# AQUATIC BIODIVERSITY PATTERNS ALONG GRADIENTS OF MULTIPLE STRESSORS AND DISTURBANCE HISTORIES: INTEGRATION OF PALEOECOLOGY AND MOLECULAR TECHNIQUES

Amanda Katherine Winegardner

Department of Biology

McGill University, Montréal

July, 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

## DEDICATION

For my friends who inspire youth with nature. Your passion fuels the science I create.

Fieldwork for this thesis took place on the traditional lands of the Naskapi and Innu-Montagnais First Nations.

McGill University is located on unceded Kanien'kehá:ka traditional territory.

**"Whiskey is for drinking, water is for fighting over"**

- **Mark Twain**

**TABLE OF CONTENTS**

## ABSTRACT

Modern geological time is commonly referred to as the Anthropocene; a designation recognizing the extent to which humans dominate processes and life on Earth. Within this context, a major theme of biodiversity research is to model and predict species losses due to land exploitation and use. However, in order to more completely understand the effect of human stressors on biodiversity, species losses and gains along with biodiversity change over varied temporal and spatial scales need to be considered in concert. My research seeks to fulfill two main objectives related to both biodiversity trends throughout the Anthropocene and the expansion of paleolimnological techniques for biodiversity science. Firstly, by paying closer attention to the way in which beta diversity can uncover trends previously missed when examining alpha or gamma diversity alone, my work helped improve our understanding of freshwater biodiversity responses to the anthropogenic stressors that have accelerated over the last ~ 150 years. Secondly, by integrating paleolimnological data with data collected from contemporary timescales, and with the application of DNA-based approaches to paleolimnology, I answered questions novel to both paleolimnology and biodiversity science. In my first chapter, I used diatom assemblage data from the U.S. Environmental Protection Agency's National Lakes Assessment (NLA) program to compare the variation in diatom assemblages across environmental and spatial gradients, using both water-column and surface sediment data. Here I showed that diatom assemblages from both types of sampling were characterized by environmental and spatial gradients in similar ways. In my second chapter, I extended this work with the NLA data, examining modern and historical (pre-1850 CE) timeframes, and showed that beta diversity responded strongly to national-scale land use gradients, with turnover hotspots in regions with low forest cover. My third chapter focused on a specific stressor for aquatic

biodiversity, metal contamination in an iron-ore mining region of northern Québec, and showed that the beta diversity of zooplankton communities responded strongly to heavy metal loading. Finally, in my fourth chapter I used metabarcoding approaches to more fully characterize microbial eukaryote communities from sediment cores in this same mining region and showed substantial temporal beta diversity in both diatoms and green algae. This final chapter was the capstone for this work, continuing the integration of paleolimnological data with DNA-based approaches, capturing a more complete representation of aquatic biodiversity than possible with individual proxies. I also demonstrated how beta diversity is an important way to characterize diversity in systems experiencing multiple stressors. In general, this research provides insight into the importance of multi-scale and multi-metric methods in the study of aquatic biodiversity, while illuminating key drivers of aquatic assemblages through time.

## RÉSUMÉ

Le temps géologique moderne est communément appelé l'Anthropocène en reconnaissant l'effet important des humains sur nombre processus terrestre, et même la vie sur Terre. C'est dans ce contexte que la recherche sur la biodiversité se focalise en grande partie sur la modélisation et la prévision des pertes d'espèces dues à la surexploitation des terres. Cependant, afin de mieux comprendre l'effet des changements anthropiques sur la biodiversité, nous devons examiner parallèlement les pertes et les gains d'espèces ainsi que le changement de la biodiversité selon diverses échelles temporelles et spatiales. Ma recherche vise donc à répondre à deux objectifs principaux liés à la fois aux tendances de la biodiversité au cours de l'anthropocène et l'application des techniques paléolimnologiques à la science de la biodiversité. Tout d'abord, en me focalisant sur comment la diversité beta peut découvrir des réponses précédemment manqué avec seule la diversité alpha ou gamma, mon travail a contribué à améliorer notre compréhension des patrons de biodiversité dans les eaux douces face aux stress anthropiques des ~150 dernières années. Deuxièmement, en intégrant des données paléolimnologiques et contemporaines aux approches basées sur l'ADN, j'ai pu répondre à des questions nouvelles à la paléolimnologie ainsi qu'à la science de la biodiversité. Dans mon premier chapitre, j'ai utilisé des données d'assemblage de diatomées échantillonné par le programme d'évaluation national des lacs (ENL) de l'agence américaine de protection de l'environnement pour comparer la variation des assemblages de diatomées parvenant de la colonne de l'eau à ceux des sédiments de surface. Ici, j'ai montré que les gradients environnementaux et spatiaux ont eu des effets comparable sur l'assemblage des diatomées parvenant de ces deux types d'échantillonnage. Dans mon deuxième chapitre, j'ai étendu ce travail avec les données de l'ENL, en comparant la diversité beta des diatomées moderne et

historique (pré-1850 EC), et montré que la diversité beta a répondu fortement aux gradient d'utilisation des terres à l'échelle nationale, avec des points chauds de renouvèlement d'espèce en régions avec moins couvert forestier. Mon troisième chapitre a porté sur une contrainte de la biodiversité spécifique dans les écosystèmes aquatiques; la contamination métallique dans une région d'exploitation de minerai de fer du nord du Québec, et montré que la diversité beta des communautés de zooplancton a répondu fortement à l'accumulation des métaux lourds. Enfin, dans mon quatrième chapitre, j'ai utilisé des approches de «metabarcoding» pour mieux caractériser les communautés eucaryotes microbiens de cette même région minière, et j'ai montré une diversité beta temporelle importante chez les diatomées et les algues vertes. Ce dernière chapitre représente un point culminant de mon travail, qui a appliqué des approches basées sur l'analyse de l'ADN aux données paléolimnologiques, capturé une représentation plus complète de la biodiversité aquatique que possible avec des indicateurs individuelles, et examiné la diversité beta comme outil important pour caractériser la diversité dans les systèmes éprouvant multiples facteurs de stress. En général, cette recherche nous a permis de mieux comprendre l'importance de l'approche multi-échelles et des méthodes multi-métriques dans l'étude de la biodiversité aquatique, tout en éclairant les principaux facteurs de changement dans les assemblages aquatique à travers le temps.

## ACKNOWLEDGEMENTS

There are a lot of people who have helped make this thesis a reality, continually supporting and pushing me to abandon my dreams of a goat farm and to get this thing done. I gratefully appreciate the guidance, motivation, enthusiasm and time investment by my co-supervisors, Dr. Irene Gregory-Eaves and Dr. Beatrix Beisner. Every scientific decision I make in the future will have an aspect of your influence as I have grown into my scientist role because of you both. Thank you to the rest of my supervisory committee, Dr. Andrew Gonzalez and Dr. Gregor Fussmann for their insight and to Dr. Alison Derry for her advice regarding DNA barcoding and the Schefferville chapters. I was very fortunate to have had Dr. Pierre Legendre as a sort of unofficial mentor. Dr. Legendre created many of the R functions used in the beta diversity analyses and helped develop new approaches and analyses for beta diversity that shaped my thinking on this topic throughout the writing of this thesis. Pierre, merci beaucoup pour ton attention et l'intérêt dans mon travail. Tes conseils et des critiques ont grandement amélioré mes analyses.

Thank you to Dr. Amina Pollard, Dr. Marina Potapova, Dr. Mihaela Enache, Dr. Don Charles and Dr. Alan Herlihy of the U.S. Environmental Protection Agency for their help accessing samples at the Academy of Natural Sciences (Philadelphia), interpretation of U.S. EPA data and constructive comments on analyses and manuscripts related to the NLA chapters. Thank you to Dr. Paul del Giorgio, Annick St-Pierre and the CARBASS crew for allowing me to tag along on their field campaign in Schefferville in 2012. Thank you to Dr. Shadi Shokralla, Jennifer Spall and Stéphanie Boilard for facilitating lab work at the Biodiversity Institute of Ontario and teaching me that I somehow missed some essential microbiology skills in my

undergrad. Dr. Isabelle Domaizon and Eric Capo at UMR-CARRTEL in Thonon les Bains, France hosted me while completing analyses for the High-throughput sequencing project.

I am indebted to the undergraduate and independent project students who worked on various aspects of these projects. Natasha Salter worked on aspects of the second chapter for her undergraduate honours' thesis as well as an NSERC USRA. Natasha, thank you for going above and beyond making and identifying slides of subfossil cladocerans and for your help in the field (even when that included 12 hour train rides to Schefferville and 4am wake-ups). Carolyn Duthie also worked tirelessly on her undergraduate thesis, developing many R scripts to work with the NLA diatom data. Carolyn, thank you for your positive attitude and help organizing the NLA data. Kathryn Han was the Jack-of-all-Trades for all things lab work in this thesis. Kathryn, thank you for all your time spent extracting resting stages from sediment cores and for the hundreds of DNA extractions. You advanced this project more than you can imagine. Thank you also to Ally Noel and Elisa Cheng for their contributions to the Schefferville projects.

I received financial support through a variety of organizations and agencies. I am grateful for funding from: an NSERC Post-graduate scholarship, the W. Garfield Weston Foundation (Award for Northern Research – Doctoral), the Northern Scientific Training Program (Canadian Polar Commission), Québec Centre for Biodiversity Science (Excellence Awards and a Seed Grant), McGill International Research and Conference awards, NSERC Discovery grant (supervisors' grants), FRQNT (supervisors' grants), and the U.S. Environmental Protection Agency (National Aquatic Resource Surveys Campus Challenge).

Thank you to Dr. Zofia Taranu for continually inspiring me throughout this PhD, especially with her integrity, statistical knowledge and data visualization skills. Personnel of the Gregory-Eaves lab, past and present, have contributed innumerable times throughout the years,

especially Katherine Velghe, Katrine Turgeon, Marieke Beaulieu, Jacob Ziegler, Leanne Elchyshyn, Jennifer Barrow and Joanna Gauthier. Thank you to Sarah Baldwin and Andrew Blakney for writing nights, as well as Dr. Russell Steele and the students involved in the Stats-Biology (S-Bex) exchange group for perspective and very nerdy stats talks. Students and staff of the QCBS network also enriched my PhD experience.

My parents, Mike and Peggy Winegardner have never wavered in their support and understanding of my educational goals. Thank you for never doubting my ideas or life direction. I love you both.

Thank you to my travel agent and friend, Brandon Laforest, for being a very large determinant in completing this thesis. He's also got some great lab efficiency tips. Thank you to Dr. LeeAnn Fishback for always believing in my abilities and giving me many opportunities to explore my non-thesis interests in Churchill. Emma Hodgson and Adrienne Davidson provided great material for side projects and I hope to work with you both many more times in the future. Thank you to Jeff Kennedy for crib breaks as well as Kate Jeffery, Zach Camozzi, Marielle Camozzi (and Fynn!), Diana Trull, Cheryl Patten, Antoine Boyer, Katie FishWard, Mike Ward, Rachel Berry, Lauren Vredenburg, Sara Kuleza, Anne Corkery and Anastasia Richardson for your friendship- it means so much to me to know that you were always there for me, even if a solvent had exploded or an R session aborted. Welcome to the world Finley Ward- I hope you'll know a world where access to freshwater is never an issue. Finally, thank you to Guillaume Lord. Guillaume, you came late to the game, but are nevertheless an essential part of my support team. Now it's your turn to finish!

# PREFACE

## Thesis format and style

This is a manuscript-based thesis in accordance with thesis regulations for McGill University. The four manuscripts are as follows:

Winegardner A.K., Beisner B.E., Legendre P., & Gregory-Eaves I. (2015). Are the landscape-level drivers of water column and surface sediment diatoms different? *Freshwater Biology*, **60**, 267-281.

Winegardner A.K., Legendre P., Beisner B.E., & Gregory-Eaves I. (In revision, *Global Ecology and Biogeography*). Diatom diversity patterns over the past ~ 150 years across the conterminous United States: identifying mechanisms behind beta diversity.

Winegardner A.K., Salter N., Aebischer S., Pientiz R., Derry A.M., Wing B., Beisner B.E., & Gregory-Eaves I. (In preparation). Cladoceran zooplankton diversity in lakes from a northern mining region: responses to multiple stressors characterized by alpha and beta diversity.

Winegardner A.K., Capo E., Domaizon I., Debroas D., Hajibabei M., Shokralla S., Wing B., Beisner B.E., & Gregory-Eaves, I. (In preparation). Microbial eukaryotic biodiversity dynamics during the Anthropocene from a northern mining region: an exploration using High-Throughput Sequencing.

In order to make the formatting style consistent across all four manuscripts as well as general and connecting sections, I have opted to use formatting consistent with *Freshwater Biology*, the journal in which Chapter 1 has been published. Note that the use of first person plural in the individual chapters represents their status as manuscripts with co-authors. First person singular is used for all other sections of the thesis.

**Contribution of Authors**

While all of my thesis chapters have been close collaborations with my co-authors, I was responsible for the design of the projects and the formulation of the hypotheses with the input and guidance of my co-supervisors, Dr. Irene Gregory-Eaves and Dr. Beatrix Beisner. Chapters 1 and 2 make use of open-access data available from the U.S. Environmental Protection Agency, however I performed considerable cleaning and management of this data in order to make it useable for these studies. I conducted the fieldwork for Chapters 3 and 4, and either completed or co-supervised undergraduate students in completing the in-laboratory work necessary for these chapters. I worked as part of a team at the Biodiversity Institute of Ontario under the supervision of Dr. Shadi Shokralla to complete the high-throughput sequencing for Chapter 4 and received assistance from Dr. Didier Debroas to process the Chapter 4 sequencing data through bioinformatic pipelines. I performed the statistical analyses for all of the chapters, working collaboratively with Dr. Pierre Legendre to further develop the temporal analyses used in Chapter 2. I wrote all of the manuscripts in consultation with my supervisors. Specific acknowledgements are laid out in each individual chapter.

**Statement of Originality**

This thesis is a creative effort to integrate paleolimnological and contemporary ecological data, effectively using paleolimnological approaches as a tool for studying ecological dynamics as well as discussing caveats for its use with contemporary (neo-ecological) data. This thesis also quantifies biodiversity change in human-impacted systems; at a large scale across the conterminous United States, as well as a local scale, in a heavily mined northern system. The thesis focuses on both spatial and temporal scales, with the latter three chapters making strong contributions to beta diversity literature.

*Chapter 1*

There is literature focusing on the comparison of paleolimnological and contemporary ecological data, i.e. how similar are assemblages (e.g., diatoms, cladocerans etc.) characterized using samples from sediments versus those characterized from water-column samples (e.g., for macrophytes, Zhao *et al*., 2006). However, for Chapter 1, I was not solely interested in whether the organisms observed in one sampling type matched the other. In fact, there are many reasons as to why you might expect assemblages to not be congruent, even when comparing surface sediments with a contemporary sample (as was the case in this first chapter), including temporal integration of sediment samples and taphonomic biases. The novelty of the approach taken in Chapter 1 was that I compared the conclusions on environmental and spatial variation of diatom assemblages from these two different sampling types as opposed to comparing observed species.

I wanted to know, if a researcher wanted to explore how diatoms varied across both environmental and spatial gradients, but was approaching the study with only one type of sampling, would they reach the same general conclusions as they would have if they had used

the other type of sampling. Building on that same theme, if a researcher wanted to ask this question across an area or region where they could improve the completeness of their sampling size (i.e. the number of lakes or sites included in the study) by combining data from both paleolimnological (surface sediment) and water-column sampling; would it be advisable to do so? And what aspects of data integration would they need to pay attention to? A hypothetical example best illustrates why this was an important focus for Chapter 1, both for conceptual as well as applied/management reasons. Imagine a scenario whereby a researcher wants to better understand how algal assemblages vary across a pollution gradient in a region of interest. The researcher may have the time, equipment and financial resources to take water-column samples from a few lakes near the pollutant source that they have identified. However, their understanding of the effect of the pollutant would be greatly enhanced if they could include additional lakes in the region. If additional lakes in the area had coincidentally already been sampled by a local government agency for another reason (but say using paleolimnological methods because that was that agency's protocol), and the data was publicly available, the work in my first chapter would help this researcher to understand how and under what circumstances they could consider using both (the water-column samples they will collect themselves and the available paleolimnological samples) data sources for their analyses. While perhaps a banal consideration for many, this is a nontrivial consideration for many lake managers and local watershed authorities working with limited resources and financial constraints. Of course, there could be study-specific differences between water-column and sediment samples, but these could be addressed within the context of a mixed effect model.

The questions asked in this chapter are similar to those in Levi *et al*., (2014), and were developed simultaneously and without prior knowledge of the Levi *et al*., (2014) study. I became

aware of the Levi *et al*., (2014) analysis just prior to submission of this first chapter for review in *Freshwater Biology*. My approach in Chapter 1, while echoing Levi *et al*., (2014), improves on the statistical methods by incorporating co-inertia analyses (in this case using the RV coefficient; a method for coupling data matrices (Dray *et al*., 2003). The use of RV coefficients ensured that the comparison of the sediment and water-column samples was not directional in nature, i.e. one type of sampling was not used as a template when in comparison to another, thereby allowing for increased flexibility when forming hypotheses and statistical predictions. Additionally, the Levi *et al*., (2014) study was focused on macrophytes and used a smaller dataset (35 lakes), so it is was unclear whether there conclusions would have translated across trophic levels and to a larger dataset.

*Chapter 2*

The field of beta diversity forms an important component of community ecology study, with extensive debate surrounding methods, interpretation and relevance of this diversity component well described in current literature (e.g., see Legendre *et al*., 2005; Tuomisto & Ruokolainen *et al*., 2006; Soininen *et al*., 2007; Legendre *et al*., 2008; Tuomisto & Ruokolainen *et al*., 2008; Anderson *et al*., 2011). In this Chapter, I chose to analyze beta diversity as represented by pairwise (site by site) comparisons of similarity and focused on quantifying both spatial and temporal hotspots of beta diversity across the conterminous United States. This chapter joins a relatively small amount of articles that have included both spatial and temporal beta diversity analyses in the same study (e.g. Jones *et al*., 2012), and fewer still that have studied spatial beta diversity of the same landscape at different temporal points (for example, Baselga *et al*., 2015). Additionally, this is one of the first studies to use indices designed to detect

significant contributions of a lake to both spatial and temporal beta diversity along with abiotic variables to explain variation in spatial and temporal beta diversity. This approach makes this study a novel contribution to the beta diversity literature for the methodological aspect alone, however the results described also highlight drivers of both spatial and temporal beta diversity for diatoms, information useful in terms of land management and conservation.

*Chapter 3*

Although there is a growing body of literature (e.g., Dixit *et al*., 1996; Doig *et al*., 2015) examining the effect of mining on historical biodiversity, Chapter 3 is an important contribution to the understanding of zooplankton diversity through time in the iron-ore mining region of Schefferville, Québec because of its thoroughness. I devoted considerable resources to considering various radiometric dating models as opposed to relying on a single model. Additionally, I took care to develop geochemical profiles of the lakes specific to this study. The care taken in the abiotic elements of this chapter, ensure that the reconstruction of cladoceran diversity is sound and relatable to the history of the region. Furthermore, it complements what is already known about the impacts of mining in the Schefferville region (Laperrière *et al*., 2008; Aebischer *et al*., 2015). Overall, Chapter 3 presents a comprehensive analysis of zooplankton community responsive to stress, and develops a conceptual framework for the comparison of alpha and beta diversity that could be applied in studies going forward.

*Chapter 4*

The integration of DNA-based approaches with paleolimnology has already demonstrated important advances in ecology, from the differentiation of cryptic species (Bissett *et al*., 2005) to

the amplification of ancient DNA (Coolen *et al*., 2013), and will likely continue to do so as methods are refined and calibrated. In order to apply metabarcoding approaches to the study of aquatic diversity, I chose to continue to work in a well-studied system (Schefferville), where I could consider the results of the paleolimnological metabarcoding study with a diversity of other approaches because I had a lot of historical knowledge to relate to any diversity changes observed with the metabarcoding work. Calibration of metabarcoding approaches for use in paleolimnology are important because increasingly researchers need to be able to reconcile results from traditional taxonomy and DNA-based approaches. For example, if a biodiversity reconstruction using these two different methods indicate different biological histories within a lake, how do we compare and contrast these results? This question and others will be at the forefront of metabarcoding related research in paleolimnology. Chapter 4 makes a contribution to this theme by reflecting on beta diversity trends in the Schefferville system using both traditional paleolimnological techniques and high-throughput sequencing. Additionally, despite caveats related DNA concentrations and primer choice (further elaborated on in Chapter 4), I showed that the microbial eukaryotic assemblages derived from DNA-based approaches in this chapter show similar changes in both alpha diversity and beta diversity through time as was observed in cladocerans in Chapter 3 (traditional taxonomy) and diatoms (traditional taxonomy, Laperrière *et al*., 2008). As such, this work provides additional evidence and verification to the intense biodiversity changes experienced in the Schefferville region.

**Literature Cited**


Aebischer S., Cloquet C., Carignan J., Maurice C., & Pienitz R. (2015). Disruption of the geochemical metal cycle during mining: Multiple isotope studies of lake sediments from Schefferville, subarctic Québec. *Chemical Geology,* **412**, 167-178.

Anderson M.J., Crist T.O., Chase J.M., Vellend M., Inouye B.D., Freestone A.L., Sanders N.J., Cornell H.V., Comita L.S., Davies K.F., Harrison S.P., Kraft N.J.B., Stegen J.C., & Swenson N.G. (2011). Navigating the multiple meanings of ß diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19-28.

Baselga A., Bonthoux S., & Balent G. (2015). Temporal beta diversity and bird assemblages in agricultural landscapes: Land cover change vs. stochastic processes. *PLoS ONE*, **10**, e0127913.

Bissett A., Gibson J.A.E., Jarman S.N., Swadling K.M., & Cromer L. (2005). Isolation, amplification, and identification of ancient copepod DNA from lake sediments. *Limnology and Oceanography: Methods*, **3**, 533-542.

Coolen M.J.L., Orsi W.D., Balkema C., Quince C., Harris K., Sylva S.P., Filipova-Marinova M., & Giosan L. (2013). Evolution of the plankton paleome in the Black Sea from Deglacial to Anthropocene. *Proceedings of the National Academy of Sciences,* 10.1073/pnas.1219283110.

Dixit A.S., Dixit S.S., & Smol J.P. (1996). Long-term water quality changes in Ramsey Lake (Sudbury, Canada) as revealed through paleolimnology. *Journal of Environmental Science and Health. Part A: Environmental Science and Engineering and Toxicology: Toxic/Hazardous Substances and Environmental Engineering*, **31**, 941-956

Doig L.E., Schiffer S.T., & Liber K. (2015). Reconstructing the ecological impacts of eight decades of mining, metallurgical, and municipal activities on a small boreal lake in northern Canada. *Integrated Environmental Assessment and Management,* **11**, 490-501.

Dray S., Chessel D., & Thioulouse J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology,* **84**, 3078-3089.

Jones S.E., Cadkin T.A., Newton R.J., & McMahon K.D. (2012). Spatial and temporal scales of aquatic bacterial beta diversity. *Frontiers in Microbiology*, **3**, 1-9.

Laperrière L., Fallu M-A., Hausmann S., Pienitz R., & Muir D. (2008). Paleolimnological evidence of mining and demographic impacts on Lac Dauriat, Schefferville (subarctic Québec, Canada). *Journal of Paleolimnology,* **40**, 309-324.

Legendre P., Borcard D., & Peres-Neto P.R. (2005). Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435-450.

Legendre P., Borcard B., & Peres-Neto P.R. (2008). Analyzing or explaining beta diversity? Comment. *Ecology,* **89**, 3238-3244.

Levi E.E., Çakiroğlu A.İ., Bucak T., Odgaard B.V., Davidson T.A., Jeppesen E., & Beklioğlu M. (2014). Similarity between contemporary vegetation and plant remains in the surface sediment in Mediterranean lakes. *Freshwater Biology*, **59**, 724-736.

Soininen J., McDonald R., & Hillebrand H. (2007). The distance decay of similarity in ecological communities. *Ecography,* **30**, 3-12.

Tuomisto H., & Ruokolainen K. (2006). Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology*, **87**, 2697-2708.

Tuomisto H., & Ruokolainen K. (2008). Analyzing or explaining beta diversity? Reply. *Ecology,* **89**, 3244-3256.

Zhao Y., Sayer C.D., Birks H.H., Hughes M., & Peglar S. (2006). Spatial representation of aquatic vegetation by macrofossils and pollen in a small and shallow lake. *Journal of Paleolimnology,* **35**, 335-350.

## LIST OF TABLES

**TABLES**

**CHAPTER 1**



**CHAPTER 2**

## LIST OF FIGURES

**FIGURES**

**CHAPTER 1**



**CONNECTING STATEMENT 1**

**CHAPTER 2**

**CHAPTER 3**

## LIST OF APPENDICES

**APPENDIX C – SUPPORTING INFORMATION FOR CHAPTER 3:**



**APPENDIX D – SUPPORTING INFORMATION FOR CHAPTER 4:**

**APPENDIX E – RAW DATA AND INFORMATION ON DATA SOURCES:**

## GENERAL INTRODUCTION

Species richness and diversity are not distributed evenly among freshwater systems, making the understanding of the spatial distribution and maintenance of diversity an important research focus, and one linked with the health and integrity of aquatic resources. Both local and continental scale drivers of biodiversity have been described in the literature. For example, a negative relationship between the number of species found in an area or site with increasing latitude has been well documented for many different organismal groups, from mammals to plants (Rosenzweig, 1995). Despite their apparent ubiquity, microorganisms also often show this same latitudinal pattern (e.g., freshwater phytoplankton (Stomp *et al*., 2011) and freshwater zooplankton (Pinel-Alloul *et al*., 2013)). We have much to learn about the mechanisms behind the distribution of biodiversity across the Earth, both past and present.

Modern geological time is commonly referred to as the Anthropocene; a designation recognizing the extent to which humans dominate processes and life on the Earth (Crutzen, 2002). Although there is still some debate over its exact onset date (Lewis & Maslin, 2015; responses), the Anthropocene concept has been quickly taken up by both the scientific community and the general public. Thus, much recent biodiversity research has sought to understand the distribution and drivers of biodiversity loss documented over the Anthropocene (Steffen *et al*., 2011) and which drivers are expected to compromise ecosystem functioning and services (Cardinale *et al*., 2012). However, several studies, including recent meta-analyses, have shown that species richness trends (local or regional) may be flat (e.g. Vellend *et al*., 2013; Dornelas *et al*., 2014) or show only modest losses (Newbold *et al*., 2015) over the last few centuries. Nonetheless, it is more generally accepted that there have been substantial shifts in community composition (Vellend *et al*., 2013; Dornelas *et al*., 2014; and Newbold *et al*., 2015)

implicating both species losses *and* gains in biodiversity as well as associated ecosystem shifts (Wardle *et al*., 2011; McGill *et al*., 2015). Advancing our understanding in this area is particularly important in freshwater ecosystems, which hold a disproportionately concentrated diversity relative to other environments (Strayer & Dudgeon, 2010).

Gregory-Eaves and Beisner (2011) promote paleolimnology as an emerging field for the study of biodiversity; highlighting that data generated from the analyses of sediment cores can be used to address many of the same questions as contemporary (also referred to as neo-) population, community or ecosystem ecology, and can also provide a long-term perspective ranging from decades to millennia. Paleolimnology has much to offer biodiversity research through the reconstruction of historical conditions (e.g., past lake water pH) and biological assemblages. However, in order to better integrate these fields it is important to understand: the comparability of contemporary and paleolimnological data sets, how different metrics of biodiversity can be applied to paleolimnological (and non-paleolimnological) data, and the potential for molecular tools and DNA-based methods to address current limitations of paleolimnological approaches.

*Paleoecological data for biodiversity research*

Using sediment records, paleoecologists have reconstructed species diversity across many aquatic systems. For example, paleoecological studies of marine plankton have identified latitudinal diversity patterns over millennial time scales (e.g., see Yasuhara *et al*., 2012). Rull (2014) posits the assertion that the lack of integration between paleoecology and contemporary ecology is due to a "psychological barrier" between researchers working in these respective fields. While perhaps a nuanced statement, Rull (2014) does point out seven reasons that there is not more interchange between these different ecological approaches, summarized as: (a) time

resolution of reconstructions in paleoecology; (b) gaps in paleoecological records; (c) low taxonomic resolution in paleoecological data; (d) issues with equating fossil or subfossil based metrics with contemporary ecological metrics; (e) the prevailing view amongst non-paleoecologists that paleoecology/paleolimnology is largely a qualitative or descriptive field and (f) the focus in paleoecology on the reconstruction of historical environments as opposed to historical ecological dynamics. These are very real barriers in terms of increasing the uptake of both paleoecological data by those working as neo-ecologists as well as the integration of data sources from these two domains together, and are perhaps most adequately addressed by focusing on the latter two points ((e) and (f)). Indeed, strong quantitative methods can be (and are being) used for paleolimnological analysis and have identified important ecological dynamics (beyond a simple reconstruction of past environmental conditions; e.g., see Quinlan *et al*., 2005; Nevalainen *et al*., 2011; Taranu *et al*., 2015; and Thienpont *et al*., 2015). Because paleolimnological studies often involve long time series, they are particularly well suited to studies of biodiversity change (beta diversity), yet this potential has not been fully realized in the existing literature.

*Beta diversity as an important biodiversity metric in the Anthropocene*

Beta diversity can be defined as the total variance across a site-by-species data table, but there are additional nuances to these calculations. Beta diversity can be considered either in a directional context, such as across temporal, spatial or environmental gradients (in which case it can be referred to as turnover) or in a non-directional frame of reference. In the spatial context, non-directional beta is simply the variance among sites, whereas in the temporal context it may be calculated as the variance among surveys at each site (Legendre & Gauthier, 2014) or, when

considering two surveys only, as the multivariate dissimilarity between these surveys (Legendre & Salvat, 2015).

Recent work has drawn attention to the insight gained when beta diversity is partitioned into different explanatory components (e.g., Podani & Schmera, 2011; Schmera & Podani, 2011; Baselga & Orme, 2012; Podani *et al*., 2013; and Legendre, 2014). Such analyses allow investigators to quantify the contributions of varying ecological processes to beta diversity, and not just the magnitude of a beta diversity index. The three most common components of beta diversity are species replacement, richness or abundance difference, and nestedness. Species replacement refers to the "simultaneous gain and loss of species (or individuals belonging to a particular species, or biomass) along an ecological gradient" (Podani & Schmera, 2011; Baselga & Orme, 2012). This means that as species (measured as individuals or biomass) are lost from sites (space, time, etc.) others take their place. Richness difference (or abundance difference when using abundance data) means that one site (or sampling unit) has more unique species than another (Podani & Schmera, 2011; Baselga & Orme, 2012). Nestedness is essentially a special case of richness difference where the species found in one site (or sampling unit) are a subset of the species found in another (higher richness) site (Baselga & Orme, 2012).

Partitioning methods can also be used for temporal beta diversity. Legendre and Salvat (2015) developed equations for partitioning temporal beta diversity into species loss and gain components, and applied this to mollusc communities recovering from nuclear testing activity. In another application of partitioning methods, Baselga *et al*., (2015) examined spatial beta diversity of avian communities at various time points and also portioned temporal turnover into components representing species replacement or nestedness (loss and gains). They found that spatial beta diversity did not change significantly between time points, demonstrating that bird

turnover across the landscape of study stayed more or less constant through time, despite habitat change. These studies demonstrate ways in which to include elements of both spatial and temporal beta diversity into the same study, further expanding the ways in which we can use beta diversity to understand biodiversity patterns and dynamics across the Anthropocene.

*The potential for DNA-based approaches in paleolimnology*

Paleolimnological techniques have traditionally been limited to organismal groups that leave distinguishable fossils in the sediment record. However, novel and improved molecular techniques make it possible to extract and amplify both intracellular and extracellular DNA from lake sediments, even degraded DNA that may be thousands of years old (e.g., Giguet-Covex, 2014). The careful integration of DNA-based approaches with paleolimnological techniques will allow ecologists to study more members of aquatic communities and potentially at a higher taxonomic resolution. For example, bacteria are ubiquitous in aquatic systems and play an important role in ecosystem functioning through various trophic pathways. However, since most bacteria (with the exception of cyanobacteria and a few other groups) do not leave fossilized or subfossil structures in sediments, the characterization of most species was generally not possible with paleolimnology. Gaining access to organismal data such as those from bacteria, soft phytoplankton, parasites and bacterivores can shore up information related to historical biodiversity data, and allow us to fill in gaps in knowledge related to temporal dynamics of aquatic food webs.

The power of DNA-based approaches in paleolimnology is not only in their ability to detect species that are not characterized with traditional techniques, but also in the ability to target specific genes with key functions. This has been shown using the mcyD gene (Pal, 2015), which is responsible for microcystin production (Davis *et al*., 2009). Microcystin is the

component produced by cyanobacteria that causes the toxicity of algal blooms (Orihel *et al*., 2012) and hence tracking its appearance through time in sediments can provide important information about historical prevalence and toxicity of bloom events. The targeting of specific genes may have many other implications in future paleolimnological studies (e.g., identification in time and space of colonists, mutation accumulation etc.).

*Overall directions of objectives of this thesis*

This thesis seeks to fulfill two main objectives related to biodiversity trends in the Anthropocene and the expansion of paleolimnological techniques for biodiversity science. First, I wanted to more fully illuminate freshwater biodiversity responses to the anthropogenic stressors that have characterized the last ~ 150 years, paying close attention to the ways in which beta diversity can uncover trends that are not demonstrated by the study of only alpha or gamma diversity. Second, I have integrated paleolimnological data with neolimnological data as well as with DNA-based approaches in order to advance the science of paleolimnology and more generally, biodiversity science. My four chapters relate to one or both of these main themes.

Chapter 1 focuses on diatom assemblage data from a large biomonitoring program administered by the United States Environmental Protection Agency (U.S. EPA); the National Lakes Assessment Program. The 2007 National Lakes Assessment program (NLA) includes data collected both from the water-column as well as from sediment cores from a set of 1200+ lakes across the conterminous United States. In this first chapter, I worked with the diatom assemblages characterized from water-column samples as well as assemblages characterized from the surface sediments of sediment cores collected from these same lakes. Using this paired-sampling design, I asked the central question "would both of these types of samples provide

congruent results on diatom variation across environment and spatial gradients?" While essentially methodological in nature, this question is very important as it examines a rarely talked about divide within aquatic ecology. Studies that collect both types of data across a wide set of lakes are rare, and meta-analyses that use datasets collected with both of these approaches are rarer still. This means that there are many research directions that could benefit from the direct integration of contemporary and paleolimnological data sets if there was a better understanding of how these data sets differed in terms of the results they yield, and of the caveats associated with their integration.

Using the understanding of how contemporary and paleolimnological datasets differed in Chapter 1, Chapter 2 makes use of an additional part of the 2007 NLA database; diatom assemblages characterized from historical sediments. I used diatom assemblages from both historical and surface sediments to examine the extent of both spatial and temporal beta diversity across the US landscape, focusing on regional differences, as well as relationships with water quality variables and modern land cover. To my knowledge, this is the first time that spatio-temporal beta diversity trends have been quantified in the same study, while at the same time considering the components that contribute to beta diversity (both to spatial and temporal beta diversity).

To examine how temporal beta diversity might change over the Anthropocene with more than two sampling points (as in Chapter 2), and how beta diversity could be related to specific drivers in a region, I conducted two studies of an upstream-downstream lake system that have a well-documented but contrasting disturbance history. In Chapter 3, I examined the relationship between cladoceran zooplankton diversity and metal contamination in a historically iron-ore mining region in northern Québec. The town of Schefferville, located in the Labrador Trough

region that straddles the provinces of Québec and Labrador has been a hub for mining activities in the region since the early 1950s (Laperrière *et al*., 2008). Iron-ore extraction in the region has gone through typical boom and bust cycles since the construction of the town in 1954, with heavy extraction from the 1950s to late 1970s, followed by an initial decommissioning of the mines in 1982. In 2009, mining experienced a resurgence, but has declined again in recent years. I used sediment cores from two lakes in the Schefferville region, both exposed to ambient metal loading, but one with a distinct history of additional wastewater contamination to track alpha and beta diversity through time in cladoceran assemblages, and to explore the effect of resting stage deposition on overall Cladoceran richness and beta diversity.

