(α,ω)-dioic acids
n-alkanoic acids
(α,ω)-dioic acids, 2-hydroxy
alkanoic acids (branched)
n-alkanoic acids, 2-hydroxy
alkanoic acids (branched), 2-hydroxy
alkanoic acids, 3-hydroxy
alkanoic acids, 3-hydroxy
alkanoic acids, (ω-1)-oxo
alkandiols
alkanols
functionalized (α,ω)-dicarboxylic acids
lignin photo dimers
benzene polycarboxylic acids (BPCAs)
α,1-monoketone dimers
lignin carboxy/formyl monomers
lignin monomers
cutin-derived acids
+methoxy
+OCH3 (M+ + 30)
+methylene
+CH2 (M+ + 14)
+OTMS (M+ + 88)
+alcohol
Sl
Sn
Vl
Vn
Sd
Pn
Vd
Pl
Pd
alkanoic acids, (ω-1)-oxo
Lignin Dimers (Fig. 2.3)
Lignin Photo Dimers (Fig. 2.4)
(α,ω)-dioic acids
(α,ω)-dioic acids, 2-hydroxy
alkanediols
n-alkanoic acids, 3-hydroxy
n-alkanoic acids
1-alkanols
n-alkanoic acids, 2-hydroxy
Secondary Rentention Time (s); RXI-17
Primary Retention Time (min); RTX-5MS
1.4
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6
5.0
5.4
5.8
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32

Figure 2.3: The elution pattern for lignin phenolic dimers in 2-D retention time space, as colored by the indicated linkage types: 5,5'-, β,1-diketone, α,1-monoketone, α-5-monoketone, and α,2-methyl. Red dotted lines show RT shifts caused by substituting 3- or 5- position ring hydrogens with $–OCH_3$ groups. Horizontal or sub-horizontal black lines represent replacement of an aldehydic proton with a $–CH_3$. Downward sloping blue lines represent replacement of an aldehydic proton with an –OH group, which is converted into a nonpolar –OTMS group prior to GC analysis. For all dimer types, substitution of an aldehydic proton with a $–CH_3$ consistently led to an increase in ^{1}RT, although the magnitude of this increase was smaller for monoacidic 5,5'-dimers. Methoxy group addition consistently led to increases in both ^{1}RT and ^{2}RT across all linkage types, although the magnitude of the ^{1}RT shift was less when a OCH_3 group was added to either a mono and diacidic (vs. non-acidic) 5,5'- dimer. The effect of methoxy addition was particularly evident for the α,1-monoketone dimer set. This set has five members, corresponding to the mono-, di-, tri-, tetra, and non-methoxylated versions of the core dimer structure. The β,1-diketone dimer series showed a similar pattern with regular increases in ^{1}RT and ^{2}RT that was driven by increasing degrees of methoxylation of the core dimer structure. Across all linkage types, replacement of an aldehydic proton with an –OH group led to an increase in ^{1}RT that was accompanied by a sharp decrease in ^{2}RT. Systematic investigation and visualization of these retention time patterns performed in tandem with mass spectra interpretation led to the discovery of eight new lignin dimer types (indicated by ǂ) that have hitherto gone unobserved. An additional series of substituted cyclobutane lignin dimers derived from photodimerization of *p*-coumaric (pCd) and ferulic acids (Fd) was also observed for the first time in CuO-oxidized OM (Fig. 2.4). The background selected sum ion chromatogram of the CuO hydrolysate of UTR SPE-DOM was derived from the following masses which correspond to m/z ions produced from fragmentation of lignin monomers and dimers: 193 223 253 327 357 371 308 318 388 401 415 418 431 445 447 448 459 453 461 475 477 478 491 519 533 549 565.


Secondary Rentention Time (seconds); RXI-17
Primary Retention Time (min); RTX-5MS
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
22
23
24
25
26
27
28
29
30
31
32
5,5'
β,1-diketone
α,1-monoketone
α,5-monoketone
α,2-methyl
lignin photo dimers
+ methoxy
+ OCH3 (M+ + 30)
+ methylene
+ CH2 (M+ + 14)
+ OTMS (M+ + 88)
+ alcohol
β,1-diketone dimers
R1, R2, R3, R4 = H, OCH3
α,2-methyl dimers
R1 = H, CH3, OH
R2 = H, OCH3
α,5-monoketone dimers
R1 = H, CH3, OH
R2, R3 = H, OCH3
5,5' dimers
R1, R2 = H, CH3, OH
R3, R4 = H, OCH3
5,5' dimers
R1 = H, CH3, OH
R2 = C2O3O
R3 = R4 = OCH3
α,1-monoketone dimers
R1, R2, R3, R4 = H, OCH3
pCd-pCd
head-to-tail
pCd-Fd
Fd-Fd
pCd-pCd
head-to-head
‡ truxinic acid dimer
‡ truxillic acid dimers

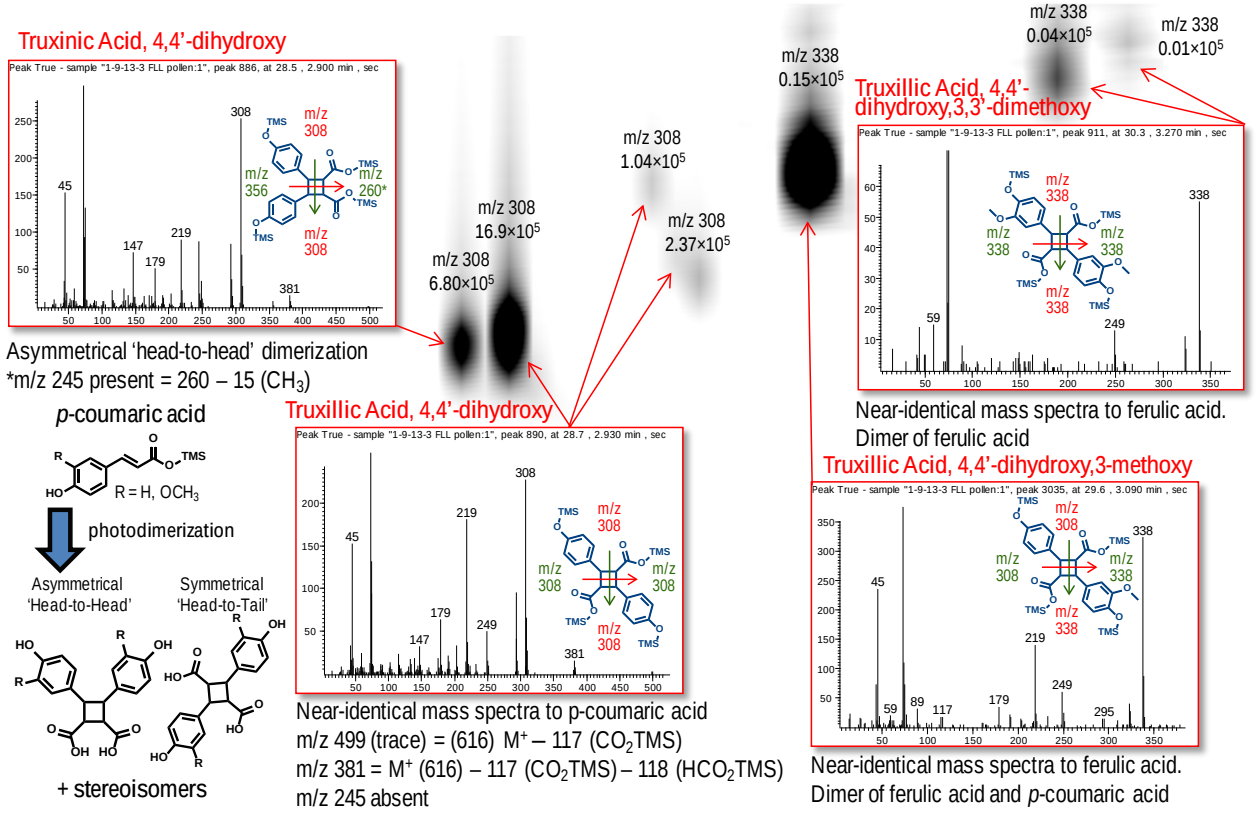


Figure 2.4: A unique set of truxinic and truxillic acid-like lignin cyclobutane dimers were newly identified in CuO-oxidized OM. These truxinic- and truxillic-acid-like compounds ostensibly derive from photo-dimerization of *p*-coumaric and ferulic acids (Ford and Hartley 1989) and exist in an array of chromatographically resolvable stereoisomers. This dimerization has been previously demonstrated for *trans*-cinnamic acid (Bernstein and Quimby 1943). Mass spectra show asymmetrical or symmetrical fragmentation, depending on whether dimerization occurs 'head-to-head' or 'head-to-tail', respectively. Counts for the indicated m/z values are shown above peaks, which were obtained from the CuO hydrolysate of pollen skimmed from the surface of Fallen Leaf Lake adjacent to UTR. A large set (n=9) of these compounds was also observed in the CuO hydrolysate of SPE-DOM from UTR.

Figure 2.5: Retention time shifts for benzene polycarboxylic acids (BPCA) and their hydroxylated counterparts are shown. These have been the subject of closer inspection in prior studies (Dickens et al. 2007). All possible positional isomers are observed: 3 each for benzene dicarboxylic (2-BPCA) and tricarboxylic acids (3-BPCA), and six each for their hydroxylated counterparts. The separation of individual isomers was more pronounced in ^{2}RT space than it was in ^{1}RT space. Separation of isomers by ^{2}RT appeared to be controlled by steric influences. Isomers with crowded substituents and greater strain energies appeared to have greater ^{2}RTs than those with less crowded substituents and lower strain energies. Numbers adjacent to yellow diamonds correspond to the Connolly Molecular Surface Area (Å^{2}) computed for each derivatized isomer. These surface areas were determined using a spherical (r = 1.4 Å) probe at the molecule's prognosed global conformational energy minima, which was determined from extensive cycles of heating and molecular relaxation using the MM2 molecular dynamics force field model, which was implemented within Chem3D Pro. Within the BPCA and BPCA-OH isomer sets examined, compounds with smaller molecular surface areas had greater ^{2}RTs. These observations, which were enabled by use of authentic standards for 2-BPCA, 3-BPCA, and 2-BPCA-OH isomers, guided the assignment of isomeric structures to the 3-BPCA-OH isomer set for which standards were unavailable and for which structures had not been previously ascertainable (Dickens et al. 2007). The addition of –OH (red lines) to either set of BPCAs caused consistent increases in ^{1}RT while the resulting shifts in ^{2}RT were more variable, but trended towards lower values. The background selected sum ion chromatogram of the CuO hydrolysate of UTR SPE-DOM was derived from the following masses which encompass mass fragments common to the BPCA suite of compounds: 514 499 439 425 426 411 398 397 395 383 367 337 310 309 295 293 251 221 191 177 147.

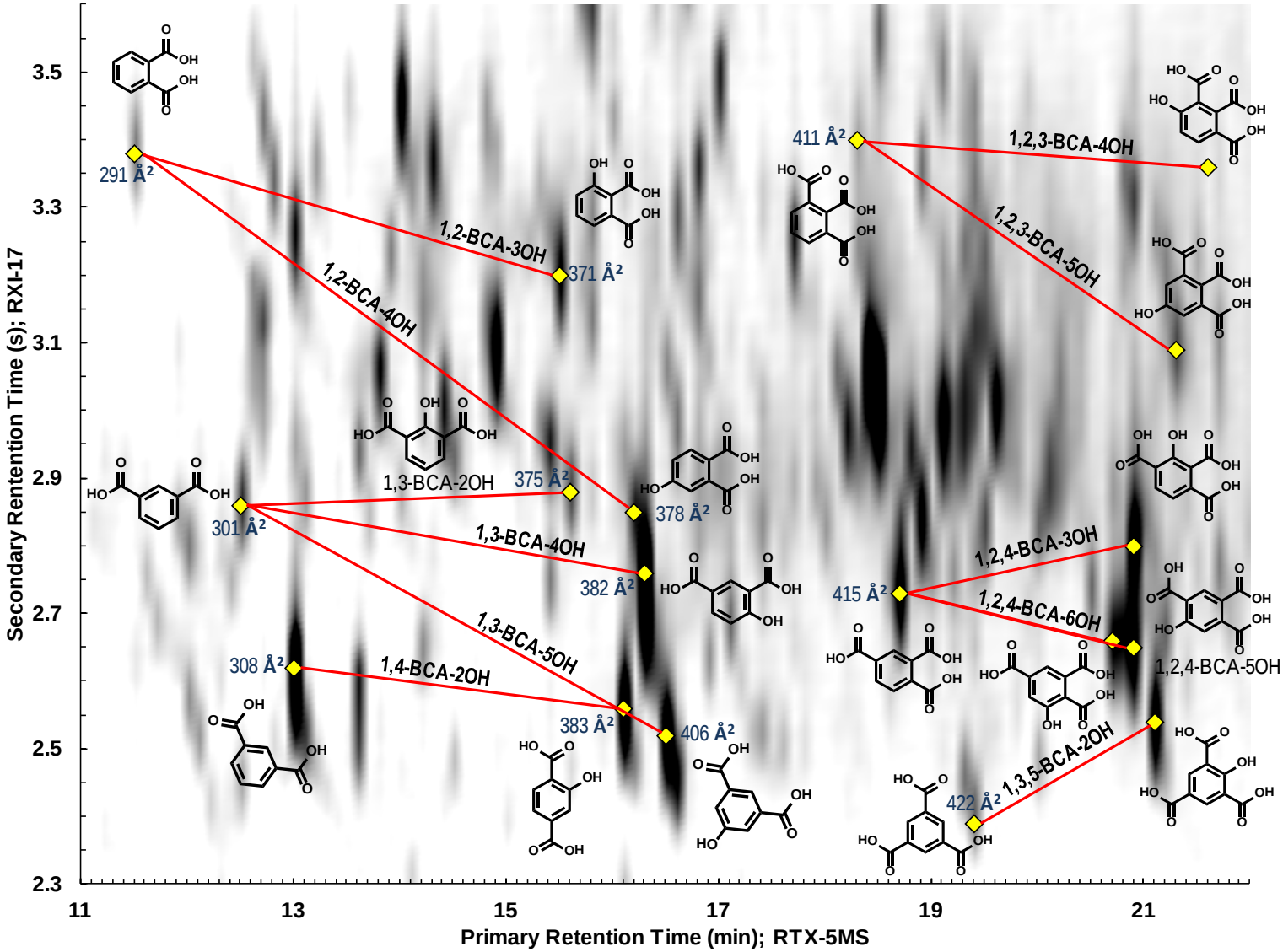
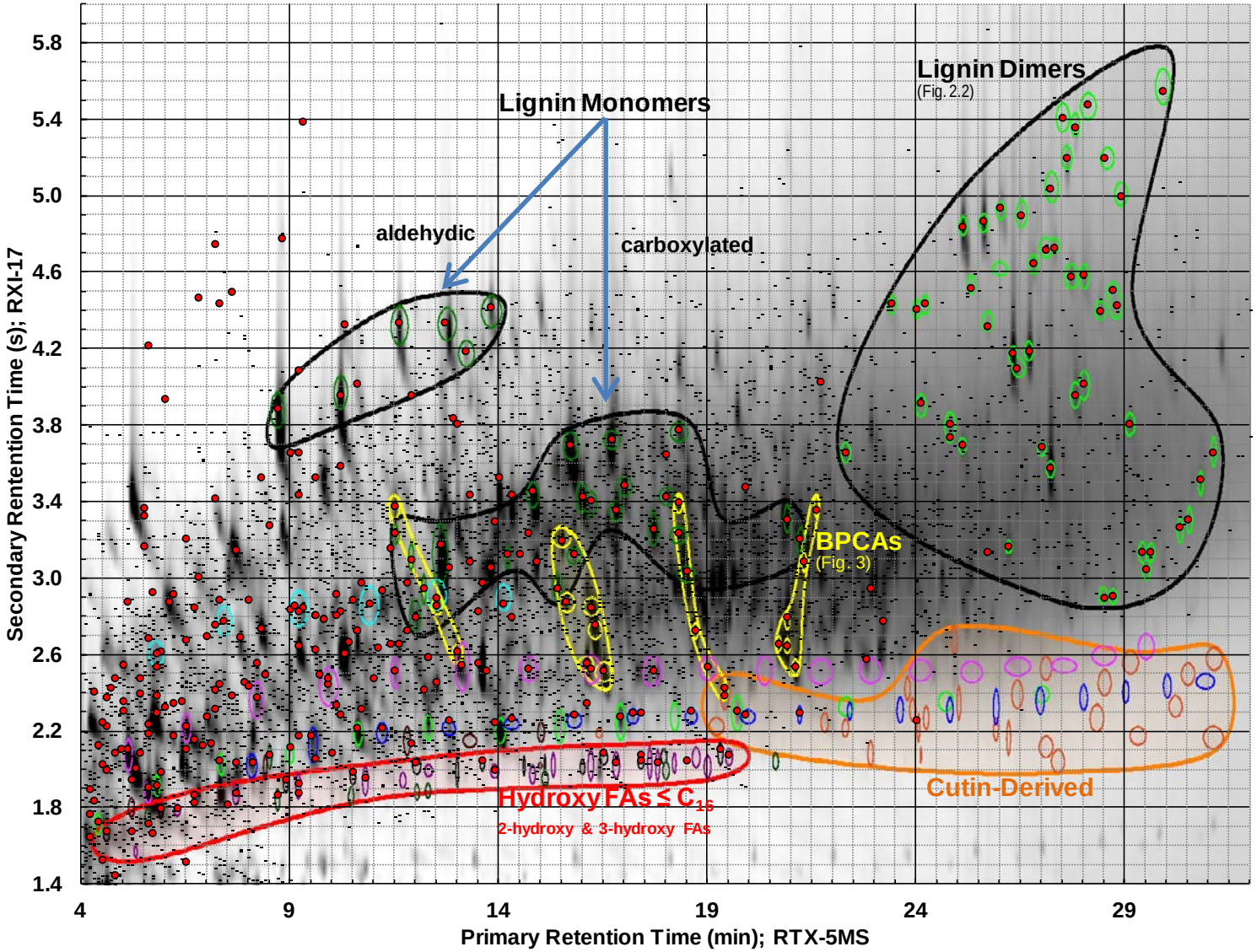

Secondary Rentention Time (s); RXI-17
Primary Retention Time (min); RTX-5MS
3.5
3.3
3.1
2.9
2.7
2.5
2.3
11
13
15
17
19
21
291 Å²
371 Å²
301 Å²
375 Å²
378 Å²
382 Å²
308 Å²
383 Å²
406 Å²
411 Å²
415 Å²
422 Å²
1,2-BCA-3OH
1,2-BCA-4OH
1,3-BCA-2OH
1,3-BCA-4OH
1,3-BCA-5OH
1,4-BCA-2OH
1,2,3-BCA-4OH
1,2,3-BCA-5OH
1,2,4-BCA-3OH
1,2,4-BCA-6OH
1,2,4-BCA-5OH
1,3,5-BCA-2OH

Figure 2.6: Peaks identified by GC×GC-TOF-MS of the TMS-derivatized CuO hydrolysate of SPE-DOM from the Upper Truckee River (UTR). The selected sum ion chromatogram is constructed from the same m/z values used in Fig. 2.2. Black dots (n = 4,789) represent discretely resolvable peaks identified at a signal to noise (S/N) of 10 or greater. The 313 red dots indicate peaks that have been tied to structures (Table 2.1) based on the identification criteria outlined in §2.3.5. These peaks collectively represented 6.5 % of the total identifiable peaks. Classification regions show the elution regions of indicated compound classes, including the lignin monomers, which separated into aldehydic and carboxylated monomers, the lignin dimers, the BPCAs, the hydroxy fatty acids, and the cutin-derived acids, most of which were polyhydroxylated $C_{\geq 16}$ fatty acids and α,ω-dioic acids. The presence/absence and localization of red dots within the enclosed classification regions combined with the visual intensity of peaks beneath the red dots and within these regions, provided an immediate impression of SPE-DOM's molecular-level chemical composition.


Lignin Monomers
aldehydic
carboxylated
Lignin Dimers
(Fig. 2.2)
BPCAs
(Fig. 3)
Cutin-Derived
Hydroxy FAs ≤ C16
2-hydroxy & 3-hydroxy FAs
Secondary Rentention Time (s); RXI-17
Primary Retention Time (min); RTX-5MS
5.8
5.4
5.0
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
4
9
14
19
24
29

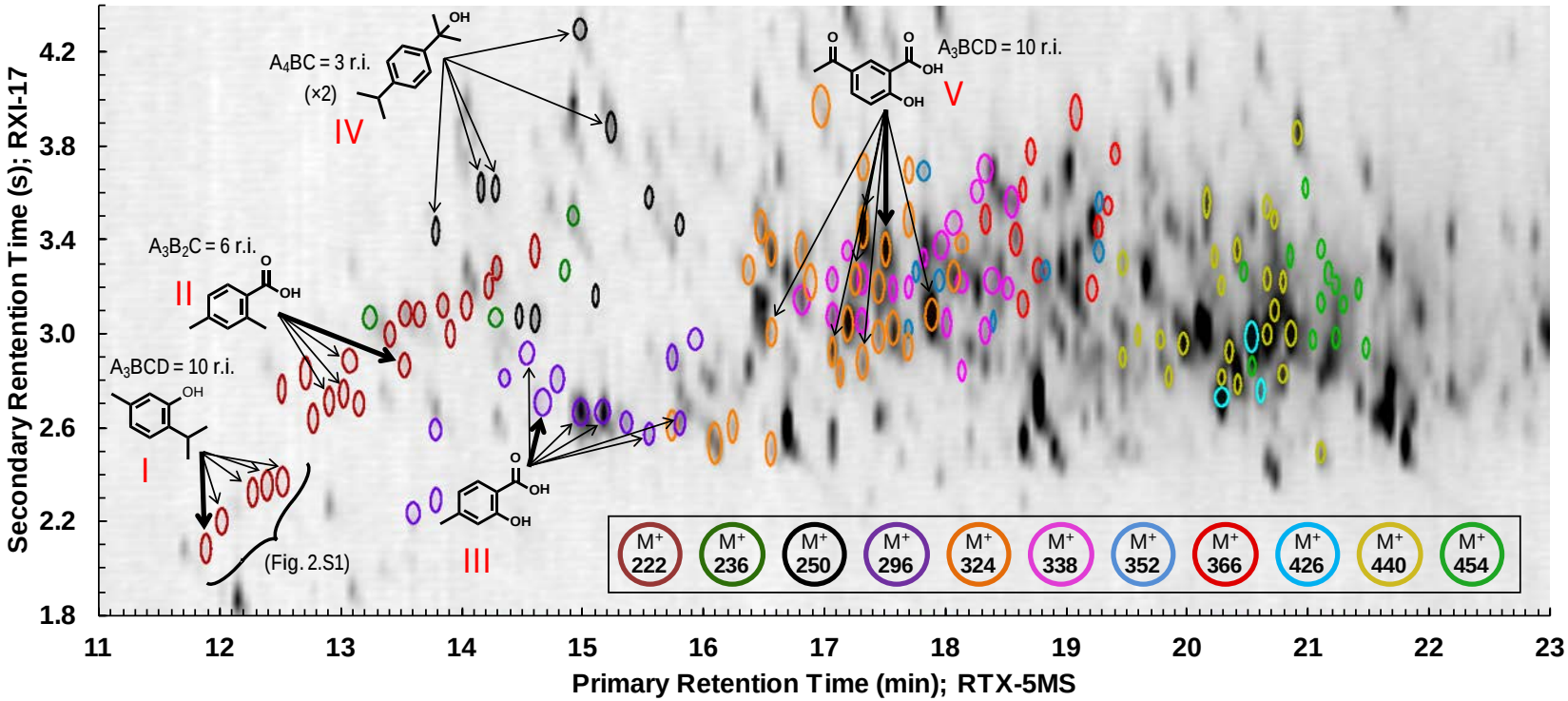


Figure 2.7: The isomeric diversity of CuO-oxidized organic matter as revealed by GC×GC-TOF-MS. The eleven color-coded circles denote peaks of identical molecular weights. Bold arrows indicate peaks whose structures have been determined based on retention time alignment *and* mass spectra matching while narrower arrows originating from the same compound indicate peaks of the *same compound* as determined by mass spectra alone, but which are present at different retention times, suggesting the presence of positional isomers of these compounds. Different numbers of positional ring isomers (PRI) are possible depending on the ring substitution pattern (Evans and Quesne 1950), which is indicated above Compounds I–V in addition to the number of ring isomers (r.i.) enabled by the substitution pattern. While compound III only allows 3 ring isomers, additional isomerism can occur due to additional isomerism in either the isopropyl or 2-propanol substituents. This would enable *at least* six possible isomers. In no instance, therefore, does the number of observed isomers exceed the number of possible isomers. The background image consists of a sum selected ion chromatogram of TMS-derivatized, CuO-oxidized SPE-DOM from the Upper Truckee River created using the intensities of the following m/z ions representing the above M^+ values –15 ($-CH_3$): 439 425 411 351 337 323 309 296 235 221 207.

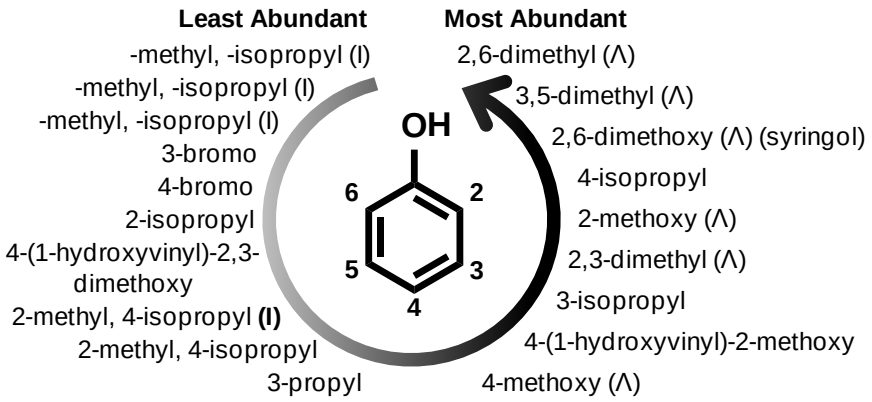


Figure 2.8: Observed substitutions for phenol compounds identified by GC×GC-TOF-MS analysis of CuO-oxidized SPE DOM from UTR. This schematic excludes phenols that contain carboxy, formyl, or acetyl substitutions and instead highlights the substitutions which appear to have escaped previous detection in CuO-hydrolyzed OM. The compounds are listed in order of increasing abundance (counter-clockwise), as indicated by the arrow. Compounds that have been previously identified by pyrolysis-MS of dissolved humic acids, fulvic acids, or bog lake water (Schulten 1999) are indicated by Λ. (I) = isomer of indicated compound.

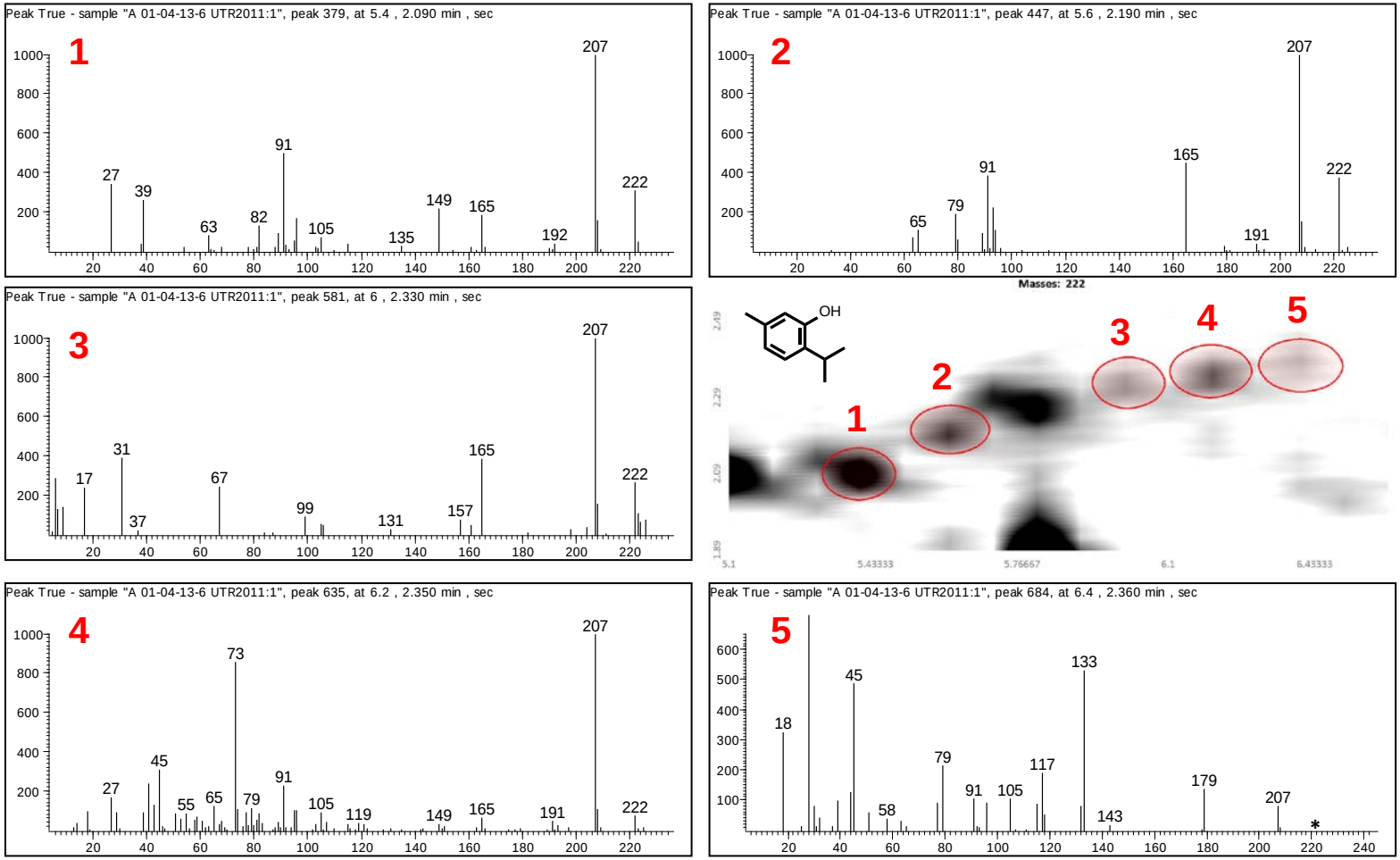


Figure 2.S1: The indicated zoom area of Fig. 2.7 is shown here and contains five peaks with a M+ of 222, 1–5. The peaks correspond to isomers of the monoterpene compound thymol and the differences in their retention times exhibit consistent spacing between this particular set of structural isomers. Ten possible positional ring structural isomers exist and five are observed here.

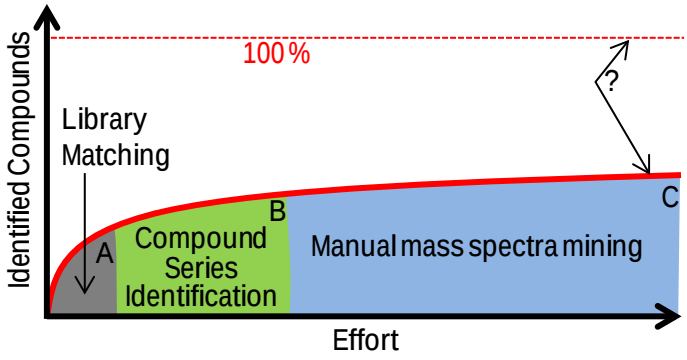


Figure 2.S2: Schematic illustrating the relationship between the number of identified compounds vs. effort for peaks identified by GC×GC-TOF-MS. Thousands of peaks are identified, but only a small subset can be assigned structures. In this study the asymptotic red line was 6.5 %. Compound identification is rapidly enabled with large EI-MS spectra libraries and automated searches (A). When homologous compound series are discovered, new sets of compounds (of varying size) can usually be found in quick succession (B). *De novo* interpretation of mass spectra data by manual mass spectra mining is time consuming and return on time invested is very small (C). It is not clear where these two lines will intersect (?), or if they ever will. The synthesis of standards by combinatorial chemistry approaches may represent one approach as commercially available standards are not likely to ever be available in the diversity that would be needed to elucidate complex mixtures peak by peak. Expanded mass spectra libraries, which continue to grow each year, probably offer the best and most economical route. Another route, combining GC×GC with mass detectors that offer both high speed and high mass resolution would further assist by revealing both the molecular formula and EI-MS mass spectra of each peak. Nevertheless, 100 % elucidation still seems unlikely, even if the abundance threshold is limited to a S/N of 10. At lower S/N values, new and larger seas of peaks will surely be found.

2.7 References

Aiken, G., D. McKnight, K. A. Thorn, and E. M. Thurman. 1992. Isolation of hydrophilic organic acids from water using nonionic macroporous resins. Organic Geochemistry **18**: 567–573.

Azam, F., T. Fenchel, J. Field, J. S. Gray, L. A. Meyer-Reil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. Marine Ecology Progress Series **10**: 257–263.

Benner, R. 2002. Chemical Composition and Reactivity, p. 56–90. *In* Biogeochemistry of Marine Dissolved Organic Matter.

Bernstein, H., and W. Quimby. 1943. The photochemical dimerization of trans-cinnamic acid. Journal of the American Chemical Society **65**: 1845–1846.

Blomberg, J., P. J. Schoenmakers, J. Beens, and R. Tijssen. 1997. Comprehensive two-dimensional gas chromatography (GCxGC) and its applicability to the characterization of complex (petrochemical) mixtures. Journal of High Resolution Chromatography **20**: 539–544.

Carlson, C. A., and H. W. Ducklow. 1995. Dissolved organic carbon in the upper ocean of the central equatorial Pacific Ocean, 1992: Daily and finescale vertical variations. Deep Sea Research Part II: Topical Studies in Oceanography **42**: 639–656.

Chang, H. M., and G. G. Allan. 1971. Oxidation, p. 916. *In* K.V. Sarkanen and C.H. Ludwig [1st ed.], Lignins: Occurrence, Formation, Structure, and Reactions. Wiley-Interscience.

Coble, P. G. 2007. Marine optical biogeochemistry: the chemistry of ocean color. Chemical reviews **107**: 402–18.

Dickens, A. F., J. A. Gudeman, Y. Gélinas, J. A. Baldock, W. Tinner, F. S. Hu, and J. I. Hedges. 2007. Sources and distribution of CuO-derived benzene carboxylic acids in soils and sediments. Organic Geochemistry **38**: 1256–1276.

Dittmar, T., and B. P. Koch. 2006. Thermogenic organic matter dissolved in the abyssal ocean. Marine Chemistry **102**: 208–217.

Druffel, E. R. M., P. M. Williams, J. E. Bauer, and J. R. Ertel. 1992. Cycling of Dissolved and Particulate Organic Matter in the Open Ocean. Journal of Geophysical Research **97**: 15639–15659.

Evans, R. F., and J. Le Quesne. 1950. A calculation of the number of positional isomers in some aromatic systems. The Journal of Organic Chemistry **15**: 19–23.

Falkowski, P., R. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hógberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffen. 2000. The global carbon cycle: a test of our knowledge of earth as a system. Science **290**: 291–296.

Ford, C. W., and R. D. Hartley. 1989. GC/MS characterisation of cyclodimers from p-coumaric and ferulic acids by photodimerisation—a possible factor influencing cell wall biodegradability. Journal of the Science of Food and Agriculture **46**: 301–310.

Frysinger, G. S., and R. B. Gaines. 1999. Comprehensive two-dimensional gas chromatography with mass spectrometric detection (GC×GC/MS) applied to the analysis of petroleum. Journal of High Resolution Chromatography **22**: 251–255.

Goñi, M. A. 1992. The use of CuO reaction products to characterize organic matter in the marine environment. University of Washington.

Goñi, M. A., and J. I. Hedges. 1990. Cutin-derived CuO reaction products from purified cuticles and tree leaves. Geochimica et Cosmochimica Acta **54**: 3065–3072.

Goñi, M. A., and J. I. Hedges. 1992. Lignin dimers: structures, distribution, and potential geochemical applications. Geochimica et Cosmochimica Acta **56**: 4025–4043.

Goñi, M. A., and J. I. Hedges. 1995. Sources and reactivities of marine-derived organic matter in coastal sediments as determined by alkaline CuO oxidation. Geochimica et Cosmochimica Acta **59**: 2965–2981.

Goñi, M. A., and S. Montgomery. 2000. Alkaline CuO oxidation with a microwave digestion system: lignin analyses of geochemical samples. Analytical Chemistry **72**: 3116–3121.

Goñi, M. A., K. Ruttenberg, and T. Eglinton. 1998. A reassessment of the sources and importance of land-derived organic matter in surface sediments from the Gulf of Mexico. Geochimica et Cosmochimica Acta **62**: 3055–3075.

Hansell, D. A., and C. A. Carlson. 1998. Deep ocean gradients in the concentration of dissolved organic carbon. Nature **395**: 263–266.

Hartley, R. D., W. H. Morrison III, F. Balza, and G. H. N. Towers. 1990. Substituted truxillic and truxinic acids in cell walls of Cynodon dactylon. Phytochemistry **29**: 3699–3703.

Hedges, J. I., and J. R. Ertel. 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. Analytical Chemistry **54**: 174–178.

Hedges, J., and P. Parker. 1976. Land-derived organic matter in surface sediments from the Gulf of Mexico. Geochimica et Cosmochimica Acta **40**: 1019–1029.

Hertkorn, N., C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt, and P. Schmitt-Kopplin. 2007. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. Analytical and Bioanalytical Chemistry **389**: 1311–1127.

Kallio, H., R. Nieminen, S. Tuomasjukka, and M. Hakala. 2006. Cutin composition of five finnish berries. Journal of Agricultural and Food Chemistry **54**: 457–462.

Kim, S., A. J. Simpson, E. B. Kujawinski, M. A. Freitas, and P. G. Hatcher. 2003. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18 solid phase disk. Organic Geochemistry **34**: 1325–1335.

Koch, B. P., M. Witt, R. Engbrodt, T. Dittmar, and G. Kattner. 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Geochimica et Cosmochimica Acta **69**: 3299–3308.

Kujawinski, E. B., M. A. Freitas, X. Zang, P. G. Hatcher, K. B. Green-Church, and R. B. Jones. 2002. The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. Organic Geochemistry **33**: 171–180.

Liu, Z., and J. B. Phillips. 1991. Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface. Journal of Chromatographic Science **31**: 199–202.

Lu, H., W. B. Dunn, H. Shen, D. B. Kell, and Y. Liang. 2008. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. TrAC Trends in Analytical Chemistry **27**: 215–227.

Meyers-Schulte, K., and J. I. Hedges. 1986. Molecular evidence for a terrestrial component of organic matter dissolved in ocean water. Nature **321**: 61–63.

Novak, J. 1988. Quantitative analysis by gas chromatography, 2nd ed. Marcel Dekker, New York, NY.

Otto, A., and M. J. Simpson. 2006. Evaluation of CuO oxidation parameters for determining the source and stage of lignin degradation in soil. Biogeochemistry **80**: 121–142.

Pearl, I. A. 1942. Vanillin from lignin materials. Journal of the American Chemical Society **64**: 1429–1431.

Schulten, H.-R. 1999. Analytical pyrolysis and computational chemistry of aquatic humic substances and dissolved organic matter. Journal of Analytical and Applied Pyrolysis **49**: 385–415.

Sleighter, R. L., and P. G. Hatcher. 2007. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. Journal of Mass Spectrometry **42**: 559–574.

Ventura, G. T., F. Kenig, C. M. Reddy, G. S. Frysinger, R. K. Nelson, B. Van Mooy, and R. B. Gaines. 2008. Analysis of unresolved complex mixtures of hydrocarbons extracted from Late Archean sediments by comprehensive two-dimensional gas chromatography (GC×GC). Organic Geochemistry **39**: 846–867.

Wardlaw, G. D., J. S. Arey, C. M. Reddy, R. K. Nelson, G. T. Ventura, and D. L. Valentine. 2008. Disentangling oil weathering at a marine seep using GC×GC: Broad metabolic specificity accompanies subsurface petroleum biodegradation. Environmental Science & Technology **42**: 7166–7173.

Williams, P. M., and E. R. M. Druffel. 1987. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. Nature **330**: 246–248.

CHAPTER III

Dissolved organic matter (DOM) composition and molecular diversity in a snowmelt-affected system as elucidated by comprehensive two-dimensional time-of-flight mass spectrometry (GC×GC–TOF–MS) and proton nuclear magnetic resonance spectroscopy (^{1}H-NMR)

3.1 Abstract

The carbon content of Earth's dissolved organic matter (DOM) reservoir approaches that of atmospheric CO_2. Its chemical complexity rivals that of kerogen and its complete structural and molecular-level elucidation remain similarly elusive. To more fully elucidate the molecular constitution of this chemically heterogeneous carbon pool, comprehensive two-dimensional GC×GC time-of-flight mass-spectrometry (GC×GC-TOF-MS) was developed and implemented as a technique to investigate molecular-level changes across a suite of DOM samples. The samples were isolated by solid phase extraction (SPE) on DAX-8 resin from several environmentally contiguous, high altitude, sub-alpine aquatic settings in the vicinity of Lake Tahoe, California-Nevada, USA.

All SPE extracts were CuO-oxidized prior to analysis as trimethylsilyl derivatives. The combination of dissimilar, but orthogonal axes of chromatographic separation, cryo-modulation and cryo-focusing of peaks with LN_2, and an ultra-fast TOF detector, led to a level of *chromatographic* resolution that approached the *mass* resolution of Fourier Transform Ion-Cycle Resonance Mass Spectrometers (FT-ICR-MS) with respect to the number of individually resolvable peaks. By this method, 6,250–9,707 peaks were resolved per sample at a S/N > 10, affording a unique DOM fingerprint for each sample. Each peak was defined by a unique 2-D retention time that was reflective of its volatility (first dimension) and surface polarity (second dimension) as well as by its electron-impact (EI) mass spectra, together enabling structural interpretation of individual peaks. By this method, peaks were identified at specified intensities and partitioned into physicochemically defined regions reflective of aggregate chemical characteristics, such

as their proportion of –OH, –COOH, –CHO –Aromatic, =O, or –$(CH_2)_n$– moieties. Quantitative means were then devised, tested, and implemented to compare these chemically fractionated peak populations to one another on a pair-wise basis across samples. Dissimilarity matrices were generated, from which similarity relationships between samples were derived based on the homology of their compared peak populations. The derived similarity relationships were then validated by comparison to other measured parameters, including ^{1}H-nuclear magnetic resonance spectroscopy (^{1}H-NMR) of the CuO hydrolysate, as well as to other environmental and bulk-level elemental (C/N), isotopic (δ^{13}C, δ^{15}N), and compound-specific organic geochemical data (lignin, fatty acids). The complementarity of these data sets yielded a rich, molecular-level context for the interpretation of bulk-level changes in DOM composition while providing a detailed molecular view into the spatiotemporal dynamics of DOM composition in snowmelt-affected oligotrophic lakes.

3.2 Introduction

3.2.1 Dissolved organic matter (DOM)

The global reservoir of dissolved organic carbon (DOC) occupies a unique place within the planet's carbon cycle. Its size (670 Gt C) (Hansell and Carlson 1998) is comparable to the atmospheric inventory of carbon dioxide (760 Gt C) as well as to that of the planet's living terrestrial biomass (~800 Gt C) (Falkowski et al. 2000), yet the reasons underlying its thermodynamically unstable persistence, from centuries to millennia (Goldberg et al.; Druffel et al. 1992), in the oxygenated water columns of

Earth's lakes and oceans remain elusive (Jiao et al. 2010). In part, full deduction of the dynamics of DOM has failed due to its vast chemical complexity, which rivals that of kerogen (Hertkorn et al. 2007). As such, studies of DOM's molecular-level composition are limited to analyses of what few components can be individually identified and quantified following their oxidative hydrolysis from DOM. In DOM, these components exist within chemically heterogeneous oligomers assembled from constituents that are ultimately derived from the genetically encoded biomolecules of living organisms that may have undergone varying extents of degradation subsequent to their synthesis. In the deep ocean, for instance, less than 3% of DOC can be attributable to individually characterizable components, such as hydrolyzable amino acids, amino sugars, neutral sugars, and lignin, a uniquely terrestrial biomarker (Benner et al. 1992; Hernes and Benner 2002). Because such quantification comes at considerable analytical effort, the preferred mode of DOM analyses is at the bulk level by means of elemental (C/N), isotopic ($\delta^{13}C$, $\delta^{15}N$), spectrophotometric (chromophoric DOM), excitation-emission matrix spectroscopy (EEMS), and occasionally ^{1}H or ^{13}C nuclear magnetic resonance (NMR) spectroscopy. While attractive in their straightforwardness, bulk-level analyses occlude the molecular and structural diversity of DOM, and preclude analysis of the underlying molecular drivers of this variability. Recently, growing appreciation of these limitations, combined with coincident advances in analytical instrumentation, have led to the wide adoption of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) (Koch et al. 2005; Sleighter and Hatcher 2008; Flerus et al. 2012) to analyze DOM composition and diagenetic alternation on a molecular basis. This technique obtains highly resolved (ppm) exact masses representing unique combinations

of C, H, and O, and other heteroatoms, from which thousands of molecular formulae can be aligned, and their abundances compared and interpreted across samples (Stenson et al. 2002). This technique's principal limitations, however, derive from its selective analysis of only ionizable compounds (usually by ESI) and from its inability to differentiate molecular structures with identical masses. Another technique that has seen wider adoption in the last decade is comprehensive GC×GC time-of-flight mass spectrometry (GC×GC-TOF-MS). The present study represents the first attempt to reveal aspects of DOM's structural and compositional diversity by this method.

GC×GC-TOF-MS affords a level of *chromatographic* resolution commensurate to the level of *mass* resolution afforded by FT-ICR-MS. Whereas FT-ICR-MS resolves individual compounds by formulae, GC×GC-TOF-MS resolves individual compounds by structure with the added advantage of yielding an EI mass spectra for each compound. The similar resolving powers of both types of instrumentation lead to the identification of thousands of peaks per sample. As such, both modes of analysis yield vast datasets whose successful distillation into geochemically and environmentally relevant interpretations ultimately underlies the utility of any method. Here, the utility of GC×GC-TOF-MS is used to disentangle molecular drivers of bulk chemical variability for a set of DOM samples isolated from snowmelt-affected aquatic bodies in the vicinity of Lake Tahoe, CA (Fig. 3.1). Sample sites included Lake Tahoe, Fallen Leaf Lake (FLL), Upper Angora Lake (UAL), and the Upper Truckee River (UTR) (Fig. 3.S1). These three lakes are all oligotrophic and all vary significantly in size, maximum depth, volume, drainage basin, and hydraulic residence time (Table 3.S1). The samples were obtained by solid phase extraction (SPE) onto DAX-8 resin during 2009–2010. Samples from this set were used

previously (Goldberg et al.) to show the accumulation of a ^{14}C-depleted proteinaceous background component within Lake Tahoe DOM. The SPE-DOM sample set spans a continuum of $\delta^{13}C$, $\delta^{15}N$, and C/N (Fig. 3.2) values whose compositions appeared to be predominantly driven by seasonal allochthonous inputs from terrestrially derived runoff (Fig. 3.3) and autochthonous inputs derived from *in-situ* aquatic production.

Lakes are globally important biogeochemical and metabolic hotspots in terrestrial ecosystems where extensive degradation of terrestrial organic matter occurs. This is particularly true for Lake Tahoe, which, despite being surrounded by extensive coniferous forests and terrestrial organic matter, has lignin concentrations (Σ_8) that are of similar magnitude (only 3× higher) to open-ocean values measured at Station ALOHA in the central Pacific (Hernes and Benner 2002), which is hundreds of miles removed from potential lignin sources.

3.2.2 Analysis of DOM's chemical complexity using GC×GC-TOF-MS

Thousands of peaks are identified by GC×GC-TOF-MS, but only a very small subset of these peaks has been directly tied to actual structures. Here, peaks are assigned to structures on the basis of (1) the similarity of their mass spectra and (2) retention time indices relative to those obtained for standards or published elsewhere in the literature or in NIST libraries, and on the basis of (3) diagnostic two-dimensional retention time-shifts that reveal either the continuation of a homologous compound series or the addition/removal of chemical functional groups from an identified core structure (Jean-Marie 2004). Beyond those compounds whose structures can be definitively assigned, however, are thousands of peaks which currently resist a structural assignment, but whose

presence underpins DOM's myriad chemical richness. These thousands of unknown peaks paint a unique chemical landscape whose finer details are more fully revealed when the chemical structures of individual peaks are determined and their occurrence frequencies across contrasting environmental samples are examined.

Distilling meaningful information from these thousands of peaks per sample, and from tens to hundreds of thousands of peaks across sample sets, is prohibitively time-consuming using manual methods. Therefore much work has been directed towards developing high-throughput programmatic approaches capable of quickly processing these data in interpretatively meaningful ways. For example, so as to quickly identify compounds of possible diagnostic significance by rapidly screening individual peak abundances across samples. For all algorithmic methodologies, however, the first problem that must be overcome involves that of peak alignment across samples. Small variations in chromatographic conditions, as well as larger systemic variations deriving from reduced column performance, as well as differences in the sample matrix, can cause the retention time of individual peaks to shift slightly over successive chromatographic runs, complicating alignment of peaks across samples. For GC×GC data, peak alignment is complicated by the sheer number of peaks and by the existence of two retention times for each peak. As such, algorithmic techniques have been adapted to this task. The byte size of the raw data files (100s of MB to >2GB) results in long data processing times, while the dual dimensionality of the GC×GC data is not immediately amenable to traditional retention-time alignment algorithms designed for alignment of 1-D data. Reducing the dataset's dimensionality into a single dimension by linearization is therefore a necessary first step required to access the capabilities of 1-D alignment algorithms.

Software capable of performing this alignment includes MetAlign (Lommen 2009; Lommen et al. 2012), MZmine (Katajamaa et al. 2006), and TagFinder (Luedemann et al. 2008), as well as the commercially available Statistical Compare made by LECO Corp. (St. Joseph, MI) as a supplementary component to its proprietary ChromaTOF software. A different approach is taken by GC Image, another GC×GC alignment and data processing software package. This software avoids data linearization completely and instead relies on direct processing of 2-D data using image processing techniques (Reichenbach et al. 2004).

The strengths, weaknesses, and limitations of these different alignment algorithms have been reported in some depth elsewhere (Koh et al. 2010). All alignment methodologies have subtle advantages and disadvantages, however no methodology is inherently superior; all fail to align peaks correctly 100 % of the time. This can lead to the incorrect identification of peaks (false positives and negatives) of potential interest, and emphasizes the need for manual verification of a peak's correct alignment and identification when such programmatic tools are used to rapidly screen sample sets for peaks of possible diagnostic significance. For applications of non-directed analysis, Statistical Compare has been successfully used to rapidly screen and identify volatile organic compounds diagnostic of advanced tissue decomposition (Stadler et al. 2012), metabolite markers associated with increased *M. tuberculosis* virulence (Meissner-Roloff et al. 2012), and up-regulated cellular metabolites in a strain of micro algae, *Chlamydomonas reinhardtii*, subjected to sulfur deprivation-induced H_2 production (Doebbe et al. 2010).

Here, GC×GC-TOF-MS is applied as a novel tool to access the rich molecular diversity of CuO-hydrolyzed SPE-DOM. The analyses of these hydrolysate mixtures generate complex and detail-rich 2-D chromatograms which represent unique organic geochemical fingerprints of these samples. The thousands of peaks identified per chromatogram have unique 1- and 2-dimension retention times (^{12}RT) which are shown to be reflective of these peaks' physicochemical properties in terms of describing aspects of the size, functional group constitution, and molecular polarity of the largely unknown compounds these peaks represent. The alignment capabilities of Statistical Compare are then evaluated and employed as a tool with which the homology of physicochemically-defined populations of chromatographic peaks are examined and quantified across samples. This enables unique relationships of similarity to be drawn for chemically defined peak populations across samples and provides a new and complementary view into the molecular changes that accompany the commonly observed bulk-level transformations in DOM composition.

3.3 Methods

3.3. 1 Isolation of SPE-DOM with DAX-8 resin

Large volume (200–1300 L) whole freshwater samples from each aquatic setting were collected in carboys and transported back to the lab where they were immediately filtered through 0.2 μm filters using acid-washed (10 % HCl) Pt-cured silicone tubing (Cole-Parmer) and into an acid-washed, Teflon-lined 55 gallon barrel using a peristaltic pump. Inside the barrel, filtered water was acidified to pH ≈ 1.5–2 using 12.1 *N* HCl.

Acidified lake water was extracted onto 1 kg of macroporous, hydrophobic DAX-8 resin at a flow rate of 10 L hr^{-1}. The DAX-8 resin was thoroughly cleaned following the procedure described by (Aiken et al. 1992). Briefly, the resin was Soxhlet extracted for ≥ 50 cycles with methanol followed by acetonitrile. After the resin was poured into the column, it was again cleaned with 20 L of 0.1 *N* NaOH, followed by 1-2 L of milliQ to neutralize the pH, followed by 20 L of 0.1 *N* HCl. The resin-filled column was then stored in 0.1 *N* HCl prior to DOM extraction. Organic carbon recoveries were monitored for each sample and calculated following: $(1\text{-}DOC_{effluent}/DOC_{influent}) \times 100$ % . DOC recoveries ranged from 6-50 %. All DOC concentrations were measured on a Shimadzu TOC-V CSN analyzer using 7-point standard calibration curves ($R^2 \geq 0.99$). Columns were eluted in 3:1 acetonitrile:water (milliQ); organic matter (OM) residues/powders were produced through rotary evaporation of acetonitrile and sublimation of the remaining water. Lyophilized organic matter was analyzed for % TOC, % TON, C/N, $\delta^{13}C$, and $\delta^{15}N$ following established techniques at the Scripps Institution of Oceanography's Analytical Facility.

3.3.2 Alkaline CuO Oxidation

SPE-DOM samples were oxidized by alkaline CuO oxidation (Hedges and Ertel 1982) adapted (Goñi and Montgomery 2000) for microwave (CEM Corp.) hydrolysis using perfluoroalkoxy (PFA) hydrolysis vessels. Briefly, SPE-DOM equivalent to 5 mg C was combined under an anaerobic atmosphere (N_2) with 500 mg CuO, 50 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ and 20 mL 2 *N* NaOH and reacted at 150 °C for 90 min. Following oxidation, the vessels were cooled to room temperature, the hydrolysate was

spiked with dual internal recovery standards (*trans*-cinnamic acid and ethyl vanillin), acidified to pH ≈ 2 with 12.1 *N* HCl, extracted into ethyl acetate (3×), dried over Na_2SO_4, and gently blown to dryness under N_2 before immediately redissolving the hydrolysate residue in 400 μL of pyridine. The redissolved hydrolysate residue, pyridine, and N,O-BSTFA + 1% TMCS were combined 1:1:2 into 400 μL GC insert vials and heated at 60 °C for 15 min to swap exchangeable hydrogens in the hydrolysate with trimethylsilyl (TMS) groups, making these compounds less polar and more volatile and therefore more amenable to gas-phase separation. The lignin phenols are a small suite of 11 phenols that are often liberated when natural organic matter mixtures (NOM) containing lignin is CuO-oxidized and their total volumetric concentration (Σ_8) is determined as the sum of vanillyl- (V), syringyl- (S), and cinnamyl-type (C) phenols.

3.3.3 Nuclear Magnetic Resonance (NMR) Spectroscopy

One dimensional ^1H NMR spectra were obtained on an Avance III Bruker 500 MHz spectrometer using a three channel 1.7 mm liquid state probe. 1024 scans and 8 dummy scans were used for acquisition and 65k points were transformed from the FID using a zero fill factor of 2. The spectra were obtained with a spectral width of 10 kHz. d_5-pyridine was selected as an NMR solvent because of its ability to solvate complex, chemically diverse mixtures (Demarco et al. 1968) such as the CuO hydrolysate, and especially its acidic components. Spectra were baseline corrected, and ppm shift ranges corresponding to the resonance shifts of specific hydrogen functional groups were integrated and sum normalized. Absorbances from residual H_2O and pyridine's chemical shifts at 7.22, 7.58, and 8.7 ppm were omitted from integration. Small integrals were

most sensitive to changes in baseline, especially the acids and the aldehydes, and particular care was taken to integrate these shift regions both accurately and consistently across spectra. Sum-normalized integrals were then analyzed by principal component analysis (PCA) in Matlab following 0–to–1 normalization. The integration domain for individual proton chemical environments was selected to provide a first-order quantification of protons specific to 'alkyl' (δ 0.8–1.4 ppm), 'aliphatic' (δ 1.5–1.8 ppm), 'benzylic' (δ 1.8–2.4 + 2.7–3.2 ppm), 'acetophenone' (δ 2.4–2.7 ppm), 'alcohol + saccharides + amines' (δ 3.4–3.6, 3.9–4.8 ppm), 'aromatics + vinylics' (δ 5.2–8.3 ppm), 'highly substituted and heteroaromatics' (δ 8.3–9.0 ppm), 'aldehydes' (δ 9.5–10.5 ppm), and 'carboxylic acids' (δ 10.5–13.5 ppm).

3.3.4 Fatty acid analysis

Fatty acids were extracted from lyophilized SPE-DOM following the procedure of (Sun et al. 1997) with minor alternations to the GC oven program. Samples were weighed (~10 mg) into combusted 20 mL borosilicate scintillation vials and re-suspended in 2 mL of 0.5 N KOH in 100% methanol. Nonadecanoic acid ($C_{19:0}$; Sigma Aldrich) was added as an internal standard before saponification and was used for quantification of fatty acids. Individual samples were N_2 flushed, sealed in an ampoule, saponified for 120 min at 110 °C, and cooled. Neutral lipids were then extracted 3× using hexane. The saponified extract was acidified to pH = 1 with 6 *N* HCl and protonated fatty acids were extracted 3× into hexane. The hexane was dried to completion under N_2, methylated using BF_3/CH_3OH (12 % CH_3OH; Sigma-Aldrich), and transferred to hexane prior to GC analysis. All samples were analyzed using an Agilent 7890A GC and a 5975C triple axis

quadrupole EI-MSD using a 60 m HP-5MS column. Column effluent was split evenly between FID and MSD. The oven program entailed an initial hold of 0.5 min at 50 ºC, ramps of 10 ºC min^{-1} to 275 ºC and then 5 ºC min^{-1} to 320 ºC and then a final 8 min hold. Fatty acid peaks were identified by retention time alignment and comparison of mass spectra to standards (37n FAME mix; Sigma-Aldrich), to NIST MS library data, and to previously published spectra. The following 12 fatty acids were quantified: $C_{12:0}$, $C_{14:0}$, $C_{15:0}$, $C_{16:1}$, $C_{16:0}$, $C_{18:1a}$, $C_{18:1b}$, $C_{18:0}$, $C_{20:0}$, $C_{22:0}$, $C_{23:0}$, $C_{24:0}$.

3.3.5 GC×GC time-of-flight mass spectrometry (GC×GC-TOF-MS)

3.3.5.1 Gas Chromatography

The GC×GC-TOF-MS system was composed of an Agilent 7890A gas chromatograph operated with a 7893A Automated Liquid Sampler (ALS) and was coupled to a LECO Pegasus 4D time-of-flight (TOF) mass spectrometer. The primary column (1°) was a 10.0 m, 0.18 mm i.d., 0.2 μm df RTX-5MS column and the secondary column (2°) was a 1.20 m, 0.10 mm i.d., 0.1 μm df RXI-17 column. The columns were joined in the main oven using a Siltek-coated Vu2 Union® (Restek). The Agilent multimode inlet was operated at 300 ˚C with 100 ml min^{-1} purge flow for 60 s following injection and contained a Restek gooseneck splitless deactivated glass liner with deactivated glass wool. The column was operated at a constant flow of 0.8 ml min^{-1} using He as the carrier gas. Both primary (100–300 ˚C) and secondary (105–305 ˚C) ovens had 0.20 min initial hold times and were programmed to rise at 6.0 ˚C min^{-1} until holding at their final temperatures for 1.0 min. The method runtime was only 34.5 minutes. Based

on the results of extensive method optimization that sought to maximize the spread of eluting peaks across the two-dimensional chromatogram, the modulation period selected was 6.0 seconds with a hot pulse time of 1.32 s and a cool time between stages of 1.68 s. The transfer line temperature was set to 280 ˚C and the modulator offset was always maintained 20 ˚C above that of the primary oven to ensure complete volatilization of compounds following cryogenic focusing with LN_2 during modulation. Injection volume was 1 μL. Larger injections were used for the purposes of obtaining improved spectral quality of low abundance peaks. All samples destined for intercomparison by non-targeted alignment analysis were analyzed sequentially in < 18 hours to minimize nonparameterizable variability in the S/N value of detected peaks that has been observed to occur on a day-to-day and week-to-month basis, both in our lab and elsewhere (Kiefl et al. 2012). Constraints on the variability expected for the ~18 hour analysis period were obtained by examination of alignment scores of triplicate injections, as discussed later.

3.3.5.2 TOF-MS

The acquisition, or solvent, delay was 4.0 min, the mode of ionization was electron impact (EI) at 70 eV, the source temperature was 250 ˚C, the acquisition rate and m/z range were 100 spectra $second^{-1}$ and 50-800, respectively. The detector voltage was operated at 1600 V. System optimization, source focusing, ion optic focusing, mass calibrations, tune, and leak checks were all performed immediately prior to sample analysis. All samples were analyzed within 24 hours. File sizes under these acquisition parameters were 263 MB per sample. A doubling of the acquisition rate and runtime

would have resulted in a quadrupling of file size, resulting in unwieldy files larger than 1 GB each.

3.3.6 Data Processing

Data processing was conducted within ChromaTOF using a baseline offset of 0.7, 3 point data smoothing, an expected 1° column peak width of 18 s until 15 min, and 24 s thereafter, and an expected 2° column peak width of 0.15 s. A similarity score of 700 was required to combine peaks that may have been cut during modulation and a signal-to-noise (S/N) of 6 was defined as the limit for sub-peak detection.

3.3.7 Identification of Peaks

Whole chromatograms or individual chromatographic quadrants (Fig. 3.4) were processed to identify peaks at S/N ratios greater than either 10 ($S/N_{\geq 10}$) or 100 ($S/N_{\geq 100}$). More peaks are identified at lower S/N ratios. Peaks are automatically identified within LECO Corp's (St. Joseph, MI) ChromaTOF above a defined minimum S/N threshold by the identification of all unique apexing masses at *discrete* second dimension retention times (^{2}RTs), which are highly consistent across chromatograms. Typical 1σ variability in ^{2}RT was observed to be 32 ms for a set of 100 compounds whose ^{2}RTs were examined across multiple chromatograms. Identification of the ions that rise and fall coincidentally with the unique ion allows the EI mass spectra for the individual peak to be determined. Once a peak is identified by a unique apexing ion, its contribution to the total ion chromatogram is removed for the width of the peak; meanwhile, the other m/z ions that rise and fall coincidentally with the unique ion are also subtracted from the TIC in

proportion to their abundance relative to the unique ion. The process is repeated for all unique ions found down to a specified S/N ratio. If multiple unique ions are present, the one with the greatest S/N is selected. As such, a complete deconvolution of the TIC signal is effected whereby co-eluting compounds can be discretely and consistently detected and their full EI mass spectra extracted from the TIC. This deconvolution is similar to what is ultimately achieved using the NIST-available Automated Mass Spectral Deconvolution and Identification System (AMDIS) (Stein 1999), however the deconvolution algorithm implemented in ChromaTOF greatly benefits from the rapid full spectra (m/z = 1-1000) acquisition rates inherent to TOF detectors, which reach speeds of up to 500 spectra s^{-1}. These full spectral acquisition rates are much higher than those of more common quadrupole detectors, which achieve full scan (m/z 1-1000) speeds of only 2-3 s^{-1}. Fast acquisition speeds are needed to resolve the narrow peak widths created during GC×GC modulation. For a normal chromatographic peak to be detected it must be sampled 10 times across its width (Novak 1988) while for a deconvolution to be robust, a peak must be sampled 18-20 times across its width. The absence of spectral skewing in TOF spectra compared to quadrupole data (there is no scanning in TOF) further assists with robust deconvolutions by better affording the symmetrical, Gaussian-shaped peaks for which standard deconvolutions are most ideally suited. The effectiveness of ChromaTOF's deconvolution algorithm has been evaluated elsewhere and compared to other available mass spectral deconvolution packages with respect to the quality of the deconvoluted mass spectra it produces and the rates of false positives and negatives it encounters (Lu et al. 2008). Collectively, the augmented chromatographic and mass acquisition advantages of comprehensive GC×GC-TOF-MS facilitate the identification of

1–2 orders of magnitude more compounds compared to 1-D methods (Table 3.1); moreover, compounds are better resolved and are detectable across an expanded dynamic range.