Finally, Chapter 4 of this thesis acts as a capstone to the work by merging my two overall objectives together. The study sites of Chapter 4 are again the Schefferville lakes (same as Chapter 3), but here I applied High-throughput Sequencing (a metabarcoding method) to characterize temporal microbial eukaryotic diversity in these two lakes; this technique provides a much broader taxonomic perspective than the traditional approach of focusing in on individual groups or species. This drastically scales up the work from Chapters 1-3, each of which considered only a single trophic level, allowing us to see the dominance of particular eukaryotic groups in the Schefferville system. Additionally, I compare my results from the High-throughput Sequencing work to my results in Chapter 3 with taxonomically identified cladoceran assemblages and to existing diatom data (also taxonomically identified, Laperrière *et al*., 2008). Overall, through this thesis, I contribute to increasing the scale of investigations of aquatic biodiversity trends in the Anthropocene, spatially, temporally and across more taxonomic groups.

**Literature Cited**


Baselga A., & Orme C.D.L. (2012). betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808-812.

Baselga A., Bonthoux S., & Balent G. (2015). Temporal beta diversity of bird assemblages in agricultural landscapes: land cover change vs. stochastic processes. *PLoS ONE*, 10.1371.

Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D.A., Kinzig A.P., Daily G.C., Loreau M., Grace J.B., Larigauderie A., Srivastava D.S., & Naeem S. (2012). Biodiversity loss and its impact on humanity. *Nature,* **486**, 59-67.

Crutzen P.J. (2002). Geology of mankind. *Nature*, **415**, 23.

Davis T.W., Berry D.L., Boyer G.L., & Gobler C.J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Harmful algae,* **8**, 715-725.

Dornelas M., Gotelli N.J., McGill B., Shimadzu H., Moyes F., Sievers C., & Magurran, A.E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, **344**, 296-299.

Giguet-Covex C., Pansu J., Arnaud F., Rey P-J., Griggo C., Gielly L., Domaizon I., Coissac E., David F., Choler P., Poulenard J., & Taberlet P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications,* **5**, 3211.

Gregory-Eaves I., & Beisner B. (2011). Paleolimnological Insights for Biodiversity Science: An Emerging Field. *Freshwater Biology*, **56**, 2653-2661.

Laperrière L., Fallu M-A., Hausmann S., Pientiz R., & Muir D. (2008). Paleolimnological evidence of mining and demographic impacts on Lac Dauriat, Schefferville (subarctic Québec, Canada). *Journal of Paleolimnology*, **40**, 309-324.

Legendre P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324-1334.

Legendre P. & Gauthier O. (2014). Statistical methods for temporal and space-time analysis of community composition data. *Proceedings of the Royal Society B*, **281**, 20132728.

Legendre,P., & Salvat B. (2015). Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.

Lewis S.L., & Maslin M.A. (2015). Defining the Anthropocene. *Nature*, **519**, 171-180.

Pal S. (2015). Temporal and spatial trends in toxic cyanobacteria as identified through lake sediment DNA. Doctoral Thesis, University of Ottawa.

McGill B.J., Dornelas M., Gotelli N.J., & Magurran A.E. (2015). Fifteen forms of biodiversity trend in the Antropocene. *Trends in Ecology and Evolution*, **30**, 104-113.

Nevalainen L., Luoto T.P., Levine S., & Manca M. (2011). Paleolimnological evidence for increased sexual reproduction in chydorids (Chydoridae, Cladocera) under environmental stress. *Journal of Limnology,* **70**, 155-262.

Newbold T., Hudson L.N., Hill S.L.L., Contu S., Lysenko I., *et al*. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45-50.

Orihel D.M., Bird D.F., Brylinsky M., Chen H.R., Donald D.B., Huang D.Y., Giani A., Kinniburgh D., Kling H., Kotak B.G., Leavitt P.R., Nielsen C.C., Reedyk S., Rooney R.C., Watson S.B., Zurawell R.W., & Vinebrooke R.D. (2012). High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1457-1462.

Pinel-Alloul B., André A., Legendre P., Cardille J., Patalas K., & Salki A. (2013). Large-scale geographic patterns of diversity and community structure of pelagic crustacean zooplankton in Canadian lakes. *Global Ecology and Biogeography*, **22**, 784-795.

Podani J., & Schmera D. (2011). A new conceptual and methodological framework for exploring and explaining patterns in presence-absence data. *Oikos*, **120**, 1625-1638.

Podani J., Ricotta C., & Schmera D. (2013). A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity*, **15**, 52-61.

Quinlan R., Douglas M.S.V., & Smol J.P. (2005). Food web changes in arctic ecosystems related to climate warming. *Global Change Biology*, **11**, 1381-1386.

Rosenzweig M.L. (1995). Species diversity in space and time. Cambridge University Press, United Kingdom.

Rull V. (2014). Time continuum and true long-term ecology: from theory to practice. *Frontiers in Ecology and Evolution,* **2**, 1-7.

Schmera D., & Podani J. (2011). Comments on separating components of beta diversity. *Community Ecology*, **12**, 153-160.

Steffen W., Grinevald J., Crutzen P., & McNeill J. (2011). The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society A*, **369**, 842-867.

Stomp M., Huisman J., Mittelbach G.G., Litchman E., & Klausmeier C.A. (2011). Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology*, **92**, 2096-2107.

Strayer D.L., & Dudgeon D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, **29**, 344-358.

Taranu Z.E., Gregory-Eaves I., Leavitt P., Bunting L., Buchaca T., Catalan J, Domaizon I., Guilizonni P, Lami L., McGowan S., Moorhouse H., Morabito G., Pick F., Stevenson M.A., Thompson P.L., & Vinebrooke R.D. (2015). Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecology Letters,* **18**, 375-384.

Thienpont J.R., Korosi J.B., Cheng E.S., Deasley K., Pisaric M.J., & Smol J.P. (2015). Recent climate warming favours more specialized Cladoceran taxa in western Canadian Arctic lakes. *Journal of Biogeography,* **42**, 1553-1565.

Vellend M., Baeten L., Myers-Smith I.H., Elmendorf S.C., Beauséjour R., Brown C.D., De Frenne P., Verheyen K., & Wipf S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*, **110**, 19456-19459.

Wardle D.A., Bardgett R.D., Callaway R.M., & Van der Putten W.H. (2011). Terrestrial ecosystem responses to species gains and losses. *Science*, **332**, 1273-1277.

Yasuhara M., Hunt G., Dowsett H.J., Robinson M.M., & Stoll D.K. (2012). Latitudinal species diversity gradient of marine zooplankton for the last three million years. *Ecology Letters*, **15**, 1174-1179.

# CHAPTER 1

## Are the landscape-level drivers of water column and surface sediment diatoms different?

Amanda K. Winegardner[1,2], Beatrix E. Beisner[2,3], P. Legendre[2,4], and Irene Gregory-Eaves[1,2]

[1]Deptment of Biology, McGill University, Montréal, Québec, Canada
[2]Québec Centre for Biodiversity Science, Montréal, Québec, Canada
[3]Department of Biological Sciences, University of Québec at Montréal, Montréal, Québec, Canada
[4]Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

Reprinted with permission from Winegardner A.W., Beisner B.E., Legendre P., & Gregory-Eaves I. (2015). Are the landscape-level drivers of water column and surface sediment diatoms different? *Freshwater Biology,* **60**, 267-281. Some formatting modifications were made for this thesis.

**Abstract**

Threats to biodiversity are fostering new collaboration between aquatic ecologists and paleolimnologists, who have traditionally asked ecological questions on different time scales. While the differences between surface sediment and water-column or snapshot sampling are well understood, less so are the consequences of comparing the predominant drivers of aquatic assemblages resulting from these two types of sampling. Using diatom data from the 2007 U.S. EPA National Lakes Assessment program (468 lakes), we compared the main environmental and spatial drivers of diatom community composition between samples derived from the water column and surface sediments. We hypothesized that, in explaining community variation across the conterminous United States, the effect of environment would be stronger in diatom assemblages preserved in surface sediments because of the inclusion of benthic members and temporal integration. We used a combination of ordination overlays and variation partitioning to examine differences in community drivers between paleolimnological (surface sediment) and water-column sampling. We found that these two types of sampling were significantly correlated with respect to the drivers of community composition in addition to having congruent patterns of ordination. Congruency between sampling methods further increased when the water column data were temporally integrated and may be explained by variation in seasonally-dynamic taxa. To our knowledge, this is the first study that has tested for differences in environmental structuring patterns between paleolimnological and water-column samples using such a highly replicated and landscape-level approach. Based on our results, we encourage ecologists to consider the joint analysis of these two types of datasets where data are available.

## Introduction

Globally, freshwater systems provide habitat for about 10% of known species, despite covering less than 1% of the Earth's surface (Strayer & Dudgeon, 2010). To answer questions about how environmental change affects aquatic communities, scientists have adopted two key types of sampling methods: sampling of the water column (live organisms) and sampling lake sediment records (biotic indicators; paleolimnology). Studies focusing on a diverse array of ecological questions make up the history of paleolimnology, but recent attention has focused on historical reconstructions and the use of transfer functions to infer past conditions from subfossil assemblages (Smol, 2008; Birks *et al.*, 2012). Nonetheless, there has also been increasing interest in the long-term ecological perspective that paleolimnology studies can offer (Flessa & Jackson, 2005; Heino, 2009). Gregory-Eaves and Beisner (2011) explicitly champion this approach and discuss the great potential of paleolimnology to contribute to studies of aquatic biodiversity.

As a result of growing interest in using paleolimnological data for new types of questions (Seddon *et al.* 2014) and the movement towards coupling water-column and paleolimnological datasets (Battarbee *et al*. 2005), there is a critical need to understand when they can effectively be used together (i.e. in joint analyses) and how conclusions drawn from these two data sources may differ (i.e. comparability). A recent example of this kind of study was reported by Levi *et al.,* (2014) who quantified the macrophyte community of 35 Mediterranean lakes by surveying both the present-day vegetation and their remains in surface sediments. Their work moved beyond "do we find the same species in present-day and sediment samples?" and asked instead "what variables explain variation in the vegetative communities across these lakes?" and "are the same dominant drivers of variation identified using data from both types of sampling?" This last

question remains unanswered for diatom communities (as are other questions related to this type of comparison), and is especially important because while joint structuring of communities by both environmental and spatial variables is found in many water-column studies (e.g. Cottenie, 2005), paleolimnological studies often show stronger support for environmental structuring (e.g. Verleyen *et al.* 2009).

Clearly, there are fundamental differences between water-column samples and surface sediments, with the latter integrating habitats from across an entire lake and through time, and usually representing multiple years of sediment accumulation (Brothers *et al*., 2008). Some studies have used subfossil assemblages from surface sediment samples to ask about the relative contributions of environmental and spatial factors to community composition, though far fewer than with traditional water-column data. For example, Verleyen *et al*. (2009) used diatom surface sediment calibration sets to show that factors related to local environment explained a median of 21% of diatom variation, whereas spatial variables explained a median of only 5.5%, while pure space (without any influence of environment) did not explain any significant variation in diatom communities. Bennett *et al*. (2010) provides another example of a diatom study conducted across a very large spatial scale, reporting that variables related to dispersal limitation were important at an inter-continental scale, but that pH exhibited an omnibus effect at regional spatial scales. More studies of this nature are needed to better understand drivers of community composition as preserved in sediment samples and to provide data for larger syntheses.

The main focus of our study was to compare the dominant drivers of diatom assemblages delineated from paleolimnological (surface sediment) samples and two types of water-column sampling methods: single-visit samples and temporally-averaged values over repeated summer samplings. For this work we relied on the USEPA National Lake Assessment from 2007

(USEPA, 2009) because it represented both a large sample size, and included both water-column and surface sediment sampling for many of the lakes. Our specific questions were:

(1) How different are diatom assemblage compositions between surface sediment and single visit water-column samples, as well as between surface sediment and temporally averaged water-column samples?

(2) What are the dominant drivers of diatom variation across these lakes? Do conclusions about these drivers depend on whether surface sediment or water-column data are used?

(3) What are the implications of including only planktonic taxa when comparing the ecological conclusions drawn from these different sampling methods?

We hypothesized that, because lake surface sediment samples have a degree of temporal integration, thus providing a longer interval for the immigration, emigration and colonization of taxa than do water-column samples, assemblages within surface sediment would be more strongly explained by environmental gradients. However, we expected that when seasonally averaged water-column samples are used, ecological patterns would more closely match those of surface sediments. We also hypothesized that excluding benthic species from sediment samples would increase the congruence between surface sediment and water-column samples because only planktonic species are being compared between the two types of sampling. Surface sediment samples also integrate spatially across several zones in a lake (Smol, 2008) so that habitat type (benthic or planktonic species) may be an important structuring factor of diatom assemblages that becomes apparent when comparing surface sediment and water-column samples. However, the removal of benthic species from surface sediment samples may decrease the amount of variation explained by environment relative to spatial drivers because benthic species are known to match closely to environmental conditions (Philibert & Prairie, 2002). To

our knowledge, this is the most exhaustive analysis that addresses these questions with freshwater diatoms, and with a common set of lakes for the two types of sampling methods.

## Methods

*Description of National Lakes Assessment (NLA) dataset*

The NLA program, administered by the United States Environment Protection Agency and partnered with state environment and resource agencies, is part of the National Aquatic Resource Surveys program and involves intensive sampling of lakes and reservoirs of the conterminous USA (the lower 48 states) every five years. A full description of the program is available in the "NLA Field Operations Manual" from both 2007 and 2012 (http://water.epa.gov/type/lakes/lakessurvey_index.cfm) and is further summarized in Beaulieu, Pick & Gregory-Eaves (2013). Metadata and raw data from the 2007 campaign is available from the USEPA: http://water.epa.gov/type/lakes/NLA_data.cfm. From the approximately 1000 lakes sampled, we retained 468 sites that had diatom data for both water-column and surface sediment samples.

Field teams from the NLA collected diatoms from the water column using an integrated water sampler (a PVC tube with a length of 2 m and diameter of 3.2 cm) over the entire depth of the euphotic zone ($\geq$ 2m) at the deepest point of each lake. The sampler was deployed twice and the samples mixed together. One litre from the pooled sample was preserved in Lugol's solution for later enumeration (US EPA, 2011-2012). Because of the length of the integrated water sampler, "water column" samples for the purpose of this study refer to two integrated samples of the top 2 m of lake water. Surface sediment was collected using a modified Kajak-Brinkhurst corer (Glew, 1989), again at the deepest point of the lake. Where possible, a 45 cm sediment core

was collected and the top 1 cm section saved for diatom enumeration (minus a 1cm$^3$ subsection from the centre of the sediment slice). Up to 500 diatom valves were enumerated using standardized methods from the U.S. Geological Survey National Water Quality Assessment (NAQWA) (Charles, Knowles & Davis, 2003; USEPA, 2011-2012). Quality Control procedures involved re-identification of a random 10% subset of each sample by a second taxonomist to minimize differences in enumeration and taxonomic disagreement (USEPA, 2011-2012). An explanation of water quality data collection is found in the 2007 field manual (USEPA, 2007).

Sediment cores were collected only once during the sampling period (May to mid-October 2007). Most water samples were collected at the same time as the sediment cores, but for a smaller subset of the 468 lakes, water-column samples were collected (and enumerated for diatoms) both early in the sampling period (May-June; during the sediment coring) as well as later (August to October). We used these revisited sites for comparisons of surface sediment samples to water-column samples, whereby averaging of counts was done post enumeration.

*Data management and pre-processing of NLA data*

Some preprocessing of the open access NLA data was required before statistical analyses could be conducted (Appendix S1). Diatoms were aggregated to species and genus levels and we performed analyses using both of these resolutions. We removed species that did not reach at least a 5% relative abundance in a minimum of a single sample from the dataset as a whole so that abundances were not influenced by rare species. We then transformed diatom species abundances to relativized values using the Hellinger transformation; this transformation is the square root of the relative abundance values per sample (Legendre & Gallagher, 2001). This transformation made the community composition data suitable for beta diversity study (Legendre & De Cáceres, 2013). After screening the environmental data of variables that were strongly

collinear, the environment variables considered in our analyses were: pH (from the field), conductivity (μS $cm^{-1}$), turbidity (NTU), dissolved organic carbon (DOC (mg $L^{-1}$)), ammonium ($NH_4$ (μeq $L^{-1}$)), nitrate + nitrite by flow injection analysis ($NO_3/NO_2$ (mg N/L)), total nitrogen (TN (μg $L^{-1}$)), total phosphorus (TP (μg $L^{-1}$)), chloride ($Cl^-$ (μeq $L^{-1}$)), sulphate ($SO_4$ (μeq $L^{-1}$)), calcium ($Ca^+$ (μeq $L^{-1}$)), magnesium ($Mg^{+2}$(μeq $L^{-1}$)), colour (PCU), silica ($SiO_2$ (mg $L^{-1}$)), hydrogen ions (from pH measured in the lab; μeq $L^{-1}$), hydroxide (from pH measured in the lab; μeq $L^{-1}$), ion balance using acid neutralizing capacity (ANC (%)), chlorophyll *a* concentration (μg $L^{-1}$) and mean Secchi depth (m) (Table 1). Bathymetric maps were not available for all sampled lakes, and so maximum depth ($Z_{max}$) was found using a depth finder. This approximate $Z_{max}$ in metres was also included as a variable.

*Statistical analyses*

Broadly, we were interested in comparing diatom assemblages between sampling types and identifying the dominant drivers of diatom assemblage variation across the set of lakes and with each sampling type (see Figure 1 for an overview of the statistical analyses with corresponding hypotheses). As such, we used Redundancy Analysis (RDA) to identify relationships between the diatom and predictor (i.e. environmental or spatial matrices) datasets, and used co-inertia analysis to quantify the degree of common structure between the water column and surface sediment datasets (Legendre & Legendre, 2002).

Diatom species in the water-column and surface sediment datasets

Diatom species from both the water-column and surface sediment samples were classified as planktonic, benthic and tychoplanktonic using sources from both the primary literature and online databases (Appendix S2). Tychoplanktonic refers to species that are generally benthic, but that will also live in planktonic form if conditions allow (Wehr & Sheath,

2003). Generally, species found in the water-column samples were only planktonic or tychoplanktonic, but species from the surface sediment samples are also often in the benthic class. We created two sets of diatom data, one with planktonic, benthic and tychoplanktonic species and one with purely planktonic species (resulting in a total species richness reduction of 74%, i.e. 26% of the species remained).

To identify the main axes of variation across the species by site matrix, we performed Principal Components Analysis (PCA) for the 468 lakes, for each of the following datasets: (1) the water-column (all species) diatoms; (2) the surface sediment (all species) diatoms; (3) the planktonic-only water-column diatoms; and (4) planktonic-only surface sediment diatoms. To quantify correlation between the assemblage data, we then computed an RV coefficient between the first PCA axis of the water-column diatom matrix and the surface sediment diatom matrix considering all species ($n$ = 468), as well as between these respective matrices with only the planktonic species. For completeness, we also computed the RV coefficient for the full assemblage (see Picazo *et al*. 2012 for a similar approach). The RV coefficient is a multivariate generalization of the Pearson correlation that correlates two matrices with corresponding rows (sites). It produces values between 0 (no correlation) and 1 (perfect correlation). The RV coefficient between two vectors of quantitative data is the square of the Pearson correlation; between two matrices, it is thus homologous to an $R^2$ (Legendre and Legendre, 2012).

Environmental and spatial drivers of diatom variation

We were interested in both identifying the most parsimonious set of environmental and spatial variables that explained the greatest variation in each of the diatom datasets and comparing these results between water column and the surface sediment assemblages. Given that the NLA dataset included numerous environmental variables, we applied forward selection to the

suite of potential predictors, after screening for collinearity. To test for the potential influence of variables related to dispersal, we generated spatial variables for this dataset using the site coordinates and then selected significant variables using both the water-column and surface sediment diatoms (see *Software*). The cut-off for variable retention within the context of the forward selection process was the adjusted $R^2$ of the model containing all variables; its value was 0.082 for the water-column environmental variables, 0.055 for the water-column spatial variables, and 0.090 for the surface sediment spatial variables. Using these reduced sets of environmental and spatial variables, we independently partitioned the variation in water-column or surface sediment assemblages into fractions that were uniquely explained by space (spatial variables) and environment, including shared fractions.

Given that diatom assemblages can be highly seasonal, we also wanted to know whether the important environmental variables driving diatom assemblage variation would change depending on when the lakes were sampled during the growing season. To examine this, we extracted a subset of 51 lakes having surface sediment samples (from a single core sample), as well as water-column diatoms enumerated from a first sampling visit and a second set of water-column data from a second sampling visit (see Appendix S1). For this section, we focused only on the environmental variables. We again ran forward selection on the full set of environmental variables, this time for these 51 sites, using the first visit (visit 1) set of water-column diatom community data as well as the second visit (visit 2). We again used a cut-off criterion for forward selection that reflected the adjusted $R^2$ of the model containing all the variables, which was 0.084 for visit 1 and 0.071 for visit 2.

To quantify relationships between the water-column diatom assemblages and our reduced set of environmental variables, we performed a Redundancy Analysis (RDA) with the (Hellinger

transformed) water-column diatom data and forward-selected environmental variables. To examine whether surface sediment diatoms showed similar patterns within the RDA and to examine relationships between the environment and surface sediment diatoms, we then performed a RDA with the Hellinger transformed surface sediment diatom data and the same environmental variables. This reflects the approach taken in many paleoecological studies using surface sediments; that is, to quantify subfossil organisms from surface sediments, while measuring environmental variables from the water column (e.g. Kurek *et al.* 2011). This RDA approach was done for all species ($n$ = 468) as well as for the planktonic-only datasets ($n$ = 468), and for the visit 1 and visit 2 datasets, as well as an average of both visits (averaged post enumeration) ($n$ = 51). We then repeated these RDAs for the reduced set of spatial variables, again to identify relationships between spatial variables and diatom assemblages (for the $n$ = 468 datasets only).

Correlation between environmental and spatial ordinations

After performing the RDAs, we quantified the resemblance of the water-column site scores and surface sediment site scores from the various RDAs. To do this, we extracted the water-column site scores, as well as the surface sediment site scores from each RDA. We then computed an RV coefficient for the water-column versus surface sediment site scores for: (1) the environmental RDA including all species; (2) the spatial RDA including all species; (3) the environmental RDA including only planktonic species, (4) the spatial RDA including only planktonic species; (5) the visit 1 water-column (environmental) RDA including all species; (6) the visit 2 water-column (environmental) RDA including all species; and (7) the mean of visit 1 and 2 water-column (environmental) RDA including all species.

Comparing sampling types

We used partial RDA, to find out if there were significant differences in diatom assemblage composition between the water column and the surface sediments. This form of analysis, which is the multivariate equivalent of a paired *t*-test, used the sampling methods as the explanatory variable (water column or surface sediment) and the lake identifiers as covariables. Contrary to co-inertia analysis, the two data sets to be compared were placed one on top of the other, matching the species columns, while keeping the lakes in the same order in the two parts of the combined data sets. We carried out this analysis for the matrix of 468 lakes with all species, the matrix of 468 lakes with only the planktonic species (sampling methods being either water-column or surface sediment), and the matrix of 51 lakes with all species (one analysis where the sampling methods being water-column visit 1, water-column visit 2 and surface sediment; and another where the sampling methods were the mean of the water-column visits and the surface sediment).

*Software*

For all statistical analyses and the majority of data pre-processing, we used R v. 3.0.2 (R Core Team, 2013). We tested environmental variables for normality using the Shapiro-Wilk test (*Shapiro.test*() (`stats`)) and *skewness*() in `moments` (Komsta & Novomestky, 2013). The Box-Cox transformation was used to normalize non-normal variables by applying *boxcox.fit*() (`geoR`) (Ribeiro & Diggle, 2013). Spatial variables were generated by developing a matrix of synthetic Moran's Eigenvector Maps (i.e. distance-based MEM) using the geographic coordinates of the lakes on a Cartesian plane and the appropriate functions from packages `PCNM`, `ade4` (Dray *et al*., 2013), `spacemakeR` (Dray *et al*., 2013) and `packfor` (Dray *et al., 2013*). Third-order polynomials were also generated for the site coordinates to act as more

simple spatial variables. Forward selection of both environmental and spatial variables was completed using *forward.sel*() in `packfor` and verified using forward-backward-stepwise selection using *ordistep*() in `vegan`. The water-column diatom matrix was detrended for use in the selection of environmental variables. Detrended water-column and surface sediment diatoms were used for the selection of spatial variables. We used *varpart*() in `vegan` (Oksanen *et al.,* 2013) to perform the variation partitioning analysis, and used the *rda*(), *predict.rda*(), and *scores*() functions of that same package for the ordination work. We computed RV coefficients using *coeffRV*() in package `FactoMineR` (Husson *et al.* 2014).

**Results**

Diatom species in the water-column and surface sediment datasets

We performed analyses at both the species and genus levels, but only the species-level results are shown, as genus-level results did not differ greatly (see Appendix S3). After removing species with low abundance from the data matrices (species with less than 5% relative abundance in any sample), the total number of diatom species used in analyses was 456 (338 benthic or tychoplanktonic forms and 118 planktonic species). Total species richness was 228 in the water-column diatom matrix and 382 for the surface sediment. Including species from all habitats, the water-column diatom matrix was significantly correlated with the surface sediment diatom matrix, with an RV coefficient of 0.23 for the first axis of variation in the PCA ($P$<0.001; see Table 2). With planktonic-only species, the RV coefficient was 0.24 ($P$<0.001). Independent PCAs of the two types of diatom datasets enabled us to ascertain the similarity in the distribution of taxa across sites. Qualitatively we observed that *A. granulata* and *F. crotonensis* were the dominant taxa driving the first and second PC axes in both datasets (Fig. 2a,b).

Environmental and spatial drivers of diatom variation

From the original set of 20 environmental variables, 14 were retained by the selection procedure to explain diatom community structure in the water-column dataset: mean Secchi, $Z_{max}$, $SO_4$, DOC, Mg, conductivity, Ca, TP, $NH_4$, Cl, TN, turbidity, chl *a* and colour (see Appendix S4 for information on the Box-Cox transformed environmental variables). A similar set of environmental predictors were identified when the surface sediment dataset was used as the response matrix. Forward selection using subsets of the water column data resulted in few significant environmental predictors. When only the 118 planktonic species from the water column dataset were considered, the significant environmental variables were conductivity, Ca, mean Secchi, Cl, chlorophyll *a* and observed $Z_{max}$. The forward selected variables identified using only visit 1 water-column diatoms, or only visit 2 water-column diatoms, yielded only three significant variables: TP, conductivity and turbidity (the same for each of the visit data).

Diatom assemblages from the water column were mostly explained by productivity-related variables (i.e. mean Secchi depth and chl *a* with the water column samples) as well as $Z_{max}$ (RDA1 = 0.43 variation explained) and to a lesser extent by variables related to lake identity or catchment chemistry (RDA2 = 0.14) (Fig. 3a). Diatom assemblages from the surface sediment were also mostly explained by productivity-related variables (RDA1 = 0.41 variation explained) (Fig. 3b). Planktonic-restricted RDA biplots showed similar sorting patterns to the complete diatom assemblage plots with a primary axis related to chlorophyll *a,* colour and mean Secchi for both the water-column diatoms and surface sediment diatoms (figures not shown). The first and second RDA axes explained 60% and 16% of variation for the water-column diatoms, and 57% and 17% for the surface sediment diatoms. RDAs using the significant variables for the 51 lake dataset of the visit 1, visit 2, averaged visit samples and surface

sediment samples showed turbidity along the primary axis of all four ordinations (figures not shown). RDA1 values were 0.48, 0.57, 0.41 and 0.45 for visit 1, visit 2, mean visits and surface sediment, respectively.

The spatial RDAs were similar to the environmental RDAs, in that the surface sediment site scores (Fig. 4a) displayed a similar pattern in the ordination to the water-column site scores (Fig. 4b), for both the RDA with all species as well as the planktonic-only RDA. The first RDA values were 0.35 and 0.44 for the water-column spatial RDA and the surface sediment spatial RDA respectively, with the second RDA values being 0.17 and 0.20 (RDAs with all species). For planktonic species only, the first and second RDA values were 0.35 and 0.18 for the water-column RDA and 0.49 and 0.19 for the surface sediment RDA. As evident from both the environmental and spatial RDAs, sites appear to be structured in the same way for both surface sediment and water-column samples across both types of variables.

Contrary to expectations, the amount of variation explained by space did not differ substantially between the water-column and surface sediment samples. Pure space explained about 3.8% of variation in the water-column diatoms and about 5.6% in surface sediment diatoms. Pure environment explained about 4.9% of variation in water-column diatoms and about 5.4% in surface sediment diatoms, with 88% of variation being unexplained for water-column diatoms and 85% for surface sediment diatoms. Removing benthic species from the diatom matrices (such that only 118 planktonic or tychoplanktonic species remained) did not change the proportion of total variation explained by either environmental or spatial variables.

Correlation between environmental and spatial ordinations

The first axis of variation in the environmental RDA explained 43% of variation in the water-column assemblage and 41% of the surface sediment assemblage. The RV between site

scores of the water-column diatom assemblage and the site scores from the surface sediment RDA for this first axis was 0.54 ($P < 0.001$; RV = 0.17 for full set of scores). As we found with the environmental matrix, there was consistency between the spatial predictors when RDAs were performed using either the water-column or surface sediment datasets (Fig. 4). The RV coefficient value for the correlation between water-column RDA1 (35% variance explained) site scores and surface sediment RDA1 (44% variance explained) site scores was 0.54 for the spatial RDA ($P<0.001$; 0.16 for the full set of scores). As such, there appears to be quantifiable congruence between the species-by-site data from both the water-column and surface sediment samples.

For the planktonic only analyses, RV coefficients between RDA1 water-column site scores and RDA1 surface sediment site scores were 0.50 for the environmental RDA (using both actual surface sediment site scores and predicted) and 0.53 for the spatial RDA 1 ($P < 0.001$; RV = 0.22 (env) and 0.10 (spatial) for full set of scores). The strength of the correlation between averaged water-column samples and surface sediment samples was stronger than the correlation between single snapshot and surface sediment samples. For the water-column visit 1 comparison to surface sediment, the RV coefficient was 0.38 for the first axis and 0.03 for all the axes ($P <0.001$). The correlation was weaker for the water-column visit 2 comparison with surface sediment with an RV coefficient of 0.07 ($P = 0.001$; 0.25 for the full set of axes). However, the strength of the correlation increased when comparing the averaged water-column data to the surface sediment data, RV = 0.5 ($P<0.001$; 0.47 for the full set of axes).

Comparing sampling types

Partial RDAs constrained diatom data to sampling method while controlling for the variation among lakes, since the lake sites were the same for both the water-column and surface sediment

samples. The proportion of variation explained by sampling type (water column or surface sediment) for the $n = 468$ dataset of all diatom taxa was 0.015 (adj. $R^2 = 0.014$; pseudo $F$-value = 13.9; $P = 0.005$). The proportion of variation explained by sampling type with planktonic species only was 0.013 (adj. $R^2 = 0.012$; pseudo $F$-value = 12.7; $P = 0.005$). The proportion of variation explained by sampling type when considering water-column visit 1, water-column visit 2 and surface sediment was also 0.013 (adj. $R^2 = 0.010$; pseudo $F$-value = 3.2; $P = 0.005$). The proportion of variation explained by sampling type when considering the mean of the water-column visits and surface sediment was 0.009 (adj. $R^2 = 0.010$; pseudo $F$-value = 3.2; $P = 0.015$). These results were consistent with our other analyses, in that there was a negligible effect associated with the sampling method.

**Discussion**

Paleolimnology has been used extensively in tracking long-term environmental change. However, there are also numerous examples of how paleolimnological approaches can be applied to ecological questions on more contemporary time-scales: response to nutrient reduction (Battarbee *et al.* 2005), tracking invasive species (Hawryshyn *et al.* 2012), space for time substitutions (Blois *et al.* 2013) and questions related to human-environment interactions, biogeochemical cycling and combining multiple records (Seddon *et al.* 2014). An awareness of both the shared attributes of these sampling types, as well as their differences is crucial when using these data in concert. Perhaps more critically, an understanding of where there is the potential to draw different conclusions about the effects of environmental variation on aquatic community composition is necessary, especially as data from these two different sampling

methods are increasingly being integrated into ecological research (e.g., Gregory-Eaves & Beisner, 2011; Velghe and Gregory-Eaves, 2013).

We found that both types of datasets yielded similar relationships with the environmental and spatial predictors, despite a low amount of explained variation. The most prominent environmental variables related to this 468 lake dataset were mean Secchi depth, $Z_{max}$, chlorophyll *a* and colour, irrespective of whether benthic taxa were included, or whether water-column or surface sediment samples were considered. When a smaller subset of lakes with multiple sampling dates was considered, the main environmental variables were conductivity, total phosphorus and turbidity. Thus with this large set of study lakes, researchers analyzing environmental data would have drawn a similar (RV = 0.54, $P < 0.001$ relationship between water-column and surface-sediment RDA scores) conclusion, regardless of whether they had access to surface sediment or water-column diatom counts; diatom variation was mainly structured by lake primary productivity. It is worth noting that RV coefficients were lower when looking at the correlative structure amongst matrices representing the full set of RDA scores, but still significant. We also found that variation across spatial variables was similar between surface sediment and water-column diatoms, and that the significant spatial structure identified for both types of sampling was reflective of regional scale processes. This result echoes findings from a few other studies that have identified the importance of space across larger scales (Verleyen *et al.* 2009, Bennett *et al.* 2010).

Our original hypotheses emphasized differences between water-column and surface sediment samples, mostly with respect to the integration across habitat types and time scales (seasonal or even annual). Our rationale was that environmental variables would more fully explain variation in diatom communities preserved in surface sediment than captured from the

water column for two primary reasons: first, because benthic diatom species more closely track environmental conditions than do planktonic species (Philibert & Prairie, 2002), and second because paleolimnological samples integrate over a longer time period. This means that surface sediment diatom assemblages would reflect communities observed over a longer period of time (at least an entire growing season), capturing species that are temporally transient or may show a patchy distribution in a system, thereby resulting in more complete species sorting across environmental gradients. We found only an approximately similar relationship between the water-column environmental RDA and surface sediment environmental RDA when including only planktonic species (RV = 0.50 versus RV = 0.54 with all species); however our study did provide an insight into why we instead found congruence between these datasets despite differences in species, and this information could be useful when planning sampling methods or combining data, a main goal of the study.

Interestingly, variation explained by the environment was slightly greater in surface sediments than in water-column sediments. Variation explained by space was also slightly greater in surface sediments when compared to water-column sediments, resulting in overall lower unexplained variation in diatom assemblages when using surface sediment samples. These differences were minor though (e.g. 88% unexplained variation for water-column diatoms, 85% unexplained variation for surface sediment diatoms). This means that there was but weak support for our hypothesis that diatoms from surface sediment samples would be more strongly structured by environmental variables than by spatial eigenfunctions (as studied using dbMEM spatial variables). While the large amount of unexplained variation in the variation partitioning analyses necessitates a cautious interpretation of these results, the amount of variation explained is not disproportionate to other large surveys.

We did find evidence that the relationships of diatom assemblages from surface sediment samples to environmental variation was significantly more similar to seasonally-averaged water-column samples than to individual snapshot samples (regardless of time in the growing season). In particular, we saw a higher RV coefficient between scores from an environmental RDA where surface sediment diatoms were compared to water-column samples where diatom species collected from two visits from early and later in the growing season were averaged. This could relate to an aspect of time integration that we did not consider in our initial hypotheses. In particular, surface sediment samples (and time-integrated water-column samples) are probably more similar to each other because both more accurately reflect cyclical changes of abundance amongst diatom species. This is true even if their variation is not explained in a significantly higher proportion by the environmental variables.

Diatom species are highly dynamic and generally display two peaks in abundance throughout the growing season, in spring and early autumn. It is well known that environmental variables related to lake productivity are also seasonally cyclical. This family of variables (e.g. Secchi depth, chlorophyll *a*) most effectively represented variation in our diatom communities, and thus may explain why integrating over the growing season results in a closer match between surface sediment samples and temporally averaged water-column samples. Our PCAs of diatom assemblages identified *Cyclotella* spp. and *Fragilaria* spp. (*sensu stricto*) as key diatom taxa. Both of these genera contain species known to show large seasonal peaks (at least in ponds), with autumn being an important month in temperate systems for some of these species (Köster & Pienitz, 2006). *Cyclotella* spp. also show periodicity in paleolimnological records (Saros & Anderson, 2014). The set of 468 lakes used for the main analysis and the first visit samples for the smaller subset of 51 lakes, which consisted of water-column samples collected in May or

June, would only capture at best one of the large seasonal peaks of phytoplankton, whereas the surface sediment samples collected at the same time would have included diatoms from peaks in abundance of the previous growing season. For genera like these two examples, timing of water-column sampling can result in different community compositions, altering the conclusions drawn about metacommunity composition. As a result, while both water-column and surface sediment samples yielded the same environmental signals, their assemblage resemblance appears to depend on the timing of sampling and can be enhanced by comparing surface sediment samples to averaged data from the water column over multiple sampling points.

Recent studies of environmental drivers of water-column diatom composition have found that many variables are significant contributors, including lake productivity, longitude, nitrate, nitrate/nitrite levels, pH, phosphate, silica, stratification, TP and percent surrounding vegetation (Vanormelingen, Verleyen & Vyverman, 2008; Soininen & Weckström, 2009; Ptacnik *et al*., 2010; Gottschalk & Kahlert, 2012). A similarly wide set of environmental variables have been found to affect diatom surface sediment composition (core samples or sediment traps), including Ca, chlorophyll *a*, Cl, conductivity, elevation, K, lake circulation, Mg, Na, pH, surface area, TN and TP (Dixit *et al*., 1999; Köster & Pienitz, 2006; Hausmann & Pienitz, 2009; Leira *et al*., 2009; Verleyen *et al*., 2009; Hájek *et al*., 2011; Bennett *et al*., 2010). In some cases, the connectivity of habitats and variables related to dispersal limitation have been identified as important predictors at certain scales (Vyverman *et al*., 2007; Vanormelingen *et al*., 2008). With such a wealth of knowledge present in the literature, the challenge is not in finding studies to corroborate the importance of a candidate variable, but in realizing that with different gradients and different measured environmental variables, many outcomes are possible with respect to drivers of diatom community composition. The relevance to this study is that, for any of these

different study examples, the same conclusion about the important environmental variables could probably have been reached regardless of using water-column or surface sediment diatom samples. This is shown by correlated ordination structure, but also by forward selection of environmental variables using the different sampling types.

Unlike many other sampling programs where a particular environmental gradient is targeted, the primary goal of the NLA survey was to randomly sample from all lakes in the continental USA that were deeper than 1 m and larger than 1 ha in surface area. As such, the relatively high nutrient status evident in this dataset (i.e. the median values for total phosphorus is indicative of eutrophic conditions and mesotrophic based chlorophyll *a*) is reflective of the average trophic state of most US lakes. Previous research in more oligotrophic (nutrient poor) systems have shown lake pH to be a dominant structuring variable for diatoms (e.g., Ginn *et al.*, 2007; Valois *et al.*, 2011), with acidification resulting in a loss of planktonic taxa (Battarbee *et al.*, 1984) and close tracking of benthic taxa to environmental gradients. If pH had been a more important variable for diatom communities across the lakes in this study, we may not have come to the same conclusions about diatom variation with both the water-column and surface-sediment samples; instead the different types of samples may have yielded different results (not in taxonomic composition, but in the key structuring variables identified). While we recognize that different regions with particularly low or high pH levels may see a significant effect on diatom communities, pH may not be as important at the continental scale as it was in the 1980s (e.g. Wigington *et al.*, 1992).

A further impetus behind our study was to provide insights into some of the perceived challenges associated with comparing contemporary and paleolimnological studies, including their joint use. We think this type of work is central to collaborative research in the aquatic

sciences. The general message from our analyses is that there are broadly similar patterns from the analyses of diatom communities as captured by the surface sediments and water-column samples, although the greatest similarity is evident when water-column samples are pooled across time to reflect a time-integrated sample. Data sharing is one way in which these two branches of aquatic ecology can work together more concretely, and this is especially important as both the availability of data and requirements for data storage evolve in ecology (Hampton *et al*., 2013). Nonetheless, there are many reasons why researchers may choose one type of sampling over another for a given study. For example, paleolimnology studies have been very useful in quantifying environmental change in a large number of lakes (i.e. upwards of 50) through analysing pre-industrial (pre-1850 CE) and surface sediments (e.g., Dixit *et al*. 1999). On the other hand, direct water-column sampling can allow for a more thorough representation of total algal community diversity (as opposed to just diatoms which are often the target of paleolimnological studies).

*Future research directions and potential implications for monitoring programs*

This study was singular in its focus on diatom taxa. However, as is evident in many of the works cited herein, this question is also important for other organismal groups. As such, follow-up studies could conduct similar analyses with zooplankton subfossils and compare results to those presented in this work, and with studies utilizing sediment traps and net samples to track seasonal changes in zooplankton communities (e.g. see Nykänen *et al*., 2009 and Alric & Perga, 2011). While our study did focus on the effect of averaging water-column samples across a season, we were not able to compare time series data from more than one year to a full core sediment record. This would be a logical extension to our work where we have shown

comparisons between early visit and later visit water-column sampling to surface sediment samples.

**Acknowledgements**

The authors wish to thank Amina Pollard of the NLA program for help with data access and interpretation, as well as Mihaela Enache and Marina Popatova at The Patrick Center for Environmental Research (Academy of Natural Sciences Philadelphia; ANSP) for diatom information and help procuring samples. Members of the Gregory-Eaves lab at McGill contributed greatly to the development of this work. We also thank two anonymous reviewers for their comments on an earlier version of this manuscript. Funding for this project was made possible by an FRQNT Equipe grant awarded to IGE, BB and PL, as well as an NSERC PGS-D scholarship awarded to AKW. Travel to collect samples from the ANSP was supported by a McGill International Travel Grant. This project has also benefitted from the support of a Québec Centre for Biodiversity Science excellence award to AKW. This study is a contribution to the FQRNT-funded Québec Centre for Biodiversity Science (QCBS-CSBQ).

**Literature Cited**


Alric B., & Perga M.E. (2011). Effects of production, sedimentation and taphonomic processes on the composition and size structure of sedimenting cladoceran remains in a large deep subalpine lake: paleo-ecological implications. *Hydrobiologia,* **676**, 101-116.

Battarbee R.W., Thrush B.A., Clymo R.S., Le Cren E.D., Goldsmith P., Mellanby K., Bradshaw A.D., Chester P.F., Howells G.D., & Kerr A. (1984). Diatom analysis and acidification of lakes. *Philosophical Transactions of the Royal Society B,* **305**, 451-477.

Battarbee R.W., Anderson N.J., Jeppesen E., & Leavitt P.R. (2005). Combining paleolimnological and limnological approaches in assessing lake ecosystem response to nutrient reduction. *Freshwater Biology,* **50**, 1772-1780.

Beaulieu M., Pick F., & Gregory-Eaves I. (2013). Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. *Limnology and Oceanography,* **58**, 1736-1746.

Bennett J.R., Cumming B.F., Ginn B.K., & Smol J.P. (2010). Broad-scale environmental response and niche conservation in lacustrine diatom communities. *Global Ecology and Biogeography*, **19**, 724-732.

Birks H.H., Jones V.J., Brooks S.J., Birks H.J.B., Telford R.J.S., & Peglar S.M. (2012). From cold to cool in northernmost Norway: Lateglacial and early Holocene multi-proxy environmental and climate reconstructions from Jansvatnet, Hammerfest. *Quaternary Science Reviews,* **33**, 100-120.

Blois J.L., Williams J.W., Fitzpatrick M.C., Jackson S.T., & Ferrier S. (2013). Space can substitute for time in predicting climate change effects on biodiversity. *Proceedings of the National Academy of Sciences,* **110**, 9374-9379.

Brothers S., Vermaire J.C., & Gregory-Eaves I. (2008). Empirical models for describing recent sedimentation rates in lakes distributed across broad spatial scales. *Journal of Paleolimnology*, **40**, 1003-1019.

Charles D.F., Knowles C., & Davis R.S. (2003). Protocols for the analysis of algal samples as collected as part of the U.S. Geological Survey National Water-Quality Assessment program. Patrick Center for Environmental Research, The Academy of Natural Sciences. Report No. 02-06.

Cottenie K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175-1182.

Dixit S.S., Smol J.P., Charles C.F., Hughes R.M., Paulsen S.G., & Collins G.B. (1999). Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Canadian Journal of Fisheries and Aquatic Sciences,* **56**, 131-152.

Dray S., Legendre P., Oksanen J., Blanchet F.G., Solymos P., Thioulouse J., Simpson G., & Jombart T. (2013). Spatial ecological data analysis with R. sedaR project, R-Forge: rforge.r-project.org/projects/sedar

Flessa K.W., & Jackson S.T. (2005). Forging a Common Agenda for Ecology and Paleoecology. *BioScience*, **55**, 1030-1031.

Ginn B.K., Cumming B.F., & Smol J.P. (2007). Diatom-based environmental inferences and model comparisons from 494 northeastern North-American Lakes. *Journal of Phycology*, 43, 647-661.

Glew J., (1989). A new trigger mechanism for sediment samplers. *Journal of Paleolimnology*, 2, 241–243.

Gottschalk S., & Kahlert M. (2012). Shifts in taxonomical and guild composition of littoral diatom assemblages along environmental gradients. *Hydrobiologia*, **694**, 41-56.

Gregory-Eaves I., & Beisner B.E. (2011). Paleolimnological insights for biodiversity science: an emerging field. *Freshwater Biology*, **56**, 2653-2661.

Hájek M., Roleček J., Cottenie, Kintrová K., Horsák M., Poulíčková A., Hájková P., Fránková M., & Dítě D. (2011). Environmental and spatial controls of biotic assemblages in a discrete semi-terrestrial habitat: comparison of organisms with different dispersal abilities sampled in the same plots. *Journal of Biogeography*, **38**, 1683-1693.

Hampton S.E., Strasser C.A., Tewksbury J.J., Gram W.K., Budden A.E., Batcheller A.L., Duke C.S., & Porter J.H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment,* **11**, 156-162.

Hausmann S., & Pienitz R. (2009). Seasonal water chemistry and diatom changes in six boreal lakes of the Laurentian Mountains (Québec, Canada): impacts of climate and timber harvesting. *Hydrobiologia*, **635**, 1-14.

Hawryshyn J., Rühland K.M., Julius M., & Smol J.P. (2012). Absence of evidence is not evidence of absence: Is *Stephanodiscus binderanus* (Bacillariophyceae) an exotic species in the Great Lakes region? *Journal of Phycology*, **48**, 270-274.

Heino J. (2009). Species co-occurrence, nestedness and guild-environment relationships in stream macroinvertebrates. *Freshwater Biology*, **54**, 1947-1959.

Husson F., Josse J., Le S., & Mazet J. (2014). Package 'FactoMineR': http://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf.

Komsta L., & Novomestky F. (2013). Package 'moments': www.komsta.net

Köster D., & Pienitz R. (2006). Seasonal diatom variability and paleolimnological inferences – a case study. *Journal of Paleolimnology,* **35**, 395-416.

Kurek J., Weeber R.C., & Smol J.P. (2011). Environment trumps predation and spatial factors in structuring cladoceran communities from Boreal Shield lakes. *Canadian Journal of Fisheries and Aquatic Sciences,* **68**, 1408-1419.

Legendre P, & Gallagher E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271-280.

Legendre P., & De Cáceres M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951-963.

Legendre P., & Legendre L. (2012). Numerical Ecology, 3rd edition. Elsevier Academic Press.

Leira M., Chen G., Dalton C., Irvine K., & Taylor D. (2009). Patterns in freshwater diatom taxonomic distinctness along an eutrophication gradient. *Freshwater Biology*, **54**, 1-14.

Levi E.E., Çakiroğlu A.İ., Bucak T., Odgaard B.V., Davidson T.A., Jeppesen E., & Beklioğlu M. (2014). Similarity between contemporary vegetation and plant remains in the surface sediment in Mediterranean lakes. *Freshwater Biology*, **59**, 724-736.

Nykänen M., Vakkilainen K., Liukkonen M., & Kairesalo T. (2009). Cladoceran remains in lake sediments: a comparison between plankton counts and sediment records. *Journal of Paleolimnology,* **42**, 551-570.

Oksanen J, Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L, Solymos P., Stevens H.H, & Wagner H. (2013). Package 'vegan': cran.r-project.org/web/packages/vegan/vegan.pdf.

Philibert A., & Prairie Y.T. (2002). Diatom-based transfer functions for western Quebec lakes (Abitibi and Haute Mauricie): the possible role of epilimnetic $CO_2$ concentration in influencing diatom assemblages. *Journal of Paleolimnology*, **27**, 465-480.

Picazo F., Millán A., & Dolédec S. (2012). Are patterns in the taxonomic, biological and ecological traits of water beetles congruent in Mediterranean ecosystems? *Freshwater Biology,* **57**, 2192-2210.

Ptacnik R, Andersen T., Brettum P., Lepistö L., & Willén E. (2010). Regional species pools control community saturation in lake phytoplankton. *Proceedings of the Royal Society B*, **277**, 3755-3764.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org/

Ribeiro P.J., & Diggle P.J. (2013). Package 'geoR': http://cran.r-project.org/web/packages/geoR/geoR.pdf.

Saros J.E., & Anderson N.J. (2014). The ecology of the planktonic diatom *Cyclotella* and its implications for global environmental change studies. *Biological Reiews,* doi: 10.1111/brv.12120.

Seddon A.W.R., Mackay A.W., Baker A.G. *et al.* (2014). Looking forward through the past: identification of 50 priority research questions in paleoecology. *Journal of Ecology*, **102**, 256-267.

Smol J.P. (2008). Pollution of Lakes and Rivers: A Paleoenvironmental Perspective, 2nd ed. Wiley-Blackwell.

Soininen J., & Weckström J. (2009). Diatom community structure along environmental and spatial gradients in lakes and streams. *Fundamental and Applied Limnology,* **174**, 205-213.

Strayer, D.L., & Dudgeon D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, **29**, 344-358.

US EPA. (2007). Survey of the Nation's Lakes. Field Operations Manual. EPA 841-B-07-004. U.S. Environmental Protection Agency, Washington, DC, USA.

US EPA. (2009). National Lakes Assessment: A Collaborative Survey of the Nation's Lakes. EPA 841-R-09-001. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC, USA.

US EPA. (2011-2012). National Lakes Assessment. Field Operations Manual. EPA 841-B-11-003. U.S. Environmental Protection Agency, Washington, DC, USA.

Valois A.E., Keller W.B., & Ramcharan C.W. (2011). Recovery in a multiple stressor environment: using the reference condition approach to examine zooplankton community change along opposing gradients. *Journal of Plankton Research*, **33**, 1417-1429.

Vanormelingen P., Verleyen E., & Vyverman W. (2008). The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, **17**, 393-405.

Velghe K., & Gregory-Eaves I. (2013). Body size is a significant predictor of congruency in species richness patterns: A meta-analysis. *PLoS ONE*, **8**, e57019.

Verleyen E., Vyverman W., Sterken M., Hodgson D.A., De Wever A., Juggins S., Van de Vijver B., Jones V.J., Vanormelingen P., Roberts D., Flower R., Kilroy C., Souffreau C., & Sabbe K. (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos,* **118**, 1239-1249.

Vyverman W., Verleyen E., Sabbe K., Vanhoutte K., Sterken M., Hodgson D.A., *et al.* (2007). Historical processes constrain patterns in global diatom diversity. *Ecology,* **88**, 1924-1931.

Vanormelingen P., & De Wever A. (2007) Historical processes constrain patterns in global diatom diversity. *Ecology*, **88**, 1924-1931.

Wehr J.D., & Sheath R.G. (2003) Freshwater Algae of North America: Ecology and Classification. Academic Press, Elsevier Science, New York, USA.

Wigington P.J. Jr, Davies T.D., Tranter M., & Eshleman K.N. (1992) Comparison of episodic acidification in Canada, Europe and the United States. *Environmental Pollution,* **78**, 29-35.

**Tables**

**Table 1: Mean, median, range and standard deviation of (non-transformed) environmental variables measured from the integrated water-column sample (*n* = 468).** All variables included in selection procedures are included here.

| Variable | Mean | Median | Range | Standard deviation |
|---|---|---|---|---|
| **pH** | 8.1 | 8.2 | 4.7 – 10.3 | 0.8 |
| **Conductivity (µS $cm^{-1}$)** | 470 | 255.4 | 12.9 – 9751 | 961.8 |
| **Turbidity (NTU)** | 13.2 | 3.4 | 0.3 – 312 | 31.7 |
| **DOC (mg $L^{-1}$)** | 9.2 | 5.3 | 0.3 – 290.6 | 20.9 |
| **$NH_4$ (µeq $L^{-1}$)** | 2.9 | 1.3 | 0.3 – 122.0 | 8.3 |
| **$NO_3$ + $NO_2$ (mg N $L^{-1}$)** | 0.09 | 0.005 | 0 – 5.6 | 0.4 |
| **Total Nitrogen (µg $L^{-1}$)** | 1185.1 | 543.5 | 70 – 26100 | 2481.1 |
| **Total Phosphorus (µg $L^{-1}$)** | 107.9 | 24.5 | 1 – 2147 | 259.2 |
| **Cl (µeq $L^{-1}$)** | 740.2 | 219.7 | 1.5 – 22890.4 | 1977.7 |
| **$SO_4$ (µeq $L^{-1}$)** | 1840 | 203.5 | 2.5 – 133210.8 | 8414.9 |
| **Ca (µeq $L^{-1}$)** | 1341.7 | 1201.0 | 61.0 – 17095.7 | 1357.3 |
| **Mg (µeq $L^{-1}$)** | 1425.1 | 554.1 | 16.1 – 60703.9 | 3842.2 |
| **Colour (PCU)** | 16.6 | 11.0 | 0 – 93 | 15.5 |
| **$SiO_2$ (mg $L^{-1}$ $SiO_2$)** | 8.6 | 5.4 | 0.03 – 91.9 | 10.6 |
| **$H^+$ (µeq $L^{-1}$)** | 0.07 | 0.006 | 0 – 15.1 | 0.7 |
| **$OH^-$ (µeq $L^{-1}$)** | 3.2 | 1.7 | 0.001 – 123.0 | 8.2 |
| **Ion balance (ANC %)** | -0.9 | -1.2 | -13.7 – 20.3 | 2.8 |
| **Chl *a* (µg $L^{-1}$)** | 27.9 | 7.4 | 0.1 – 871.2 | 73.9 |
| **Secchi depth (m)** | 2.06 | 1.6 | 0.05 – 12.5 | 1.9 |
| **$Z_{max}$ (m)** | 9.9 | 6.5 | 0.5 – 60.3 | 10.3 |

**Table 2: RV coefficients from comparisons of diatom assemblage matrices and lake positions within RDAs.** "WC" refers to water-column samples and "SSed" to surface sediment samples. The "Ordination overlay" column lists the two matrices, of which their structure is compared *symmetrically* using an RV coefficient. The column for the RV coefficient of the full set of fitted scores refers to the comparison of scores from all of the axes within an ordination versus the RV coefficient of the 1st axis scores, which is the correlation between the main axis of variation in one matrix and the main axis of variation in another (synonymous with Ordinary Least Squares Regression).

| Ordination overlay | RV coefficient of 1st axis of fitted scores (*P*-value) | RV coefficient of (full set) of fitted scores (*P*-value) |
|---|---|---|
| **Matrix A:** Site scores from WC diatom assemblage PCA<br>**Matrix B:** Site scores from SSed diatom assemblage PCA<br>(all species) | 0.23 ($P<0.001$) | 0.63 ($P<0.001$) |
| **Matrix A:** Site scores from WC diatom assemblage PCA<br>**Matrix B:** Site scores SSed diatom assemblage PCA<br>(planktonic only) | 0.24 ($P<0.001$) | 0.23 ($P<0.001$) |
| **Matrix A:** Site scores from WC environmental RDA<br>**Matrix B:** Site scores from SSed environmental RDA<br>(all species) | 0.54 ($P<0.001$) | 0.17 ($P<0.001$) |
| **Matrix A:** Site scores from WC spatial RDA<br>**Matrix B:** Site scores from SSed spatial RDA<br>(all species) | 0.54 ($P<0.001$) | 0.16 ($P<0.001$) |
| **Matrix A:** Site scores from WC environmental RDA<br>**Matrix B:** Site scores from SSed environmental RDA<br>(planktonic only) | 0.50 ($P<0.001$) | 0.22 ($P<0.001$) |

| | | |
|---|---|---|
| **Matrix A:** Site scores from WC spatial RDA<br>**Matrix B:** Site scores from SSed spatial RDA<br>(planktonic) | 0.53 ($P$<0.001) | 0.10 ($P$<0.001) |
| **Matrix A:** Site scores from WC environmental RDA (visit 1, 51 sites)<br>**Matrix B:** Site scores from SSed environmental RDA (51 sites)<br>(all species) | 0.03 ($P$ = 0.3) | 0.38 ($P$<0.001) |
| **Matrix A:** Site scores from WC environmental RDA (visit 2, 51 sites)<br>**Matrix B:** Site scores from SSed environmental RDA (51 sites)<br>(all species) | 0.25 ($P$ = 0.0002) | 0.07 ($P$<0.001) |
| **Matrix A:** Site scores from WC environmental RDA (mean visits, 51 sites)<br>**Matrix B:** Site scores from SSed environmental RDA (51 sites)<br>(all species) | 0.47 ($P$<0.001) | 0.5 ($P$<0.001) |

**Figures**

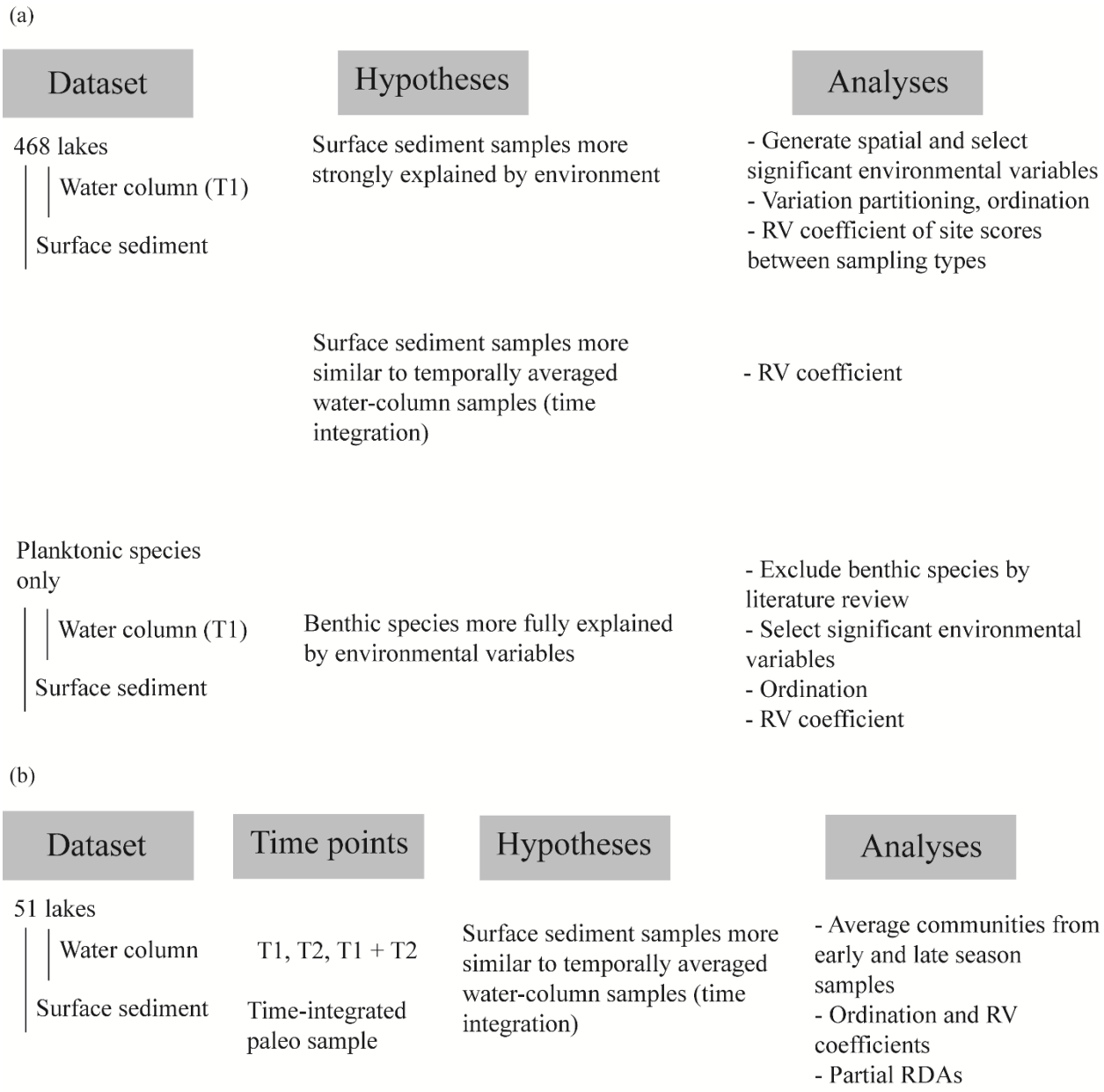



**Figure 1: Visual overview of statistical analyses and associated hypotheses.**

**Overview of data set-up, hypotheses and associated statistical analyses for (a) analyses encompassing the entire lake dataset (*n* = 468) and (b) a subset of 51 lakes with the inclusion of a second water-column sampling visit.** For both (a) and (b), samples were paired so that each lake was represented by a water-column sample derived from an integrated water

sample and a surface sediment sample derived from the top 1 cm of a sediment core. “T1” and “T2” refer to sampling time points included in each analysis, with “T1” occurring in June or early July 2007 and “T2” in August 2007. All analyses were conducted with both species- and genus-level community data.

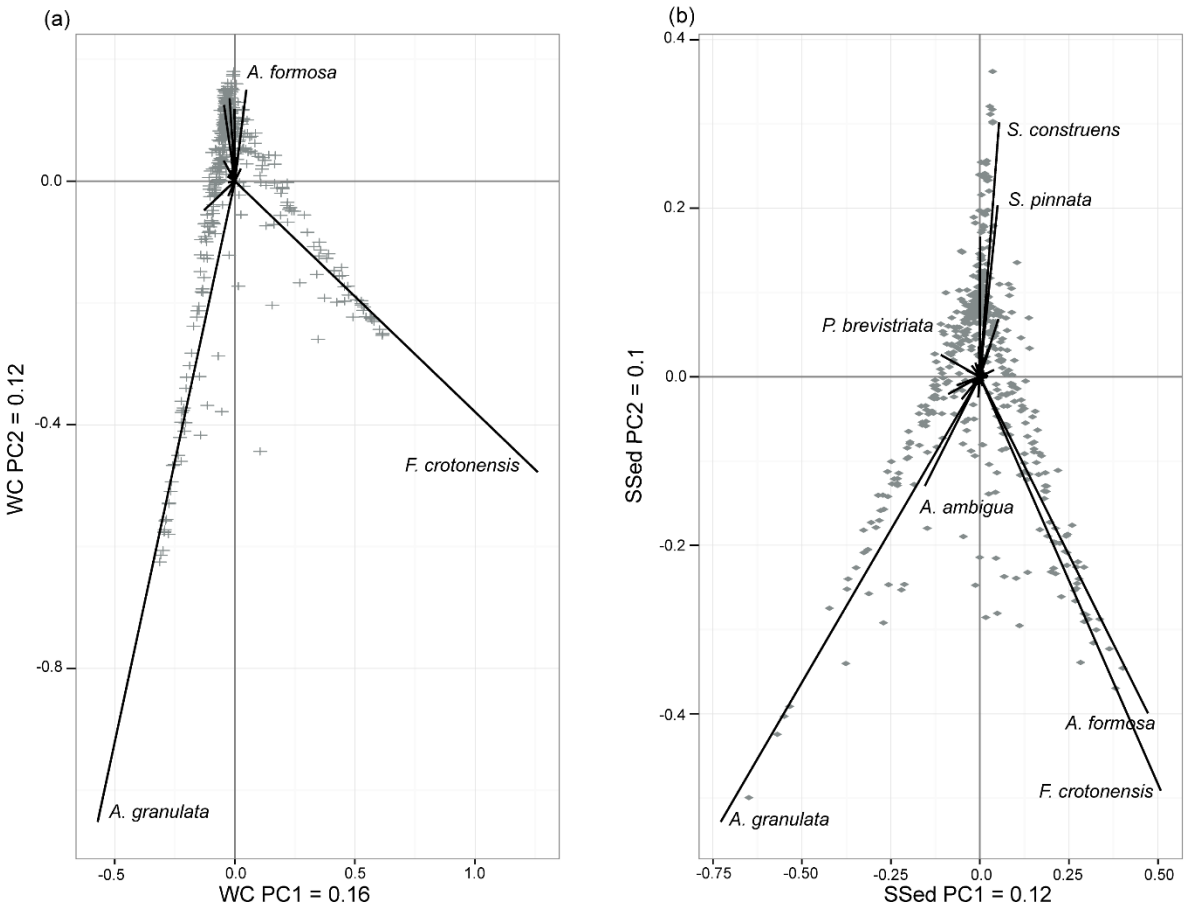



**Figure 2: PCA biplots of (a) water-column diatom species and (b) surface sediment diatoms.** Species shown in the PCA biplots are those with vectors greater than or approaching 0.2 units: *A. formosa* (*Asterionella formosa*), *A. ambigua* (*Aulacoseira ambigua*), *A. granulata* (*Aulacoseira granulata*), *F. crotonensis* (*Fragilaria crotonensis*), *S. contruens* (*Staurosira construens*), and *S. pinnata* (*Staurosirella pinnata*). Water column samples are represented by "+" symbols, while the surface sediment samples represented by filled shapes.

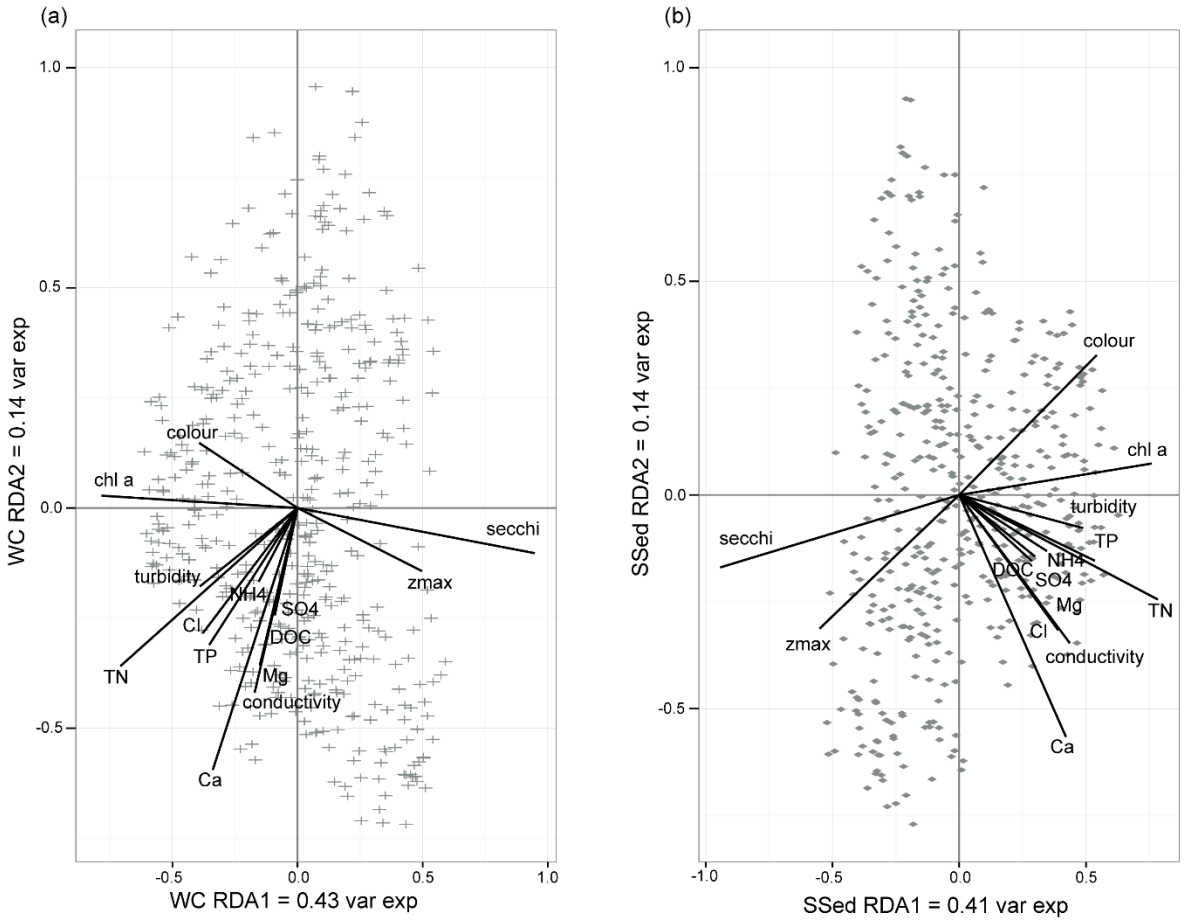



**Figure 3: Biplots of the first two axes from redundancy analysis (RDA) of environmental variables using diatom species from the 468 study lakes: (a) using water-column diatoms, and (b) using surface sediment diatoms.** Water column samples are represented by "+" symbols, while the surface sediment samples represented by filled shapes. The environmental variables depicted by the arrows were selected using forward selection. Variables left untransformed were: $Z_{max}$ (i.e. maximum observed depth), conductivity, Ca, Mg, $SO_4$, $NH_4$, TP, turbidity, DOC and colour (PCU). Box-Cox transformed variables were TN, Chl *a* and Secchi.

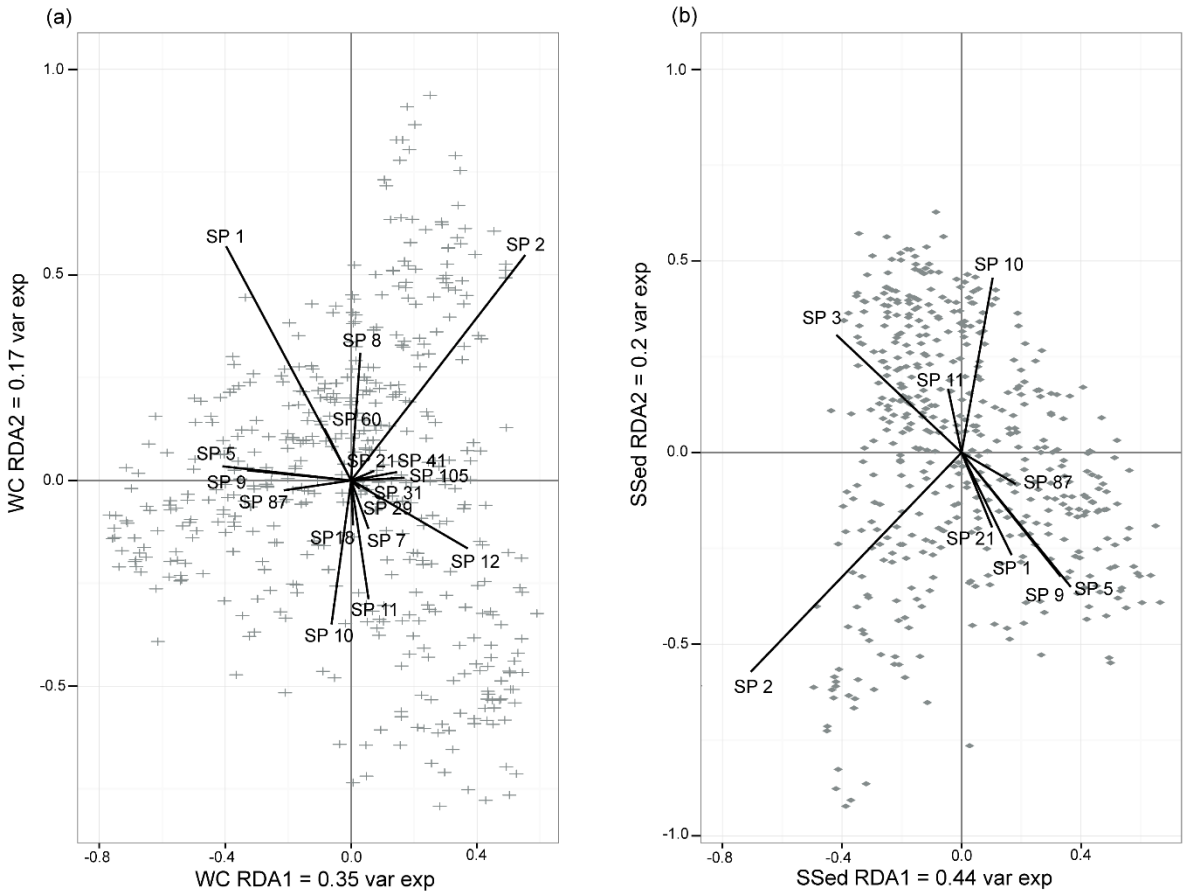



**Figure 4: Biplots of (a) the redundancy analysis (RDA) of 17 spatial PCNM predictors from the 468 study lakes, using the water column diatom data and (b) the 9 spatial PCNM predictors from the 468 study lakes, using the surface sediment diatom data.** The spatial variables consisted of PCNM predictors ("SP") selected using forward selection from 105 PCNM predictors.