In these chromatograms, 299 compounds were identified on the basis of standards, published relative retention time (RRT) information, and mass spectral data obtained from the main and replicate 2011 NIST MS library and from elsewhere in the published literature (Goñi 1992; Goñi and Hedges 1992, 1995; Dickens et al. 2007).

3.3.8 Pair-wise alignment of peaks across samples

The orthogonal nature of GC×GC separations is such that the total peak capacity is the product of each column's respective peak capacity. Peak capacity is then increased even further by the narrowing of peak widths during cryo-modulation. Finally, the rapid acquisition speeds (500 spectra s^{-1}), the large dynamic range, and the robust spectral deconvolutions that are enabled by TOF detectors collectively lead to the unveiling of thousands of peaks per sample where only 100–300 could be previously resolved using 1-D methods (Table 3.1), representing an increase of 1–2 orders of magnitude. These large amounts of data, however, quickly create a bottleneck with data processing. The vast majority (> ~95 %) of the peaks cannot be identified definitively on the basis of commercially available standards or by matches to NIST EI-MS spectral libraries, but these peaks nevertheless represent unique compounds that collectively comprise a geochemical fingerprint for each DOM sample. The presence/absence and relative abundance of these unknown peaks across samples can assist with more fully unveiling

the myriad molecular drivers that underlie the bulk-level changes in DOM that are commonly measured, such as % TOC, %TON, C/N, $\delta^{13}C$, or $\delta^{15}N$.

Peaks were aligned across samples using Statistical Compare, a powerful add-on feature to LECO Corp.'s ChromaTOF. ChromaTOF both drives the GC×GC-TOF-MS instrumentation and provides a workspace environment for data processing and analysis. Within Statistical Compare, peaks from individual samples were first grouped into classes and then peak populations of different intensities were aligned. Peak alignments across samples were performed on the basis of (1) retention time, or (2) a combination of both retention time and spectral similarity of aligned peaks. Fewer aligned peaks were found when a minimum similarity match was used. This was mainly attributable to the greater variability of the deconvoluted mass spectra (PeakTrues) for low abundance peaks which, in turn, led to reduced similarity scores when the mass spectra for two identical, but low abundance peaks were compared. Analysis of this sample set showed that peak alignment on the basis of retention time alone was more likely to accurately gauge the number of shared peaks between two samples while the addition of a minimum spectral similarity threshold to the retention time alignment criteria was likely to underestimate the number of peaks shared between two samples. For example, when peaks derived from triplicate injections of the same sample were aligned by retention time alone, 88.73 ± 0.36 % ($S/N_{\geq 10}$, 1σ) and 87.29 ± 0.76 % ($S/N_{\geq 100}$, 1σ) of the peaks were successfully aligned. Meanwhile, when the stringency of the alignment criteria was increased and a minimum spectral similarity score of 700 was added, only 21.40 ± 0.49 % ($S/N_{\geq 10}$) and 65.41 ± 1.12 % ($S/N_{\geq 100}$) of the peaks were aligned. This showed that adding a minimum mass spectral similarity score of 700 caused the number of aligned

peaks to drop by a factor of 4.1 and 1.3 for peaks identified at $S/N_{\geq 10}$ and $S/N_{\geq 100}$, respectively. These data showed that low abundance peaks are erroneously discarded at higher rates than high abundance peaks during alignment when a mass spectra similarity threshold was applied. This arose because lower S/N peaks had more variable PeakTrue mass spectra than higher S/N peaks, which are of comparatively higher quality. Because adding a mass spectrum criterion artefactually lowers the actual number of shared peaks, especially for low S/N peaks, the results of the retention time alignments will be of greater subsequent focus.

The alignment criteria used for all alignments specified a retention time alignment window of ± 1 modulation period (6 s) for ^{1}RT and ± 0.15 s for ^{2}RT. The window for ^{1}RT was chosen as it was the minimum non-zero value possible that afforded some flexibility in ^{1}RT. Meanwhile, the ^{2}RT window of 0.15 s was chosen following manual inspection of the observed ^{2}RT shifts (Fig. 3.5B) for one hundred known compounds identified across two chromatograms. The standard deviation (1σ) of these observed shifts was 0.065 s. Therefore, to encompass 98 % of the anticipated variability in ^{2}RT across all peaks, a ^{2}RT interval of 2.32 σ was needed and so a ^{2}RT of 2.32 $\sigma \times 0.065$ s/σ = 0.15 s was selected.

3.3.9 Determination of intersection set size for aligned peaks

Quadrant-partitioned chromatograms or whole chromatograms (all quadrants) were processed to identify peaks with S/N values greater than or equal to (or between) 10 or 100. Each sample was then aligned manually on a *pair-wise* basis to all other samples and the set of peaks {X} intersecting ($\cap$) the two sample sets, notated by {A} and {B},

were determined as shown in Fig. 3.6 where {X} = {A} ∩ {B}. The number of peaks in the intersection set, n{X}, was normalized to the sum of peaks in both sets {A} and {B}, or n{A} + n{B}. Because {X} is a subset (⊆) of both set {A} and set {B}, the number of intersecting peaks can be expressed as a normalized percentage,

$$s\ (\%) = \frac{2 \cdot n(\{A\} \cap \{B\}) \cdot 100}{n\{A\} + n\{B\}} = \frac{2 \cdot n\{X\} \cdot 100}{n\{A\} + n\{B\}}, \qquad \text{(Equation 1)}$$

which is analogous to a calculated similarity score between samples. A matrix, [S], of pair-wise similarity scores, *s*, was thus created when this pair-wise alignment was performed across all samples that were processed with identical data processing parameters. This methodology, devised independently here, is equivalent to constructing a matrix of Sørensen–Dice coefficients (Sørensen 1948), a metric originally conceived to determine similarity coefficients from presence/absence data.

3.3.10 Errors in computed similarity scores (σ_{CV})

To evaluate the significance of differences in similarity scores obtained from pair-wise sample alignments across the SPE-DOM sample set examined, the sensitivity of the computed similarity score to (1) implementation of the alignment algorithm, (2) replicate oxidations, and (3) replicate injections was examined and quantified, the results of which are summarized individually below and in Table 3.2.

3.3.10.1 Non-commutative alignment algorithm (σ_{AA})

A different number of aligned peaks were sometimes found depending on the order by which the samples were compared within Statistical Compare. For instance, if {A} was aligned with {B}**,** n{X} aligned peaks were found whereas if {B} was aligned with {A}**,** n{Y} aligned peaks were found. Alternately stated, under any particular alignment criteria where {X} = {A} ∩ {B} and {Y} = {B} ∩ {A}, n{X} ≅ n{Y}. This non-commutative feature of the alignment algorithm was initially disconcerting. Inspection of similarity score variability for pair-wise comparisons executed in forward and reverse order revealed that while it was *sometimes* the case that n{X} = n{Y}, it was *often* the case that n{X} ≠ n{Y}. However, when n{X} ≠ n{Y}, they differed from one another by only a very small amount (± 0.098 %, 1σ) (Fig. 3.5A).

3.3.10.2 Oxidation replicates (σ_{OR})

Identical, replicate oxidations of individual SPE-DOM samples were examined to parameterize the variability in similarity scores that could derive from the CuO-oxidation of identical samples. The oxidation replicates were analyzed by GC×GC-TOF-MS, the retention times of the resultant peaks were aligned pair-wise to one another, and similarity scores were computed. The mean score for these replicate oxidations was 89.17 ± 0.51 % (1σ, n=6). As such, the 1σ variability in similarity scores that arose from replicate oxidations was ~5× greater than the variability introduced by the non-commutativeness of the alignment algorithm.

3.3.10.3 Replicate Injections (σ_{RI})

The variability in similarity scores derived from replicate injections was also examined. For peaks at $S/N_{\geq 100}$ the mean similarity score was 87.29 ± 0.73 % (1σ) while for peaks at $S/N_{\geq 10}$ it was 88.73 ± 0.36 % (1σ). As can quickly be seen, 100 % alignment is never completely achieved, even for replicate injections, principally because the number of identified peaks from replicate oxidations or replicate injections can vary. 100% alignment requires, at a minimum, that n{A} = n{B}, but this was never observed (Table 3.S2).

3.3.11 Peak counting in partitioned and non-partitioned chromatograms

When a chromatogram was processed *by quadrant* at a S/N of either 10 or 100 and peaks were aligned across samples *by quadrant*, the total number of peaks across all quadrants (as well as the total number of aligned peaks) was greater than the number of peaks that would have been found if the sample been processed as a single chromatogram. This is because peaks that are identical on the basis of spectral similarity can be automatically combined if they elute over multiple modulations. This peak combining, however, cannot occur *across* quadrants if individual chromatographic quadrants are processed separately. Therefore, a greater number of peaks were identified *per chromatogram* when it was processed by quadrant than when it was processed whole (Table 3.S1). However, since the quadrant partitions are identically for all samples, this discrepancy will not adversely affect the fidelity of the computed alignment scores.

3.3.12 Examining molecular similarity using hierarchical clustering

Similar to how large datasets of peptide and nucleotide sequences can be rapidly compared to one another after they are reduced to a distance matrix that reflects pair-wise distances between sequences, the thousands of peaks obtained by GC×GC-TOF-MS can be similarly reduced and interpreted following alignment, and hierarchical relationships between populations of peaks can be constructed. As such, distance, or dissimilarity matrices (1 - [S]) were computed from matrices of similarity scores. This mode of analysis enabled the *quantitative* comparison of one chromatogram, or 2-D chromatographic quadrant, to another. Absent this methodology, comparison of chromatograms could only otherwise be done on a qualitative and cumbersome visual basis.

Dissimilarity matrices were then subjected to agglomerative hierarchical clustering in Matlab using the Euclidean distance metric and the unweighted average distance with arithmetic mean (UPGMA) linkage method to compute distances between clusters. In the UPGMA linkage method, clustering proceeds step-wise as the nearest clusters are joined before higher-level clusters are joined. Following each clustering iteration, distances to other nodes are determined from the average of the previous cluster's component nodes. The method was described by Sokal and Michener in 1958 as, "a method for grouping species, and regrouping the resulting assemblages, to form a classificatory hierarchy most easily expressed as a treelike diagram of relationships" (Sokal and Michener 1958). These resulting linkage data, derived from tens of thousands of peaks across samples, were then used to construct dendrograms that enabled the

comprehensive profiling of conserved peak homology across samples which, in turn, led to the clustering of samples by their proportion of shared peaks

Here, the UPGMA linkage method was used to profile the molecular-level similarities and differences across peak populations. Making efforts to represent the data as precisely as possible, other common linkage methods were explored, including the complete (Defays 1977) and single linkage (Sibson 1973) methods. Irrespective of the linkage method used, however, the most distantly related clusters were consistently separated. The choice of linkage method affected the composition of lower distance clusters more significantly than it did the composition of higher-distance clusters. The single linkage method was not chosen as it tended to produce a skewed hierarchical clustering, or a 'chaining' (Murtagh 1983) of nodes, which did not reflect the underlying dissimilarity matrix with as high a fidelity as either the complete or UPGMA linkage methods. Both the complete and UPGMA linkage methods produced improved clustering topologies over those obtained from the single linkage method, however since no averaging is used in the complete linkage method, undue significance can be given to small differences in distances as clustering proceeds. This sensitivity to noise in the complete linkage method was deemed undesirable given the small parameterized errancy of alignment scores (σ_{CV}) and, as such, the UPGMA linkage method remained as the most appropriate. The appropriateness of this choice was confirmed quantitatively by computing the cophenetic correlation coefficients (Sokal and Rohlf 1962) between each dissimilarity matrix and dendrograms constructed from this matrix using each of the three linkage types. Of the three linkage types, the UPGMA linkage method consistently yielded the highest cophenetic scores and the single linkage method yielded the lowest.

Cophenetic correlation coefficients provide a measure of the fidelity with which a given tree topology represents the distance data from which it was derived. Different linkage methods can generate different topologies and some reflect the underlying dissimilarity or distance data more faithfully than do others. As such, an external, quantitative metric (the cophenetic score) was used to evaluate the appropriateness of the UPGMA linkage method and the faithfulness with which it yielded a hierarchical topology that was reflective of the underlying distance data from which it was derived.

This hierarchical clustering and linkage analysis led to distinct patterns of sample clustering and some clusters were more distinctly defined than others. The number of clusters per dendrogram were determined by defining a distance of significance below which node branching points were deemed likely to be insignificant. This distance was determined by identifying the distance at which link heights became inconsistent with the link heights below it. As such, an inconsistency score was calculated for each link (Zahn 1971) as the difference of the inspected link height to the average height of its subordinate three (up to) links, divided by the standard deviation of these subordinate link lengths. Nodes were considered likely to be insignificant if they were defined by inconsistency scores ≤ 1. This was done to assist with consistent delineation of sample groupings.

3.4 Results

3.4.1 CuO-Oxidized SPE-DOM GC×GC Chromatograms

The GC×GC-TOF-MS method parameters were successfully optimized subject to the following constraints: (1) the absence of chromatographic 'wrap-around', (2) the maximization of the ratio of the peak elution area to the chromatographic area, (3) the selection of a spectral acquisition rate that afforded a minimum of 20 sampling points across a 200 ms peak to enable adequate determined peak deconvolutions, (4) the minimization of method runtime to conserve LN_2, and (5) the limitation of file sizes to facilitate data handling and accelerate processing times. The chromatograms acquired using this optimized method are shown, identically scaled, in Fig. 3.7. Darkened blobs correspond to most abundant peaks. Most peaks eluted within the window: $1s < {}^2RT < 5s$. Samples enriched in polar moieties such as aldehydes, ketones, aromatic rings, and methoxy groups produced chromatograms that occupied a larger range of ^{2}RTs than did samples that were depleted in these groups and which were instead enriched in aliphatic carbon, alcohols (–OTMS), and acids ($–CO_2TMS$) and which therefore spanned a narrower ^{2}RT range. More terrestrially influenced SPE-DOM hydrolysates (evidenced by $\downarrow\delta^{13}C$,↑lignin, ↑C/N, Fig. 3.2) had higher proportions of compounds with these polar moieties while SPE-DOM hydrolysates lacking strong terrestrial influences had lower proportions of such compounds and were instead dominated by oxidation products of presumably autochthonous algal material which was enriched in alcohols (–OTMS), acids ($–CO_2TMS$), and aliphatic carbon. Differences in the extent and intensity of the unresolved complex mixture (UCM) (manifested as 'blurriness'), the number and

localization of peaks, and the generalized, visual appearances of each of the chromatograms were apparent on a qualitative, visual basis alone. Visual comparisons are enabled because all derivatized CuO hydrolysates were derived from the same amount of SPE DOM (5 mg C). The areal extent and intensity of the elution space appears greater in FLL and UAL isolates compared to those from Tahoe or UTR. Among Tahoe isolates, blurriness appears greatest for the July sample, for FLL it appears greatest for the June sample, and for UAL it appears greatest for the August sample. Lignin dimers are chromatographically well-resolved (Fig. 3.7K) and are present at visually varying abundances in all samples except for those from Tahoe. As summarized in Table 3.S2, at either $S/N_{\geq 10}$ or $SN_{\geq 100}$, the Lake Tahoe chromatograms (Fig. 3.7A–C) consistently yielded the fewest number of detectable peaks while those from UAL (Fig. 3.7G→I) yielded the greatest number; meanwhile, FLL (Fig. 3.7D→F) and UTR (Fig. 3.7K) yielded intermediate numbers of peaks. The average number of peaks for each of the four locations at different time-points for peaks identified at $S/N_{\geq 10}$ and $S/N_{\geq 100}$, respectively, were 6562 ± 511 and 2083 ± 274 for Tahoe, 8617 ± 105 and 3394 ± 161 for FLL, 9420 ± 356 and 3808 ± 251 for UAL, and 8807 and 3357 for UTR (1σ, n = 3) (Table 3.S2).

3.4.2 Partitioning GC×GC chromatograms into quadrants

Quadrant-specific data processing led to the binning of peaks by chromatographic quadrant. Normalized quadrant bin sizes (B_Q) were determined as the number of peaks in a particular quadrant divided by the total number of peaks identified across all quadrants. A consistent relationship was observed for B_Q across samples, regardless of the S/N of identified peaks, whereby $B_{Q3} > B_{Q1} > B_{Q4} > B_{Q2}$ (Fig. 3.8A→C, Table 3.S2). Across

samples, B_{Q2} varied the most and B_{Q3} varied the least while variability was of similar magnitude for B_{Q1} and B_{Q4} (Fig. 3.8A→C). B_{Q4} exhibited greater variability across samples when processed with a S/N threshold of 100 instead of 10, however the variability of other quadrant bin sizes appeared relatively unaffected by the specified S/N threshold ($S/N_{\geq 10}$, $S/N_{\geq 100}$, or $S/N_{10\text{-}100}$) for peak identification. Across samples, some bin sizes expanded or contracted at the expense or gain of others as the S/N criterion for identifying peaks was changed, reflecting quadrant-specific variability in the S/N dispersion of peaks. This effect was most pronounced in the case of B_{Q3} whose size increased at the expense of B_{Q2} across all samples when low intensity peaks were discarded. A significant fraction of peaks in B_{Q2} were present at a S/N between 10 and 100 (hence B_{Q2} is larger in Fig. 3.8C than A or B) while a greater fraction of the peaks in B_{Q3} are present at a $S/N_{\geq 100}$ (hence B_{Q3} is larger in Fig. 3.8B than in A or C), offering a useful basis for evaluating the S/N dispersion of binned peaks.

3.4.3 Physicochemical selectivity of defined chromatographic partitions

In GC×GC, compounds are separated on the basis of their volatility (boiling point) and polarity. More volatile compounds elute earlier than semi-volatile compounds with higher boiling points. In the current configuration, the primary column (RTX-5MS) separates analytes principally by boiling point while the more polar secondary column (RXI-17) separates compounds by polarity under isothermal conditions. In the second dimension, analytes with greater polar surface area interact more strongly with the electron rich conjugated π-systems of the diphenyl groups and the progress of the analyte through the column is impeded and its retention time is increased. As such, compounds

are separated on both a physical (boiling point) and a chemical basis (polarity). Each compound's two-dimensional elution time, ^{12}RT, is therefore a reflection of both its molecular size (∝ boiling point) and surface polarity (∝ chemical composition). For instance, peaks with high ^{2}RTs contain a greater ratio of polar to non-polar surface area than do peaks with low ^{2}RT's. Peaks are thus sorted, or binned, into regions of 2-D RT space on a psychicochemically defined basis. This binning provides a basis for examining the differential occurrence (presence/absence) of physicochemically defined peak classes across samples.

Following this approach, GC×GC-TOF-MS chromatograms were partitioned into quadrants (Fig. 3.4). Quadrant assignment was informed by a subset of 299 peaks whose chemical structures had already been definitively identified. The partitioning was performed using the Classification feature of ChromaTOF. Single quadrants were processed individually by excluding all other quadrants during data processing. Compounds with polar chemical functionalities such as aldehydes, aromatics, ketones, and methoxy groups eluted with later ^{2}RTs while compounds that were more aliphatic in nature eluted with lower ^{2}RT times. Because all acids and alcohols were derivatized with TMS groups, these compounds consistently eluted with lower ^{2}RTs. As such, Quadrant 1 (Q_1) was defined as 'small-polar', Quadrant 2 (Q_2) as 'large-polar', Quadrant 3 (Q_3) as 'small-apolar', and Quadrant 4 (Q_4) as 'large-apolar'. While innumerable other physicochemically relevant chromatogram partitionings are conceivable – with some perhaps even being narrowly constrained to encompass a particular class of compounds – the described quadrant-based partitioning was chosen because it (1) encompassed all detected peaks, because it (2) yielded a workable number of partitions on which to

perform pair-wise sample alignments and because it (3) did not partition the peaks on an unreasonably narrow basis that relied on preconceptions of their structures that were overly specific. This is why, for instance, peaks in the lignin monomer elution region were not independently aligned: other compounds also eluted in this region. The following is a brief synopsis detailing the types of compounds that were observed to elute in each quadrant based on a set of 299 compounds identified in these quadrants.

3.4.3.1 TMS Groups (Fig. 3.9A):

Q_4 had the greatest number of TMS groups ($\tilde{x} = 3$) per compound while Q_1 had the fewest ($\tilde{x} = 1$) ($p < 0.05$); Q_2 and Q_3 both had similar numbers of TMS groups per compound ($\tilde{x} = 2$) (Fig. 3.9A). TMS groups are large and nonpolar and therefore compounds enriched in –OH and –COOH groups in the DOM hydrolysate become enriched in non-polar TMS groups once derivatized. The ^{2}RTs of these compounds are correspondingly reduced. As such, the quadrants characterized by lower ^{2}RTs, Q_3 and Q_4, were alcohol and acid-enriched relative to their upper adjacent quadrants, Q_1 and Q_2.

3.4.3.2 Carbonyl & Ester & Ether Oxygens (Fig. 3.9B):

Carbonyl, ester, and ether oxygens represent those oxygens that are not derivatized with –TMS. More oxygen lone pair electrons led to retarded elution from the secondary column. Q_2 had significantly ($p < 0.05$) more of these groups than did Q_4. This was consistent with the naming of each quadrant as either polar or nonpolar. Q_2 contained more non-TMS oxygens than did Q_1 because this quadrant was enriched in lignin dimers, which contain 2× the methoxy content of lignin monomers.

3.4.3.3 Molecular Weights (Fig. 3.9C):

Expectedly, compounds that eluted in Q_2 and Q_4 had significantly ($p < 0.05$) higher molecular weights than did compounds that eluted in either Q_1 or Q_3. Compounds that eluted in Q_2 and Q_4 were thus termed 'large' while those that eluted in Q_1 and Q_3 were termed 'small'. No significant differences in the molecular weights of compounds were apparent for vertically adjacent quadrants.

3.4.3.4 Atomic C/H ratios (Fig. 3.9D):

Atomic C/H ratios were computed for derivatized compounds. Compounds in Q_3 and Q_4 had lower C/H ratios ($p < 0.05$) than the compounds in their vertically adjacent counterparts, Q_1 and Q_2, respectively.

3.4.3.5 Quadrant partitioning of chemical functional groups (Fig. 3.9E):

The partitioning of compounds to specific quadrants was also examined on the basis of the 299 compounds' constituent functional groups. Ketones, aromatics, aldehydes, and methoxy groups were found to be enriched in Q_1 and Q_2 relative to Q_3 and Q_4. Q_3 and Q_4, meanwhile, were enriched in alcohols and acids (and therefore also TMS groups, c.f. Fig. 3.9A), relative to Q_1 and Q_2. It is noticeable that aldehydes, methoxy, and aromatics are present in small quantities in Q4. This is because these polar functional groups are constituents of compounds, principally di-acidic lignin dimers, which also contain 3–4 nonpolar –OTMS groups that reduce the compound's ^{2}RT value, pushing these few compounds from Q_2 to Q_4. All functional groups were of lower

abundance in Q_4 because the fewest number of compounds were identified in this quadrant.

3.4.4 Quadrant-based and whole-chromatogram alignments

Peaks identified at $S/N_{\geq 10}$, $S/N_{\geq 100}$, and $S/N_{10–100}$ were aligned either across whole chromatograms or across individual quadrants and the number of aligned peaks was recorded. This was then used to compute a similarity score (*s*) for each pair-wise alignment, 45 of which constituted $^1/_2$ of [S]'s non-diagonal elements. The other half of [S] was transposed across the diagonal of 1's, indicating 100 % peak homology for samples aligned to themselves. In total, using five chromatographic bins ($Q_{1,2,3,4,All}$) subject to three different peak intensity criteria and two different alignment criteria, 25 100-element (10×10) [S] matrices were constructed, each of which had 45 unique elements derived from manually implemented pair-wise alignments within Statistical Compare.

3.4.5 Similarity score variability resulting from whole-chromatogram and quadrant-based peak alignments

Matrices of [S] for all chromatographic quadrants are shown in Tables 3.S3–17 for peaks identified at $S/N_{\geq 100}$, $S/N_{\geq 10}$, and $S/N_{10\text{-}100}$. The distribution of similarity scores, *s*, within each [S] varied significantly from one matrix of [S] to the next (Fig. 3.10). Inspection of box plots (Fig. 3.10) revealed that the range and magnitude of alignment scores across the sample set varied appreciably on a quadrant-to-quadrant basis

and that these values were modulated significantly when alignments were performed only for specific quadrants. Several other observations were made concerning how the centering, sizing, and shape of the alignment score distributions varied by quadrant, by S/N dispersion, and by increasing the stringency of the alignment criteria to include a high threshold spectral similarity test.

When low intensity peaks ($S/N_{\leq 100}$) were omitted from alignment, the inter-quartile ranges (IQRs) for $Q_{1–4}$ expanded (Fig. 3.10A→3.10B), indicating that samples became more distinguishable from one another when solely their high intensity ($S/N_{\geq 100}$) peaks were aligned. This indicated that lower intensity peaks, defined by S/N values of 10–100 (Fig. 3.10C), offered less differentiation between samples. This was evidenced by smaller IQRs when low ($S/N_{10\text{-}100}$) vs. higher intensity peaks ($S/N_{\geq 100}$) were aligned. This was not the case, however, for Q_2 where alignment of lower intensity peaks ($S/N_{10\text{-}100}$) afforded a level of sample differentiation that was commensurate to that provided by alignment of higher abundance peaks ($S/N_{\geq 100}$) in this quadrant. While the size, or span, of the IQRs was clearly affected by the S/N dispersion of peaks, the relative sizing of each quadrant's IQR was consistently preserved under all RT alignment conditions such that $IQR_{Q2} >> IQR_{Q4} > IQR_{Q1} > IQR_{Q3}$. The one exception was when a minimum spectral match threshold of 700 was applied as an additional alignment criterion for peaks with a $S/N_{\geq 10}$.

Median alignment scores for Q_1 and Q_3 were consistently and significantly ($p < 0.05$) higher than those for either Q_2 or Q_4. Median alignment scores for Q_1 were always significantly higher than those for Q2; meanwhile, those for Q_4 were significantly higher than those for Q_2 only for peaks aligned at $S/N_{\geq 10}$ or $S/N_{10\text{-}100}$, but not for those aligned

at $S/N_{\geq 100}$, or when a spectral match threshold for alignment was applied to peaks identified at $S/N_{\geq 100}$. Outliers were most commonly observed for Q_3, whose distribution of alignment scores was consistently the narrowest and therefore also the least discriminative, but whose skew was also the largest (negative) compared to the other observed distributions.

Increasing the stringency of the alignment to include a spectral match criteria of 700 (Fig. 3.10D–E) enormously reduced the IQRs and medians of alignment scores for all quadrants. Comparing Fig. 3.10A ($SN_{\geq 10}$) to Fig. 3.10D ($SN_{\geq 10}$+spectral match), IQRs for Q_1 Q_2, Q_3, and Q_4 were reduced by an average of 5.3×, including a reduction of 15× in the IQR of Q_2. The magnitude of the alignment score decrease resulting from the addition of a spectral match criteria was less pronounced when applied to peaks identified at $S/N_{\geq 100}$ (vs. $S/N_{\geq 100}$) where an average decrease in IQR of only 2.2× was observed.

3.4.6 Agglomerative hierarchical clustering and linkage analysis of alignment data

Agglomerative hierarchical clustering was performed on dissimilarity matrices, [D] = 1- [S], and dendrograms were constructed using the UPGMA linkage method and Euclidian distances. Results of this clustering are shown in Fig. 3.11 for peaks identified at $S/N_{\geq 100}$ across whole chromatograms (Fig. 3.11A), $S/N_{\geq 10}$ (Fig. 3.11B), and $S/N_{10\text{-}100}$ (Fig. 3.11C) and for individual quadrants for peaks identified at $S/N_{\geq 100}$ (Fig. 3.12), $S/N_{\geq 10}$ (Fig. 3.S2), and $S/N_{10\text{-}100}$ (Fig.3.13). In these analyses, two high distance primary clusters consistently formed in all but two instances, $Q_{1\&3}$ at $S/N_{10\text{-}100}$ (Fig. 3.13A,C). The two main high-distance clusters that consistently formed in these analyses differentiated Tahoe samples (Jan, June, July) from all other samples. In four instances

these primary clusters additionally differentiated just the Tahoe Jan and June (but not July) samples from all others (Q_1 at $S/N_{\geq 10}$ and $Q_{2,4,All}$ at $S/N_{10\text{-}100}$, Figs. 3.S2A, 3.13B,D, 3.11C). In one instance, Q_1 at $S/N_{10\text{-}100}$ (Fig. 3.13A), Tahoe Jan separated most significantly from all others. In addition to these two primary clusters, secondary, lower-distance clusters were readily identifiable and these were principally comprised of the non-Tahoe samples. Overall, the number of identifiable clusters varied by the quadrant that was aligned as well as by the S/N threshold at which peaks were identified in each quadrant.

3.4.6.1 Whole-chromatogram alignments of $S/N_{\geq 10}$, $S/N_{\geq 100}$, $S/N_{10\text{-}100}$ peaks

When peaks that were not partitioned by quadrant were aligned, five sample groupings were identifiable irrespective of the defined S/N criteria (Fig. 3.11). The sample groupings were unchanged for peaks defined by $S/N_{\geq 10}$ or $S/N_{\geq 100}$, but differed when peaks identified at $SN_{10\text{-}100}$ were aligned; in this latter case, the samples clustered by origin: FLL samples into one group and UAL + UTR samples into another. For peaks identified at $S/N_{\geq 100}$, Tahoe July and Jan were most similar (Fig. 3.11A). However, for peaks identified at $S/N_{\geq 10}$, Tahoe June and Jan samples were most similar (Fig. 3.11B). Meanwhile, Tahoe July was more similar to any sample except the other two Tahoe samples on the basis of peaks identified at $S/N_{10\text{-}100}$. The distances that separated samples were greatest when peaks identified at $S/N_{\geq 100}$ were aligned, followed by peaks identified at $S/N_{\geq 10}$ and $S/N_{10\text{-}100}$.

3.4.6.2 Quadrant partitioning of $S/N_{\geq 100}$ peaks

Peaks identified at $S/N_{\geq 100}$ were aligned according to the physicochemically defined quadrant in which they were contained. Hierarchical relationships of sample similarities were then constructed from the resultant alignment-derived [S] matrices, the results for which are shown in Fig. 3.12 as dendrograms. The discriminative capacity of each quadrant, as measured by the number of discrete, identifiable sample clusters or groupings, was as follows: Q_2 (7 groupings) > Q_4 (6 groupings) > Q_1 (5 groupings) > Q_3 (4 groupings). Groups 4,5 and 6,7 in Q_2 were reduced to two separate groups in Q_1 and groups 4,5,6 in Q_4 were reduced to one group in Q_3. Quadrants selecting for 'polar' or 'large' compounds yielded more discernible sample groupings than did those that selected for 'apolar' or 'small' compounds. As such, some quadrant-based alignments provided greater sample differentiation than was revealed by whole-chromatogram alignments.

In order to provide some context to the groupings that resulted from the most discriminating alignments (i.e. those in Q_2 and Q_4), ^{1}H-NMR functional group abundance data obtained from the CuO hydrolysate were examined and correspondences to alignment-derived sample groupings were observed. The ^{1}H-NMR parameters that showed the greatest correspondences with sample groupings derived by GC×GC quadrant alignments were methoxy (δ 3.6–3.9 ppm) and aldehydic protons (δ 9.5–10.5 ppm) for Q_2 and $H_n–C_\alpha–(OH,NR_3)$ (δ 3.9-4.8) protons for Q_4, the latter of which is characteristic of saccharide, alcohol, and amine functional groups. Indeed, these correspondences conformed to the types of compounds expected to elute in each GC×GC quadrant (Fig. 3.9E). Grouping of samples based on quadrant-specific alignments for peaks identified at

$S/N_{\geq 100}$ were better explained by ^{1}H-NMR-elucidated functional group abundances than by any other measured bulk-level property (e.g. $\delta^{13}C$, $\delta^{15}N$, %TOC, C/N). Alignments at $S/N_{\geq 100}$ also showed that samples that were spatially removed from one another could be more similar than samples isolated from the same setting, but at a different time. For instance, in Q_2 UAL Aug is more similar to FLL July than to the other UAL samples.

3.4.6.3 Quadrant partitioning of $S/N_{\geq 10}$ peaks

When peaks identified at $S/N_{\geq 10}$ were aligned by quadrant, the number of identified clusters generally decreased and the composition of these clusters (Fig. 3.S2) changed relative to those that formed from alignment of peaks identified at $S/N_{\geq 100}$. In clusters derived from alignment of $S/N_{\geq 10}$ peaks, Q_2 and Q_4 ('large') contained 6 and 7 groups, respectively, while Q_1 and Q_3 ('small') contained 4 and 2 groups, respectively. As with $S/N_{\geq 100}$-based quadrant alignments, peaks in Q_3 had the least discriminative capacity of the four quadrants as evidenced by the formation of only two weakly differentiated (low distance) groups that separated the Tahoe samples from all others. In the $S/N_{\geq 10}$ quadrant-based alignment for Q_1, the Tahoe July sample was more similar to the FLL, UAL, and UTR samples than it was to the other two Tahoe samples due to the inclusion of lower abundance peaks that were more greatly shared with the terrestrial samples than with the Tahoe samples.

The $S/N_{\geq 10}$ analysis includes peaks identified at $S/N_{\geq 100}$. Alignment of these peaks therefore profiles how the addition of $S/N_{10\text{-}100}$ peaks affects alignment results. Across all samples, 59–71 % of whole-chromatogram peaks were defined by $S/N_{10\text{-}100}$ while 29–41% were defined by $S/N_{\geq 100}$.

3.4.6.4 Quadrant partitioning of $S/N_{10\text{-}100}$ peaks

When higher abundance peaks defined by $S/N_{\geq 100}$ were excluded and only peaks identified at $S/N_{10\text{-}100}$ were aligned, sample groupings changed markedly. Q_3 again yielded the fewest number of groups and thus differentiated samples the least, and it only did so for the UTR (riverine) sample, which it separated from all other lacustrine samples (Fig. 3.13C). As with all other quadrant-based alignments, more groups were apparent in 'large' and 'polar' quadrants as opposed to in their 'small' and 'apolar' counterparts, indicating the heightened ability of these quadrants to differentiate samples not capable of being differentiated on the basis of whole-chromatogram alignments alone. The clustering of $S/N_{10\text{-}100}$ peaks did not mirror the clustering patterns observed for any other measured elemental, isotopic, or NMR-related variable, however notable similarities in clustering patterns were apparent when groupings in Q_4 were compared to groupings of fatty acid-derived PCA scores and concentrations of $C_{16:0}$. This is consistent with knowledge of Q_4, which elutes 'large apolar' compounds (Fig. 3.7A) including unfunctionalized, hydroxylated, and carboxylated fatty acids

3.4.7 Analysis by Nuclear Magnetic Resonance Spectroscopy (^{1}H)

3.4.7.1 ^{1}H-NMR spectra of CuO-oxidized DOM

Unlike the smooth, broad humps characteristic of bulk DOM, the ^{1}H-NMR spectra of CuO-oxidized DOM isolates resolved a high level of structure (Figs. 3.14, 3.S3). These spectra were fine enough to resolve *J*-couplings for protons within discrete

compounds. The oxidative liberation of monomeric constituents from precursor oligomeric DOM led to the concomitant production of protons in the aldehydic (δ 9.5–10.5), acetophenonic (δ 2.4–2.7), and carboxylic (δ 10.5–13.5) regions of the spectrum via cleavage of oxidatively labile bonds, including amides, esters, and most ethers. Proton resonances characteristic of other functional groups were also more clearly resolved in CuO-oxidized versus unoxidized ^{1}H-NMRs of SPE-DOM isolates (Figs. 3.14, 3.S3). Visual comparison of ^{1}H-NMR spectra revealed FLL Aug to be most similar to the Tahoe samples and the Tahoe sample to be, in turn, most similar to one another. Meanwhile, the FLL June and July samples closely resembled one another, as did the UAL and UTR samples. Beyond qualitative visual comparisons, however, discrete and often subtle differences in functional group abundances were quantifiable across all samples (Fig. 3.15, Table 3.S19). Alkyl protons (δ 0.8–1.4) were of highest abundance in the Tahoe + FLL Aug samples (33.6 ± 4.7 %) and were of lowest abundance in UTR (19.9 %). Aliphatic proton (δ 1.5–1.8) abundances were relatively invariant across samples, averaging 10.4 ± 0.8 %, while they were lowest (8.9 %) for UTR 2010. Abundances of benzylic and acetophenone protons averaged 26.6 ± 3.8 % and 8.9 ± 1.0 %, respectively, and although trends in the abundances of these proton-types were not strongly and consistently evident across samples, the abundances of these two groups were lowest for the three Tahoe samples, averaging 21.9 ± 1.4% and 7.8 ± 0.2%, respectively. Resonances characteristic of protons attached to carbons α to –OH moieties of alcohols, or α to $–NR_3$ moieties of amines, varied 2.6-fold across samples, averaged 5.8 ± 1.3 %, and were of lowest abundance for Tahoe June (4.9 %) and FLL Aug (3.2 %). Methoxy hydrogens showed significant variability in abundance (3.4-fold) across

samples even though their abundances were always < 6% of total integrated protons. Methoxy protons were most abundant in the UAL + UTR samples (5.2 ± 0.6 %), followed by FLL June and July (3.6 ± 0.5 %), and were of lowest abundance in the Lake Tahoe and FLL Aug samples (1.9 ± 0.2 %). Aromatic and vinylic proton abundances were lowest for FLL, averaging 11.7 ± 1.0 %, compared to the remainder of the samples for which they averaged 16.6 ± 2.0 %. Interestingly, aromatic proton abundances in the Tahoe samples were similar (17.4 ± 2.4 %) to those for the terrestrial UAL and UTR (16.0 ± 1.8%) samples. Chemical shifts from δ 8.3–9.0 occur for protons attached to highly substituted (principally by –OH and –COOH) aromatics, or belonging to sterically hindered aromatic ring structures. Such structures and low-field protons can be found in some lignin dimers and carboxy monomers, or in heteroaromatics, such as picolinic acid (δ=8.8), a catabolite of tryptophan (Melillo et al. 1993) identified in the Lake Tahoe samples. These protons varied 3.4× in abundance across samples; they were lowest for the Tahoe and FLL Aug samples (1.7 ± 0.5 %), were higher for UAL (3.4 ± 0.2 %), and were highest for UTR (4.5 %), suggesting a predominant terrestrial source for compounds containing these shifts. Aldehyde, or formyl, protons were of consistently low abundance (0.3–1.3 %) across samples. These low-field protons showed significant (4×) variability across samples, and were most abundant for the UTR and UAL isolates (1.0 ± 0.3 %), followed by FLL June and July (0.6 ± 0.04 %), and finally by the Tahoe samples + FLL Aug (0.4 ± 0.1%). Acid protons were of highest abundance in the Tahoe June and Jan samples (1.4 ± 0.2%), and were significantly lower for the rest of the samples (0.1–0.5 %); acid protons exhibited the greatest variability (15.8x) of any functional group across samples. Acid/Aldehyde ratios (as determined from NMR of the total hydrolysate) varied

33× across samples, were highest for the Tahoe and FLL Aug samples (2.2 ± 0.8), were lowest for the UTR sample (0.09), and were of intermediate values for the remaining samples (0.27 ± 0.11),.

3.4.7.2 Interrogating ^{1}H-NMR data by principal component analysis (PCA)

In examining changes of individual functional group abundances across samples (Fig. 3.15, Table 3.S19) it was apparent that some functional groups co-varied strongly while some did so only weakly, if at all, and in ways that were too subtle to discern on a qualitative, visual basis alone. Therefore, to systematically identify the sign and magnitude of these co-variations in functional group abundances, the dimensionality of the data's variability was reduced by projection onto its inherent principal components using PCA following 0–to–1 normalization (Fig. 3.16). The first three principal components collectively explained 90.8 % of the original data set's underlying variability. PC1, PC2, and PC3 explained 63.1 %, 20.8 %, and 6.9 % of the data, respectively.

PC1 (63.1 %) differentiated samples that were enriched in benzylic, acetophenone, methoxy, substituted aromatic, heteroaromatic, and aldehydic protons from those that were correspondingly depleted in these functional groups, but which were instead enriched in alkyl and acid protons. PC1 distinctly divided the samples into two main groups, separating the Tahoe + FLL Aug samples from the rest (Fig. 3.16). Tahoe June and UAL June represented the extremes of these two PC1-delineated groups. When PC1 scores were plotted versus 12 other measured sample variables (Figs. 3.S4,5–PC1), PC1 scores correlated (R^2) most significantly with $\delta^{13}C$ (–0.95), C:N (0.85), $\delta^{15}N$ (–0.83), and Ad/Al_{NMR} (-0.82) and correlated more weakly with ∑vanillyl (0.68), ∑syringyl

(0.48), ∑cinnamyl (0.43), syringyl/vanillyl (–0.21) and ∑ lignin phenol monomers (0.62). No significant correlations were observed for PC1 scores and lignin acid/aldehyde ratios, either for syringyl-type (0.27) or vanillyl-type (–0.13) phenols (Fig. 3.S5).

Inclusion of PC2 (20.8 %) additionally differentiated samples that were aliphatic-enriched/aromatic-depleted from those that were aliphatic-depleted/aromatic-enriched, leading to the further separation of samples that were otherwise too similar to one another to be clearly resolved by PC1 alone. PC2 thus led to a clear separation of the FLL Aug from the Tahoe samples, primarily owing to the comparatively lower aromaticity of the FLL Aug sample relative to the Tahoe samples (10.9 % vs. 17.4 ± 2.4 %). PC2 also separated the FLL June and July samples, primarily owing to their depletion in aromatic protons (12.1 ± 0.9 %) relative to UAL + UTR (16 ± 1.8 %) as well as to their marginal enrichment in aliphatic protons (11.5 ± 0.04 vs. 10.0 ± 0.8 %). The aromatic/aliphatic proton abundance ratio of UTR was 3.1 versus an average of 1.5 ± 0.4 for all other samples; this feature separated UTR in PC2 from the UAL samples with which it was grouped in PC1. PC2 was weakly and negatively correlated (Figs. 3.S4,5–PC2) with lignin concentration parameters (∑V, ∑S, ∑C, $\sum_8$), whose R^2 values ranged from –0.20 to –0.37, although a moderately negative correlation of PC2 with S/V was observed (–0.52).

PC3 explained only 6.9 % of the total variability and, as such, its explanatory power was subordinate to either PC1 or PC2. Separation along this third PC was principally driven by changes in acid and amine-associated protons relative to alkyl protons. Here, Tahoe Jan and July separated further from Tahoe June than they had in PC2 owing to Tahoe June's relative depletion (4.9 vs. 5.7 ± 0.5 %) of amine-associated

protons and concomitant enrichment in alkyl protons (37.0 % vs. 29.6 ± 0.9). PC3 scores were found to correlate very weakly ($-0.14 < R^2 < 0.08$) with other measured elemental, isotopic, or organic geochemical parameters (Figs. 3.S4,5–PC3), although a stronger relationship was observed for PC3 and VAd/V ($R^2 = 0.28$) than had been observed for either PC1 or PC2 and this variable.

The parallelity of variable loadings in PCA space also revealed which variables correlated most significantly with one another. For instance, ^{1}H-NMR-derived PCA plots showed that variable loadings for methoxy, hindered/highly-substituted/heteroaromatics, and aldehydic proton were oriented similarly to one another and cross-correlations of these same variables were both strong and positive ($0.66 < R^2 < 0.86$). Aromatic proton abundances correlated very weakly with all functional groups except for aliphatic protons where the relationship was modest and negative ($R^2 = -0.40$). Meanwhile, benzylic and acetophenone proton abundances co-varied strongly and positively ($R^2 = 0.85$), but both were negatively correlated with acid protons ($R^2 = -0.58$ to -0.54, respectively). These relationships are quantified explicitly in Table 3.S18 for all possible functional group comparisons.

3.4.8 Fatty Acid Variability in SPE-DOM

The fatty acid fingerprints of SPE-DOM isolates varied spatially and temporally across samples, both in composition and in the relative abundances of individual components (Table 3.S20). $C_{16:0}$ was the most abundant FA in all samples, except for UAL Aug for which $C_{18:0}$ was most abundant. On average, $C_{16:0}$ represented 30 % of all quantified lipids across samples while $C_{18:0}$ represented 23 % and $C_{14:0}$ represented 22 %.

Longer chain lipids ($C_{20:0}$ and above) were most abundant in the UAL, UTR and Tahoe July samples, were highest for UAL June, and were lowest for the Tahoe June 2009 and FLL Aug. Other FAs were represented in smaller and more variable quantities in ways that were not immediately obvious. Therefore, PCA was also applied to sum-normalized, 0–to–1 normalized fatty acid data to reduce the dimensionality of these data's variances.

PC1, PC2, and PC3, explained 36.9 %, 21.5 %, and 16.9 % of the variability in the fatty acid data, respectively, collectively accounting for 75.3 % of the observed variability. PC1 differentiated the samples into four apparent groups (Fig. 3.17A,B). Group 1 (FLL Aug, Tahoe June+Jan) was defined by higher proportions of $C_{12:0}$ FAs, Groups 3 and 4 (UAL Aug, UAL June, respectively) were conversely defined by longer ($C_{20:0} \rightarrow C_{24:0}$) FAs. Meanwhile, Group 2 (Tahoe July + UAL July + UTR + FLL June + July) was not differentiated by PC1 alone. Lipid-derived PC1 scores correlated much more weakly with other measured SPE-DOM parameters (e.g. max $|R^2| = 0.52$ with respect to ∑cinnamyl) compared to the stronger correlations (max $|R^2| = 0.96$ for $\delta^{13}C$) which were observed between these parameters and NMR-derived PC1scores (Fig. 3.S6,7-PC1). Other measurements that showed intermediate correlative strength (R^2) with FA-derived PC1 scores were ∑lignin (0.48), ∑vanillyl (0.47), $\delta^{15}N$ (0.42), and ∑syringyl (0.41) phenols.

PC2 (21.5 %) additionally differentiated FA PC1-defined Groups 1 and 2 into Groups 1b, and 2b, respectively, and separated samples primarily on the basis of their ratio of $(C_{14:0}+C_{16:1})/C_{18:1b}$. Group 2b comprised UTR + UAL July while Group 1b consisted solely of Tahoe Jan. Lipid-derived PC2 scores correlated (R^2) weakly with DOC (0.26), S/V (–0.28), and SAd/S (R^2=0.34) (Figs. 3.S6,7–PC2).

PC3 (16.9 %) weakly differentiated UAL June from UAL Aug based on UAL Aug's higher proportion of $C_{22:0}$ FAs (Fig. 3.17B). PC3 also differentiated FLL June from Group 1 based on its enrichment in $C_{15:0}$ and $C_{16:0}$ lipids relative to other samples. PC3 correlated weakly and positively with [DOC] ($R^2 = 0.21$) and weakly and negatively with all lignin parameters, R^2's for which ranged from –0.35 to –0.25.

No significant relationships were observed when FA-derived PCs 1,2, or 3 were plotted against relative functional group abundance data obtained by ^{1}H-NMR (Fig. 3.S8), suggesting that the variability in lipid composition exists at too low a concentration (< 1 μg L^{-1}) for detection by NMR, an instrument that lacks the sensitivity that is achieved by compound-specific analyses like GC×GC-TOF-MS.

3.5 Discussion

3.5.1 Visual inspection of GC×GC chromatograms of CuO-oxidized DOM

A GC×GC-TOF-MS method was successfully developed, optimized, and used for the first time to interrogate the chemical complexity of CuO-oxidized organic matter. The separation of compounds by polarity in addition to volatility along two orthogonal chromatographic dimensions yielded 2-D chromatograms that immediately revealed a level of chemical heterogeneity that has not been previously profiled by traditional 1-D methods. The additional, real-time separation of compounds on the basis of their polarity in the second dimension enabled the broad chemical constitution of a sample to be inferred from its 2-D chromatogram alone. Compounds enriched in aldehydes, aromatics, ketones, and methoxy groups eluted with distinctly different (greater) ^{2}RT times than

those that were depleted in these groups, but which were instead enriched in aliphatic, alkyl, alcohol, and carboxylic acid groups. While alcohols and carboxylic are indeed polar, their polarity is completely masked by TMS such that the compounds are made more nonpolar. As such, clear differences between chromatograms readily reveal the relative contributions of allochthonous terrestrial material vs. in-situ produced autochthonous material to aquatic SPE-DOM extracts. Here, it is assumed that algal material consists primarily of lipids, carbohydrates, and protein; it is inherently methoxy poor, does not contain appreciable amounts of compounds yielding aldehydes upon oxidation, and its aromatic amino acids do not yield methoxylated aromatic compounds upon oxidation (Goñi and Hedges 1995). Compared to aromatic and methoxy rich terrestrial OM precursor material, there are limited aquatic sources for compounds that would elute with high ^{2}RTs. Of the few compounds that did elute in this region, however, most were observed in the Tahoe samples and included some aromatic heterocycles, a few benzofuranones, 2- and ω-1-oxo fatty acids, and pyroglutamate, a lactam formed by cyclization of glutamic acid.

Lignin yields 4-hydroxy, 3-, and 3-5-methoxylated phenolic and diphenolic aldehydes, acetophenones, and acids upon CuO oxidation (Chang and Allan 1971). These compounds have distinctly elevated ^{2}RTs, especially the dimers, which makes them exceedingly well-resolved from other chromatographic eluents in the second dimension. Two lignin phenolics, vanillin and syringaldehyde, as well as the elution region of the lignin dimers, are shown in Fig. 3.7K. The ion intensities corresponding to these compounds visibly change across chromatograms, providing a measure of the terrestrial material in each hydrolysate sample. The lignin dimers (annotated in Fig. 3.7K), for

instance, were distinctly observable in the UAL + UTR (Fig 8G→K) samples, were more weakly visible in the FLL samples (Fig. 3.7D→F), and were visibly absent in the Tahoe samples (Fig. 3.7A→C).

It also appeared that chromatographic complexity could perhaps provide a qualitative measure of a sample's biodegradation. Of the FLL samples, it was greatest for FLL June, when this lake was removed from either the influence of peak snowmelt (Fig. 3.3) and fresh terrestrial OM inputs (FLL July) or the influence of fresh phytoplankton inputs (FLL Aug) as was suggested by ^{13}C and alkyl proton enrichment (Figs. 3.2, 3.15). For Tahoe, it appeared greatest in July when a transient pulse of terrestrial OM (evidenced by ↑lignin, ↑methoxy, ↑aromatics, ↓^{13}C) influenced the lake and was degraded in a shallow and highly stratified epilimnion removed from new and substantial algal inputs (low alkyl protons). Supporting this, the Tahoe July CuO hydrolysate had significantly more detectable peaks at $S/N_{\geq 10}$ (7152) than did the other two Tahoe samples (6268 ± 25, 1σ). For UAL, the visual extent of sample complexity appeared greatest for Aug '09, a period when runoff was low compared to June '09 and July '10, two samples which captured portions of peak runoff (Fig. 3.3). Peak count was also the highest in the UAL Aug '09 sample (Table 3.1). These data suggest that DOM's molecular complexity increases with temporal isolation from fresh terrestrial inputs, an interpretation consistent with biodegradation and generation of UCM-like chromatographic regions (Gough and Rowland 1990).

3.5.2 Examining the physicochemical selectivity of each quadrant

The quadrants that were conceived and delineated in Fig. 3.4 partitioned peaks on a chemically meaningful basis (§3.4.3, Fig. 3.9). Across samples, Q_3 consistently contained the highest proportion of peaks, followed by Q_1 and Q_4, while Q_2 contained the lowest proportion, regardless of the S/N used to identify peaks (Fig. 3.8). Q_3 was comprised of compounds that were termed 'small' and 'apolar' which were typically enriched in alcohol (–OTMS) and carboxyl ($–CO_2TMS$) groups, and which were depleted in aldehydes, ketones, and methoxylated compounds. Compounds observed in this quadrant included compounds such as methyl- and hydroxy-benzoic acids, hydroxy-phenylpropanoic acids, small, unfunctionalized, hydroxy-, methyl-, α,ω-diacids ($\leq C_{12}$), unsubstituted and monohydroxy benzene dicarboxylic acids, hydroxy- and normal fatty acids ($\leq C_{16}$), methyl-substituted phenols, and aliphatic diols ($\leq C_{12}$). Many of these compound types were additionally present in a rich (and often complete) array of positional isomers. The samples with the greatest proportion of peaks in Q_3 were the Tahoe Jan and June samples, showing that oxidation products of presumed autochthonous vs. terrestrially allochthonous aquatic OM disproportionately accounted for peaks in this quadrant. Correspondingly, CuO-hydrolysates of these two samples also had the highest levels of carboxylic protons by ^{1}H-NMR.

Q_1 was comprised of compounds including normal, carboxy, and formyl lignin monomers, methoxylated benzoic acids, hydroxy-, carboxy-, and methyl- acetophenones, a few benzofuranones and nitrogen heterocycles, methyl-, and isopropyl-substituted phenols, and oxo-FAs. The proportion of peaks in Q_1 was similar for all samples, but the proportion was lowest for Tahoe Jan and June at $S/N_{\geq 100}$. Meanwhile at $S/N_{10\text{-}100}$, the

proportion of peaks in Q_1 was highest for the Tahoe samples, in part because lignin-like phenolic monomers are *not* present at low ($S/N_{10\text{-}100}$), but instead only at high, concentrations in the terrestrially influenced samples.

Q_2 was primarily comprised of a rich array of aldehyde, acid, and ethanone-containing phenolic lignin dimers of varying linkage types, including the 5,5', β,1-diketone, α-1-,α-5-monoketone, and α-2 methyl dimers (Sarkanen and Ludwig 1971; Goñi and Hedges 1992). Across samples, Q_2 exhibited the most variability in the number of peaks it contained and the proportion of peaks in Q_2 was greatest for the terrestrially influenced samples. This finding is consistent with the dominance of uniquely terrestrial lignin dimers identified here.

Lastly, Q_4 contained a greater proportion of peaks in the terrestrially influenced samples at $S/N_{\geq 100}$, which is consistent with longer chain length lipids observed in these samples (Table 3.S20). Long chain FAs and hydroxylated FAs have been previously assigned to terrestrial sources (Meyers and Ishiwatari 1995). At $S/N_{10\text{-}100}$ for Q_4, little differentiation between samples was apparent. Q_4 was comprised of acid- and alcohol-rich compounds including three 4-O-5 vanillic-containing lignin dimers whose diphenyl ether linkage was predicted to resist CuO-oxidation (Chang and Allan 1971). These compounds appear to have previously escaped detection in CuO-oxidized OM. An apparently new 5,5' diacidic lignin dimer was also found (dehydrovanillic-p-hydroxybenzoic acid) in Q_4. Other compounds observed to elute in Q_4 included longer chain FAs ($\geq C_{18}$), all possible positional isomers of benzene tricarboxylic acids as well as their full suite of hydroxylated derivatives (Dickens et al. 2007), positional isomers of carboxylated vanillic acid, and longer chain hydroxy fatty acids ($\geq C_{14}$), such as those

derived from the cuticular waxes of plants (Eglinton and Hunneman 1968; Kallio et al. 2006).

3.5.3 Quantification of DOM-DOM correlations by molecular fingerprinting

Sørensen–Dice similarity coefficients (*s*) were computed from pair-wise alignments of peaks across samples for 25 different sets of alignment criteria [S] defining the types of peaks subject to alignment. Examined criteria included (1) peak intensity ($S/N_{\geq 10}$, $S/N_{\geq 10}$, $S/N_{10\text{-}100}$), (2) chromatographic peak localization (All, Q_1, Q_2, Q_3, Q_4), and (3) a minimum mass spectral similarity threshold (700) for aligned peaks. Using a minimum mass spectral similarity threshold was deemed inappropriate for reasons discussed in §3.3.8.

Inspection of Fig. 3.10 immediately revealed that alignments made on the basis of physicochemically defined quadrants captured a level of molecular-level variability between 2-D chromatograms of CuO hydrolysates that was occluded when whole chromatograms were aligned without partitioning. Quadrants exhibiting the greatest differences in conserved peak homology across samples had the largest IQRs. Q_2 had the largest IQR and also had the most visually striking peak differences between samples (Fig. 3.7) and indicated that analysis of the distribution of alignment scores for a particular quadrant provided an integrated and quantitative measure of visible features that differentiated chromatograms which were otherwise only visually describable.

The IQRs characterizing elements in [S] matrices revealed that alignment of peaks with a $S/N_{\geq 100}$ provided a greater level of discrimination between samples than did peaks defined by $S/N_{10\text{-}100}$, revealing that low intensity ($S/N_{10\text{-}100}$) peaks varied much less

across samples, regardless of quadrant, than did higher intensity peaks. This indicated that CuO-oxidized samples were *less* differentiable from one another on the basis of their lower concentration components (which were more numerous) than on the basis of their higher concentration components (which were less numerous) (Table 3.S2). This observation is consistent with the notion that DOM composition structurally diversifies to higher entropic states as it tends towards lower concentrations (Hertkorn et al. 2008) while becoming more indistinguishably homogenous. Indeed, for all quadrants except Q_2, significantly higher ($p \leq 0.05$) median similarity scores are obtained when low intensity ($S/N_{10\text{-}100}$) vs. higher intensity ($S/N_{\geq 100}$) peaks are aligned. Meanwhile, the IQRs for aligned low intensity peaks drop several-fold compared to the IQRs derived from alignment of high intensity peaks, indicating that samples become dramatically less distinguishable from one another on the basis of their more numerous, and hence structurally diverse, but lower concentration components. Q_2 may circumvent this rule because this quadrant primarily contains lignin dimers which undergo 3- and/or 5-demethylation during photochemical diagenesis (Azadfallah et al. 2008) to form an aromatic vicinal diol (o-benzoquinone) susceptible to ring rupture (Leenheer 2007). Any compounds produced from this process would create compounds whose elution times would shift them out of Q_2 as polar methoxy and aromatic moieties were lost while more aliphatic ones (–OTMS and $–CO_2TMS$) were gained. Thus, oxidative diagenesis will, in effect, operate to 'scrub' compounds from Q_2 and Q_1 as methoxy and aromatic compounds are removed.

Q_3 showed the least variability in alignment scores across samples compared to other quadrants. This intimated that the CuO oxidation of organic matter of disparate

biosynthetic origin and chemical composition (e.g. terrestrial vs. algal) yielded similar suites of oxidation products upon CuO-oxidation and these types of compounds appeared to be well-captured by the delineations of Q_3, which was also consistently the quadrant with the greatest number of peaks. This chemically convergent pool of oxidation products would be consistent with the types of compounds observed in this quadrant, which included small phenolics, functionalized fatty acids, and diacids as described more fully §3.5.2, all of which have numerous possible precursor and authigenic sources, including many core metabolic acids common to both plant and algal sources. For example, amino acids can yield aromatic phenolics and small acids upon CuO oxidation following hydrolysis and deamination (Goñi and Hedges 1995). Muconic acid derivatives can also be found here in Q_3, which can result from aromatic ring rupture following demethylation of lignin phenolics and coincident opening of the resulting labile *o*-benzoquinones. While concentrations of individual components in Q_3 could theoretically disentangle and suggest stronger vs. weaker end-member sources for compounds found in this quadrant, such an exercise was beyond the objectives of this analysis.

3.5.4 Examination of chemical similarity between DOM molecular fingerprints

The previous section examined the magnitude and variability of peak homology that was present across whole and partitioned chromatograms for three peak populations defined by varying intensities. This was done by examining the distribution of elements within [S]. The elements (*s*) within [S] quantified the similarity of each sample to every other sample and thus provided a basis for quantifying the chemical distance (1-*s*) separating two samples. This procedure effected distance matrices that were amenable to

linkage analysis and the creation of dendrograms in which samples were grouped into clusters based on the homology of their peak populations.

3.5.4.1 Whole-chromatogram similarities

The results of whole-chromatogram alignments of unpartitioned peaks identified at three intensities ($S/N_{\geq 10}$, $S/N_{\geq 100}$, $S/N_{10\text{-}100}$) are represented as dendrograms in Fig. 3.11A→C. Peaks identified at $S/N_{\geq 100}$ and $S/N_{10\text{-}100}$ represent exclusive peak populations and the alignment analyses of these peaks will therefore be of greater subsequent focus. (Peaks identified at $S/N_{\geq 10}$ (Fig. 3.11B, 3.S2) are inclusive of those identified at both $S/N_{\geq 100}$ and $S/N_{10\text{-}100}$.)

For both peak intensity criteria, two high distance clusters consistently formed. These clusters separated Tahoe Jan and Tahoe June from all other samples. When high intensity peaks were aligned, all Tahoe samples clustered together. However, when low intensity peaks were aligned, the Tahoe July sample was more similar to the terrestrially influenced samples. The Tahoe July sample corresponded to the period of peak snowmelt for Lake Tahoe. During this time, the collective discharge of Tahoe tributaries rose over an order of magnitude (Fig. 3.3) and terrestrial inputs were substantially elevated. Plumes of intense heterotrophic bacterial activity have been previously reported to occur at the confluence of the Upper Truckee River and Lake Tahoe (Paerl and Goldman 1972), indicating that microbial respiration plays an important role in addition to photo-oxidation in the removal of terrestrial OM from Lake Tahoe. Degradation of lignin ultimately leads to the production of low molecular weight acids that have been reported to include low molecular weight acids such as malonic, oxalacetic, hydroxysuccinic,

maleic, oxalic and muconic-like acids (Chang and Allan 1971). The potential presence of degraded lignin-derived material at low abundance in Tahoe July was consistent with the clustering of this sample with the terrestrially influenced samples when low intensity ($S/N_{10\text{-}100}$) peaks were aligned. This was because, like the Tahoe July sample, the other terrestrially influenced samples similarly contained low-level abundances of degraded terrestrial OM, including degraded lignin. Since lignin phenols and dimers almost all exceeded a S/N of 100 in the terrestrially influenced samples, it is *not*, in fact, the lignin that is driving the correspondence of the Tahoe July sample to the more terrestrially influenced samples. Consistent with this explanation, concentrations of DOC (43.6 vs. 35.7 ± 1.5μM C) and Σ_8 lignin (188 vs. 93 ± 23 ng L^{-1}) are elevated in Tahoe July compared to the other Tahoe samples, but mass balance indicates that the magnitude of the $\delta^{13}C$ depletion (-20.5 vs. -18.3 ±1.3‰, Fig. 3.2) is much too large to even be marginally explained by a near-doubling of the measured lignin concentrations in the July sample. This indicates that there are likely other terrestrially derived compounds beyond intact lignin that lead to the measured depletion of ^{13}C in this sample. Additionally, these compounds with terrestrial origins must be of low abundance because alignment of high abundance peaks ($S/N_{\geq 100}$) reveals Tahoe July to be more distantly related to the terrestrial samples and instead most closely related to the Tahoe Jan sample. These data again lend support to degraded terrestrial components driving the variability in the $S/N_{10\text{-}100}$ peaks for the Tahoe July sample, leading this sample to appear more similar to the more terrestrially influenced samples than with the other two Tahoe samples on the basis of its low abundance peaks.

The FLL Aug sample also clustered differently when low vs. higher abundance peaks were aligned. When dendrograms were constructed from alignments performed on peaks defined by $S/N_{\geq 100}$, this sample clustered separately from the Tahoe samples as well as from the more terrestrially influenced samples. On the basis of peaks defined by $S/N_{10\text{-}100}$, however, the sample was not significantly distinguishable from the terrestrially influenced ones. This likely derived from the addition of algal (putatively diatom) derived organic matter to FLL Aug, as was evidenced by $\delta^{13}C$ enrichment, lipid, and ^{1}H-NMR data. This input caused the FLL Aug DOM sample to become enriched in constituents that were of high abundance, but low diversity, and caused the GC×GC molecular fingerprint of this sample to evolve towards the composition of the Tahoe samples for higher intensity compounds.

Sample clustering based on homology of aligned peaks at $S/N_{\geq 100}$ was mirrored by the measured values for $\delta^{13}C$ of bulk SPE-DOM extracts. The $\delta^{13}C$ of SPE-DOM samples spanned a neat continuum bracketed by a ^{13}C-depleted terrestrial end-member and a more ^{13}C-enriched aquatic end-member. The correspondence of clustering patterns observed for $\delta^{13}C$ values and for dendrograms derived from alignment of higher intensity peaks ($S/N_{\geq 100}$) (Fig. 3.11A) showed that the isotopic signature of DOM was disproportionately driven by compounds present at elevated abundances in the DOM matrix, but which were of lower diversity (less structurally numerous). This correspondence in clustering patterns between $\delta^{13}C$ and molecular alignment data also showed that the relationships of chemical similarity derived from alignment of these molecular data could be independently validated by an externally measured parameter

and enabled the examination of the chemical composition of compounds driving isotopic variability.

The clustering of samples based on alignment of peaks at $S/N_{10\text{-}100}$ did not scale with $\delta^{13}C$ as it did for alignment of peaks with $S/N_{\geq 100}$.The clustering pattern observed here was instead most closely approximated by the DOC concentrations (Fig. 3.11C) of the aquatic settings from which the samples were isolated.

Lastly, it is notable that while the Tahoe samples are generally most molecularly similar to each other, the distances that separate them from one another are generally greater than the distances separating any two other samples from one another (regardless of the S/N of identified peaks). This suggests that the molecular composition of the Tahoe DOM is the most temporochemically dynamic of the examined aquatic settings, followed by FLL and UAL, in a manner that scales inversely with [DOC]. Unsurprisingly, the more dilute a DOC reservoir, the more sensitive its bulk composition will be to OM additions, whether they are allochthonous (e.g. peak runoff) or autochthonous (e.g. diatom bloom).

3.5.4.2 Quadrant-based similarities

The alignments of peaks across whole chromatograms revealed very few definitive details about the *chemical composition* of compounds that differentiated one sample from another. The alignment of peaks across individual physicochemically defined chromatographic regions sought to address this limitation by assessing the relationships of chemical distance across samples that were derived from the homology of their peak populations. As a unique feature to GC×GC, the localization of a compound

to a particular quadrant is dependent on the polarity of that compound which, in turn, is dependent on the types of functional groups it contains. As such, peaks were chemically fractionated by quadrant. A breakdown of the types of compounds found in each quadrant is provided in Fig. 3.7A→E and described more extensively in §3.4.3. Alignments for individual quadrants were analyzed for peaks found at $S/N_{\geq 100}$ (Fig. 3.12), $S/N_{10\text{-}100}$ (Fig. 3.13), and $S/N_{\geq 10}$. (Fig. 3.S2) within these quadrants. This enabled the quantification of peak homology across individual quadrants in addition to across the whole chromatograms (Fig. 3.12A→D vs. Fig. 3.11A).

Complementing the findings of the box plot analyses of *s*-element distributions for different sets of alignment criteria [S], the quadrants that showed the greatest variability in alignment scores across samples yielded the highest number of resolvable sample clusters, or groupings. All permutations of alignment criteria (quadrants, peak intensities) resulted in sample groupings that were unique, indicating that the similarity of one CuO-oxidized DOM fingerprint to another is dependent on the peak set chosen for comparison. The quadrants that selected most strongly for individual functional groups exhibited sample groupings that were corroborated by independent quantification of these functional group abundances by ^{1}H-NMR analysis of CuO-hydrolyzed DOM.

3.5.4.2.1 High intensity ($S/N_{\geq 100}$) peaks

3.5.4.2.1.1 Quadrant 2 'Large Polar'

Q_2 contained compounds that were 'large' and 'polar'. While most of the peaks present in this quadrant were unidentified, like with all quadrants, a preponderance of lignin dimers elute in Q_2. In fact, the hitherto unidentified peaks in this quadrant were

most probably large, lignin-like methoxylated aromatic aldehydes, acetophenones, and acids. Each lignin dimer has up to 4 methoxy groups and up to two formyl groups and is comprised of various couplings of p-hydroxyphenyl, vanillyl, and syringyl monomers which contain 0, 1, and 2 methoxy groups, respectively. The grouping of samples derived from alignment of peaks in Q_2 (Fig. 3.12B) could not be readily explained on the basis of trends in elemental, isotopic, or measured organic geochemical measurements, including lignin concentrations. While the division of Tahoe from the other samples (Fig. 3.12B) was broadly consistent with the sample divisions (i.e. C/N, $\delta^{13}C$) seen in the data set, a basis for the observed groupings of the non-Tahoe samples resisted identification.

The possibility therefore arose that these alignment-derived sample groupings may lack significance, or that they may have simply arisen artefactually during linkage analysis. Proton NMRs of the CuO hydrolysate, however, revealed a correspondence of methoxy and aldehydic hydrogen content with the sample grouping patterns observed for alignments of $S/N_{\geq 100}$ peaks in Q_2. The sum of methoxy and aldehydic proton abundances are annotated in Fig. 3.12B to show this correspondence, which broadly corresponded to the sample groupings that were derived from the Q_2 alignments. None of the other functional groups profiled by ^{1}H-NMR yielded a pattern that was similar. As such, the alignment of peaks in Q_2 corresponded with the overall methoxy and aldehyde content of the sample, which was consistent with the selective partitioning of such compounds into this quadrant (Fig. 3.7E). This provided an independent, external means to corroborate the validity of the sample clusters that resulted from alignments of peaks in Q_2 across samples.

3.5.4.2.1.2 Quadrant 1 'Small Polar'

This quadrant also eluted lignin and non-lignin derived methoxylated phenolics, but not as singularly as did Q_2, compared to which it eluted larger amounts of acidic compounds (Fig. 3.7E). The groupings of samples for Q_1 alignments (Fig. 3.12A) condensed Groups 4,5 and Groups 6,7 that differentiated Q_2. Thus, Q_1 offered poorer definition between samples than did Q_2.

3.5.4.2.1.3 Quadrant 4 'Large Apolar'

Alignment of $S/N_{\geq 100}$ peaks in Q_4 produced sample clusters (Fig. 3.12D) distinct from those in Q_2.These sample clusters closely approximated the variability in the abundance of protons attached to carbons α to polar substituents such as alcohols and amines that were measured in the CuO hydrolysates by ^{1}H-NMR. Compounds with these functional groups are expected to be enriched in this quadrant. Addition of –OH groups (–OTMS, m/z = 89) leads to significant increases in a compound's mass and hydrophobicity, chemical traits for which Q_4 selects. The strength of the correspondence of NMR resonances to the sample groupings in Q_4 was less pronounced than the correspondence of methoxy and aldehydic NMR resonances that mirrored alignment-derived sample groupings in Q_2. This weaker correspondence may be linked to the lower relative proton contribution to NMR resonances characteristic of alcohols and amines vs. methoxy groups, which contain more hydrogens per functional group moiety. The correspondence of sample groupings for peak alignments in Q_4 and by ^{1}H-NMR provided additional evidence that the relationships of chemical similarity identified between samples for $S/N_{\geq 100}$ peaks in Q_4 have an identifiable chemical basis.

3.5.4.2.1.4 Quadrant 3 'Small Apolar'

Compared to alignments of peaks in other quadrants, alignment of peaks in Q_3 differentiated samples from one another most poorly (Fig. 3.12C). Sample groupings by alignment in Q_3 represented condensations of the sample groupings that were observed for Q_4 (Fig. 3.12D). Compounds identified in Q_3 were substantially enriched in alcohols and acids (Fig. 3.7E) and the alignment-derived sample grouping pattern for this quadrant appeared to roughly correspond to acidic proton content as measured by ^{1}H-NMR (c.f. Table 3.S19).

3.5.4.2.2 Low intensity ($S/N_{10\text{-}100}$) peaks

Peaks identified at $S/N_{10\text{-}100}$ and aligned by quadrants yielded new sample groupings whose origins were more difficult to definitively disentangle. As was shown in Fig. 3.10B-C, similarities between samples were significantly ($p < 0.05$) higher across all quadrants when alignments were performed on low ($S/N_{10\text{-}100}$) vs. higher intensity ($S/N_{\geq 100}$) peaks. indicating that variations in alignment scores between samples extended beyond 'noise', whose stochasticity would act to reduce, and not increase, similarity scores. Compared to the $S/N_{\geq 100}$ quadrant alignments, the number of resolvable groups decreased in the $S/N_{10\text{-}100}$ alignments as the distances separating samples from one another decreased. The sample groupings observed for $S/N_{10\text{-}100}$ alignments could not be corroborated in any way by the ^{1}H-NMR data as was possible for the groupings based on the $S/N_{\geq 100}$ alignments. This is understandable because the analytical window of ^{1}H-NMR is limited to detecting the protons of only the most abundant compounds. These same compounds are captured by GC×GC-TOF-MS analysis for peaks with a S/N greater

than at least 100, but are not captured for lower concentration components, even though they are more structurally numerous (there are a greater number of peaks at $S/N_{10\text{-}100}$ vs. $S/N_{\geq 100}$ for all quadrants for all samples).

The sample groupings for peaks at $S/N_{10\text{-}100}$ aligned by quadrant did not correspond to groupings of any other measured environmental, elemental or isotopic parameters. However, the sample groupings that resulted from alignments of peaks identified in Q_4, which elutes aliphatic, hydroxylated, carboxylated, and lipid-related compounds, showed a correspondence with concentrations of palmitic ($C_{16:0}$) acid and with principal component scores (Fig. 3.13D) derived from PCA performed on the SPE-DOM samples' fatty acid fingerprints (Fig. 3.17), which were analyzed independently. This suggested that alignments of $S/N_{10\text{-}100}$ peaks across quadrants were in fact driven by real, underlying chemical differences between these samples. The variability in alignments for $S/N_{10\text{-}100}$ peaks in Q_4 appears to have arose from variations in a low abundance, structurally diverse, and chemically convergent background pool of hydroxylated and carboxylated aliphatics whose diversity presumably arises from both the innate biosynthetic precursors as well as from degradation of fresh OM precursors.

3.5.5 Elucidation of CuO hydrolysates by ^{1}H-NMR

Hundreds of studies have used CuO oxidations to explore the origin, composition, and diagenetic fate of natural organic matter (NOM) in the environment. However, the chemical heterogeneity of this hydrolysate mixture has not been elucidated by ^{1}H-NMR. This is despite this technique's relative simplicity and otherwise wide adoption for examining the chemical constitution of unoxidized organic matter. This is likely a

reflection of the very specific role CuO oxidation serves in most studies, which is to break down chemically heterogeneous NOM oligomers to enable the chromatographic quantification of individual compounds diagnostic of NOM's origin and diagenesis in the environment, primarily pursuant to the quantification of lignin phenols. As such, only the HPLC or GC-amenable fractions of the CuO hydrolysate have been hitherto examined.

Proton NMR spectra of NOM samples have smooth, broad humps (Fig. 3.14) that poorly or indistinctly resolve individual functional group, confounding their quantification. These broad, indistinct features arise due to fast spin-lattice relaxation times caused by the close spatial proximity of compounds within the NOM ultrastructure, leading to nonhomogeneous peak broadening (Simpson et al. 2004; Masoom et al. 2013) and more diffuse, non-Lorentzian peak shapes and massive signal overlap (Hertkorn et al. 2013). The hydrolytic disentanglement of individual compounds from the NOM ultrastructure leads to slower spin-lattice relaxation times for excited protons which enables the resolution of individual proton chemical shifts and *J*-coupling constants specific to discrete compounds (Figs. 3.14, 3.S3). The resolution of discrete compounds in the CuO hydrolysate by ^{1}H-NMR additionally enables a closer examination of the suitability of functional group chemical shift assignments for unoxidized NMR spectra. For instance, the methoxy group has a fairly narrow and consistent chemical shift of δ 3.6-3.9 ppm. However, inspection of Fig. 3.14A reveals that the proton absorbances in this shift region are greater for the Tahoe June than for the UTR sample, despite the latter's significant enrichment in methoxylated aromatics such as lignin (Fig. 3.2). As such, ^{1}H-NMR of bulk DOM cannot resolve differences in the methoxy content of these two samples because of the broad humps arising from the diffusiveness of other

functional group absorbances in this chemical shift range, which here includes aminic-linked protons. Following oxidation, however, the methoxy groups become very well resolved (Fig. 3.14, 3.S3) due to (1) the hydrolytic disentanglement of methoxy-containing compounds from the NOM ultrastructure and concomitant slowing of spin-lattice relaxation times and (2) the inertness of the methoxy group and lability of amines and saccharides to the CuO oxidation. These processes result in greatly enhanced spectral simplification of the methoxy functional group and enable its accurate quantification in the CuO hydrolysate using ^{1}H-NMR, as was done here.

^{1}H-NMR fingerprints of CuO hydrolysates also provided a molecular-level basis for disentangling changes in functional group abundances across samples. For instance, acetophenone-type (Fig. 3.S3.5) protons (δ 2.4–2.7), which represented 7.6 and 10.2 % of total quantified protons, would have neatly differentiated the terrestrial and aquatic end-members had it not been for the significant enrichment of a single compound represented by a distinct triplet (Fig. 3.S3.5) (J = 7.4 Hz) in the Tahoe and FLL Aug (aquatically influenced) samples that was otherwise absent from the terrestrially influenced samples. Such a molecular-level interpretation of DOM-derived ^{1}H-NMR data is only enabled when DOM is chemically oxidized and absorbances and J-couplings for individual compounds are measurable. ^{1}H-NMR fingerprints of CuO-hydrolyzed DOM could additionally enable rapid screening of molecular differences across NOM samples in a manner similar to what is now done for metabolomic studies (Raamsdonk et al. 2001; Nicholson et al. 2002; Krishnan et al. 2005), but bulk NOM NMR spectra are too ill-resolved to accomplish this on a molecular basis.

3.5.6 Compositional dynamics of DOM composition in a snowmelt affected system

^{1}H-NMR integrations of CuO hydrolysate data revealed substantial variability in ^{1}H-NMR functional group abundances across samples (Fig. 3.15); no two samples were identical. In fact, if an additional sample were analyzed blindly, its origin would be difficult to definitively ascertain. These ^{1}H-NMR data complemented the geochemical data in interpretable ways, assisted with deconvoluting the influence of OM source material to samples in this set following PCA, and provided a basis for identifying patterns of sample clustering derived from alignment of GC×GC peaks partitioned by quadrant ($S/N_{\geq 100}$). The magnitude of the variations of various functional group abundances, especially alkyl, benzylic, acetophenone, methoxy, hindered/heteroaromatic, aldehydes, acids, and the ratio of Ad/Al_{NMR}, showed correlations with the magnitude of variability observed for $\delta^{13}C$, $\delta^{15}N$ and C/N of these samples. These correlations ranged from $0.90 \geq |R2| \geq 0.53$ (Fig. 3.S9–A×B).

Examination of the ^{1}H-NMR data by PCA assisted with identifying the principal drivers of bulk-level chemical variability across these samples. Most (63.2 %) of the bulk-level chemical variability could be attributed to a process that most likely reflected mixing of two-distinct OM end-members (Fig. 3.16) in this pristine, sub-alpine system. One end-member is defined by an enrichment in methoxy, acetophenones, aldehydes, and hetero/hindered-aromatics and is strongly consistent with a terrestrial origin while the other end-member lacks these groups and is instead enriched in acids and alkyl protons following CuO-oxidation. This latter end-member is most consistent with an aquatic autochthonous source dominated by lipids and proteins, which account for the alkyl resonances. The proteins in particular likely account for the acid enrichment because

peptide hydrolysis leads to greater proportions of carboxylic acids (Goñi and Hedges 1995) than other types of OM precursors do upon CuO oxidation. For instance, terrestrial OM is enriched in lignin, which contains ether linkages (particularly the β-O-4 and σ-O-4 types) that yield greater amounts of aldehydic vs. acidic phenolics upon CuO oxidation (Chang and Allan 1971). On this basis alone, lignin OM is a weaker potential source of acids following CuO oxidation compared to protein. ^{1}H-NMR-derived PC1 sample scores also correlated highly ($R^2 = -0.95$) with $\delta^{13}C$, providing external validation that PC1 was, in fact, representing mixing of two distinct end-members while highlighting the utility of PCA for disentangling multivariate data along environmentally relevant dimensions.

Upper Angora Lake ^{1}H-NMR-elucidated functional group variability was greatest for protons corresponding to aldehydes, methoxy, and alcohol/amine-related groups, whose increases reflected the addition of fresh, terrestrial OM (e.g. ↑lignin ↑methoxy). In addition to the bulk isotope data, NMR data complemented the compound-specific organic geochemical data. For instance, UAL Aug '09 and UAL July '10 had near-identical lignin concentrations, 2.68 and 2.62 μg L^{-1}, respectively, and so it was expected that their methoxy contents should also be near identical as lignin is considered the predominant source of methoxy protons. However, their methoxy contents were discernibly divergent. The methoxy content of UAL July '10 (5.41 %) was 1.3× higher than UAL Aug '09 (4.30 %), a discrepancy that was only reconciled with the observation that the S/V ratio of UAL July '10 was similarly elevated (1.4×), perhaps reflecting more recent terrestrial OM inputs as snowmelt was over an order of magnitude higher in UAL July '10 than it was in August '09 (Fig. 3.3). Syringyl phenols have 2× the methoxy protons of vanillyl phenols and at high lignin concentration are more labile than vanillyl

phenols to photo-oxidation (Opsahl and Benner 1998; Hernes and Benner 2003). This conceivably caused the lower S/V ratios of UAL Aug '09.

Fallen Leaf Lake DOM ^{1}H-NMR variability was greatest for alkyl, hindered/hetero-aromatic, methoxy, benzylic, acetophenonic, aldehydic and aliphatic protons and reflected the influence of a snowmelt-coincident, terrestrial OM pulse (FLL July, ↑Σ_8 ,↑methoxy) that was superimposed onto a markedly divergent DOM regime (FLL Aug) whose composition appeared to be controlled by an aquatic end-member (↑alkyl ↑acid, ↑$C_{12:0+14:0}$, ↑Ad/Al_{NMR}, methoxy, ↑$\delta^{13}C$, ↑$\delta^{15}N$) that contained only weakly discernible, remnant traces of unambiguously terrestrial OM in the form of lignin (Σ_8). Somewhat remarkably, and despite considerable differences in size (87× smaller) and proximity to terrestrial OM sources, the DOM composition of FLL Aug was nearly indistinguishable from that of the Tahoe samples by ^{1}H-NMR (Fig. 3.15), reflecting the existence of similar aquatic end-members exerting control over the bulk DOM composition in each of these lakes. The Tahoe samples were, however, resolved from FLL Aug samples by PC2 which explained 20.8 % of the ^{1}H-NMR variability, primarily owing to the very significant aromaticity of the Tahoe samples relative to the FLL Aug sample (Fig. 3.16A). FLL June had a DOM composition that was intermediate between the allochthonous-terrestrial and autochthonous-aquatic dominant DOM modes of FLL July or FLL Aug, respectively.

Variability in ^{1}H-NMR functional group abundances for Lake Tahoe was greatest for acid, alkyl, and hindered/hetero-aromatic protons, although methoxy proton abundances also varied discernibly and were highest for Tahoe July, coincident with the summer snowmelt (Fig. 3.3) and elevated lignin concentrations. Tahoe June '09 lacked

the terrestrial influence of the Tahoe July '10 sample likely because the pulse of terrestrial OM-rich snowmelt waters had not yet reached the sampling station at the center of the lake. The higher lignin, methoxy, and hindered/hetero-aromatic content of the January sample relative to the June sample are perplexing given the low runoff during this time, but could be related to a deepening of the epilimnion to ~110 m in January 2010 (TERC 2011) and mixing with hypolimnion waters containing a moderately enriched terrestrial component that may accumulate in the absence of elevated sunlight and microbial metabolic processes present at shallower depths.

Notably, the aromatic content of Lake Tahoe SPE-DOM samples was most similar to that of UTR, despite the exceedingly low concentrations of lignin in Lake Tahoe. Lignin is the principal source of aromaticity in terrestrial DOM. This suggests that Lake Tahoe's aromaticity is due to some other component and the principal alternative candidates are the aromatic amino acids tyrosine, phenylalanine, and tryptophan. The presence of these amino acids in these SPE-DOM samples was confirmed by 2-D NMR (Goldberg et al.) and additional corroboration was obtained here with the observation of ^{1}H-NMR resonances in the CuO hydrolysate that corresponded to the protons of the aromatic portions of tyrosine and phenylalanine (Fig. 3.S3) whose aromatic splitting patterns were discretely and individually resolvable. Absorbances for protons of these two aromatic moieties were substantially enriched in the Tahoe samples while being markedly absent from the terrestrially influenced samples, as shown explicitly in Fig. 3.S3-6,7–A→J. Additionally, GC×GC-TOF-MS analysis of the Tahoe CuO hydrolysates revealed them to be depleted in the methoxylated phenolics characteristic of lignin. What is nevertheless apparent is that this aromatic component is slowly accumulating in Lake

Tahoe, as evidenced by its appreciable presence in all Tahoe samples compared to the seasonal dynamics this aromatic component undergoes in FLL where it is either present at low abundance (recently produced, FLL Aug) or it is absent because it has been degraded (FLL June, July). Further evidence that the source of this proteinaceous material was algal came from the observation that increases in this aromatic component (FLL Aug) (Fig. 3.S3) were coincident with $\delta^{13}C$- and alkyl proton enrichment (Figs. 3.2, 3.16A), which revealed the fixation of atmospherically equilibrated CO_2 by aquatic algae and concomitant generation of algal OM, which is enriched in short-chain aliphatic lipid material (Fig. 3.17A) relative to terrestrial OM. This proteinaceous material may well accumulate at similar rates in both Tahoe and FLL, but the long, 650-year hydraulic residence time of Lake Tahoe (Jassby et al. 1999) allows it to accumulate to greater concentrations than can occur in FLL, which has a much shorter residence time of only 8 years (Table 3.S1). In the end, the residence time of Lake Tahoe, exceptionally long by comparison to most lakes, led to aromatic protons being a very poor discriminator of terrestrial allochthonous vs. aquatic autochthonous carbon despite the ubiquity of highly aromatic lignin-rich terrestrial organic matter in this high altitude, sub-alpine coniferous ecosystem. This was shown clearly in Fig. 3.16A where the loading of aromatic protons was near-orthogonal to PC1, indicating its poor prognostic utility towards differentiating the terrestrial and aquatic end-members that were otherwise so clearly delineated by PC1.

3.6 Conclusions

A unique set of samples reveals the intense, seasonally driven temporochemical dynamics of DOM composition in several sub-alpine lacustrine (+ one riverine) settings surrounding Lake Tahoe, CA, USA. Comprehensive GC×GC-TOF-MS was applied as a novel tool to interrogate the molecular complexity of CuO-oxidized DOM isolated from these pristine environments. Thousands of peaks corresponding to unique structures were revealed per sample, 299 of which were known to this study. This yielded a molecularly defined DOM fingerprint within which compounds were physicochemically fractionated on the basis of their surface polarity using an orthogonal chromatographic dimension. Quantitative measures for aligning peaks across chromatograms were developed, described, assessed, and successfully implemented using Statistical Compare, and errors associated with these procedures were quantifiably parameterized. A quantitative metric was then devised to compute pair-wise dissimilarities between whole or sectioned chromatograms based on the retention time homology of their shared peaks. Uniquely enabled by GC×GC-TOF-MS, this newly developed methodology elucidated molecular-level distance relationships between populations of peaks from different samples that were defined by specific physicochemical and abundance attributes. Using this approach, chemical distance relationships for the following compound types were drawn and analyzed across samples: (1) 'small polar', (2) 'large polar', (3) 'small apolar', and (4) 'large apolar'. These classifications were, respectively, shown to broadly correspond to (1) methoxylated/aldehydic aromatics and benzofuranones, (2) methoxylated di-aromatics, (3) small hydroxylated and/or carboxylated aromatics or small hydroxylated

and/or carboxylated aliphatics, and (4) poly-hydroxylated and/or poly-carboxylated aromatics and large aliphatics. Samples in this set were most differentiated from one another by the contents of their 'large polar' peaks (2), followed by 'large apolar' peaks (4), 'small polar' peaks (1), and 'small apolar' peaks (3). The chemical distance relationships drawn across samples for each respective peak type were corroborated and interpreted in the context of independently obtained ^{1}H-NMR spectra of the CuO hydrolysates, a common hydrolysate mixture whose chemical richness appears to have not been previously defined by ^{1}H-NMR. Chemical distance relationships for low abundance peaks were tentatively corroborated and interpreted using FA-derived PCA scores and fatty acid fingerprints, and sample clustering in Q_4 reflected the alignments of carboxylated and hydroxylated aliphatics. Alignment of low abundance peaks revealed a low-level, structurally diverse, numerically large pool of compounds that had high peak homology across samples and could reflect the entropy-driven convergence of DOM towards a structurally diverse, isomer-enriched, composition. This convergence renders DOM increasingly indistinguishable on the basis of resolution by unique retention times, as revealed by higher levels of cross-sample alignment for lower abundance peaks. It was also shown in this system that aromatic protons were extremely poor discriminants of terrestrial and algal organic matter due to the peculiarities of this system associated with the unusually long residence time of Lake Tahoe and the accumulation of a refractory algal component enriched in protein. Some evidence suggests that accumulation of nitrogen may be assisted by P-limitation (Guillemette et al. 2013). It is interesting that a lake with an apparently refractory proteinaceous pool of organic matter also appears to accumulate aromatic amino acids, whose conjugated π-systems impart to them a

susceptibility to photo-oxidation which, in part, underlie the vanishingly low levels of lignin present in Lake Tahoe. This would imply that hydrophobic aromatic amino acids are occluded from direct ultraviolet (UV)-light exposure while the similarly aromatic, but comparatively more polar lignin is not. This could be due to the surface occlusion of hydrophobic amino acid residues within DOM macromolecules (or within loosely associating macromolecular complexes) which act to shield these photoreactive residues from UV radiation. Hydrophobic associations and hydrogen bonding have been previously implicated as important mechanisms in the long-term preservation of protein (Nguyen and Harvey 2001).

Table 3.1: Variability in peak count across CuO-oxidized Tahoe Basin SPE-DOM samples (derivatized with TMS) as counted from chromatograms obtained from 1-D GC-(quadrupole)-MS (60m, 0.25 mm i.d., 0.5 μm df DB-5 column, 4 ˚C min^{-1} from 100–320 °C, 1 and 10 min initial and final holds, respectively) and as counted from chromatograms obtained from analysis by GC×GC-TOF-MS (10 m RTX-5MS, 0.18 mm i.d., 0.2 μm df + 1.2 m RXI-17, 0.10 mm i.d., 0.1 μm df column set, 6 ˚C min^{-1}, 34.5 min runtime) for peaks with a S/N of 10 or greater.

	Number of Peaks		
SPE-DOM Isolate	GC-Q-MS	GC×GC-TOF-MS	Fold-Increase
Tahoe June 2009	103	6250	61
Tahoe Jan 2010	117	6285	54
Tahoe July 2010	141	7152	51
FLL Aug 2009	150	8577	57
FLL June 2009	173	8538	49
FLL July 2010	217	8737	40
UAL Aug 2009	225	9707	43
UAL June 2009	200	9022	45
UAL July 2010	218	9532	44
UTR July 2010	220	8807	40

Table 3.2: Factors external to CuO-oxidized SPE-DOM compositional differences that contribute variability to computed pair-wise similarity scores include variability associated with the implementation order of the alignment algorithm (σ_{AA}), oxidation replicates (σ_{OR}), and replicate injections (σ_{RI}), the combined variability (σ_{CV}) of which can be computed as $\sigma_{CV} = \sqrt{\sigma_{AA}^2 + \sigma_{OR}^2 + \sigma_{RI}^2}$.

External Influence	**Similarity Score 1σ (%)**
Alignment algorithm	0.098
Oxidation replicates	0.51
Injection replicates	0.76
Combined variability	0.92

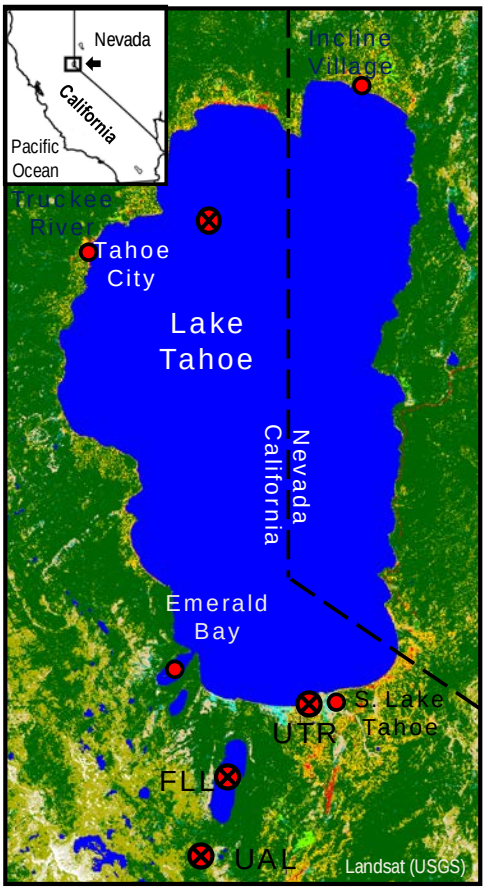


Figure 3.1: The location of the upper Lake Tahoe Basin on the California-Nevada border with selected features atop a Landsat background (USGS). DOM isolates (⊗) were taken from Lake Tahoe, Fallen Leaf Lake (FLL), Upper Angora Lake (UAL), and the Upper Truckee River (UTR).

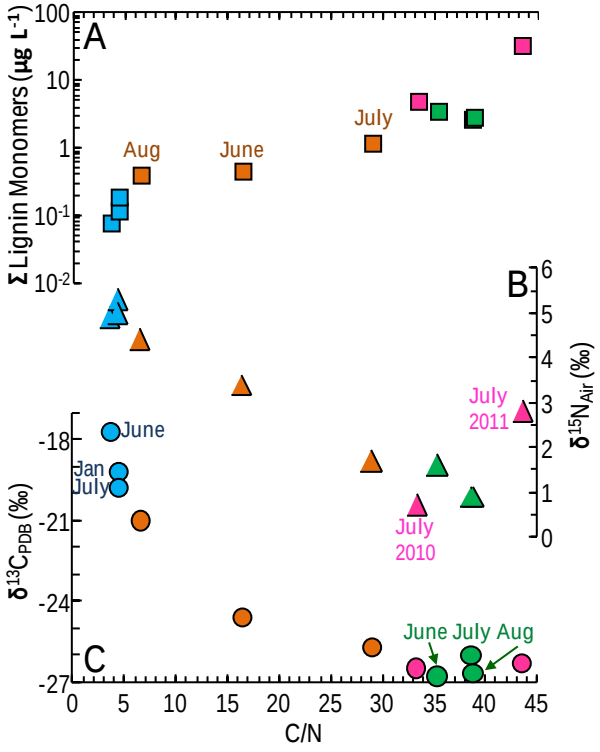


Figure 3.2: Organic geochemical, elemental, and isotopic space occupied by the SPE-DOM isolates studied here showing the correspondences between the sums of vanillyl, syringyl, and cinnamyl lignin phenolic monomers (A ■), $\delta^{15}N_{AIR}$ (B ▲) and $\delta^{13}C_{PDB}$ (C ●) to C/N ratios. SPE-DOM isolates are annotated by time-point and are coded by color, with Lake Tahoe in blue, Fallen Leaf Lake (FLL) in orange, Upper Angora Lake (UAL) in green, and Upper Truckee River (UTR) in magenta. Data are described in (Goldberg et al.) with the exception of UTR 2011, which has been added here.

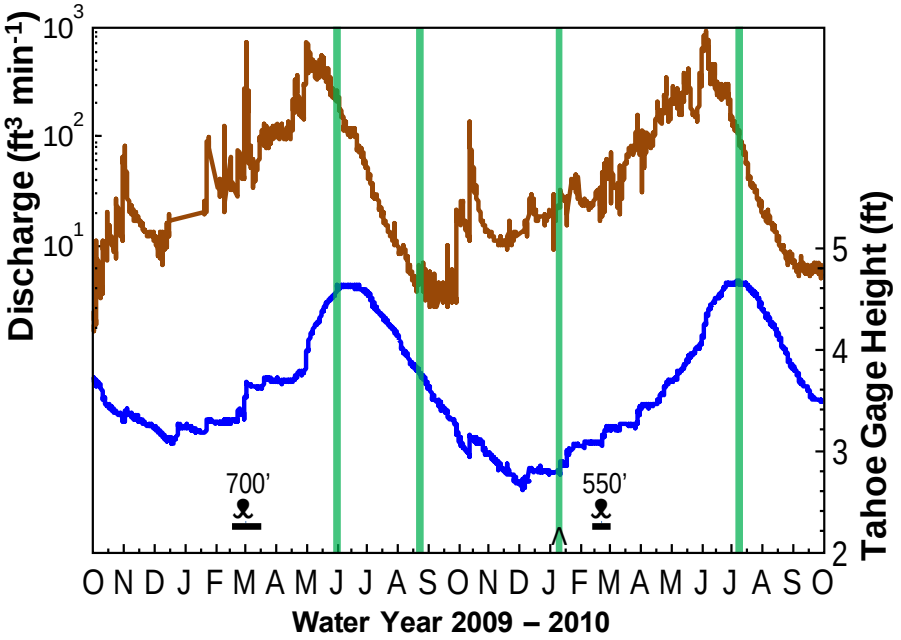


Figure 3.3: The timing and shape of the discharge profile for the Upper Truckee River, which accounts for roughly 25 % of runoff into Lake Tahoe, mirrors that of Lake Tahoe's other tributaries, all of which are controlled by snowmelt-derived runoff which increases with warmer temperatures through early-to-late-spring before peaking and subsiding through the fall. The height of Lake Tahoe increases correspondingly with discharge as do the heights of both FLL and UAL (data not shown). Discharge data for UTR is from USGS station 10336610 at South Lake Tahoe, CA. Gage height data is from USGS station 10337000 at Tahoe City, CA and is reported as height above 6,220'. Vertical green bars indicate time intervals during which Lake Tahoe, UAL, FLL, or UTR were sampled. The depth and time of annual mixing are indicated by (𓀟) in feet. Lake Tahoe had mixed to ~360' when the Jan 2010 sample was collected (Λ).

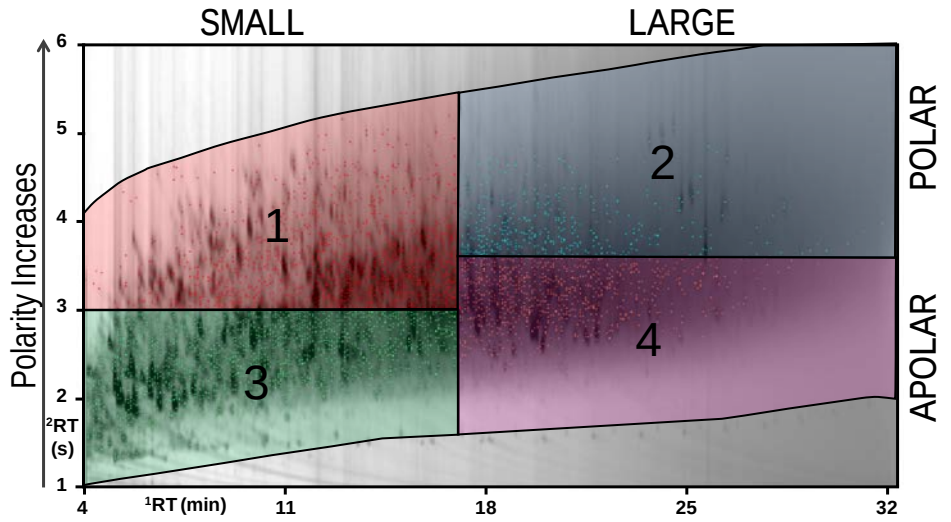


Figure 3.4: A representation of the GC×GC chromatogram partitioning scheme employed to differentiate peaks on a physicochemical basis prior to alignment. The dots denote peaks with U^+ S/N intensities greater than 10 and are overlain on a GC×GC-TOF-MS (UTR) summed extracted ion chromatogram comprised of m/z ions 73 (–TMS), 89 (–OTMS), and 117 (–CO_2TMS). The chromatogram is delineated into four physicochemically defined quadrants on the basis of molecular size (x-axis) and polarity (y-axis).

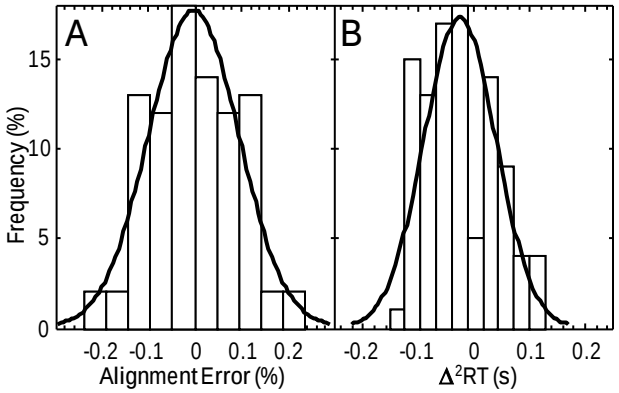


Figure 3.5: Error in alignment-derived similarity score (A) resulting from ordering of samples prior to pair-wise peak alignments performed in Statistical Compare showing the non-commutative nature of the alignment algorithm and its conformance to normal error statistics (overlaid, fitted Gaussian) characterized by a 1σ of 0.098 % (n = 90). Observed ^{2}RT variability (B) for 100 compounds compared across two chromatograms (overlaid, fitted Gaussian), used to determine the specified ^{2}RT threshold window for alignment. To capture ≥ 98% of the expected variability in ^{2}RT, it was determined that a minimum ^{2}RT window of 0.15 s was needed, or 2.32σ, where 1σ = 0.065 s.

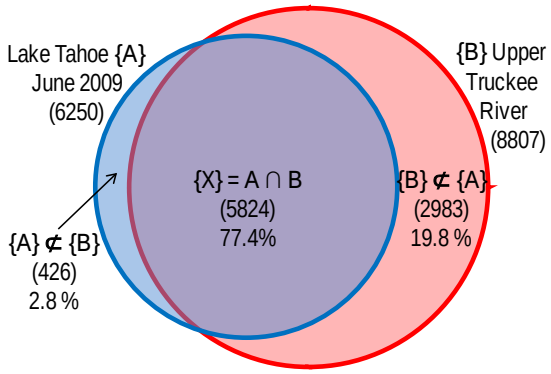


Figure 3.6: Venn diagram illustrating the methodology used for comparing two sets of peaks, {A} and {B}, whereby the peaks unique to set {A} are identified as {A} ⊄ {B}, and the peaks unique to set {B} are identified as {B} ⊄ {A}, and the peaks intersecting sets {A} and {B} are identified as {X} = {A} ∩ {B} where in the above example {A} and {B} refer to the peaks contained in the Lake Tahoe June 2009 and Upper Truckee River 2010 CuO-oxidized DOM samples. Peak count is in parenthesis above set-normalized percentages.

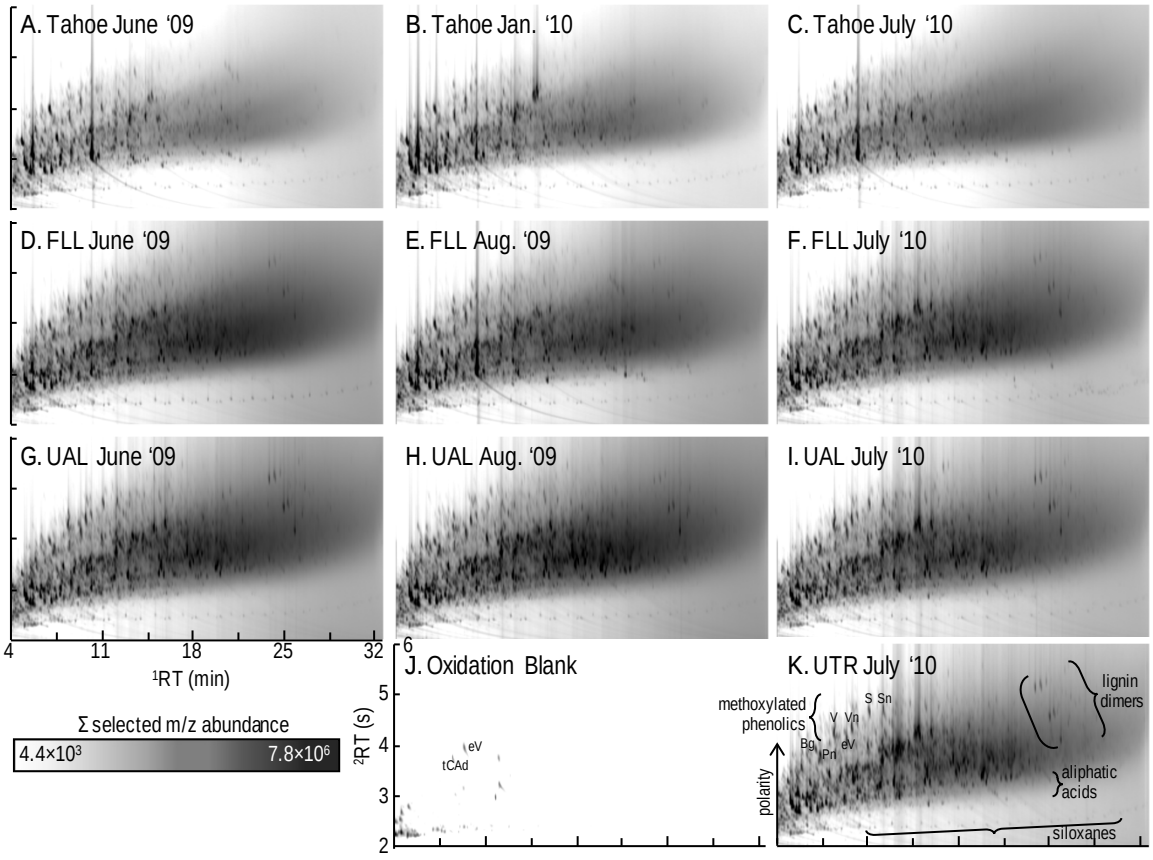


Figure 3.7: Tahoe Basin CuO-hydrolyzed, TMS-derivatized SPE-DOM GC×GC-TOF-MS chromatograms constructed from identically scaled $\log_{10}$ summation of m/z ions 57 ($–CH_2CH_2CH_3$), 73 (–TMS), 89 (–OTMS), and 117 ($–CO_2TMS$). Clear differences between the chromatograms are immediately apparent. For instance, the lignin dimers are visually absent from the Lake Tahoe samples but are present in all others samples. Similarly, the diversity and abundance of small, polar phenolics varies substantially across samples, as does the size, shape, and position of other peaks. The intensity and diffuseness of the UCM-like region also varies. Some compounds are indicated next to their identifiable 'blobs', including V = vanillin, Vn = acetovanillone, S = syringaldehyde, Sn = acetosyringone, eV = ethyl vanillin (IS), Bg = phenylglyoxylic acid, Pn = *p*-hydroxy acetophenone, tCAd = *trans*-cinnamic acid (IS). Aliphatic acids (FAs, α,ω-dioic acids, and oxo-fatty acids) each form individual homologous compound series related by $–CH_2$ units.

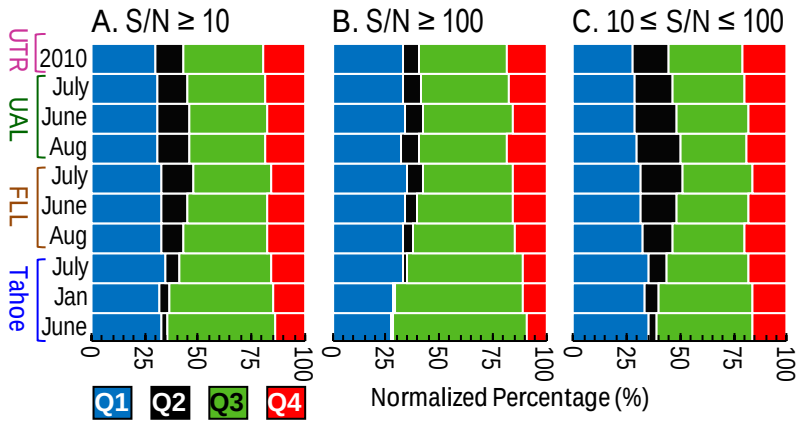


Figure 3.8: The S/N dispersion of peaks binned by quadrant shown by normalization of quadrant bin size to the total number of peaks in Q_1–Q_4 for each sample for peaks identified at $S/N_{\geq 10}$ (A), $S/N_{\geq 100}$ (B), and $S/N_{10\text{-}100}$ (C), determined as the difference in peak count between A and B. Q_1 = small polar, Q_2 = large, polar, Q_3 = small apolar, Q_4 = large apolar. The quadrant with the greatest number of peaks was Q_3, followed by Q_1, Q_4, and Q_2. Peak counts are shown in Table 3.S2.

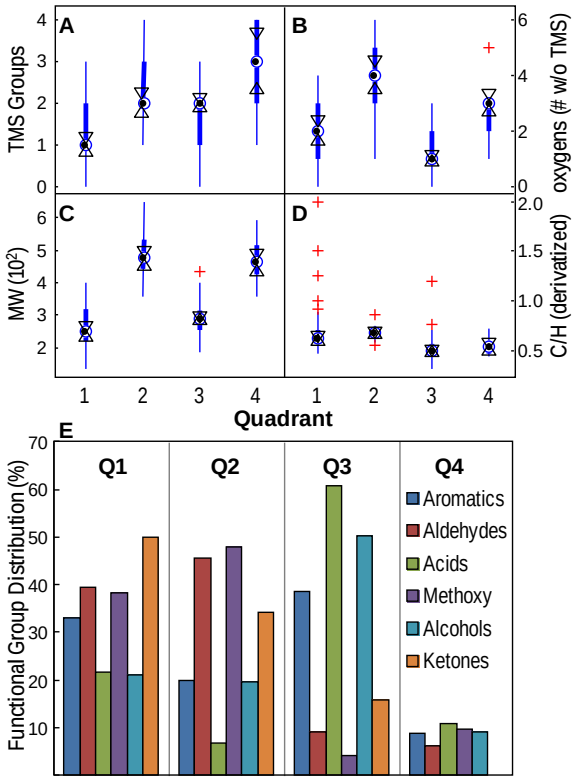


Figure 3.9: Compilation of the physicochemical attributes of compounds (n = 299) identified in GC×GC-TOF-MS chromatograms from all samples according to the chromatographic quadrant in which they eluted. Examined attributes included the number of TMS groups per compound (A), the number of TMS-free oxygens (B), molecular weight (g mol^{-1}) (C), atomic C/H ratio of derivatized compounds (D), and the distribution of aromatic, aldehyde (–CHO), acid ($–CO_2TMS$), methoxy ($–OCH_3$), alcohol (–OTMS), and ketone functional groups (E). In the box plots, the black dot denotes the median ($\tilde{x}$, q_2), the box edges denote the bounds of the first (q_1) and third (q_3) quartiles, and the whisker lengths encompass ± 2.7 σ, or 99.3% of the data if they were normally distributed, which is not the case for (D). Significant outliers are denoted by the red "+". The span of the self-pointing triangled notches denotes the interval $q_2 \pm 1.57(q_3 - q_1)/\sqrt{n}$) and indicates that two medians are significantly ($p \leq 0.05$ or ≥ 95 % confidence interval) different from one another when the intervals of two medians do not overlap (Nelson 1989). Identified structures are binned by their constituent functional groups (as they existed before derivatization) and the partitioning of these functional groups between physicochemical quadrants is shown in (E) expressed on a sum-normalized basis to the total number of compounds across all quadrants containing a particular functional group. For instance, 86 % of compounds containing methoxy groups are located in either Q_1 (38 %) or Q_2 (48%). Of the 299 identified compounds, 182 (61%) had aromatics, 33 (11 %) had –CHO, 204 (68 %) had –COOH (24 %), 73 had $–OCH_3$ (24 %), 185 had –OH (62 %), and 70 (23 %) had ketones ($\Sigma \neq 100\%$). 76 (25.4 %) of the compounds were in Q_1, 37 (12.4 %) were in Q_2 (, 164 (54.8 %) were in Q_3, and 22 (7.4 %) were in Q_4 ($\Sigma = 100\%$).

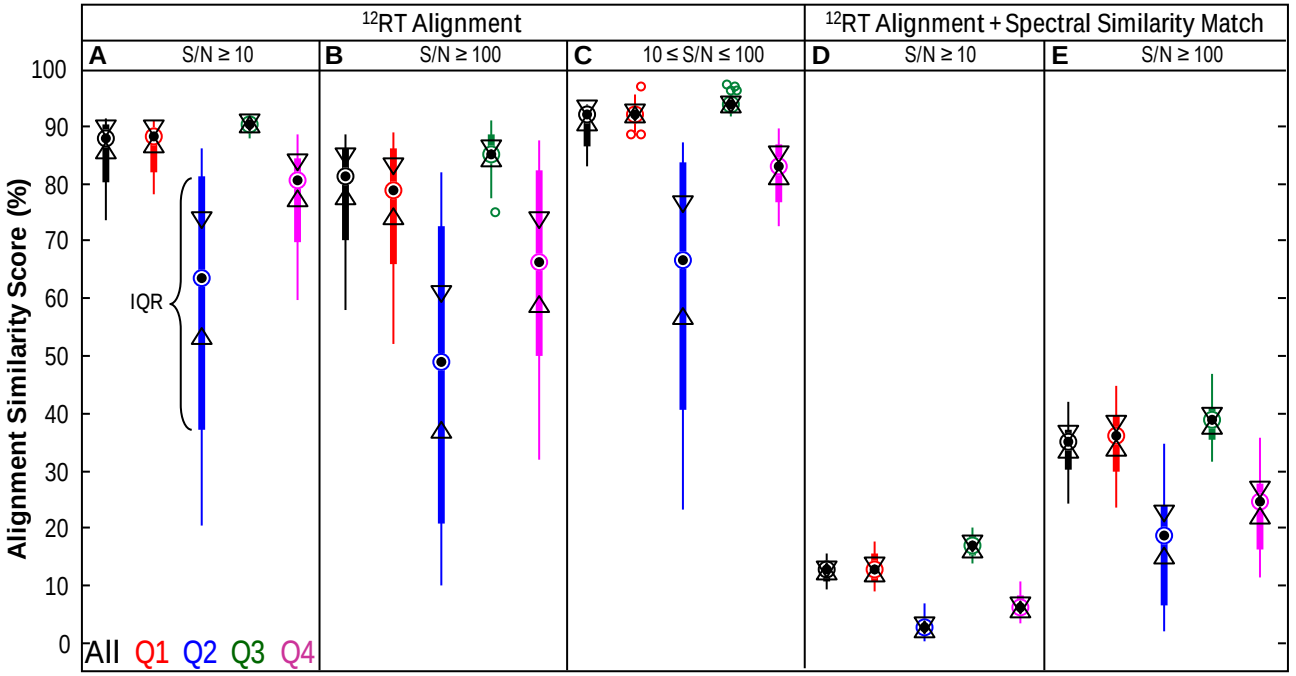


Figure 3.10: The observed variability of alignment scores resulting from pair-wise alignments of all samples for whole-chromatogram (All) and quadrant-partitioned (Q_{1-4}) chromatograms (as keyed by color) for peaks detected at varying abundances, as discerned by their U^+ S/N ratio. Alignments were performed for peaks identified with U^+ S/N ≥ 10 (A), S/N ≥ 100 (B), and 10 ≤ S/N ≤ 100 (C) and with the added alignment criteria of a minimum mass spectral similarity match of 700 for peaks identified with U^+ S/N ≥ 10 (D) and S/N ≥ 100 (E). In the box plots, box edges encompass quartiles one (q_1) and three (q_3), the black dot denotes the median ($\tilde{x}$, q_2) whiskers correspond to ± 2.7 σ, or 99.3 % of the variability of normally distributed data, and '○' denotes significant outliers. Whisker asymmetry reveals the skew of the alignment scores across samples. The span of the self-pointing triangled notches denotes the interval $q_2 \pm 1.57(q_3 - q_1)/\sqrt{n}$) and indicates that two medians are significantly ($p \leq 0.05$ or ≥ 95 % confidence interval) different from one another when the intervals of two medians do not overlap (Nelson 1989). As there are 10 inter-compared isolates (k), n = $\sum_{k=1}^{10}(10 - k) = 45$ for all boxes.

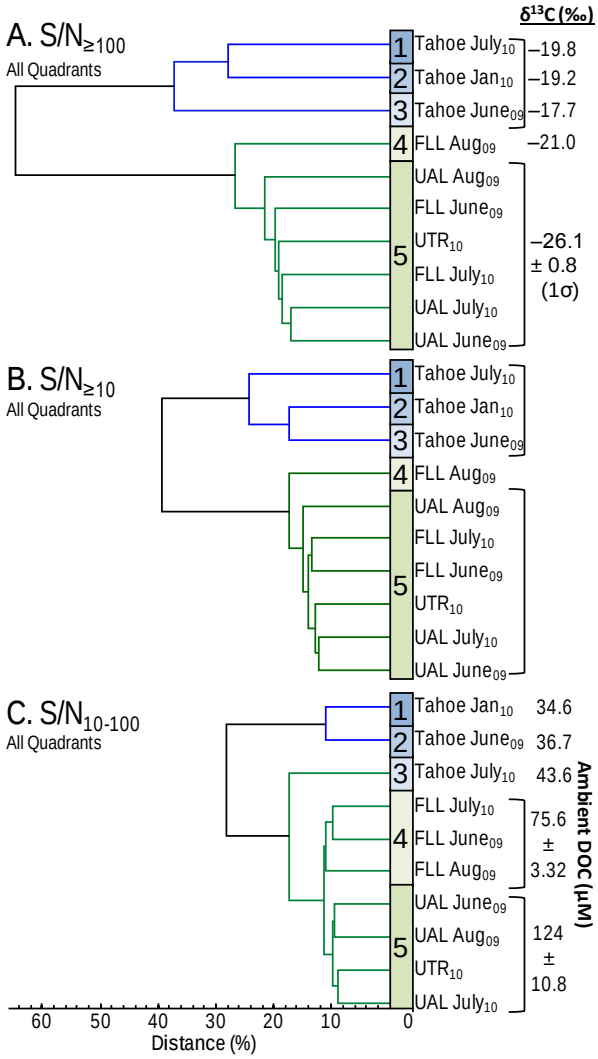


Figure 3.11: Dendrograms produced from linkage analysis of dissimilarity score matrices derived from whole-chromatogram pair-wise alignments of peaks identified at $S/N_{\geq 100}$ (A), $S/N_{\geq 10}$ (B), and $S/N_{10\text{-}100}$ (C) with colors denoting the most significant clusters, which neatly discriminate Tahoe samples from all others, although other significant sub-clusters are also present. The discriminative power of the hierarchical clustering (average distance between clusters) is markedly attenuated (from A to C to B) as lower S/N peaks participate in the alignment.

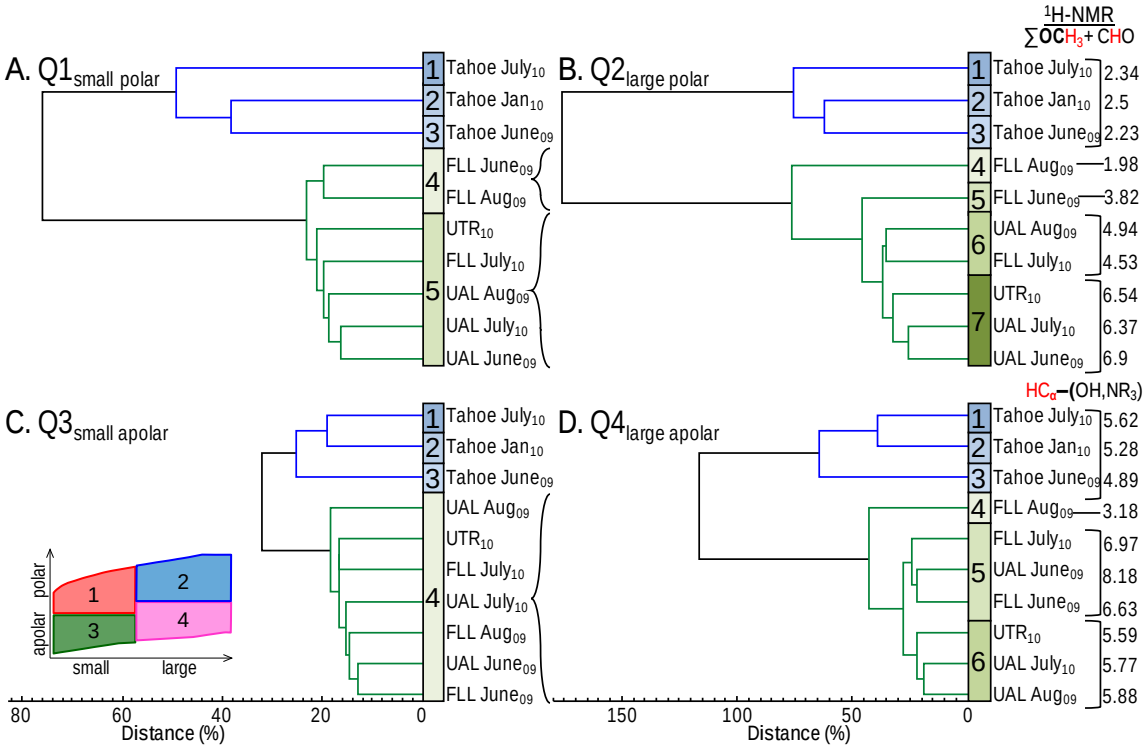


Figure 3.12: Hierarchical relationships derived from alignments performed on quadrant-partitioned (A–D) peaks identified at $S/N_{\geq 100}$. The ^{1}H-NMR normalized areas that best corresponded to the indicated sample groupings are shown to the right of B and D. For B, groupings best corresponded to the observed abundances of methoxy and aldehydic protons while for D groupings best corresponded to the abundances of protons α to –OH or $–NR_3$ groups, such as for those present in alcohols, saccharides, and amines. Groups 4,5 and 6,7 in Q_2 condense into single groups in Q_1 and groups 5,6 in Q_4 condense to a single group in Q_3, showing the existence of a greater number of distinguishing peaks in Q2 and Q4 quadrants than in their respective quadrants Q_1 and Q_3.

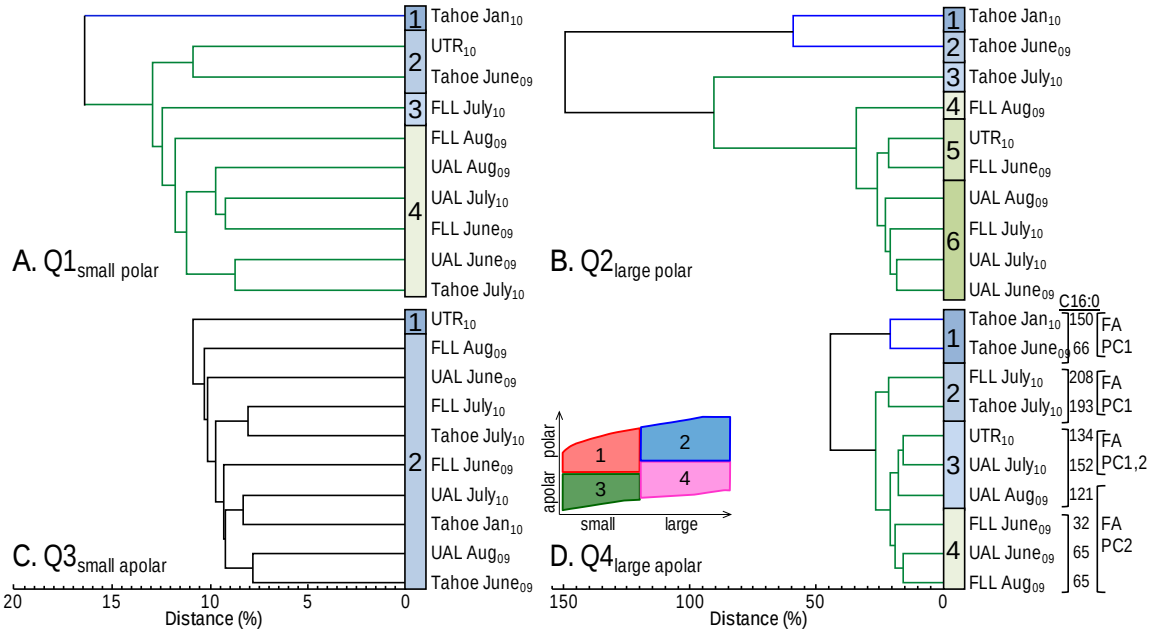


Figure 3.13: Hierarchical relationship of alignments performed on quadrant-partitioned (A–D) peaks identified at $S/N_{10\text{-}100}$. $C_{16:0}$ concentrations and fatty acid principal components (PC) 1 and 2 showed clustering patterns that were similar to those which arose from GC×GC peak-based alignments. No clustering trends were observed when $S/N_{10\text{-}100}$ peaks were identified and aligned in other quadrants.

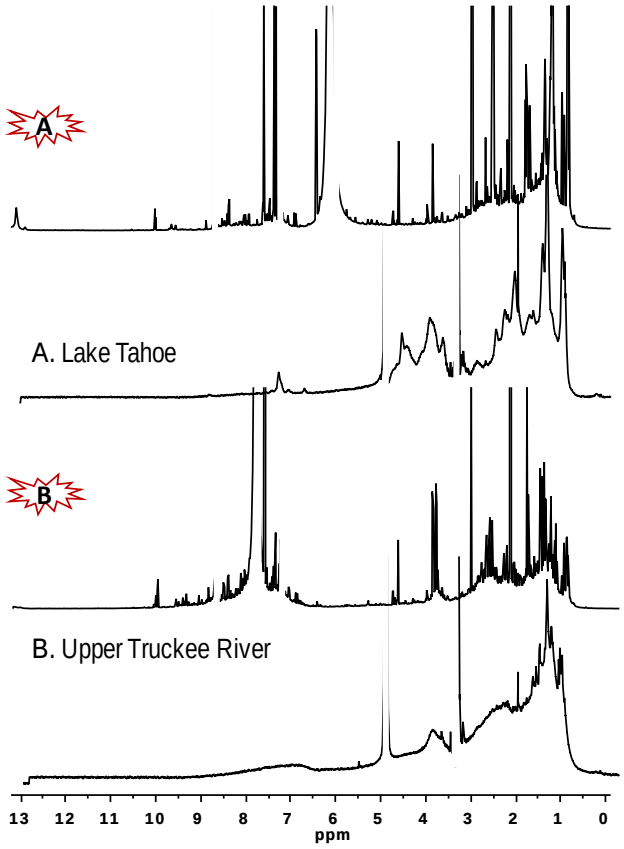


Figure 3.14: ^{1}H-NMR spectra are 'cracked' open by oxidation as DOM constituents are oxidatively liberated from the OM matrix in which they are chemically entwined, yielding NMR spectra with sharp, jagged peaks from which chemical shift and *J* coupling constants of individual protons within discrete compounds are resolved. Tahoe June 2009 (A) and Upper Truckee River July 2010 (B). Red, jagged oval denotes the ^{1}H-NMR of CuO-hydrolyzed samples. H_2O and solvent peaks have been excised from these spectra.