## CONNECTING STATEMENT 1

In the Introduction of this thesis, I identified the importance of characterizing overall trends in biodiversity in the Anthropocene (McGill *et al*., 2015). Meta-analyses such as Vellend *et al*., (2013) and Dornelas *et al*., (2014) have emphasized no net change in species alpha diversity through time, while at the same time showing considerable beta diversity. However, many of the studies included in these meta-analyses did not exceed 50 years in duration, resulting in relatively limited temporal scales of study. As an example, only 11% of the studies (19 of 168) included in the Vellend *et al*., (2013) meta-analysis were greater than 50 years in length (maximum length = 261 years). For Dornelas *et al*., (2014), 4% of the studies had more than 50 years of data (4 of 100 studies, maximum length = 129 years). While these meta-analyses are very comprehensive, the relatively 'short' duration of the studies included in them preclude the investigation of an ancient 'biodiversity cliff' (*sensu* McGill *et al*., 2015) or significant losses or changes in beta diversity during time periods more than 50 years in the past. Because of the focus in this thesis on freshwater biodiversity, I emphasize the use of paleolimnological approaches to extend the temporal scale of our biodiversity investigations.

When comparing the temporal scales at which contemporary sampling, monitoring and paleolimnological studies operate at, one can make a few observations (see Fig. C1). First, there are some differences in the temporal scales that are accessible by these different methods. For example, paleolimnology is not useful in situations where one requires temporal resolution to be at the hourly, daily, weekly, monthly or even in some cases, yearly scale. Second, the temporal scale that monitoring studies accurately cover takes up a small proportion of the geological time that one might be interested in, in terms of describing biodiversity or environmental change. However, there is also some overlap in the temporal scales that these two approaches can cover. For example, depending on the location/study - a decadal temporal scale may be accessible using

both long-term monitoring records *and* paleolimnological approaches and thus one could answer questions like "is one tracking different things when using these different approaches?" Attempting to resolve these types of questions was the focus of Chapter 1. I found that, in the case of freshwater diatoms, that environmental and spatial drivers of community variation were congruent between surface sediments and water-column samples collected via the 2007 U.S. EPA National Lakes Assessment (NLA). Although there is the potential for each surface sediment sample to represent a different degree of temporal integration, depending on the lake, there was significant congruence across sample matrices that varied in this temporal integration. This provides some strong support for working more closely with these same surface sediment samples in Chapter 2.

In Chapter 2, I continued work with the diatom data from the 2007 NLA, focusing this time on another way to extend the examination of biodiversity in the Anthropocene, with the use of different biodiversity metrics. In particular, I was interested in comparing alpha and gamma diversity metrics with beta diversity results given the finding by Vellend *et al*., (2013) and Dornelas *et al*., (2014) who recorded no net change in alpha diversity, and significant amounts of beta diversity. As such, in Chapter 2, I explore beta diversity patterns of freshwater diatoms across the conterminous United States, examining the extent to which including additional metrics of biodiversity (beyond alpha diversity) can assist in illuminating drivers of freshwater biodiversity. Chapter 1 takes a very quantitative approach to describing diatom communities enumerated from surface sediments. Chapter 2 continues with these quantitative methods, and focuses on reconstructions of ecological dynamics as opposed to the environment, a gap identified in Rull (2014).

## Literature Cited

Dornelas M., Gotelli N.J., McGill B., Shimadzu H., Moyes F., Sievers C., & Magurran, A.E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science,* **344**, 296-299.

McGill B.J., Dornelas M., Gotelli N.J., & Magurran A.E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology and Evolution*, **30**, 104-113.

Pollard A. (Personal communication).

Rull V. (2014). Time continuum and true long-term ecology: from theory to practice. *Frontiers in Ecology and Evolution,* **2**, 1-7.

Smol J.P. (2008). Pollution of Lakes and Rivers: A Paleoenvironmental Perspective 2nd edition. Blackwell Publishing, Massachusetts, USA

Vellend M., Baeten L., Myers-Smith I.H., Elmendorf S.C., Beauséjour R., Brown C.D., De Frenne P., Verheyen K., & Wipf S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*, **110**, 19456-19459.

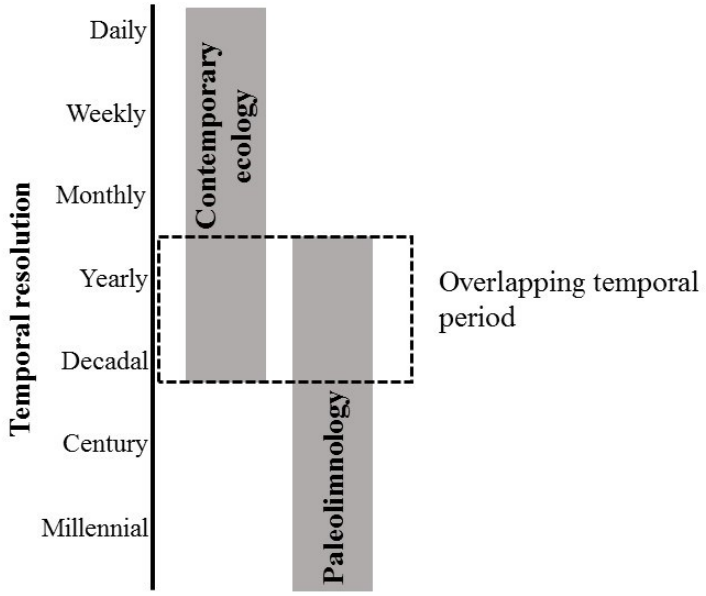



**Figure C1: Time scales of study for both paleolimnological and monitoring studies.** Adapted from Rull (2014) and Smol (2008).

# CHAPTER 2

## Diatom diversity patterns over the past ~ 150 years across the conterminous United States: identifying mechanisms behind beta diversity

Amanda K. Winegardner[1,2], Beatrix E. Beisner[2,3], P. Legendre[2,4], and Irene Gregory-Eaves[1,2]

[1]Deptment of Biology, McGill University, Montréal, Québec, Canada
[2]Québec Centre for Biodiversity Science, Montréal, Québec, Canada
[3]Department of Biological Sciences, University of Québec at Montréal, Montréal, Québec, Canada
[4]Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

**Abstract**

Understanding the magnitude and drivers of freshwater diversity change over the last 150 years provides essential insights for developing scenarios of future change. Here we quantify and identify drivers of spatial and temporal beta diversity in diatom assemblages between historical and modern times. Using sedimentary genus-level diatom data from 176 lakes, and species-level data for 59 lakes, we computed spatial beta diversity across all lakes and within ecoregions for 2007 and pre-1850 CE time points. We also computed local contributions to beta diversity (LCBD) and analysed them with respect to environmental variables. Total beta diversity was partitioned into replacement and abundance difference components to identify mechanisms possibly responsible for spatial beta at each time point. Temporal Beta diversity Indices (TBI) were also computed for each lake by comparing the diatom data of all lakes at the two time points. TBIs were decomposed into taxon losses and gains to facilitate interpretation. TBIs and their components were related to contemporary land cover variables. Temporal beta diversity varied significantly with forest cover and longitude, with higher values in western regions. Spatial beta diversity was similar between the historical and 2007 time points, with genus replacement explaining almost all variation. There was no systematic pattern in lakes that contributed to spatial beta diversity; however local contributions were explained by a specific structure of water quality and land cover variables. Spatial beta diversity of diatoms across the US does not appear to have changed between pre-1850 CE and 2007, suggesting that broad-scale land use and hydrological alteration of the landscape has not homogenized these communities. Temporal beta diversity occurred through genus gains and losses and was significantly related to forest and agriculture cover in watersheds, with genus replacement dominating beta diversity at

both time points. These analyses, pairing spatial and temporal beta diversity, provide insight into the mechanisms behind diatom diversity.

## Introduction

Modern geological time is commonly referred to as the Anthropocene, a designation recognizing the extent to which humans dominate processes and life on the Earth (Crutzen, 2002). Although there is still some debate over its exact onset date (Lewis & Maslin, 2015; responses, Hamilton, 2015; Zalasiewicz *et al*. 2015), the Anthropocene concept has been quickly taken up by the scientific community and the general public. Thus, much recent biodiversity research has sought to understand the distribution and drivers of biodiversity loss concomitant with habitat losses documented over the Anthropocene (Steffen *et al*., 2011), which are expected to compromise ecosystem functioning and services (Cardinale *et al*., 2012). However, several studies, including recent meta-analyses, have shown that temporal species richness trends (local or regional) may be flat (e.g. Vellend *et al*., 2013; Dornelas *et al*., 2014) or show only modest losses (Newbold *et al*., 2015) over the last few centuries. Nonetheless, it is more generally accepted that there have been substantial shifts in community composition (Vellend *et al*., 2013; Dornelas *et al*., 2014; Newbold *et al*., 2015) implicating both species losses *and* gains in biodiversity and associated ecosystem shifts (Wardle *et al*., 2011; McGill *et al*., 2015). Advancing our understanding in this area is particularly important in freshwater ecosystems, which hold a disproportionately concentrated diversity relative to other environments (Strayer & Dudgeon, 2010).

Recent literature has highlighted the importance of different metrics of biodiversity and scales of study. Many biodiversity studies have focused on species richness in a site (alpha diversity), or across a region or several sites (gamma diversity). Beta diversity focuses on the differentiation of communities among sites or through time in terms of number of species and composition (Whittaker, 1972). Spatial beta diversity is particularly interesting because it can

identify sites and regions that are exceptional across a landscape owing to degraded or enhanced diversity. Temporal beta diversity can also indicate degrees of change in composition through time at single locations.

Despite the usefulness of considering numerous biodiversity metrics at multiple scales, very few studies have attempted to consider more than a few trends simultaneously (McGill *et al*., 2015). The majority of biodiversity change studies have also been largely restricted to the last 50 years, with only a handful extending beyond this time frame to encompass the entire Anthropocene. This literature has been recently synthesized to develop predictions of various metric-scale combination trends during the Anthropocene (McGill *et al*., 2015, Table 1), together with a call to test these predictions.

Aquatic systems provide a unique opportunity to quantify several biodiversity trends over different spatial and temporal scales through the use of the historical archives found in lake sediments (Gregory-Eaves & Beisner, 2011). We analysed a large paleolimnological dataset to quantify diversity trends across both space and time and explore mechanisms behind diatom assemblage variation across the conterminous US during the Anthropocene. We used sediment cores collected as part of the 2007 National Lakes Assessment (NLA), United States Environmental Protection Agency (U.S. EPA; http://water.epa.gov/type/lakes/lakessurvey_index.cfm) to examine patterns of diatom genera and species richness, as well as alpha, beta and gamma diversities for 176 lakes. To identify beta diversity hot spots and relate these to water quality and land cover, we computed spatial beta diversity across all lakes at both the historical (pre-1850 CE) and modern (2007) time points. We also calculated temporal beta diversity within each lake to identify compositional variation between the historical and modern time points (Fig. 1), asking:

1) What is the relative magnitude of spatial beta diversity across different ecoregions at each time point?
2) What is the magnitude of temporal beta diversity for each lake?
3) Which lakes contribute the most to spatial beta diversity? Which lakes show significant change through time?
4) Is spatial beta diversity explained mostly by replacement of genera/species or differences in richness and abundances of genera/species across space?
5) What are the environmental drivers of spatial beta diversity?
6) Is the change through time explained by loss or gain of taxa abundances?
7) What are the environmental drivers of temporal beta diversity?

We developed hypotheses based on the trends identified in McGill *et al*., (2015), adapting these predictions to the three scales used in this study: continental US, ecoregion and individual lakes (Table 1).

**Methods**

*Description of the 2007 National Lakes Assessment (NLA)*

The 2007 NLA survey took place from May to October 2007, sampling over 1000 lakes and reservoirs from the conterminous U.S. (U.S. EPA, 2009). Lakes were all >10 acres and deeper than 1 m (U.S. EPA, 2009). The lakes were randomly selected using a combination of probabilistic design and specifically targeted 'reference' (pristine/undisturbed) lakes identified by state and tribal partners (U.S. EPA, 2011-2012). After sampling, lakes were re-classified as necessary to three categories representing disturbance levels: least disturbed, intermediate, and highly disturbed relative to other lakes in the same ecoregion (Herlihy, Pers. Comm.). The 2007

NLA data includes water quality measurements, land use metrics, and compositional data for zooplankton and phytoplankton, including diatom assemblages from both water column and lake sediment samples. The data are publicly available from: http://water.epa.gov/type/lakes/NLA_data.cfm. Further details of the sampling are summarized in Beaulieu *et al*., (2013) and Winegardner *et al*., (2015).

*Sediment core screening*

NLA field teams collected sediment cores from a subset of the sampled lakes, mainly where sampling teams estimated that the sediment bank at the deepest point of the lake would be undisturbed, such that the bottom part of the core would reach sediments representing pre-1850 CE (pre-industrial) conditions. Cores were collected from the deepest point of the lake using a modified Kajak-Brinkhurst corer (Glew, 1989). Both the top and bottom 1 cm interval of the sediment was saved for diatom enumeration, with up to 500 diatom valves counted using standardized methods from the U.S. Geological Survey National Water Quality Assessment (Charles *et al*., 2003; U.S. EPA, 2011-2012). Thus, for each lake where sediment coring occurred, there are diatom assemblage data for both modern (2007) conditions as well as bottom of the core (historical) conditions. To account for different core lengths and variation in sedimentation across lakes, we used screening criteria to select lakes for our study, ensuring that the bottom sediment samples used represented historical (pre-1850 CE) conditions. We used a three-fold approach to estimate age of sediment cores (described in S1) and identified 179 sites for which we were confident that their bottom samples represented pre-1850 CE conditions. Three Coastal Plains lakes that were very isolated by a few 1000s of kilometres from others in the same ecoregion were removed. The final set of lakes used in this study (n = 176) is shown in Figure 2a, along with ranges of key limnological variables in Figure 2b-e. Note that there was a

geographic bias with bottom core samples were not collected from many lakes from the mid- and southern portions of the country where reservoirs dominate.

*Diversity analyses*

All statistical analyses were completed in R v. 3.1.2 (R Core Team, 2014). Genus-level was the taxonomic level associated with a higher level of consistency between lab groups enumerating the diatom data for the 2007 NLA (Pollard, Pers. Comm.). As such, we conducted the analyses on the full set of lakes (n = 176) with the diatom classified to genus-level. For a smaller set of the lakes (n = 59), we completed all the analyses again at the species level data because for this smaller set, all samples were enumerated by a single lab group (The Academy of Natural Sciences). We calculated diatom species or genus richness in three different ways: total richness for all taxa, rarefied richness for all taxa, and rarefied richness for dominant taxa only. "Dominant" species or genera were those having greater than 2% relative abundance in at least one sample from either surface or historical sediments. We calculated rarefied taxa richness using the *rarefy()* function in `vegan` (Oksanen *et al*., 2015), correcting for the total number of valves counted for each sample, setting all samples to a cut-off of 300 valves (150 individuals). Thus, the single sample with ~300 valves could be compared to samples with a higher total abundance (the majority had >500 valves counted). All subsequent analyses were done using only dominant (>2% relative abundance) diatom taxa. We calculated alpha diversity using a Shannon index and Simpson's index (evenness) using *diversity()* in `vegan` and further transformed Shannon-Weiner diversity (*H*) to an effective diversity number (exp(*H*)) as suggested by Ellison (2010).

We used four source functions developed by P. Legendre: *beta.div()* (Legendre & De Cáceres, 2013, App. S4; Legendre, 2013a), *beta.div.comp()* (Legendre, 2014, App. S3),

*LCBD.comp()* (Legendre, 2013b; Legendre, 2014, App. S5), and *TBI()* (Legendre, 2015) for use in R. We used *beta.div()* to compute spatial beta diversity for all sites (either n = 176 or n = 59) across the landscape at both the historical and contemporary time points, using non-transformed matrices of diatom abundances and the percentage difference method (Legendre & De Cáceres, 2013). We also computed spatial beta diversity at the two different time points within six ecoregions, as defined by Herlihy *et al*., (2008): Coastal Plains, Northern Appalachians, Southern Plains, Temperate Plains, Upper Midwest, and Western Mountains (Xeric was excluded from spatial beta diversity computations because n = 1) (Fig. 2a, n = 176). The function *beta.div* () also provided a 'Local Contribution to Beta Diversity' (LCBD index) for each lake as well as a permutational p-value indicating, the significance of each LCBD value. LCBD values provide a metric to assess the individual importance of each lake on total beta diversity (Legendre & De Cáceres, 2013). We identified the sites with significant LCBD that were the same or different between the historical and 2007 sediments and performed a chi-square test of the null hypothesis that processes producing significant LCBD were independent at the two time points. To assess individual species' influences on beta diversity, we computed 'Species Contributions to Beta Diversity' (SCBD indices) using a Hellinger distance measure (Legendre & De Cáceres, 2013). SCBD allowed us to identify species that were important contributors to spatial beta diversity, both historically and in modern samples.

We were interested in the mechanisms generating spatial beta diversity at both time points. Using *beta.div.comp()* and the percentage difference index (which is the quantitative form of the Sørensen coefficient), we partitioned total beta diversity into replacement and percentage difference (Podani decomposition) components such that, in addition to a mean total beta diversity value for the landscape and each individual ecoregion, we also knew, for each time

point, the proportions of the total (spatial) beta diversity that are explained by replacement and percentage difference. We then computed LCBD indices from the replacement and abundance difference matrices using function *LCBD.comp()*.

To explain the variation in the extent to which lakes contribute to spatial beta diversity (e.g. the magnitude of LCBD indices), we employed a univariate regression tree (URT) approach. URTs split a response variable (here the vector of LCBD indices) along gradients in explanatory variables, creating groups or partitions with similar values of the response (De'ath, 2002). We computed regression trees using two different sets of variables. First, we built a regression tree with each lake's LCBD values of the 2007 data as a response variable, and the following explanatory variables (Appendix S3a for PCA), recorded for each lake in 2007: latitude, longitude, ecoregion, mean Secchi depth (m), chlorophyll a concentration (μg $L^{-1}$), total phosphorus (TP; μg $L^{-1}$), total nitrogen (TN; μg $L^{-1}$), mean temperature (°C), specific conductivity (μS $cm^{-1}$ at 25°C) and pH. Second, we built a tree using basin-level land-use variables (summarized by the NLA and based on the 1992 National Land Cover Database; http://www.mrlc.gov/nlcd1992.php). These variables included percent land cover in each lake basin for the following land cover types: developed (split into low intensity residential, medium intensity residential, high intensity residential, open space), barren, forest (split into deciduous forest, coniferous forest, mixed forest), grasslands, agriculture (split into pasture and row crops), and wetland (split into woody wetlands and emergent herbaceous wetlands) (USGS, 1992). We used functions *rpart()* and *prune()* from the `rpart` package (Therneau *et al*., 2015). We then performed these same URTs with transformed Shannon diversity (exp(*H*)) for the 2007 sediments as a response variable, to determine whether alpha diversity showed similar patterns in the URTs as those with LCBD indices. Pruning of the trees was done using the lowest cross-

validation relative error (CVRE) for the set of variables, meaning that a higher number of variables would be retained than if we had selected a tree with fewer splits whose cross-validation relative error value is within one standard error of the smallest CVRE value.

We then analysed temporal beta diversity between the two time points for each individual lake. We used function *TBI(),* with the percentage difference option, to compare the surveys of all lakes at the two time points. This function computes Temporal Beta diversity Indices (TBI) and tests them for significance through a permutation test (permutations = 99999), identifying the lakes where temporal change is exceptionally large between the two surveys (Legendre, pers. comm.). Multiple testing is adjusted for using Holm's procedure, which is less severe than a standard Bonferroni correction (see Legendre & Legendre, 2012, p. 23). The TBI are somewhat similar to LCBD indices in spatial beta analysis, highlighting sites experiencing statistically significant community composition changes through time.

Analogous to analysing the relationship between spatial LCBD and land use variables, we examined the relationship between temporal beta diversity values (and their taxa gain and loss components) and contemporary land use to test the hypothesis that developed areas (see list of variables used in URT analyses) were associated with higher temporal beta diversity. We used logistic regression to test the relationship between significant temporal beta diversity and highly developed land cover. We also regressed the TBIs onto a composite land cover variable (Principal Component 1 scores of a Principal Components Analysis of the land cover variables, S3b), and created a regression tree (using the same approach as with spatial LCBD) with either TBI indices or taxa gain in temporal beta diversity as the response variable and land cover types as the explanatory variables.

**Results**

Results are presented for both the genus-level analyses (n = 176) and the species-level analyses (n = 59). For simplicity, tables and figures for the genus-level analyses are presented in the main body of the study, while all species-level tables and figures are shown in Appendices S4 and S5.

Genus-level results

*Spatial diversity results*

We found that the different diversity metrics gave considerably distinctive information about landscape level patterns in diatom assemblages. For example, the range of rarefied genus richness differed in some ecoregions between modern and historical sediments. In the Upper Midwest, rarefied richness was about 50% as large for modern as for historical sediments (Table 2). In contrast, alpha diversity, and gamma diversity were approximately equivalent across all ecoregions (Table 2).

Total spatial beta diversities across the landscape (*beta.div()*) were similar at both time points (Table 2), with taxa replacement consistently the dominant mechanism of compositional change among lakes (Table 3). Across all sites, spatial beta diversity using the percentage difference metric was 0.36 (72% of the maximum percent difference index value of 0.5) for both historical, and modern sediments. At the ecoregion level, spatial beta diversity also varied little at the two time points (Table 2). The genus replacement component dominated in importance for all ecoregions, for both historical and 2007 sediments (Table 3). It is important to note that in this case, spatial beta diversity is comparable amongst ecoregions, even with different sample

sizes, because spatial beta diversity was computed from a dissimilarity index that has an upper bound of 1 for the percentage diference index (Legendre & De Cáceres, 2013).

We observed considerable variation in LCBD values across the landscape, with some consistency between the two time points (Fig. 3), and some of the 2007 variation could be explained by environmental variables. Overall, 22 sites (13% of total) contributed significant LCBD indices in historical assemblages, while 18 sites (10%) contributed significant LCBDs in the modern assemblages; eight of these sites overlapped in the two time surveys. There was no clear geographic pattern in the distribution of higher relative LCBD values at either time point, or in the sites that had significant LCBD in common between the two time points. A McNemar test on a two by two table (number of sites significant in both time points, number of sites significant historically but not in 2007, number of sites not significant historically but significant in 2007 and number of sites not significant in both), resulted in $P = 0.5$, meaning that we cannot reject a null hypothesis of no effect of time period on significant LCBD values. A simple correlation between the historical and 2007 LCBD values showed a significant ($P < 0.05$) correlation of 0.38. Lakes that were exceptional in historical times (significant LCBD only at the historical time point) were classified as intermediate or highly disturbed based on the U.S. EPA's classification system (i.e. a system that is based on thresholds for TP, TN, chloride, $SO_4$, turbidity and agricultural/developed land cover (Herlihy, Pers. Comm.).

The univariate regression tree (URT) using water quality variables as predictors explained 54% of the variation in modern LCBD values, and only retained conductivity (Fig. 4a). When a similar analysis was run on the 2007 alpha diversity (both Shannon diversity ($H$) and exp($H$)), only water temperature was retained as a positive predictor in the model, which had $R^2_{adj}$ of 0.15 for $H$ and 0.14 for exp($H$). Using land cover measures instead of water quality

variables in a URT of the 2007 LCBD values resulted in a slightly weaker model ($R^2_{adj} = 0.3$; Figure 4b). With land cover variables, alpha diversity was found to be greater in sites with higher proportions of wetland and forest cover in the watersheds ($R^2_{adj} = 0.30$).

*Temporal diversity results*

Mean temporal beta diversity computed at the genus level ranged from 0.4 to 0.6 across the ecoregions (Table 4). The importance of genera loss and gain through time was approximately the same across ecoregions. Twenty-one of the 176 lakes showed significant temporal beta diversity (Fig. 3c; however, after correcting for multiple testing, we failed to identify any significant sites). Only two of the 21 lakes identified as sites of significant TBI were considered "Reference" (i.e. pristine or undisturbed) by US EPA prior to sampling; the others were assumed to have been disturbed and were selected using probabilistic methods. Based on the post-sampling classification of lakes into disturbance categories by the US EPA, most (15 of 21) of the lakes with significant temporal beta diversity were classified as having intermediate disturbance. Significant relationships were found between temporal beta diversity (used as a binary variable; significant or not) and the contribution of the genera loss and genera gain (Figure 5) components, with *z*-values of 4.8 (genera gain, pseudo $P < 0.05$) and 5.3 (genera loss, pseudo $P < 0.05$ using logistic regression.

The logistic relationship between TBI significance and percent agriculture (in 2007) was significant ($P = 0.04$), meaning that lakes in regions where agriculture developed in the 20$^{th}$ century were more likely to show large differences in diatom community composition between historical times and 2007. The URT constructed with total temporal beta diversity as a response variable and % land cover variables as input retained % agriculture, % developed, % grassland, % forest, % shrubland, and % wetland. After pruning the tree, % forest was the only significant

variable retained to explain total temporal beta diversity ($R^2_{adj}$ = 0.36; Fig. 6). The URT constructed with the genus gain component of temporal beta diversity retained % forest with a $R^2_{adj}$ = 0.40, % shrubland and % wetland had been pruned out.

Species-level results

We were interested in whether species-level data would allow for additional conclusions to be drawn regarding beta diversity. Working with a smaller set of lakes located in both the Northern Appalachians (n = 52) and Coastal Plains (n = 7) ecoregions, we found landscape patterns similar to our genus-level results, where alpha and spatial beta diversities were approximately equivalent between time points. Because we were working at the species-level, Species' Contribution to Beta Diversity (SCBD) could also be examined in relation to spatial beta diversity. The ten species with the highest contributions to SCBD in historical and 2007 assemblages were mostly planktonic and five of them were key contributors to SCBD in both data sets (see Appendix S5). In terms of LCBD, sites with significant LCBD were clustered in the Coastal Plains ecoregion in historical times, but more evenly distributed between the two ecoregions in modern times (Appendix S4, Fig. 1). Univariate of modern LCBD showed longitude, temperature and pH to be key determinants of high LCBD in these lakes (Adj. $R^2$ = 0.6).

Similar to the genus level analyses, temporal turnover of diatom species in this reduced set of lakes was between 50-60%. After correcting for multiple testing, no lakes had significant TBI, though this is likely a result of the reduced sample size of this dataset. Additionally, the TBI analyses could be run with 999999 permutations in an effort to improve the stability of the *P*-values during a correction for multiple testing.

**Discussion**

Our analyses identified patterns and drivers of diatom diversity across the conterminous United States over the last ~150 years and provided support for hypothesized Anthropocene trends previously identified (McGill *et al*., 2015), as well as some unexpected ones (Table 5). Because of the larger sample size and geographic coverage, our genus-level results are the most robust for drawing conclusions and are discussed at length here. Across all sites, we found that genus richness increased in the modern sediments over historical ones at the continental scale; however this was mostly driven by the marked changes observed in the Upper Midwest. This pattern may be the result of introductions of new genera and dispersal aided by hydrological modification (Alig *et al*., 2004); this is consistent with McGill *et al*., (2015). For spatial beta diversity, we found that historical and modern beta diversity measured at the ecoregion level were similar, although at the site level, there were substantial changes in diversity through time. Sites that contributed significantly to spatial beta diversity were not always those that experienced significant temporal changes in diatom composition. This is not surprising, given that in the spatial beta diversity analyses, the lakes with significant LCBD indices are those that are most exceptional at one time. In the temporal analysis, the lakes with the largest TBIs are those that have changed the most in diatom composition between time points. In this study, sites identified as 'reference' lakes prior to sampling were less likely to have experienced significant temporal changes in beta diversity. However, this pattern was less clear when examining the post-sampling U.S. EPA lake classification, for which significant temporal beta diversity was found in all three disturbance categories (least disturbed, intermediate disturbance and highly disturbed).

In terms of the mechanism underlying the changes in spatial and temporal beta diversity, we consistently found evidence for replacement-dominated spatial beta diversity within each time point. While the components computed in the spatial beta diversity analyses are not comparable with the taxa loss and gain components computed in the temporal analysis, these analyses taken separately provide evidence for strong changes in community composition, without strong changes to genus richness. Finally, while there does not appear to be a clear regional pattern in site contributions to modern spatial beta diversity, site-specific variables associated with water quality and land cover were significant predictors of the spatial variation in LCBD.

*Beta diversity as an important metric of biodiversity*

Alterations to biodiversity are not just via species or genera losses. While the loss of taxa is intuitively and empirically important, taxa gains and compositional shifts can also fundamentally alter ecosystems. In a meta-analysis of plant biodiversity, Vellend *et al*., (2013) found that local-scale species diversity was as likely to increase or decrease through time. More recently, Dornelas *et al*., (2014) analyzed time series data from various ecosystems and found distinct differences between alpha and beta diversity, with considerable beta diversity variation, but no net loss in alpha diversity. Taken together with our results, two aspects are highlighted: first, that some critical biodiversity patterns may only be visible through the lens of beta diversity, pointing to the importance of this estimate along with alpha and gamma diversity; and secondly, that beta diversity could be a key component of the biodiversity 'crisis' of the Anthropocene. Indeed, we found beta diversity to illuminate key patterns in diatom composition across land use gradients, providing more information than had we focused only on differences in alpha diversity alone. For example, our regression tree analyses showed different driving

variables for alpha diversity than components of beta diversity (particularly for water quality). Similar to Dornelas *et al*., (2014), beta diversity (both spatial and temporal) was significantly related to land cover change, a factor itself related to water quality (Taranu & Gregory-Eaves, 2008). Additionally, we saw diatom genera richness increase between historical and modern sediments, which could be related to climatic shifts; this is not unlike studies that demonstrated an increase in species richness with temperature (e.g. Stomp *et al*., 2011).

*Sites versus ecoregion-level responses*

Individual lakes and their contributions to either spatial or temporal beta diversity varied from site to site, even when variation across the ecoregions did not occur (based either on historical or 2007 sediments). For example, we found that lakes with very low specific conductivity (less than 18 $\mu S\ cm^{-1}$) had the highest LCBD values (spatial), indicating that these sites are particularly unique across the landscape. In our dataset, conductivity was positively correlated (corr = 0.6) with silica concentration, such that sites with low silica had the most distinctive diatom assemblages. The low conductivity and low silica results concord with an extensive body of literature documenting the responsiveness of diatom assemblages to water chemistry (Smol & Stoermer, 2010), especially ionic strength (Fritz *et al*., 2010). Spatial LCBD values were greater in basins with lower human (i.e. urban) development, indicating that less disturbed lakes may contribute more diatom heterogeneity to the landscape. Indeed, nutrient enriched lakes (a by-product of most development) can have a homogenizing effect on aquatic assemblages (Olden *et al.,* 2004; Donahue *et al*., 2009). We identified a very low threshold for the proportion of development in the watershed (Fig. 4) and suggest that this is due to the tendency for humans to modify shorelines even when the rest of the watershed may be intact (e.g., by removing riparian and littoral vegetation; Kaufmann *et al*., 2014), which in turn have

substantial consequences for diatoms (Velghe *et al*., 2012) and other organisms (Strayer & Findlay, 2010). The relationship with forest cover indicates that less forest cover may produce conditions that favour specific species or genera and hence representing more of the total beta diversity.

Land use variables were also informative in describing the observed temporal variation in diatom assemblage, which was higher in sites with lower forest cover in their basin. While the contemporary forest cover variable does not inform as to whether a change in cover occurred since historical times, it does indicate that lakes with higher basin forest cover currently have been buffered to diatom assemblage change through time. Overall, diatom communities were more likely to remain unaltered through time in sites with higher forest cover than where forest cover was less. This result is intuitive, and is reflected in the relationships between land cover variables and the genus gain component of temporal beta diversity; higher genus gain was associated with lower basin percentages of forest, shrubland, and wetland. Essentially, there was higher temporal beta diversity (more genus-level change) with reduced forest cover, driven by more gains at the genera-level in low forest, shrubland and wetland lakes.

*Dominance of the replacement component in spatial beta diversity*

Replacement was always the dominant component of spatial beta diversity, regardless of the time point. While a separate analysis, our temporal beta diversity results shed some light onto why replacement is such a dominant component of spatial beta diversity, both historically and in present times. The method used to compute the significance of temporal change for each lake (through permutation) was designed to detect significant changes through time when beta diversity is *not* uniform across space, i.e. if all lakes experienced the same amount of change in community composition through time, one could not identify significant temporal changes. Thus,

the temporal changes detected in our study resulted specifically from disproportionate temporal change across the landscape, which could also represent genus replacement in spatial beta diversity. Genus replacement is also more prevalent in regions with variable environments from lake to lake, due to factors like geographic isolation or poor dispersal ability (Leprieur *et al*. 2011). Broad-scale climatic changes (climate warming) could create a uniform temporal beta diversity pattern across the U.S. landscape. However, we noted that not all sites had significant temporal beta diversity, indicating that temporal changes differed from lake to lake, presumably due to local or regional factors. It is also important to note that climate change in the U.S. has not manifested itself to the same extent across all regions (Kennedy, 2014). Thus, it would be reasonable to expect that spatial beta diversity is explained by replacement via the mechanism described by Leprieur *et al*., (2011). Additionally, both alpha and gamma diversities were equivalent across ecoregions, further suggesting replacement as the dominant mechanism of spatial beta diversity.

Gain and loss of abundances on a genus-by-genus basis explain a nearly equivalent proportion of total beta diversity variation between our time points, as expected since the sum of these two components equals total beta diversity (Fig. 5a,b). This could mean that both the gain and loss of abundance of specific taxa are important to temporal change, and supported by the mechanisms identified for spatial beta diversity. An alternative explanation is that abundance gain and loss on a genus-by-genus basis was important at different intervals through time. This is shown in Legendre and Salvat (2015), where temporal beta diversity of mollusc communities was partitioned into genus gain and loss for five different time periods (each time period consisting of 5-20 years), with the importance of the different components alternating between the time periods and from site to site. With only two time points, like our historical and modern

points, these finer-scale alterations between genus gain and loss cannot be observed. Environmental variables and mechanisms behind temporal beta diversity may change from time period to time period as well as across temporal scales.

*Beta diversity results within the metacommunity concept*

Our study has delineated some specific diversity patterns in the different ecoregions and revealed mechanisms behind these (Table 6). The entire landscape, as well as the majority of ecoregions, show higher historical beta diversity, with the exception of the Coastal Plains and the Western Mountains, which both had lower modern spatial beta diversity. We can derive potential mechanisms driving beta diversity by considering both spatial and temporal patterns in the ecoregions. For example, in the Northern Appalachians ecoregion, the pattern of heterogeneous diatom communities across the landscape, both historically and currently, could have arisen because taxa losses have matched taxa gains across sites within the ecoregion metacommunity. On the other hand, diatom assemblages across the modern landscape may be more heterogeneous than historically, as they are in the Temperate Plains ecoregion, because dispersal of taxa through time is sufficiently high to allow for genera or species (environmental) sorting across the landscape. In the Upper Midwest, we can envision yet another scenario: diatoms may be more homogeneously distributed across space currently because population dynamics have resulted in genera being added to sites in a non-genera sorting (*sensu* mass effects; Leibold *et al*., 2004) way. Diatoms may also be more homogeneous across space currently because the taxa removed over time are always the same, regardless of site as could be the case in the Western Mountains ecoregion. These potential scenarios can now be investigated further by ecoregion through more detailed study within each, using metacommunity theory for guidance.

*Conclusions*

Patterns in both spatial and temporal beta diversity varied both across sites and ecoregions, via several possible mechanisms, of which we have highlighted some of the most important. By considering temporal and spatial beta diversity in the same study, we have been able to provide a holistic view of diatom biodiversity patterns and test hypotheses and predictions for change during the Anthropocene (McGill *et al*., 2015). While the majority of our work was at the genus-level, we have laid out a useful and robust framework for future beta diversity study. Future work considering other trophic levels, as well as data that incorporates multiple historical time points will be most insightful for generating a broader perspective of biodiversity change at multiple scales.

**Acknowledgements**

The authors wish to thank Amina Pollard of the NLA programme for help with questions about field sampling, as well as Mihaela Enache and Marina Popatova at the Patrick Center for Environmental Research for help procuring sediment material for sediment dating. Julien Gogot and Bassam Ghaleb from GEOTOP-UQÀM completed the radiometric dating. Funding for this project was made possible by an FRQNT Equipe grant awarded to IGE, BEB and PL, as well as an NSERC PGS-D scholarship awarded to AKW. AKW also received additional funding from a McGill International Travel Grant and an Excellence Award from the Québec Centre for Biodiversity Science. This project also received funding from the NARS Campus Challenge award. This study is a contribution to the FRQNT-funded Québec Centre for Biodiversity Science (QCBS-CSBQ).