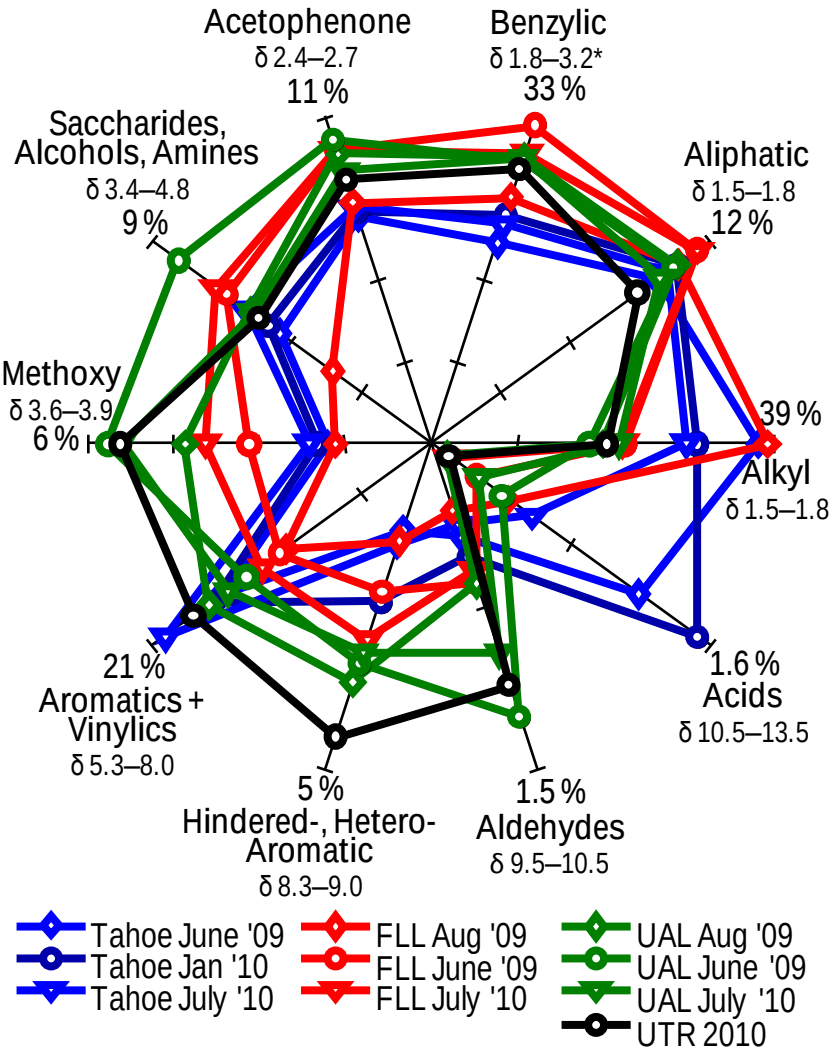


Figure 3.15: Spider plot of ^{1}H-NMR functional group abundances for CuO-oxidized SPE DOM isolates. Functional groups are defined by annotated δ ranges and normalized abundances are shown on independently scaled axes. The origin is 0%. For each sample, the sum of its functional group abundances = 100 %. * = benzylic δ range is exclusive of acetophenone δ range.

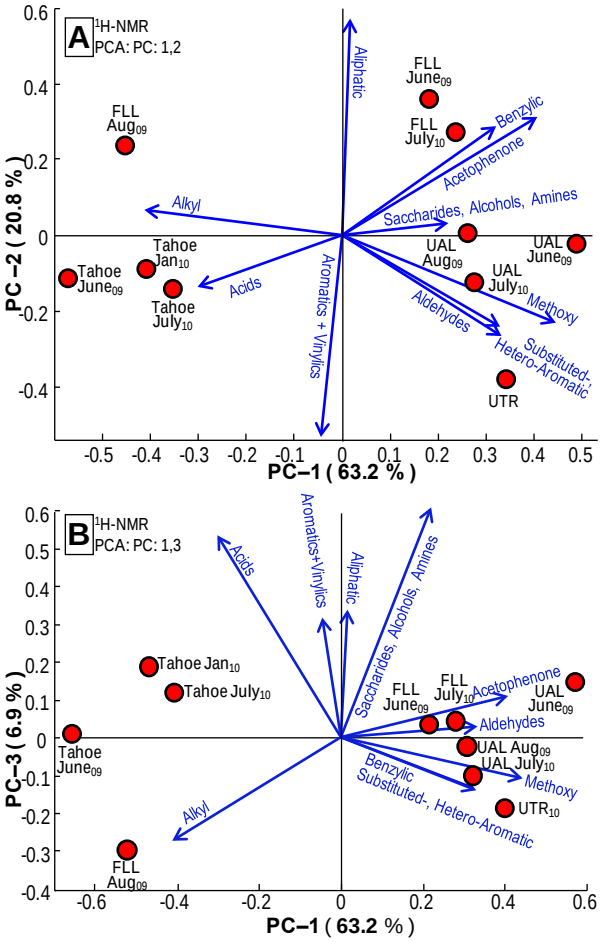


Figure 3.16: Biplots of PC1, PC2 (A) and PC1, PC3 (B) displaying range-normalized sample scores (circles) and component (variable) loadings (vectors) derived from PCA of sum-normalized, 0–to–1 transformed, ^{1}H-NMR functional group abundance data for the CuO-oxidized DOM isolates. 90.9 % of proton functional group variability in the data was modeled by the first three principal components. PC1 explained 63.2 % of the observed variability and neatly differentiated samples by their isotopically (δ^{13}C, δ^{15}N), elementally (TOC%, C/N), and environmentally ([DOC]) defined end-members (Fig. 3.S4). PC2 explained 20.8 % of the variability and further differentiated samples by their proportion of aliphatic to aromatic carbon while PC3 only explained an additional 6.9 % of the variability.

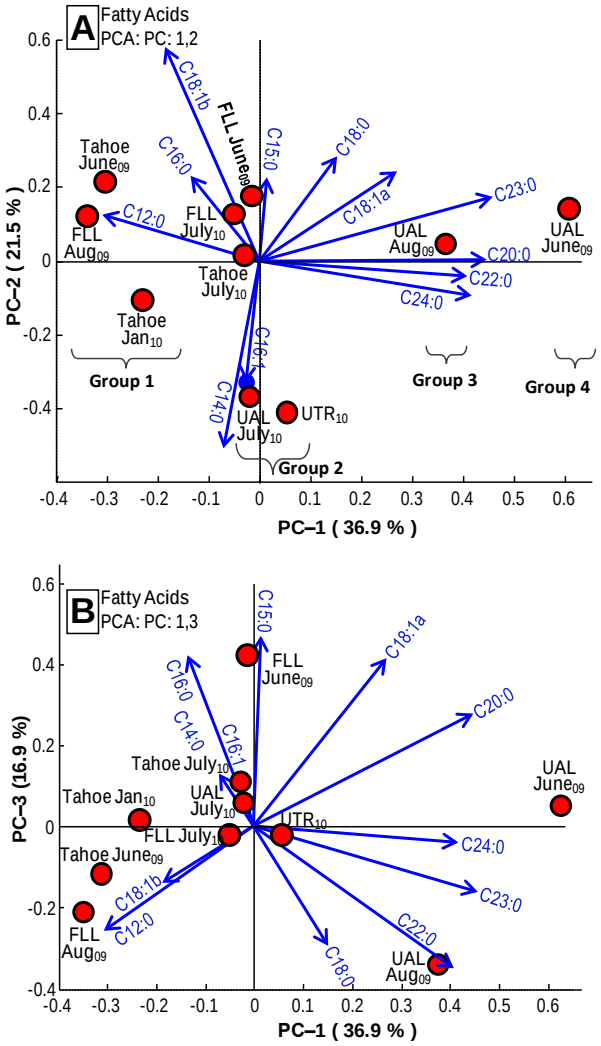


Figure 3.17: Biplots of PC1, PC2 (A) and PC1, PC3 (B) displaying range-normalized sample scores (circles) and component (variable) loadings (vectors) derived from PCA of sum-normalized, 0–to–1 transformed, FA data analyzed independently from the Lake Tahoe Basin SPE-DOM isolates.

Table 3.S1: Comparative overview of key features characterizing the contrasting aquatic settings sampled in this study. Data have been compiled from the many referenced sources.

	Lake Tahoe	Fallen Leaf Lake	Upper Angora Lake	Upper Truckee River
Surface Area (km^2)	495 [8]	5.67 [1]	0.057 [&]	-
Elevation (m)	1,897 [8]	1,942	2,270 [&]	1,897–3,067 [12]
Max depth (m)	501 [8]	120 [2]	15 [14]	3 [11]
Max length (km)	35 [8]	5.0 [&]	0.30 [&]	37 [21]
Max width (km)	19 [8]	1.6 [&]	0.30 [&]	1 - 30 [&]
Average depth (m)	305 [8]	53 [2]	10 [#]	1.4 [11]
Approx. hydraulic residence time (yrs)	650 [4]	8 [7]	1 [#]	0.003 [#]
Volume (km^3)	156 [15]	0.30	0.00057 [^]	-
Drainage basin (km^2)**	803	39 [10]	0.85 [&]	146 [6]
Drainage basin to lake surface area	1.6 [7]	6.9	15 [^]	–
Lake trophic level	ultra-oligotrophic	meso-oligotrophic [2]	oligotrophic[#]	–
Type	sub-monomictic[⊩]	sub-monomictic[⊩]	polymictic	–
Biology	diatoms (40-60%), chrysophytes (10-30%), trace cyanobacteria [17]	diatom-dominated, no blue-green algae [1]	–	–
Nitrate (µM)	1.3 [17]	0.071[3]–0.286 [1]*	–	0.42 [17]
Ammonium (µM)	trace to absent [18]	0.229 [1]–0.442 [1]*	–	–
Phosphate (µM)	0.06 - 0.10 [19]	0.161[5]–0.258 [1]*	–	0.1 [17]
Geologic setting	Tectonic; down-dropped graben [9]	recessional glacial moraine-bounded [20]	confined to glacial cirque	alpine/sub-alpine stream
Sampling Location	39.110350 -120.032158	38.88392 - 120.06592	38.862618 -120.067520	38.922879 -119.991657

1. (Lico 2004) 2. (Al-Chokhachy et al. 2009) 3. (Fuller 1975) 4. (Jassby et al. 1999) 5. (Hill 1992) 6. (Rowe et al. 2002) 7. (Kleppe 2005) 8. (USGS 2012) 9. (Herdendorf 1982) 10. (Kleppe et al. 2011) 11. (USGS 2013) 12. (USGS 2000) 13. (Oris et al. 1999) 14. (Needham and Sumner 1942) 15. (Alin and Johnson 2007) 16. (Schuster and Grismer 2004) 17. (TERC 2012) 18. (Paerl et al. 1975) 19. (Chang et al. 1992) 20. (Brothers et al. 2009) 21. (Schlesinger and Holst 2000) *Represents the range in medians observed across several studies. ** Excludes lake area. [&]Determined in Google Earth. [#] Estimated. [^] Calculated. [⊩] deep, but incomplete annual vertical mixing. [¯] Data not found or not applicable.

Table 3.S2: Number of peaks identified at $S/N_{\geq 10}$ and $S/N_{\geq 100}$ for whole chromatograms (All) and for quadrant-partitioned chromatograms. The sum of the peaks identified in quadrants 1–4 ($\Sigma Q_{1,2,3,4}$) always exceeded by a small amount (Σ/All) the number of peaks found when a chromatogram was processed without partitioning, likely because peaks could not be automatically joined across quadrants when processing was done by quadrant. At either S/N threshold, the number of peaks identified per quadrant consistently broke down as follows: $Q_3 > Q_1 > Q_4 > Q_2$.

SPE–DOM Isolate	**$S/N_{\geq 10}$**							**$S/N_{\geq 100}$**						
	All	Q_1	Q_2	Q_3	Q_4	$\Sigma Q_{1,2,3,4}$	Σ/All‡(%)	All	Q_1	Q_2	Q_3	Q_4	$\Sigma Q_{1,2,3,4}$	Σ/All‡(%)
Tahoe June 2009	6250	2095	186	3165	894	6340	1.44	1795	494	22	1137	156	1809	0.78
Tahoe Jan 2010	6285	2098	285	3220	964	6567	4.49	2113	606	30	1274	230	2140	1.28
Tahoe July 2010	7152	2576	488	3217	1165	7446	4.11	2340	778	44	1275	260	2357	0.73
FLL Aug 2009	8577	2925	924	3448	1583	8880	3.53	3232	1101	145	1568	481	3295	1.95
FLL June 2009	8538	2890	1095	3400	1533	8918	4.45	3397	1158	204	1546	536	3444	1.38
FLL July 2010	8737	2961	1340	3289	1457	9047	3.55	3554	1248	288	1518	553	3607	1.49
UAL Aug 2009	9707	3067	1573	3518	1883	10041	3.44	4012	1304	380	1681	737	4102	2.24
UAL June 2009	9022	2873	1424	3421	1633	9351	3.65	3528	1224	304	1502	566	3596	1.93
UAL July 2010	9532	3008	1402	3606	1872	9888	3.73	3884	1323	312	1638	679	3952	1.75
UTR July 2010	8807	2714	1204	3376	1810	9104	3.37	3357	1132	247	1413	640	3432	2.23

SPE–DOM Sample	**$S/N_{10\text{-}100}$**						
	All	Q_1	Q_2	Q_3	Q_4	$\Sigma Q_{1,2,3,4}$	Σ/All‡(%)
Tahoe June 2009	4455	1601	164	2028	738	4531	1.71
Tahoe Jan 2010	4172	1492	255	1946	734	4427	6.11
Tahoe July 2010	4812	1798	444	1942	905	5089	5.76
FLL Aug 2009	5345	1824	779	1880	1102	5585	4.49
FLL June 2009	5141	1732	891	1854	997	5474	6.48
FLL July 2010	5183	1713	1052	1771	904	5440	4.96
UAL Aug 2009	5695	1763	1193	1837	1146	5939	4.28
UAL June 2009	5494	1649	1120	1919	1067	5755	4.75
UAL July 2010	5648	1685	1090	1968	1193	5936	5.10
UTR July 2010	5450	1582	957	1963	1170	5672	4.07

‡ $= 100\% \times (\Sigma(Q1\text{–}4)/\text{All}\text{–}1)$

Table 3.S3: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 100}$** across **all** quadrants.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	1795	2113	2340	3232	3397	3554	4012	3528	3884	3357
Tahoe June 2009	1795	100	81.68	79.23	67.75	64.64	62.33	58.00	62.37	60.08	64.36
Tahoe Jan 2010	2113	81.68	100	81.92	75.43	71.62	70.58	65.37	69.28	66.33	70.86
Tahoe July 2010	2340	79.23	81.92	100	79.76	76.24	74.21	69.46	73.42	71.72	73.90
FLL Aug 2009	3232	67.75	75.43	79.76	100	86.83	84.32	81.23	83.31	83.45	83.78
FLL June 2009	3397	64.64	71.62	76.24	86.83	100	87.07	85.09	87.97	86.25	86.11
FLL July 2010	3554	62.33	70.58	74.21	84.32	87.07	100	86.28	87.35	87.12	86.44
UAL Aug 2009	4012	58.00	65.37	69.46	81.23	85.09	86.28	100	86.31	87.77	85.66
UAL June 2009	3528	62.37	69.28	73.42	83.31	87.97	87.35	86.31	100	88.53	86.91
UAL July 2010	3884	60.08	66.33	71.72	83.45	86.25	87.12	87.77	88.53	100	87.70
UTR 2010	3357	64.36	70.86	73.90	83.78	86.11	86.44	85.66	86.91	87.70	100

Table 3.S4: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 10}$** across **all** quadrants.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	6250	6285	7152	8577	8538	8737	9707	9022	9532	8807
Tahoe June 2009	6250	100	89.33	85.88	80.22	78.39	76.81	73.81	76.06	75.68	77.36
Tahoe Jan 2010	6285	89.33	100	88.31	82.94	81.89	80.08	76.99	79.24	78.07	80.36
Tahoe July 2010	7152	85.88	88.31	100	87.20	86.46	84.56	81.54	83.69	82.41	83.89
FLL Aug 2009	8577	80.22	82.94	87.20	100	90.70	88.91	87.69	89.29	89.12	88.97
FLL June 2009	8538	78.39	81.89	86.46	90.70	100	90.92	89.54	91.24	90.46	90.49
FLL July 2010	8737	76.81	80.08	84.56	88.91	90.92	100	89.69	90.91	89.99	90.06
UAL Aug 2009	9707	73.81	76.99	81.54	87.69	89.54	89.69	100	90.62	91.13	90.02
UAL June 2009	9022	76.06	79.24	83.69	89.29	91.24	90.91	90.62	100	91.45	90.74
UAL July 2010	9532	75.68	78.07	82.41	89.12	90.46	89.99	91.13	91.45	100	91.44
UTR 2010	8807	77.36	80.36	83.89	88.97	90.49	90.06	90.02	90.74	91.44	100

Table 3.S5: Pair-wise alignment scores for peaks identified at **S/N$_{10\text{-}100}$** across **all** quadrants.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	4455	4172	4812	5345	5141	5183	5695	5494	5648	5450
Tahoe June 2009	4455	100	92.80	88.85	86.61	85.83	84.85	82.86	83.39	84.45	84.12
Tahoe Jan 2010	4172	92.80	100	91.47	87.15	87.96	85.84	84.20	85.06	85.23	85.76
Tahoe July 2010	4812	88.85	91.47	100	91.29	92.35	90.67	88.84	89.54	88.78	89.44
FLL Aug 2009	5345	86.61	87.15	91.29	100	93.15	91.87	91.94	93.02	92.79	92.14
FLL June 2009	5141	85.83	87.96	92.35	93.15	100	93.51	92.58	93.37	93.30	93.29
FLL July 2010	5183	84.85	85.84	90.67	91.87	93.51	100	92.06	93.27	91.96	92.41
UAL Aug 2009	5695	82.86	84.20	88.84	91.94	92.58	92.06	100	93.52	93.47	92.90
UAL June 2009	5494	83.39	85.06	89.54	93.02	93.37	93.27	93.52	100	93.39	93.15
UAL July 2010	5648	84.45	85.23	88.78	92.79	93.30	91.96	93.47	93.39	100	93.89
UTR 2010	5450	84.12	85.76	89.44	92.14	93.29	92.41	92.90	93.15	93.89	100

Table 3.S6: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 100}$** in **Quadrant 1**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	494	606	778	1101	1158	1248	1304	1224	1323	1132
Tahoe June 2009	494	100	78.73	75.63	60.31	57.51	55.22	52.17	55.06	52.94	58.43
Tahoe Jan 2010	606	78.55	100	78.03	69.71	66.55	63.86	62.09	64.04	60.96	66.74
Tahoe July 2010	778	73.58	77.89	100	79.51	76.03	72.95	70.80	72.43	71.49	74.03
FLL Aug 2009	1101	60.31	69.60	79.51	100	87.38	84.38	83.49	84.39	84.08	83.65
FLL June 2009	1158	57.51	66.55	76.03	87.38	100	86.28	85.62	89.00	86.34	85.94
FLL July 2010	1248	55.22	63.97	72.95	84.38	86.28	100	85.66	86.57	87.05	85.38
UAL Aug 2009	1304	52.50	62.09	70.80	83.33	85.62	85.66	100	87.58	86.94	86.04
UAL June 2009	1224	55.06	64.04	72.53	84.39	89.00	86.17	87.58	100	88.97	86.76
UAL July 2010	1323	52.94	61.38	71.49	83.91	86.34	87.05	86.94	88.97	100	86.27
UTR 2010	1132	58.67	66.74	73.93	83.56	85.85	85.38	86.04	86.76	86.27	100

Table 3.S7: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 100}$** in **Quadrant 2**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	22	30	44	145	204	288	380	304	312	247
Tahoe June 2009	22	100	57.69	48.48	25.15	15.04	12.26	9.95	11.04	12.57	15.61
Tahoe Jan 2010	30	57.69	100	51.35	33.14	21.37	18.87	14.15	16.17	17.54	21.66
Tahoe July 2010	44	48.48	51.35	100	42.33	28.23	22.89	19.34	23.56	23.60	26.12
FLL Aug 2009	145	25.15	33.14	42.33	100	69.91	59.12	49.14	54.79	58.21	62.76
FLL June 2009	204	15.04	21.37	28.23	69.91	100	73.98	65.41	73.23	71.32	73.61
FLL July 2010	288	12.26	18.87	22.89	59.12	73.98	100	77.54	76.35	76.33	72.52
UAL Aug 2009	380	9.95	14.15	19.34	49.14	65.41	77.25	100	78.07	79.19	72.41
UAL June 2009	304	11.04	16.17	23.56	54.79	73.23	76.35	78.07	100	82.14	76.95
UAL July 2010	312	12.57	17.54	23.60	58.21	71.32	76.33	79.19	82.14	100	81.22
UTR 2010	247	15.61	21.66	26.12	62.76	73.61	72.52	72.41	76.95	81.22	100

Table 3.S8: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 100}$** in **Quadrant 3**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	1137	1274	1275	1568	1546	1518	1681	1502	1638	1413
Tahoe June 2009	1137	100	84.86	85.16	79.33	79.39	78.72	74.95	79.27	77.26	80.63
Tahoe Jan 2010	1274	84.86	100	86.62	84.45	83.69	84.53	80.34	83.79	81.18	84.26
Tahoe July 2010	1275	85.16	86.62	100	85.19	85.15	84.21	81.12	84.48	83.21	84.38
FLL Aug 2009	1568	79.33	84.45	85.19	100	90.49	88.20	87.35	89.25	89.15	88.02
FLL June 2009	1546	79.39	83.69	85.15	90.49	100	89.16	88.63	91.01	89.57	88.54
FLL July 2010	1518	78.72	84.53	84.21	88.20	89.16	100	88.47	88.21	88.53	88.50
UAL Aug 2009	1681	74.95	80.34	81.12	87.35	88.63	88.47	100	88.09	89.42	86.75
UAL June 2009	1502	79.27	83.79	84.48	89.25	91.01	88.21	88.09	100	90.70	88.71
UAL July 2010	1638	77.26	81.18	83.21	89.15	89.57	88.53	89.42	90.70	100	88.50
UTR 2010	1413	80.63	84.26	84.38	88.02	88.54	88.50	86.75	88.71	88.50	100

Table 3.S9: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 100}$** in **Quadrant 4**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	156	230	260	481	536	553	737	566	679	640
Tahoe June 2009	156	100	66.32	64.90	44.90	38.15	38.36	32.03	37.40	34.25	36.43
Tahoe Jan 2010	230	66.32	100	73.88	59.07	51.96	52.36	43.85	50.25	46.20	49.43
Tahoe July 2010	260	64.90	73.88	100	65.32	57.54	57.32	49.15	54.72	51.12	53.56
FLL Aug 2009	481	44.90	59.07	65.32	100	78.86	78.72	72.25	75.26	74.31	76.00
FLL June 2009	536	38.15	51.96	57.54	78.86	100	82.64	79.34	85.12	81.65	82.82
FLL July 2010	553	38.36	52.36	57.32	78.72	82.64	100	80.00	83.47	82.63	83.15
UAL Aug 2009	737	32.03	43.85	49.15	72.25	79.34	80.00	100	81.96	87.43	84.68
UAL June 2009	566	37.40	50.25	54.72	75.26	85.12	83.47	81.96	100	85.46	83.58
UAL July 2010	679	34.25	46.42	51.12	74.31	81.65	82.63	87.43	85.46	100	86.58
UTR 2010	640	36.43	49.43	53.56	76.00	82.82	83.15	84.68	83.58	86.58	100

Table 3.S10: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 10}$** in **Quadrant 1**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	2095	2098	2576	2925	2890	2961	3067	2873	3008	2714
Tahoe June 2009	2095	100	88.15	85.72	80.88	80.44	78.88	78.26	80.23	78.23	81.80
Tahoe Jan 2010	2098	88.10	100	85.41	82.14	81.96	80.13	79.03	81.11	79.87	83.04
Tahoe July 2010	2576	85.81	85.32	100	87.91	88.58	85.61	85.88	87.98	85.78	87.49
FLL Aug 2009	2925	80.88	81.94	87.91	100	89.87	88.58	89.19	89.65	89.47	89.09
FLL June 2009	2890	80.44	81.96	88.44	89.87	100	89.76	90.01	90.86	90.61	89.65
FLL July 2010	2961	78.88	80.06	85.61	88.58	89.76	100	89.52	89.34	89.63	88.92
UAL Aug 2009	3067	78.23	79.03	85.88	88.92	90.01	89.52	100	90.03	90.80	88.84
UAL June 2009	2873	80.23	81.11	87.65	89.65	90.86	89.34	90.03	100	90.60	90.10
UAL July 2010	3008	78.23	79.75	85.78	89.43	90.61	89.63	90.80	90.60	100	89.34
UTR 2010	2714	81.76	83.04	87.07	89.13	89.26	88.92	88.84	90.10	89.34	100

Table 3.S11: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 10}$** in **Quadrant 2**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	186	285	488	924	1095	1340	1573	1424	1402	1204
Tahoe June 2009	186	100	66.24	48.66	32.25	27.32	23.59	20.35	21.99	23.30	25.90
Tahoe Jan 2010	285	66.24	100	63.65	45.99	39.71	34.46	30.14	32.18	33.43	37.88
Tahoe July 2010	488	48.66	63.65	100	63.46	58.50	51.53	45.90	48.85	50.16	54.73
FLL Aug 2009	924	32.25	45.99	63.46	100	81.72	76.15	70.48	74.02	75.92	78.01
FLL June 2009	1095	27.32	39.71	58.50	81.72	100	80.82	78.49	80.98	80.98	83.17
FLL July 2010	1340	23.59	34.46	51.53	76.15	80.82	100	82.46	83.29	84.10	82.70
UAL Aug 2009	1573	20.35	30.14	45.90	70.48	78.49	82.32	100	84.48	82.82	80.01
UAL June 2009	1424	21.99	32.18	48.85	74.02	80.98	83.29	84.48	100	86.20	83.26
UAL July 2010	1402	23.30	33.43	50.16	75.49	80.98	84.10	82.82	86.20	100	84.34
UTR 2010	1204	25.90	38.01	54.73	82.52	83.17	82.70	80.01	87.52	84.34	100

Table 3.S12: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 10}$** in **Quadrant 3**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	3165	3220	3217	3448	3400	3289	3518	3421	3606	3376
Tahoe June 2009	3165	100	90.59	90.25	89.37	88.93	89.34	87.77	88.22	88.38	88.40
Tahoe Jan 2010	3220	90.59	100	90.82	89.71	89.40	89.91	88.48	89.63	89.07	88.72
Tahoe July 2010	3217	90.25	90.82	100	89.90	90.16	90.65	89.03	89.48	89.37	88.88
FLL Aug 2009	3448	89.37	89.71	89.90	100	91.56	90.52	90.21	91.37	91.61	90.42
FLL June 2009	3400	88.93	89.40	90.16	91.56	100	91.31	91.41	91.95	91.92	91.12
FLL July 2010	3289	89.34	89.91	90.65	90.52	91.31	100	90.73	91.33	90.99	91.43
UAL Aug 2009	3518	87.77	88.48	89.03	90.21	91.41	90.73	100	91.22	91.52	90.92
UAL June 2009	3421	88.22	89.63	89.48	91.37	91.95	91.33	91.22	100	91.82	91.04
UAL July 2010	3606	88.38	89.07	89.37	91.61	91.92	90.99	91.52	91.82	100	92.04
UTR 2010	3376	88.40	88.72	88.88	90.42	91.12	91.43	90.92	91.04	92.04	100

Table 3.S13: Pair-wise alignment scores for peaks identified at **S/N$_{\geq 10}$** in **Quadrant 4**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	894	964	1165	1583	1533	1457	1883	1633	1872	1810
Tahoe June 2009	894	100	82.24	75.96	68.07	64.03	64.57	59.78	62.60	61.46	62.87
Tahoe Jan 2010	964	82.13	100	80.60	71.93	70.56	71.21	65.05	67.92	65.30	66.62
Tahoe July 2010	1165	75.96	81.26	100	78.38	77.02	76.58	71.52	75.55	71.85	72.07
FLL Aug 2009	1583	68.07	71.93	78.38	100	84.34	82.57	81.77	84.95	83.76	82.64
FLL June 2009	1533	64.03	70.56	77.02	84.34	100	86.02	84.31	87.05	84.05	82.98
FLL July 2010	1457	64.57	71.21	76.58	82.57	86.02	100	81.50	85.31	81.35	82.95
UAL Aug 2009	1883	59.78	65.05	71.52	81.77	84.31	81.50	100	86.01	87.51	87.25
UAL June 2009	1633	62.60	67.92	75.55	84.95	87.05	85.31	86.01	100	86.28	85.27
UAL July 2010	1872	61.46	64.88	71.85	83.76	84.05	81.35	87.51	86.28	100	88.43
UTR 2010	1810	62.87	66.62	72.07	82.64	82.98	82.95	87.25	85.27	88.43	100

Table 3.S14: Pair-wise alignment scores for peaks identified at **S/N$_{10\text{-}100}$** in **Quadrant 1**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	1601	1492	1798	1824	1732	1713	1763	1649	1685	1582
Tahoe June 2009	1601	100	91.50	89.50	90.45	91.81	91.31	92.21	93.54	92.21	93.75
Tahoe Jan 2010	1492	91.50	100	88.51	88.54	90.38	89.55	88.97	91.05	91.34	92.26
Tahoe July 2010	1798	90.38	88.45	100	92.27	95.47	92.91	94.69	97.01	94.40	95.09
FLL Aug 2009	1824	90.45	88.30	92.27	100	91.45	91.38	93.00	93.18	93.19	92.66
FLL June 2009	1732	91.81	90.38	95.24	91.45	100	92.19	93.10	92.16	93.71	92.21
FLL July 2010	1713	91.31	89.36	92.91	91.38	92.19	100	92.35	91.37	91.58	91.47
UAL Aug 2009	1763	91.97	88.97	94.69	92.67	93.10	92.35	100	91.85	93.74	90.88
UAL June 2009	1649	93.54	91.05	96.43	93.18	92.16	91.67	91.85	100	91.84	92.54
UAL July 2010	1685	92.21	90.90	94.40	93.25	93.71	91.58	93.74	91.84	100	91.64
UTR 2010	1582	93.56	92.26	94.50	92.78	91.61	91.47	90.88	92.54	91.64	100

Table 3.S15: Pair-wise alignment scores for peaks identified at **S/N$_{10\text{-}100}$** in **Quadrant 2**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	164	255	444	779	891	1052	1193	1120	1090	957
Tahoe June 2009	164	100	67.30	48.68	33.51	29.95	26.48	23.43	24.77	26.16	28.37
Tahoe Jan 2010	255	67.30	100	64.95	48.16	43.46	38.26	34.67	36.07	37.47	41.58
Tahoe July 2010	444	48.68	64.95	100	66.72	64.12	57.89	52.78	54.48	56.32	60.67
FLL Aug 2009	779	33.51	48.16	66.72	100	84.19	80.17	76.17	78.57	80.26	81.45
FLL June 2009	891	29.95	43.46	64.12	84.19	100	82.55	82.15	82.94	83.49	85.50
FLL July 2010	1052	26.48	38.26	57.89	80.17	82.55	100	83.92	85.17	86.27	85.42
UAL Aug 2009	1193	23.43	34.67	52.78	76.17	82.15	83.83	100	86.38	83.92	82.23
UAL June 2009	1120	24.77	36.07	54.48	78.57	82.94	85.17	86.38	100	87.33	84.93
UAL July 2010	1090	26.16	37.47	56.32	79.72	83.49	86.27	83.92	87.33	100	85.20
UTR 2010	957	28.37	41.75	60.67	86.98	85.50	85.42	82.23	90.32	85.20	100

Table 3.S16: Pair-wise alignment scores for peaks identified at **S/N$_{10\text{-}100}$** in **Quadrant 3**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	2028	1946	1942	1880	1854	1771	1837	1919	1968	1963
Tahoe June 2009	2028	100	94.06	93.35	96.32	95.52	96.76	97.13	94.20	96.10	93.36
Tahoe Jan 2010	1946	94.06	100	93.57	93.62	93.63	93.95	94.85	93.82	94.94	91.79
Tahoe July 2010	1942	93.35	93.57	100	93.41	93.89	95.50	95.21	93.08	93.96	91.98
FLL Aug 2009	1880	96.32	93.62	93.41	100	92.45	92.47	92.71	93.08	93.66	92.27
FLL June 2009	1854	95.52	93.63	93.89	92.45	100	93.13	93.85	92.71	93.88	93.11
FLL July 2010	1771	96.76	93.95	95.50	92.47	93.13	100	92.74	93.88	93.07	93.73
UAL Aug 2009	1837	97.13	94.85	95.21	92.71	93.85	92.74	100	93.88	93.35	94.32
UAL June 2009	1919	94.20	93.82	93.08	93.08	92.71	93.88	93.88	100	92.72	92.79
UAL July 2010	1968	96.10	94.94	93.96	93.66	93.88	93.07	93.35	92.72	100	94.79
UTR 2010	1963	93.36	91.79	91.98	92.27	93.11	93.73	94.32	92.79	94.79	100

Table 3.S17: Pair-wise alignment scores for peaks identified at **$S/N_{10\text{-}100}$** in **Quadrant 4**.

		Tahoe June 2009	Tahoe Jan 2010	Tahoe July 2010	FLL Aug 2009	FLL June 2009	FLL July 2010	UAL Aug 2009	UAL June 2009	UAL July 2010	UTR 2010
	peaks	738	734	905	1102	997	904	1146	1067	1193	1170
Tahoe June 2009	738	100	86.41	78.76	76.09	74.35	75.88	72.93	72.69	73.23	73.90
Tahoe Jan 2010	734	86.28	100	82.61	76.91	78.80	80.22	75.96	75.74	74.31	74.47
Tahoe July 2010	905	78.76	83.47	100	83.21	85.17	85.24	82.40	84.28	81.12	80.10
FLL Aug 2009	1102	76.09	76.91	83.21	100	86.99	84.55	86.92	89.63	88.54	85.92
FLL June 2009	997	74.35	78.80	85.17	86.99	100	87.95	87.26	88.08	85.39	83.06
FLL July 2010	904	75.88	80.22	85.24	84.55	87.95	100	82.44	86.35	80.59	82.84
UAL Aug 2009	1146	72.93	75.96	82.40	86.92	87.26	82.44	100	88.39	87.56	88.77
UAL June 2009	1067	72.69	75.74	84.28	89.63	88.08	86.35	88.39	100	86.73	86.19
UAL July 2010	1193	73.23	73.59	81.12	88.54	85.39	80.59	87.56	86.73	100	89.46
UTR 2010	1170	73.90	74.47	80.10	85.92	83.06	82.84	88.77	86.19	89.46	100

Table 3.S18: Cross-correlations (R^2) of proton functional group abundances across Lake Tahoe Basin samples. Some groups are highly correlated (e.g. acetophenone with benzylic, aldehydes and methoxy hydrogens) while others exhibit significant anti-correlation (e.g. methoxy with alkyl and acids).

R^2 values	Alkyl	Aliphatic	Benzylic	Acetophenone	Alcohol + Saccharides + Amines	Methoxy	Aromatics + Vinylics	Highly Substituted-, Hetero-aromatics	Aldehydes	Acids
Alkyl	1	0.00	-0.63	-0.71	-0.62	-0.79	-0.01	-0.71	-0.60	0.42
Aliphatic	0.00	1	0.17	0.19	0.07	-0.06	-0.40	-0.08	-0.12	-0.01
Benzylic	-0.63	0.17	1	0.85	0.28	0.46	-0.22	0.37	0.28	-0.58
Acetophenone	-0.71	0.19	0.85	1	0.53	0.54	-0.16	0.37	0.36	-0.54
Alcohol + Saccharides + Amines	-0.62	0.07	0.28	0.53	1	0.38	0.00	0.20	0.37	-0.10
Methoxy	-0.79	-0.06	0.46	0.54	0.38	1	0.00	0.79	0.86	-0.42
Aromatics + Vinylics	-0.01	-0.40	-0.22	-0.16	0.00	0.00	1	0.03	0.00	0.04
Highly Substituted-, Hetero-aromatics	-0.71	-0.08	0.37	0.37	0.20	0.79	0.03	1	0.66	-0.31
Aldehydes	-0.60	-0.12	0.28	0.36	0.37	0.86	0.00	0.66	1	-0.17
Acids	0.42	-0.01	-0.58	-0.54	-0.10	-0.42	0.04	-0.31	-0.17	1

Table 3.S19: ^{1}H-NMR integrations of CuO-hydrolysates expressed as sum-normalized percentages (%). ∑ = 100 %. Integrations were performed according to the indicated ppm chemical shift (δ) ranges. These ranges broadly correspond to the hydrogen chemical environment within, or proximate to, the functional group types listed.

DOM Isolate	Hydrogen Functional Group Environment										
	Alkyl[1]	Aliphatic[2]	Benzylic[3]	Acetophenone	Alcohols, Saccharides, Amines[4]	Methoxy	Aromatics, Vinylics	Highly Substituted or Heteroaromatics	Aldehydes	Acids	$\underline{Ad/Al_{NMR}}$
ppm shift	*0.8–1.4*	*1.5–1.8*	*1.8–2.4 2.7–3.2*	*2.4–2.7*	*3.4–3.6 3.9–4.8*	*3.6–3.9*	*5.2–8.3*	*8.3–9.0*	*9.5–10.5*	*10.5–13.5*	-
Tahoe June '09	36.98	9.76	20.31	7.61	4.89	1.80	15.68	1.33	0.43	1.20	2.81
Tahoe Jan '10	30.28	10.60	23.10	7.78	5.28	1.98	16.50	2.42	0.52	1.54	2.97
Tahoe July '10	28.96	10.20	22.14	7.95	6.03	2.14	20.10	1.54	0.37	0.58	1.55
FLL Aug '09	38.23	10.70	24.90	8.17	3.18	1.67	10.89	1.49	0.31	0.45	1.46
FLL June '09	21.99	11.48	32.13	9.94	6.63	3.18	11.48	2.27	0.64	0.26	0.41
FLL July '10	21.69	11.54	29.48	9.93	6.97	3.94	12.75	2.99	0.59	0.11	0.19
UAL Aug '09	19.49	10.62	28.84	9.75	5.88	4.30	16.74	3.64	0.64	0.10	0.15
UAL June '09	17.91	10.50	28.64	10.22	8.18	5.65	13.87	3.37	1.25	0.41	0.33
UAL July '10	21.26	9.89	28.68	9.20	5.77	5.41	15.35	3.21	0.96	0.28	0.29
UTR July '10	19.88	8.95	27.67	8.87	5.59	5.43	17.91	4.49	1.11	0.10	0.089

1. Methyl, methylene ($H–C_{>\gamma}–(O,N)$),
2. $H_n–C_{\beta,\gamma}–(O,N)$ or methyne
3. Benzylic, $H–C_{\beta}–O$
4. $H_n–C\alpha–(O,NR_3)$

Table 3.S20: Fatty acid (FA) concentrations quantified for SPE-DOM isolates.

DOM Isolate	Fatty Acid Concentration (ng L^{-1})															
	C12:0	C14:0	C15:0	C16:1	C16:0	C18:1a	C18:1b	C18:0	C20:0	C22:0	C23:0	C24:0	$\sum C_{12\text{-}15}$	$\sum C_{16\text{-}18}$	$\sum C_{20+}$	$\sum$FAs
Tahoe June '09	26.6	29.8	5.7	6	65.6	6.2	13.8	69.5	0	0	0	0	62.1	161.1	0	223
Tahoe Jan '10	47.3	140.1	10.9	12.2	149.6	11	7.6	64	4.3	3.6	0	0	198.3	244.4	7.9	451
Tahoe July '10	17.7	85.7	14.7	22.1	207.9	13.9	11.6	131.2	11	6.6	0	7.2	118.1	386.7	24.8	530
FLL Aug '09	46.5	41.1	5.7	12.3	65.4	6	16.9	38.4	0	3.6	0	0	93.3	139	3.6	236
FLL June '09	2.8	15.6	3.7	3.2	31.7	3.8	3	19.8	3	0	0	0	22.1	61.5	3	86.6
FLL July '10	24.8	92	9.1	26.2	192.7	20	37.6	130.8	9.4	8.1	0	7.3	125.9	407.3	24.8	558
UAL Aug '09	12.8	48.8	9.7	12.7	121.2	12.4	9	259	17.7	18.1	6.5	9.6	71.3	414.3	51.9	538
UAL June '09	11.1	36.6	6.9	9.6	64.9	11.9	8.8	49.2	9.6	8.2	4.9	10	54.6	144.4	32.7	232
UAL July '10	25.4	153.2	9.6	50.3	151.5	12.8	0	68.7	10.2	7.1	0	7.4	188.2	283.3	24.7	496
UTR July '10	33.3	327.7	14.7	35.1	134.2	21	0	92.7	13.8	11.4	0	13.9	375.7	283	39.1	698



Figure 3.S1: Photographs including the water bodies sampled in the upper Lake Tahoe Basin. Upper Angora Lake, Lower Angora Lake, Fallen Leaf Lake, and Lake Tahoe are viewable in (A) from bottom to top. Upper Angora Lake is shown in (B), Fallen Leaf Lake in (C), and the Upper Truckee River in (D). A–C were taken by Judith Hildinger and are used here with permission. D was taken by Jeff Hunt.

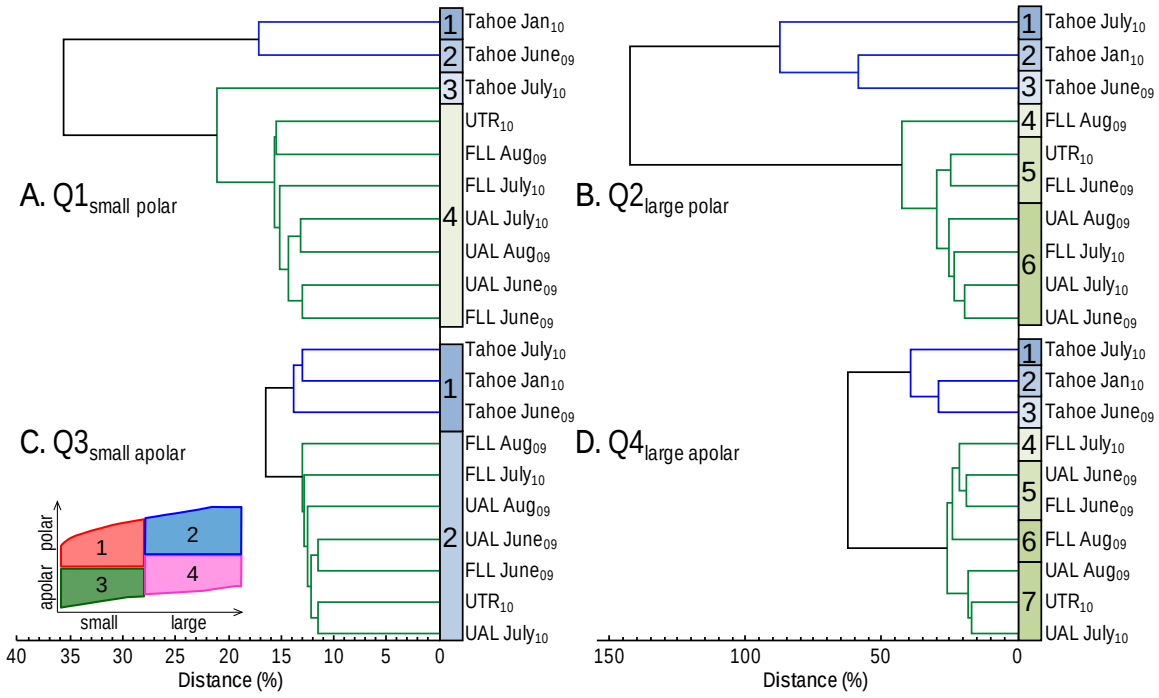


Figure 3.S2: Hierarchical clustering of alignment-derived chemical distances performed on quadrant-partitioned (A–D) peaks identified at $S/N_{\geq 10}$. Peaks with $S/N_{\geq 10}$ included those with a $S/N_{\geq 100}$. As such, the peaks aligned here represent those defined by $S/N_{10\text{-}100}$ + $S/N_{\geq 100}$.

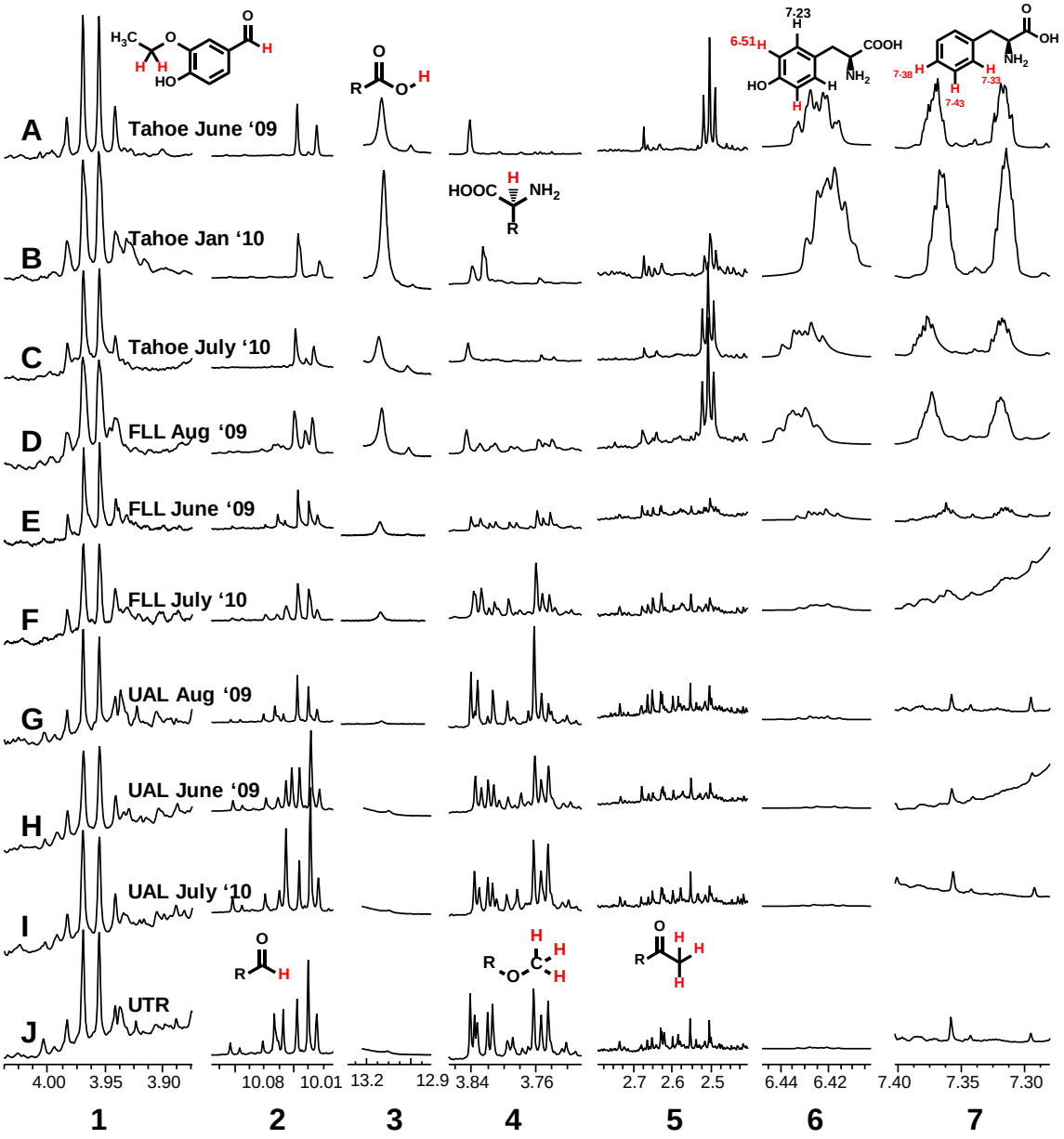


Figure 3.S3: ^{1}H-NMR (d_5-pyridine) for seven shift regions of interest that show variability between CuO-oxidized SPE-DOM isolates (A–J) that were otherwise unresolved prior to oxidation. All intensities are scaled to the height of ethyl vanillin's ethoxy group quartet (1), a standard spiked into the CuO-hydrolysate mixture following oxidation. Shift regions are shown for aldehydic and carboxylic protons in (2) and (3), respectively. Methoxy protons are shown in (4) and are most abundant in terrestrially influenced samples (F–J) while their abundances generally decrease from E→A where they are replaced by the aminic protons of amino acids. Acetophenone protons (5) are enriched in terrestrially influenced samples (F–J), however other functional groups replace them in (E–A), including one with a very distinct triplet (J = 7.4 Hz). Shift regions for aromatic protons in tyrosine and phenylalanine are shown in (6) and (7). These two amino acids are most abundant in A-D, are highest in B (Tahoe Jan), and are not resolvable in the more terrestrially influenced samples (F–J).

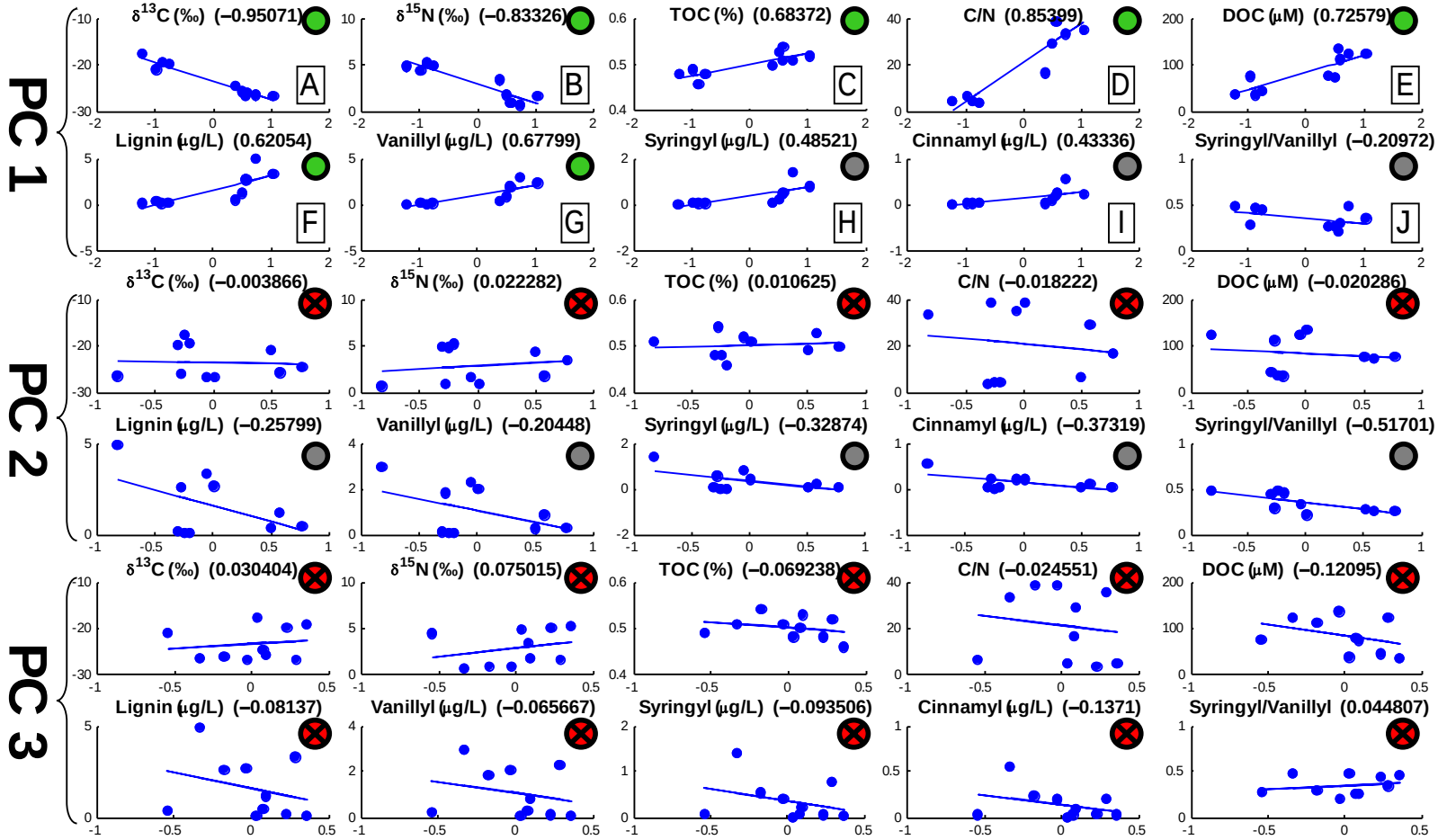


Figure 3.S4: Scaling of NMR-derived PCs 1,2 and 3 (*x*-axis) with other measured elemental, isotopic, and organic geochemical parameters (*y*-axis), including $\delta^{13}C$ (A), $\delta^{15}N$ (B), TOC (C), C/N (D), DOC (E), ∑ lignin monomers (F), ∑ vanillyl monomers (G), ∑ syringyl monomers (H), ∑ cinnamyl monomers (I), and S/V (J). R^2's are in parentheses. Green circle = $|R^2| \geq 0.6$, grey circle = $0.2 < |R^2| < 0.6$, red circle = $|R^2| < 0.2$.

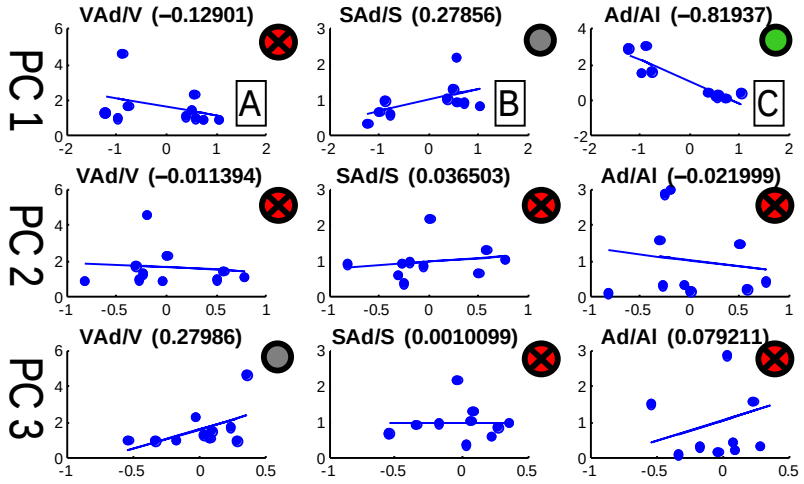


Figure 3.S5: Scaling of NMR-derived PCs 1, 2 and 3 with VAd/V (A), SAd/S (B), and Ad/Al_{NMR} (C) where Ad/Al_{NMR} is of the bulk hydrolysate as determined by ^{1}H-NMR. R^2 values are in parentheses. The most significant relationship is between PC1 and Ad/Al_{NMR}. SAd/S is weakly correlated ($R^2 = 0.28$) with PC1 while VAd/V ($R^2 = 0.28$) is weakly correlated with PC3. Green circle = $|R^2| \geq 0.6$, grey circle = $0.2 < |R^2| < 0.6$, red circle = $|R^2| < 0.2$.

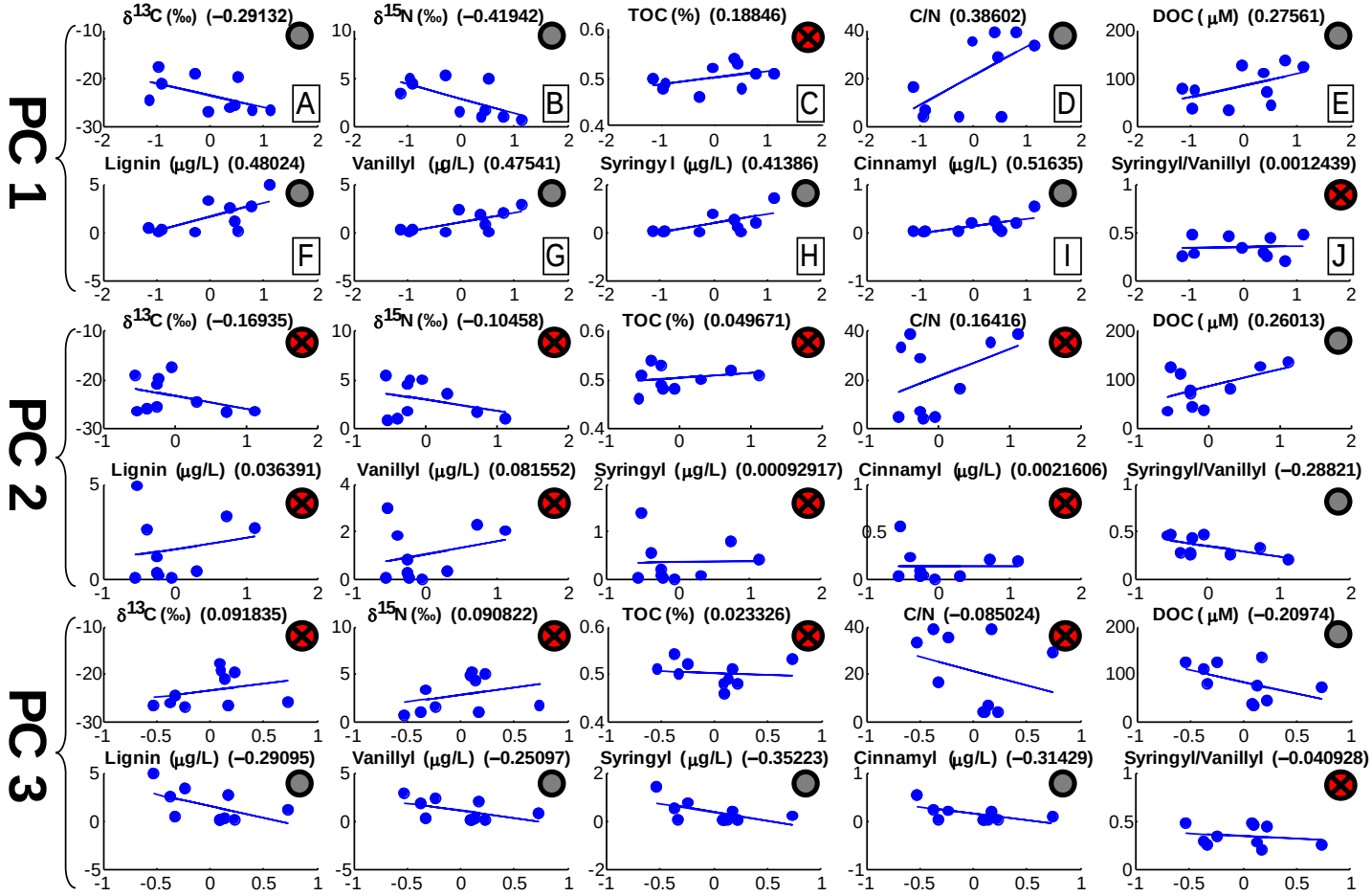


Figure 3.S6: Scaling of FA-derived PCs 1, 2 and 3 (*x*-axis) with other measured elemental, isotopic, and organic geochemical parameters (*y*-axis), including $\delta^{13}C$ (A), $\delta^{15}N$ (B), TOC (C), C/N (D), DOC (E), ∑ lignin monomers (F), ∑ vanillyl monomers (G), ∑ syringyl monomers (H), ∑ cinnamyl monomers (I), and S/V (J). R^2's are in parentheses. Grey circle = $0.2 < |R^2| < 0.5$, red circle = $|R^2| < 0.5$.

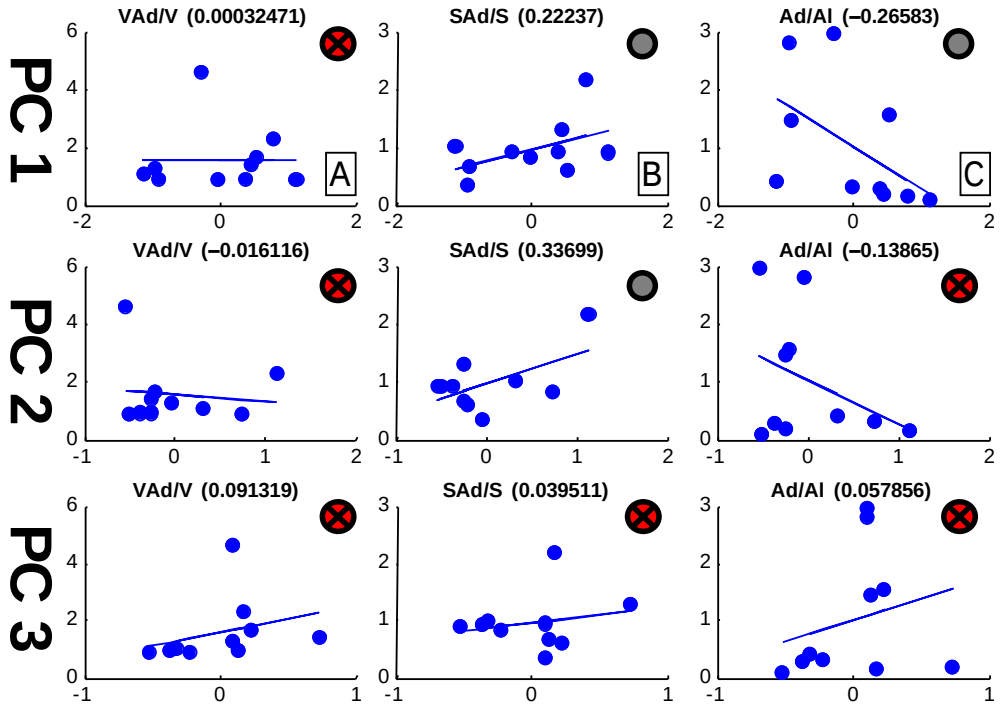


Figure 3.S7: Scaling of FA-derived PCs 1, 2 and 3 with VAd/V (A), SAd/S (B), and Ad/Al_{NMR} (C) where Ad/Ad is of the bulk hydrolysate as determined by ^{1}H-NMR. R^2 values are in parentheses. No strong correlations are present, although some correlations of intermediate strength are present, such as for PC2 SAd/S (R^2 = 0.34), PC1 SAd/S (0.22), and PC1 Ad/Al (–0.27). Grey circle = 0.2 < $|R^2|$ < 0.5, red circle = $|R^2|$ < 0.5.