**Literature Cited**


Alig R.J., Kline J.D., & Lichtenstein M. (2004). Urbanization on the US landscape: looking ahead in the 21st century. *Landscape and Urban Planning*, **69**, 219-234.

Beaulieu M., Pick F., & Gregory-Eaves I. (2013). Nutrients and water temperature are significant predictors of cyanobacterial biomass in an 1147 lakes data set. *Limnology and Oceanography*, **58**, 1736-1746.

Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D.A., Kinzig A.P., Daily G.C., Loreau M., Grace J.B., Larigauderie A., Srivastava D.S., and Naeem S. (2012). Biodiversity loss and its impact on humanity. *Nature*, **486**, 59-67.

Charles D.F., Knowles C., & Davis R.S. (2003). Protocols for the analysis of algal samples as collected as part of the U.S. Geological Survey National Water-Quality Assessment program. Patrick Center for Environmental Research, The Academy of Natural Sciences. Report No. 02-06.

Crutzen P.J. (2002). Geology of mankind. *Nature*, **415**, 23.

De'ath G. (2002). Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology*, **83**, 1105-1117.

Donahue I., Jackson A.L., Pusch M.T., & Irvine K. (2009). Nutrient enrichment homogenizes lake benthic assemblages at local and regional scales. *Ecology*, **90**, 3470-3477.

Dornelas M., Gotelli N.J., McGill B., Shimadzu H., Moyes F., Sievers C., & Magurran A.E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science,* **344**, 296-299.

Ellison A.M. (2010). Partitioning diversity. *Ecology,* **91**, 1962-1963.

Fritz S.C., Cumming B.F., Gasse F., & Laird K.R. (2010). Diatoms as indicators of hydrologic and climatic change in saline lakes in "The Diatoms: Applications for the Environmental and Earth Sciences 2nd edition". Cambridge University Press, Cambridge UK.

Glew J. (1989). A new trigger mechanism for sediment samples. *Journal of Paleolimnology*, **2**, 241-243.

Hamilton C. (2015). Getting the Anthropocene so wrong. *The Anthropocene Review*, **2**, 102-107.

Herlihy A.T., Paulsen S.G., Van Sickle J., Stoddard J.L., Hawkins C.P., & Yuan L.L. (2008). Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *Journal of the North American Benthological Society*, **27**, 860-877.

Herlihy A.T. (2015). Personal Communication.

Kaufmann P.R., Peck D.V., Paulsen S.G., Seeliger C.W., Hughes R.M., Whittier T.R., & Kamman N.C. (2014). Lakeshore and littoral physical habitat structure in a national lakes assessment. *Lake and Reservoir Management*, **30**, 192-215.

Kennedy C. (2014). Does "global warming" mean it's warming everywhere? National Oceanic and Atmospheric Administration. Available online: https://www.climate.gov/news-features/climate-qa/does-global-warming-mean-it%E2%80%99s-warming-everywhere. Accessed 31 August 2015.

Legendre P. (2013a). R function: beta.div(). Available online: http://adn.biol.umontreal.ca/~numericalecology/Rcode/.

Legendre P. (2013b). R function: LCBD.comp(). Available online: http://adn.biol.umontreal.ca/~numericalecology/Rcode/.

Legendre P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324-1334.

Legendre P. (2015). R function: TBI(). Available online: http://adn.biol.umontreal.ca/~numericalecology/FonctionsR/.

Legendre,P. (2015). Personal Communication.

Legendre P., & De Cáceres M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters,* **16**, 951-963.

Legendre P., & Legendre L. (2012). Numerical Ecology, 3rd ed. Elsevier Science BV, Amsterdam.

Legendre P., & Salvat B. (2015). Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.

Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F., Holt R.D., Shurin J.B., Law R., Tilman D., Loreau M., & Gonzalez A. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601-613.

Leprieur F., Tedesco P.A., Hugueny B., Beauchard O., Dürr H.H., Brosse S., & Oberdorff T. (2011). Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325-334.

Lewis S.L., & Maslin M.A. (2015). Defining the Anthropocene. *Nature*, **519**, 171-180.

McGill B.J., Dornelas M., Gotelli N.J., & Magurran A.E. (2015). Fifteen forms of biodiversity trend in the Antropocene. *Trends in Ecology and Evolution*, **30**, 104-113.

McGill B.J. (2015). Land use matters. *Nature*, **520**, 38-39.

Newbold T., Hudson L.N., Hill S.L.L., Contu S., *et al*. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*, 45-50.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., & Wagner H. (2015). Package 'vegan'. Available: http://cran.r-project.org/web/packages/vegan/vegan.pdf.

Olden J.D., Poff N.L., Douglas M.R., Douglas M.E., & Fausch K. (2004). Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution,* **19**, 18-24.

Pollard A. (Personal communication).

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org/

Smol J.P., & Stoermer E.F. (eds) (2010). The Diatoms: Applications for the Environmental and Earth Sciences 2nd edition. Cambridge University Press, Cambridge UK.

Steffen W., Grinevald J., Crutzen P., & McNeill J. (2011). The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society A*, **369**, 842-867.

Stomp M., Huisman J., Mittelbach G.G., Litchman E., & Klausmeier C.A. (2011). Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology*, **92**, 2096-2107.

Strayer D.L., & Dudgeon D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society,* **29**, 344-358.

Strayer D.L., & Findlay S.E.G. (2010). Ecology of freshwater shore zones. *Aquatic Sciences*, **72**, 127-163.

Taranu Z.E., & Gregory-Eaves I. (2008). Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. *Ecosystems*, **11**, 715-725.

Therneau T., Atkinson B., & Ripley B. (2015). Package 'rpart'. Available: http://cran.r-project.org/web/packages/rpart/rpart.pdf.

U.S. EPA. (2009). National Lakes Assessment: A Collaborative Survey of the Nation's Lakes. EPA 841-R-09-001. U.S. Environmental Protection Agency, Washington, DC, USA.

U.S. EPA. (2011-2012). National Lakes Assessment. Field Operations Manual. EPA 841-B-11-003. U.S. Environmental Protection Agency, Washington, DC, USA.

USGS. (1992). The USGS Land Cover Institute: NLCD 92 Land Cover Class Definitions. Available: http://landcover.usgs.gov/classes.php.

Velghe K., Vermaire J.C., & Gregory-Eaves I. (2012). Declines in littoral species richness across both spatial and temporal nutrient gradients: A palaeolimnological study of two taxonomic groups. *Freshwater Biology*, **57**, 2378-2389.

Vellend M., Baeten L., Myers-Smith I.H., Elmendorf S.C., Beauséjour R., Brown C.D., De Frenne P., Verheyen K., & Wipf S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences,* **110**, 19456-19459.

Wardle D.A., Bardgett R.D., Callaway R.M., & Van der Putten W.H. (2011). Terrestrial ecosystem responses to species gains and losses. Science, **332**, 1273-1277.

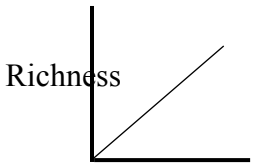
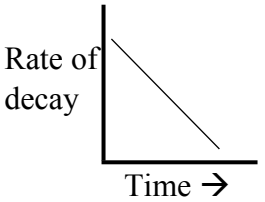
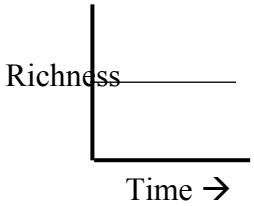
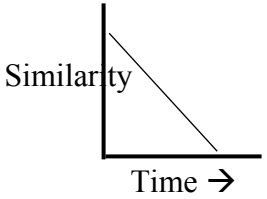
Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, **21**, 213-251.

Winegardner A.K., Beisner B.E., Legendre P., & Gregory-Eaves I. (2015). Are the landscape-level drivers of water column and surface sediment diatoms different? *Freshwater Biology,* **60**, 267-281.

Zalasiewicz J., Waters C.N., Barnosky A.D., Cearreta A., *et al*. (2015). Colonization of the Americas, 'Little Ice Age' climate, and bomb-produced carbon: Their role in defining the Anthropocene. *The Anthropocene Review*, **2**, 117-127.

**Tables**

**Table 1: Hypotheses and predictions for alpha and beta diversity analyses based on McGill *et al*. (2015, Figure 2) framework.**

| **Spatial scale** | **α- or γ-diversity** | **Temporal ß-diversity** | **Spatial ß-diversity** |
|---|---|---|---|
| Continental<br><br>(Biogeographic) |  <br>Higher gamma diversity in modern times | No prediction |  <br>Less spatial beta diversity in modern times (homogenization) |
| Ecoregion<br><br>(Metacommunity) | Higher alpha diversity in modern times | No prediction | Less spatial beta diversity in modern times (homogenization- as above) |
| Lake (Local) |  <br>Alpha diversity stays constant on local scale |  <br>Beta diversity increases between observations as they are father apart in time | |

Spatial ß decrease as move down scale

**Table 2: Spatial beta (ß) diversity, mean rarefied genus richness (G), mean alpha (α) diversity and gamma (γ) diversity for each ecoregion (Shannon diversity of genus' sums).** Beta diversity was calculated using total variance computed using *beta.div()* based on percentage difference matrices. "Hx" refers to the historical sediments. Genus richness was rarefied after rare genera (<2% relative abundance) had been excluded. The Xeric ecoregion was excluded in these mean values because there was only one site in that ecoregion.

| | **ß-diversity** | | **Rarefied G** | | **α-diversity (Shannon)** | | **α-diversity (Simpson)** | | **γ-diversity** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Ecoregion** | Hx | 2007 | Hx | 2007 | Hx | 2007 | Hx | 2007 | Hx | 2007 |
| All | 0.36 | 0.36 | 18.0 | 24.3 | 2.3 | 2.2 | 0.9 | 0.8 | 3.3 | 3.2 |
| Coastal Plains | 0.37 | 0.32 | 19.2 | 18.4 | 1.7 | 2.0 | 0.8 | 0.8 | 2.5 | 2.7 |
| Northern Appalachians | 0.30 | 0.31 | 29.9 | 27.5 | 2.5 | 2.4 | 0.9 | 0.8 | 3.2 | 3.1 |
| Southern Plains | 0.28 | 0.28 | 25.0 | 24.6 | 2.3 | 2.4 | 0.8 | 0.8 | 2.8 | 2.8 |
| Temperate Plains | 0.32 | 0.31 | 25.7 | 22.2 | 2.2 | 2.1 | 0.8 | 0.8 | 2.8 | 2.7 |
| Upper Midwest | 0.36 | 0.34 | 17.0 | 24.0 | 2.2 | 2.1 | 0.9 | 0.8 | 3.2 | 3.1 |
| Western Mountains | 0.37 | 0.37 | 25.8 | 21.8 | 2.1 | 1.9 | 0.9 | 0.7 | 3.2 | 3.0 |

**Table 3: Explanatory components for historical and 2007 spatial beta diversity,** as computed using *beta.div.comp()*. 'Repl' refers to the replacement component; 'AbDiff' refers to the abundance difference component; and 'Repl/Total' and 'AbDiff/Total' are these two components with total beta diversity as the denominator.

| **Ecoregion** | **Repl** | **AbDiff** | **Repl/Total** | **AbDiff/Total** |
|---|---|---|---|---|
| *Historical* | | | | |
| All | 0.31 | 0.04 | 0.88 | 0.12 |
| Coastal Plains | 0.36 | 0.01 | 0.97 | 0.03 |
| Northern Appalachians | 0.29 | 0.01 | 0.96 | 0.04 |
| Southern Plains | 0.26 | 0.02 | 0.93 | 0.07 |
| Temperate Plains | 0.29 | 0.03 | 0.90 | 0.09 |
| Upper Midwest | 0.29 | 0.07 | 0.81 | 0.19 |
| Western Mountains | 0.33 | 0.04 | 0.91 | 0.09 |
| *2007* | | | | |
| All | 0.34 | 0.01 | 0.96 | 0.04 |
| Coastal Plains | 0.32 | <0.01 | 0.10 | <0.01 |
| Northern Appalachians | 0.31 | <0.01 | 0.10 | 0.01 |
| Southern Plains | 0.28 | 0 | 1 | 0 |
| Temperate Plains | 0.28 | 0.03 | 0.92 | 0.08 |
| Upper Midwest | 0.33 | 0.02 | 0.95 | 0.05 |
| Western Mountains | 0.35 | 0.02 | 0.94 | 0.07 |

**Table 4: Mean (and standard deviation in parentheses) temporal beta diversity components for the ecoregions.** 'Total beta' refers to the mean value of the temporal beta diversity in each region region (mean value of the D column in the 'BCD' table provided by the function *TBI()*. It was computed using the percentage difference index applied to the diatom abundance data; values are in the [0,1] range. Total beta is the sum of 'Genera loss' and 'Genera gain'. Genera loss refers to the component representing loss of abundances on a genus by genus basis between the historical and 2007 time points. Genera gain refers to the component representing gain of abundances on a genus by genus basis between the historical and 2007 time points. These components were computed on a lake-by-lake basis and then averaged for each ecoregion.

| **Ecoregion** | **Loss of genera** | **Gain of genera** | **Total beta** |
|---|---|---|---|
| Coastal Plains | 0.28 (0.08) | 0.27 (0.08) | 0.55 (0.16) |
| Northern Appalachians | 0.20 (0.08) | 0.18 (0.08) | 0.38 (0.16) |
| Southern Plains | 0.21 (0.09) | 0.19 (0.06) | 0.40 (0.14) |
| Temperate Plains | 0.23 (0.04) | 0.22 (0.08) | 0.46 (0.09) |
| Upper Midwest | 0.26 (0.10) | 0.26 (0.13) | 0.51 (0.17) |
| Western Mountains | 0.29 (0.09) | 0.25 (0.08) | 0.54 (0.17) |

**Table 5: Summary of our results (in italics) in relation to the McGill *et al*. (2015) framework**

| Spatial scale | α- or γ-diversity | Temporal ß-diversity | Spatial ß-diversity |
|---|---|---|---|
| Continental (Biogeographic) | Rarefied genus richness higher in modern times<br><br>*Matches trend* | No prediction | Spatial beta diversity approximately equivalent between historical and modern times<br><br>*Does not match trend* |
| Ecoregion (Metacommunity) | Rarefied genus richness higher in modern times in some ecoregions<br><br>*Partial match* | No prediction | Spatial beta diversity less in modern times than historical times in some ecoregions<br><br>*Partial match* |
| Lake (Local) | Rarefied genus richness higher in modern times for some lakes<br><br>*Partial match* | 12% of lakes showed significant temporal beta diversity<br><br>*Direct comparison not possible with only two time points* | 13% (historical) and 10% (2007) of lakes made significant contributions to spatial beta diversity (did not test in lake spatial beta) |

**Table 6: Summary of temporal and spatial beta diversity across ecoregions.** The mechanisms column for temporal beta diversity shows when abundance loss (genus by genus) or gains (genus by genus) are equivalent or which one dominates.

| **Ecoregion** | **Mechanism explaining *temporal* beta diversity (genus by genus)** | **Observation from *spatial* beta diversity** |
|---|---|---|
| Coastal Plains | = | Lower in contemporary |
| Northern Appalachians | Abundance loss | ~= between contemporary and historical |
| Southern Plains | Abundance loss | = between contemporary and historical |
| Temperate Plains | Abundance loss | ~= between contemporary and historical |
| Upper Midwest | = | Lower in contemporary |
| Western Mountains | Abundance loss | = between contemporary and historical |

**Figures**

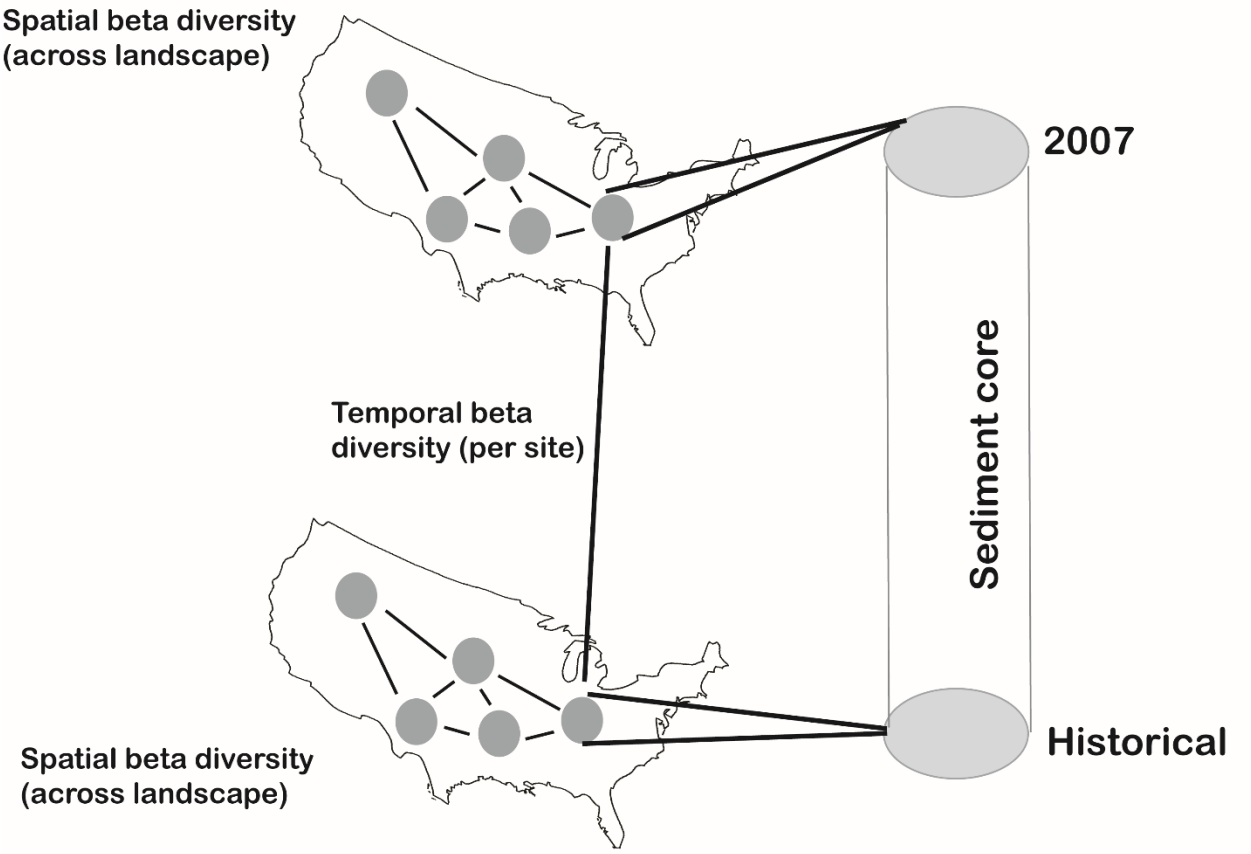



**Figure 1: Conceptual diagram of the different forms of beta diversity analysed using surface and bottom sediment core samples from the 2007 National Lakes Assessment (genus- and species-level analyses).** Beta diversity *between lakes* (represented by the grey circles) across the landscape was investigated using both bottom (historical) and modern (2007) sediments from cores; resulting in both historical and contemporary estimates of spatial beta diversity. Temporal beta diversity between historical and contemporary conditions was investigated for *each lake*.

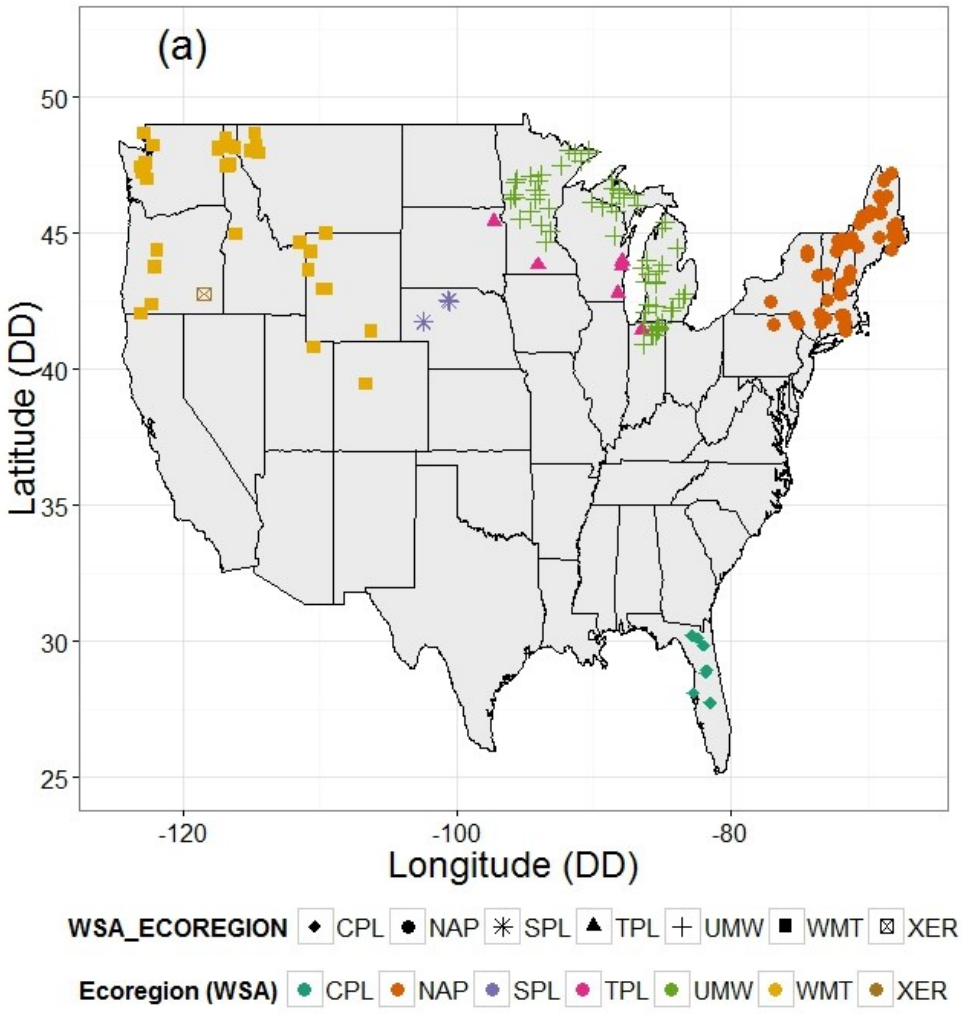

(a)
50
45
40
35
30
25
Latitude (DD)
-120
-100
-80
Longitude (DD)
WSA_ECOREGION CPL NAP SPL TPL UMW WMT XER
Ecoregion (WSA) CPL NAP SPL TPL UMW WMT XER

**Figure 2**

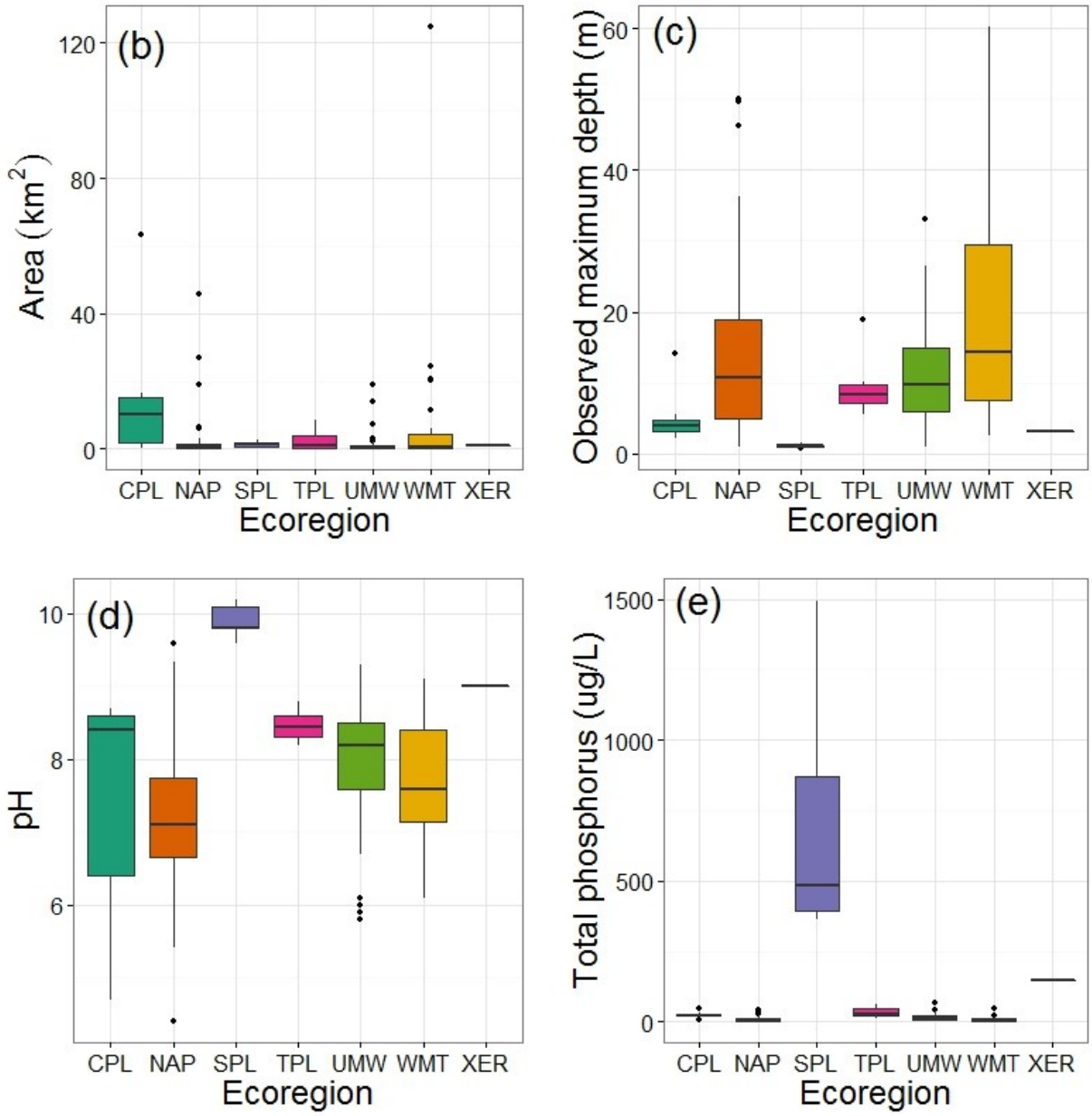



**Figure 2: Location of lake sites used in this study for genus-level analyses and key limnological variables.** (a) Lake sites are classified into seven distinct ecoregions: Coastal Plains (CPL, n = 7), Northern Appalachians (N. Appalachians/NAP, n = 53), Southern Plains (S. Plains/SPL, n = 5), Temperate Plains (TPL, n = 6), Upper Midwest (UMW, n = 69), Western Mountains (W. Mountains/WMT, n = 35), and Xeric (XER, n = 1). Boxplots show the range and median of (b) lake surface area ($km^2$), (c) observed maximum depth (m), (d) pH and (e) TP (µg

$L^{-1}$), where the centre horizontal line is the median, the lower horizontal line the $25^{th}$ percentile, the upper horizontal line the $75^{th}$ percentile and points represent outliers.

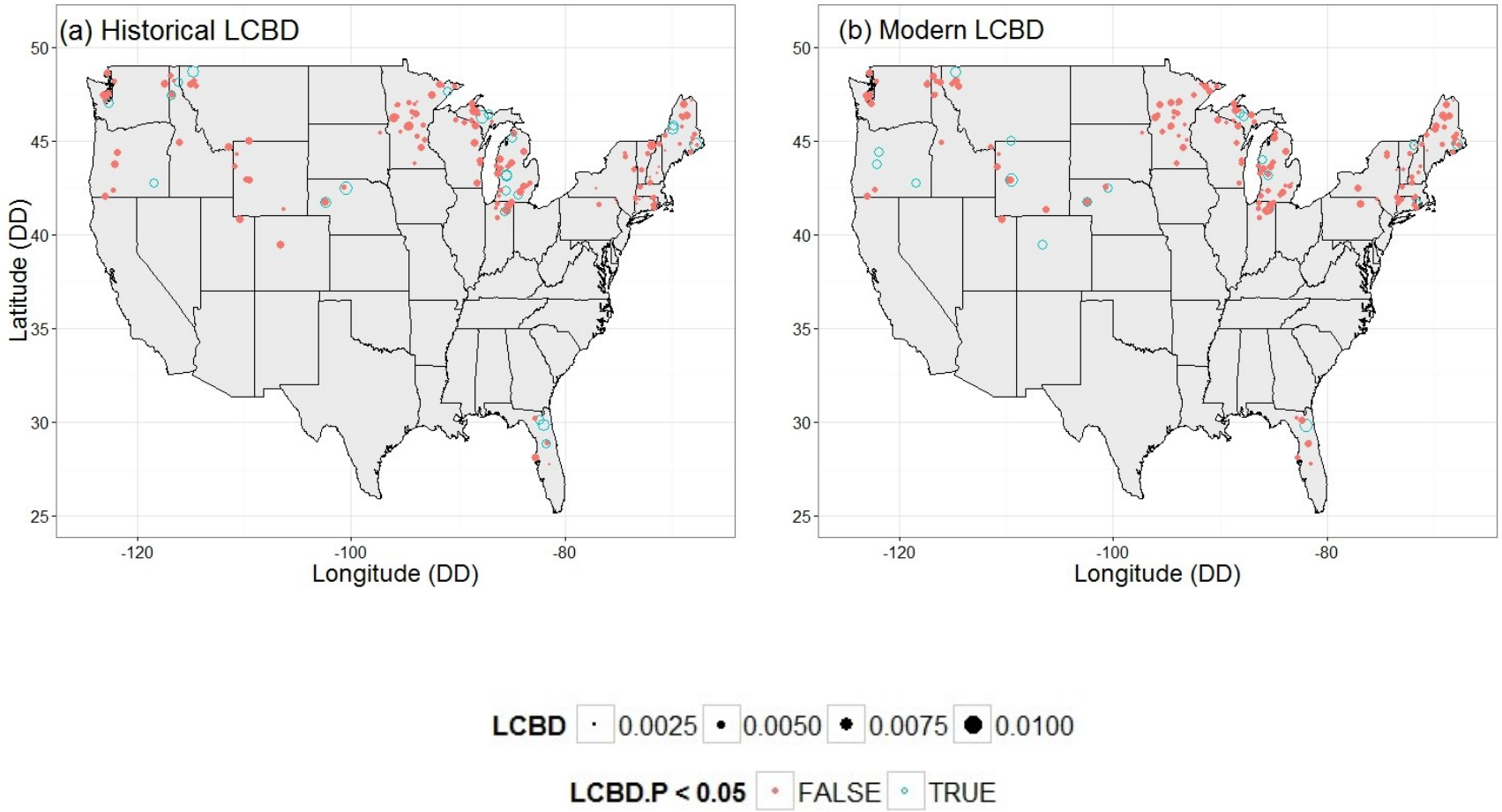

(a) Historical LCBD
(b) Modern LCBD
Latitude (DD)
Longitude (DD)
50
45
40
35
30
25
-120
-100
-80
LCBD
0.0025
0.0050
0.0075
0.0100
LCBD.P < 0.05
FALSE
TRUE

**Figure 3**

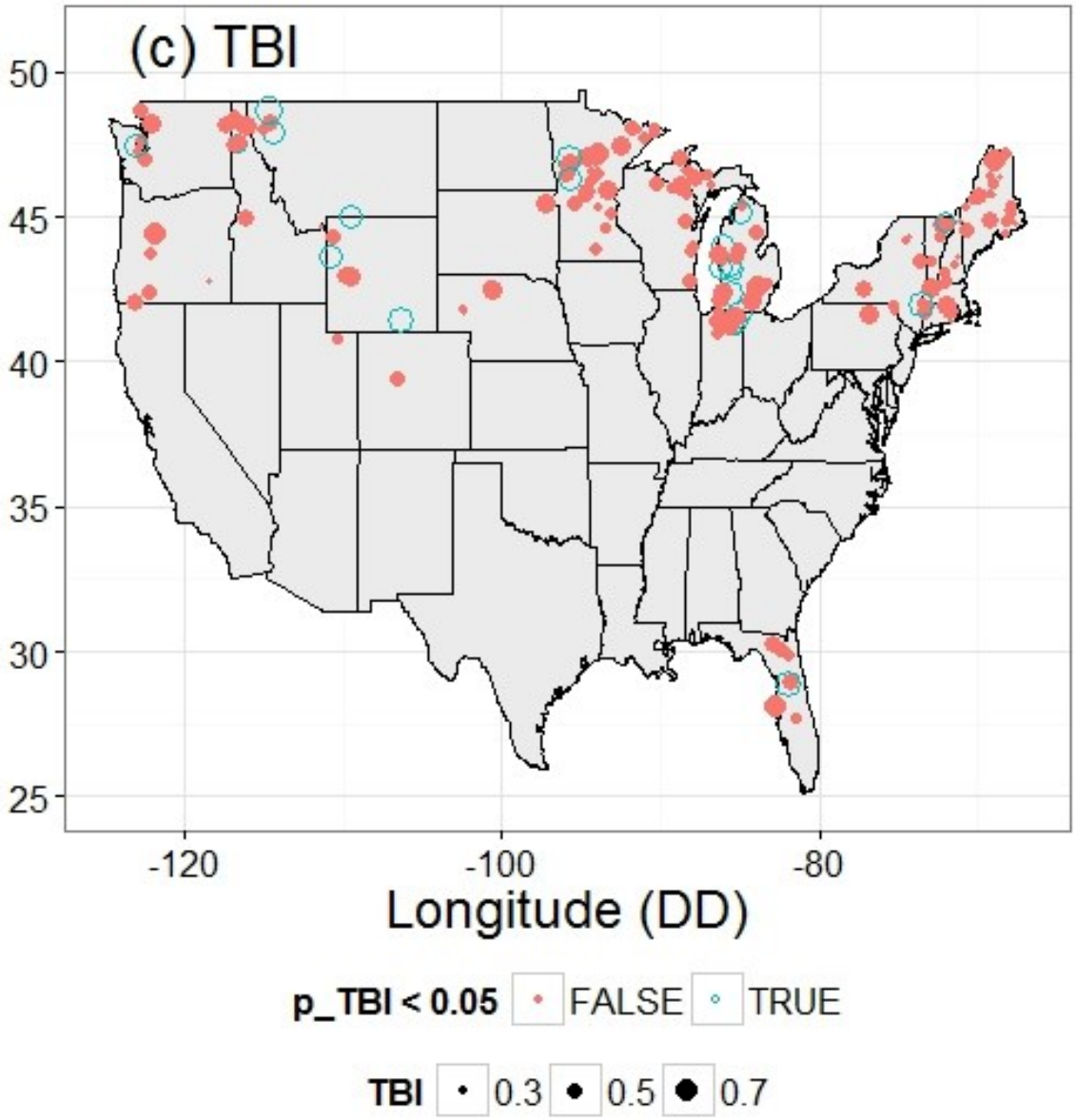



**Figure 3: LCBD values for (a) historical spatial beta diversity and (b) 2007 spatial beta diversity and exceptional sites (c) for temporal beta diversity (TBI) (all genus-level).** A lake has a significant LCBD value if $P < 0.05$, and is coded as "True" (open circles). LCBD values across all sites (in either historical or 2007) sum to 1. A lake has significant total temporal beta diversity if $P < 0.05$ before correction for multiple testing and is also coded as "True" (open circles) in (c). Note that no sites had significant TBI after correction for multiple testing.