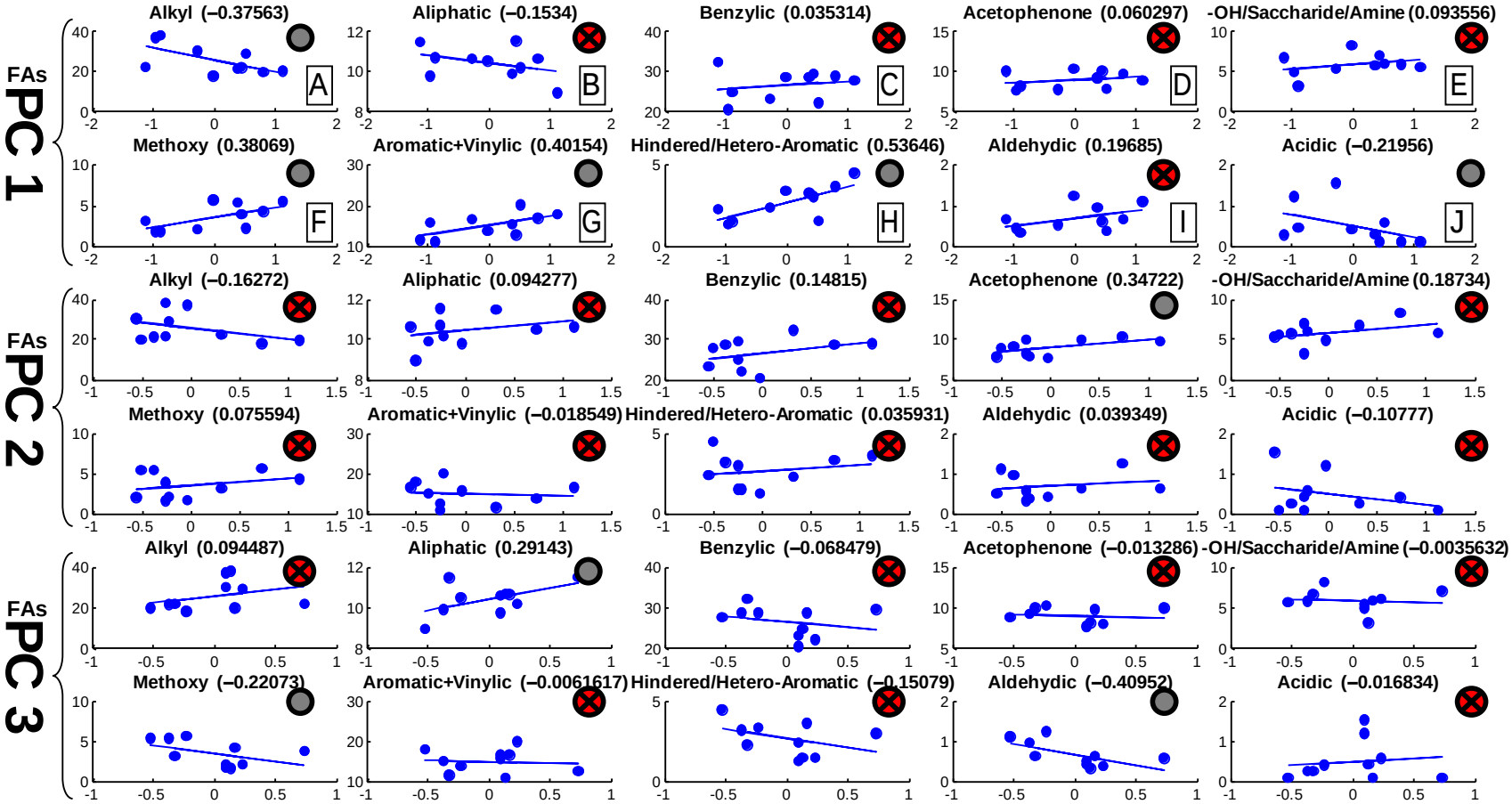


Figure 3.S8: Scaling of FA-derived PCs 1, 2 and 3 (*x*-axis) with ^{1}H-NMR sum-normalized functional group abundances (*y*-axis), for alkyl (A), aliphatic (B), benzylic (C), acetophenone (D), saccharide/alcohol/amine (E), methoxy (F), aromatic/vinylic (G), hindered/hetero-aromatic (H), aldehydic (I), and acid (J) protons. R^2's are in parentheses. No strong relationships are observed. Grey circle = $0.2 < |R^2| < 0.5$, red circle = $|R^2| < 0.5$.

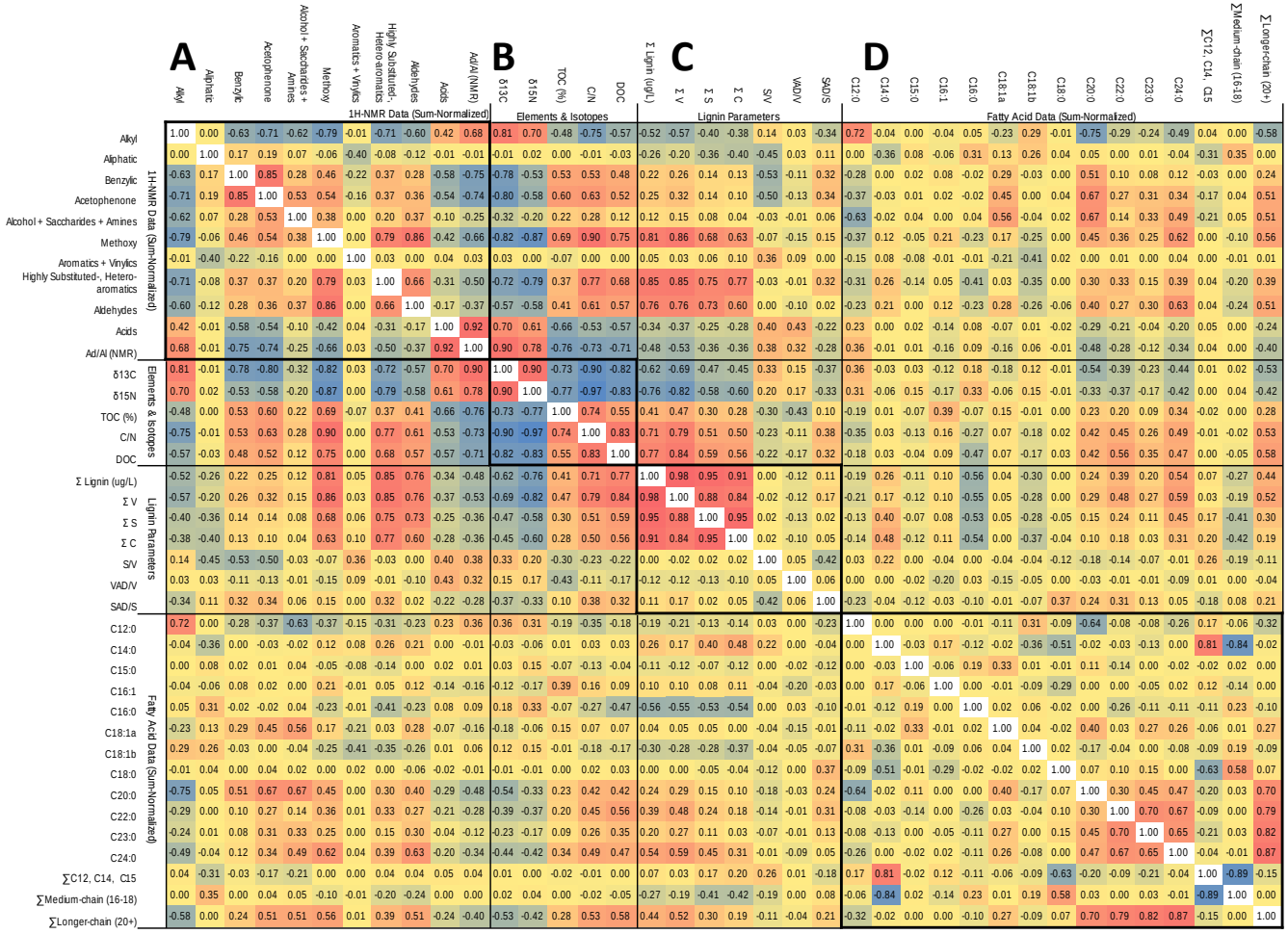


Figure 3.S9: Heat-plot of correlation coefficients (R^2) obtained from cross-correlation of all measured sample parameters for the upper Tahoe Basin SPE-DOM data set. ^{1}H-NMR (A), elemental and isotopic (B), lignin (C), and fatty acid data (D) are demarcated by A, B, C, and D, respectively. The color scale grades from reds (max R^2=0.98), to blues (min R^2= –0.97) according to the value of R^2.

3.7 References

Aiken, G., D. McKnight, K. A. Thorn, and E. M. Thurman. 1992. Isolation of hydrophilic organic acids from water using nonionic macroporous resins. Organic Geochemistry **18**: 567–573.

Al-Chokhachy, R., M. Peacock, L. G. Heki, and G. Thiede. 2009. Evaluating the Reintroduction Potential of Lahontan Cutthroat Trout in Fallen Leaf Lake, California. North American Journal of Fisheries Management **29**: 1296–1313.

Alin, S. R., and T. C. Johnson. 2007. Carbon cycling in large lakes of the world: A synthesis of production, burial, and lake-atmosphere exchange estimates. Global Biogeochemical Cycles **21**: 12.

Azadfallah, M., S. A. Mirshokraei, and A. J. Latibari. 2008. Photodegradation of acidolysis lignin from BCMP. Molecules **13**: 3129–39.

Benner, R., J. Pakulski, M. McCarthy, J. I. Hedges, and P. G. Hatcher. 1992. Bulk chemical characteristics of dissolved organic matter in the ocean. Science **255**: 1561–1564.

Brothers, D. S., G. M. Kent, N. W. Driscoll, S. B. Smith, R. Karlin, J. A. Dingler, A. J. Harding, G. G. Seitz, and J. M. Babcock. 2009. New constraints on deformation, slip rate, and timing of the most recent earthquake on the West Tahoe-Dollar Point Fault, Lake Tahoe Basin, California. Bulletin of the Seismological Society of America **99**: 499–519.

Chang, C. C. Y., J. S. Kuwabara, and S. P. Pasilis. 1992. Phosphate and iron limitation of phytoplankton biomass in Lake Tahoe. Canadian Journal of Fisheries and Aquatic Sciences **49**: 1206–1215.

Chang, H. M., and G. G. Allan. 1971. Oxidation, p. 916. *In* K.V. Sarkanen and C.H. Ludwig [1st ed], Lignins: Occurrence, Formation, Structure, and Reactions. Wiley-Interscience.

Defays, D. 1977. An efficient algorithm for a complete link method. The Computer Journal **20**: 364–366.

Demarco, P. V., E. Farkas, D. Doddrell, B. L. Mylari, and E. Wenkert. 1968. Pyridine-induced solvent shifts in the nuclear magnetic resonance spectra of hydroxylic compounds. Journal of the American Chemical Society **90**: 5480–5486.

Dickens, A. F., J. A. Gudeman, Y. Gélinas, J. A. Baldock, W. Tinner, F. S. Hu, and J. I. Hedges. 2007. Sources and distribution of CuO-derived benzene carboxylic acids in soils and sediments. Organic Geochemistry **38**: 1256–1276.

Doebbe, A., M. Keck, M. La Russa, J. H. Mussgnug, B. Hankamer, E. Tekçe, K. Niehaus, and O. Kruse. 2010. The interplay of proton, electron, and metabolite supply for photosynthetic H_2 production in Chlamydomonas reinhardtii. The Journal of Biological Chemistry **285**: 30247–30260.

Druffel, E. R. M., P. M. Williams, J. E. Bauer, and J. R. Ertel. 1992. Cycling of dissolved and particulate organic matter in the open ocean. Journal of Geophysical Research **97**: 15639–15659.

Eglinton, G., and D. H. Hunneman. 1968. Gas chromatographic-mass spectrometric studies of long chain hydroxy acids - I. The constituent cutin acids of apple cuticle. Phytochemistry **7**: 313–322.

Falkowski, P., R. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hógberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffen. 2000. The global carbon cycle: a test of our knowledge of earth as a system. Science **290**: 291–296.

Flerus, R., O. J. Lechtenfeld, B. P. Koch, S. L. McCallister, P. Schmitt-Kopplin, R. Benner, K. Kaiser, and G. Kattner. 2012. A molecular perspective on the ageing of marine dissolved organic matter. Biogeosciences **9**: 1935–1955.

Fuller, R. H. 1975. Selected water-quality data from Fallen Leaf Lake, El Dorado County, California, USGS, June through October 1974.

Goldberg, S. J., G. I. Ball, B. Allen, G. Schladow, A. Simpson, H. Soong, Masoom, and L. I. Aluwihare. Seasonal variation in the source and composition of solid phase extracted DOM from Lake Tahoe and shorter residence time lakes in the Sierra Nevada.

Goñi, M. A. 1992. The use of CuO reaction products to characterize organic matter in the marine environment. University of Washington.

Goñi, M. A., and J. I. Hedges. 1992. Lignin dimers: structures, distribution, and potential geochemical applications. Geochimica et Cosmochimica Acta **56**: 4025–4043.

Goñi, M. A., and J. I. Hedges. 1995. Sources and reactivities of marine-derived organic matter in coastal sediments as determined by alkaline CuO oxidation. Geochimica et Cosmochimica Acta **59**: 2965–2981.

Goñi, M. A., and S. Montgomery. 2000. Alkaline CuO oxidation with a microwave digestion system: lignin analyses of geochemical samples. Analytical Chemistry **72**: 3116–3121.

Gough, M. A., and S. J. Rowland. 1990. Characterization of unresolved complex mixtures of hydrocarbons in petroleum. Nature **344**: 648–650.

Guillemette, F., P. A. del Giorgio, and S. M. Leigh. 2013. Differentiating the degradation dynamics of algal and terrestrial carbon within complex natural dissolved organic carbon in temperate lakes. Journal of Geophysical Research: Biogeosciences **118**: 1–11.

Hansell, D. A., and C. A. Carlson. 1998. Deep ocean gradients in the concentration of dissolved organic carbon. Nature **395**: 263–266.

Hedges, J. I., and J. R. Ertel. 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. Analytical Chemistry **54**: 174–178.

Herdendorf, C. E. 1982. Large Lakes of the World. Journal of Great Lakes Research **8**: 379–412.

Hernes, P. J., and R. Benner. 2002. Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. Deep Sea Research Part I: Oceanographic Research Papers **49**: 2119–2132.

Hernes, P. J., and R. Benner. 2003. Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. Journal of Geophysical Research **108**: 3291.

Hertkorn, N., M. Frommberger, M. Witt, B. P. Koch, P. Schmitt-Kopplin, and E. M. Perdue. 2008. Natural organic matter and the event horizon of mass spectrometry. Analytical Chemistry **80**: 8908–19.

Hertkorn, N., M. Harir, B. P. Koch, B. Michalke, and P. Schmitt-kopplin. 2013. Functions of natural organic matter in changing environment, p. 13–17. *In* J. Xu, J. Wu, and Y. He [eds.], Functions of Natural Organic Matter in Changing Environment. Springer Netherlands.

Hertkorn, N., C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt, and P. Schmitt-Kopplin. 2007. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. Analytical and Bioanalytical Chemistry **389**: 1311–1127.

Hill, K. J. 1992. 1992 Annual water quality report: Zephyr Cove, NV, Tahoe Regional Planning Agency.

Jassby, A. D., C. R. Goldman, J. E. Reuter, and R. C. Richards. 1999. Origins and scale-dependence of temporal variability in the transparency of Lake Tahoe, California - Nevada. Limnology and Oceanography **44**: 282–294.

Jean-Marie, D. D. 2004. Comprehensive 2-D GC provides high-performance separations in terms of selectivity, sensitivity, speed and structure. Analytical Chemistry **76**: 167A–174A.

Jiao, N., G. J. Herndl, D. A. Hansell, R. Benner, G. Kattner, S. W. Wilhelm, D. L. Kirchman, M. G. Weinbauer, T. Luo, F. Chen, and F. Azam. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nature Reviews Microbiology **8**: 593–599.

Kallio, H., R. Nieminen, S. Tuomasjukka, and M. Hakala. 2006. Cutin composition of five finnish berries. Journal of Agricultural and Food Chemistry **54**: 457–462.

Katajamaa, M., J. Miettinen, and M. Orešič. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics **22**: 634–636.

Kiefl, J., C. Cordero, L. Nicolotti, P. Schieberle, S. E. Reichenbach, and C. Bicchi. 2012. Performance evaluation of non-targeted peak-based cross-sample analysis for comprehensive two-dimensional gas chromatography-mass spectrometry data and application to processed hazelnut profiling. Journal of chromatography. A **1243**: 81–90.

Kleppe, J. ., D. S. Brothers, G. M. Kent, F. Biondi, S. Jensen, and N. W. Driscoll. 2011. Duration and severity of Medieval drought in the Lake Tahoe Basin. Quaternary Science Reviews **30**: 3269–3279.

Kleppe, J. A. 2005. A study of ancient trees rooted 36.5m (120') below the surface level of Fallen Leaf Lake, California. Journal of the Nevada Water Resources Association 29–40.

Koch, B. P., M. Witt, R. Engbrodt, T. Dittmar, and G. Kattner. 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Geochimica et Cosmochimica Acta **69**: 3299–3308.

Koh, Y., K. K. Pasikanti, C. W. Yap, and E. C. Y. Chan. 2010. Comparative evaluation of software for retention time alignment of gas chromatography/time-of-flight mass spectrometry-based metabonomic data. Journal of Chromatography A **1217**: 8308–8316.

Krishnan, P., N. J. Kruger, and R. G. Ratcliffe. 2005. Metabolite fingerprinting and profiling in plants using NMR. Journal of Experimental Botany **56**: 255–65.

Leenheer, J. A. 2007. Progression from model structures to molecular structures of natural organic matter components. Annals of Environmental Science **1**: 57–68.

Lico, M. 2004. Nutrient Concentrations in Upper and Lower Echo, Fallen Leaf, Spooner, and Marlette Lakes and Associated Outlet Streams, California and Nevada, 2002–03, USGS.

Lommen, A. 2009. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Analytical Chemistry **81**: 3079–3086.

Lommen, A., H. J. van der Kamp, H. J. Kools, M. K. van der Lee, G. van der Weg, and H. G. J. Mol. 2012. metAlignID: a high-throughput software tool set for automated detection of trace level contaminants in comprehensive LECO two-dimensional gas chromatography time-of-flight mass spectrometry data. Journal of Chromatography. A **1263**: 169–78.

Lu, H., W. B. Dunn, H. Shen, D. B. Kell, and Y. Liang. 2008. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. TrAC Trends in Analytical Chemistry **27**: 215–227.

Luedemann, A., K. Strassburg, A. Erban, and J. Kopka. 2008. TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics **24**: 732–7.

Masoom, H., D. Courtier-Murias, H. Farooq, R. Soong, M. J. Simpson, W. Maas, R. Kumar, M. Monette, H. Stronks, and A. J. Simpson. 2013. Rapid estimation of nuclear magnetic resonance experiment time in low-concentration environmental samples. Environmental Toxicology and Chemistry / SETAC **32**: 129–36.

Meissner-Roloff, R. J., G. Koekemoer, R. M. Warren, and D. T. Loots. 2012. A metabolomics investigation of a hyper- and hypo-virulent phenotype of Beijing lineage M. tuberculosis. Metabolomics **8**: 1194–1203.

Melillo, G., G. W. Cox, D. Radzioch, and L. Varesio. 1993. Picolinic acid, a catabolite of L-tryptophan, is a costimulus for the induction of reactive nitrogen intermediate production in murine macrophages. The Journal of Immunology **150**: 4031–4040.

Meyers, P. A., and R. Ishiwatari. 1995. Organic Matter Accumulation Records in Lake Sediments, p. 279–328. *In* Physics and chemistry of lakes. Springer Berlin Heidelberg.

Murtagh, F. 1983. A survey of recent advances in hierarchical clustering algorithms. The Computer Journal **26**: 354–359.

Needham, P. R., and F. K. Sumner. 1942. Fish management problems of high western lakes with returns from marked trout planted in Upper Angora Lake California. Transactions of the American Fisheries Society **71**: 249–269.

Nelson, L. S. 1989. Evaluating overlapping confidence intervals. Journal of Quality Technology **21**: 140–141.

Nguyen, R. T., and H. R. Harvey. 2001. Preservation of protein in marine systems: Hydrophobic and other noncovalent associations as major stabilizing forces. Geochimica et Cosmochimica Acta **65**: 1467–1480.

Nicholson, J. K., J. Connelly, J. C. Lindon, and E. Holmes. 2002. Metabonomics: a platform for studying drug toxicity and gene function. Nature Reviews Drug Discovery **1**: 153–161.

Novak, J. 1988. Quantitative analysis by gas chromatography, 2nd ed. Marcel Dekker, New York, NY.

Opsahl, S., and R. Benner. 1998. Photochemical reactivity of dissolved lignin in river and ocean waters. Limnology and Oceanography **43**: 1297–1304.

Oris, J. T., S. I. Guttman, J. A. Bailer, J. E. Reuter, and G. C. Miller. 1999. Multi-level indicators of ecosystem integrity in alpine lakes of the Sierra Nevada, EPA.

Paerl, H., and C. R. Goldman. 1972. Heterotrophic assays in the detection of water masses at Lake Tahoe, California. Limnology and Oceanography **17**: 145–148.

Paerl, H. W., R. C. Richards, R. L. Leonard, and C. R. Goldman. 1975. Seasonal nitrate cycling as evidence for complete vertical mixing in Lake Tahoe, California-Nevada. Limnology and Oceanography **20**: 145–148.

Raamsdonk, L. M., B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology **19**: 45–50.

Reichenbach, S. E., M. Ni, V. Kottapalli, and A. Visvanathan. 2004. Information technologies for comprehensive two-dimensional gas chromatography. Chemometrics and Intelligent Laboratory Systems **71**: 107–120.

Rowe, T. G., D. K. Saleh, S. A. Watkins, and C. R. Kratzer. 2002. Streamflow and water-quality data for selected watersheds in the Lake Tahoe Basin, California and Nevada, Through September 1998 (USGS).

Sarkanen, K. V., and C. H. Ludwig. 1971. Lignins: occurrence, formation, structure, and reactions, 1st ed. K.V. Sarkanen and C.H. Ludwig. Wiley-Interscience.

Schlesinger, M. D., and E. M. Holst. 2000. Lake Tahoe watershed assessment: Appendix C (USFS). **II**: 386.

Schuster, S., and M. E. Grismer. 2004. Evaluation of water quality projects in the Lake Tahoe Basin. Environmental Monitoring and Assessment **90**: 225–42.

Sibson, R. 1973. SLINK: an optimally efficient algorithm for the single-link cluster method. The Computer Journal **16**: 30–34.

Simpson, M. J., A. J. Simpson, and P. G. Hatcher. 2004. Noncovalent interactions between aromatic compounds and dissolved humic acid examined by nuclear magnetic resonance spectroscopy. Environmental Toxicology and Chemistry / SETAC **23**: 355–62.

Sleighter, R. L., and P. G. Hatcher. 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Marine Chemistry **110**: 140–152.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. The University of Kansas Science Bulletin **38-2**: 1409–1438.

Sokal, R. R., and F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. Taxon **11**: 33–40.

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Kongelige Danske videnskabernes selskab: Biologiske Skrifter **5**: 1–34.

Stadler, S., P.-H. Stefanuto, M. Brokl, S. L. Forbes, and J.-F. Focant. 2012. Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. Analytical Chemistry **85**: 998–1005.

Stenson, A. C., W. M. Landing, A. G. Marshall, and W. T. Cooper. 2002. Ionization and fragmentation of humic substances in electrospray ionization Fourier transform-ion cyclotron resonance mass spectrometry. Analytical chemistry **74**: 4397–409.

Sun, M.-Y., S. G. Wakeham, and C. Lee. 1997. Rates and mechanisms of fatty acid degradation in oxic and anoxic coastal marine sediments of Long Island Sound, New York, USA. Geochimica et Cosmochimica Acta **61**: 341–355.

TERC. 2011. Tahoe Environmental Research Center. Tahoe: State of the lake report 2011.

TERC. 2012. Tahoe Environmental Research Center. Tahoe: State of the lake report 2012.

USGS. 2000. WRIR 00-4001: Surface- and ground water characteristics in the Upper Truckee River and Trout Creek watersheds.

USGS. 2012. Facts about Lake Tahoe. http://tahoe.usgs.gov/facts.html. Accessed April 10, 2013. Facts About Lake Tahoe

USGS. 2013. USGS Current Conditions for California. National Water Information System: Web Interface. http://nwis.waterdata.usgs.gov/ca/nwis/uv/. Accessed April 10, 2013

Zahn, C. T. 1971. Graph-theoretical methods for detecting and describing Gestalt Clusters. IEEE Transactions on Computers **C-20**: 68–86.

Chapter IV

Organic geochemistry of late Quaternary sediments from Fallen Leaf Lake, California Sierra Nevada, USA

4.1 Abstract

Three piston cores from Fallen Leaf Lake, California Sierra Nevada, USA were analyzed for $\delta^{13}C$, $\delta^{15}N$, TOC, C:N, and TN to examine the possible preservation of Holocene climate signals preserved within these lacustrine sediments. Calibrated, macrofossil-derived radiocarbon (^{14}C) ages revealed the record to span from present day to ca. 13.5 kya, capturing all of the Holocene as well as the end-Pleistocene. Lignin analyses were performed on early-Holocene/late-Pleistocene sediments spanning ca. 9.0-13.5 kya and provided insights into the evolution of terrestrial flora surrounding Fallen Leaf Lake following the end-Tioga glaciation. The most striking features of the FLL geochemistry record are the presence of turbidite intervals interspersed within a generally smooth and gradually increasing trend in TOC, TN, and C:N throughout the Holocene, reflecting increased organic matter inputs that became increasingly terrestrial in character. $\delta^{15}N$ was the only measured geochemical parameter that was unaffected by turbidites and it showed a trend through the Holocene towards more depleted values. The $\delta^{13}C$ record can be divided into four major sections and the end of Section III at ca. 3.8 kya coincides with the cessation of drought-like conditions in the Great Basin and a return to more enriched $\delta^{13}C$ values. $\delta^{13}C$, TOC, TN, and C:N were also examined across graded turbidites and these parameters appeared driven by changes in particle sizes within the turbidite that effected organic matter sorting by size fraction.

4.2 Introduction

4.2.1 Holocene climatic variability

Evidence for long and recurring droughts in the Sierra Nevada and Great Basin of North America exist for the Holocene (Stine 1994). Evidence for these droughts comes from multiple records, including submerged, ≥ 200 year-old trees (Kleppe et al. 2011) and desiccated terminal lake basins, including those of Owens Lake and Pyramid Lake (Benson et al. 2002; Mensing et al. 2004). Sub-alpine lakes within the endorheic Tahoe basin are potentially well-suited for studies of Holocene climate because they represent systems that are constrained by known water budgets and are sensitive to variability in precipitation because of the influence snowpack exerts in the supply of water to these systems. Additionally, these systems are surrounded by vegetation and, as such, abundant macrofossils can be found in their sediments that, in turn, enable the construction of robust age models that allow sedimentologically observed events to be translated into the time domain, allowing for their interpretation within the context of other paleoclimate records available for the Sierra Nevada and Great Basin. To date, most studies in the Tahoe Basin have focused on its terminus, Pyramid Lake, Nevada. Lake Tahoe is itself poorly suited to climate studies because of its low basin-to-lake surface ratio of 1.67:1, its low sedimentation rates (~20cm/kyr), and because of its sediment record, approximately 50 % of which is comprised of debris flows (Smith et al. 2013). Adjacent Fallen Leaf Lake (FLL) (Fig. 4.1) is much smaller, has higher sedimentation rates, less debris flows, and has a greater drainage area to lake area ratio (6.9) and, as such, its sedimentary record should be more sensitive to climatological influences.

4.2.3 Fallen Leaf Lake (FLL)

Fallen Leaf Lake (FLL) is a pristine, sub-alpine lake bounded by recessional glacial moraines, which are remnants of the Tioga glaciation (Brothers et al. 2009). The lake is located in El Dorado County, California at a surface elevation of 1,942 m and it is immediately adjacent and hydraulically connected to Lake Tahoe (Fig. 4.1), which makes its lake level significantly more sensitive than Lake Tahoe's to changes in precipitation, or snowpack (Kleppe et al. 2011). The lake has a maximum length of 5.0 km (N-S), a maximum width of ~1.6 km, a surface area ~ 5.67 km^2 (Lico 2004), an average depth of 53 m, and a maximum depth of 120 m (Al-Chokhachy et al. 2009), and a resultant volume of about ~ 0.30 km^3.The ratio of its drainage basin to its surface area is about 6.9 and the lake is considered meso-oligotrophic (Al-Chokhachy et al. 2009). The lake is sub-monomictic, mixing fully less than once a year. Its hydraulic residence has been reported to be 8 years (Hanes 1981).

4.2.2 Organic geochemical proxies in lake sediments

Analyses of organic geochemical parameters in lake sediments can provide information pertaining to the composition and magnitude of past organic matter (OM) sources to lakes. Carbon to nitrogen (C:N) ratios of bulk sedimentary organic matter generally reflect the original proportions of algal and land-derived material while carbon isotopes can be a useful tool that can gauge past levels of lake productivity and the extent of carbon recycling (Meyers and Ishiwatari 1993). Compound-specific approaches can provide additional detail into the composition of organic matter accumulating in lake sediments. The lignin phenols (Fig. 4.2), which are uniquely derived from vascular plant

material, can be particularly useful for ascertaining terrestrial vegetation changes through time (Huang et al. 1999; Tareq et al. 2004; Ishiwatari et al. 2006). Gymnosperm plants synthesize lignin that is almost entirely devoid of syringyl phenols whereas angiosperms synthesize both vanillyl and syringyl phenols. Meanwhile, nonwoody plant tissues, such as pollen, leaves, and pine needles, are preferentially enriched in cinnamyl-type phenols. Collectively, the relative abundances of these individual lignin phenol types be used quantify changes in terrestrial vegetation through time.

4.3 Methods

4.3.1 Coring

A coring campaign by the National Lacustrine Core Facility (LacCore) was completed in November 2010 and a suite of piston and gravity cores were collected. Three piston cores (**1**A, **2**D, and **3**A) were analyzed for their geochemistry. Core 3A (120°3’35.028”, 38°54’19.8354”) was 4.6 m in length and was obtained from a water depth of 94.9 m. Core 1A (120°3’47.268”, 38°53’36.3834”) was 9.84 m in length and was obtained from a water depth of 115 m. Core 2D (120°3’56.808”, 38°53’30.9834”) was 11.54 m in length and was obtained from a water depth of 111 m. The cores were originally processed and stored at LacCore and subsamples of these cores were returned to University of Nevada, Reno (UNR) for subsequent elemental and isotopic analyses and to Scripps Institution of Oceanography (SIO) for lignin analyses.

4.3.2 Age Control

Age control for cores 1A, 2D and 3A were provided by ^{14}C analyses of 28 macrofossils (pine needles) picked from these three cores as well as by the Tsoyowata ash marker bed present in cores 1A and 2D. This ash bed has been dated to 7470-7620 cal years BP (Hallett et al. 1997) and was produced from the eruption of Mt. Mazama, present day Crater Lake, Oregon. Macrofossils were analyzed for their natural abundance radiocarbon (^{14}C) content by accelerator mass spectrometry (AMS) at Lawrence Livermore National Laboratory and were converted to radiocarbon ages. Radiocarbon ages were then converted to calibrated calendar (cal.) ages using OxCal v4.1.7 (Bronk Ramsey 2009) and the IntCal04 terrestrial radiocarbon age calibration curve for the last 0–26,000 cal. years before present (Reimer et al. 2004), where present is defined as AD 1950. Age models were constructed for each core. This enabled the construction of a composite core.

4.3.3 Elemental abundance (C,N) and stable isotope ($\delta^{13}C$, $\delta^{15}N$) analysis.

Cores 1A, 2D, and 3A were subsampled at LacCore and subsamples were returned to UNR where they were analyzed at the Nevada Stable Isotope Lab at UNR for TOC (%), TN (%), $\delta^{13}C$, and $\delta^{15}N$ following established procedures at this facility. $\delta^{15}N$ and $\delta^{13}C$ values are reported relative to either air or PeeDee Belemnite (PDB), respectively. Briefly, subsamples (~1 g wet sediment) were soaked in HCl fumes overnight to ensure the absence of carbonates, were dried at 60 °C for ~48 hours (or until weights stabilized, whichever was longer) were ground to a fine powder, were rolled into tins, and were then analyzed by elemental analysis isotope ratio mass spectrometry (EA-

IRMS). Standards (acetanilide) of known elemental and isotopic composition, and of measured weights, were run every 10-15 samples. Each tin prepared for analysis contained ~ 25–35mg of sediment (dry weight). The sediment was diatom rich and was extremely hygroscopic and, as such, samples were always weighed immediately after drying and their subsequent exposure to moisture was limited. Reproducibility, as determined from replicate sample analyses, for TOC (%) was ± 0.03 (1σ), for TON (%) it was < ± 0.01 (1σ), for $\delta^{13}C$ it was ± 0.04 ‰ (1σ), and for $\delta^{15}N$ it was significantly higher, ± 0.15 ‰ (1σ). 15 samples, plus five replicates, were analyzed from sections 7 and 8 of core 1A, representing a 20 cm sampling resolution. 130 samples, plus 22 replicates, were analyzed from all sections (1–8) of core 2D, representing an 8–12 cm sampling resolution throughout the whole core. 36 samples, plus 6 replicates, were analyzed from section 1 of core 3A, representing a 4 cm sampling resolution. In total, 214 measurements were made which corresponded to 181 depth measurements for $\delta^{13}C$, $\delta^{15}N$, TOC, and TN across the three cores. Core 3A, section 1, was 154 cm in length and spanned ca. 3.8–0.9 kya as it over-penetrated the lake floor. Core 1A sections 7 and 8 were together 288 cm in length and spanned ca. 13.5–9.0 kya. Core 2D was 11.54 m in length and spanned ca. 10.5 kya to present day (i.e. 2010).

4.3.4 Lignin phenol analysis

4.3.4.1 Plant tissue and sediment samples selected for analysis

Dried, homogenized sediment aliquots of the 15 samples originally prepared for elemental and isotopic analysis for sections 7 and 8 of core 1A were analyzed for their

lignin phenol contents. Sections 7 and 8 of core 1A were analyzed in order to provide a basis for explaining the large elemental and isotopic excursions observed in these sections. A total of 13 plant tissue and forest litter samples were also analyzed to determine their lignin contents such that the lignin compositions of potential organic matter (OM) sources to present day Fallen Leaf Lake could be surveyed. The samples were collected June 9, 2012 to be reflective of the vegetation surrounding Fallen Leaf Lake.

The plant tissues analyzed included needles, bark, and pinecone from *Pinus jeffreyi* (Jeffrey pine), needles and pinecone from *Abies concolor* (white fir), leaves and twigs from *Quercus vacciniifolia* (huckleberry oak), and leaves from *Arctostaphylos patula* (greenleaf manzanita). Forest litter samples were collected from adjacent Glen Alpine Creek (GAC) and from adjacent to Lily Lake (LL). The sample from LL was divided into a coarse and a fine fraction. The fine fraction contained particles less than approximately 0.5 mm in diameter while the coarse fraction contained larger fragments, including small, visible pieces of pine needles. All plant and forest litter samples were dried at 60 °C for 72 hours and then pulverized into a fine powder using a mortar and pestle.

4.3.4.2 Cupric oxide (CuO) oxidation

Organic matter samples were oxidized by cupric oxide (CuO) oxidation following the method of (Hedges and Ertel 1982) as adapted by (Goñi and Montgomery 2000) for microwave-assisted alkaline hydrolysis. Briefly, material equivalent to between 3.0–5.0 mg OC was mixed with 500 mg CuO, 50 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, and 20 mL 2 *N*

NaOH into perfluoroalkoxy (PFA) hydrolysis vessels. The vessels were placed under vacuum and the headspace was replaced with N_2 at ~ 30 psi. This process was repeated 5× in total to ensure the removal of molecular oxygen from the vessels whose presence in alkaline media at elevated temperature can otherwise lead to super-oxidation effects. After the fifth evacuation, the headspace was replaced with ~7 psi N_2. The mixture was heated to 150 °C during a 5.0 min temperature ramp and was then held at this temperature for 90 min. Then, the oxidation vessels were immersed in water and cooled to room temperature, depressurized, opened, and spiked with dual internal recovery standards, *trans*-cinnamic acid and ethyl vanillin. The oxidation mixture was acidified to pH ≈ 2 using 12.1 *N* HCl. The aqueous acidic solution was extracted 3× with ethyl acetate, the ethyl acetate was first dried over Na_2SO4 and then blown to complete dryness under N_2 after which the hydrolysate residue was immediately redissolved into 400 μL of pyridine. The dissolved hydrolysate residue, pyridine, and N,O-BSTFA + 1% TMCS were then combined 1:1:2 inside 400 μL GC glass insert vials. The mixture was heated for 15 min at 60 ° C to derivatize functional groups containing exchangeable protons (principally acids and alcohols) with nonpolar trimethylsilyl groups (TMS). This derivatization procedure increases the volatility while reducing the polarity of analytes. This makes the analytes more amenable to gas chromatographic separation. Samples were analyzed using a 7890A Agilent gas chromatograph (GC) coupled to an Agilent 5975C triple-axis quadrupole-MS and flame ionization detector (FID) with a 1:1 split between detectors. A 60 m, 0.25 mm i.d., 0.5μm df DB-5 capillary column was used. The oven program was 100–320 °C at 4.0 °C min^{-1} with initial and final holds of 1 and 10 minutes, respectively, and the total runtime was 66 min. The system was operated in

constant flow mode at 1.5 mL min^{-1} with He as the carrier gas. Inlet and transfer line temperatures were both set to 280 °C. Detector temperature was 320 °C. Injections were always 1 μL and were made using an Agilent 7683 automatic liquid sampler (ALS). Quantification of lignin phenols was performed against 5-point calibration curves obtained from simultaneous GC analysis of derivatized external standards for each compound using the FID signal. R^2 values were always > 0.999 for quantified compounds. Ethyl vanillin was used as the recovery standard for the aldehydic compounds while *trans*-cinnamic acid was used as the recovery standard for all other compounds. Lignin concentrations are reported normalized to total carbon, as mg (lignin phenol) per 100 mg organic carbon (OC). Higher lignin concentrations therefore mean more of the carbon in a particular sample can be accounted for as lignin. All plant and forest litter samples were run in duplicate and all sediment samples were in *at least* duplicate, as indicated. All errors are expressed as ±1σ. The lignin phenol vegetation index (LPVI) was calculated according to (Tareq et al. 2004). Lake Washington Mud (LWM) was CuO-oxidized and analyzed in concert with the other samples as an external environmental reference standard. Lignin parameters for LWM have been compared between labs previously (Goñi and Montgomery 2000). The lignin parameters for LWM obtained in the course of these analyses are shown in the context of previously reported lignin parameters for this sample (LWM) (Fig. 4.S1) and no statistically significant differences are observed.

4.4 Results

4.4.1 Depth-age relationships, sedimentation rates, and stratigraphic event context

Depth-age relationships are shown in Fig. 4.3 for those core sections of 1A, 2D, and 3A from which geochemical data were obtained. The sediment accumulation rates in cores 2D and 3A were similar (~0.51 mm $year^{-1}$) up until approximately ca. 3.25 ka. at which time the sediment accumulation rate in 2D increased approximately two-fold, to ~0.95 mm $year^{-1}$, over that of 3A. Large spikes in sediment accumulation for core 2D exist and are associated with gravity driven events. One spike, at ~ ca. 5.0 ka is associated with the most recent (seismic) event (MRE) as reported by (Brothers et al. 2009). This event resulted in a turbidite deposit that was preserved in 2D. The thickness and areal extent of this turbidite was subsequently mapped using compressed high intensity radar pulse (CHIRP) imaging, was termed Fallen Leaf Lake Slide 1 (FLLS1), and was assigned a preferred age of 4.57–4.85 ka by (Maloney et al. 2013). The sediment accumulation rate in core 2D spiked again at ca. 7.8 ka. This event has been termed FLLS2 and was assigned a preferred age of 7.62–7.90 ka (Maloney et al. 2013). FLLS2 directly overlies the Tsoyowata ash marker bed. Following FLLS2, sediment accumulation rate in core 2D remained at 1.65 mm $year^{-1}$. A distal portion of a turbidite termed FLLS3 (ca. 11.24–11.64 ka) (Maloney et al. 2013) was observed in CHIRP data to intersect with core 1A. This turbidite, however, appeared too thin to elicit an observable spike in sediment accumulation rate in core 1A. A fourth slide (FLLS4) was also observed by (Maloney et al. 2013) in CHIRP data, however it was not penetrated by any of the these three cores. The sediment accumulation rate in core 1A is consistently

lower than in core 2D. Core 1A extends the furthest back in time. Determination of these core-specific age-depth relationships enables the partial geochemistry records from cores 1A, 2D, and 3A to be merged into a composite core and creates a continuous geochemical record for Fallen Leaf Lake from Late Pleistocene through Recent.

4.4.2 C and N elemental and isotopic stratigraphy

The geochemistry data analyzed from core 1A, 2D, and 3A were merged into a composite core (Fig. 4.4) using the age-depth relationships for each core (Fig. 4.3). This composite core spanned the length of the Holocene and the last ~1800 years of the Pleistocene and enabled a direct inter-comparison of the isotope records observed in each core. The composite core shows clear excursions in TOC (%), TN (%), $\delta^{13}C$, and C:N that are associated with FLLS1 (MRE) and FLLS2 (above the Tsoyowata ash). These turbidite events are the most apparent features in the isotope stratigraphy of the core. Unlike the other measured C and N parameters, $\delta^{15}N$ shows no correlation to the turbidite intervals. Aside from these event-driven disturbances, TOC (%) and C:N show gradual increases through the Holocene, starting at values of 1.5 and 8.3 %, respectively, in the Pleistocene and increasing to values as high as 6.4 and 11.8 %, respectively, at the core-top of 2D. TN (%) also exhibits a consistent increase through the Holocene, ranging from approximately 0.2 to 0.5 %. Meanwhile, $\delta^{15}N$ shows a trend towards less enriched values through the Holocene. Other smaller trends were also apparent. Following the MRE, from ca. ~5.0 kya to ca. ~3.8 kya, TOC and TN both decrease and $\delta^{13}C$ becomes more depleted; however, no corresponding shift was observed for either $\delta^{15}N$ or for C:N in this interval. A second interval that shows significant shifts in geochemical parameters is at

the base of core 1A, which crosses the Holocene-Pleistocene boundary. TOC, TN, and C:N increase rapidly moving into the Holocene and $\delta^{15}N$ also shows an increase, although it is less pronounced. During this same interval $\delta^{13}C$ undergoes a sinuous transition from depleted values (-25.9 ‰, ca. 13.5 kya), to more enriched ones (-24.4 ‰, ca. 13.1 kya), back to depleted values (-26.4 ‰, ca. 12.0 kya), and back again to slightly more enriched values (-25.8 ‰, 11.7 kya) at the very end of the Pleistocene before stabilizing at more depleted values (< -27.0 ‰) in the early Holocene.

Cross-plots of elemental and isotope data from each of the three cores is shown in Fig. 4.5. In Fig. 4.5A, the TN (%) is plotted vs. TOC (%). Here, the *y*-intercept is positive for each core and suggests the presence of nitrogen in the absence of carbon, or inorganic nitrogen (TIN) in the analyzed sediments. The magnitude of the TIN contribution to each core appears to vary based on their unique *y*-intercepts. If the differences in the *y*-intercepts between cores are deemed significant, the core with the lowest sedimentation rate, 1A, had the lowest levels of TIN (0.03 %) while the core from the shallowest water depth (3A) contained levels which were roughly 3× higher (0.10 %). Across all samples, higher $\delta^{15}N$ appears to correlate with lower $\delta^{13}C$ (Fig. 4.5B). C:N exhibits strong and positive relationships with TOC for all cores (Fig. 4.5C) and C:N ratios observed for core 3A were slightly elevated relative to those observed for core 2D. Sediment samples enriched in nitrogen (lower C:N) tended to have higher $\delta^{15}N$ (Fig. 4.5D). TN showed high correlation to TOC ($R^2 = 0.87$)

4.4.3 TOC, TN, $\delta^{13}C$, and C:N across the MRE turbidite

The turbidite (FLLS1) caused by the MRE is present in core 2D from 3.75 to 3.86 m and leaves a visible imprint on the elemental and isotope stratigraphy through this section (Fig. 4.6) that is complemented by a spike in magnetic susceptibility (MS) at the base of the turbidite and a smaller, less-defined peak in MS at its top. The turbidite has a dark, dense, silty base and an organic rich top capped by a light gray clay. Its top is irregular and mottled in with overlying olive gray clay. TOC and TN were lowest in the clay interval at the top of the turbidite (0.73 and 0.12 %, respectively), highest in the middle of the turbidite (3.25 and 0.33 %, respectively), and are of intermediate values at the base of the turbidite (2.21 and 0.19 %, respectively). $\delta^{13}C$ is highest at the base of the turbidite (-24.9 ‰) and lowest at its top (-26.5 ‰). C:N is highest (11.9) at the base of the turbidite and lowest at its top (6.1).

4.4.4 Lignin phenols analyses

4.4.4.1 Lignin phenols in plant tissues and forest litter

Lignin yields for the survey of 13 analyzed plant tissue and soil litter samples encompassed a wide range of values. They are listed in Table 1 and are shown in Fig. 4.7, listed by species name. The three angiosperm tissue samples yielded S-type phenols in the highest yield while the forest litter samples yielded subordinate quantities of this phenol-type, indicating a small angiosperm contribution to the forest litter samples. The LPVI values (1-15) of the forest litter samples also suggests a predominant woody to nonwoody gymnosperm source. S-type phenols are largely absent from gymnosperm

tissue samples except for needles from *Abies concolor* (white fir) where S-type phenols are present, but at very low abundances ($\Sigma S = \sim 50$ μg 100 mg OC^{-1}). C-type phenols are highest for nonwoody plant tissues, such as for leaves analyzed from *Arctostaphylos patula* (green leaf manzanita) and for pollen skimmed from FLL. As such, these two samples have the highest C/V ratios (1.742 and 0.805, respectively). The nonwoody angiosperm samples have the highest LPVI values while the woody samples have the lowest LPVI values, both of which were consistent with this index. The bark and cone from *Pinus jeffreyi* (Jeffrey pine) has the highest yields of vanillin and the lowest C/V ratios, making these two tissue types the most woody of those analyzed. The presence of S-type phenols in the pollen from FLL indicates that the pollen sample contains pollen from both gymnosperm and angiosperm plant species. The Vd/Vl ratios of all plant samples are below 0.22, consistent with non-degraded lignin, and are highest for the forest litter samples (0.222 – 0.396). The small grain size forest litter sample contains significantly ($p<0.01$) higher Vd/Vl (0.396) and Sd/Sl (0.358) ratios compared to the larger grain-size fraction, which has values of 0.260 and 0.252, respectively, possibly suggesting a higher state of degradation for fine-fraction soil lignin, or partitioning of degraded, acid-enriched lignin to smaller grain sizes (Bergamaschi et al. 1997).

4.4.4.2 Lignin phenols across the Pleistocene-Holocene transition

Results of lignin analyses for sediments from sections 7 and 8 of core 1A (688–976 mblf, ca. 9.0-13.5 kya) are shown in Table 2 and are plotted with plant and forest litter-derived S/V and C/V values in Fig. 4.8 for the purposes of comparison. Lignin concentrations derived from repeated (n=4) analyses of reference sediment

(LWM) analyzed alongside these samples were statistically indistinguishable from concentrations reported elsewhere (Goñi and Montgomery 2000) for this same material (Fig. 4.S1).

The lignin phenol composition of all sediment samples across the Pleistocene-Holocene transition are broadly reflective of lignin predominantly derived from gymnosperm sources based on their low S/V values (~ 0.28) with more limited angiosperm contributions. Rescaling the S/V axis (Fig. 4.8B) reveals more clearly that the C/V ratios of lignin phenols in core 1A show significantly greater variability than do the S/V ratios. In Fig. 4.9, lignin parameters from core 1A sediments are plotted vs. time (ca. kya) alongside TOC, $\delta^{13}C$, and magnetic susceptibility. The Vd/Vl and Sd/Sl ratios both show substantial variability between replicate oxidations (as evidenced by large errors bars) and were therefore considered unsuitable for interpretation. C/V and P/V ratios show almost identical profiles which, in turn, closely resemble the LPVI profile. As such, the LPVI index in this system appears to predominantly discriminate woody vs. nonwoody plant tissue types. Additionally, while the P-type phenols have sources other than lignin, such as protein (Goñi and Hedges 1995), their close correspondence with C/V ratios suggests that these phenols are, in fact, predominantly lignin-derived from nonwoody tissues, such as pollen, leaves, and pine needles. In the analyzed section of core, total lignin concentrations (Λ_8) vary between 1.65 – 2.19 mg (100 mg OC)$^{-1}$, except for an abrupt increase in Λ_8 at ca. 12.6–13.0 kya, where Λ_8 rises to between 2.51–2.69 mg (100 mg OC)$^{-1}$. This increase in Λ_8 lignin is also associated with an abrupt increase in $\delta^{13}C$; meanwhile, during this increase TOC remains low (< 1.5 %). In this interval, C/V, P/V, S/V, and LPVI also all drop markedly and collectively suggest a significant, and

predominantly woody, gymnosperm source to this interval of the core. This interval also corresponds with high magnetic susceptibility values (Fig. 4.9) indicative of elevated clastics.

After this apparent gymnosperm pulse, however, Λ_8 lignin concentrations quickly drop to between 1.65–2.19 mg (100 mg OC)$^{-1}$ beginning at ca. 12.3 kya. Concurrent with this drop in Λ_8 were sharp increases in C/V, P/V, S/V and LPVI, which collectively suggests an abrupt transition to a nonwoody lignin source consistent with a mixture of gymnosperm and angiosperm pollen. Gymnosperm pollen from *Pinus* species has been reported to have C/V ratios as high as 13.6 and S/V values of zero (Ishiwatari et al. 2006). The pollen sample measured here integrated both angiosperm and gymnosperm pollen as evidenced by its non-zero S/V value. Then, after ca. 12.3 kya, C/V, P/V, and LPVI begin a gradual decrease until about ca. 10.9 kya at which point these values stabilized. During this period, S/V ratios increased only slightly, from ~0.23 to ~0.29.

4.5. Discussion

4.5.1 Carbon and nitrogen elemental and isotope record

Fallen Leaf Lake (FFL) sediments reveal substantial changes in the composition and amount of accumulating carbon and nitrogen over the past ca. 13.5 kya. Present day TOC (6.36 %) and TON (0.54 %) are about 4× and 2× higher than they were at the base of the composite core. Over this period, excluding turbidite intervals, $\delta^{13}C$ has varied (-28.5 to -24.9 ‰) in magnitude by 3.6 ‰ and $\delta^{15}N$ has varied (0.3 to 2.1 ‰) by 1.8 ‰. The long term elemental and isotopic trends preserved in the FLL record show TOC, TN,

and C:N to be steadily increasing throughout the Holocene with minimal interruption. This same trend was also recently observed in Lake Tahoe (Smith et al. 2013). While post-burial organic matter degradation may well occur, this process alone seems unlikely to account for this trend. TOC:TN are highly correlated to one another through the sediment record in FLL (Fig. 4.5E), indicating that no significant preferential degradation of carbon or nitrogen is occurring through time. Additionally, if organic matter degradation alone accounted for the Holocene-long trends in TOC, TN, and C:N, a more consistent and systematic attenuation of these parameters with depth might well be expected. As such, these trends most likely reflect variations in the composition and amount of organic source material being delivered to the sediments. The amount of OM being delivered to the sediments appears to be increasing as does its C:N ratio, suggesting that the sediments have received increasing quantities of terrestrial OM through the Holocene.

Turbidites, which punctuate the FLL geochemistry record, leave distinct elemental and isotopic imprints of their own (Fig. 4.6). The turbidite associated with the MRE was examined in detail. A large increase in magnetic susceptibility is present at its base, presumably due to ferromagnetic material associated with coarser grained sands that settled quickly following their original suspension. The coarser grained material at the base of the turbidite is ^{13}C-enriched compared to the finer grained material overlying it, which is more ^{13}C-depleted. The white/gray clay capping the turbidite is the most ^{13}C-depleted portion of the turbidite. In Lake Tahoe, (Smith et al. 2013) observed that the basal turbidite sands contained < 10 % of biogenic opal while the clay at the top of the turbidite frequently contained opal in excess of 35 %. If sorting of algal and terrestrial

OM is achieved by turbidite grading, this would suggest that algal OM is more ^{13}C-depleted than terrestrial OM by *at least* 1.6 ‰, and by greater amounts if sorting of algal from terrestrial OM by grain size is considered here to be incomplete. At the FLLS2 turbidite overlying the Tsoyowata ash, the difference between the ^{13}C-enriched basal turbidite and the ^{13}C-depleted clay cap was similar (2.0 ‰) to what was observed for the MRE/FLLS1 turbidite.

The $\delta^{15}N$ record in FLL trends towards lighter, less depleted values throughout the Holocene and this record is unique compared to the other measured geochemical parameters as it was the only parameter that appeared unaffected by the turbidites. The same trend towards lower $\delta^{15}N$ values throughout the Holocene, as well as the insensitivity of $\delta^{15}N$ to turbidite intervals, was also observed in Lake Tahoe (Smith et al. 2013).

The interpretation of the $\delta^{13}C$ of sedimentary organic matter in freshwater lake sediments can be complex. Freshwater aquatic algae can either fix dissolved CO_2 that is in equilibrium with the atmosphere ($\delta^{13}C \approx -8$) or from CO_2 that is derived from the aqueous respiration of isotopically light terrestrial carbon. Meanwhile, if the utilized carbon is first equilibrated to bicarbonate, the resulting algae will be ~ 7 ‰ higher than if it had used CO_2 directly (Meyers 1994). This latter scenario is the case for most marine algae, which use bicarbonate as their carbon source. Lastly, the expression of kinetic isotopic effects can be suppressed as the supply of dissolved CO_2 or bicarbonate is exhausted, such as during blooms when rapid update of CO_2 or HCO_3^- occurs.

The $\delta^{13}C$ record in FLL can be divided into four sections. These are labeled I–IV in Fig. 4.4. In Section I, from ca 10.9–12.6 kya, $\delta^{13}C$ values decrease. In Section II, from

ca. 5.0-10.9 kya, $\delta^{13}C$ remains stable, but undergoes a very gradual ~1 ‰ increase. In Section III, from ca 3.8-5.0 kya, a sharp decrease in $\delta^{13}C$ occurs following the MRE. Finally, in Section IV, from ca. 3.8 kya to present, a sharp increase in $\delta^{13}C$ occurs followed by a slow and tapered progression to lower values approaching present day.

There does not appear to be close correlation between the timing of $\delta^{13}C$ shifts in FLL and the climate shifts in the Great Basin and Pyramid Lake system proposed by (Benson et al. 2002). For example, the boundary between Sections III and IV at ca. 3.8 kya coincides with the end of Holocene drought conditions as identified by (Benson et al. 2002). During these drought conditions up until ca. 3.8 kya, Owens Lake became desiccated, Lake Tahoe stopped spilling into the Truckee River, and Pyramid Lake began to desiccate. If indeed this isotope excursion is due to climate-induced drought conditions, it is not clear why the negative isotope excursion only appears after the MRE turbidite, or why the earlier part of the mid-Holocene is characterized by $\delta^{13}C$ values that are generally 2 ‰ higher. More work is clearly needed if the drivers of the observed shifts in $\delta^{13}C$ are to be accurately deciphered. Performing additional compound-specific geochemical analyses can complement the interpretation of bulk-level OC isotopic shifts by providing an additional explanatory dimension.

4.5.2 Lignin across the Pleistocene-Holocene boundary

The analysis of lignin phenols through section 1A provides an independent means of assessing the drivers of the $\delta^{13}C$ variations observed in this core. The positive $\delta^{13}C$ excursion at the base of the core is correlated with increased lignin (Λ_8) concentrations and lignin parameters including low C/V, S/V, and low LPVI values (Figs. 4.8, 4.9),

suggesting a dominant woody gymnosperm source. The increase in Λ_8 lignin, enrichment in ^{13}C, decreases in both C/V and S/V, and the high magnetic susceptibility of this interval, could be consistent with a glacial outburst event (annotated on Fig. 4.9) that mobilized woody gymnosperm lignin preserved beneath the ice during its original advance. Selective degradation of C-type phenols may result in anomalously low C/V values, especially if the lignin is associated with fine soil or sediment particles (Goñi et al. 1997), presumably including glacial rock flour. Artificially low C/V values have also been reported for lignin analyzed from lacustrine sediments from the Last Glacial Maximum (Huang et al. 1999). FLL CHIRP data contain chaotic and strongly reflective deposits that have been interpreted to represent glacial outburst events (Maloney et al. 2013).

At ca. 12.3 kya, following the deposition of this high MS sediment package near the base of core 1A, the organic matter accumulating in FLL presumably began to integrate OM sources from its terrestrial surroundings. These surroundings conceivably changed with time following glacial retreat, Holocene warming, habitat colonization and species succession. After ca. 12.3 kya, LPVI, C/V, and P/V decreased steadily before stabilizing at ca. 10.9 kya. Meanwhile, S/V showed little variation. Interpretation of these parameters suggests the expansion of gymnosperms during this time and the stabilization of these parameters at ca. 10.9 kya may represent the time at which the FLL watershed first reached a steady-state population of gymnosperms, approximately ~1,400 years after the apparent glacial outburst event. The positive $\delta^{13}C$ shift, elevated Λ_8 lignin and magnetic susceptibility, and low C/V at the base of 1A, however, present a small puzzle as lignin derived from C_3 plants is generally very ^{13}C-depleted (e.g. -34.5 to -24.9 ‰ for

pine needles) (Goñi and Eglinton 1996). Lignin derived from C_4 plants, meanwhile, is more ^{13}C-enriched enriched, however it is difficult to reconcile a possible C_4 plant source with lignin phenol data that strongly evidence a gymnosperm source. The C_4 metabolic pathway arose in the angiosperm plant lineage and, as such, no C_4 gymnosperms exist. It is also not clear that extensive C_4 plants existed in the vicinity of Lake Tahoe or Fallen Leaf Lake prior to the last glaciation. A final explanation may be that preferential degradation of isotopically light lignin occurred prior to its entrainment in FLL sediments, but the effects of degradation on the compound-specific $\delta^{13}C$ of lignin phenols appear to have escaped previous study and therefore the likelihood of this hypothesis cannot be evaluated.

4.6 Conclusions

The construction of a composite core enabled the creation of an organic geochemical record for Fallen Leaf Lake that extended to ca. 13.5 kya. This core spanned the whole of the Holocene and the Holocene-Pleistocene boundary. In the past ca. 13.5 ky, FLL sediments have recorded substantial changes in the amount and composition of OM input to this lake. The most striking features of these records are the elemental and isotope excursions associated with the turbidite sequences in these cores. The first turbidite, FLLS1, was associated with the MRE while the second was directly underlain by the Tsoyowata ash. These turbidite intervals, while abruptly punctuating the otherwise gradual trends in organic geochemical parameters, contained resolvable C and N patterns of their own. C and N analyses through turbidites showed evidence of hydrodynamic particle sorting that fractionated fines from coarser sediments, consistent with what has

been observed for turbidite sequences in Lake Tahoe (Smith et al. 2013). This particle sorting appeared to affect a partitioning of organic matter whereby ^{13}C-enrichment and elevated C:N occurred in the sandy bases of the turbidites while ^{13}C-depletion and lower C:N values occurred in the fine clays capping the turbidite, consistent with a predominantly aquatic (vs. terrestrial) source for this latter carbon. The magnitude of the isotope fractionation affected by this particle sorting across these intervals was consistent, approximately 1.6 ‰ for the MRE turbidite and 2.0 ‰ for the FLLS2 turbidite. Throughout the Holocene, TOC, TN, and C:N have been increasing and this trend likely reflects increasing amounts of OM being delivered to FLL sediments. The gradually increasing C:N through the core may indicate that the inputs of terrestrial OM have grown more rapidly than inputs of algally derived OM through the Holocene. The mechanistics that underlie the several interesting $\delta^{13}C$ shifts observed in the FLL record are difficult to definitively disentangle and are likely confounded by the existence of multiple sources of dissolved CO_2 to the lake, which can be both atmospheric and terrestrially derived. Without the aid of additional proxies, it remains difficult to definitively ascertain the mechanisms underlying many of these shifts. Although Section III appears to terminate with the cessation of drought-like conditions in the Great Basin, a definitive linking mechanism seems to be lacking. For instance, in Section III, a shift to more ^{13}C-depleted values may suggest increased algal inputs, however no corresponding decrease in C:N is observed, although its rate of increase does slow during this period. Analysis of lignin through section 1A revealed lignin and its related parameters to vary substantially. An interval characterized by elevated lignin, ^{13}C-enrichment, low C/V, low P/V, and low TOC may reflect a glacial outburst event that contains lignin with a

different provenance from that which accumulated in the rest of 1A (Fig. 4.8B). Following this event, at ca. 12.3 ka, lignin immediately becomes very nonwoody in character, characterized by high C/V values that decrease before stabilizing at approximately ca. 10.9 ky. This transition ostensibly records an ecosystem shift characterized by increasing amounts of woody vs. nonwoody gymnosperm lignin through time consistent with the expansion of coniferous flora following the end-Tioga glaciation.

Table 4.1: Survey of lignin concentrations in potential source materials. Plant and tissue specimens were collected from Fallen Leaf Lake and its immediate surroundings. Concentrations are expressed as mg per 100 mg OC, assuming 50 % OC in these terrestrial materials. g = gymnosperm needles, c = gymnosperm cones; a = angiosperm leaf; b = gymnosperm bark; tw = angiosperm twig; GAC = sample from adjacent Glen Alpine Creek, LL = sample from adjacent Lily Lake. The granular forest litter samples contained small fragments of conifer needles. LPVI = lignin phenol vegetation index. Σ (%V + %S + %C) = 100 %. Errors (± 1σ) derived from duplicate oxidations are shown below the values to which they correspond. [I] = pollen was skimmed from the surface of Fallen Leaf Lake. Pl = *p*-hydroxy benzaldehyde, Pn = *p*-hydroxy acetophenone, Pd = *p*-hydroxy benzoic acid, Vl = vanillin, Vn = acetovanillin, Vd = vanillic acid, Sl = syringaldehyde, Sn = acetosyringone, Sd = syringic acid, Cd = *p*-coumaric acid, Fd = ferulic acid.

Sample	Pl	Pn	Pd	Vl	Vn	Vd	Sl	Sn	Sd	Cd	Fd
1. *Pinus jeffreyi* (g)	0.272	0.074	0.318	4.405	0.667	0.679	0.016	0.017	0.000	0.553	0.331
	0.109	0.004	0.015	0.326	0.059	0.122	0.003	0.024	0.000	0.004	0.003
2. *Abies concolor* (g)	0.132	0.055	0.200	3.100	0.447	0.498	0.098	0.042	0.037	0.496	0.143
	0.039	0.004	0.009	0.159	0.023	0.012	0.015	0.002	0.004	0.016	0.006
3. *Abies concolor* (c)	0.266	0.065	0.094	3.268	0.627	0.665	0.021	0.022	0.007	0.341	0.136
	0.017	0.006	0.010	0.183	0.049	0.065	0.000	0.003	0.000	0.016	0.008
4. *Quercus vacciniifolia* (a)	0.225	0.053	0.338	2.117	0.358	0.431	1.926	0.438	0.397	0.629	0.169
	0.018	0.003	0.004	0.137	0.004	0.042	0.189	0.019	0.051	0.030	0.002
5. *Pinus jeffreyi* (b)	0.437	0.106	0.106	5.488	1.158	1.072	0.018	0.022	0.000	0.027	0.566
	0.012	0.003	0.001	0.006	0.007	0.021	0.002	0.006	0.000	0.003	0.006
6. *Pinus jeffreyi* (c)	0.359	0.078	0.133	4.865	0.877	1.070	0.020	0.020	0.000	0.030	0.125
	0.032	0.000	0.030	0.079	0.069	0.336	0.001	0.001	0.000	0.017	0.022
7. *Arctostaphylos patula* (a)	0.042	0.035	0.092	0.354	0.066	0.071	0.685	0.186	0.126	0.800	0.053
	0.006	0.001	0.005	0.007	0.004	0.004	0.005	0.010	0.024	0.015	0.001
8. *Arctostaphylos patula* (tw)	0.080	0.336	0.092	0.924	0.156	0.151	2.466	0.288	0.425	0.407	0.142
	0.017	0.047	0.014	0.099	0.021	0.019	0.091	0.345	0.039	0.064	0.015
9. *Quercus vacciniifolia* (tw)	0.073	0.017	0.083	2.175	0.365	0.377	4.641	1.029	0.835	0.171	0.363
	0.002	0.002	0.003	0.154	0.013	0.061	0.663	0.043	0.144	0.001	0.007
10. Forest litter (GAC, granular)	0.307	0.095	0.144	4.203	0.891	0.931	0.045	0.045	0.038	0.163	0.199
	0.047	0.010	0.011	0.296	0.050	0.044	0.029	0.002	0.023	0.011	0.005
11. Forest litter (LL, fine)	0.127	0.045	0.085	0.909	0.278	0.360	0.201	0.084	0.072	0.142	0.093
	0.020	0.001	0.001	0.022	0.006	0.010	0.008	0.002	0.001	0.004	0.000
12. Forest litter (LL, granular*)	0.163	0.054	0.122	2.102	0.474	0.546	0.587	0.181	0.147	0.189	0.177
	0.040	0.004	0.005	0.018	0.003	0.014	0.078	0.023	0.008	0.004	0.000
13. FLL lake surface pollen [II]	0.392	0.078	0.193	1.436	0.360	0.486	0.562	0.198	0.171	1.618	0.215
	0.049	0.012	0.039	0.107	0.022	0.002	0.059	0.014	0.019	0.070	0.071

Table 4.1: Survey of lignin concentrations in potential source materials, continued . . .

Sample	Σ V	Σ S	Σ C	Σ P	Tot. (Λ_8) V+S+C	% V	% S	% C	Vd/V	Sd/S	S/V	C/V	LPVI
1. *Pinus jeffreyi* (g)	5.751	0.033	0.884	0.665	6.668	86.2	0.5	13.3	0.153	0.000	0.006	0.154	3
	0.507	0.027	0.002	0.128	0.482	1.4	0.4	0.9	0.016	0.000	0.005	0.013	0
2. *Abies concolor* (g)	4.045	0.178	0.639	0.387	4.862	83.2	3.6	13.2	0.161	0.383	0.044	0.158	4
	0.194	0.021	0.011	0.025	0.226	0.1	0.3	0.4	0.004	0.022	0.003	0.005	0
3. *Abies concolor* (c)	4.561	0.050	0.477	0.425	5.087	89.6	1.0	9.4	0.203	0.321	0.011	0.105	2
	0.298	0.003	0.024	0.032	0.325	0.1	0.0	0.1	0.009	0.004	0.000	0.001	0
4. *Quercus vacciniifolia* (a)	2.905	2.761	0.797	0.616	6.464	45.0	42.7	12.3	0.204	0.208	0.950	0.274	191
	0.100	0.157	0.032	0.019	0.289	0.5	0.5	0.1	0.033	0.047	0.021	0.002	7
5. *Pinus jeffreyi* (b)	7.719	0.040	0.593	0.649	8.352	92.4	0.5	7.1	0.195	0.000	0.005	0.077	2
	0.022	0.004	0.003	0.008	0.024	0.0	0.1	0.1	0.004	0.000	0.001	0.001	0
6. *Pinus jeffreyi* (c)	6.812	0.040	0.155	0.570	7.007	97.2	0.6	2.2	0.219	0.000	0.006	0.023	1
	0.484	0.003	0.038	0.003	0.525	0.4	0.0	0.4	0.065	0.000	0.000	0.004	0
7. *Arctostaphylos patula* (a)	0.490	0.997	0.854	0.169	2.341	20.9	42.6	36.5	0.199	0.183	2.033	1.742	5414
	0.015	0.038	0.016	0.012	0.069	0.0	0.4	0.4	0.007	0.033	0.015	0.022	37
8. *Arctostaphylos patula* (tw)	1.232	3.179	0.549	0.508	4.959	24.8	64.1	11.1	0.164	0.172	2.608	0.445	1001
	0.139	0.215	0.079	0.045	0.004	2.8	4.4	1.6	0.003	0.009	0.470	0.014	102
9. *Quercus vacciniifolia* (tw)	2.918	6.505	0.533	0.173	9.956	29.3	65.3	5.4	0.175	0.184	2.228	0.183	306
	0.107	0.562	0.008	0.002	0.660	0.9	1.3	0.4	0.040	0.057	0.111	0.010	1
10. Forest litter (GAC, granular)	6.025	0.127	0.362	0.546	6.515	92.5	2.0	5.6	0.222	1.256	0.021	0.060	1
	0.390	0.007	0.016	0.068	0.399	0.3	0.2	0.1	0.005	1.318	0.003	0.001	0
11. Forest litter (LL, powdery)	1.547	0.356	0.235	0.257	2.138	72.3	16.7	11.0	0.396	0.358	0.230	0.152	14
	0.038	0.009	0.004	0.020	0.044	0.3	0.1	0.4	0.001	0.019	0.000	0.006	1
12. Forest litter (LL, granular*)	3.123	0.915	0.366	0.339	4.404	70.9	20.7	8.3	0.260	0.252	0.293	0.117	15
	0.007	0.109	0.004	0.049	0.113	1.6	2.0	0.3	0.009	0.019	0.034	0.002	2
13. FLL lake surface pollen [‖]	2.282	0.931	1.833	0.663	5.046	45.2	18.4	36.4	0.340	0.304	0.408	0.805	266
	0.131	0.093	0.002	0.100	0.225	0.6	1.0	1.6	0.024	0.001	0.017	0.046	3

Table 4.2: Lignin concentrations of BOLLY Core 1A, Sections 7 & 8. Concentrations are in mg phenol per 100 mg OC while ratios are unitless.

1A Section	Section Depth (cm)	Core Depth (cm)	cal. Age BP	TOC (%)	$\delta^{13}C_{PDB}$ (‰)	Pl	Pn	Pd	Vl	Vn	Vd	Sl	Sn	Sd	Cd	Fd
7	0	688	8993	4.42	-26.9	0.14	0.03	0.23	0.67	0.21	0.22	0.16	0.08	0.05	0.57	0.06
n=2						0.00	0.00	0.03	0.02	0.00	0.01	0.01	0.01	0.01	0.03	0.01
7	20	708	9379	4.39	-27.2	0.15	0.03	0.22	0.63	0.20	0.21	0.15	0.08	0.05	0.57	0.06
n=2						0.00	0.00	0.00	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00
7	40	728	9684	4.29	-27.2	0.14	0.03	0.19	0.56	0.18	0.19	0.14	0.09	0.05	0.53	0.05
n=2						0.00	0.00	0.01	0.04	0.00	0.01	0.01	0.00	0.00	0.01	0.01
7	60	748	9989	3.17	-27.0	0.14	0.03	0.19	0.63	0.20	0.21	0.16	0.09	0.05	0.48	0.06
n=2						0.01	0.00	0.01	0.05	0.01	0.01	0.01	0.00	0.00	0.03	0.01
7	80	768	10295	4.13	-27.4	0.16	0.02	0.22	0.55	0.18	0.20	0.14	0.09	0.04	0.61	0.04
n=2						0.01	0.00	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.03	0.00
7	100	788	10600	4.76	-27.4	0.16	0.02	0.20	0.47	0.17	0.21	0.11	0.08	0.04	0.54	0.04
n=2						0.03	0.00	0.03	0.12	0.01	0.04	0.02	0.00	0.02	0.01	0.01
7	120	808	10905	3.61	-27.6	0.17	0.02	0.20	0.65	0.20	0.23	0.20	0.11	0.05	0.56	0.04
n=2						0.01	0.00	0.02	0.02	0.01	0.02	0.00	0.01	0.01	0.01	0.00
7	140	828	11210	3.49	-27.2	0.20	0.03	0.31	0.52	0.20	0.24	0.13	0.08	0.06	0.78	0.06
n=2						0.03	0.00	0.04	0.05	0.00	0.03	0.01	0.01	0.01	0.03	0.01
8	0	839	11378	2.70	-26.4	0.20	0.03	0.31	0.53	0.19	0.21	0.14	0.07	0.05	0.95	0.05
n=2						0.03	0.00	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.08	0.01
8	20	859	11683	2.50	-25.8	0.18	0.02	0.23	0.45	0.19	0.23	0.12	0.08	0.06	0.64	0.04
n=3						0.02	0.01	0.04	0.03	0.00	0.02	0.00	0.01	0.00	0.05	0.00
8	40	879	11989	2.19	-26.4	0.16	0.03	0.28	0.43	0.15	0.16	0.08	0.05	0.03	1.14	0.04
n=2						0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.06	0.00
8	60	899	12299	1.63	-26.2	0.16	0.03	0.29	0.31	0.12	0.13	0.06	0.05	0.03	1.01	0.04
n=2						0.01	0.00	0.02	0.03	0.00	0.01	0.00	0.00	0.00	0.03	0.00
8	80	919	12646	1.23	-24.4	0.21	0.03	0.27	0.95	0.32	0.42	0.09	0.05	0.04	0.76	0.05
n=2						0.04	0.01	0.00	0.08	0.01	0.08	0.00	0.00	0.01	0.04	0.00
8	100	939	13075	1.33	-24.4	0.15	0.03	0.27	0.97	0.30	0.30	0.08	0.06	0.03	0.70	0.06
n=3						0.01	0.01	0.04	0.02	0.03	0.03	0.01	0.01	0.01	0.05	0.00
8	120	959	13533	1.49	-25.9	0.14	0.03	0.23	0.40	0.14	0.16	0.08	0.05	0.03	0.89	0.05
n=2						0.00	0.01	0.03	0.01	0.00	0.01	0.00	0.01	0.00	0.02	0.01

Table 4.2: Lignin concentrations of Bolly Core 1A, Sections 7 & 8, continued . . .

1A Section	Section Depth (cm)	Core Depth (cm)	cal. Age BP	Σ V	Σ S	Σ C	Σ P	Tot. (Λ_8) V+S+C	% V	% S	% C	Vd/V	Sd/S	S/V	C/V	P/V	LPVI
7	0	688	8993	1.10	0.30	0.63	0.40	2.02	54.2	14.7	31.1	0.33	0.30	0.27	0.57	0.36	100
n=2				0.01	0.03	0.02	0.03	0.03	1.6	1.1	0.5	0.02	0.01	0.03	0.03	0.04	19
7	20	708	9379	1.04	0.28	0.62	0.40	1.94	53.3	14.6	32.1	0.33	0.33	0.27	0.60	0.39	106
n=2				0.05	0.01	0.01	0.00	0.06	0.7	0.0	0.7	0.00	0.04	0.00	0.02	0.02	7
7	40	728	9684	0.94	0.28	0.58	0.36	1.80	52.1	15.7	32.2	0.34	0.33	0.30	0.62	0.38	125
n=2				0.03	0.01	0.02	0.01	0.06	0.0	0.1	0.1	0.03	0.04	0.00	0.00	0.02	1
7	60	748	9989	1.03	0.30	0.54	0.35	1.87	55.4	15.9	28.7	0.33	0.33	0.29	0.52	0.34	93
n=2				0.04	0.01	0.03	0.00	0.02	1.4	0.4	1.9	0.05	0.03	0.00	0.05	0.01	11
7	80	768	10295	0.93	0.26	0.65	0.40	1.84	50.4	14.4	35.2	0.36	0.25	0.29	0.70	0.43	138
n=2				0.00	0.02	0.03	0.00	0.06	1.5	0.7	0.8	0.05	0.06	0.02	0.04	0.01	23
7	100	788	10600	0.85	0.23	0.57	0.38	1.65	51.2	14.0	34.7	0.47	0.37	0.27	0.68	0.44	127
n=2				0.07	0.00	0.02	0.00	0.05	2.8	0.7	2.1	0.21	0.26	0.03	0.08	0.04	36
7	120	808	10905	1.07	0.36	0.60	0.39	2.03	52.8	17.8	29.4	0.35	0.23	0.34	0.56	0.37	127
n=2				0.05	0.00	0.01	0.04	0.06	0.7	0.4	0.3	0.02	0.03	0.01	0.01	0.02	10
7	140	828	11210	0.95	0.26	0.83	0.54	2.04	46.6	12.7	40.7	0.46	0.44	0.27	0.87	0.56	170
n=2				0.02	0.01	0.02	0.02	0.01	0.6	0.5	1.1	0.11	0.10	0.01	0.04	0.01	3
8	0	839	11378	0.93	0.27	1.00	0.54	2.19	42.4	12.2	45.4	0.41	0.37	0.29	1.07	0.59	233
n=2				0.04	0.03	0.08	0.01	0.15	1.1	0.4	0.6	0.04	0.03	0.02	0.04	0.02	29
8	20	859	11683	0.87	0.26	0.68	0.43	1.81	47.9	14.3	37.7	0.52	0.46	0.30	0.79	0.50	169
n=3				0.02	0.01	0.05	0.04	0.05	1.2	1.1	1.8	0.07	0.05	0.02	0.05	0.05	19
8	40	879	11989	0.74	0.17	1.18	0.48	2.09	35.5	8.2	56.3	0.38	0.43	0.23	1.59	0.64	273
n=2				0.01	0.00	0.06	0.01	0.06	0.8	0.3	1.1	0.01	0.03	0.00	0.07	0.01	9
8	60	899	12299	0.56	0.14	1.06	0.47	1.76	32.0	8.0	60.0	0.42	0.51	0.25	1.88	0.84	355
n=2				0.04	0.01	0.03	0.04	0.08	0.7	0.2	0.9	0.02	0.05	0.00	0.07	0.01	13
8	80	919	12646	1.70	0.18	0.81	0.50	2.69	63.0	6.9	30.2	0.44	0.46	0.11	0.48	0.30	29
n=2				0.01	0.01	0.05	0.05	0.06	1.1	0.0	1.0	0.12	0.07	0.00	0.02	0.03	3
8	100	939	13075	1.57	0.18	0.77	0.45	2.51	62.4	7.1	30.6	0.31	0.43	0.11	0.49	0.29	31
n=3				0.08	0.03	0.05	0.06	0.06	1.6	1.0	2.6	0.03	0.05	0.01	0.05	0.02	2
8	120	959	13533	0.69	0.16	0.94	0.40	1.79	38.6	9.0	52.4	0.39	0.41	0.23	1.36	0.58	234
n=2				0.01	0.00	0.03	0.03	0.04	0.1	0.2	0.3	0.00	0.03	0.00	0.01	0.05	2

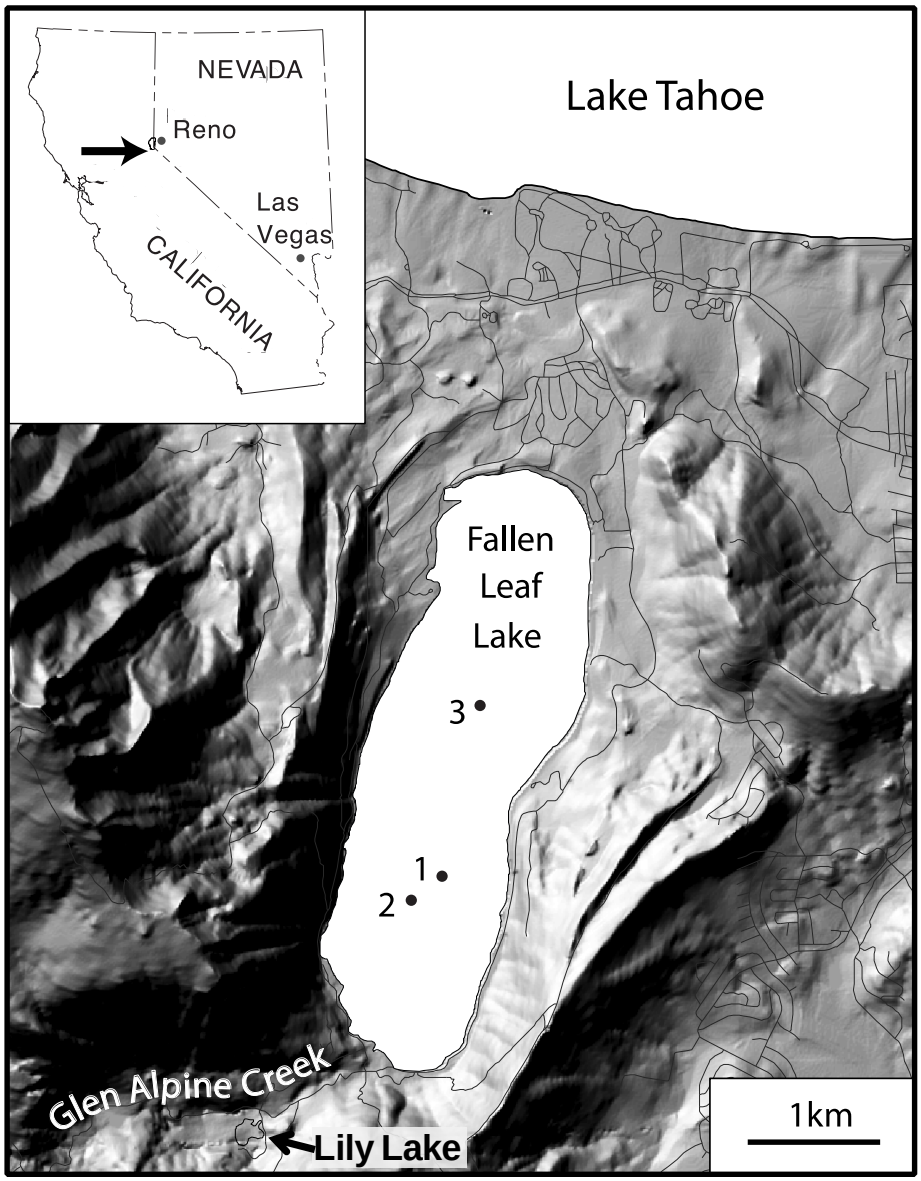


Figure 4.**1:** Location of Fallen Leaf Lake (FLL), El Dorado County, California, relative to Lake Tahoe. FLL's main tributary is Glen Alpine Creek (GAC). Lily Lake (LL) is also annotated. The locations of the three piston cores analyzed geochemically in this study are shown.

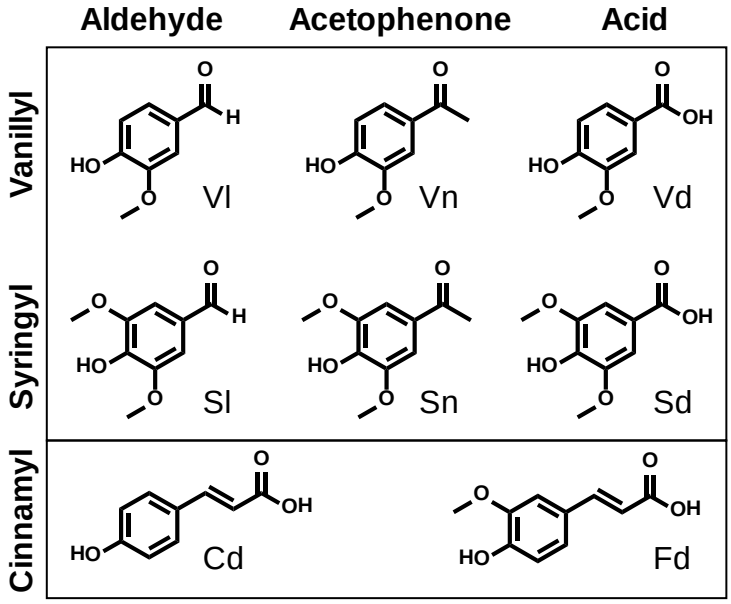


Figure 4.**2:** The eight phenol monomers unique to lignin that are produced upon its oxidation by alkaline CuO. Vl = vanillin; Vn = acetovanillone; Vd = vanillic acid; Sl = syringaldehyde; Sn = acetosyringone; Sd = syringic acid; Cd = *p*-coumaric acid; Fd = ferulic acid. The syringyl (S) phenols contain two methoxy groups while the vanillyl (V) phenols contain only one. Both S- and V-type phenol monomers are observed in their aldehydic, acetophenonic, and acidic forms. The p-hydroxy phenols (P), which contain zero methoxy groups, are also produced upon oxidation of lignin, but they have other sources as well. The P-type phenols also have three forms: *p*-hydroxy benzaldehyde (Pl), *p*-hydroxy acetophenone (Pn), and *p*-hydroxy benzoic acid (Pd).

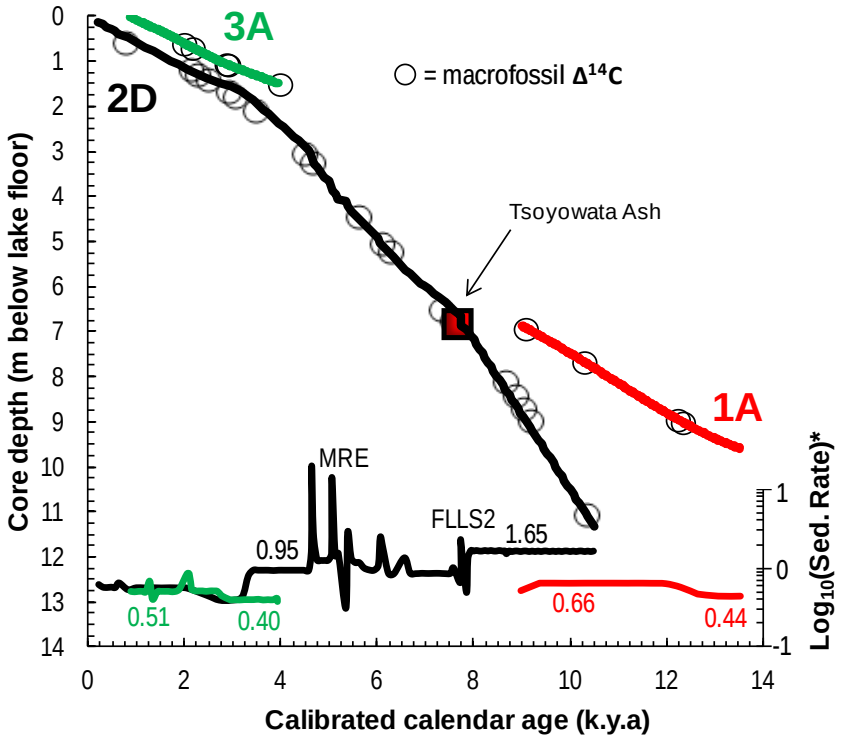
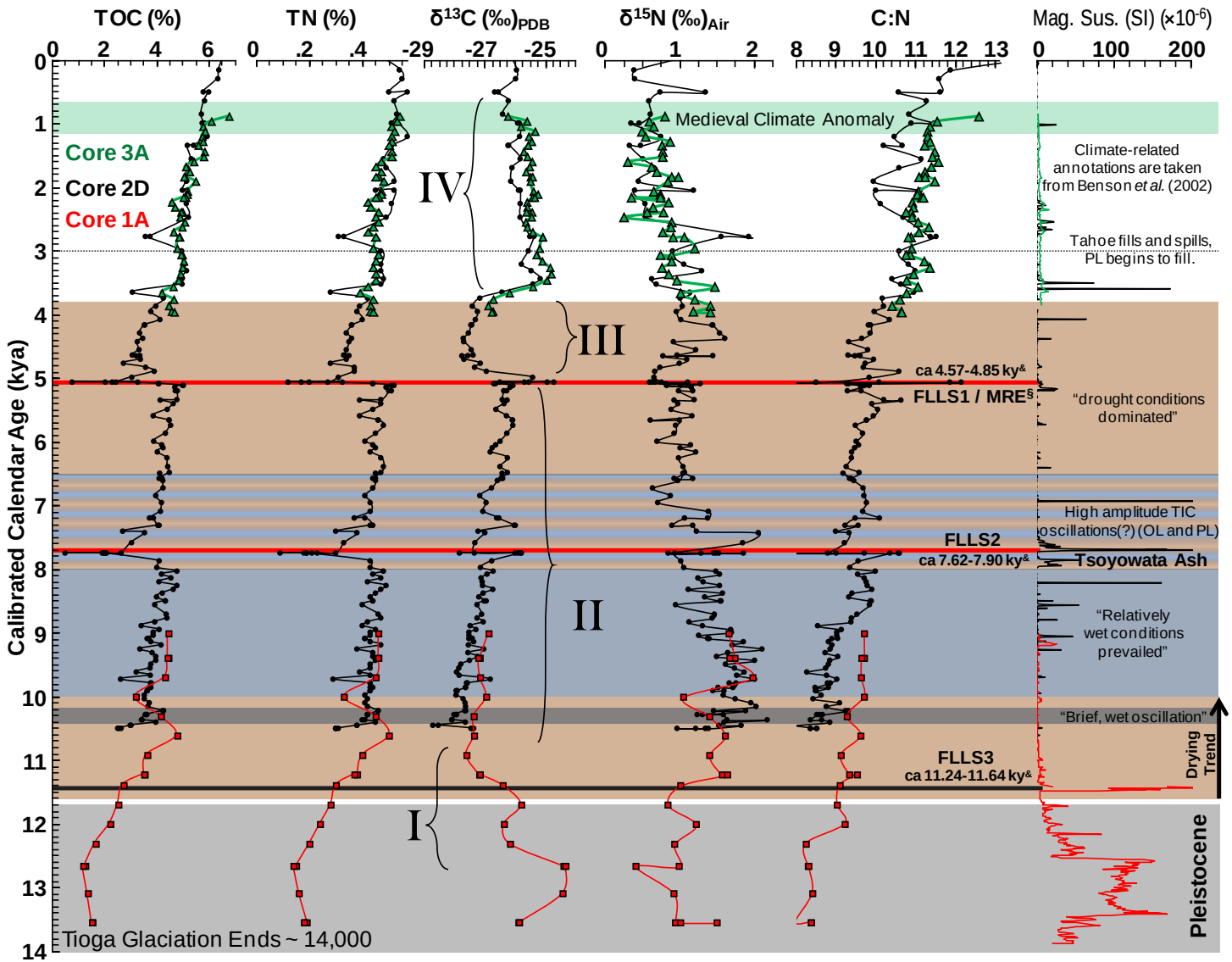


Figure 4.3: Shown are the age-depth relationships for each core, 3A (green), 2D (black), and 1A (red). Macrofossils were picked and analyzed for radiocarbon (^{14}C) by accelerator mass spectrometry (AMS). Calibrated calendar ages were determined for these macrofossil samples using OxCal v4.1.7 (Bronk Ramsey 2009) and the IntCal04 terrestrial radiocarbon age calibration curve for the last 26,000 cal years before present (Reimer et al. 2004). Sediment accumulation rates, expressed as $\log_{10}$ (mm $year^{-1}$), are shown for the sampled sections of each core. Several spikes are apparent in the sediment accumulation rate for core 2D. Two of these spikes correspond to previously described events, the most recent (seismic) event (MRE) (Brothers et al. 2009) and FLLS2 (Maloney et al. 2013). The measured sections of cores 3A and 1A had lower sediment accumulation rates than did core 2D. Core 2D contained the Tsoyowata ash marker bed, which underlies FLLS2. The ash's source has been attributed to the eruption of Mt Mazama, present day Crater Lake, Oregon and its age has been reported to be 7470–7620 cal years BP (Hallett et al. 1997).

Figure 4**.4:** Composite core composed of the analyzed sections of BOLLY cores 1A (red), 2D (black), and 3A (green) showing measured total organic carbon (TOC %), total nitrogen (TN), $\delta^{13}C_{PDB}$ (‰), $\delta^{15}N_{Air}$, C:N (mass), and magnetic susceptibility (SI units × 10^{6}). Measurements were performed throughout all of 2D, the topmost section of 3A, and the bottom two sections of 1A. The section analyzed from 3A fully complemented the analyses from 2D while 1A partially complemented the analyses of 2D while also extending the record of the composite core into the Pleistocene, as shown. & = FLLS1, FLLS2, and FLLS3 and corresponding preferred event ages are those reported by (Maloney et al. 2013). The elemental and isotope measurements sampled are through FLLS1 and FLLS2. FLLS3, which was thinner (~1 cm) in core 1A, was not sampled. Shaded regions are reflective of the changing wet and dry climatic conditions of the Great Basin during the Holocene as described by (Benson et al. 2002) based on records from Pyramid Lake (PL) and Owens Lake (OL) sediments. The indicated striped oscillations in the figure are not to be interpreted literally. The Medieval Climate Anomaly (MCA), the approximate end-point of the Tioga glaciation, and the Tsoyowata ash bed are also indicated for reference. § = Most recent (seismic) event (MRE) as identified by (Brothers et al. 2009) and the ostensible trigger for the FLLS1 turbidite. The $\delta^{13}C$ record has been broken into four sections, I–IV. MCA dates are from (Kleppe et al. 2011).


TOC (%)
TN (%)
δ13C (‰)PDB
δ15N (‰)Air
C:N
Mag. Sus. (SI) (×10-6)
Calibrated Calendar Age (kya)
Core 3A
Core 2D
Core 1A
Medieval Climate Anomaly
IV
III
II
I
ca 4.57-4.85 ky&
FLLS1 / MRE§
FLLS2
ca 7.62-7.90 ky&
FLLS3
ca 11.24-11.64 ky&
Climate-related annotations are taken from Benson et al. (2002)
Tahoe fills and spills, PL begins to fill.
"drought conditions dominated"
High amplitude TIC oscillations(?) (OL and PL)
Tsoyowata Ash
"Relatively wet conditions prevailed"
"Brief, wet oscillation"
Drying Trend
Pleistocene
Tioga Glaciation Ends ~ 14,000

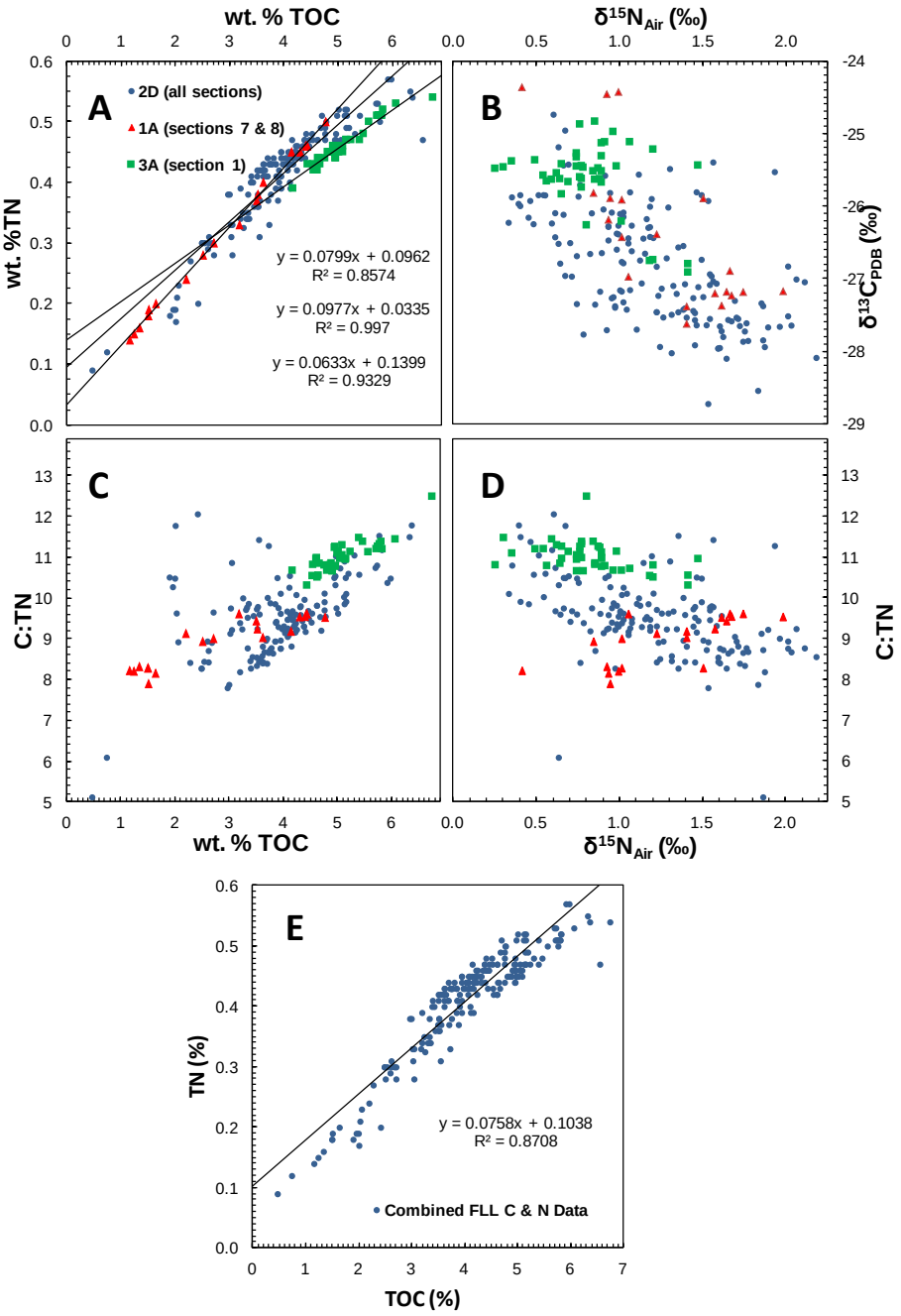


Figure 4.5**:** Cross-plots of bulk measured organic geochemical parameters are shown as follows and are broken down by core for (A–D) and include all samples for (E): (A) TN (%) vs. TOC (%), (B) $\delta^{15}N_{Air}$ (‰) vs. $\delta^{13}C_{PDB}$ (‰), (C) C:TN (mass) vs. TOC (%), (D) $\delta^{15}N_{Air}$ (‰) vs. TOC (%), and (E) TN (%) vs. TOC (%) for all samples. The intercepts of the linear regressions shown in (A) indicate the amount of nitrogen assumed to be present in the absence of carbon and, as such, provide a rough estimate of average total inorganic nitrogen (TIN) present in samples analyzed from each core.

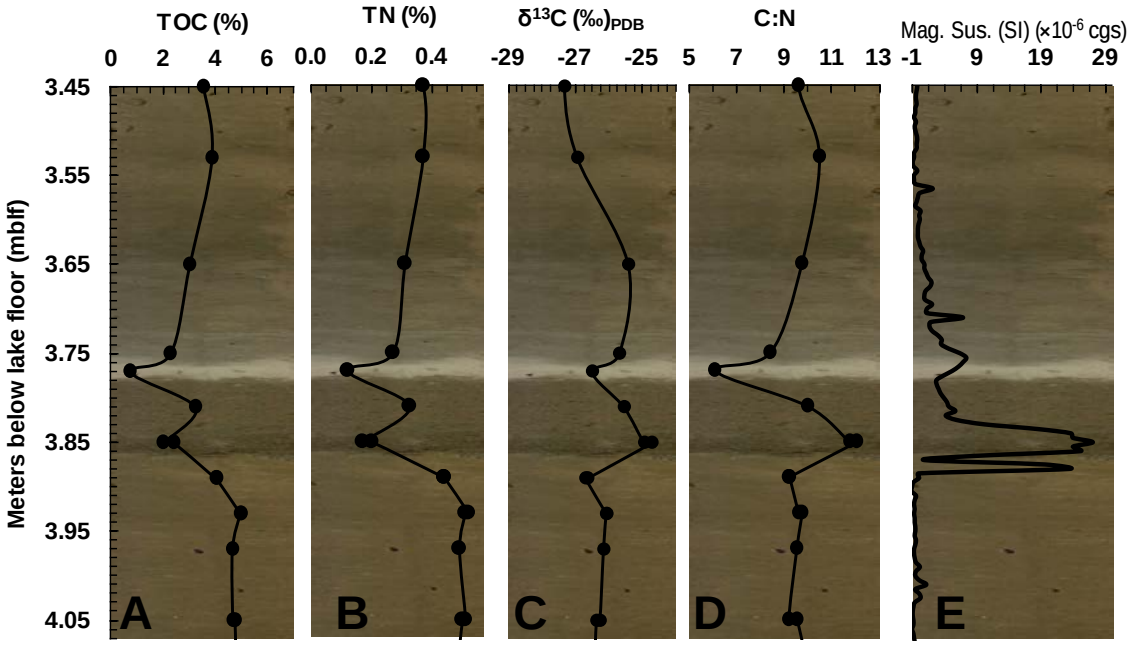


Figure 4.6: Elemental and isotopic analyses across the MRE turbidite (3.75–3.86 mblf) as preserved in core 2D, plotted by depth showing (A) TOC (%), (B) TN (%), (C) $\delta^{13}C$ (‰), (D) $\delta^{15}N$ (‰), and (E) magnetic susceptibility. The plots are each underlain by the same photograph of the MRE turbidite as it was visible in 2D. The turbidite has a dark, silty base and an organic rich top (3.75-3.78 mblf) overlain by a gray clay. $\delta^{15}N$ values did not appear to be significantly affected across the turbidite.

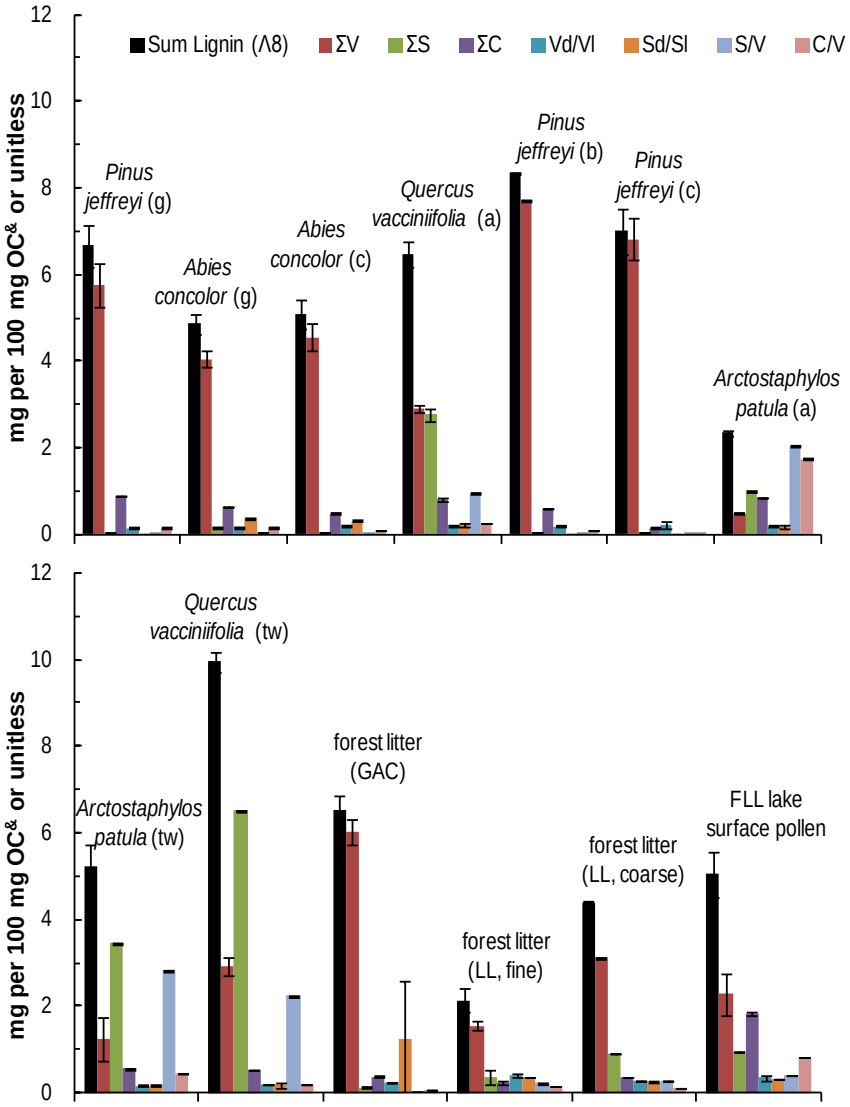


Figure 4.7: Lignin concentrations mg phenol (100 mg OC)$^{-1}$ of modern-day potential OM source materials to Fallen Leaf Lake's (FLL) sediments. All ratios are unitless. LL = Lily Lake; GAC = Glen Alpine Creek. $^{\&}$ = 50% OC was assumed for these terrestrial samples. g = gymnosperm needles (nonwoody); c = gymnosperm cone (woody); a = angiosperm leaf (nonwoody); b = gymnosperm bark (woody); tw = twig. Both twigs were woody tissues from angiosperms. ΣV = sum of vanillyl-type phenols; ΣS = sum of syringyl-type phenols; ΣC = sum of cinnamyl-type phenols; Vd/Vl = ratio of vanillic acid to vanillin; Sd/Sl = ratio of syringic acid to syringaldehyde; S/V = ratio of S- to V-type phenols; C/V = ratio of C-type to V-type phenols.

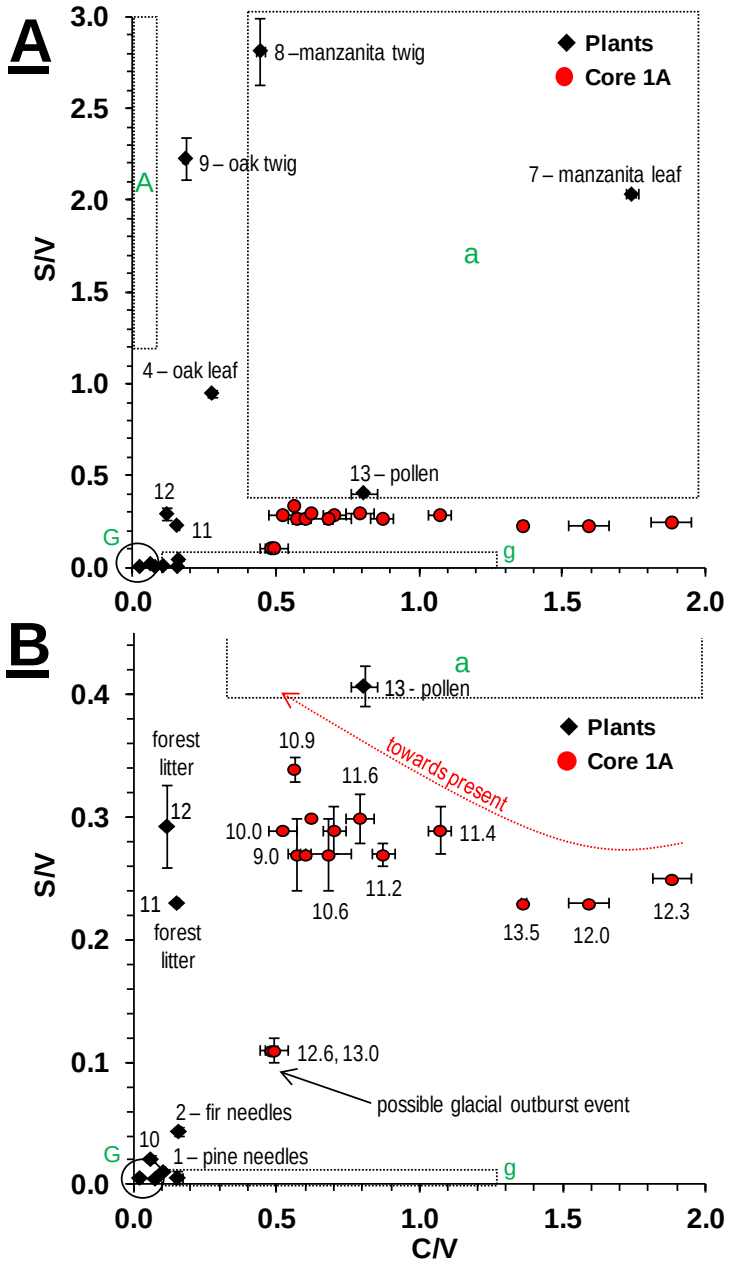


Figure 4.8: Syringyl/vanillyl (S/V) vs. cinnamyl/vanillyl (C/V) ratios are plotted for plant and soil specimens (diamonds, black) and for analyzed intervals of sections 7 and 8 of BOLLY core 1A (circles, red) (A). Data are contained in Tables 1 and 2. An enlargement of (A) emphasizing the samples from core 1A is shown in (B). The green letters and their corresponding regions broadly encompass the S/V and C/V ratios characteristic of the nonwoody and woody plant types of angiosperms and gymnosperms, including woody angiosperm tissue (A), woody gymnosperm tissue (G), nonwoody angiosperm tissue (a), and nonwoody gymnosperm tissue (g). The regions shown have been expanded from those originally reported by (Hedges and Mann 1979) to encompass the high observed C/V ratio of (7), a nonwoody angiosperm tissue sample from *Arctostaphylos patula* (greenleaf manzanita). Numbers corresponding to samples (Table 1) are shown next to black diamonds (see Table1). Ages (cal. kya BP) are shown next to red circles.

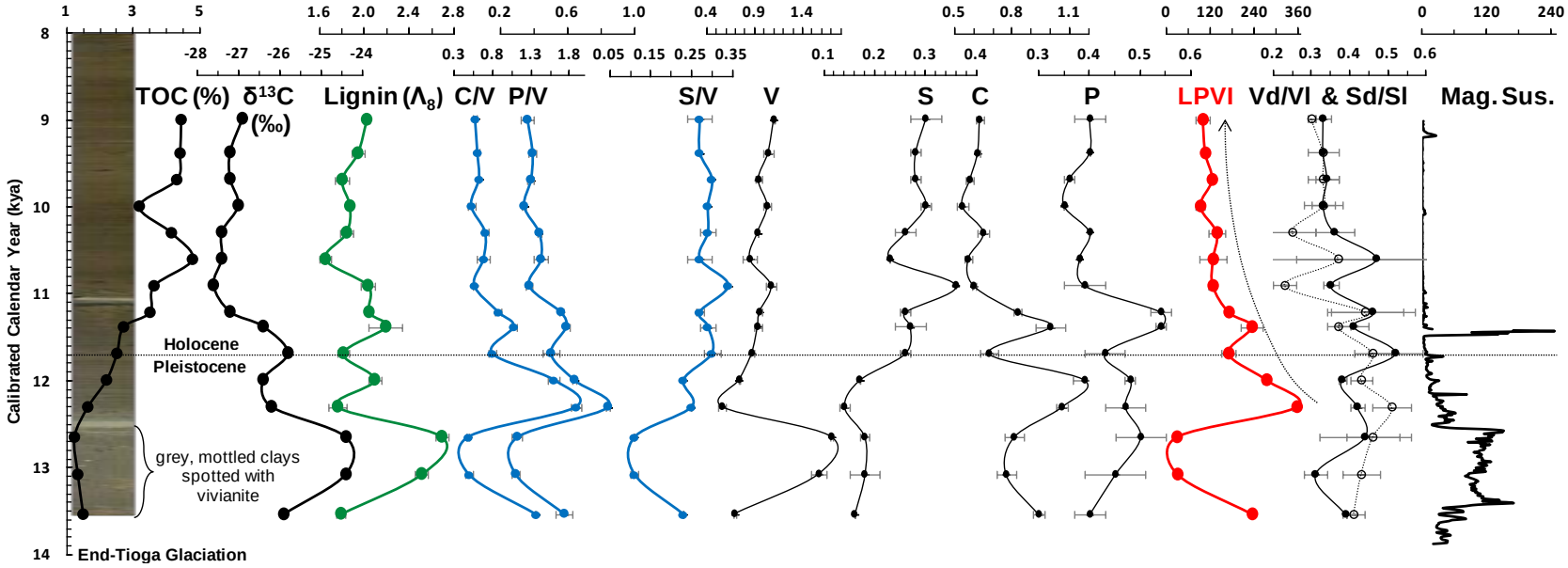


Figure 4.9: Lignin and organic geochemical parameters for 15 samples analyzed across sections 7 and 8 of core 1A, plotted vs. calibrated calendar age. Shown are TOC (%), $\delta^{13}C_{PDB}$ (‰), total lignin (Λ_8) (mg per 100 mg OC), which is the sum of the vanillyl (V), syringyl (S), and cinnamyl (C) lignin phenols. Carbon-normalized concentrations (mg per 100 mg OC) of the V, S, and C-type phenols are also shown, plus those of the p-hydroxyphenyl monomers (P). Weight ratios of C/V, P/V and S/V, which show similar profiles to one another, are also shown. All ratios are unitless. The lignin phenol vegetation index (LPVI), expressed as LPVI = [S(S+1)/(V+1)+1]×[C(C+1)/(V+1)+1] was used by (Tareq et al. 2004) to broadly apportion plant source types to sedimentary organic matter (OM) based on the relative abundances of V-, S-, and C-type phenol types observed in sediments. Lower LPVI numbers are characteristic of woody (LPVI = 1) and nonwoody gymnosperms (12-27) while higher numbers are reflective of woody angiosperms (67-415), and the highest numbers are reflective of nonwoody angiosperms (378-2782). The vanillin to vanillic acid (Vd/Vl) and syringaldehyde to syringic acid (Sd/Sl) weight ratios are both shown overlain on top of one another. Magnetic susceptibility is shown in the final column (SI units × 10^6). At left, and partially overlain by TOC (%), are spliced photographs of sections 7 and 8 from core 1A as they visually appear. Error bars for all lignin parameters were derived from at least duplicate oxidations and denote ± 1σ. Error bars were too small to see for either TOC or $\delta^{13}C$.

Table 4.S1: BOLLY C and N elemental and isotope data. All data collected for cores 2D, 1A, and 3A are shown, including replicates. Abundances are by weight. cmblf = cm below lake floor.

Core	Top (cmblf)	Base (cmblf)	Median Age top (BP)	TOC (%)	TN (%)	$\delta^{13}C_{PDB}$ (‰)	$\delta^{15}N_{AIR}$ (‰)	C:N (mass)
2D	12.0	12.5	-55	6.55	0.47	-26.1	1.0	13.94
2D	21.0	21.5	146	6.36	0.54	-26.0	0.4	11.78
2D	29.0	29.5	285	6.32	0.55	-26.0	0.4	11.49
2D	41.0	41.5	492	5.76	0.50	-26.6	1.4	11.52
2D	41.0	41.5	492	5.97	0.57	-26.7	0.7	10.47
2D	49.0	49.5	631	5.81	0.52	-26.2	0.6	11.17
2D	61.0	61.5	840	5.69	0.53	-26.4	0.6	10.74
2D	69.0	69.5	981	5.80	0.51	-25.9	0.5	11.37
2D	69.0	69.5	981	5.72	0.53	-25.9	0.4	10.79
2D	81.0	81.5	1191	5.91	0.57	-25.9	0.8	10.37
2D	89.0	89.5	1331	5.39	0.51	-26.2	0.5	10.57
2D	89.0	89.5	1331	5.15	0.51	-26.2	0.3	10.10
2D	101.0	101.5	1542	5.30	0.48	-25.8	0.5	11.04
2D	109.0	109.5	1682	5.14	0.49	-26.1	0.7	10.49
2D	121.0	121.5	1894	5.12	0.52	-26.2	0.5	9.85
2D	128.0	128.5	2035	4.96	0.45	-25.9	1.2	11.02
2D	129.0	129.5	2035	5.15	0.52	-25.8	0.4	9.90
2D	129.0	129.5	2245	5.11	0.51	-25.8	0.5	10.02
2D	136.5	137.0	2460	5.19	0.49	-25.8	0.6	10.59
2D	148.5	149.0	2762	3.72	0.33	-25.5	1.9	11.27
2D	148.5	149.0	2762	3.54	0.31	-25.4	1.6	11.42
2D	157.5	158.0	2989	4.93	0.47	-25.6	0.9	10.49
2D	167.0	167.5	3197	5.04	0.47	-25.8	1.1	10.72
2D	175.0	175.5	3300	5.12	0.47	-25.5	1.3	10.89
2D	187.0	187.5	3426	4.95	0.48	-25.2	0.6	10.31
2D	195.0	195.5	3510	4.95	0.47	-25.4	0.7	10.53
2D	207.0	207.5	3637	3.04	0.28	-26.4	1.1	10.86
2D	216.0	216.5	3731	4.23	0.42	-27.2	1.0	10.07
2D	228.0	228.5	3857	3.94	0.39	-27.4	1.0	10.10
2D	236.0	236.5	3941	3.75	0.38	-27.2	1.0	9.87
2D	248.0	248.5	4068	4.10	0.40	-27.3	1.0	10.25
2D	256.0	256.5	4152	3.50	0.36	-27.4	1.4	9.72
2D	256.0	256.5	4152	3.52	0.36	-27.5	1.5	9.78
2D	268.0	268.5	4278	3.32	0.34	-27.5	1.5	9.76
2D	276.0	276.5	4362	3.44	0.36	-27.7	1.6	9.56
2D	281.0	281.5	4415	3.23	0.35	-27.7	0.9	9.23
2D	293.0	293.5	4541	3.30	0.34	-27.5	1.2	9.71
2D	301.0	301.5	4625	3.04	0.33	-27.4	1.0	9.21
2D	309.0	309.5	4634	3.33	0.35	-27.6	1.5	9.51
2D	317.0	317.5	4637	3.19	0.34	-27.8	0.8	9.38
2D	329.0	329.5	4691	3.35	0.34	-27.7	1.1	9.85
2D	337.0	337.5	4754	2.70	0.28	-27.2	1.0	9.64
2D	345.0	345.5	4817	3.55	0.37	-27.3	0.7	9.59
2D	353.0	353.5	4880	3.88	0.37	-27.0	0.7	10.49
2D	365.0	365.5	4975	3.02	0.31	-25.4	0.7	9.74
2D	375.0	375.5	5050	2.27	0.27	-25.7	0.8	8.41
2D	377.0	377.5	5050	0.73	0.12	-26.5	0.6	6.08
2D	381.0	381.5	5050	3.25	0.33	-25.6	1.1	9.99

Table 4.S1: BOLLY C and N elemental and isotope data, continued . . .

Core	Top (cmblf)	Base (cmblf)	Median Age top (BP)	TOC (%)	TN (%)	$\delta^{13}C_{PDB}$ (‰)	$\delta^{15}N_{AIR}$ (‰)	C:N (mass)
2D	385.0	385.5	5050	2.00	0.17	-25.0	0.7	11.76
2D	385.0	385.5	5050	2.41	0.20	-24.7	0.6	12.05
2D	389.0	389.5	5074	4.04	0.44	-26.7	0.8	9.18
2D	389.0	389.5	5074	4.06	0.44	-26.7	1.3	9.23
2D	393.0	393.5	5102	5.00	0.52	-26.1	1.0	9.62
2D	393.0	393.5	5102	4.97	0.51	-26.1	0.8	9.75
2D	397.0	397.5	5130	4.67	0.49	-26.2	1.2	9.53
2D	405.0	405.5	5186	4.69	0.51	-26.3	1.2	9.20
2D	405.0	405.5	5186	4.77	0.50	-26.4	1.0	9.54
2D	413.0	413.5	5325	4.75	0.47	-26.4	1.2	10.11
2D	425.0	425.5	5332	4.11	0.39	-26.3	0.9	10.54
2D	433.0	433.5	5388	4.61	0.47	-26.4	1.0	9.81
2D	445.0	445.5	5478	4.38	0.44	-26.6	0.9	9.95
2D	457.0	457.5	5579	3.84	0.39	-26.3	1.2	9.85
2D	465.0	465.5	5646	4.45	0.46	-26.1	0.6	9.67
2D	465.0	465.5	5646	4.45	0.46	-26.1	1.0	9.67
2D	473.0	473.5	5728	4.51	0.48	-26.1	1.0	9.40
2D	485.0	485.5	5853	4.31	0.45	-26.3	0.9	9.58
2D	497.0	497.5	5979	3.86	0.41	-26.5	0.7	9.41
2D	505.5	506.0	6035	4.17	0.44	-26.6	1.2	9.48
2D	517.0	517.5	6081	4.22	0.45	-26.8	1.0	9.38
2D	525.0	525.5	6147	4.00	0.43	-26.8	1.2	9.30
2D	533.0	533.5	6239	4.37	0.47	-26.3	1.0	9.30
2D	545.0	545.5	6377	4.40	0.48	-26.5	1.1	9.17
2D	553.0	553.5	6469	4.46	0.47	-26.3	1.1	9.49
2D	554.0	554.5	6562	4.07	0.44	-26.5	0.9	9.25
2D	561.0	561.5	6562	4.18	0.45	-26.4	1.2	9.29
2D	561.0	561.5	6481	4.09	0.45	-26.4	1.1	9.09
2D	564.0	564.5	6596	4.21	0.45	-26.6	1.0	9.36
2D	574.0	574.5	6712	4.22	0.44	-26.8	0.6	9.59
2D	584.5	585.0	6833	3.94	0.41	-27.2	0.9	9.61
2D	594.0	594.5	6942	4.16	0.43	-27.0	0.7	9.67
2D	606.0	606.5	7080	4.12	0.43	-27.1	1.4	9.58
2D	615.0	615.5	7184	3.70	0.37	-26.5	1.4	10.00
2D	615.0	615.5	7184	3.85	0.41	-26.6	1.1	9.39
2D	625.0	625.5	7299	4.02	0.44	-26.0	0.9	9.14
2D	625.0	625.5	7299	4.07	0.43	-26.0	1.2	9.47
2D	633.0	633.5	7392	2.67	0.30	-27.2	1.2	8.90
2D	635.0	635.5	7415	3.51	0.38	-27.0	2.1	9.24
2D	655.0	655.5	7573	3.01	0.33	-27.3	1.9	9.12
2D	663.0	663.5	7735	2.61	0.30	-27.3	0.9	8.70
2D	671.0	671.5	7735	2.05	0.23	-25.9	1.5	8.91
2D	671.0	671.5	7735	2.02	0.21	-26.0	1.3	9.62
2D	676.5	677.0	7735	0.46	0.09	-27.8	1.9	5.11
2D	679.0	679.5	7735	1.89	0.18	-25.8	1.5	10.50
2D	679.0	679.5	7735	1.99	0.19	-25.8	1.5	10.47
2D	679.0	679.5	7735	1.95	0.19	-25.8	0.9	10.26
2D	692.5	693.0	7858	4.07	0.43	-26.8	1.0	9.47
2D	710.0	710.5	7962	3.98	0.43	-27.1	1.1	9.26

Table 4.S1: BOLLY C and N elemental and isotope data, continued . . .

Core	Top (cmblf)	Base (cmblf)	Median Age top (BP)	TOC (%)	TN (%)	$\delta^{13}C_{PDB}$ (‰)	$\delta^{15}N_{AIR}$ (‰)	C:N (mass)
2D	720.0	720.5	8019	4.75	0.48	-26.7	1.5	9.90
2D	728.0	728.5	8066	4.16	0.44	-26.9	1.5	9.45
2D	737.5	738.0	8126	4.52	0.47	-27.1	1.3	9.62
2D	748.0	748.5	8186	4.06	0.42	-27.1	1.3	9.67
2D	757.5	758.0	8245	4.75	0.49	-27.1	1.5	9.69
2D	768.0	768.5	8305	4.61	0.47	-26.7	1.1	9.81
2D	777.5	778.0	8364	4.19	0.45	-27.1	1.6	9.31
2D	788.0	788.5	8424	3.97	0.43	-27.2	1.3	9.23
2D	797.5	798.0	8487	4.31	0.44	-27.0	1.6	9.80
2D	797.5	798.0	8487	4.30	0.44	-27.0	1.6	9.77
2D	808.5	809.0	8546	3.90	0.40	-27.3	1.0	9.75
2D	833.0	833.5	8695	4.34	0.46	-27.1	1.5	9.43
2D	843.0	843.5	8755	4.37	0.47	-27.3	1.4	9.30
2D	853.0	853.5	8816	3.90	0.42	-27.1	1.1	9.29
2D	863.0	863.5	8876	3.38	0.40	-27.5	1.3	8.45
2D	873.0	873.5	8936	4.07	0.45	-27.3	1.7	9.04
2D	880.0	880.5	8979	3.84	0.43	-27.5	1.7	8.93
2D	888.0	888.5	9027	3.84	0.43	-27.5	1.3	8.93
2D	896.0	896.5	9075	3.67	0.41	-27.4	1.4	8.95
2D	896.5	897.0	9078	3.61	0.41	-27.5	1.4	8.80
2D	904.5	905.0	9127	3.75	0.43	-27.4	1.9	8.72
2D	913.0	913.5	9178	4.14	0.47	-27.5	1.7	8.81
2D	923.5	924.0	9242	3.33	0.38	-27.0	2.1	8.76
2D	933.0	933.5	9299	3.80	0.44	-27.6	1.7	8.64
2D	943.0	943.5	9359	3.94	0.44	-27.2	1.5	8.95
2D	953.0	953.5	9420	3.94	0.45	-27.5	2.0	8.76
2D	963.0	963.5	9480	3.75	0.43	-27.8	1.6	8.72
2D	973.0	973.5	9541	3.72	0.43	-27.9	1.7	8.65
2D	983.5	984.0	9604	3.19	0.39	-27.9	1.9	8.18
2D	992.0	992.5	9655	3.73	0.43	-27.8	1.8	8.67
2D	1002.0	1002.5	9716	2.59	0.29	-26.8	2.0	8.93
2D	1012.0	1012.5	9776	3.74	0.43	-27.6	1.7	8.70
2D	1012.0	1012.5	9776	3.77	0.43	-27.6	1.8	8.77
2D	1023.5	1024.0	9846	3.69	0.44	-27.6	1.7	8.39
2D	1023.5	1024.0	9846	3.75	0.43	-27.6	1.5	8.72
2D	1032.0	1032.5	9897	3.62	0.43	-27.9	1.5	8.42
2D	1032.0	1032.5	9897	3.61	0.43	-27.9	1.6	8.40
2D	1043.5	1044.0	9966	3.50	0.41	-28.0	1.6	8.54
2D	1053.0	1053.5	10024	3.50	0.42	-27.7	1.8	8.33
2D	1063.5	1064.0	10087	3.69	0.41	-27.7	2.0	9.00
2D	1073.0	1073.5	10145	3.64	0.42	-27.6	2.0	8.67
2D	1084.0	1084.5	10211	4.21	0.46	-27.6	1.5	9.15
2D	1084.0	1084.5	10211	4.24	0.46	-27.7	1.9	9.22
2D	1094.0	1094.5	10272	3.60	0.42	-28.0	1.3	8.57
2D	1094.0	1094.5	10272	3.57	0.42	-27.9	1.2	8.50
2D	1094.0	1094.5	10272	3.54	0.42	-28.0	1.6	8.43
2D	1104.5	1105.0	10355	3.39	0.41	-28.1	1.6	8.27
2D	1104.5	1105.0	10355	3.42	0.40	-28.1	2.2	8.55
2D	1114.0	1114.5	10392	3.94	0.45	-27.6	1.6	8.76

Table 4.S1: BOLLY C and N elemental and isotope data, continued . . .