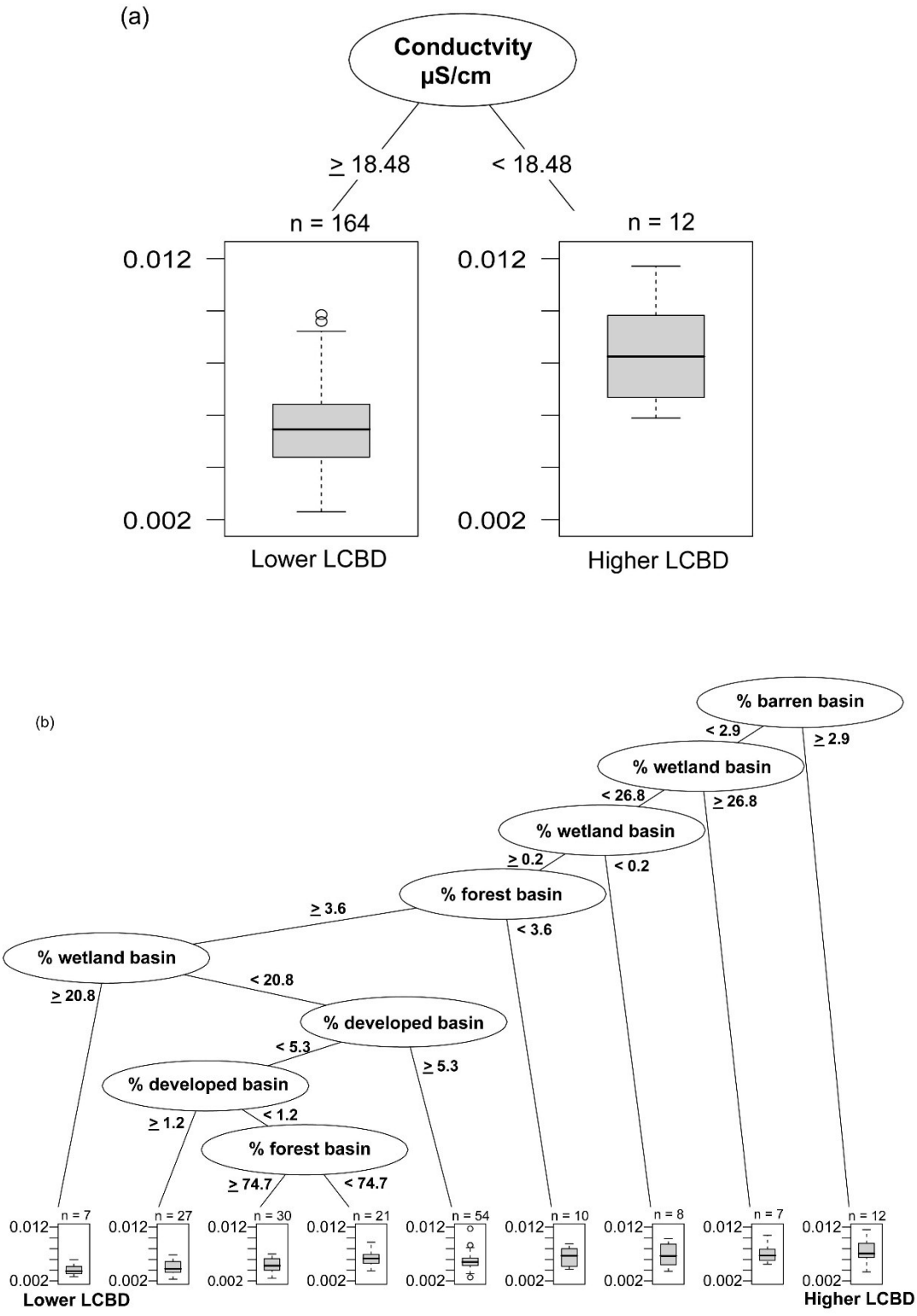



**Figure 4: Univariate regression tree of 2007 spatial LCBD (n = 176; genus-level) explained by (a) 2007 the retained water quality variable and (b) land cover variables.** The units for

the water quality variable id: Conductivity: µS $cm^{-1}$. The cumulative $R^2_{adj}$ for the model is 0.5. The land cover variables are percentage of each lake basin. The cumulative $R^2_{adj}$ for the model is 0.3.

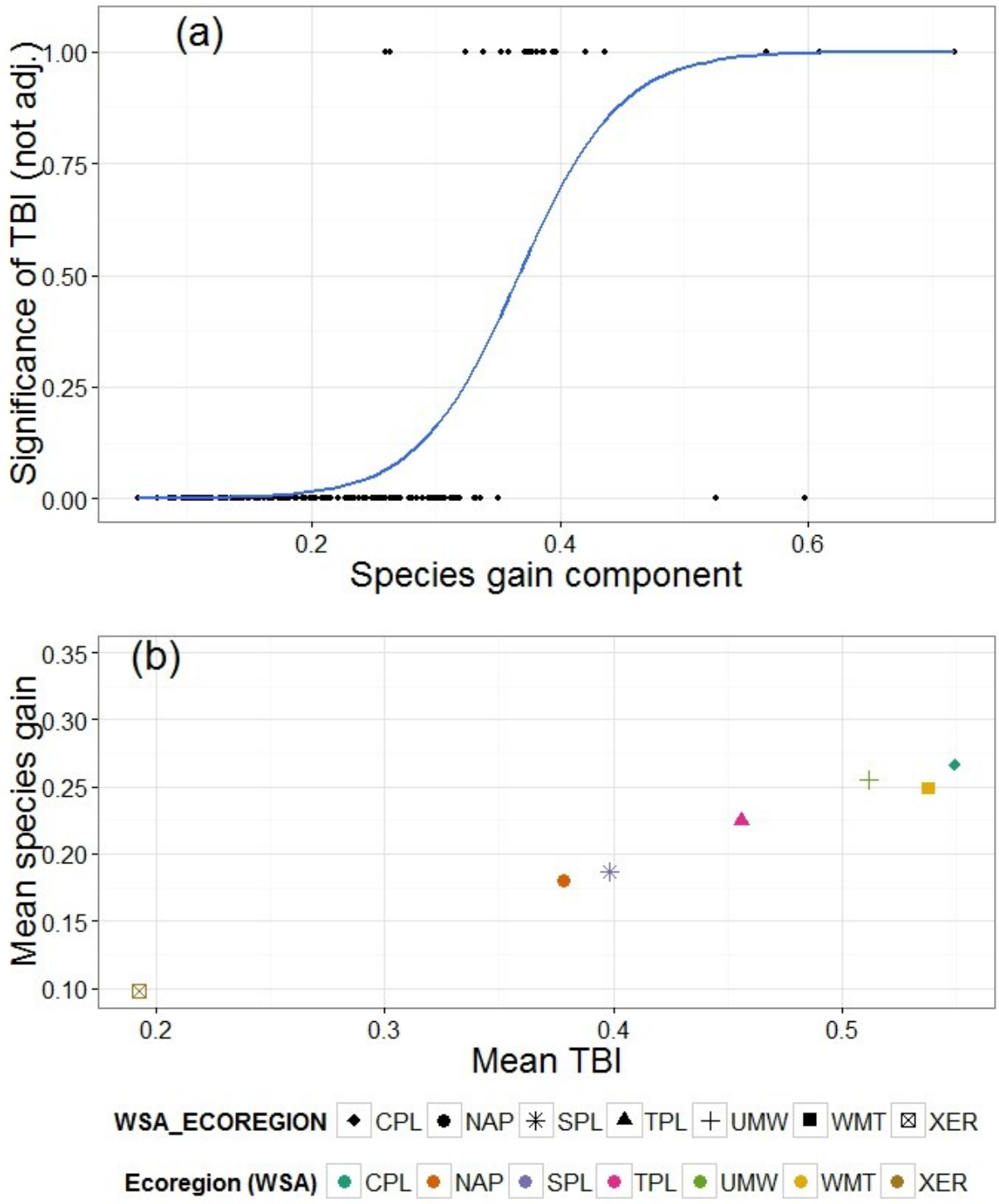



**Figure 5: Temporal beta diversity explained by genus gain component (n = 176; genus-level).** Panel (a) shows the relationship between genera gain and whether a particular lake site

experienced significant ($P < 0.05$) beta diversity (TBI) between historical and 2007 conditions using logistic regression, where "1" means that the temporal beta diversity was significant (not adjusted for multiple testing). Panel (b) shows the eco-regional relationships between mean temporal beta diversity and the proportion of temporal beta diversity explained by variation in genera gain.

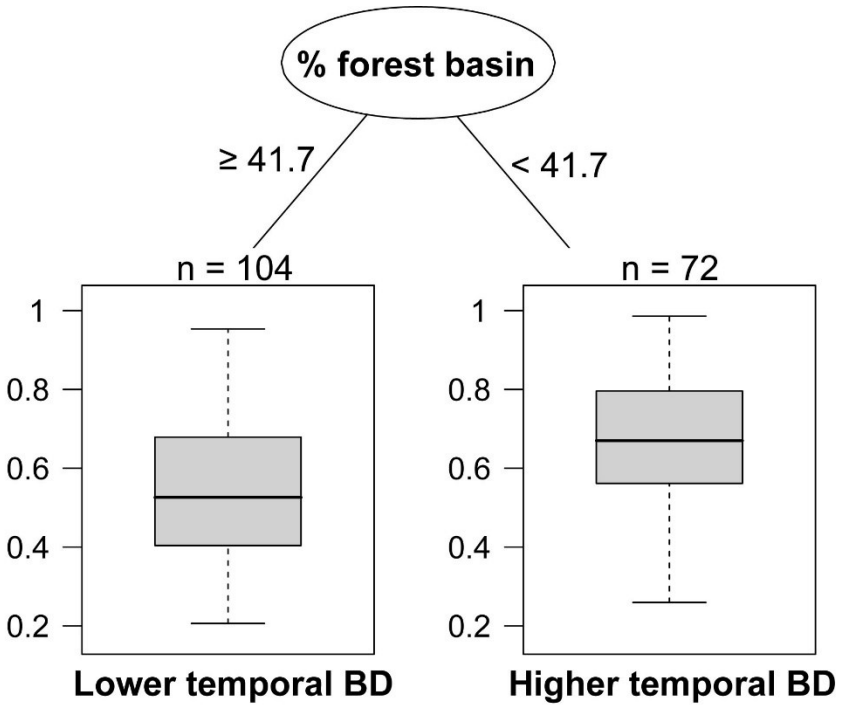



**Figure 6: Univariate regression tree of total temporal beta diversity explained by percent forest in a basin (n = 176; genus-level).** The $R^2_{adj} = 0.36$.

## CONNECTING STATEMENT 2

Temporal beta diversity was a major focus in Chapter 2, where I found that lakes varied in their magnitude of diatom beta diversity through time. I also found that there was no geographical pattern to this variation, but that temporal beta diversity was high in some disturbed (as classified by the U.S. EPA) lakes, and that forest cover in a basin was a key driver of this temporal beta diversity. However, our approach to quantifying temporal beta diversity in Chapter 2 was limited by the sediment core data available in the NLA, where only samples from the bottom and surface cores were kept and enumerated for their diatom assemblages. While I worked to overcome the limitation this posed in terms of accurately dating bottom core samples (and the variation in their ages), there is still the issue that the temporal beta diversity findings from Chapter 2 include only two time points, with significant time elapsing between them. I wanted to better characterize how temporal beta diversity might differ through time and expand my thinking on how different drivers might influence temporal beta diversity through time. Additionally, I wanted to more fully test the temporal beta diversity trend of interest identified in McGill *et al*., (2015), where temporal beta diversity is hypothesized to increase through time, with increasing turnover as time elapses (i.e. turnover increases with more temporal distance between samples, but also, the magnitude of temporal turnover continues as you progress through time in the Anthropocene). I was also interested in adopting a similar approach to the study of beta diversity dynamics to what is described in Legendre and Salvat (2015), where the authors quantified the relative importance of explanatory components behind temporal beta diversity.

In order to truly investigate these trends in Chapter 3, I needed to work with paleolimnological data that included multiple time points from sediment cores. As such, I used full sediment cores from two lakes from the iron-ore mining region of Schefferville, Québec to

track temporal beta diversity across multiple time points. I also focused on cladoceran zooplankton communities, switching from primary producers (diatoms in Chapters 1 and 2) to a primary consumer, a group that has been shown to be responsive to both biotic and abiotic drivers (e.g. Arnott & Vanni, 1993; Binks *et al*., 2005).

**Literature Cited**

Arnott S.E., & Vanni M.J. (1993). Zooplankton assemblages in fishless bog lakes: Influence of biotic and abiotic factors. *Ecology,* **74**, 2361-2380.

Binks J.A., Arnott S.E., & Sprules G.W. (2005). Local factors and colonist dispersal influence crustacean zooplankton recovery from cultural acidification. *Ecological Applications*, **15**, 2025-2036.

Legendre P., & Salvat B. (2015). Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.

McGill B.J., Dornelas M., Gotelli N.J., & Magurran A.E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology and Evolution*, **30**, 104-113.

# CHAPTER 3

## Cladoceran zooplankton diversity dynamics in lakes from a northern mining region: responses to multiple stressors characterized by alpha and beta diversity

Amanda K. Winegardner[1,2*], Natasha Salter[1], Stéphane Aebischer[3], Reinhard Pienitz[3], Alison M. Derry[2,4], Boswell Wing[5], Beatrix E. Beisner[2,4], and Irene Gregory-Eaves[1,2]

1. Department of Biology, McGill University, Montréal, Québec
2. Québec Centre for Biodiversity Science, Montréal, Québec
3. Centre d'Études Nordiques, Université Laval, Québec, Québec
4. Department of Biological Sciences, University of Québec at Montréal, Montréal, Québec
5. Department of Earth and Planetary Sciences, McGill University, Montréal, Québec

**Abstract**

The presence of multiple stressors is now a reality in the majority of aquatic systems. Hydrological modification, urban development, agriculture, pollution and climate change have the potential to affect aquatic organisms through many different mechanisms, and their combined impacts may be synergistic, antagonistic, additive or compensatory. The lakes surrounding the iron-ore mining region of Schefferville, Québec sit within a landscape of multiple stressors, which have been relatively well documented over time; this combination of factors (perturbations) makes this area an ideal site for studying the effects of multiple stressors on cladoceran zooplankton. Based on the analysis of sediment cores, we used cladoceran subfossil assemblages from two lakes located within the town of Schefferville to track both alpha and beta diversity over the last 100+ years (during which time urban development, pollution and iron-ore extraction activities have varied substantially). We showed that high metal concentrations were correlated with decreased zooplankton diversity and that the site that experienced both direct wastewater input and atmospheric metal loading (Lake Dauriat) had the greatest declines cladoceran richness. In both lakes, turnover in zooplankton assemblages was highest during the mining period. During the period of mine closures and improvement of wastewater treatment, Lake Dauriat showed some decreases in metal enrichment in the sediments and increases in cladoceran richness, but these values have not returned to pre-industrial conditions. Overall, our combined use of species richness and beta diversity metrics provide key insights into understanding the dynamics and drivers of biodiversity change in northern freshwater ecosystems.

## Introduction

Mining and other extractive industries modify landscapes in many different ways. In Canada, many of these industries are located in freshwater-rich regions (Natural Resources Canada, 2014), as well as in regions thought to be susceptible to ongoing climatic change (Prowse *et al*., 2009). Perturbation of freshwater systems is of general concern because of their importance in providing ecosystem services, their role as biodiversity hotspots (Strayer & Dudgeon, 2010), and the global prevalence of water insecurity (Vörösmarty *et al*., 2010). Although local and/or regionally significant anthropogenic disturbances are now the norm, it is all the more important to understand how, where and why aquatic biodiversity resists change, or shows resilience.

The mining region surrounding Schefferville, Québec (Canada) has experienced an intense history of multiple stressors, through iron-ore mining and the development of associated infrastructure. The town of Schefferville is similar to many other communities developed around extractive industries, where the infrastructure needed to manage community and (in some cases) industrial wastes has not necessarily kept pace with its rapid development. Indeed, wastewater treatment facilities were not installed in Schefferville until 1975, despite the founding of the town site in the early 1950s (Adams, 2007; Aebischer *et al*., 2015). The approach of dealing with restoration of mining sites after the fact, as opposed to via proactive planning has been viewed as a "pay-it-backward" approach and has been common across the mining industry in Québec (Hamilton *et al*., 2015), in addition to being very costly.

While many mining regions around the world have experienced large-scale acidification of their surface water bodies (such as the well-studied nickel mining region of Sudbury-Killarney; Dillon *et al*., 1987), the sedimentary rock geology of the Labrador Trough has

prevented acidification of aquatic systems around Schefferville (Aebischer *et al*., 2013; Aebischer *et al*., 2015). As such, these lakes provide an opportunity to better understand the effects of metal pollution and increased nutrient inputs on zooplankton diversity in a boreal mining region, without the compounding effect of acidification; especially as the interactions of stressors may be lessened when one stressor has an omnibus effect (such as pH) (Burton & Johnston, 2010).

Increased nutrient and metal inputs have varied through time and across the Schefferville landscape (especially between the lakes within the town), arising not only from mining sites, but also from commercial infrastructure and residential development to support the mining sector (Lapèrierre *et al*., 2008; Aebischer *et al*., 2015). Laperrière *et al*., (2008) was one of the first studies to demonstrate an effect of mining and mining-related development on the aquatic structure and diversity of Lake Dauriat (previously named Lake Pearce). Their study emphasized that, while the majority of mining activities occurred between the mid-1950s and late 1970s, the legacy effects of these activities were still occurring in Lake Dauriat, more than 20 years later. Metal loading to sediments means that the negative effects of metal contamination can persist long past a perturbation event and continue to cause pulses of pollutants during what would otherwise be considered recovery phases (Burton & Johnston, 2010; Tropea *et al*., 2010). Evidence of long-lasting effects on aquatic systems in mining areas has been echoed in other regions, such as the Sudbury nickel belt, where mining effects on biodiversity have persisted past the point at which lakes have been considered chemically recovered (Keller *et al*., 2002; Tropea *et al*., 2010; Valois *et al*., 2011; Gray *et al*., 2012; Labaj *et al*., 2015). For the Schefferville region, long-term effects are important to consider, as there was a resurgence of various mining activities in 2009 and 2011, though mining activities are presently slowing again in this region.

Studying other lakes in the Schefferville area is also important as the region has been subject to a gradient of metal loading and urban development (Aebischer *et al*., 2015), which are two stressors known to have additive effects in streams (Merriam *et al*., 2011).

Given that metal contamination to lakes in the Schefferville region has varied both spatially and temporally (Aebischer *et al*., 2015), the aquatic biodiversity responses may also be dynamic and thus we quantified the changes in alpha and beta diversity from two lakes with distinct histories. Whereas many investigators have analysed alpha diversity trends (often measured as the number of species (richness) in each lake), beta diversity is emerging as a more sensitive diversity metric as it characterizes the degree of change in species composition that may occur without changes in the total number of species present. Keeping in mind the unique characteristics of the Schefferville systems, such as absence of acidification, combined presence of metal loading and nutrient inputs, and discontinuity of mining activities in the region, this study focused on the following main objectives:

(1) To characterize the effect of metal loading on zooplankton (cladoceran) species richness and composition, in sites where acidification was not a simultaneous stressor.

(2) To examine the extent to which patterns in cladoceran community species richness through time are congruent (or not congruent) with temporal beta diversity.

To our knowledge, the examination of temporal beta diversity (i.e. the turnover in assemblages, measured over multiple time points in a sediment record) to quantify cladoceran responses to mining is a novel application of this relatively new method. We hypothesized that metal loading would decrease cladoceran species richness in these lakes as metal-sensitive species are pushed towards the limits of their tolerance and due to direct toxicity effects. However, there could be

site-specific differences in cladoceran responses because the study lakes vary in their metal and nutrient loading histories.

## Methods

*Description of study site and field sampling*

The Labrador Trough region, which straddles Québec and Labrador (Fig. 1a), has long been part of the traditional territory of both the Naskapi and Innu (Innu-Montagnais) peoples (Boutet, 2012). These lands were used frequently for travel, trapping and hunting, although a permanent settlement in Schefferville did not occur until (often forced) relocations in the early 1950s; Naskapi came from Kuujjuaq and the Ungava region, and Innu came from Maliotenam (near Sept-Îles QC). The history surrounding the relocation of both the Naskapi/Iyiyiw and Innu and subsequent negotiation and establishment of the reserve communities of Kawawachikamach, Matimekush, and Lac John is as important as it is complex (Boutet, 2012). The Iron Ore Company of Canada (IOC) began ramping up mining operations between 1939 and 1947 (Laperrière *et al*., (2008); Aebischer *et al*., 2015) and the town of Schefferville was built in 1954 by the IOC. The first official shutdown of mining operations occurred between 1977 and 1982 (Laperrière *et al*., 2008; Aebischer *et al*., 2015). A timeline of both town development and mining activities in the region is shown in Appendix S1.

Our two study lakes, Lake Dauriat and Lake Knob, are situated within the town site of Schefferville (Fig. 1b; Table 1). Lake Knob is upstream (southeast) of the town site and serves as the drinking water source for Schefferville. A small outflow from Lake Knob connects it to Lake Dauriat, which is situated northwest and downstream of the town site. Lake Dauriat received raw sewage effluent from the town from its inception until 1975 when a water treatment plant was

built, and continues to receive both treated wastewater and surface runoff during rain events. Laperrière *et al*., (2008) used sediment cores from Lake Dauriat in the centre of Schefferville to show paleolimnological evidence for negative effects of both mining activities and the discharge of sewage directly into this ecosystem. They found pronounced changes in diatom community composition and increases in diatom-inferred total phosphorus (DITP) from the 1940s to late 1970s (which partly overlapped a monitoring record from the lake (Choulik & Moore, 1991)), followed by partial recovery (Fig. 2c).

Based on analyses of epiphytic lichens growing on trees in the area, as well as sediment cores taken from Lake Dauriat and a lake farther from the town (Lake Oksana), Aebischer *et al*., (2015) showed that the concentration of Pb decreases linearly (and significantly) with increased distance from both the Schefferville town site and active mining sites. This work demonstrated that lakes have received significant metal loading from both mining projects as well as the town itself, in addition to a background loading of elements like Pb from distant sources via atmospheric deposition (e.g. leaded gasoline in the past, distant smelter activities). Furthermore, this same study used variation in stable isotope ratios for lead, iron and zinc to delineate separate sources of metal loading to the Schefferville lakes. They found four distinct sources of metal loading (for Pb, Fe and Zn) to the lakes: geogenic (occurring prior to mining in this region), town development (corresponding to the mining period), town development after the mining period, and long distance atmospheric transport.

Field sampling occurred between 24-30 July 2012 and 4-14 September 2013. In July 2012, we collected an approximately 40 cm long (diameter = 6.5cm) sediment core from both Lakes Dauriat and Knob using a Maxi-Glew gravity corer, at the deepest observed part of each lake. We sectioned these cores at 0.25 cm intervals and froze these samples for transport back to

the laboratory. In September 2013, we used a gravity corer to collect 20-40 cm long sediment cores along a ~50 m transect in each lake. Starting at the observed $Z_{max}$, we collected one core at 5 m intervals across the lake, to a maximum of six cores for each lake. From these cores, one from each lake was sectioned at 0.25 cm intervals and frozen, while the remaining cores were sectioned at 2 cm intervals before freezing.

*Radiometric dating and heavy metal geochemistry*

The 2012 cores from Lakes Dauriat and Knob served as references for chronology and geochemical analyses. We first freeze-dried these cores and then measured magnetic susceptibility on each of the 0.25 cm sample slices by subsampling a constant volume. We then selected 15 evenly spaced intervals in each core for radiometric dating ($^{210}$Pb), and sent these to the GEOTOP facility at the University of Québec at Montreal. The selection of age models based on $^{210}$Pb, $^{226}$Ra, and $^{137}$Cs activity is outlined in Appendix S2.

An additional 15 intervals were selected for geochemical analyses (again evenly spaced and adjacent to the 15 intervals selected for dating). These dried samples were analyzed at a commercial laboratory, Actlabs (Ancaster, Ontario, Canada). Using inductively coupled plasma-optical emission spectrometry (ICP-OES), concentrations of the following heavy metals were obtained for each interval: Ag, Al, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, Li, Mg, Mn, Mo, Na, Ni, P, S, Sb, Sc, Sr, Te, Ti, U, V, W, Y, Zn, and Zr. Mercury concentrations were obtained through cold vapour Flow Injection Mass Spectrometry (FIMS). Total organic carbon (TOC) was quantified for these samples using a Carbon-Sulphur combustion analyzer.

*Subfossil cladocera*

For each interval selected, slides of cladoceran subfossils were prepared following a procedure adapted from Korhola and Rautio (2001). Approximately 0.1 g (this mass was increased for intervals with low subfossil abundance) of sediment was digested in 50 mL of potassium hydroxide (KOH) for 30 minutes at 65 ºC. The solution was manually stirred every few seconds. The solution was removed from heat and 5 mL of 10% hydrochloric acid (HCl) was added to eliminate carbonates, before sieving through a 36 µm mesh. Sediments were washed for 15 minutes to remove any remaining dissolved organic matter and transferred to a Falcon tube using as little water as possible for a total volume between 5 and 10 mL. Permanent slides were then prepared by pipetting 0.05 mL of cleaned slurry material, which was then fixed with glycerine-safranin jelly. Kurek *et al*., (2010) established that a minimum of 70 to a 100 individuals must be counted to reliably represent a cladoceran assemblage, though it was also noted that 50 individuals may be an appropriate count number in species-poor lakes to record the majority of taxa with relative abundance over 1% of the total assemblage (Kurek *et al*., 2010). As such, while we attempted to reach a minimum of 70 individuals per interval, some low abundance intervals were only counted to 50 individuals.

Cladoceran remains typically preserved in lake sediments are the chitinized body parts: carapaces, post-abdomens, post-abdominal claws, head shields and mandibles (Korhola & Rautio, 2001). For each species or species group, subfossil remains were counted and the most frequently occurring subfossil was used to calculate species abundance for each interval. However, there were some cases where we were unable to resolve subfossils to species. For example, in some samples we found *Bosmina* spp. carapaces and head shields with their pore location covered and therefore could not resolve these individuals to species. In these cases, we

split this aggregated count between *Bosmina longirostris* and *Eubosmina longispina* based on the proportion of individuals from these species that had already been accurately identified in that sample. In the same way, carapaces of large *Alona* species that could not be separated were split between *Alona affinis* and *Alona quadrangularis*. Poorly preserved *Chydorus* spp. individuals were also split proportionately between the identified chydoriids in a sample. Identifications were based on the following taxonomic guides: Frey (1959 & 1962), Megard (1967), Sweetman & Smol (2006), Szeroczynska & Sarmaja-Korjonen (2007), and Korosi & Smol (2012 [a and b]).

*Statistical analyses*

Taxa richness

All statistical analyses were completed in R version 3.1.2 (R Core Team, 2014). We calculated rarefied taxon richness for each interval to determine richness standardized by the minimum number of individuals counted in any of the intervals, using *rarefy()* in `vegan` (Oksanen, 2015).

Beta diversity

We quantified temporal beta diversity between each time interval within each lake core, computing temporal beta diversity directionally such that the oldest time point was compared to the second oldest time point; the second oldest time point compared to the third oldest and so on. We used the R function *TBI*() developed by Pierre Legendre (Legendre, 2015), to produce a measure of assemblage differentiation between each of the intervals based on the relative abundances of taxa recorded at each interval. We used the percentage difference option in the *TBI*() function which computes dissimilarity using the quantitative (not presence-absence) form of the Sørensen index, also called a Bray-Curtis index (Legendre, 2015). We also extracted the proportion of total temporal beta diversity that was explained either by the loss or gain of

species' abundances (Legendre & Salvat, 2015), to better understand mechanisms behind the temporal differentiation of assemblages. For example, a comparison where total temporal beta diversity was explained mostly by a loss component would mean that the assemblage change occurring between time intervals at that point is mostly the result of individual species experiencing decreased relative abundance.

Linear mixed effect models

To test the effect of metal loading on cladoceran taxa richness, we employed a mixed effect modelling approach to allow lake identity to be included as random factors in the model. This meant that we pooled the data from both lakes to test the relationship between metals and cladoceran assemblages. The response variables considered for these models were rarefied taxa richness, as well as the first principal component (PC1) scores from a Principal Components Analysis (PCA) of the cladoceran relative abundance data. We Hellinger-transformed relative abundance values for the cladoceran assemblages and performed a PCA using *rda()* in `vegan`. We used *scores()* (`vegan`) to extract the "site" scores for the first principal axis for this PCA, though these represent scores for each individual interval (lake/time combination) as opposed to individual sites.

The explanatory variables for these models were either a cumulative metal enrichment factor (EF) or the PC1 from a PCA of heavy metal concentrations. We calculated an EF for each heavy metal using the following formula (from Bhuiyan *et al*., 2010):

$$\frac{([\text{Element}]_{\text{Sample X}} / [\text{Al}]_{\text{Sample X}})}{([\text{Element}]_{\text{Background}} / [\text{Al}]_{\text{Background}})}$$

where each element's EF is calculated with reference to aluminum concentrations, and 'Background' refers to the oldest interval in a core. After calculating each element's individual EF, we then summed the EFs to result in total EFs, keeping the background EF as a constant. We

performed a PCA on heavy metal data for the following elements: Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, S, Sb, Sc, Sr, Te, V, Y, Zn, and Zr (standardized by aluminum). We again extracted "site" scores from the PCA to represent scores for each individual sample (interval).

We tested the relationship between metal loading and enrichment on cladoceran taxa richness with the models outlined in S3, using *lmer*() in the R package `lme4` (Bates *et al*., 2015). We then assessed each model via Akaike information criteria *AICc*() from the package `AICcmodavg` (Mazerolle, 2015). Because the observations from our two cores have the potential to be significantly autocorrelated (samples are from the same cores but different time points), we used the method outlined in Simpson and Anderson (2009) to test for autocorrelation amongst our intervals. In particular, we constructed linear models using *gls*() in `nlme` (Pinheiro *et al*., 2015), including one model with the CAR1 correlation structure (based on year estimates for each interval). We then used an ANOVA to test for significant differences between the linear model without the autocorrelation structure included and the model with autocorrelation structure.

## Results

### *Geochemical loading to the four study lakes*

Dauriat and Knob lakes showed contrasting histories in terms of metal loading and other abiotic drivers (see raw plots in Appendix S4 for details). Metal loading to Lake Dauriat varied greatly through time, with the elements iron, aluminum, copper, cobalt and nickel showing both pronounced peaks and troughs during pre-mining and during mining time periods. Metal concentrations in the Dauriat sediments generally exceeded both International Quality Guidelines

for the Protection of Aquatic Life (ISQG) as well as Probable Effect Levels (PEL) (Canadian Council of Ministers of the Environment, 2014) (Table 2). Temporal variation in specific metals measured from the Lake Dauriat core, combined with relatively constant background levels of other elements (such as arsenic, cadmium, mercury and zinc), resulted in considerable variation in the scores from the 1st PC axis of the PCA of metal concentrations (Fig. 2). Metal enrichment in the Dauriat record peaked in the 1960s, followed by the second highest peak in the 1980s when compared to historical conditions. Metal loading in Lake Knob was less pronounced, despite the presence of the regional metal stressors. The majority of heavy metals measured in Knob remained fairly constant through time with the exception of manganese, which showed peaks in recent intervals. Metal concentrations in the Knob sediments generally exceed ISQG levels, but do not always exceed PELs.

*Biodiversity*

Across the two lakes, we enumerated a total of 37 taxa. However, many of these taxa were relatively rare and once we removed taxa that did not make up at least 5% of a sample's assemblage in at least one interval in two lakes, we were left with 14 taxa (Figs. 3 & 4). In Dauriat, cladoceran species richness showed considerable fluctuations through time, with a fairly extensive reduction in rarefied richness circa 1960s (Fig. 5). Species richness then increased, but only the most recent sediment interval comes close to the pre-mining values. In comparison to Lake Dauriat, cladoceran species richness trends from Lake Knob were relatively stable, even during the mining period (Fig. 5). Species richness in the Lake Knob record was highest at the start of extraction activities, followed by fluctuations in richness during and after the mining period. Estimated sedimentation rates were considered constant through time in both Lake Knob

and Lake Dauriat because of the selection of a Constant Initial Concentration age model, and as such cladoceran species richness could not be correlated to estimates of sedimentation rate.

The community assemblages of the two lakes were different and remained so through time (Fig. 6). In Dauriat, Bosminids (*Bosmina longirostris* and *Eubosmina longispina*) maintained their relative abundances throughout the mining period, whereas taxa such as *Alona circumfimbria* and *Alona* spp. became much more abundant after the main mining period (Fig. 3). In Knob, *B. longirostris* and *E. longispina* were also quite abundant, with other species present in low relative abundances throughout the core profile (Fig. 4). In comparing both records with a PCA, we observed that the earliest Dauriat assemblages (1920/1930s) were most similar to the Knob assemblages, with more recent Dauriat assemblages differing substantially from Knob by the prevalence of *Alona* spp., and *A. circumfimbria*. Lake Knob assemblages tended to be characterized by taxa such as *Alona affinis* and *Acroperus harpae*. *Chydorus* cf. *sphaericus* was found in intervals throughout both the Lake Dauriat and Knob records, although it was present in higher abundance in the more recent samples of the Knob record. The PCA also demonstrates that there was less variation in cladoceran assemblages through time in Lake Knob record relative to that of Lake Dauriat.

The most pronounced changes in temporal beta diversity were observed in the Lake Dauriat record compared to Lake Knob (Fig. 7). However, beta diversity reached peaks during the period of mining and town construction in both lake records (from the late 1930s to the 1970s). In both Lake Dauriat and Knob, changes in total beta diversity were largely explained by a loss of species (i.e., the loss component which measures the changes in abundances on a species by species basis).

*Cladoceran response to metal loading*

The best fit model for the relationship between cladoceran species richness and metal loading was the simple linear model of $Clad\_S_i \sim Metal\ EF_i + \varepsilon$ without lake included as a random factor (lowest AIC as well as the highest explained variation (Adj. $R^2 = 0.69$)) amongst the null linear models; Table 3; Fig. 8a). The second lowest AIC values were from the following models: 1) $Clad\_S_{ij} \sim Metal\ EF_i + Lake_j + \varepsilon$ with lake included as a random factor (Fig. 8b). For both these latter models, the slope for the mixed effect model was significantly different from zero, as indicated by the range of upper and lower confidence intervals (i.e., did not include zero)-0.05 and -0.08, respectively). Although we found that measurements on adjoining sediment intervals were autocorrelated ($\phi = 0.6$), we failed to detect a significant difference based on an ANOVA of the linear model run with autocorrelated structure, and the model run without accounting for this autocorrelation. The set of models with cladoceran richness as a response and metal PC1 as an explanatory variable had higher AIC and a lower Adj. $R^2$ value than the models with metal EF as an explanatory variable (Table 3). Linear models with cladoceran PC1 as a response variable explained very low amounts of variation (Adj. $R^2 = 0.09$ with metal EF as an explanatory variable and 0.02 with metal PC1 as an explanatory variable).

**Discussion**

Examining the effects of metal contamination on biodiversity dynamics within the Schefferville region demonstrates the complexity associated with understanding aquatic community and ecosystem responses in a multiple stressor context. Both Lake Dauriat and Lake Knob are situated within the limits of the same town site and are exposed to the same sources of atmospheric metal deposition. Despite this similarity, peaks in heavy metals and metal

enrichment varied considerably between these two lakes, with Dauriat experiencing higher sediment metal concentrations enrichment values compared to background conditions (where T = 0, bottom of sediment core). Based on isotopic analyses, Aebischer *et al*., (2015) have recently shown that the sediments of Lake Dauriat track a variety of pollutant sources, including those from mining activities as well as municipal waste and sewage. However, given that the primary source of metal loading to Lake Knob was through an atmospheric pathway and that this lake record contained numerous sediment intervals that were above quality guidelines for aquatic life, the extent of metal enrichment derived from mining sources in both these lakes appears to be ecologically important. Consistent with our hypothesis, we found that higher metal loading was associated with decreased cladoceran richness. Secondly, we revealed that the contribution of loss and gain components to temporal beta diversity differ through time and between lakes. To further advance knowledge concerning zooplankton responses to metal contamination, we have developed a conceptual framework for explaining the cladoceran richness and temporal beta diversity trends.

Based on our modeling work, we found that metal contamination had an overarching effect on cladoceran species richness. Specifically, we found that a simple linear relationship between cladoceran species richness and metal enrichment factor was the most robust model. The observed negative effect of metal loading on cladoceran richness is consistent with the reported sensitivity of zooplankton to increased heavy metal concentrations based on many laboratory toxicity studies (e.g., Biesinger & Christensen, 1972; Bossuyt & Janssen, 2005, among others). This effect did not appear to have been mitigated in Lake Dauriat by other factors (e.g. nutrient inputs that one might have predicted would have helped cladoceran taxa to persist despite metal contamination), as troughs in cladoceran richness in Lake Dauriat were lower than

that of Lake Knob. While sedimentation rate may also have an effect on these results, we think that these results are not simply due to changes in sedimentation rates (as per Smol 1981), because the best age model for Lake Dauriat was one where sedimentation rate was held relatively constant (and even when a different age model was used (i.e. CRS), we still failed to find a significant effect of sedimentation rates on cladoceran species richness).