Core	Top (cmblf)	Base (cmblf)	Median Age top (BP)	TOC (%)	TN (%)	$\delta^{13}C_{PDB}$ (‰)	$\delta^{15}N_{AIR}$ (‰)	C:N (mass)
2D	1122.0	1122.5	10441	2.96	0.38	-28.7	1.5	7.79
2D	1122.0	1122.5	10441	2.99	0.38	-28.5	1.8	7.87
2D	1130.0	1130.5	10489	2.53	0.30	-27.4	1.4	8.43
2D	1130.0	1130.5	10489	2.48	0.30	-27.4	1.0	8.27
2D	1130.0	1130.5	10489	2.61	0.31	-27.5	1.2	8.42
2D	1130.0	1130.5	10489	2.48	0.30	-27.5	1.4	8.27
Core 1A								
1A	688.0	688.5	8993	4.42	0.46	-26.9	1.7	9.61
1A	708.0	708.5	9379	4.42	0.46	-27.2	1.7	9.61
1A	708.0	708.5	9379	4.39	0.46	-27.2	1.7	9.54
1A	728.0	728.5	9684	4.29	0.45	-27.2	2.0	9.53
1A	748.0	748.5	9989	3.17	0.33	-27.0	1.1	9.61
1A	768.0	768.5	10295	4.13	0.45	-27.4	1.4	9.18
1A	788.0	788.5	10600	4.76	0.50	-27.4	1.6	9.52
1A	808.0	808.5	10905	3.61	0.40	-27.6	1.4	9.03
1A	828.0	828.5	11210	3.51	0.38	-27.2	1.6	9.24
1A	828.0	828.5	11210	3.49	0.37	-27.2	1.6	9.43
1A	839.0	839.5	11378	2.70	0.30	-26.4	1.0	9.00
1A	859.0	859.5	11683	2.50	0.28	-25.8	0.8	8.93
1A	879.0	879.5	11989	2.19	0.24	-26.4	1.2	9.13
1A	899.0	899.5	12299	1.63	0.20	-26.2	0.9	8.15
1A	919.0	919.5	12646	1.23	0.15	-24.4	1.0	8.20
1A	919.0	919.5	12646	1.15	0.14	-24.4	0.4	8.21
1A	939.0	939.5	13075	1.33	0.16	-24.4	0.9	8.31
1A	959.0	959.5	13533	1.50	0.19	-25.9	0.9	7.89
1A	959.0	959.5	13533	1.49	0.18	-25.9	1.0	8.28
1A	959.0	959.5	13533	1.49	0.18	-25.9	1.5	8.28
Core 3A								
3A	1.5	2.0	870	6.74	0.54	-26.3	0.8	12.48
3A	5.5	6.0	949	6.06	0.53	-25.6	0.6	11.43
3A	9.5	10.0	1029	5.74	0.51	-25.8	0.7	11.25
3A	13.5	14.0	1108	5.82	0.52	-25.4	0.5	11.19
3A	17.5	18.0	1187	5.71	0.51	-25.6	0.5	11.20
3A	21.5	22.0	1267	5.74	0.51	-25.5	0.9	11.25
3A	25.5	26.0	1326	5.56	0.50	-25.6	0.8	11.12
3A	29.5	30.0	1425	5.80	0.51	-25.7	0.8	11.37
3A	33.5	34.0	1505	5.77	0.51	-25.6	0.8	11.31
3A	37.5	38.0	1584	5.39	0.47	-25.5	0.3	11.47
3A	41.5	42.0	1663	5.08	0.45	-25.5	0.6	11.29
3A	45.5	46.0	1742	5.23	0.47	-25.7	0.7	11.13
3A	49.5	50.0	1822	5.05	0.46	-25.5	0.9	10.98
3A	49.5	50.0	1822	5.01	0.45	-25.4	1.0	11.13
3A	53.0	53.5	1891	5.46	0.48	-25.5	0.8	11.38
3A	61.5	62.0	2048	5.15	0.47	-25.4	0.7	10.96
3A	65.5	66.0	2096	5.14	0.47	-25.3	0.7	10.94
3A	69.5	70.0	2144	5.08	0.46	-25.4	0.7	11.04
3A	69.5	70.0	2144	4.99	0.45	-25.4	0.4	11.09

Table 4.S1: BOLLY C and N elemental and isotope data, continued . . .

Core	Top (cmblf)	Base (cmblf)	Median Age top (BP)	TOC (%)	TN (%)	$\delta^{13}C_{PDB}$ (‰)	$\delta^{15}N_{AIR}$ (‰)	C:N (mass)
3A	73.5	74.0	2219	4.54	0.42	-25.6	0.9	10.81
3A	77.5	78.0	2297	4.66	0.43	-25.6	0.6	10.84
3A	81.5	82.0	2375	4.96	0.46	-25.7	0.6	10.78
3A	81.5	82.0	2375	4.90	0.46	-25.5	0.8	10.65
3A	85.5	86.0	2453	4.86	0.45	-25.5	0.3	10.80
3A	89.5	90.0	2532	5.04	0.46	-25.7	0.9	10.96
3A	93.5	94.0	2610	4.94	0.44	-25.6	0.9	11.23
3A	97.5	98.0	2688	4.61	0.42	-25.4	0.8	10.98
3A	101.5	102.0	2767	4.85	0.45	-25.1	0.9	10.78
3A	101.5	102.0	2767	4.82	0.45	-25.1	1.1	10.71
3A	109.5	110.0	2945	4.75	0.44	-25.2	1.2	10.80
3A	113.5	114.0	3045	4.95	0.46	-25.3	0.9	10.76
3A	113.5	114.0	3045	4.90	0.46	-25.3	0.7	10.65
3A	117.5	118.0	3145	5.00	0.45	-25.1	0.9	11.11
3A	121.5	122.0	3246	4.95	0.44	-24.9	0.8	11.25
3A	125.5	126.0	3346	4.88	0.45	-24.8	0.9	10.84
3A	129.5	130.0	3446	4.80	0.45	-25.0	1.0	10.67
3A	133.5	134.0	3546	4.60	0.42	-25.4	1.5	10.95
3A	137.5	138.0	3645	4.16	0.39	-26.2	1.0	10.67
3A	141.5	142.0	3746	4.62	0.44	-26.7	1.2	10.50
3A	145.5	146.0	3846	4.43	0.43	-26.9	1.4	10.30
3A	149.0	149.5	3934	4.53	0.43	-26.7	1.2	10.53
3A	149.5	150.0	3946	4.64	0.44	-26.8	1.4	10.55

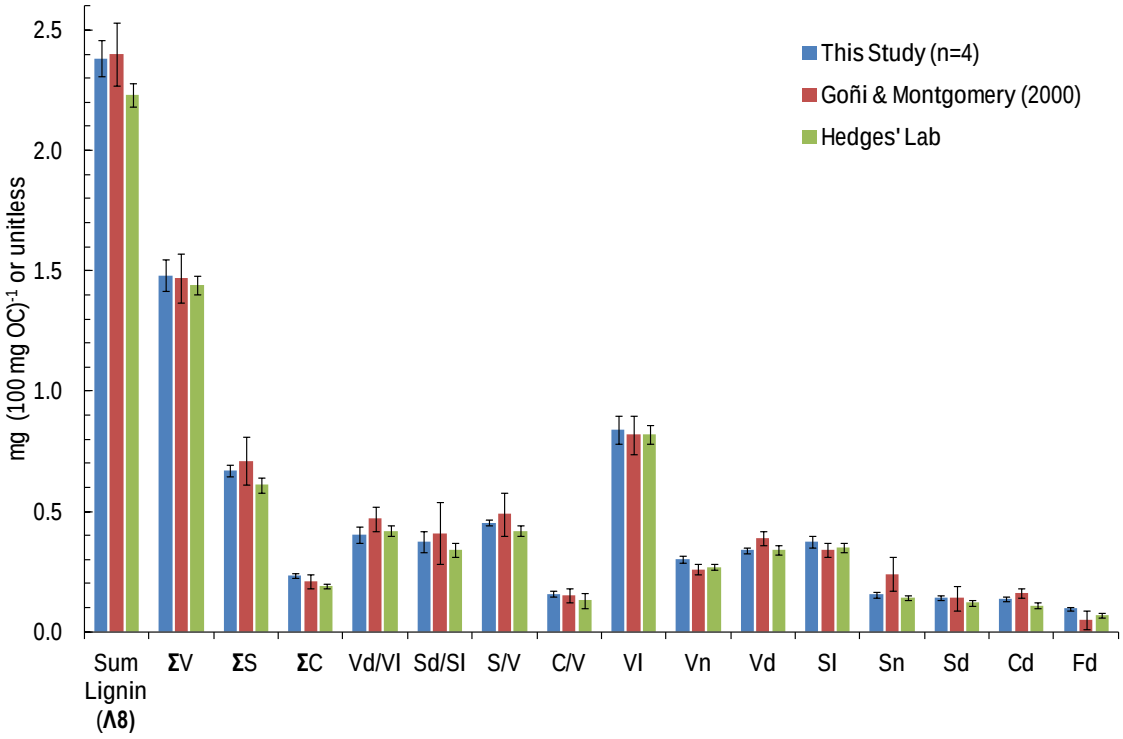


Figure 4.S1: Inter-laboratory comparison of lignin phenol parameters derived from analysis of CuO-oxidized Lake Washington Mud. Values for the two other labs were described previously (Goñi and Montgomery 2000) and the same sample was analyzed here, courtesy of Professor Rick Keil (University of Washington), concurrently with Fallen Leaf Lake (FLL) core samples. The Hedges' data were obtained from bomb-oxidation in steel vessels with diethyl ether work-up while the Goñi data were obtained from microwave-assisted CuO oxidation with work-up in ethyl acetate, as was done here. Concentrations are expressed as mg per 100 mg OC. Ratios are unitless. Error bars are ± 1σ (n=4).

4.7 References

Al-Chokhachy, R., M. Peacock, L. G. Heki, and G. Thiede. 2009. Evaluating the Reintroduction Potential of Lahontan Cutthroat Trout in Fallen Leaf Lake, California. North American Journal of Fisheries Management **29**: 1296–1313.

Benson, L., M. Kashgarian, R. Rye, S. Lund, F. Paillet, J. Smoot, C. Kester, S. Mensing, D. Meko, and S. Lindström. 2002. Holocene multidecadal and multicentennial droughts affecting Northern California and Nevada. Quaternary Science … **21**: 659–682.

Bergamaschi, B. A., E. Tsamakis, R. G. Keil, T. I. Eglinton, D. B. Montluçon, and J. I. Hedges. 1997. The effect of grain size and surface area on organic matter, lignin and carbohydrate concentration, and molecular compositions in Peru Margin sediments. Geochimica et Cosmochimica Acta **61**: 1247–1260.

Bronk Ramsey, C. 2009. Bayesian analysis of radiocarbon dates. Radiocarbon **51**: 337–360.

Brothers, D. S., G. M. Kent, N. W. Driscoll, S. B. Smith, R. Karlin, J. A. Dingler, A. J. Harding, G. G. Seitz, and J. M. Babcock. 2009. New Constraints on Deformation, Slip Rate, and Timing of the Most Recent Earthquake on the West Tahoe-Dollar Point Fault, Lake Tahoe Basin, California. Bulletin of the Seismological Society of America **99**: 499–519.

Goñi, M. A., and J. I. Hedges. 1995. Sources and reactivities of marine-derived organic matter in coastal sediments as determined by alkaline CuO oxidation. Geochimica et Cosmochimica Acta **59**: 2965–2981.

Goñi, M. A., and S. Montgomery. 2000. Alkaline CuO oxidation with a microwave digestion system: lignin analyses of geochemical samples. Analytical Chemistry **72**: 3116–3121.

Goñi, M. A., K. C. Ruttenberg, and T. I. Eglinton. 1997. Sources and contribution of terrigenous organic carbon to surface sediments in the Gulf of Mexico. Nature **389**: 275–278.

Goñi, M., and T. I. Eglinton. 1996. Stable carbon isotopic analyses of lignin-derived CuO oxidation products by isotope ratio monitoring-gas chromatography-mass spectrometry (irm-GC-MS). Organic Geochemistry **24**: 601–615.

Hallett, D. J., L. V. Hills, and J. J. Clague. 1997. New accelerator mass spectrometry radiocarbon ages for the Mazama tephra layer from Kootenay National Park , British Columbia , Canada. Canadian Journal of Earth Sciences **34**: 1202–1209.

Hanes, T. 1981. Hydrologic analysis of the Fallen Leaf Lake watershed and operations plan for Fallen Leaf Lake. United States Forest Service Report, Region 5, Lake Tahoe Basin Management Report.

Hedges, J. I., and J. R. Ertel. 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. Analytical Chemistry **54**: 174–178.

Hedges, J. I., and D. Mann. 1979. The characterization of plant tissues by their lignin oxidation products. Geochimica et Cosmochimica Acta **43**: 1803–1807.

Huang, Y., K. H. Freeman, T. I. Eglinton, and F. A. Street-perrott. 1999. δ13C analyses of individual lignin phenols in Quaternary lake sediments: A novel proxy for deciphering past terrestrial vegetation changes. Geology **27**: 471–474.

Ishiwatari, R., S. Yamamoto, and S. Shinoyama. 2006. Lignin and fatty acid records in Lake Baikal sediments over the last 130kyr: A comparison with pollen records. Organic Geochemistry **37**: 1787–1802.

Kleppe, J. ., D. S. Brothers, G. M. Kent, F. Biondi, S. Jensen, and N. W. Driscoll. 2011. Duration and severity of Medieval drought in the Lake Tahoe Basin. Quaternary Science Reviews **30**: 3269–3279.

Lico, M. 2004. Nutrient Concentrations in Upper and Lower Echo, Fallen Leaf, Spooner, and Marlette Lakes and Associated Outlet Streams, California and Nevada, 2002–03, USGS.

Maloney, J. M., P. J. Noble, N. W. Driscoll, G. M. Kent, S. B. Smith, G. C. Schmauder, J. M. Babcock, R. L. Baskin, R. Karlin, a. M. Kell, G. G. Seitz, S. Zimmerman, and J. A. Kleppe. 2013. Paleoseismic history of the Fallen Leaf segment of the West Tahoe-Dollar Point fault reconstructed from slide deposits in the Lake Tahoe Basin, California-Nevada. Geosphere

Mensing, S. A., L. V Benson, M. Kashgarian, and S. Lund. 2004. A Holocene pollen record of persistent droughts from Pyramid Lake, Nevada, USA. Quaternary Research **62**: 29–38.

Meyers, P. A. 1994. Preservation of elemental and isotopic source identification of sedimentary organic matter. Chemical Geology **114**: 289–302.

Meyers, P. A., and R. Ishiwatari. 1993. Lacustrine organic geochemistry—an overview of indicators of organic matter sources and diagenesis in lake sediments. Organic Geochemistry **20**: 867–900.

Reimer, P. J., M. G. L. Baillie, E. Bard, A. Bayliss, J. W. Beck, C. J. H. Bertrand, P. G. Blackwell, C. E. Buck, G. S. Burr, K. B. Cutler, P. E. Damon, R. L. Edwards, R. G. Fairbanks, M. Friedrich, T. P. Guilderson, A. G. Hogg, K. A. Hughen, B. Kromer, G. McCormac, S. Manning, C. Bronk Ramsey, R. W. Reimer, S. Remmele, J. R. Southon, M. Stuvier, S. Talamo, F. W. Taylor, J. van der Plicht, and C. E Weyhenmeyer. 2004. INTCAL04 terrestrial radiocarbon age calibration, 0-26 cal Kyr BP. Radiocarbon **46**: 1029–1058.

Smith, S. B., R. E. Karlin, G. M. Kent, G. G. Seitz, and N. W. Driscoll. 2013. Holocene subaqueous paleoseismology of Lake Tahoe. Geological Society of America Bulletin **125**: 691–708.

Stine, S. 1994. Extreme and persistent drought in California and Patagonia during mediaeval time. Nature **369**: 546–549.

Tareq, S. M., N. Tanaka, and K. Ohta. 2004. Biomarker signature in tropical wetland: lignin phenol vegetation index (LPVI) and its implications for reconstructing the paleoenvironment. Science of the Total Environment **324**: 91–103.

CHAPTER V

A two-dimensional, heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analyses

5.1 Abstract

Motivated by the need to develop clean, high purity preparative enrichments of individual compounds for micro-scale compound-specific natural abundance isotope and radiocarbon (^{14}C) analyses, we describe a new, two-dimensional, heart-cutting, low-bleed, three-oven, single GC preparative system, demonstrate its resolving capabilities as applied to a typically complex environmental sample matrix, and investigate the robustness with which it preserves the authigenic $^{13}C/^{12}C$ and $^{14}C/^{12}C$ ratios of individual compounds it targets for preparative enrichment. The system is comprised of a programmable temperature vaporizing (PTV) inlet, a single GC oven, two modular, door-mounted, resistively heated low thermal mass (LTM) columns, a preparative fraction collector (PFC), and a Deans pneumatic switching device which facilitates heart-cutting between the system's 1˚ and 2˚ chromatographic dimensions. Further, the system's inlet and trapping parameters are optimized for the efficient preparative enrichment of the methyl ether and ester derivatives of the lignin phenol compound class. The lignin phenols include such compounds as the vanillyl and syringyl aldehydes, ethanones, and acids and are unrivaled biomarkers of terrestrial organic matter, some of which are also important components of fragrances and flavors. Using this suite of compounds, the suitability of this augmented preparative capillary GC (PCGC) system was investigated for micro-scale compound specific (CS) stable isotope and natural abundance radiocarbon analyses (RA). Analysis of a >300 injection enrichment scheme reveals the instrument to fractionate ^{13}C in predictable ways and to preserve the authigenic $\Delta^{14}C$ of compounds it targets for preparative enrichment to within 6.7 ± 5.0‰, demonstrating the

promising new utility of such systems towards micro-scale CSRA investigations for which clean and high resolution separation techniques are prerequisite.

5.2 Introduction

Radiocarbon is naturally formed in the upper troposphere and lower stratosphere before quickly oxidizing to $^{14}CO_2$ – a molecule of which is present for every trillion (10^{12}) molecules of $^{12}CO_2$. Additionally, above-ground nuclear weapons testing in the 1950s and 1960s doubled the atmospheric ^{14}C inventory. Once oxidized into CO_2, ^{14}C acts as an isotopic tracer and radioactive clock which is entrained into all photosynthetically derived biomolecules before decaying with a 5730-year half-life. Radiocarbon's presence throughout the Earth's active atmospheric, terrestrial, and oceanic components of the Global Carbon Cycle makes compound-specific radiocarbon analysis (CSRA) a powerful tool which enables explorations into the modes and rates of carbon exchange between these reservoirs. It also has the potential to be applied to other purposes, such as towards authenticating the source of chemical additives in consumer products as either natural (^{14}C-containing) or petrochemical derived (^{14}C-dead).

The persistent obstacle to CSRA investigations, however, derives from the challenges inherent to cleanly isolating individual components from the complex environmental matrices in which they are contained (Mollenhauer and Rethemeyer 2009). For this reason, preparative enrichments for natural abundance CSRA rely upon efficient, highly resolved, and high purity chemical separations which are explicitly designed to resist contamination by exogenous carbon.

Traditionally, CSRA preparations have been performed using either HPLC (Gillespie et al. 1984) or preparative capillary gas chromatography (PCGC) (Eglinton et al. 1996). More recently, a two-step preparative HPLC technique was applied to CSRA investigations of dicarboxylic acids from aerosols (Fahrni et al. 2010) and lignin phenols isolated from sediments (Hou et al. 2010) and plant matter (Ingalls et al. 2010). The greater number of theoretical separation plates in capillary columns, compared to HPLC, has made PCGC more attractive for natural abundance CSRA analyses because of the clean and highly resolved analytical separations it affords. CSRA investigations relying on PCGC have expanded greatly in the past decade to include the somewhat routine analyses of individual fatty acids, alkanes, and long-chain (C_{37}) alkyl ketones (alkenones) isolated from oceanic (Eglinton et al. 1997; Uchida et al. 2000, 2005; Pearson et al. 2001; Druffel et al. 2010) sediments. One previous exploratory lignin phenol PCGC CSRA study was also reported (McNichol et al. 2000). PCGC-coupled stable isotope analyses have also been applied to discriminate between natural and fossil sources of polycyclic aromatic hydrocarbons (PAHs) in aerosols (Kumata et al. 2006) and in sediments (Reddy et al. 2002), to identify naturally occurring homologues to polybrominated diphenyl ethers (PBDEs) (Teuten et al. 2005), and to investigate Cl-isotope ratios in polychlorinated dibenzo-p-dioxins (PCDDs) to infer biotic versus abiotic formation mechanisms for these compounds in kaolinite clays (Holmstrand et al. 2006a).

In the case of long chain alkyl PCGC separations, urea adduction is often performed as an essential pre-PCGC clean-up step, partially owing to the limited resolving power of traditional PCGCs, which often necessitates additional upstream cleanup steps. This limited resolving power ultimately derives from traditional PCGCs'

single dimensionality, lack of heart-cutting, and their near-exclusive reliance on ‘mega-bore’ (0.53 mm i.d.) rather than narrower-bore columns, principally because of the greater loading capacities of larger-bore columns which necessitate fewer injections, leading to shortened instrument times. Such larger-bore columns are used despite their significantly reduced resolving power relative to narrower-bore capillary columns whose resolving advantages are described by an inverse-square relationship that relates a halving of column width to a corresponding four-fold increase in resolution (Grob and Barry 2004). As such, the existing separation capabilities of traditional 1D PCGC methods clearly merit improvements which will facilitate high purity, single-compound enrichments from complex samples and which may reduce the need for upstream cleanup steps.

Multi-dimensional GC provides well-known resolving advantages over traditional single-column GC (Grob and Barry 2004); two columns of varying stationary phase connected in series can resolve components at greater resolutions than are achievable using a single column alone. However, standard 2D methods are fundamentally limited by the diminishing improvements in resolution that are attained with increases in column length. This is because each doubling of column length, while doubling the number of theoretical separation plates (N), only increases peak resolution by a maximum of $\sqrt{\Delta N}$, or 41% (Grob and Barry 2004). The separation capabilities of such 2D GC methods, however, can be effectively augmented beyond this limitation by insertion of a heart-cutting device between the 1˚ and 2˚ dimensions. Additionally, as was done in this investigation, resolving power is enhanced by a factor of 4.5 by selection of 0.25 mm analytical columns in lieu of the 0.53 mm mega-bore columns traditionally used in PCGC

implementations. While the reduced sample loading capacities of narrower-bore columns necessitate a greater number of repeated injections and longer instrument times, the use of modular low thermal mass (LTM) GC ovens allows for individual GC run times to be shortened.

Such modifications yield a newly implemented, 2D, heart-cutting PCGC approach. In this setup, compounds not targeted for CSRA are discarded while targeted compounds are made to enter, separate, and elute from an additional capillary column interfaced to a preparative fraction collector (PFC). A practical implementation of such a real-time, valve-less switching device was first described by Deans in 1968 (Deans 1968), adapted to capillary columns, and has more recently been made commercially available by Agilent Technologies as a microfluidic “Deans Switch”. Since its inception, the Deans Switch has been implemented in a variety of applications, including analytical process GCs (de Alencastro et al. 2003) and 2D olfactory detection and MSD GCs (Sasamoto and Ochiai 2010; Ochiai and Sasamoto 2011). It has also been used as a duty-cycling modulator for comprehensive GC×GC (Seeley et al. 2007). Its suitability for preparative enrichments upstream of compound-specific isotope and radiocarbon analysis, however, has been neither demonstrated nor evaluated.

Preparative compound enrichments prior to CSRA are essential because accelerator mass spectrometry (AMS) ^{14}C determinations require >30 μg C if the influence of process blank contaminant carbon, usually on the order of 1–4 μg and of inconstant ^{14}C content, is to be adequately constrained (Shah and Pearson 2007; Ziolkowski and Druffel 2009). Because sample quantities are often limited and because targeted components may only be dilutely present in sample matrices, it is necessary to

optimize PCGC instrumental parameters for each targeted compound class to ensure efficient recovery of targeted compounds. In PCGC systems, sample recoveries rarely approach 100%. Analyte losses principally stem from the inlet and the trapping stages of PCGC, but can also arise due to analyte decomposition, irreversible adsorption to system components, and system leaks. The importance of optimizing PCGC collection parameters for different compound classes has been clearly demonstrated for a few model compound types (Meinert and Brack 2010), and specifically for PAHs (Mandalakis and Gustafsson 2003), but not yet for the family of phenolic lignin methyl esters and ethers (Figure 5.1) described here. Lignin phenols are used to track terrestrial organic matter in the environment (Leo and Barghoorn 1970), and many are also investigated because of their noted contributions to complex flavors and fragrances, of which they are often trace-level, but notable constituents (Guillén and Manzanos 2005; Sterckx et al. 2011). Tracking the radiocarbon signature of these components as they occur in flavors, fragrances, and in the environment offers the potential to provide provenance information of trace additives in complex formulations and to inform the timescales with which terrestrial organic carbon exists in its reduced state (sequestered from the atmosphere) before being re-oxidized to CO_2.

5.3 Experimental

5.3.1 Instrumental design

Several commercially available components from Gerstel Inc. and Agilent Technologies were combined to create a two-dimensional, heart-cutting, low-bleed, preparative GC which has been schematically outlined in Figure 5.2. The setup contains

an Agilent 7890A GC, a 7693A Automated Liquid Sampler (ALS), a programmed temperature vaporization (PTV, Gerstel) inlet, two 30 m, 0.25 mm, low thermal mass (LTM, Agilent) capillary columns (1˚: Rtx-5MS, 0.50 μm df, 2˚: DB 17MS, 0.25 μm df) with mated GC door-mounted transfer line assembles, a zero-dead volume Deans pneumatic microfluidic switching device (Agilent), and a Preparative Fraction Collector (PFC, Gerstel). Inside the Agilent GC oven, 0.25 mm deactivated fused silica tubing (Agilent) routes analytes from the PTV inlet through to the 1˚ LTM column and into the Deans Switch, which alternately directs flow to the front FID (FID-1) or into the 2˚ LTM from which it elutes into a four-way splitter. Flow at the four-way splitter is split 1:49 such that ~2% enters the back (FID-2) through a 50 μm i.d. fused silica capillary restrictor while the remaining ~98% continues through 0.32 mm deactivated fused silica tubing which routes analytes through a heated transfer line and into the PFC and its six individual component traps. A 1 m guard column interfaces the PTV with the first column to achieve extended LTM column life. Because the LTM columns are 5" coils tightly intertwined with temperature sensors and resistively heated coils, they are not readily amenable to periodic column trimming and the use of a guard column is therefore important. The attraction of LTM ovens stems from their modular design, their ability to rapidly heat and cool, and their control from within ChemStation – features that collectively shorten run-times and accelerate preparative GC throughput.

The pneumatic resistance at each Deans switching outlet (labeled ON or OFF in Figure 5.2) is equally balanced using deactivated capillary tubing of varying diameters and lengths. Flow direction through the Deans Switch is actuated via a solenoid valve positioned upstream of a shunt restrictor and is controlled within ChemStation as a

programmed runtime event. The PFC contains a heated eight-way splitter, which directs eluting analytes into pre-defined component traps by means of a downstream seven-membered solenoid valve array whose valves are programmatically actuated within ChemStation. The traps are U-shaped borosilicate tubes programmed, in this study, to be held at −20 ˚C (measured to be at −16 ˚C) using a Haake recirculating cooler. Eluting windows of interest are 'cut' from the 1˚ column and onto the 30 m 2˚ column where contaminating interferences present in the 'cut' 1˚elution stream are further resolved from those compounds targeted for enrichment. Targeted compounds eluting from the 2˚ column are then directed into individual component traps where they are frozen and further enriched over repeated injections. In the current study 308 consecutive injections were made. Meanwhile, the contaminating interferences, carryover from the 1◦ cuts, are directed to a waste trap. Following an enrichment scheme of many repeated injections, the traps can be eluted with high-purity solvents and prepped for CSRA.

5.3.2 Environmental sample and isotopic (^{13}C, ^{14}C) process standards

Lake Washington Mud (LWM, Seattle, WA) was obtained from Rick Keil (University of Washington, Seattle, WA) and was used to test the resolving capabilities of this system for a widely analyzed and inter-compared environmental reference material (Hedges and Ertel 1982; Goñi and Montgomery 2000). To investigate the efficacy of this 2D, heart-cutting PCGC for isotopic and natural abundance ^{14}C investigations, we evaluated a set of eight model compounds (Table 5.1) whose $^{13}C/^{12}C$ and $^{14}C/^{12}C$ were previously unknown to us. Benzoic acid, methyl ester (bAme, 99%) and benzoic acid, 3,4 dimethoxy-, methyl ester (VAdme, 98%) were obtained from Alfa

Aesar. Benzaldehyde, 3,4 dimethoxy (Vme, > 99%), ethanone, 1-(3,4-dimethoxyphenyl)- (AcVme, 98%), and benzaldehyde, 3,4,5-trimethoxy (SAlme) were obtained from Acros Organics. 2-Propenoic acid, 3-phenyl-, methyl ester (tCAdme, 98.7%), benzoic acid, 3,4,5-trimethoxy-, methyl ester (SAdme, 98%), and ethanone, 1-(3,4,5-trimethoxyphenyl)- (AcSme, 98%) were obtained from SigmaAldrich. tCAdme was obtained of certifiably natural origin ($\delta^{13}C = -28.96 \pm 0.04‰$, $^{14}C = 14.91 \pm 0.12$ dpm gC^{-1}). These compounds are shown in Figure 5.1 with their relation to their better-known alcoholic, aldehydic, and acidic relatives. All compounds except tCAdme and bAme are methylated ethers and esters of lignin phenolic compounds, which occur naturally in the environment. Multiple process standards were chosen partly owing to the absence of any a priori knowledge of the ^{14}C-content of these process standards, but also because we sought to have a set of process standards with variable $^{14}C/^{12}C$ ratios to enable a more robust evaluation of the $^{14}C/^{12}C$ ratio of any exogenous carbon which might be added throughout a preparative enrichment.

5.3.3 Evaluation with Lake Washington mud (LWM)

LWM was subjected to CuO oxidation following Goñi and Montgomery's (Goñi and Montgomery 2000) adaption of Hedges and Ertel (Hedges and Ertel 1982). Approximately 5 mg organic carbon (100 mg freeze dried and homogenized sediment) was combined with 500 mg CuO, 50 mg $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ and 20 mL 2 *N* NaOH in a MARS (CEM Corp.) PFA vessel for microwave hydrolysis at 150˚C for 90 min under N_2. The CuO hydrolysate was spiked with ethyl vanillin and *trans*-cinnamic acid as dual internal recovery standards before being quantitatively decanted into 40 mL EPA

vials. The hydrolysate was acidified to pH~2 with 12.1 *N* trace-metal grade HCl and then extracted 3× with ethyl acetate (Optima). Ethyl acetate extracts were dried over Na_2SO_4 and blown to dryness under ~10 psi N_2 prior to immediate resuspension in 400 μL pyridine. 50 μL of the sample was then diluted 1:1 with pyridine in 400 μL GC vial inserts and derivatized with 100 μL BSTFA + 1% TMCS (Sigma) (60 ˚C, 10 min) to swap exchangeable protons in the sample mixture with trimethylsilyl (TMS) groups. Injection volume was 5 μL, splitless time was 1 min, vent flow was 100 mL min^{-1}, vent pressure was 1 psi, vent time 0.11 min, and the initial inlet temperature was 100˚C and ramped to 280˚C at 12˚C s^{-1} following a 0.25 min initial time. Compounds were identified based on their retention times (t_R) relative to those of injected standards and were verified on a GC–MS using an identical column and method. PTV inlet pressure was 38.77 psi, Deans Switch EPC3 pressure was 28.48 psi, carrier gas was He and the system was operated in constant pressure mode. Cut times were determined based on the elution windows of compounds of interest (Figure 5.3). GC and (both) LTM ovens were programmed from 100 to 320 ˚C at 4 ˚C min^{-1} with a final hold time of 1 min.

5.3.4 PCGC optimization for lignin ethers and esters

Optimization of PCGC systems is necessary to ensure that targeted compounds are efficiently enriched. Changes in inlet-to-column transfer efficiency were examined by systematically and independently varying PTV inlet parameters while monitoring the FID peak areas resulting from these variations. Changes in PFC trapping efficiencies were also evaluated, but for a more limited set of conditions. All optimizations were performed using a prepared 0.50 mM mixed standard cocktail containing each of the eight

compounds listed in Table 5.1 dissolved in DCM, which was used because of its intermediate polarity and high volatility. Initial LTM oven and solvent venting temperatures were set to 40 ˚C as no notable improvements in peak shape or analyte transfer efficiency were observed at lower temperatures down to ambient (23 ˚C).

5.3.4.1 PTV inlet conditions

Splitless (or purge) times were varied from 0.5 to 3.0 min. FID-1 peak areas were normalized to the area produced using a 60 s splitless time (Figure 5.4A). Injection volume was varied systematically from 1 to 5 μL. The resulting changes in peak areas in FID-1 were normalized to μL injected and to the area produced by a 1 μL injection (Figure 5.4B). Evaporative concentration of analytes in the PTV inlet is accomplished with solvent venting. The goal of optimizing vent pressure and flow is to selectively evaporate solvent without sweeping targeted analytes out of the split vent. Using a vent time of 0.10 min, vent flow and vent pressure were each varied independently of one another. Vent flow was varied from 30 to 90 mL min^{-1} at a constant vent pressure of 1 psi and the resulting changes in FID areas were normalized to that produced at 30 mL min^{-1} (Figure 5.4C). Similarly, vent pressure (the column head pressure) was varied between 1 and 39 psi while vent flow was held constant at 90 mL min^{-1} and the resulting changes in areas were normalized to the area produced at a vent pressure of 1 psi (Figure 5.4D). All injection speeds were at 9600 μL min^{-1}.

5.3.4.2 Trapping conditions

Trapping efficiencies were evaluated using Gerstel Inc's "1 μL" and "100 μL" traps, as well as "100 μL" traps filled with DCM:acetone (1:1, v/v). In these evaluations, the PFC temperature was held at 325 ˚C and the trap temperatures were maintained at −16 ˚C. All other variables were held constant. To quantify preparative yields under these conditions, 10–15 repeated injections were made with each trapping configuration. The sample traps were then eluted 3× with DCM:acetone (1:1, v/v) into pre-weighed 4 mL vials. The vials were immediately capped and weighed at 22 ˚C and then rapidly aliquoted into GC vials, which were quickly capped and immediately analyzed via GC–MS. Quantification of trap contents was achieved using 5-point calibration curves of individual external standards.

5.3.5 Acquisition parameters for CSRA enrichments

The eight compounds in Table 5.1 were dissolved in DCM and combined into a 0.50 mM mixed standard cocktail, which was used as the starting material for preparative enrichment. Optimized inlet parameters were used. 5μL of standard cocktail was injected at 9600 μL min^{-1} into a baffled deactivated glass liner packed with deactivated glass wool (Gerstel). The PTV was programmed from 40 ˚C, with an initial time of 0.20 min, to 275 ˚C at 12 ˚C s^{-1} with a final hold time of 5 min. Vent flow was set to 90 mL min^{-1} for 0.10 min at a vent pressure of 1 psi. Splitless time was 2 min. The GC oven temperature was programmed from 300 ˚C (20 min hold) to 310 ˚C (3 min hold) at 21 ˚C min^{-1}. The LTM ovens were programmed with identical temperature programs: 40 ˚C (2 min hold) to 210 ˚C (1 min hold) at 40 ˚C min^{-1}, to 230 ˚C (1 min hold) at 2 ˚C min^{-1}, to

310 ˚C (1.05 min hold) at 20 ˚C min^{-1}. Inlet pressure was 38.77 psi, Deans Switch EPC3 pressure was 28.48 psi, carrier gas was He and the system was operated in constant pressure mode. Cut and trap times were carefully set based on each compound's 1˚ and 2˚ dimension elution times as detailed in Figure 5.5. Retention times and areas (via FID-2) were monitored continuously (Figure 5.6) to ensure consistency and appropriateness of cut and trap times. Throughout a week-long, 308-injection preparative run, no adjustments to either cut or trap times were needed.

5.3.6 Quantification of trap contents and purities

Traps were eluted 5× with DCM:acetone (1:1, v/v) (GC Resolv, Fisher) into pre-combusted (450 ˚C, 6 h), pre-weighed 4 mL vials. Aliquots (10–20 μL) were removed, diluted 20×, and quantified using an Agilent 7890A/5975C GC-FID/EI-MSD using the FID signal and 5-point calibration curves of external standards of the quantified compounds. PCGC recovery yields for each compound were determined following U-tube elution as follows: 100% × (ng of compound trapped)/(ng compound injected). Trap purities were assessed as the ratio of the FID peak area of the targeted compound to that of the total peak area (omitting solvent area) contained within the chromatogram. The remaining volume of eluted samples was then dried gently under N_2.

5.3.7 ^{13}C and ^{14}C analyses

In preparation for radiocarbon measurements via AMS using the closed tube combustion (CTC) method, samples were resuspended in pentane:dcm (2:1, v/v) and transferred with pre-combusted (450 ˚C, 6 h) Pasteur pipettes to pre-combusted (850 ˚C,

6 h) quartz tubes and were again dried under N_2. CuO and silver wire were added to the quartz tube, which was then evacuated and sealed while being immersed in an isopropanol/CO_2(s) slurry (−78 ˚C). The sealed tube was combusted at 850 ˚C for 6 h. The evolved CO_2 was cryogenically extracted, quantified, reduced to graphite under H_2 at 630 ˚C with Fe catalyst, pressed into a graphite AMS target, and analyzed on either a 500 kV or 3 MV Tandetron AMS system following established methods (Gagnon et al. 2000) at the National Ocean Sciences AMS (NOSAMS) facility. CO_2 splits were also taken for $^{13}C/^{12}C$ analysis using either a VG PRISM or a VG OPTIMA stable isotope ratio mass spectrometer.

5.3.8 Reporting of $\delta^{13}C$, $\Delta^{14}C$ data

$^{13}C/^{12}C$ sample ratios are reported in standard $\delta^{13}C$ 'per mil' (‰) notation relative to the Vienna Pee Dee Belemnite (VPDB) standard as follows (Coplen 1996):

$$\delta^{13}C(‰) = \left(\frac{\left(^{13}C/^{12}C\right)_{sample}}{\left(^{13}C/^{12}C\right)_{VPDB}} - 1 \right) \times 1000 \, . \qquad (1)$$

Measured $^{14}C/^{12}C$ ratios are reported by NOSAMS as fraction modern (Fm) relative to the internationally accepted 'modern' ratio of $(1.176 \pm 0.010) \times 10^{-12}$ which corresponds to 95% of the $^{14}C/^{12}C$ ratio of the 'modern' reference material NBS Oxalic Acid I (in AD 1950) normalized to a $\delta^{13}C_{VPDB}$ of −19‰. Because isotope fractionation may alter the authigenic $^{14}C/^{12}C$ ratios observed in samples, all Fm data are normalized to a $\delta^{13}C$ of −25‰ using the following equation:

$$Fm_{-25‰\ \delta 13C} = Fm\left(\frac{1-25/1000}{1+\delta^{13}C/1000}\right)^2, \quad (2)$$

where the magnitude of ^{14}C fractionation is parameterized to be exactly 2× the magnitude of the sample's observed $\delta^{13}C$ fractionation, relative to a $\delta^{13}C$ of −25 ‰. Radiocarbon results reported here follow the convention of Stuiver and Polach (Stuiver and Polach 1977) in which sample Fm values are 'age corrected' back to 1950 and expressed as a per mil deviation from the NBS international standard (in 1950) following:

$$\Delta^{14}C(‰) = \left(Fm_{-25‰\delta 13C} \times e^{\lambda(1950-CY)} - 1\right) \times 1000, \quad (3)$$

where CY is the collection year of the sample (here, 2010 for all samples) and λ is 1/(^{14}C mean life), which corresponds to 1/(8267 years). The reported uncertainties in $\Delta^{14}C$ values are the larger of either the internal machine error (derived from Poisson counting statistics) or the external error (derived from repeated AMS analysis of each graphite target). $\Delta^{14}C$ values > 0‰ reveal the incorporation of ^{14}C derived from above-ground nuclear weapons testing which peaked immediately prior to the ratification of the Nuclear Test Ban Treaty in 1963.

5.4 Results and discussion

5.4.1 LWM chromatogram

The chromatogram of the LWM hydrolysate in Figure 5.3 demonstrates the augmented resolving capabilities of this high resolution, 2D, heart-cutting PCGC implementation when confronted with a typically complex environmental sample matrix. The LWM 1˚ chromatogram (visualized uncut with FID-1 in Figure 5.3A) exhibits

marked co-elution, especially for labeled compounds AcV, SAl, and AcS. However, when these co-eluting components are cut to the 2◦ column they are observed to undergo further separation during which impurities from the 1˚ column cuts separate from the components targeted for isolation (seen in Figure 5.3C with FID-2). Meanwhile, the residual chromatogram containing uncut components is visualized by FID-1 (Figure 5.3B), whose monitoring over repeated injections can be used to monitor t_R variability and the appropriateness of programmed cut times. As in other implementations of heart-cutting systems, the performance of the 2˚ column benefits greatly because it only needs to accommodate cuts from the 1˚ column, instead of the burden of the entire injected sample (and solvent) handled by the 1˚ column. In this example, LTM columns were programmed with identical temperature programs. As a result, the full resolving capabilities of this particular setup are perhaps understated for this particular sample. A likely functional improvement to this system could derive from the insertion of a cryo-trap into which 1˚ column cuts would be directed prior to the onset of 2˚ column chromatography. Such a modification would likely find utility when trace-level compounds are targeted for CSRA as the cyrotrap facilitates both ‘cryo-concentration’ (via repeated 1˚ cuts) and ‘cryo-focusing’, which would narrow peak widths, further enhancing the instrument’s effective resolving power.

5.4.2 Optimization of inlet and trapping conditions

Improvements in PCGC recoveries resulted primarily from optimization of PTV inlet conditions and secondarily from optimization of PFC trapping conditions. Other

unobserved modes of sample loss include system leaks, analyte decomposition or irreversible adsorption to internal instrument surfaces.

5.4.2.1 Splitless times

Reasonably, longer splitless times were observed to significantly increase inlet-to-column transfer efficiency for all compounds (Figure 5.4A). Small increases in inlet-to-column transfer efficiency were noted for splitless times beyond 2 min for all compounds. However, the deterioration of peak shapes beyond times of 2 min made longer splitless times undesirable. The effect of splitless time on inlet-to-column transfer efficiency derives from the time it takes the analytes to enter the column following sample volatilization after rapid inlet heating (12 °C s^{-1}). During this time, volatilized analytes are continuously being diluted in the inlet by incoming carrier gas, which explains why peak shapes were observed to deteriorate beyond 2 min and why longer splitless times achieve maximal transfer of analytes onto the column.

5.4.2.2 Injection volume

Variations in injection volume (Figure 5.4B) indicated that improved inlet-to-column transfer efficiencies could be obtained with larger injection volumes, even after having normalized to the μL of material injected. This indicated that injection volume does not scale proportionally to the amount of material that ultimately enters the column and emphasizes the importance of optimizing this parameter in addition to other adjustable PTV settings if maximal column transfer efficiencies are to be achieved.

5.4.2.3 Vent flow and vent pressure

The objective of optimizing vent flow and vent pressure conditions is to selectively evaporate solvent while achieving evapo-concentration of targeted analytes inside the glass liner without inadvertently blowing analytes out of the solvent vent. Venting flow, pressure, temperature, and time can all be independently adjusted to tune a PTV system for a particular compound class. Consistent with other PTV inlet assessments (Mandalakis and Gustafsson 2003; Meinert and Brack 2010), shorter vent times (0.10 min) were found to provide optimal inlet-to-column transfer efficiencies. Longer venting times adversely affected column transfer efficiencies. Appropriate venting temperatures are informed by the boiling point (b.p.) of the solvent relative to the b.p. of the analytes as well as by the minimum attainable temperature of a particular PTV system (via either Peltier or LN_2 cooling). For the purposes of this work, the inlet temperature was set at 40 ˚C owing to both the proximity of this temperature to the b.p of the solvent (DCM) and to the speed with which the inlet could be cooled to its initial set-point temperature using the less cumbersome, cryogen-less Peltier PTV cooler. Using these settings, the effect of vent flow and pressure on inlet-to-column transfer efficiency was assessed. Together, the optimization of these two parameters exerted greatest control towards the two-fold goal of selectively evaporating solvent while keeping and concentrating analytes inside the inlet. Independent optimizations of both vent flow and pressure indicated that maximal balancing between selective solvent evaporation and evapo-concentration of analytes in the inlet were obtained at a vent pressure of 1 psi and a vent flow of 90 mL min^{-1}. Vent pressures of ~1 psi correspond to ‘stop-flow’ conditions because 1˚ column flow is effectively stopped when the column head pressure

is reduced to near-ambient pressures. Finally, it was observed that a compound with physicochemical attributes different from the lignin-related set analyzed here, bAme, behaved differently with respect to certain PTV parameters, including injection volume (Figure 5.4B) and vent pressure (Figure 5.4D), emphasizing the importance of performing compound-class specific PTV optimizations to ensure effective and efficient implementations of PCGC.

5.4.2.4 Trapping conditions

In a study examining PAHs, trapping efficiencies were found to be insensitive either to PFC switch temperature (280 to 340 ˚C) or to PFC trap temperature (−16 to 30 ˚C) (Mandalakis and Gustafsson 2003). It has been reported elsewhere (Meinert and Brack 2010), however, that the efficiency with which compounds are isolated using PCGC can be significantly affected by the PFC parameters selected, particularly (1) the temperature of the PFC switch, (2) the temperature of the traps, and (3) the choice of solvent (or lack thereof) within the component traps. The volatilities of the compounds in the set analyzed here were most similar to two compounds examined in the first study (Meinert and Brack 2010) whose recoveries were greatest at sub-ambient trapping temperatures while being largely unaffected by PFC switch temperatures ranging from 280 to 320 ˚C. Therefore, we chose a sub-ambient trapping temperature of −16 ˚C and a PFC switch temperature of 325 ˚C because it was 5 ˚C above the maximum programmed method temperature and minimized the undesirable possibility of analyte condensation within the PFC switch. With these settings in place, our recoveries ranged from 60 to 72%.

Because the choice of solvent-filled versus empty traps has been previously shown to significantly affect observed PFC recoveries (Meinert and Brack 2010) and because we sought to bring our recoveries closer to 100%, we explored whether recoveries for the lignin-related compound set examined here could be improved by use of solvent trapping. We investigated this for traps at −16 ˚C containing DCM:acetone (1:1, v/v), but no significant increase in PFC recoveries were observed compared to dry traps also maintained at −16 ˚C. Lastly, Gerstel Inc. offers two PFC trap types (small "1 μL" and large "100 μL") which are shaped and accommodate flow differently. When tested, trapping efficiencies were indistinguishable using either trap type, however the small trap is ultimately more desirable as it requires less solvent to elute, reducing both dry-down times and solvent-derived exogenous carbon contamination.

5.4.3 t_R Stability and FID reproducibility during a 308-injection enrichment

Because cut and trap times are programmatically defined, small variations in t_R throughout a several-hundred injection run could adversely affect the purities and yields with which targeted compounds are isolated. Figure 5.6B shows that despite inclusion of a Deans Switch, an additional 30 m column, and eight additional capillary column connections, this PCGC implementation maintains low tR variability (max 1σ = 0.016 min), consistent with traditional PCGC setups (Eglinton et al. 1996).

The observed variability in FID areas (Figure 5.6A) over the seven day, 308-injection scheme ranged from 1.1 to 1.2 % ($\pm 1\sigma$). This observed injection-to-injection variability likely derives from the turbulent nature of PTV solvent venting. Such variability is undesirable if precise quantifications are to be achieved, but it poses less of

a problem for CSRA enrichments where sample enrichment is the principal objective. However, this variability could underlie observed variability in compound enrichment efficiencies over successive preparative runs. The gradual upward trends in FID areas in Figure 5.6A are an understandable consequence of slow solvent evaporation through a repeatedly pierced sample vial septa which was replaced periodically as additional sample aliquots were added.

5.4.4 Instrumental isotope fractionation

Instrumental methods seeking to discern natural isotope abundances require rigorous investigations into the degree to which they may isotopically fractionate targeted analytes. In PCGC systems, both inverse (Zencak et al. 2007) and normal isotope (Holmstrand et al. 2006b) fractionation effects have been observed for different compounds analyzed in different systems and under different chromatographic conditions. The magnitude of the fractionation has been shown to vary by as much as 5‰ for $\delta^{13}C$ (Eglinton et al. 1997), even when care is taken to ensure that whole peaks are collected. Therefore, the degree to which a particular PCGC method fractionates stable isotopes needs to be investigated if its intended enrichments are destined for stable isotope analyses. Additionally, contributions of exogenous “live” or “dead” carbon from a particular PCGC system need to be properly constrained and accounted for to ensure that exogenous carbon does not confound the already-dilute $^{14}C/^{12}C$ (10^{-12}) signal present within an isolated compound.

In all but two instances, trapped compounds were ^{13}C-enriched by an average of 0.85 ± 0.48‰ (1σ) relative to their initial $\delta^{13}C$ value (Table 5.1). This result could

reasonably arise from two sources: (1) excision of ^{13}C-depleted peak tails and (2) preferential evaporative loss of light isotopologues during the drying of samples prior to CTC. Evaporative inlet concentration might be another means of selective evaporation of light isotopologues; however, this effect is likely muted because solvent is only incompletely removed from analytes at this stage.

The first source of isotope fractionation results from the so-called 'inverse isotope effect' (Bruner et al. 1966; Matucha et al. 1991) which predicts ^{13}C-depletion in the tails, and ^{13}C-enrichment in the leads, of capillary column chromatographic peaks. That is, peak tails contain the lighter isotopologues while peak leads contain the heavier isotopologues. Therefore, if peak tails are not wholly cut to the 2˚ column, or if they are not wholly trapped by the PFC, the contents of the trap will be ^{13}C-enriched. The inverse isotope effect is thought to arise as a consequence of the greater bond strengths present within the heavier isotopologues of a particular compound. The stronger bonds effectively reduce molar volume and molecular surface area, weakening the molecule's interaction with the column's stationary phase. This effect is then amplified as a result of the thousands of repeated gas–liquid thermodynamic equilibrations which occur during capillary column chromatography. This effect is likely accentuated in this particular setup because (1) the narrower-bore columns used here contain a greater *N* value relative to traditionally used mega-bore columns and (2) peaks are effectively cut twice in this system (once during heart-cutting and again during trapping). The inverse isotope effect was not observed for SAlme, whose $\delta^{13}C_{trap} - \delta^{13}C_{initial}$ was -0.13 ± 0.14‰ (1σ). This likely resulted because the peak tail was better sampled for T5 owing to a more generously defined trap window (Figure 5.5). The waste trap (T0) also exhibited slight

^{13}C-depletion (−0.25 ± 0.14‰) relative to the ^{13}C content of the initial AcSme, which was the primary component (89%) collected within this trap. This likely resulted from the collection of ^{13}C-depleted tails of other, incompletely trapped peaks within this trap and whose presence and identity were confirmed when determining the waste trap's purity, which was expectedly low (89%). These observations show that chromatographic conditions and trap times exert observable and significant influence on the observed δ^{13}C values of trapped compounds and that the 'inverse isotope effect' appears to predominate in the 2D heart-cutting PCGC system described here. Because ^{14}C fractionation is parameterized to be exactly 2× the magnitude of the ^{13}C fractionation, and because δ^{13}C is always measured in conjunction with ^{14}C determinations, the ^{13}C fractionation observed here should not affect the authigenic Δ^{14}C value of compounds trapped by this system because all Δ^{14}C values are normalized to a δ^{13}C of −25‰ (VPDB).

5.4.5 Exogenous carbon contamination

$\Delta^{14}C_{measured} - \Delta^{14}C_{initial}$ is negative (Table 5.1) for those compounds with moderately enriched Δ^{14}C signatures (>−340‰). This suggests that some amount of exogenous ^{14}C-depleted (<−340‰) carbon is being introduced prior to AMS analysis. Meanwhile, $\Delta^{14}C_{measured} - \Delta^{14}C_{initial}$ is positive (beyond error margins) for three out of four compounds where $\Delta^{14}C_{initial}$ values indicated a lack of measurable ^{14}C (Vme, AcVme, and SAlme), suggesting that blank carbon being introduced is unlikely to be radiocarbon dead and that it has some measurable ^{14}C (i.e., Δ^{14}C content > −993‰). As described in Section 5.4.4, isotope fractionation occurring during PCGC or subsequent sample processing is not the source of this difference. To better understand the origins of

this blank contaminant carbon contribution, GC-FID quantification of trapped analytes (as μg C) was compared to the manometric CO_2 yield of each trap's contents following CTC and cryogenic CO_2 extraction. In principle, trapped components quantified using each method should agree and plot near the 1:1 line in Figure 5.7. In theory, deviations above the 1:1 line are indicative of carbon contamination during PCGC, U-tube elution, sample handling, solvent dry-down and CTC. On the other hand, deviations below the 1:1 line are indicative of sample volatilization losses during sample dry-down stages. Indeed, this was observed with tCAdme, the most volatile compound analyzed, however, contaminant carbon was still likely added, but in an amount less than that lost as a result of tCAdme's volatility. In practice, inaccuracies associated with both the manometric and GC-FID C quantification methods may undermine the quantitative meaningfulness of the y-axis of Figure 5.7. However, examined together and on a qualitative basis, six of the seven compounds lie above the 1:1 line, providing a secondary indicator beyond $\Delta^{14}C_{measured} - \Delta^{14}C_{initial}$ values that some small measure of carbon contamination is likely present in each of the samples, including tCAdme, which lies below the 1:1 line. Ultimately, if the amount and the $\Delta^{14}C$ content of this exogenous, contaminant carbon are known, the authigenic ^{14}C signal of the compound isolated for CSRA can be extracted.

For this reason, column bleed was assessed as a potential source of ^{14}C-depleted contaminant carbon by heating the 1˚ and 2˚ columns to 310˚C and trapping the column bleed eluting at this temperature continuously for 60 min – equivalent to 120 × 0.5 min trapping intervals. The column bleed sample was obtained without repeated solvent injections because solvent is not actively cut to the 2˚ column during instrument operation and therefore its injection should not affect the 2˚ column's bleed profile. The

trapped column bleed was processed in a manner identical to the sample enrichments, but yielded insufficient carbon (1.73 ± 0.17 μg C) for ^{14}C analysis. The elution of a clean U-tube, following the identical procedure, yielded a statistically indistinguishable amount of carbon (1.63 ± 0.16 μg C) while blank CTC combustions yielded 1.56 ± 0.20 μg C (1σ, n = 3). Together, these data indicate that (1) column bleed was not a major source of contaminant carbon to preparative enrichments and that (2) the observed background contaminant carbon observed here derives from the unavoidable combined processes of U-tube elution, sample handling, sample dry down, and the combustion blank. PCGC blanks deriving from a combination of these processes have been found to have dead to intermediately dead $\Delta^{14}C$ values generally ranging from −700 to −1000‰ (Ziolkowski and Druffel 2009; Santos et al. 2010). In another PCGC system using a DB-XLB capillary column (30 m, 0.53 mm i.d., 1.5 μm df) (Ziolkowski and Druffel 2009), column bleed was found to contribute up to 0.6 ± 0.3 μg C min^{-1} of trapping per 50 sample injections, an amount which dropped to 0.1 μg C min^{-1} of trapping per 50 injections when no injection was made and the column saw neither sample nor solvent. In this heart-cutting PCGC the 30 m 2˚ preparative column coupled to the PFC only comes into contact with the limited amount of sample material heart-cut from the 30 m 1˚ column (no solvent is cut), significantly reducing the potential contribution of column bleed C into sample traps. Additionally, compared to the thick films of mega bore columns used in traditional PCGC, the choice of thin-film (0.25 μm) and low bleed MS stationary phases in this system further reduces the potential for column bleed.

5.4.6 Carbon mass balance

If the blank contaminant carbon mass (C_{blank}) and ^{14}C content ($\Delta^{14}C_{blank}$) are known, their contribution to the NOSAMS-measured carbon mass ($C_{measured}$) and ^{14}C content ($\Delta^{14}C_{measured}$) can be used to calculate the unadulterated mass (C_{sample}) and ^{14}C content ($\Delta^{14}C_{sample}$) of the sample using a simple mass balance relationship:

$$\Delta^{14}C_{measured} \cdot C_{measured} = \Delta^{14}C_{sample} \cdot C_{sample} + \Delta^{14}C_{blank} \cdot C_{blank} \; . \qquad (4)$$

where $C_{sample} = C_{measured} - C_{blank}$. If the measured C_{blank} of 1.73 μg C is used in this calculation, then the $\Delta^{14}C_{blank}$ (minimum = −1000‰) can be determined for each trap (T1–T6) using the reasonable assumption that $\Delta^{14}C_{measured}$ should equal $\Delta^{14}C_{initial}$ in the absence of any contaminating blank carbon. 1.73 μg C is chosen because this C_{blank} measurement comes from the column bleed experiment and therefore integrates all contamination introduced from post-PCGC sample handling procedures, including C_{blank} introduced from PCGC column bleed, U-tube elution, sample handling, dry-down, and the combustion blank. When this calculation is performed, the average $\Delta^{14}C$ of the blank carbon contaminating each trap is determined to be −766.7‰, which falls within the −700 to −1000‰ range of PCGC blanks reported elsewhere (Ziolkowski and Druffel 2009; Santos et al. 2010).

With values for C_{blank} and $\Delta^{14}C_{blank}$ at hand, $\Delta^{14}C_{sample}$ can be computed from $\Delta^{14}C_{measured}$. When this calculation is performed, the difference ($\Delta\Delta^{14}C$) between the blank-corrected $\Delta^{14}C_{measured}$ (which is $\Delta^{14}C_{sample}$) and $\Delta^{14}C_{initial}$ drops to ≤6.7 ± 5.0‰ for all compounds (Table 5.1) collected in T1 to T6. Understandably, because the blank's

$\Delta^{14}C$ is significantly $\Delta^{14}C$-depleted, the blank C correction exerts greatest influence on the smallest (μg) and most ^{14}C-enriched compounds. For these reasons, the largest difference between $\Delta^{14}C_{sample}$ and $\Delta^{14}C_{measured}$ is observed for tCAdme.

5.5 Conclusions

Adaption of 2D heart-cutting techniques to PCGC shows promise towards resolving complex environmental mixtures which normally exhibit recalcitrance towards the baseline-resolved, high-purity separations which are essential for preparative enrichments preceding CSRA. The system is shown to contribute low amounts of blank carbon to preparative enrichments, preserving the authigenic $\Delta^{14}C$ of sample enrichments to $\leq 6.7 \pm 5.0$‰ between initial and trapped compounds, which is near the limits of AMS precision for small samples. In this implementation of heart-cutting PCGC, an inverse isotope effect is observed, and it is perhaps more pronounced here than in traditional PCGCs because of the narrower columns used (greater *N*) and because peak tails may be incompletely cut twice in this system, once during heart-cutting and again during trapping. For the lignin-related set of compounds analyzed here, no difference was found between the trapping efficiency using solvent-filled traps, the small ("1 μL"), or the large ("100 μL") U-traps used with these PCGC systems. Meanwhile, variability in inlet conditions was found to affect collection efficiency most greatly, and a particular set of conditions providing ideal inlet-to-column transfer efficiency was found, reported, and used for the evaluation of this system for its suitability towards CSRA. A further enhancement to the system would result from insertion of a cryo-trap interfacing the system's 1˚ and 2˚ chromatographic dimensions. The absence of detectable column bleed

in either blanks or samples collected during >300 repeated injections (over seven days) suggests that heart-cutting PCGC is a lower bleed alternative to traditional PCGC, making it a technique ideally suited for small, ultra micro-scale sample enrichments destined for CSRA.

5.6 Acknowledgements

We would like to thank Jeff Lippert and Jack Stuff of Gerstel Technologies for helpful discussions on the effective implementation of Gerstel components in the development of this instrument. This work was aided by the helpful staff at NOSAMS, including Mary Lardie, Al Gagnon, and Mark Roberts, all of whom provided helpful guidance with the generation and interpretation of AMS results. This work was supported by a NOSAMS student internship, the Department of Defense (DoD), National Defense Science and Engineering Graduate (NDSEG) Fellowship (32 CFR 168a), as well as grants from ACS-PRF (45608-AC2) and NSF (MRI OCE-072322, OCE 05–48275) and benefited from the comments of three anonymous reviewers. This chapter is a reprint of Ball, G.I., Xu, Li., NcNichol, A.P., Aluwihare, L.I. A multi-dimensional heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analysis. Journal of Chromatography A., 1220 (2012) 122-131. The dissertation author was the primary investigator and author for this paper.

Table 5.1: $\delta^{13}C$ and $\Delta^{14}C$ values of initial and trapped components investigated in this study in addition to trap purities and yields following the 308-injection preparative enrichment.

Samples			Initial			Measured						Differences (‰)		
short name (trap #)	Cat#, Lot #	% C	NOSAMS ID #	$\delta^{13}C$ (‰)[a]	$\Delta^{14}C$ (‰)[b,c]	NOSAMS ID #	$\delta^{13}C$ (‰)[a]	µg C as CO_2	$\Delta^{14}C$ (‰)[c,d]	Yield (%)	Purity (%)	$\Delta\delta^{13}C$[e] Measured - Initial	$\Delta\Delta^{14}C$[f] Measured - Initial	$\Delta\Delta^{14}C$[g] Sample - Initial
tCAdme (T1)	12096TH, W269816	74.06	OS-78665	-29.06	79.9 ± 3.6	OS-78867	-28.63	38.6	40.9 ± 3.4	68	100	0.43	- 39.0 ±5.0	- 1.1 ± 5.0
Vme (T2)	A0233058, 162610050	65.05	OS-78652	-33.07	-997.5 ± 0.2	OS-78851	-32.33	70.9	-988.9 ± 5.0	60	100	0.74	+ 8.6 ±5.0	+ 3.1 ± 5.0
AcVme (T3)	A01984670, 1688500050	66.65	OS-78653	-33.14	-998.5 ± 0.2	OS-78847	-32.31	72.9	-993.1 ± 5.0	68	100	0.83	+ 5.4 ±5.0	- 0.1 ± 5.0
VAdme (T4)	10107436, B23568	61.22	OS-78654	-33.55	-993.3 ± 0.3	OS-78848	-32.96	72.3	-996.9 ± 5.0	72	100	0.59	- 3.6 ±5.0	- 6.7 ± 5.0
SAlme (T5)	A0174340, 139910250	61.22	OS-78655	-36.10	-998.2 ± 0.2	OS-78833	-36.23	93.1	-990.9 ± 5.0	63	98	-0.13	+ 7.3 ±5.0	+ 3.1 ± 5.0
SAdme (T6)	S29247-308, M8,600-6	58.40	OS-78657	-46.53	-307.8 ± 2.6	OS-78841	-44.87	82.3	-318.1 ± 3.1	67	96	1.66	- 10.3 ±4.0	- 0.7 ± 4.0
AcSme (T0)	S29062-149, T6,810-1	62.85	OS-78656	-35.75	-268.6 ± 2.9	OS-78839	-36.00	86.2	-340.4 ± 2.5	69	89[h]	-0.25	- 71.8[h] ±3.8	- 63.1[h] ± 3.8

[a] Measured at NOSAMS to 1σ uncertainty of 0.1 ‰.
[b] Analyzed on the 3 Megavolt Tandetron AMS for samples > 100 µg C. ^{14}C sample splits ranged from 0.43 to 1.0 mg C.
[c] Reported error is the larger of either the internal (Poisson counting statistics) or external AMS error measured
[d] Analyzed on the NOSAMS 500 kilovolt compact AMS system. ^{14}C sample splits ranged in size from 34-90 µg C.
[e] $\Delta\delta^{13}C = \delta^{13}C_{measured} - \delta^{13}C_{initial}$. Uncertainty (1σ) = ± 0.14 ‰.
[f] $\Delta\Delta^{14}C = \Delta^{14}C_{measured} - \Delta^{14}C_{initial}$. No corrections for blank carbon made.
[g] $\Delta\Delta^{14}C = \Delta^{14}C_{sample} - \Delta^{14}C_{initial}$. $\Delta^{14}C_{sample}$ determined from $C_{measured}$ and $\Delta^{14}C_{measured}$ using measured C_{blank} of 1.73 µg and a $\Delta^{14}C_{blank}$ of -766.6 ‰ as described in section 5.4.6.
[h] This large discrepancy is expected because all cut and untrapped 2° column eluents enter this trap, four of which were 'radiodead' and identified in T0 when it was assayed for purity via GC-FID.