Community assemblage records from our two study lakes showed some similarity to studies from other mining regions. In particular, the dominance of *Bosmina longirostris* and *Eubosmina longispina* that we observed in both study lakes is similar to the findings reported from other metal contaminated lakes in northern Russia, eastern Canada and the Sudbury mining region in Canada (Lukin *et al*., 2003; Korosi *et al*., 2012; and Labaj *et al*., 2015). In these other mining regions, the prevalence of these bosminid taxa has often been attributed more to their ability to withstand acidification than metal tolerance per se, as *B. longirostris* is considered more tolerant to acidification than other cladoceran species (Doig *et al*., 2015). However, Doig *et al*., (2015) found that *Bosmina* spp. actually decreased drastically at the start of industrial development in a study of a lake in northern Manitoba (Canada), where measures were taken to mitigate pH changes that occurred following the start of industrial activities. In the case of Doig *et al*., (2015), it seems that the decrease in *Bosmina* spp. was driven by both acidification and metal contamination. In our two study lakes though, acidification was not a factor and the absence of this additional stressor may have allowed *Bosmina* spp. to show considerable tolerance to the metal contamination experienced in Lakes Dauriat and Knob.

For some taxa, there is a strong need for more clearly defined autecological information and laboratory studies could help to advance this goal. For example, our lakes also showed the consistent presence of *Chydorus* cf. *sphaericus*, a taxon that is generally associated with

macrophytes and mud (Chen *et al*., 2010) or moderately acidified waters (Belyaeva & Deneke, 2007). Based on this literature and our work, it is reasonable to suggest that this *Chydorus* species can also survive metal-contamination, without accompanying acidification, but this should be tested experimentally. For other taxa, their response to metal contamination and/or acidification is rather unclear from the literature. *Holopedium gibberum* has been observed to be present in metal-contaminated and acidified lakes of the nickel-mining region of Sudbury (Canada; Valois *et al*., 2011). However, Yan *et al*., (2004) found that *H. gibberum* was twice unsuccessful in colonizing a previously acidified and metal contaminated Sudbury lake, even after remediation through liming. While we also observed *H. gibberum* in one of our samples from Lake Knob (during the mining period), this species was present only at a very low relative abundances (less than 5% of the total assemblage), suggesting that *H. gibberum* was not tolerant of the disturbance conditions in Lakes Dauriat and Knob. Here again, experimental research would help to clearly elucidate the tolerance of this taxon to a variety of stressors.

While the relationship between metal enrichment and cladoceran diversity was similar between lakes, there are some key differences with respect to biotic and abiotic characteristics. The magnitude of metal contamination has been much higher in Lake Dauriat, and its cladoceran assemblages experienced more drastic reductions in taxon richness. Additionally, a preliminary investigation of the cladoceran resting egg bank in the two lakes also shows other potential differences between the cladoceran dynamics. For example, we have found some evidence that the cladoceran resting egg bank of Lake Dauriat shows a greater ratio of unhatched: empty ephippial cases relative to Lake Knob (Appendix S5). The predominance of empty (possibly hatched) cases as a function of the total cladoceran resting egg bank in Lake Knob may be due to greater hatching success in the lake with less severe contamination, although other mechanisms

are also possible (e.g., poorer preservation of eggs in Lake Knob). One possible explanation for the observed resting egg results is that the combination of metal-contamination and eutrophication in Lake Dauriat led to a depression in hatching success of cladoceran resting stages, whereas hatching of ephippia has continued through time in Lake Knob. Clearly, a more thorough investigation of these dynamics is necessary in order to directly compare the deposition rates of cladoceran resting eggs in these two lakes, and experimental work could be insightful to evaluate the support for different potential mechanisms.

As we have seen, the impact of stressors on aquatic systems rarely happen in isolation. Understanding the effect and mechanism of a single stressor does not necessarily aid in delineating whether its effect will be additive, antagonistic or synergistic when occurring with other stressors (Burton & Johnston, 2010). In the case of multiple stressors in mining regions, the effect of metal loading or contamination on aquatic biodiversity is affected by the bioavailability of metals, toxicokinetics, and the tolerance of organisms (Heugens *et al*., 2001), as well as non-metal related stressors. Mining activities can produce a legacy of heavy metals, as well as result in the release of many other pollutants and toxicants through the various production processes including land-use change and infrastructure development (Bernhardt & Palmer, 2011). This is exemplified by the history of Lake Dauriat, where metal contamination has come both from resource extraction and town construction; with the latter also causing substantial eutrophication. Episodic deepwater anoxia associated with this eutrophication may have contributed to the finding that cladoceran richness declines in Lake Dauriat were stronger in this site relative to Lake Knob. Local knowledge as well the lithology of the Lake Dauriat cores suggests an anoxic period from the 1950s to 1970s. During this time, metal deposition to Lake Dauriat may have had an amplified impact on biodiversity because eutrophication and subsequent deepwater

anoxia may have altered redox reactions within the sediments, making metals in the sediments more soluble and allowed them to diffuse into overlying waters (e.g. as shown in Jacobs and Emerson, 1982) and thus potentially exposing zooplankton a second time or to a more toxic form.

Temporal beta diversity varied throughout the history of both cores and consistently showed peaks during the mining period. While these peaks were mostly attributable to loss in abundance on a taxa by taxa basis (taxa loss; and this was overwhelmingly the case in Lake Dauriat), one peak in Lake Knob was attributable to gain in abundance on a taxa by taxa basis (taxa gain). Lake Knob showed two incidences of higher total beta diversity during the mining period (during the 1960s and 1970s). The first of these peaks was explained predominantly by taxa gain, whereas the point immediately following was explained predominantly by taxa loss (Fig. 7). This apparent switching in terms of explanatory components may have been due to a lag in response of cladoceran taxa in Lake Knob, or a consequence of early mining activities; the initial start to the mining period may have resulted in some taxa increasing in abundance in Lake Knob as new niches for metal- and disturbance-tolerant taxa become available, before metal-sensitive taxa were very affected. This would increase total temporal beta diversity, primarily via the establishment and proliferation of metal-tolerant species. As mining activities continued however, associated metal enrichment and disturbance could have reduced relative abundances of non-metal tolerant species and perhaps also of generalist taxa, resulting in sustained high temporal beta diversity, but this time predominantly via taxa loss (as seen for the second beta diversity peak during the mining period in Lake Knob; Fig. 7). This latter mechanism is also the most plausible in Lake Dauriat, where taxa loss consistently explained total temporal beta diversity. Mechanisms aside, this "switching" between dominant taxa loss and gain components

in temporal beta diversity has been demonstrated in at least one other study. Legendre and Salvat (2015), found that a combination of niche and neutral processes structured variation in their study of temporal beta diversity of mollusc communities, as evidenced by the alternation of taxa loss or gain in temporal beta diversity. Thus, variation in the main explanatory component of temporal beta diversity may provide key insights into the dynamics of a system. In the case of Lake Knob, we can see that the early versus mid-mining period responses in cladoceran communities differed as reflected by the explanatory components of beta diversity (Fig 7).

There are at least a few different conceptual ways in which the cladoceran richness and beta diversity findings can be interpreted in relation to each other (Fig. 9). For example, one might predict that if the gain component of beta diversity (pseudo species replacement) dominated, then temporal beta diversity would not track alpha diversity during periods of stress (i.e. metal stress in this case, assuming that metal loading would cause decreases in alpha diversity; Fig. 9a). Temporal beta diversity in these lakes had the potential to be either (relatively) high or low between each time point of comparison. High temporal beta diversity could indicate strong gain or loss components individually or in combination. If the dominant explanatory component was loss, reductions in richness should occur, as species are lost between intervals (via reduced abundances which make species extirpations more demographically likely; Fig. 9b). Conversely, high temporal beta diversity could indicate a strong taxa gain component, and could be associated with either no net change of richness between intervals (if species abundances for those already present increased, as opposed to new species being added), or an increase in richness as new species colonized over time (Fig. 9c). In our study, cladoceran taxa richness was less under high metal enrichment conditions (mixed effect model), and temporal beta diversity was highest during time periods of maximum (or near maximum) metal

enrichment. Our results thus indicate that taxa abundance loss made up the majority of total beta diversity during metal enrichment. Consideration of richness and beta diversity in this way is important, as it will improve ecologists' understanding of biodiversity trends that can be masked when only considering richness or alpha diversity

In a review of freshwater systems subjected to a wide variety of anthropogenic stressors, Niemi *et al*., (1990) documented 150 case studies where aquatic assemblages (from plankton to fish) exhibited some form of resilience to a disturbance. They found that the majority of these trended towards biological recovery in less than three years, as long as, 1) the disturbance or stressor did not physically alter aquatic habitat, 2) there were no residual pollutants remaining from the disturbance, and 3) the system was not isolated to the extent that recolonization by organisms could not occur. In our study, cladoceran communities in Lake Knob show relatively little change until the onset of the mining period (Fig. 7). For Lake Dauriat, the majority of assemblage turnover occurred during the mining period due to reductions in taxa richness associated with increased metal contamination (Fig. 7). However, even though mining activities were terminated in this region by the early 1980s, the cladoceran assemblages in Lake Dauriat have not returned to their pre-mining state, possibly because lake sediments are still metal-enriched, which may be contributing to further release of metals to the water-column. Climate change is an unlikely mechanism in this area as temperature records show very little change over the study period (Environment and Climate Change Canada, 2015). As such, the characteristics leading to successful recovery of aquatic systems laid out in Niemi *et al*., (1990) are not fully met in the Lake Dauriat system. Additionally, the Niemi *et al*., (1990) study did not explicitly examine the unique effect of multiple concurring stressors, or the effect of temporally re-occurring stressors. This was the specific intent of Moe *et al*., (2013), wherein the authors

reviewed case studies examining the interaction of climate change and toxicants, and their effects on aquatic biodiversity. They found that species and communities may be more susceptible to multiple stressors because adaptation to a single stressor often results in trade-offs that may render a species or community more susceptible to other stressors. The cladoceran assemblages in Lake Dauriat have been subject to a combination of both metal contamination and eutrophication, and these dual stressors amplify the lasting effect on cladocera in this lake.

Understanding both the effect of metal loading on cladocerans in this region, as well as their propensity for temporal turnover is important because the extent of mining continues to vary. As such, biological assemblages that might represent 'recovery' in this system post-mining are experiencing evolving stressor combinations. The strength of our study is really in our use of a temporal beta diversity lens to quantify cladoceran turnover over a 100-300 year period across a major perturbation event. In doing this, we have shown that cladoceran communities were considerably dynamic, despite a linear relationship with metal contamination. Understanding this dynamism at the community-level will provide insight into understanding evolutionary dynamics and rapid adaptation of cladoceran taxa at the species level to metal stressors (such as that demonstrated in Turko *et al*., 2016).

**Acknowledgements**

We wish to thank Kathryn (Yici) Han for her work with the ephippia removal, DNA extractions and preparation of samples to be sent for DNA barcoding, Alessandra Loria for advice with DNA extractions as well as Julien Gogot and Bassam Ghaleb (UQÀM – GEOTOP) for radiometric dating and assistance developing age models. Jean-Philippe Jenny also provided key insight relating to age models. Thank you to Paul del Giorgio and the CarBBAS lab for help

during the 2012 field season, and to Oksana Choulik at the McGill Subarctic Research Station for her hospitality and insight into Schefferville's history. This research was supported by an NSERC postgraduate scholarship to AKW, the Northern Scientific Training program (grants to AKW and NS), a QCBS Excellence Award to AKW, an NSERC USRA to NS, an NSERC Discovery Grant to IGE and a FRQNT Équipe grant to IGE, BEB, AD and Pierre Legendre.

**Literature Cited**


Adams P. (2007). Trent, McGill, and the North: A story of Canada's growth as a sovereign polar nation. Cover to Cover Publication Services: Peterborough, Ontario, Canada.

Aebischer S., Carignan J., Cloquet C., Maurice C., & Pienitz R. (2013). Le cycle géochmique des métaux de surface en période d'exploitation minière: Étude isotopique des sediments lacustres de Schefferville (rapport final). Ministère des Ressources Naturelles Québec: GM 66439.

Aebischer S., Cloquet C., Carignan J., Maurice C., & Pienitz R. (2015). Disruption of the geochemical studies of lake sediments from Schefferville, subarctic Québec. *Chemical Geology*, **412**, 167-178.

Bates D., Maechler M., Bolker B., & Walker S. (2015). Fitting Linear Mixed-Effecs Models using lme4. *Journal of Statistical Software,* **67**, 1-48.

Belyaeva M., & Deneke R. (2007). Colonization of acidic mining lakes: *Chydorus sphaericus* and other Claoceran within a dynamic horizontal pH gradient (pH 3-7) in Lake Senftenberger See (Germany). *Hydrobiologia*, **594**, 97-108.

Bernhardt E.S., & Palmer M.A. (2011). The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians. *Annals of the New York Academy of Sciences,* **1223**, 39-57.

Bhuiyan M.A.H., Parvez L., Islam M.A., Dampare S.B., & Suzuki S. (2010). Heavy metal pollution of coal mine-affected agricultural soils in the northern part of Bangladesh. *Journal of Hazardous Materials*, **173**, 384-392.

Biesinger K.E., & Christensen G.M. (1972). Effects of various metals on survival, growth, reproduction, and metabolism of *Daphnia magna*. *Journal of Fisheries Research Board of Canada,* **29**, 1691-1700.

Bossuyt B.T.A., & Janssen C.R. (2005). Copper toxicity to different field-collected cladoceran species: intra- and inter-species sensitivity. *Environmental Pollution*, **136**, 145-154.

Boutet J.S. (2012). An Innu-Naskapi ethnohistorical geography of industrial mining development at Schefferville, Québec. MA thesis, Department of Geography, Memorial University.

Burton G.A., & Johnston E.L. (2010). Assessing contaminated sediments in the context of multiple stressors. *Environmental Toxicology and Chemistry*, **29**, 2625-2643.

Canadian Council of Ministers of the Environment. (2014). Canadian Environmental Quality Guidelines. Available: http://ceqg-rcqe.ccme.ca/en/index.html#void. Accessed 01 November 2015.

Chen G., Dalton C., & Taylor D. (2010). Cladocera as indicators of trophic state in Irish lakes. *Journal of Paleolimnology*, **44**, 465-481.

Choulik O., & Moore T.R. (1991). Response of a subarctic lake chain to reduced sewage loading. *Canadian Journal of Fisheries and Aquatic Sciences,* **49**, 1237-1245.

Dillon P.J., Reid R.A., & de Grosbois E. (1987). The rate of acidification of aquatic ecosystems in Ontario, Canada. *Nature,* **329**, 45-48.

Doig L.E., Schiffer S.T., & Liber K. (2015). Reconstructing the ecological impacts of eight decades of mining, metallurgical, and municipal activities on a small boreal lake in Northern Canada. *Integrated Environmental Assessment and Management,* **11**, 490-501.

Environment and Climate Change Canada. (2015). Climate Trends and Variations Bulletin. Available online: https://www.ec.gc.ca/sc-cs/default.asp?lang=En&n=1F942323-1

Frey D.G. (1959). The taxonomic and phylogenetic significance of the head pores of the Chydoridae (Cladocera). *International Revue der gesamten Hydrobiologie und Hydrographie,* **44**, 27-50.

Frey D.G. (1962). Supplement to: The taxonomic and phylogenetic significance of the head pores of the Chydoridae (Cladocera). *International Revue der gesamten Hydrobiologie und Hydrographie*, **47**, 603-609.

Gray D.K., Arnott S.E., Shead J.A., & Derry A.M. (2012). The recovery of acid-damaged zooplankton communities in Canada Lakes: the relative importance of abiotic, biotic and spatial variables. *Freshwater Biology,* **57**, 741-758.

Hamilton P.B., Lavoie I., Alpay S., & Ponader K. (2015). Using diatom assemblages and sulfur in sediments to uncover the effects of historical mining on Lake Arnoux (Quebec, Canada): a retrospective of economic benefits vs. environmental debt. *Frontiers in Ecology and Evolution,* **3**, doi 10.3389/fevo.2015.00099

Hebert P.D.N., Cywinska A., Ball S.L., & deWaard J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313-321.

Heugens E.H.W., Hendriks A.J., Dekker T., van Straalen N.M., & Admiraal W. (2001). A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. *Critical Reviews in Toxicology,* **31**, 247-284.

Jacobs L., & Emerson S. (1982). Trace metal solubility in an anoxic fjord. *Earth and Planetary Science Letters,* **60**, 237-252.

Keller W., Yan N.D., Somers K.M., & Heneberry J.H. (2002). Crustacean zooplankton communities in lakes recovering from acidification. *Canadian Journal of Fisheries and Aquatic Sciences,* **59**, 726-735.

Korhola A., & Rautio M. (2001). Cladoceran and other branchiopod crustaceans. In Smol, J.P., Birks, H.J.B., and Last, W.M. (Eds). "Tracking Environmental Change Using Lake Sediments", Kluwer, Dordrecht.

Korosi J.B., Ginn B.K., Cumming B.F., & Smol J.P. (2012). Establishing past environmental conditions and tracking long-term environmental change in the Canadian Maritime provinces using lake sediments. *Environmental Reviews*, **21**, 15-27.

Korosi J.B., & Smol J.P. (2012a). An illustrated guide to the identification of cladoceran subfossils from lake sediments in northeastern North America, part 1: the Daphniidae, Leptodoridae, Bosminidae, Poyphemidae, Holopedidae, Sididae, and Macrothricidae. *Journal of Paleolimnology,* **48**, 571-586.

Korosi J.B., & Smol J.P. (2012b). An illustrated guide to the identification of cladoceran subfossils from lake sediments in northeastern North America, part 2: the Chydoridae. *Journal of Paleolimnology,* **48**, 587-622.

Labaj A.L., Kurek J., Jeziorski A., & Smol J.P. (2015). Elevated metal concentrations inhibit biological recovery of Cladocera in previously acidified boreal lakes. *Freshwater Biology*, **60**, 347-359.

Legendre P. (2015). R function. TBI(). Available online: http://adn.biol.umontreal.ca/~numericalecology/FonctionsR/

Legendre P., & Salvat B. (2015). Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.

Lukin A., Dauvalter V., Kashulin N., Yakovlev V., Sharov A., & Vandysh O. (2003). Assessment of copper- nickel industry impact on a subarctic lake ecosystem. *The Science of the Total Environment,* **306**, 73-83.

Matthaei C.D., Piggott J.J., & Townsend C.R. (2010). Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology,* **47**, 639-649.

Mazerolle M.J. (2015). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.0-3. Available: http://CRAN.R-project.org/package=AICcmodavg

Megard R.O. (1967). Three new species of *Alona* (Cladocera, Chydoridae) from the United States. *Internationale Revue der gesamten Hydrobiologie und Hydrographie,* **52**, 37-50.

Merriam E.R., Petty J.T., Merovich Jr, G.T., Fulton J.B., & Strager M.P. (2011). Additive effects of mining and residential development on stream conditions in a central Appalachian watershed. *Journal of the North American Benthological Society,* **30**, 399-418.

Moe S.J., De Schamphelaere K., Clements W.H., Sorensen M.T., Van den Brink P.J., & Liess M. (2013). Combined and interactive effects of global climate change and toxicants on populations and communities. *Environmental Toxicology and Chemistry,* 32, 49-61.

Natural Resources Canada. (2014). "Minerals and Mining Map", The Atlas of Canada. Available online: http://atlas.gc.ca/mins/en/index.html. Accessed: 27 August 2015.

Niemi G.J., DeVore P., Detenbeck N., Taylor D., Lima A., Pastor J., Yount J.D., & Naiman, R.J. (1990). Overview of case studies on recovery of aquatic systems from disturbance. *Environmental Management*, **14**, 571-587.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., & Wagner H. (2015). Package 'vegan'. Available: http://cran.r-project.org/web/packages/vegan/vegan.pdf

Pinheiro J., Bates D., DebRoy S., Sarkar D., & R Core Team. (2015). Nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-122. Available: https://cran.r-project.org/web/packages/nlme/index.html.

Prowse T.D., Furgal C., Chouinard R., Melling H., Milburn D., & Smith S.L. (2009). Implications of climate change for economic development in Northern Canada: Energy, resource, and transportation sectors. *A Journal of the Human Environment,* **38**, 272-281.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org/

Simpson G.L., & Anderson N.J. (2009). Deciphering the effect of climate change and separating the influence of confounding factors in sediment core records using additive models. *Limnology and Oceanography,* **54**, 2529-2541.

Smol J.P. (1981). Problems associated with the use of "species diversity" in paleolimnological studies. *Quaternary Research*, **15**, 209-212.

Strayer D.L., & Dudgeon D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society,* **29**, 344-358.

Sweetman J.N., & Smol J.P. (2006). A guide to the identification of cladoceran remains (Crustacea, Branchiopoda) in Alaskan lake sediments. *Archiv für Hydrobiologie (Supplement),* **151**, 353-394.

Szeroczynska K., & Sarmaja-Korjonen K. (2007). Atlas of subfossil Cladocera from Central and Northern Europe (1st ed). Friends of the Lower Vistula Society, Swiecie, Poland.

Tropea A.E., Paterson A.M., Keller W., & Smol J.P. (2010). Sudbury sediments revisited: Evaluation limnological recovery in a multiple-stressor environment. *Water Air & Soil Pollution,* **210**, 317-333.

Turko P., Sigg L., Hollender J., & Spaak P. (2016). Rapid evolutionary loss of metal resistance revealed by hatching decades-old eggs. *Evolution*, **70**, 398-407.

Valois A.E., Keller W.B., & Ramcharan C.W. (2011). Recovery in a multiple stressor environment: using the reference condition approach to examine zooplankton community change along opposing gradients. *Journal of Plankton Research*, **33**, 1417-1429.

Vandekerkhove J., Declerck S., Vanhove M., Brendonck L., Jeppesen E., Conde Porcuna J.M., & De Meester L. (2004). Use of ephippial morphology to assess richness of anomopods: potentials and pitfalls. *Journal of Limnology,* **63**, 75-84.

Vörösmarty C.J., McIntyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P., Glidden S., Bunn S.E., Sullivan C.A., Liermann C.R., & Davies P.M. (2010). Global threats to human water security and river biodiversity. *Nature*, **467**, 555-561.

Whittaker R.H. (1972). Evolution and measurement of species diversity. *Taxon*, **21**, 213-251.

Yan N.D., Girard R., Heneberry J.H., Keller W.B., Gunn J.M., & Dillon P.J. (2004). Recovery of copepod, but not cladoceran, zooplankton from severe and chronic effects of multiple stressors. *Ecology Letters*, **7**, 452-460.

**Tables**

**Table 1: Description of lake size characteristics and surrounding land use for each of the four study lakes.**

| Lake | Latitude (DD) | Longitude (DD) | Observed maximum depth (m) | Surface area ($km^2$) | Surrounding land use |
|---|---|---|---|---|---|
| **Dauriat** | 54°48’23.73”N | 66°49’30.85”W | 11 | 0.56 | Town site; hedge row around some of lake; used for raw sewage until 1975. |
| **Knob** | 54°37’29.13”N | 66°48’30.77”W | 5.2 | 2 | Town site surrounding part of lake; drinking water source; flow through to Dauriat. |

**Table 2: Mean and maximum sediment metal concentrations along with sediment quality guidelines for the protection of aquatic life (for freshwater systems).** All concentrations are expressed in parts per million (ppm). 'ISQG' refers to "International Sediment Quality Guidelines for the Protection of Aquatic Life" and is converted from μg $kg^{-1}$ dry weight (Canadian Council of Ministers of the Environment, 2014). 'PEL' refers to "Probable Effect Level", also converted from μg $kg^{-1}$ dry weight.

| Metal | Dauriat mean (ppm) | Dauriat maximum (ppm) | Knob mean (ppm) | Knob maximum (ppm) | ISQG (ppm) | PEL (ppm) |
|---|---|---|---|---|---|---|
| **Al** | 5746 | 7360 | 5923 | 6690 | -- | -- |
| **Fe** | 13452 | 18400 | 8272 | 9500 | -- | -- |
| **Mn** | 2176 | 4270 | 4087 | 40900 | -- | -- |
| **Zn** | 600 | 1430 | 144 | 200 | 123 | 315 |
| **As** | 10 | 21 | 11 | 14 | 5.9 | 17 |
| **Cd** | 3 | 6 | 0.9 | 2 | 0.6 | 3.5 |
| **Co** | 64 | 121 | 37 | 124 | -- | -- |
| **Cr** | 67 | 85 | 72 | 93 | -- | -- |
| **Cu** | 159 | 348 | 35 | 52 | 35.7 | 197 |
| **Hg** | 3 | 8 | 0.9 | 0.9 | 0.2 | 0.5 |
| **Ni** | 105 | 226 | 81 | 158 | -- | -- |
| **Sb** | 5 | 5 | 5 | 5 | -- | -- |

**Table 3: Mixed-effect model summaries and model assessment in order of decreasing plausibility.** The formula (as used as input in `lme4`) is shown along with the AIC value for each model. Only models with high amounts of explained variation are shown.

| Model | Random factors | AIC value | $R^2$ (for linear models) |
|---|---|---|---|
| $\mathbf{Clad\_S_{ij} \sim Metal\_EF_i + \varepsilon}$ | None | 85 | Adj. $R^2 = 0.69$ |
| $\mathbf{Clad\_S_{ij} \sim Metal\_EF_i + Lake_j + \varepsilon}$ | Lake (varying intercept) | 88 | -- |
| $\mathbf{Clad\_S_{ij} \sim Metal\_EF_i + Lake_j + \varepsilon}$ | Lake (varying intercept + slope) | 94 | -- |
| $\mathbf{Clad\_S_{ij} \sim Metal\_PC1_i + \varepsilon}$ | None | 100 | Adj. $R^2 = 0.51$ |
| $\mathbf{Clad\_S_{ij} \sim Metal\_PC1_i + Lake_j + \varepsilon}$ | Lake (varying intercept) | 102 | -- |
| $\mathbf{Clad\_S_{ij} \sim Metal\_PC1_i + Lake_j + \varepsilon}$ | Lake (varying intercept + slope) | 107 | -- |

**Figures**

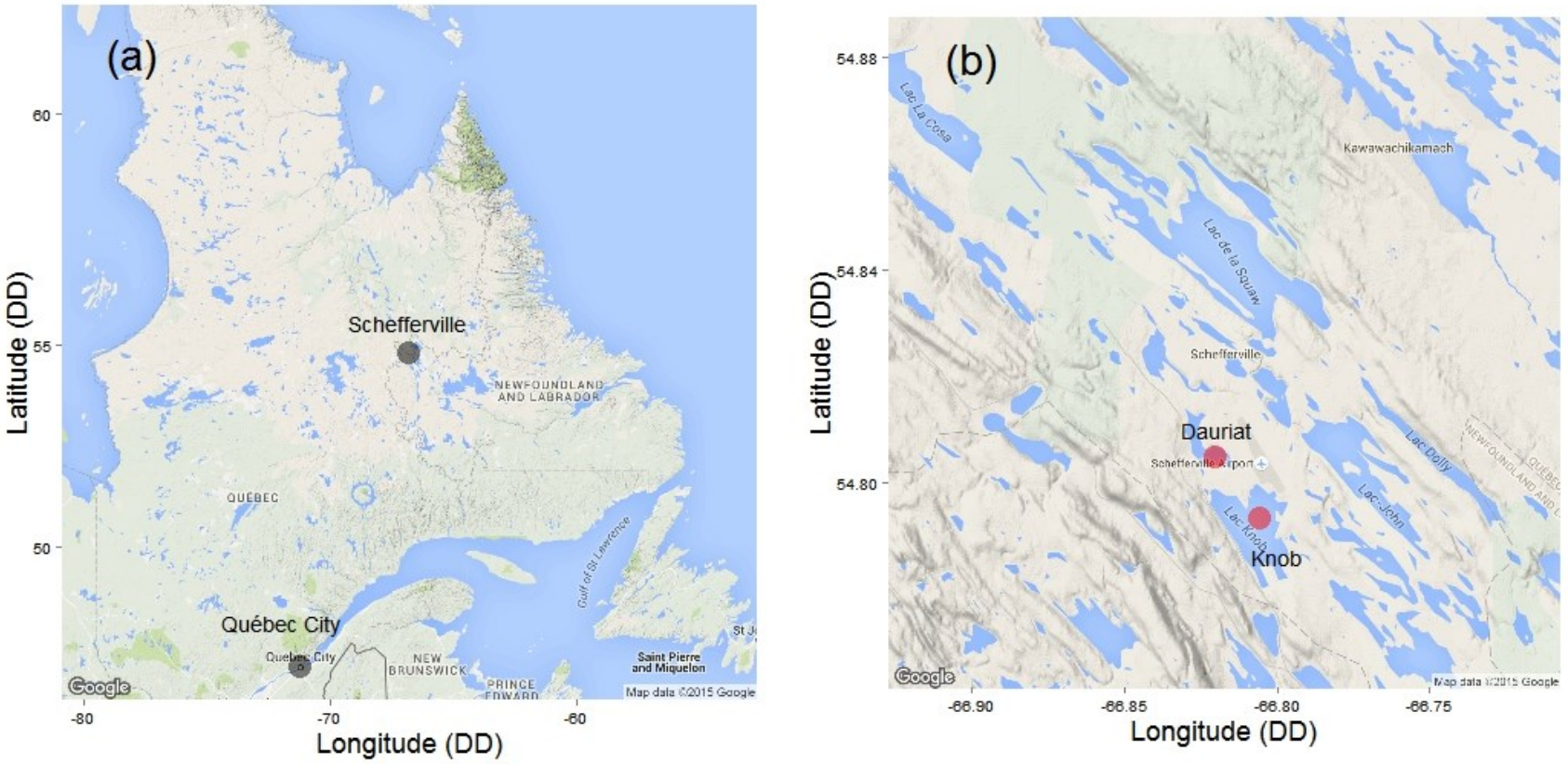



**Figure 1: Maps showing the location of (a) Schefferville in the Québec landscape and (b) the two study lakes surrounding the town of Schefferville.**

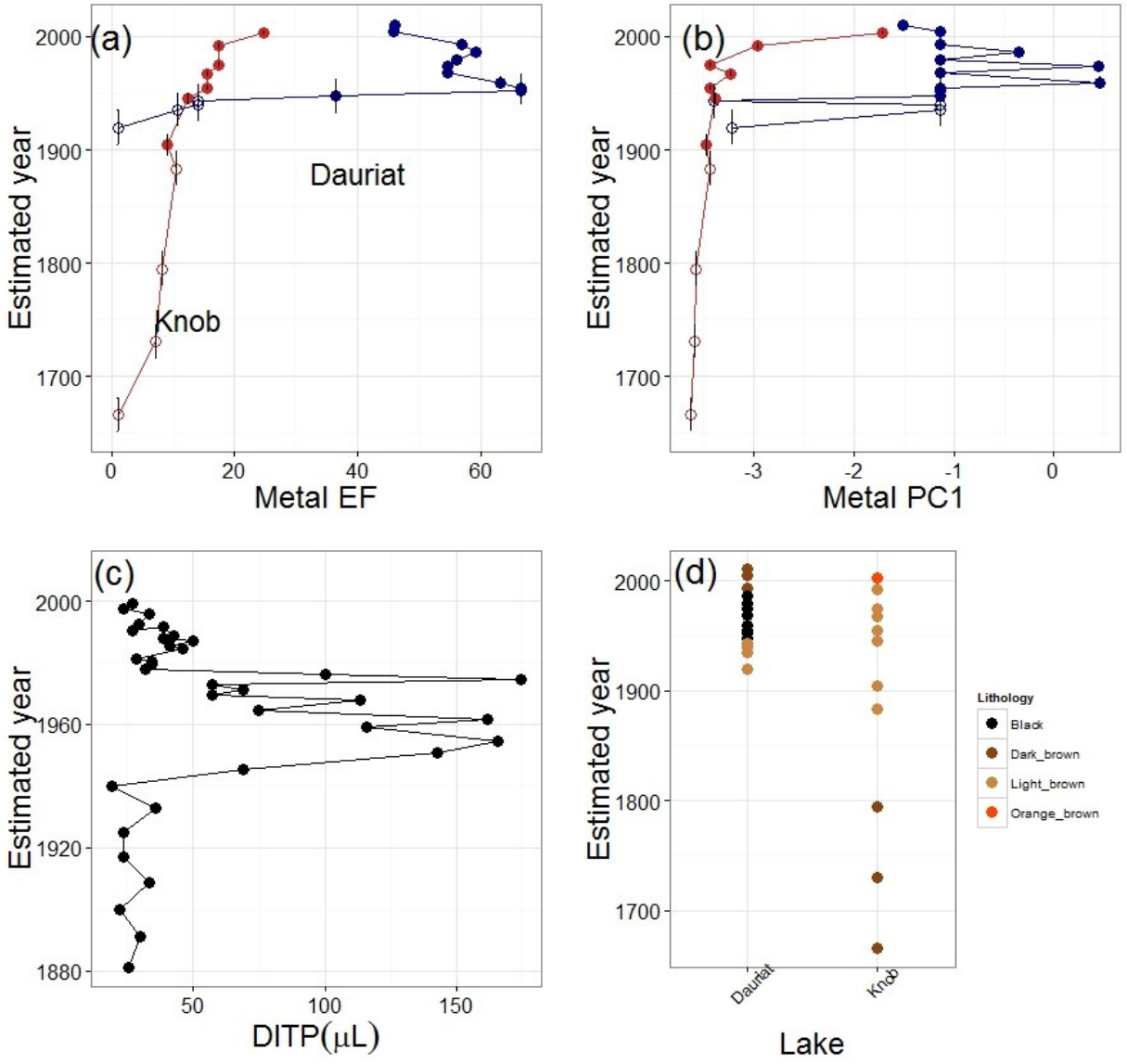



**Figure 2: Heavy metal enrichment factors (a), heavy metal PC 1 scores (b), diatom-inferred total phosphorus (c) and core lithology (d) for lakes Dauriat (blue) and Knob (red).** The Metal EFs depicted in (a) panel represent cumulative metal enrichment factors of 10 heavy metals (Ag, Ba, Cd, Co, Cu, Fe, Hg, Mo, Ni, and Pb). The Metal PC1 scores are from a PCA of the same 10 heavy metals plus a full set of elements (As, Be, Bi, Ca, Cr, Ga, Hg, K, Li, Mg, Mn, Mo, Na, P, S, Sb, Sc, Sr, Te, V, Y, Zn, and Zr). Diatom-inferred total phosphorus is shown for Lake Dauriat only and reproduced from Laperrière *et al*., (2008). Lithology demonstrate colour

changes down the cores for these two lakes. Closed points in (a) and (b) have been aged using radiometric dating estimates, whereas open points are extrapolated ages.