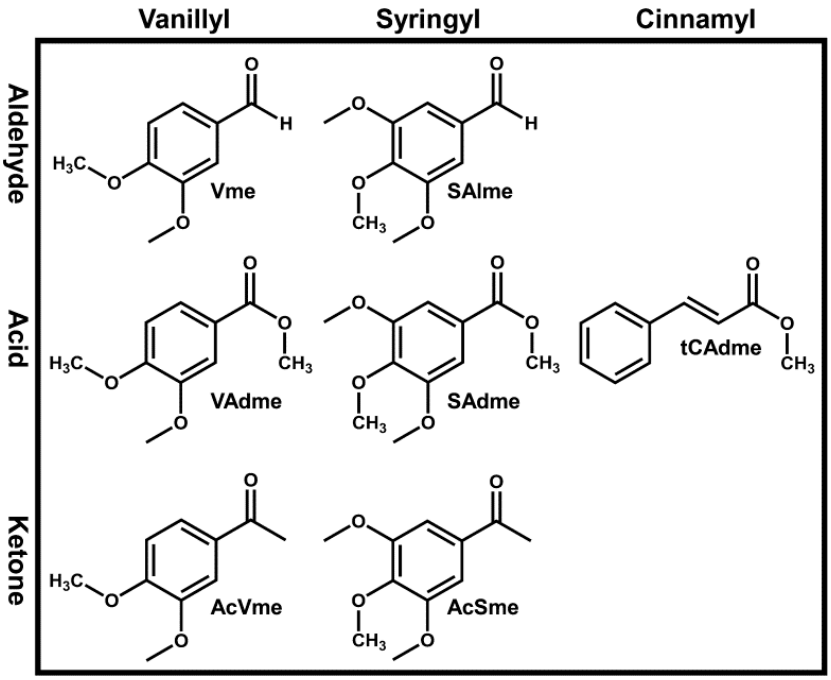


Figure 5.1: Structures of the six compounds used to optimize PCGC parameters and evaluate carbon contamination and isotope fractionation imparted by this PCGC system. All compounds except trans-cinnamic acid, methyl ester (tCAdme), are methyl esters and ethers of the lignin vanillyl and syringyl phenolic family: vanillin (V), vanillic acid (VAd), acetovanillone (AcV), syringaldehyde (SAl), syringic acid (SAd), and acetosyringone (AcS). The –CH3 notation and –me suffix for each compound abrreviation show where hydrogens replace –CH3 groups in the corresponding lignin phenols.

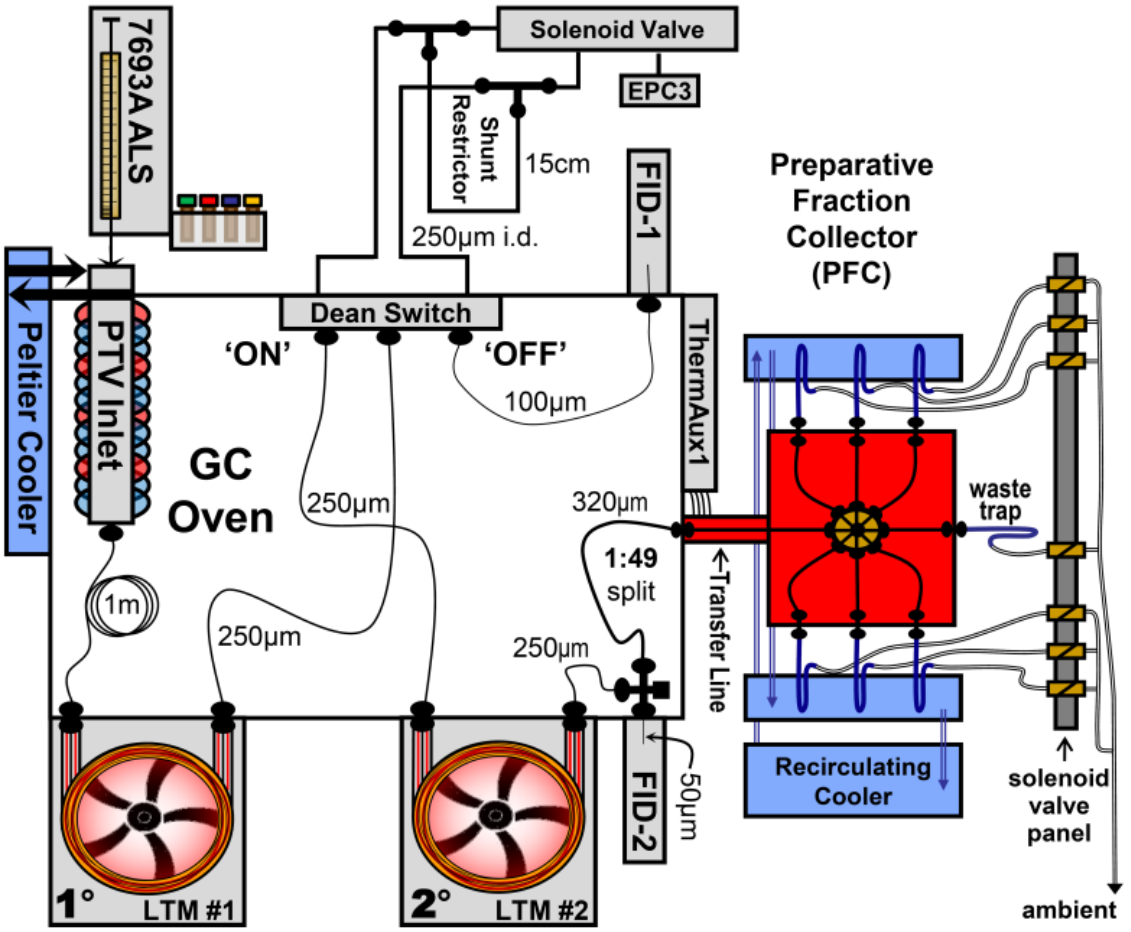


Figure 5.2: Schematic of the 2D, heart-cutting PCGC detailed in this paper and whose components are described in Section 5.2.1. The components shown here are not drawn to scale and their spatial orientation is not representative of the actual configuration. The schematic is instead constructed for ease of visualization.

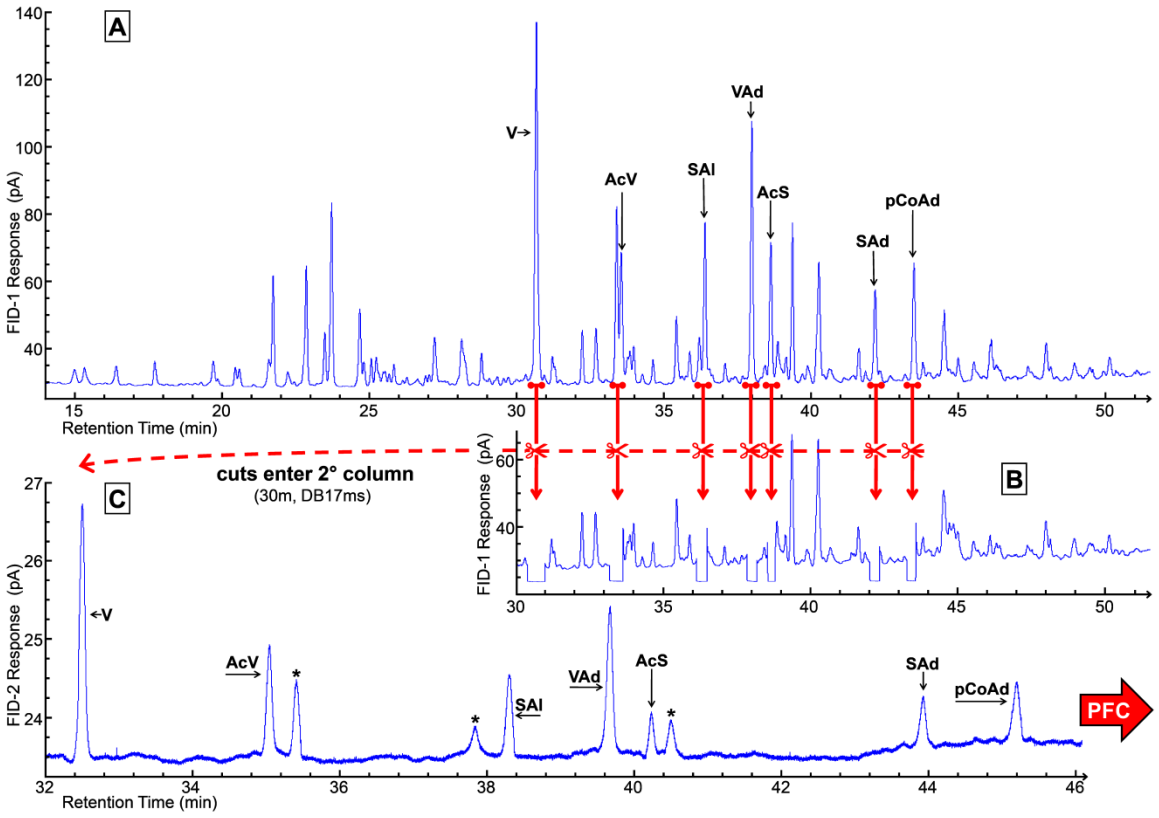


Figure 5.3: A complex environmental chromatogram of the alkaline CuO hydrolysate of Lake Washington mud derivatized with trimethylsilyl (TMS) groups. (A) The complex, uncut, original chromatogram visualized by FID-1. (B) Indicated compounds (arrows) are cut to the 2° column in real-time as their excision from the 1° elution stream is observed by FID-1. (C) The cut components enter the 2° column and are further resolved from contaminating interferences (*) before eluting into the PFC as ~2% of the 2° elution stream is diverted to FID-2 for monitoring. Cut intervals were 30.400–31.000, 33.200–33.650, 36.150–36.500, 37.875–38.200, 38.570–38.815, 42.050–42.370, and 43.300–43.600 min. Compound abbreviations follow Figure 1. pCoAd, p-coumaric acid.

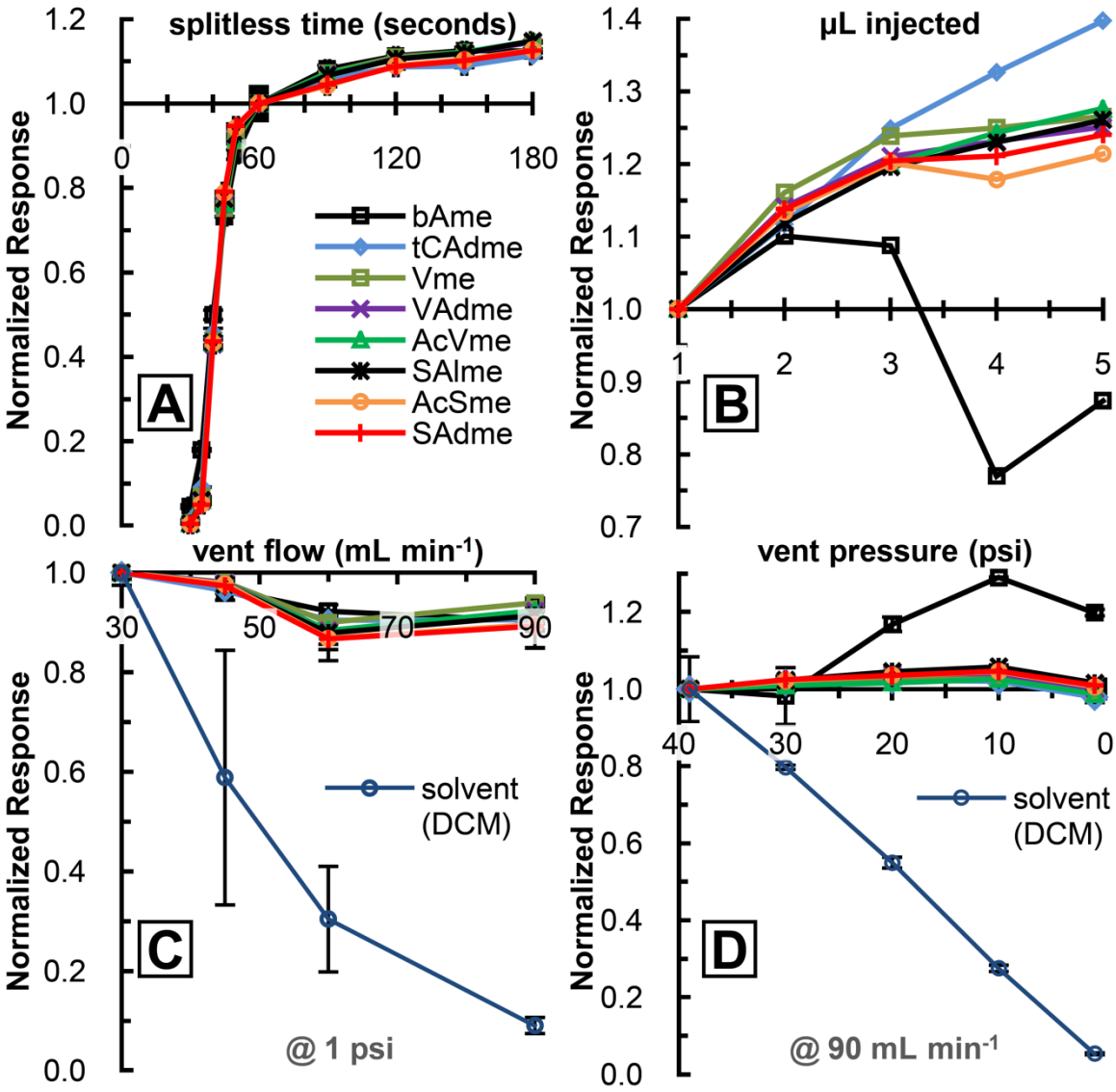


Figure 5.4: Optimization of PCGC inlet conditions to promote the efficient transfer of components into the capillary column. Methyl esters and ethers of lignin phenols are identified in Table 5.1. Analyte-to-column transfer efficiency is monitored by recording FID-1 peak areas as inlet conditions are independently varied. Response areas for four optimized parameters are shown normalized to 60 s for splitless time (A), to 0.5 nmol injected for injection volume (B), to 30 mL min^{-1} for vent flow at 1 psi (C), and to 1 psi for vent pressure at 90 mL min^{-1} (D). Benzoic acid, methyl ester (bAme) was included in the standard cocktail to monitor the response of a compound with physicochemical properties different to those for which the inlet was optimized so that the effect of varying inlet parameter could be monitored more informatively. Error bars are ±1σ.

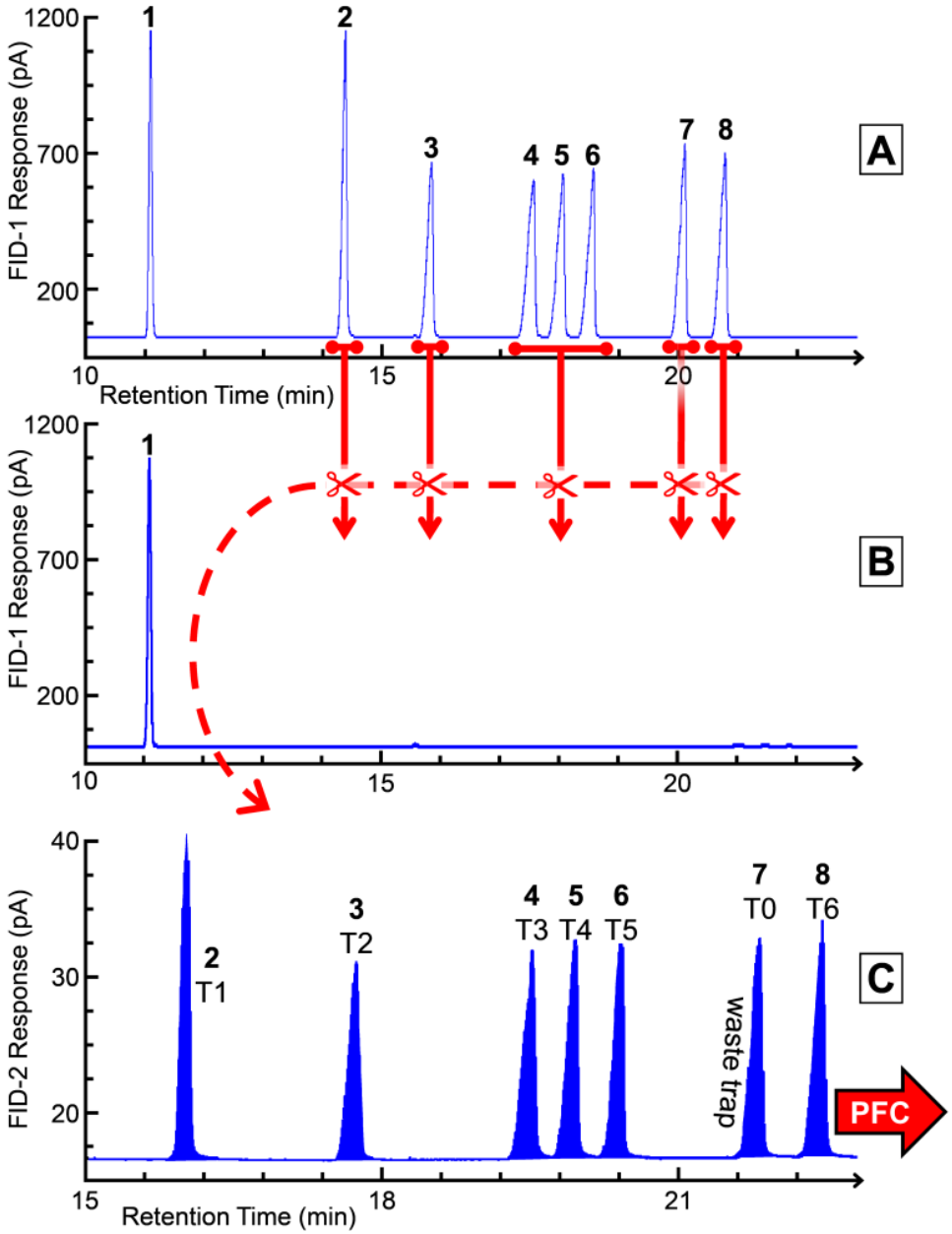


Figure 5.5: Chromatograms of the preparative enrichment scheme used for evaluating the suitability of this system for CSRA enrichments. (A) FID-1 visualizes components eluting uncut from the 1° column and (B) when cuts are made. The cuts then enter and elute from the 2° column where ~2% of the elution stream is visualized by FID-2, (C) while the remaining ~98% continues to the PFC where individual compounds are directed into separate component traps prior to CSRA. Cut intervals (arrows) were 14.160–14.900, 15.600–16.100, 17.220–19.000, and 19.770–20.900 min while trapping intervals (shaded peaks) were from 15.82 to 16.20 (tCAdme, T1), 17.45 to 18.05 (Vme, T2), 19.15 to 19.67 (AcVme, T3), 19.67 to 20.12 (VAdme, T4), 20.12 to 20.70 (SAlme, T5) and 22.13 to 22.65 min (T6, SAdme). Numbers atop peaks show corresponding peaks in the three chromatograms. Peak 1 is bAme.

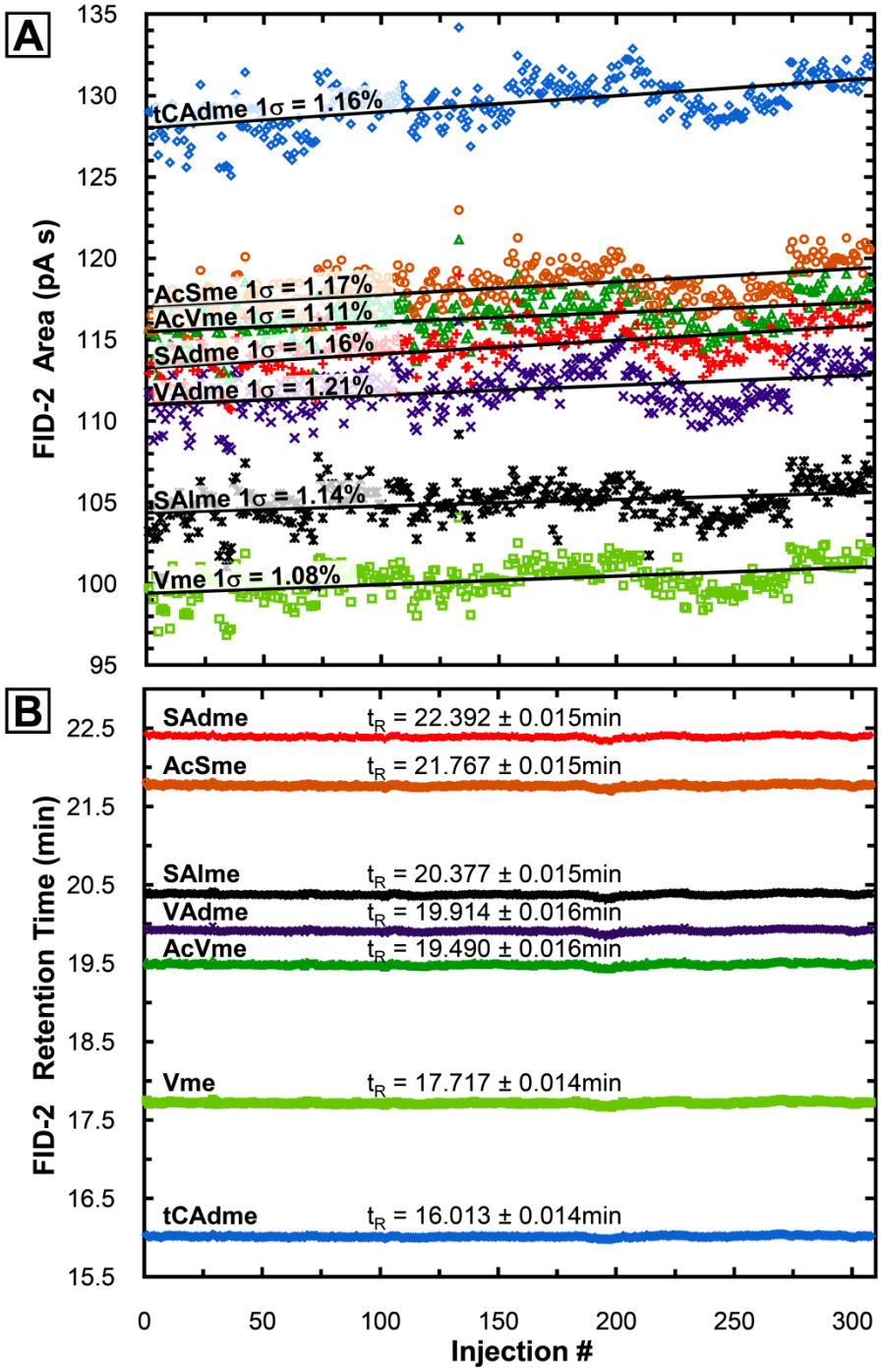


Figure 5.6: Stability of the 2D heart-cutting PCGC during the 308-injection preparative enrichment scheme. (A) FID-2 responses for each compound are indicated with their 1σ variability. (B) Stability of retention times (t_R) ± 1σ of cut components eluting from the 2° column into FID-2 and the PFC.

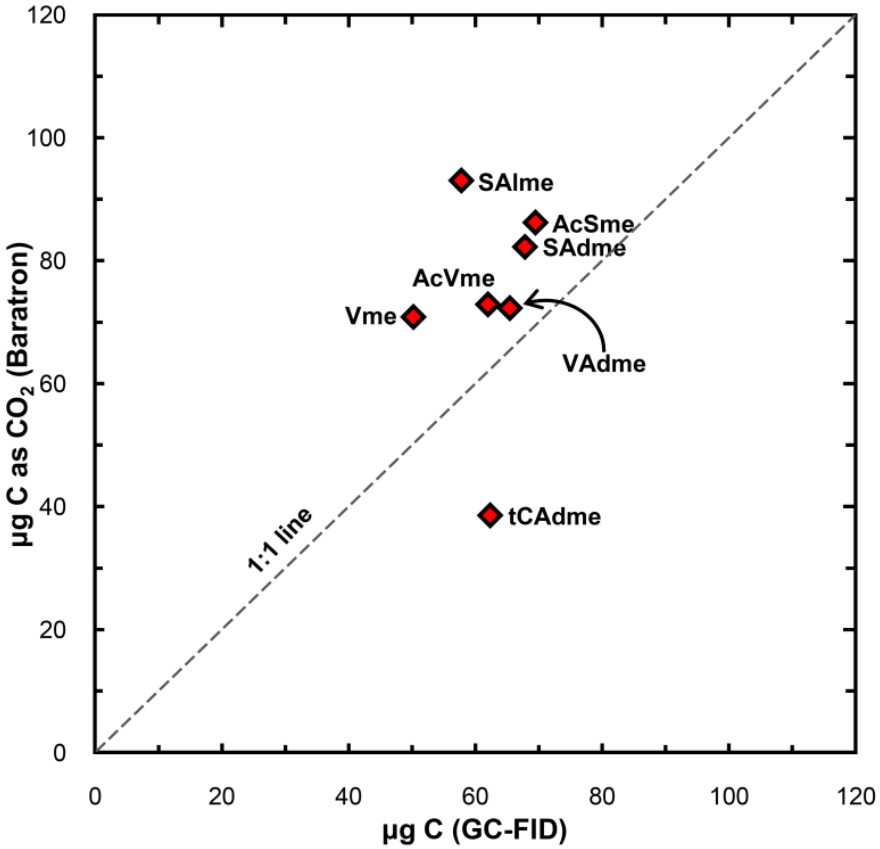


Figure 5.7: GC-FID/MS quantification (as µg C) of the trapped component vs. the manometric (Baratron) CO_2 yield of the trap's contents (as µg C) following CTC and cryogenic CO_2 extraction. The data should generally cluster in the vicinity of the 1:1 line, which suggests the absence of either marked sample contamination (most easily detected when samples lie significantly above line) or volatile sample losses (below line) following trap elution and prior to CTC.

5.7 References

De Alencastro, L. F., D. Grandjean, and J. Tarradellas. 2003. Application of multidimensional (heart-cut) gas chromatography to the analysis of complex mixtures of organic pollutants in environmental samples. CHIMIA International Journal for Chemistry **57**: 499–504.

Bruner, F., G. P. Cartoni, and A. Liberti. 1966. Gas chromatography of isotopic molecules on open tubular columns. Analytical Chemistry **38**: 298–303.

Coplen, T. B. 1996. New guidelines for reporting stable hydrogen, carbon, and oxygen isotope-ratio data. Geochimica et Cosmochimica Acta **60**: 3359–3360.

Deans, D. R. 1968. A new technique for heart cutting in gas chromatography. Chromatographia **1**: 18–22.

Druffel, E. R. M., D. Zhang, X. Xiaomei, L. A. Ziolkowski, J. R. Southon, G. M. Santos, and S. E. Trumbore. 2010. Compound-specific radiocarbon analyses of phospholipid fatty acids and n-alkanes in ocean sediments. Radiocarbon **52**: 1215–1223.

Eglinton, T. I., L. I. Aluwihare, J. E. Bauer, E. R. M. Druffel, and A. P. McNichol. 1996. Gas chromatographic isolation of individual compounds from complex matrices for radiocarbon dating. Analytical Chemistry **68**: 904–12.

Eglinton, T. I., B. C. Benitez-Nelson, A. Pearson, A. P. McNichol, J. E. Bauer, and E. R. M. Druffel. 1997. Variability in radiocarbon ages of individual organic compounds from marine sediments. Science **277**: 796–799.

Fahrni, S. M., M. Ruff, L. Wacker, N. Perron, H. W. Gäggeler, and S. Szidat. 2010. A preparative 2d-chromatography method for compound-specific radiocarbon analysis of dicarboxylic acids in aerosols. Radiocarbon **52**: 752–760.

Gagnon, A. R., A. P. McNichol, J. C. Donoghue, D. R. Stuart, and K. von Reden. 2000. The NOSAMS sample preparation laboratory in the next millenium: Progress after the WOCE program. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **172**: 409–415.

Gillespie, R., R. Hedges, and J. Wand. 1984. Radiocarbon dating of bone by accelerator mass spectrometry. Journal of archaeological science **11**: 165–170.

Goñi, M. A., and S. Montgomery. 2000. Alkaline CuO oxidation with a microwave digestion system: lignin analyses of geochemical samples. Analytical Chemistry **72**: 3116–3121.

Grob, R. L., and E. F. Barry. 2004. Modern practice of gas chromatography, 4th ed. John Wiley & Sons.

Guillén, M. D., and M. J. Manzanos. 2005. Characteristics of smoke flavourings obtained from mixtures of oak (Quercus sp.) wood and aromatic plants (Thymus vulgaris L. and Salvia lavandulifolia Vahl.). Flavour and Fragrance Journal **20**: 676–685.

Hedges, J. I., and J. R. Ertel. 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. Analytical Chemistry **54**: 174–178.

Holmstrand, H., D. Gadomski, M. Mandalakis, M. Tysklind, R. Irvine, P. Andersson, and Ö. Gustafsson. 2006a. Origin of PCDDs in ball clay assessed with compound-specific chlorine isotope analysis and radiocarbon dating. Environmental Science & Technology **40**: 3730–3735.

Holmstrand, H., M. Mandalakis, Z. Zencak, Ö. Gustafsson, and P. Andersson. 2006b. Chlorine isotope fractionation of a semi-volatile organochlorine compound during preparative megabore-column capillary gas chromatography. Journal of Chromatography A **1103**: 133–138.

Hou, J., Y. Huang, C. Brodsky, M. R. Alexandre, A. P. McNichol, J. W. King, F. S. Hu, and J. Shen. 2010. Radiocarbon dating of individual lignin phenols: a new approach for establishing chronology of late quaternary lake sediments. Analytical Chemistry **82**: 7119–7126.

Ingalls, A. E., E. E. Ellis, G. M. Santos, K. E. McDuffee, L. Truxal, R. G. Keil, and E. R. M. Druffel. 2010. HPLC purification of higher plant-dervied lignin phenols for compound specific radiocarbon analysis. Analytical Chemistry **82**: 8931–8938.

Kumata, H., M. Uchida, E. Sakuma, T. Uchida, K. Fujiwara, M. Tsuzuki, M. Yoneda, and Y. Shibata. 2006. Compound class specific ^{14}C analysis of polycyclic aromatic hydrocarbons associated with PM 10 and PM 1.1 aerosols from residential areas of suburban Tokyo. Environmental Science & Technology **40**: 3474–3480.

Leo, R. F., and E. S. Barghoorn. 1970. Phenolic aldehydes: generation from fossil woods and carbonaceous sediments by oxidative degradation. Science **168**: 582–584.

Mandalakis, M., and Ö. Gustafsson. 2003. Optimization of a preparative capillary gas chromatography-mass spectrometry system for the isolation and harvesting of individual polycyclic aromatic hydrocarbons. Journal of Chromatography A **996**: 163–172.

Matucha, M., W. Jockisch, P. Verner, and G. Anders. 1991. Isotope effect in gas-liquid chromatography of labelled compounds. Journal of Chromatography **588**: 251–258.

McNichol, A. P., J. R. Ertel, and T. I. Eglinton. 2000. The radiocarbon content of individual lignin-derived phenols: technique and initial results. Radiocarbon **42**: 219–227.

Meinert, C., and W. Brack. 2010. Optimisation of trapping parameters in preparative capillary gas chromatography for the application in effect-directed analysis. Chemosphere **78**: 416–22.

Mollenhauer, G., and J. Rethemeyer. 2009. Compound-specific radiocarbon analysis – Analytical challenges and applications. IOP Conference Series: Earth and Environmental Science **5**: 1–9.

Ochiai, N., and K. Sasamoto. 2011. Selectable one-dimensional or two-dimensional gas chromatography-olfactometry/mass spectrometry with preparative fraction collection for analysis of ultra-trace amounts of odor compounds. Journal of Chromatography A **1218**: 3180–3185.

Pearson, A., A. P. McNichol, B. C. Benitez-Nelson, J. M. Hayes, and T. I. . Eglinton. 2001. Origins of lipid biomarkers in Santa Monica Basin surface sediment: a case study using compound-specific $\Delta^{14}C$ analysis. Geochimica et Cosmochimica Acta **65**: 3123–3137.

Reddy, C. M., A. Pearson, L. Xu, A. P. McNichol, B. A. Benner, S. A. Wise, G. a Klouda, L. A. Currie, and T. I. Eglinton. 2002. Radiocarbon as a tool to apportion the sources of polycyclic aromatic hydrocarbons and black carbon in environmental samples. Environmental Science & Technology **36**: 1774–82.

Santos, G. M., J. R. Southon, N. J. Drenzek, L. A. Ziolkowski, E. Druffel, X. Xu, D. Zhang, S. Trumbore, T. I. Eglinton, and K. A. Hughen. 2010. Blank assessment for ultra-small radiocarbon samples : chemical extraction and separation versus AMS. Radiocarbon **52**: 1322–1335.

Sasamoto, K., and N. Ochiai. 2010. Selectable one-dimensional or two-dimensional gas chromatography-mass spectrometry with simultaneous olfactometry or element-specific detection. Journal of Chromatography A **1217**: 2903–2910.

Seeley, J. V, N. J. Micyus, S. V Bandurski, S. K. Seeley, and J. D. McCurry. 2007. Microfluidic deans switch for comprehensive two-dimensional gas chromatography. Analytical Chemistry **79**: 1840–1847.

Shah, S. R., and A. Pearson. 2007. Ultra-microscale (5-25μg C) analysis of individual lipids by ^{14}C AMS: Assessment and Correction for sample processing blanks. Radiocarbon **49**: 69–82.

Sterckx, F. L., J. Missiaen, D. Saison, and F. R. Delvaux. 2011. Contribution of monophenols to beer flavour based on flavour thresholds, interactions and recombination experiments. Food Chemistry **126**: 1679–1685.

Stuiver, M., and H. A. Polach. 1977. Discussion Reporting of ^{14}C Data. Radiocarbon **19**: 355–363.

Teuten, E. L., L. Xu, and C. M. Reddy. 2005. Two abundant bioaccumulated halogenated compounds are natural products. Science (New York, N.Y.) **307**: 917–20.

Uchida, M., Y. Shibata, K. Kawamura, M. Yoneda, H. Mukai, A. Tanaka, T. Uehiro, and M. Morita. 2000. Isolation of individual fatty acids in sediments using preparative capillary gas chromatography (PCGC) for radiocarbon analysis at NIES-TERRA. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **172**: 583–588.

Uchida, M., Y. Shibata, K. Ohkushi, M. Yoneda, K. Kawamura, and M. Morita. 2005. Age discrepancy between molecular biomarkers and calcareous foraminifera isolated from the same horizons of Northwest Pacific sediments. Chemical Geology **218**: 73–89.

Zencak, Z., C. M. Reddy, E. L. Teuten, L. Xu, A. P. McNichol, and Ö. Gustafsson. 2007. Evaluation of gas chromatographic isotope fractionation and process contamination by carbon in compound-specific radiocarbon analysis. Analytical chemistry **79**: 2042–2049.

Ziolkowski, L. A., and E. R. M. Druffel. 2009. Quantification of extraneous carbon during compound specific radiocarbon analysis of black carbon. Analytical chemistry **81**: 10156–10161.

CHAPTER VI

Conclusions

6.1 Original motivations of research

The quantity of organic carbon (DOC) in the ocean is similar to the quantity of carbon present in the atmosphere which is similar to the quantity of carbon present in Earth's living biosphere (Hansell and Carlson 1998; Falkowski et al. 2000). In the atmosphere, this carbon exists almost exclusively in a single form: CO_2. In the living biosphere, this carbon is partitioned into species of trees and plants, and biomolecules therein, that can be visually observed, described, and counted. In the ocean, however, <1 % of the organic carbon is present within living marine organisms (Hansell et al. 2009) while the remaining > 99 % of reduced carbon exists as DOC wherein it is partitioned into uncountable, largely unknown, and tenuously defined chemical species that resist compositional and structural elucidation. As much as 97 % of this carbon is considered refractory (Hansell 2013) and resists degradation and re-oxidation to CO_2 on timescales that exceed several millennia (Williams and Druffel 1987; Druffel et al. 1992). As such, oceanic DOC largely exists as a long-term carbon sink compared to the carbon reservoirs of either the terrestrial biosphere or of atmospheric CO_2, whose turnovers are comparatively more rapid. The reasons underlying the existence of refractory DOC within the oxygenated waters of the world ocean must ultimately derive from aspects of its chemical composition and structure, which impart to it properties that make it largely inaccessible to microbial metabolism or to abiotic oxidation, which would otherwise lead to its reconversion to CO_2 and the disappearance of this carbon pool. Types of compounds that have been identified as components of refractory DOC based on their ^{14}C-depletion include black carbon (Ziolkowski and Druffel 2010), compositionally opaque marine humics (Druffel et al. 1992), lipid-like material (Loh et al. 2004), and

amorphously defined compounds such as carboxylic rich alicyclic molecules (CRAM) (Hertkorn et al. 2006) whose structures represent theoretically self-consistent solutions to molecular, elemental, and size constraints derived from analysis of DOM by nuclear magnetic resonance (NMR) spectroscopy (^{1}H and ^{13}C) and Fourier transform ion cyclotron resonance spectroscopy (FT-ICR-MS). The motivating objectives of this dissertation were the development and application of techniques capable of more completely accessing and describing the chemical diversity of this compositionally complex dissolved reservoir of organic carbon. While terrestrial systems were the focus of this dissertation, the techniques that were developed and implemented are readily transferrable to characterizations of marine DOM. In many ways the DOM samples that were studied from oligotrophic Lake Tahoe are representative of marine DOM. DOC concentrations in both Lake Tahoe (37-44 μM) and in the deep ocean are similar and both lack significant and recognizable terrestrial inputs; DOC in both settings instead appear to be dominated by autochthonous algal inputs.

In the last decade, highly resolved characterizations of DOM have been greatly assisted by the advent of FT-ICR-MS instrumentation (Kujawinski et al. 2002; Kim et al. 2003; Koch et al. 2005; Dittmar and Koch 2006; Sleighter and Hatcher 2007). FT-ICR-MS separates compounds by highly resolved masses and identifies thousands of discrete masses within DOM mixtures, each of which is reflective of a unique chemical formula. Almost all peaks identified by FT-ICR-MS, however, correspond to compounds of unascertained identity. Additionally, compounds with the same molecular formula, but which exist as different structures, are resolved as single peaks by FT-ICR-MS even though they constitute different compounds. The technique that has been applied

extensively in this dissertation, two dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOF-MS), similarly resolved thousands of peaks per sample and the identities of the vast majority of these peaks remained similarly unknown, but the chromatographic separation of individual components in two-dimensional retention time (RT) space, plus the availability of an electron impact (EI) mass spectra that accompanied each peak, greatly assisted with tying these thousands of unknown peaks to actual chemical structures. By tying constituents of DOM to actual chemical structures, more significant and conclusive insights can be made into the origins, diagenesis, and fate of DOM in the environment.

6.2 Summary of findings

Chapter II

In Chapter II, GC×GC-TOF-MS was applied as a truly novel technique to examine the chemical complexity of CuO-oxidized natural organic matter (NOM) hydrolysate mixtures. The overriding objective of this study was to tie as many peaks as possible to actual compounds. This was done in order to describe the molecular complexity of CuO-oxidized DOM as comprehensively as possible on a compound-specific basis. The unique separative capabilities of GC×GC-TOF-MS were enabled by two short capillary columns of dissimilar stationary phases that were joined at a cryo modulator. Cryo-modulation produced extremely narrow peak widths (average FWHM = 113 ms) that, when combined with the ultra-fast, full spectra (m/z 1-1000) acquisition speeds (500 Hz) of the TOF detector, revealed several thousand peaks per sample, each

of which was accompanied by its own EI mass spectra. By contrast, traditional 1-D GC characterization of the same CuO-oxidized DOM sample revealed two orders of magnitude fewer peaks. Analysis by GC×GC-TOF-MS additionally separated compounds in two dimensions in a manner that enabled unknown peaks to be linked to their underlying chemical structures. Using a suite of samples that represented potential organic matter (OM) contributors to Upper Truckee River (UTR) DOM, elution patterns of chemically related compounds were identified and mapped. The most commonly observed 2-D RT shifts were those that corresponded to methoxy and methylene additions as well as to replacements of formyl hydrogens with methyls or alcohols. Isomers of identical chemical formulae were also shown to be chromatographically well-resolved and it was shown that molecular surface properties, which could be independently determined by molecular mechanics modeling, governed their separation. Collectively, these features, which are unique to GC×GC-TOF-MS, led to the identification of 451 compounds across all CuO-oxidized samples and this set included novel and newly observed compounds that were identified for the first time in CuO-oxidized OM. A GC×GC classification template was additionally developed to quickly profile the chemical constituents of CuO-oxidized OM in a manner that enabled rapid apportionment of OM source contributors to a particular sample. Following exhaustive analysis of a CuO-oxidized DOM sample from the Upper Truckee River, the identities of 6.5 % of the 4,789 peaks in this sample were ascertainable. This represented the most comprehensive compound-specific analysis of a DOM sample performed to date and revealed ~200 additional compounds within CuO-oxidized OM mixtures beyond the ~120 that have been previously identified during the past three decades of extensive

study of this oxidation product mixture. The described methodology for tying peaks to structures, as well as the mapped separation patterns of homologous and chemically related compounds in the CuO-oxidized OM, laid the groundwork for future analyses of this complex mixture by GC×GC-TOF-MS while providing a roadmap for the chemical identification of additional unknowns. Furthermore, the identification of >300 compounds in CuO-oxidized UTR DOM revealed the substantial and previously unaccessed compositional diversity of this natural DOM mixture. In Chapter II it was noted that more than 90 % of the peaks detected in the CuO-oxidized UTR DOM sample resisted identification. This limitation provided the impetus for Chapter III, which successfully developed and implemented a method by which the entirety of the peak data produced by GC×GC-TOF-MS was integrated, interrogated, and interpreted within an environmental context for a set of ten aquatic SPE-DOM samples from the Lake Tahoe Basin, California, USA.

Chapter III

In recognition that only a small subset of peaks identified by GC×GC-TOF-MS could be tied to known structures, Chapter III developed and implemented a methodology by which populations of peaks could be compared across samples on the basis of retention time alignment. Populations of peaks were defined by both a minimum signal to noise (S/N) intensity and by the chromatographic localization of the peaks in one of four chromatographic quadrants. Mapping of GC×GC space in Chapter II suggested that different types of compounds eluted in different chromatographic quadrants. Using a set of 299 compounds, the types of compounds that eluted in each quadrant were defined and

used as a basis for sorting unknown peaks by their chemical composition. It was shown that Quadrant 1(Q_1), Quadrant 2 (Q_2), Quadrant 3 (Q_3), and Quadrant 4 (Q_4) selected for 'small polar', 'large polar', 'small apolar', and 'large apolar' compounds, respectively. Compounds with higher second dimension retention times (^{2}RTs) eluted in Q_1 and Q_2 and were enriched in formyl, methoxy, acyl, and aromatic functional groups while compounds with lower ^{2}RTs eluted in Q_3 and Q_4 and were depleted in these groups, but were instead enriched in alcohols (–OTMS), acids (–CO_2TMS), and methylene carbons. Peaks detected above a defined threshold were then binned into $Q_{1,2,3,\text{ or }4}$ and, as such, were effectively sorted by their chemical attributes. Alignment of peak populations defined by chromatographic localization and peak intensity was then performed, *pair-wise*, across all samples. Dissimilarity matrices were generated whose elements represented quantifications of the chemical distance separating two peak populations. Linkage analysis was then used to cluster samples by their chemical distance from each other. These clusters were visualized as dendrograms and sample groupings were subsequently identified. It was shown that different populations of peaks differentiated samples to varying extents. In most cases, for example, alignment of higher abundance ($S/N_{\geq 100}$) peaks differentiated samples from one another to a greater extent than did alignment of lower abundance ($S/N_{10\text{-}100}$) peaks. Similarly, alignment of peaks identified in Q_2 differentiated samples to a greater extent than did alignments of peaks identified in Q_3, which yielded comparatively poorer differentiations between the aligned peak populations of compared samples. It was further shown that the clusters resulting from linkage analyses of dissimilarity matrices derived from alignment of high abundance peaks could be independently corroborated using quantitative ^{1}H-NMR data obtained

from analysis of the CuO hydrolysates which were originally analyzed by GC×GC-TOF-MS. Sample groupings produced from alignment of low abundance peaks, however, could not be corroborated by ^{1}H-NMR data due to its poor sensitivity for low abundance compounds. Sample groupings for low abundance peaks (in Q_4) were instead shown to correspond to FA-derived principal component analysis (PCA) scores. As such, groupings derived from the GC×GC alignment methodologies developed in this chapter could be independently corroborated by external measures, providing independent validation of the method. The higher levels of alignment of lower abundance ($S/N_{10\text{-}100}$) peaks compared to higher abundance ($S/N_{\geq 100}$) peaks suggested the existence of a low-abundance, structurally diverse, and numerically large pool of compounds that had high peak homology across samples. It was posited that this could reflect the increasing entropy of DOM composition as it ages and degrades and becomes a more complex, and hence structurally diverse, mixture that becomes increasingly unresolvable on the basis of chromatographic separation, leading to the greater alignment of peaks. The interpretative value of aligning peak populations defined by different criteria was underscored by examination of the July Lake Tahoe sample. On the basis of alignment of high abundance peaks, this sample was most closely related to the other two Lake Tahoe samples. However, on the basis of alignment of its lower abundance peaks, the Tahoe July sample was more similar to the terrestrially influenced samples than it was to the other Lake Tahoe samples, ostensibly due to the alignment of peaks corresponding to products of degraded terrestrial OM.

Additionally, in Chapter III it was shown that aromaticity (as revealed by ^{1}H-NMR) was a poor discriminator of terrestrial vs. algal organic matter in this system due

to the accumulation of a proteinaceous component in Lake Tahoe that contained appreciable quantities of hydrophobic aromatic amino acids that appeared to persist and accumulate despite the high UV flux to this high altitude lake and the susceptibility of their aromatic moieties to UV oxidation. This suggested that these amino acid residues are occluded from direct UV light exposure either within intact proteins or by macromolecular associations of DOM arising from H-bonding or Van der Waal interactions, as has been previously suggested (Nguyen and Harvey 2001). This is in contrast to the lignin phenols, which are similarly aromatic and susceptible to UV oxidation, but which are present only at vanishingly low levels in Lake Tahoe. The ^{1}H-NMR data of CuO hydrolysates revealed a high level of compound-specific detail that enabled quantification of methoxy protons, which were shown to correspond to lignin concentrations and to ratios of syringyl to vanillyl (S/V) ratios. PCA analysis of ^{1}H-NMR-derived functional group abundance data also clearly differentiated autochthonous and allochthonous OM end-members in these systems (PC1).

Chapter IV

Here, a sedimentological record of organic geochemical parameters (δ^{13}C, δ^{15}N, TOC, TN, C:N) spanning 13,500 years to Recent was constructed for Fallen Leaf Lake (FLL), encompassing the whole of the Holocene and the last reaches of the Pleistocene. Additionally, a 4,500-year lignin record was obtained that spanned the Pleistocene-Holocene boundary. FLL is immediately adjacent to Lake Tahoe. An original objective of this research was to investigate the possible preservation of drought-associated geochemical signals in this record. Paleoproxy records, including those from the

California Sierra Nevada, Owens Lake, and Pyramid Lake, (Benson et al. 2002; Mensing et al. 2004; Kleppe et al. 2011) preserve evidence for severe and recurring droughts and mega-droughts that punctuated Holocene climate in the Great Basin and the Sierra Nevada region of North America. The FLL core was interspersed with gravity-driven event deposits (turbidites) that had been previously observed in CHIRP acoustic data (Brothers et al. 2009; Maloney et al. 2013). Elemental and stable isotope analyses across these turbidite intervals showed evidence of organic matter partitioning that appeared driven by hydrodynamic particle sorting. Excluding the turbidite intervals, generally smooth and gradual changes were observed in $\delta^{15}N$, TOC, TN, and C:N throughout the Holocene and these trends were consistent with increasing delivery of total (↑TOC) and terrestrial OM (↑C:N) to FLL sediments. All parameters, except for $\delta^{15}N$, were disrupted by the turbidites. $\delta^{15}N$ values showed significant variability, but trended towards more ^{15}N-depleted values over the span of the Holocene. Comparisons were also made between the sedimentary organic geochemical records of FLL and Lake Tahoe (Smith et al. 2013) and similarities were observed. The FLL $\delta^{13}C$ record showed a trend towards more enriched $\delta^{13}C$ values throughout the Holocene, consistent with increasing terrestrial inputs, although a prolonged interval (~ 1,400 year) was observed (Section III) wherein $\delta^{13}C$ values were depleted. The end of this interval corresponded to the end of drought-like conditions identified by (Benson et al. 2002) based on analysis of paleo-records from Pyramid and Owens Lake. To first order, the $\delta^{13}C$ of sedimentary OM appeared to be controlled by the relative inputs of terrestrial vs. algal organic matter, and the latter was shown to be more ^{13}C-depleted than terrestrial OM in FLL sediments. Analysis of lignin phenols, biomarkers unique to vascular plants, across the Pleistocene-Holocene boundary

preserved evidence of flora succession following the end-Tioga glaciation. The lignin data recorded a proliferation of gymnosperms surrounding FLL that followed the last deglaciation and that lasted up until about 10,900 (calendar) years ago at which point syringyl to vanillyl (S/V) and cinnamyl to vanillyl (C/V) ratios stabilized, evidencing a stabilization in the flora composition of the FLL watershed. Additionally, an early interval of the lignin record revealed strong contributions from gymnosperm-derived lignin that suggested the presence of degraded, relict carbon in the sediments of FLL. It was suggested that this relict carbon had originally been preserved during the last glacial advance before being mobilized into the sediments of FLL during glacial retreat, either as glacial outwash or during an outburst flood event.

Chapter V

In Chapter II hundreds of new compounds were observed in CuO-oxidized DOM. In Chapter III, thousands of unknown compounds that were observed in CuO-oxidized DOM were interpreted within the context of bulk-OM parameters such as DOC, $\delta^{13}C$, and ^{1}H-NMR. Peaks were identified that were suggested to correspond to degraded terrestrial material. These peaks led the Tahoe July SPE-DOM sample to be more similar to the other terrestrially influenced samples than to the other Tahoe samples when low abundance ($S/N_{10\text{-}100}$) peaks were aligned. In Chapter IV, lignin was quantified in the context of changing C:N and $\delta^{13}C$ in a sediment core. The motivation for Chapter V was to develop a technique (Ball et al. 2012) by which individual compounds could be isolated in high purities and yields from complex environmental organic matter matrices for the purposes of subsequent analysis by elemental analysis coupled to isotope ratio

mass spectrometry (EA-IRMS) or by accelerator mass spectrometry (AMS) to measure natural abundance levels of radiocarbon (^{14}C). The isotopic fractionation of the method was quantified and the inverse isotope effect (Bruner et al. 1966) predominated. The method was shown to contribute very low levels of exogenous carbon to preparative enrichments and it was shown that the influence of this carbon could be largely negated through the use of isotope mass balance with knowledge of the amount and $\Delta^{14}C$ of this added carbon.

6.3 Future directions

Chapter II

The thousands of peaks revealed by GC×GC-TOF-MS of CuO hydrolysate mixtures presented an immediate interpretative challenge as most (> 99 %) of the identified peaks were not readily linked to chemical structures on the basis of automated mass spectra library searches. Manual review of library matches was therefore essential. As such, a substantial effort was made in Chapter II to link as many peaks as possible to actual structures. The level of compound identification that was ultimately achieved was 6.5 % of the identified peaks in a single CuO-oxidized, TMS-derivatized DOM sample. Future work may seek to identify more of these compounds. This would be assisted by identifying new series of compounds, by running additional standards (possibly synthesized by combinatorial chemistry approaches), and by employing EI-MS NIST libraries that continue to grow in size each year. Analysis of additional sample types that are known to be enriched in certain compounds would further assist with expanding EI-

MS and relative retention time (RRT) libraries that greatly accelerate the assignments of structures to peaks. Combining GC×GC with mass detectors that offer both high acquisition speeds *and* high mass resolutions would further assist with assigning structures to peaks. In such a setup, peaks that are chromatographically resolved by GC×GC, in addition to having an accompanying EI-MS spectrum, would also have a precise mass for their molecular (M^+) and fragment ions. This would reveal the peak's molecular formula and thus further assist with linking individual peaks to actual compounds. In such a method, sets of isomers could be more easily identified (they would have the same *exact* masses) and this could greatly assist with exploring the isomeric diversity of DOM, which has hitherto escaped inquiry. This may in turn assist with providing constraints on how many different isomers in fact contribute to the mass-resolved peaks identified by FT-ICR-MS.

Additional insights will be gained into the structure of DOM, and ultimately into the modes of refractory DOC formation, as the myriad molecular mixture that encapsulates this carbon reservoir becomes increasingly well-resolved and identification of its individual components is further enabled. CuO oxidation, while it significantly alters the original OM present, still preserves in its oxidation products information concerning the original OM from which it was derived. Aliphatic, alicyclic, aromatic, and unsaturated chemical moieties survive the oxidation. Substitution patterns also remain intact. The GC×GC classification scheme that was also developed in this chapter can be overlaid upon other CuO-oxidized DOM or NOM samples that have been analyzed by GC×GC. This classification scheme enables the compositional diversity of a CuO-oxidized OM sample to be rapidly and visually assessed for its proportions of cutin,

lignin, microbial, or benzene polycarboxylic acids (BPCAs). The OM sample's state of degradation can also be visually inspected, both by the abundance of BPCAs as well as by the relative intensities of ions in the elution regions corresponding to the aldehydic and acidic (carboxylated) lignin monomers.

Chapter III

Most of the peaks identified by GC×GC-TOF-MS correspond to unknown compounds. The location of these peaks within the GC×GC chromatogram, however, preserves information that informs their chemical composition. Different sets of capillary columns could be employed to create peak regions that select for different or more narrowly defined types of compounds. The peaks identified in these regions could then be aligned across a sample set, pair-wise similarities between peak populations could be constructed, and hierarchical relationships of similarity between peak populations could then be revealed – as was done in Chapter III. Alternatively, the same GC×GC method and column setup employed in Chapter III could be applied, but peaks from differently, or more narrowly defined GC×GC regions could be aligned. This technique could also be applied, for instance, to the analysis of CuO-oxidized SPE-DOM or to high molecular weight (HMW) DOM samples isolated from different marine settings, possibly separated by depth (deep vs. surface), by upwelling strength (upwelling vs. stratified), or by region (e.g. arctic vs. an oligotrophic gyre). Applying the methodology developed in this chapter to such samples would assist with the identification of (1) components of refractory marine DOM, (2) hitherto unidentified compounds (that could be species specific) associated with blooms and recent production, or (3) degraded terrestrial components of

marine DOM within arctic waters.

The use of high resolution ^{1}H-NMR to examine CuO hydrolysates was shown to reveal a high level of chemical detail that was otherwise occluded when unoxidized bulk OM fractions were analyzed by this method. The compound-specific interpretations that were afforded by analysis of ^{1}H-NMR data of CuO hydrolysate mixtures suggested that this technique could be useful when applied to other samples. Alignment of ^{1}H-NMR peaks across samples could also be performed, similar to what was done here for GC×GC peaks. Such a technique is already performed somewhat routinely in metabolomics research to identify metabolites that are differentially regulated as a result of some physiological perturbation compared to a control. Additionally, in Chapter III, it was shown that obtaining ^{1}H-NMRs of CuO-hydrolyzed DOM proved extremely useful for the purposes of corroborating the clustering patterns that emerged from alignment of high abundance peaks. As such, the complementarity of NMR and GC×GC-TOF-MS should be exploited for future studies.

Chapter IV

The high resolution organic geochemical record that was constructed for FLL preserved signals of Holocene climatic changes and vegetation shifts that occurred following the onset of the Holocene. Complementarity was observed between the sedimentary records of Lake Tahoe and Fallen Leaf Lake. Analyzing the sedimentary organic matter records from other lakes of the Sierra Nevada could assist with painting a more comprehensive portrait of Holocene drought conditions as they existed in this region, which would help inform predicative forecasts of California's future climate and

snowpack. Such a similar organic geochemical record has recently been constructed for Swamp Lake in northwestern Yosemite National Park where sedimentary records were constructed for $\delta^{13}C$ $\delta^{15}N$, TOC, TN, C:N, n-alkanes, and the deuterium (δD) content of leaf wax lipids (Roach 2010; Street et al. 2012, 2013). The interpretations made from the elemental and isotopic signals contained in the FLL record could be strengthened by increasing the resolution of the record and by measuring additional compound-specific parameters that would assist with understanding the drivers of the elemental and isotope shifts observed in these cores. The availability of a species-specific diatom record for this core could also potentially prove interpretatively useful for the purposes of tying geochemical shifts to climatic changes.

Chapter V

The method described in Chapter V should be applied to additional samples. Compounds destined for preparative enrichment can be identified following the methodologies described in this dissertation by using GC×GC-TOF-MS. While compounds that are resolved by GC×GC-TOF-MS may not be completely resolved using other chromatographic separation techniques, the augmented separation capabilities of the two-dimensional heart-cutting preparative capillary gas chromatograph described in Chapter V is uniquely suited to this task. This technique should be applied to refractory components of marine DOM once they are identified. In the meantime, it can be applied to the enrichment of other compounds destined for compound-specific stable isotope or compound-specific radiocarbon analysis.

Alternatively, this technique could be applied to the isolation of peaks identified in Chapter III that were proposed to represent degraded terrestrial material in Lake Tahoe (the July sample). Measuring the $\delta^{13}C$ of these individual compounds could, for instance, more definitively identify these compounds as terrestrially derived while ^{14}C analyses of these compounds, or of similar ones isolated from marine DOM, could help to evaluate their residence times in these systems and their possible role in the formation of refractory DOC. This technique could also be used, for instance, to isolate and measure the $\delta^{13}C$ of lignin in the glacial outwash event (Chapter IV) to see if lignin contributed to the positive $\delta^{13}C$ TOC signature that was observed in this interval, or if other fractions of OM were driving this ^{13}C enrichment.

6.4 References

Ball, G. I., L. Xu, A. P. McNichol, and L. I. Aluwihare. 2012. A two-dimensional, heart-cutting preparative gas chromatograph facilitates highly resolved single-compound isolations with utility towards compound-specific natural abundance radiocarbon (^{14}C) analyses. Journal of Chromatography A **1220**: 122–31.

Benson, L., M. Kashgarian, R. Rye, S. Lund, F. Paillet, J. Smoot, C. Kester, S. Mensing, D. Meko, and S. Lindström. 2002. Holocene multidecadal and multicentennial droughts affecting Northern California and Nevada. Quaternary Science Reviews **21**: 659–682.

Brothers, D. S., G. M. Kent, N. W. Driscoll, S. B. Smith, R. Karlin, J. A. Dingler, A. J. Harding, G. G. Seitz, and J. M. Babcock. 2009. New constraints on deformation, slip rate, and timing of the most recent earthquake on the West Tahoe-Dollar Point Fault, Lake Tahoe Basin, California. Bulletin of the Seismological Society of America **99**: 499–519.

Bruner, F., G. P. Cartoni, and A. Liberti. 1966. Gas chromatography of isotopic molecules on open tubular columns. Analytical Chemistry **38**: 298–303.

Dittmar, T., and B. P. Koch. 2006. Thermogenic organic matter dissolved in the abyssal ocean. Marine Chemistry **102**: 208–217.

Druffel, E. R. M., P. M. Williams, J. E. Bauer, and J. R. Ertel. 1992. Cycling of dissolved and particulate organic matter in the open ocean. Journal of Geophysical Research **97**: 15639–15659.

Falkowski, P., R. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hógberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffen. 2000. The global carbon cycle: a test of our knowledge of earth as a system. Science **290**: 291–296.

Hansell, D. A. 2013. Recalcitrant dissolved organic carbon fractions. Annual review of marine science **5**: 421–45.

Hansell, D. A., and C. A. Carlson. 1998. Deep ocean gradients in the concentration of dissolved organic carbon. Nature **395**: 263–266.

Hansell, D. A., C. A. Carlson, D. J. Repeta, and R. Schlitzer. 2009. Dissolved organic matter in the ocean: a controversy stimulates new insights. Oceanography **22**: 202–211.

Hertkorn, N., R. Benner, M. Frommberger, P. Schmitt-kopplin, M. Witt, K. Kaiser, A. Kettrup, J. I. Hedges, and P. Schmittkopplin. 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochimica et Cosmochimica Acta **70**: 2990–3010.

Kim, S., A. J. Simpson, E. B. Kujawinski, M. A. Freitas, and P. G. Hatcher. 2003. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18 solid phase disk. Organic Geochemistry **34**: 1325–1335.

Kleppe, J. ., D. S. Brothers, G. M. Kent, F. Biondi, S. Jensen, and N. W. Driscoll. 2011. Duration and severity of Medieval drought in the Lake Tahoe Basin. Quaternary Science Reviews **30**: 3269–3279.

Koch, B. P., M. Witt, R. Engbrodt, T. Dittmar, and G. Kattner. 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Geochimica et Cosmochimica Acta **69**: 3299–3308.

Kujawinski, E. B., M. A. Freitas, X. Zang, P. G. Hatcher, K. B. Green-Church, and R. B. Jones. 2002. The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. Organic Geochemistry **33**: 171–180.

Loh, A. N., J. E. Bauer, and E. R. M. Druffel. 2004. Variable ageing and storage of dissolved organic components in the open ocean. Nature **430**: 877–881.

Maloney, J. M., P. J. Noble, N. W. Driscoll, G. M. Kent, S. B. Smith, G. C. Schmauder, J. M. Babcock, R. L. Baskin, R. Karlin, a. M. Kell, G. G. Seitz, S. Zimmerman, and J. A. Kleppe. 2013. Paleoseismic history of the Fallen Leaf segment of the West Tahoe-Dollar Point fault reconstructed from slide deposits in the Lake Tahoe Basin, California-Nevada. Geosphere

Mensing, S. A., L. V Benson, M. Kashgarian, and S. Lund. 2004. A Holocene pollen record of persistent droughts from Pyramid Lake, Nevada, USA. Quaternary Research **62**: 29–38.

Nguyen, R. T., and H. R. Harvey. 2001. Preservation of protein in marine systems: Hydrophobic and other noncovalent associations as major stabilizing forces. Geochimica et Cosmochimica Acta **65**: 1467–1480.

Roach, L. D. 2010. Climate change in the Pacific North America region over the past millenium: development and application of novel geochemical tracers. University of California, San Diego.

Sleighter, R. L., and P. G. Hatcher. 2007. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. Journal of Mass Spectrometry **42**: 559–574.

Smith, S. B., R. E. Karlin, G. M. Kent, G. G. Seitz, and N. W. Driscoll. 2013. Holocene subaqueous paleoseismology of Lake Tahoe. Geological Society of America Bulletin **125**: 691–708.

Street, J. H., R. S. Anderson, and A. Paytan. 2012. An organic geochemical record of Sierra Nevada climate since the LGM from Swamp Lake, Yosemite. Quaternary Science Reviews **40**: 89–106.

Street, J. H., R. S. Anderson, R. J. Rosenbauer, and A. Paytan. 2013. n-Alkane evidence for the onset of wetter conditions in the Sierra Nevada, California (USA) at the mid-late Holocene transition, ~3.0ka. Quaternary Research **79**: 14–23.

Williams, P. M., and E. R. M. Druffel. 1987. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. Nature **330**: 246–248.

Ziolkowski, L. A., and E. R. M. Druffel. 2010. Aged black carbon identified in marine dissolved organic carbon. Geophysical Research Letters **37**.