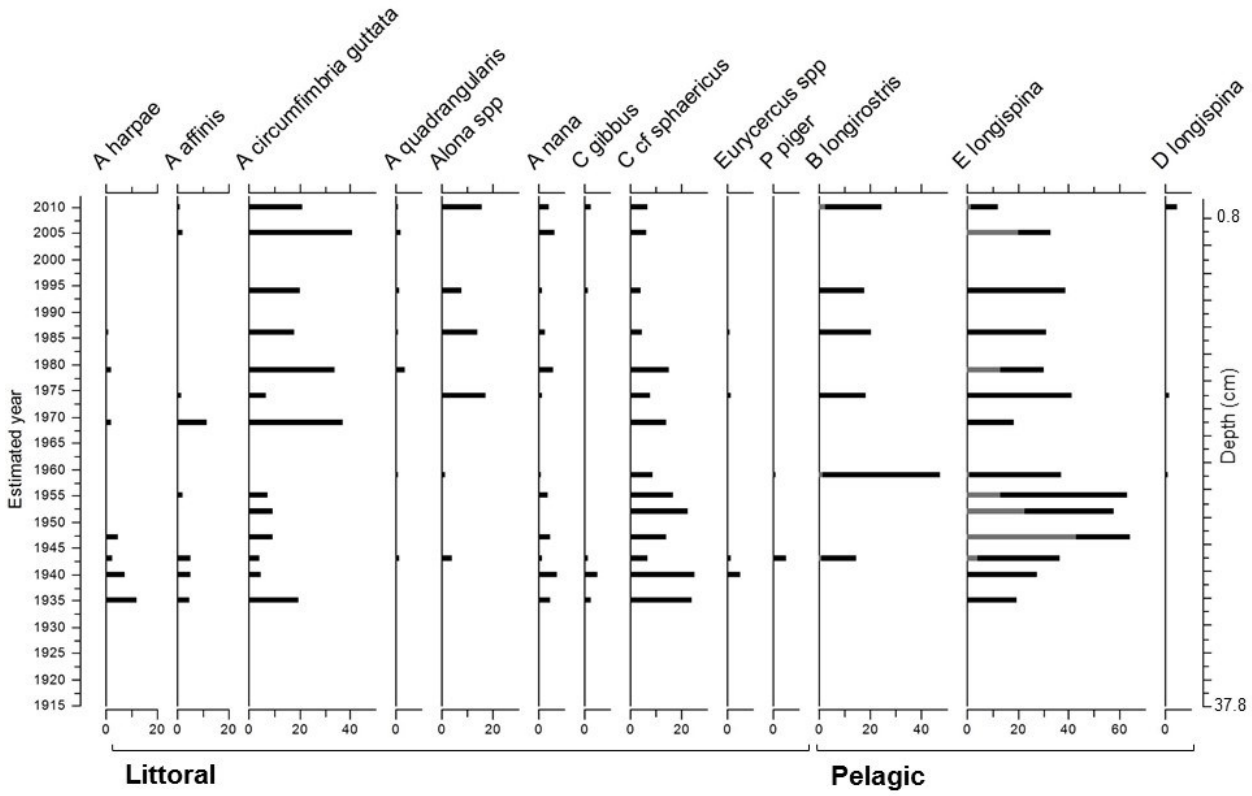



**Figure 3: Relative abundance of Cladoceran zooplankton assemblages through time in Lake Dauriat as expressed by percentage of the total Cladoceran assemblage.** Only taxa with a relative abundance of at least 5% in at least one interval are shown (rare species excluded). Cladoceran taxa are: *Acroperus harpae* (*A. harpae*); *Alona affinis* (*A. affinis*); *Alona circumfimbria*/*Alona guttata* (*A. circumfimbria guttata*); *Alona quandrangularis* (*A. quandrangularis*); *Alona* spp. (*Alona* spp.); *Alonella nana* (*A. nana*); *Chydorus gibbus* (*C. gibbus*); *Chydorus* cf. *sphaericus* (*C. sphaericus*); *Eurycercus* spp. (*Eurycercus* spp.); *Paralona piger* (*P. piger*); *Bosmina longirostris* (*B. longirostris*); *Eubosmina longispina* (*E. longispina*); and *Daphnia longispina* (*D. longispina*). Taxa are labelled as being found in either a littoral or pelagic habitat. Note that *Bosmina* spp. can be found in both open-water littoral as well as

pelagic habitat. Black bars for *B. longirostris* and *E. longispina* represent the total number of individuals, grey bars represent the fraction of that total that were determined from indistinguishable *Bosmina* spp., that were then split proportionally between the two species depending on the prevalence of individuals of each species in the sample.

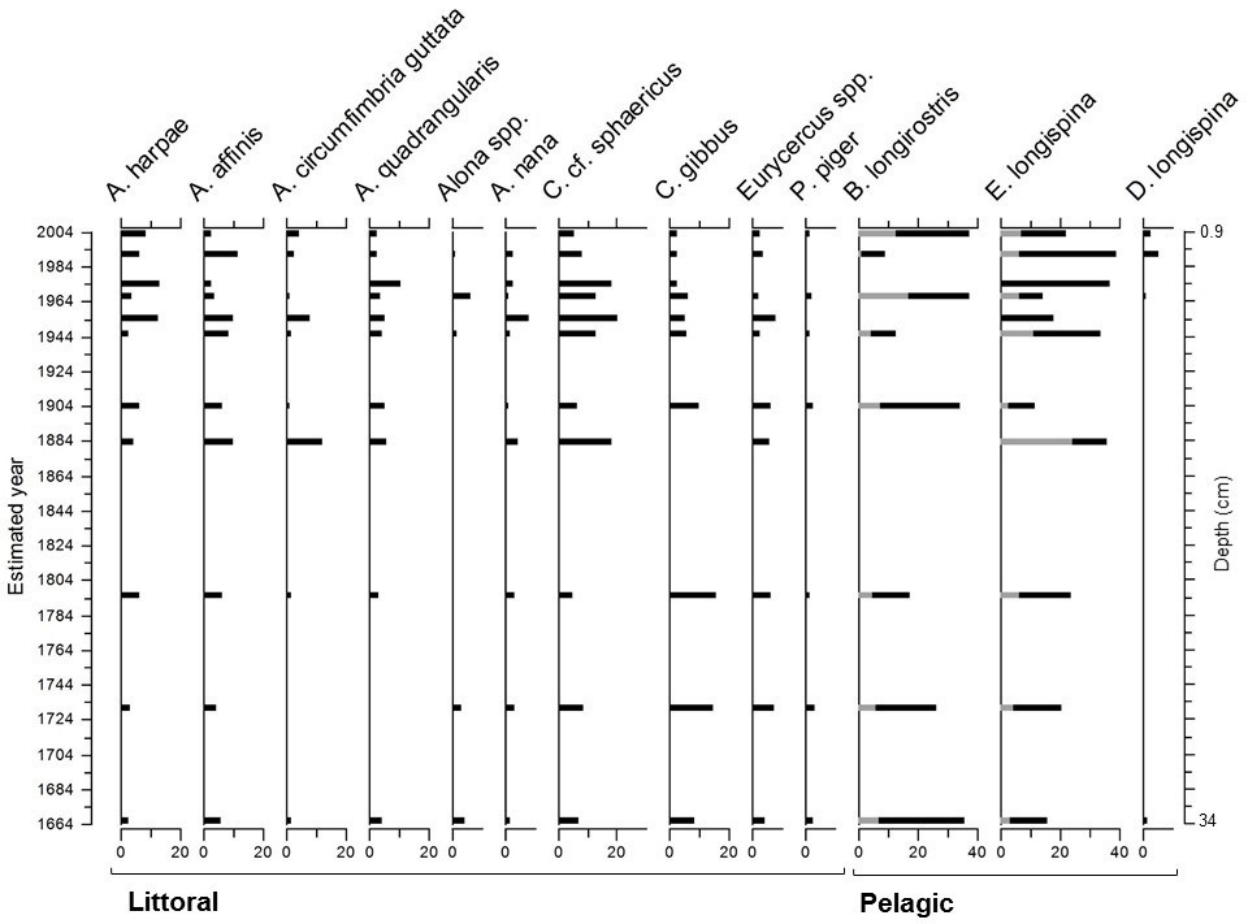



**Figure 4: Relative abundance of Cladoceran zooplankton assemblages through time in Lake Knob as expressed by percentage of the total Cladoceran assemblage.** Only taxa with a relative abundance of at least 5% in at least one interval are shown (rare species excluded). Cladoceran taxa are: *Acroperus harpae* (*A. harpae*); *Alona affinis* (*A. affinis*); *Alona circumfimbria/Alona guttata* (*A. circumfimbria guttata*); *Alona quandrangularis* (*A. quandrangularis*); *Alona* spp. (*Alona* spp.); *Alonella nana* (*A. nana*); *Chydorus gibbus* (*C. gibbus*); *Chydorus* cf. *sphaericus* (*C. sphaericus*); *Eurycercus* spp. (*Eurycercus* spp.); *Paralona piger* (*P. piger*); *Bosmina longirostris* (*B. longirostris*); *Eubosmina longispina* (*E. longispina*); and *Daphnia longispina* (*D. longispina*). Taxa are labelled as being found in either a littoral or pelagic habitat. Note that *Bosmina* spp. can be found in both open-water littoral as well as

pelagic habitat. Black bars for *B. longirostris* and *E. longispina* represent the total number of individuals, grey bars represent the fraction of that total that were determined from indistinguishable *Bosmina* spp., that were then split proportionally between the two species depending on the prevalence of individuals of each species in the sample.

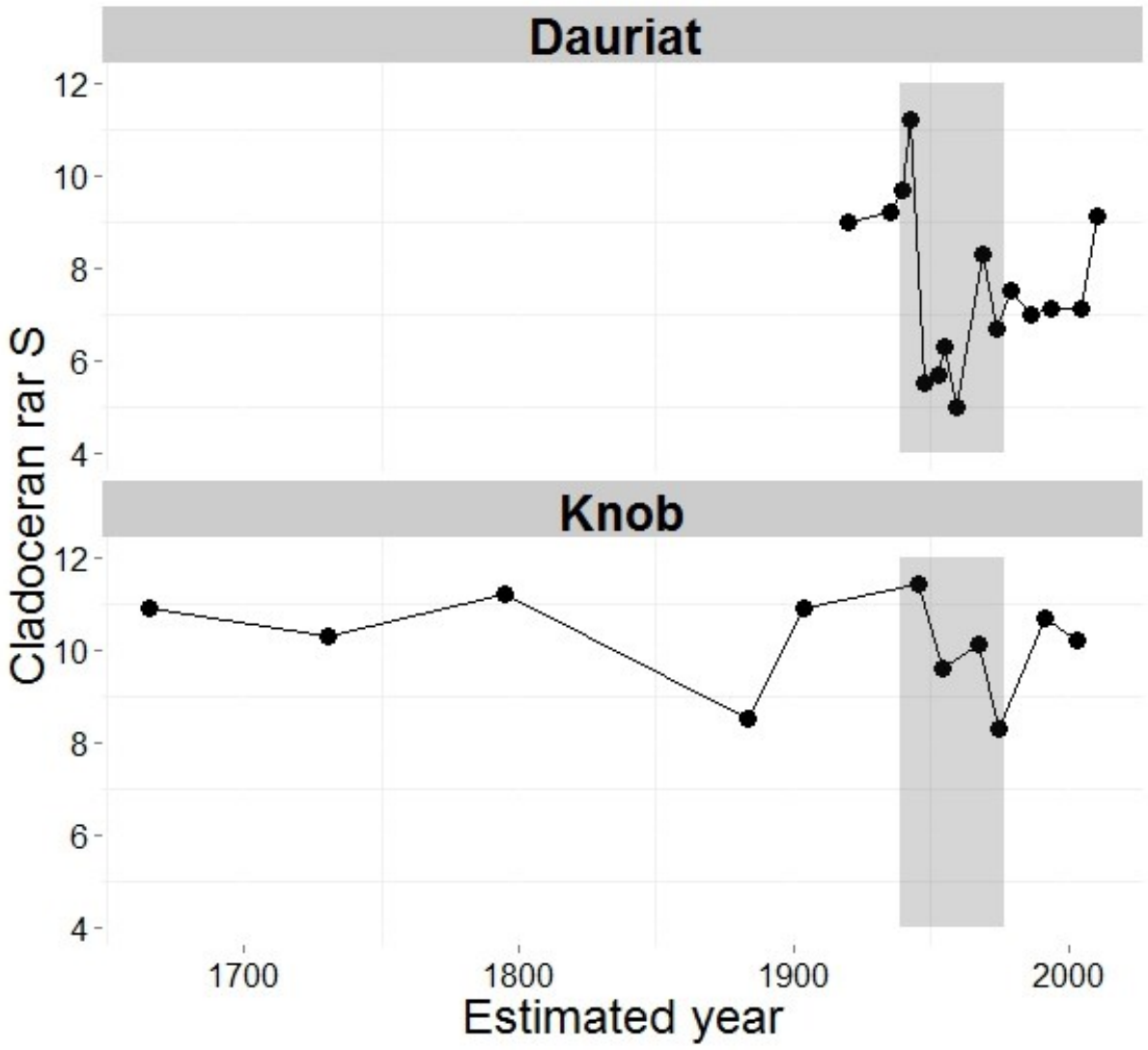



**Figure 5: Rarefied species richness through time for lakes Dauriat and Knob.** The mining period (1939-1977) is shaded.

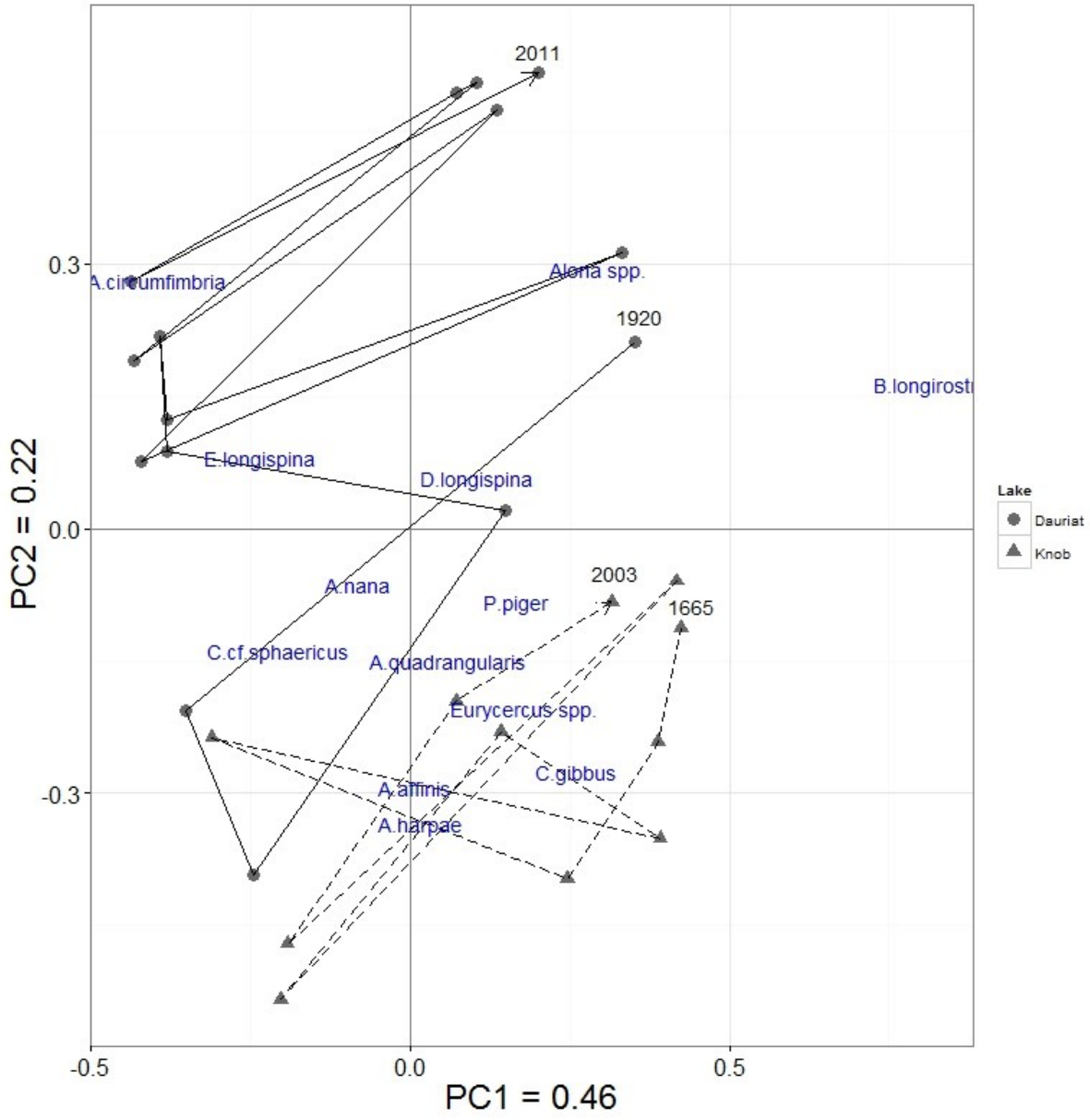



**Figure 6: PCA of the Cladoceran assemblages from lakes Dauriat (samples marked by circles) and Knob (triangles).** The temporal trajectory of the assemblages from the two lakes is connected between samples with the oldest and most recent samples marked by estimated years. Species labels are the same as in Figs. 3 and 4.

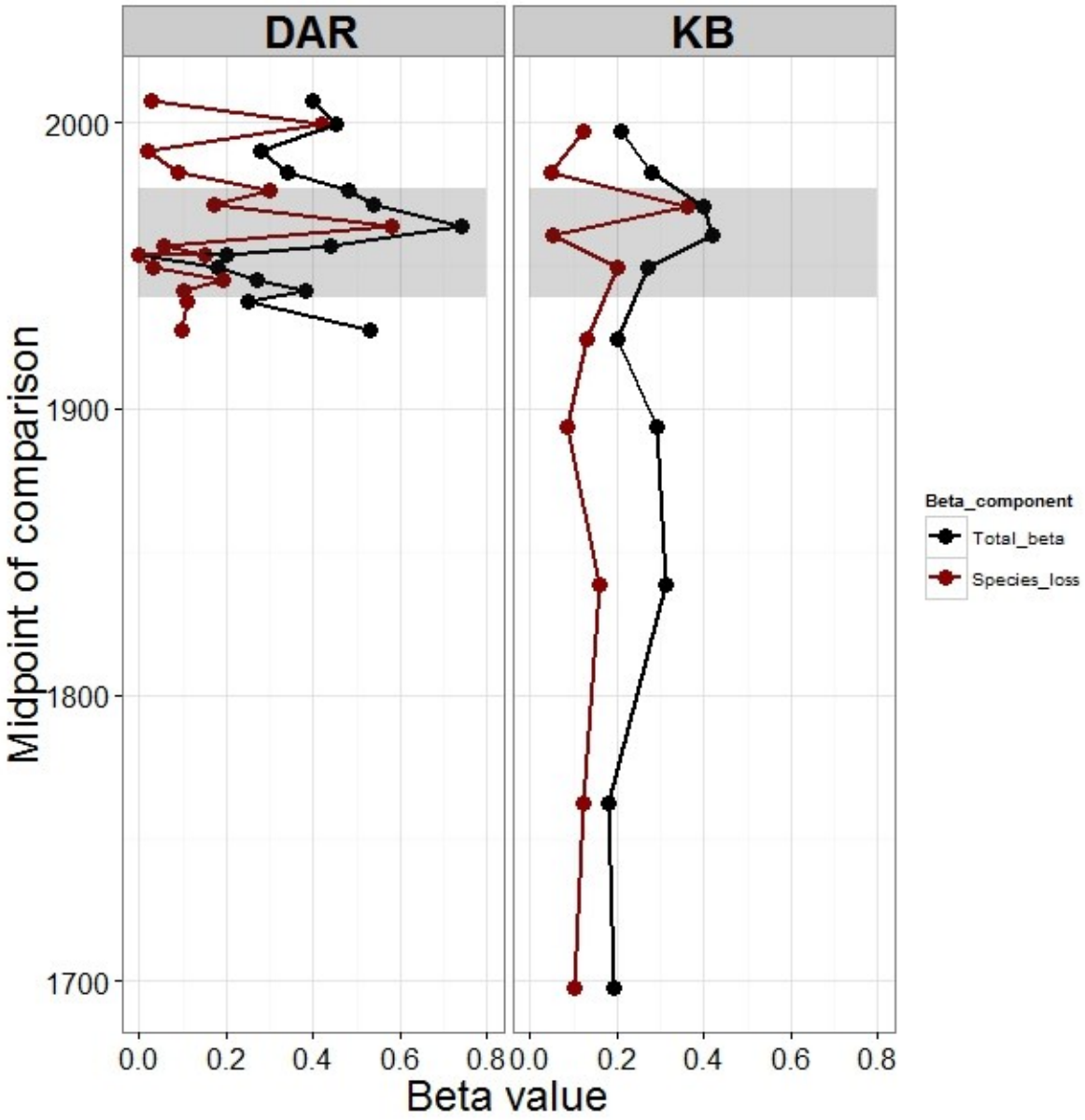



**Figure 7: Beta diversity for (left) Lake Dauriat and (right) Lake Knob, shown through time.** Each temporal beta diversity point represents a comparison between two intervals, and the year midpoint of that comparison is shown on the y-axis. "Total_beta" refers to total temporal beta diversity computed between time intervals. "Species_loss" refers to the loss of Cladoceran abundances on a taxa by taxa basis. The mining period (1939-1977) is shaded.

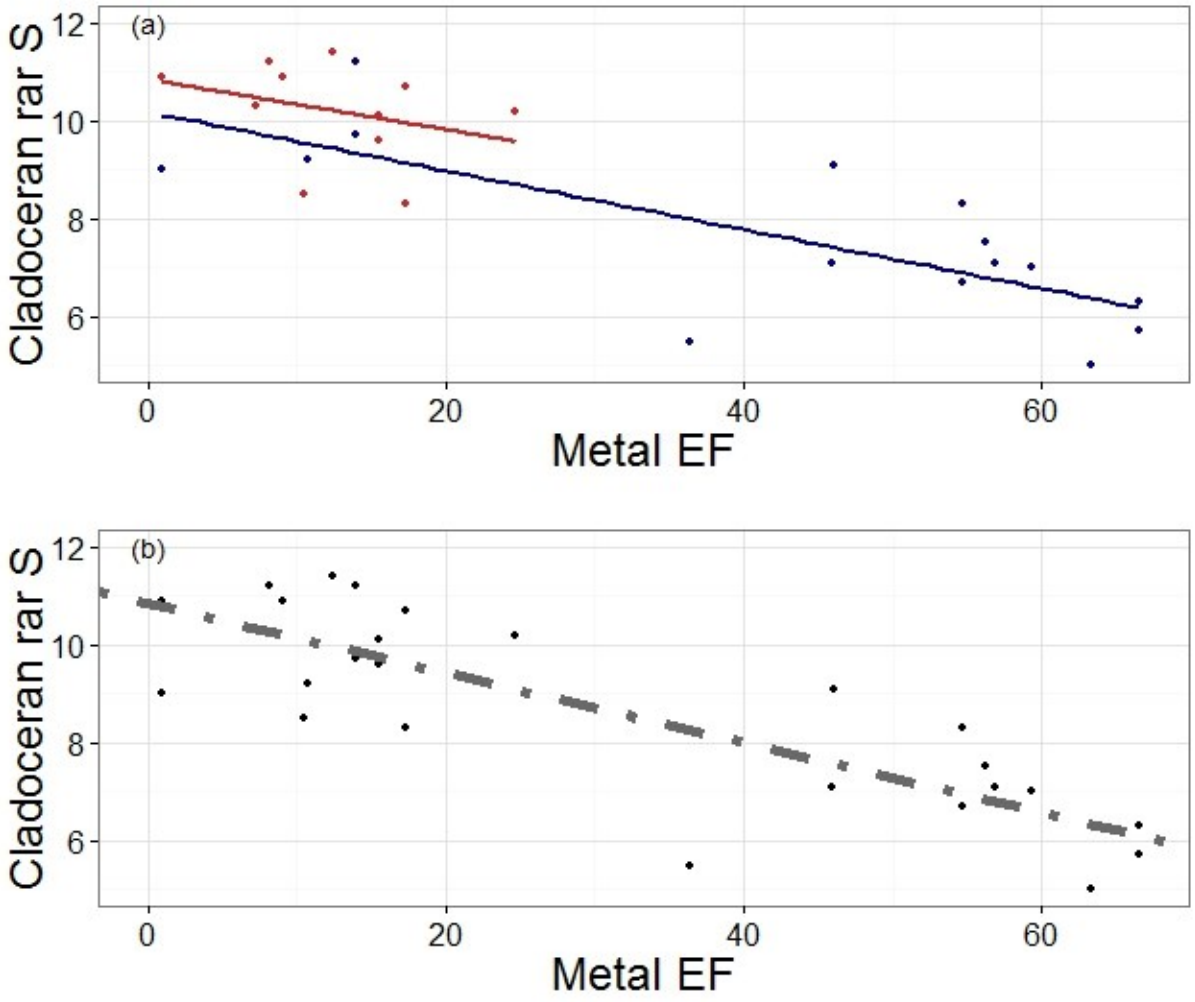



**Figure 8: Relationship between cumulative metal enrichment factor and rarefied Cladoceran species richness for the two Schefferville lakes ($n_{interval}$ = 27) using (a) a linear model and (b) mixed effect model with lake as a random factor.** For the linear model, the blue line represents Lake Dauriat and the red line Lake Knob. The adjusted R2 = 0.7. For the mixed effect model, the slope of the overall line is -0.071, whereas the upper and lower confidence intervals for this model are -0.0 and -0.08, respectively.

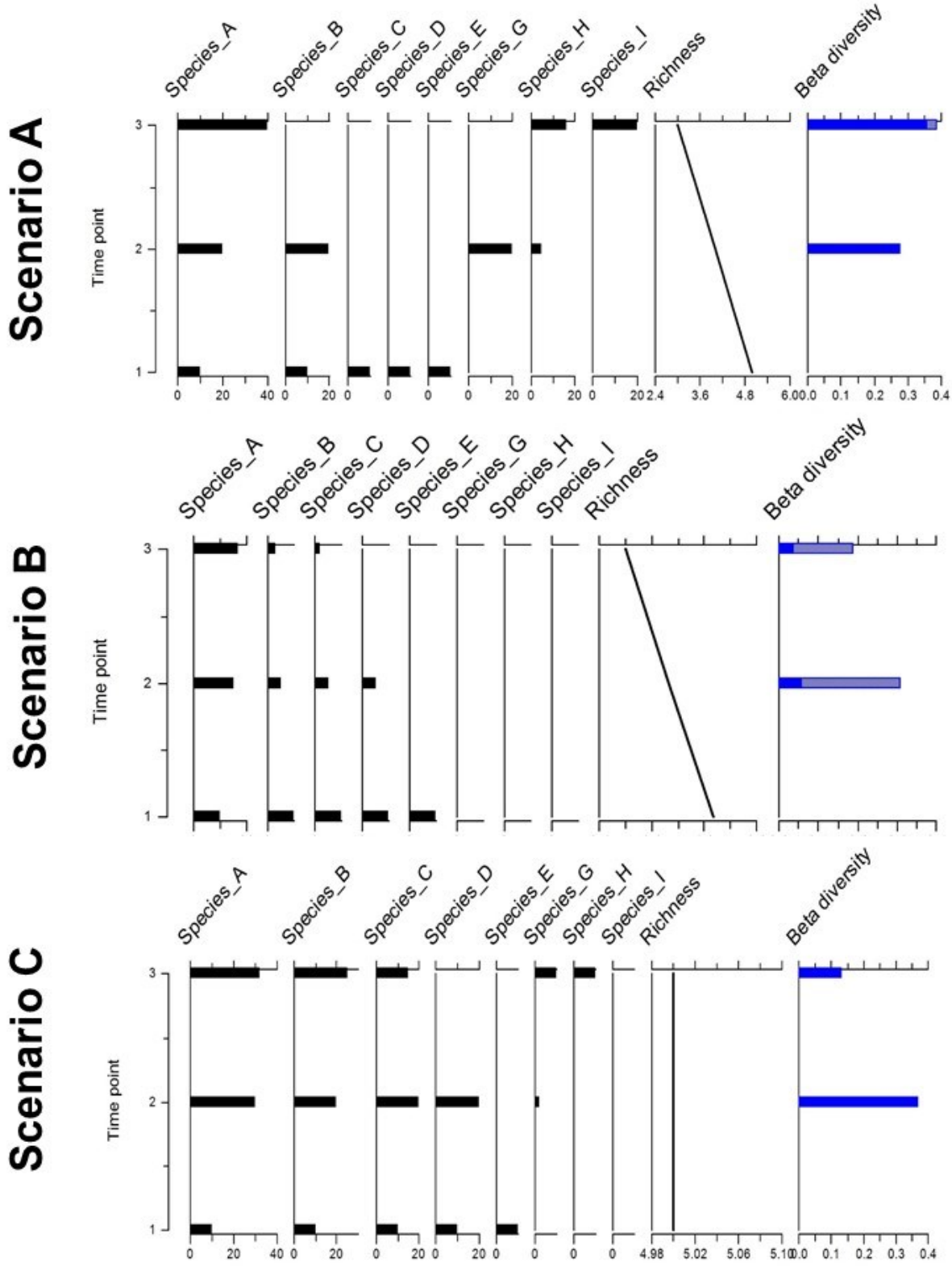

Scenario A
Species_A
Species_B
Species_C
Species_D
Species_E
Species_G
Species_H
Species_I
Richness
Beta diversity
Time point
Scenario B
Species_A
Species_B
Species_C
Species_D
Species_E
Species_G
Species_H
Species_I
Richness
Beta diversity
Time point
Scenario C
Species_A
Species_B
Species_C
Species_D
Species_E
Species_G
Species_H
Species_I
Richness
Beta diversity
Time point

**Figure 9**

**Figure 9: Three hypothetical assemblages of eight species (relative abundance) at each of three time points show ways in which patterns observed using cladoceran richness may or may not be congruent with those observed using temporal beta diversity.** Figure 9 shows three hypothetical assemblages (Scenarios A, B, C) where the abundance of eight species (A-H) is recorded at three points in time (Time point 1, 2 and 3). Beta diversity is further represented by either a loss component (hashed bar) or gain component (solid bar). Scenario A and B represent assemblages where species richness has decreased through time (for the context of this study, we will assume that this decrease through time is related to a negative linear relationship with metal contamination), whereas Scenario C shows an assemblage where species richness has stayed constant through time, despite changes in the relative abundances of individual species. In Scenario A, metal tolerant species are able to take advantage of niches opened up in high metal conditions, thus contributing to an increased gain component in temporal beta diversity because these metal tolerant species gain in abundance on an individual basis. Under this scenario, the conclusions drawn about the effect of metal contamination in a system would be different if observing species richness (decreasing) or temporal beta diversity (high gain component). Scenario B demonstrates an example of where both species richness and temporal beta diversity show congruent patterns, where cladoceran species richness decreases with higher metal concentrations and this results in a high loss component making up total temporal beta diversity, as species themselves are lost and remaining species experience losses in abundance. Finally, Scenario C shows a situation where cladoceran richness may increase or stay the same with increased metal concentrations, and this is further echoed by increase in abundances of a species by species basis in terms of temporal beta diversity. As such individual species show gains in abundance, and species richness either stays constant or increases.

## CONNECTING STATEMENT 3

Using a single group of organisms, the cladoceran zooplankton, I found evidence in Chapter 3 that metal contamination is related to reduced alpha diversity as well as temporal beta diversity. Like Chapter 1 and 2 though, these patterns were detected using single groups of organisms. The focus on individual proxies in sediment records has long been a limitation of paleolimnology, where studies are constrained by a pretty laborious analysis to generate data on organisms that leave morphological (or pigment signatures) in the subfossil or fossil record. However, a great number of other aquatic organisms have the potential to leave DNA signatures of environmental DNA (eDNA) records in sediment cores (see Fig. C2, Gregory-Eaves & Domaizon, 2014). Thus, in Chapter 4, I used DNA-based approaches (metabarcoding) to capture a fuller picture of diversity and how that relates to patterns of beta diversity. DNA-based approaches have great potential to expand paleolimnological analyses, however there are still many caveats and unknowns (Barnes *et al*., 2014; Eichmiller *et al*., 2016). As a result, I decided to continue to work in the Schefferville system, where there is a lot of information about the history of the region, and because I have already explored some ecological dynamics of freshwater biodiversity in Chapter 3.

Chapter 4 is a capstone of the thesis in a way, as it represents the final part of my exercises working across different scales of biodiversity study. In Chapter 1, I worked on the calibration of samples from both contemporary and paleolimnological samples. In Chapter 2, I worked across a large spatial scale (the whole conterminous United States) for the study of beta diversity, as well as extending to a long scale of temporal beta diversity. In Chapter 3, I used higher resolution temporal data for more extensive study of temporal beta diversity. Finally, for Chapter 4, I expanded my study to multiple taxonomic groups and trophic levels, addressing

limitations of earlier chapters that focused on single organismal groups. Finally, Chapter 4 integrates molecular ecological techniques into paleolimnology.

**Literature Citecd**

Barnes M.A., Turner C.A., Jerde C.L., Renshaw M.A., Chadderton W.L., & Lodge D.M. (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology,* **48**, 1819-1827.

Eichmiller J.J., Best S.E., & Sorensen P.W. (2016). Effects of temperate and trophic state on degradation of environmental DNA in lake water. *Environmental Science & Technology*, **50**, 1859-1867.

Gregory-Eaves I., & Domaizon I. (2014). Analysis of DNA archived in lake sediments. ASLO e-Lecture. Available: 10:4319/lol/2014.igregory-eaves.idomaizon.7

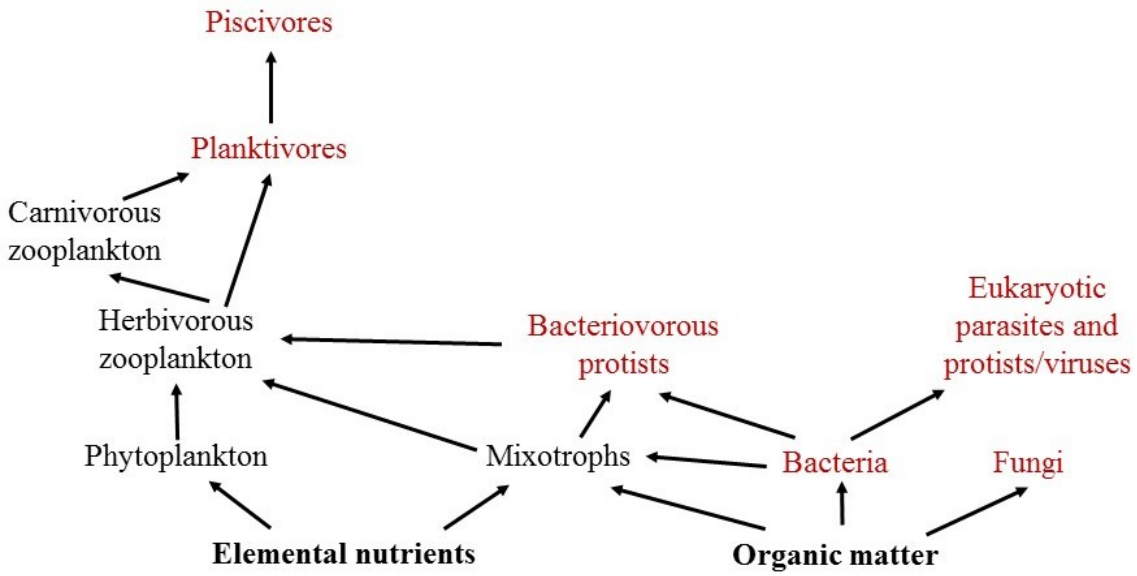



**Figure C2: Potential sources of eDNA in lake sediments from aquatic organisms.** Organisms in black are those which result in diagnostic signatures in paleolimnological sediments (subfossils, pigments etc.). Organisms in red are those which are not generally detectable in sediments, but from which a DNA signature could be left in lake sediments. Adapted from Gregory-Eaves and Domaizon (2014).

# CHAPTER 4

## Microbial eukaryotic biodiversity dynamics during the Anthropocene from a northern mining region: an exploration using High-Throughput Sequencing

Amanda K. Winegardner[1,2], Eric Capo[3], Isabelle Domaizon[3], Didier Debroas[4], Mehrdad Hajibabei[5], Shadi Shokralla[5], Boswell Wing[6], Beatrix E. Beisner[2,7], and Irene Gregory-Eaves[1,2]

1. Department of Biology, McGill University
2. Québec Centre for Biodiversity Scienc
3. CARRTEL, INRA, Université de Savoie Mont Blanc, 74200, Thonon-les-Bains, France
4. Université Clermont Auvergne, Université Blaise Pascal, Laboratoire « Microorganismes : Génome et Environnement », BP 10448, F-63000 Clermont-Ferrand, France
5. Biodiversity Institute of Ontario, University of Guelph
6. Department of Earth and Planetary Sciences, McGill University
7. Department of Biological Sciences, University of Québec at Montréal

**Abstract**

Paleoecology has provided important insight into past aquatic diversity through the use of subfossil remains, pigment analyses and transfer functions. The use of environmental DNA and high-throughput sequencing methods are able to further extend the power of paleoecology in lake systems by allowing for the reconstruction of assemblages that do not leave readily identifiable or quantifiable remains in sediments. We extracted environmental DNA from sediment cores retrieved from two lakes in the iron-ore mining region of Schefferville, Québec. These lakes represent an upstream and downstream system, with more intense eutrophication and metal inputs in the downstream site. We used primers to amplify microbial eukaryotes (protists and fungi) with 454 high-throughput sequencing to sequence environmental DNA from sediment cores. We were interested in whether the historical eukaryote assemblages differed between the two lakes, whether there were temporal changes in eukaryote assemblages, and quantifying the extent to which these changes could be associated with the multiple stressors present in the environment. We expected to find a reduction in eukaryote richness in the disturbed lake during a mining period (from the late 1930s to late 1970s) as specialist species are pushed towards the upper bound of their tolerance to metal loading. We also expected higher temporal turnover (beta diversity) between time intervals during the mining period than pre- and post-mining periods due to increased species replacement. We found that both lakes showed a breakpoint in eukaryote assemblage composition during the 1950s-1960s, corresponding to the most intense mining activities in the region. The downstream lake also showed a breakpoint in the 1980s, corresponding with the cessation of mining activities and post-installation of a wastewater treatment plant. In both lakes, temporal beta diversity peaked during the mining period. Overall,

this study demonstrates how the application of molecular techniques in paleoecology can provide additional insights into understanding the dynamics of historically contaminated systems.

## Introduction

Contributions from the field of paleolimnology have had increasingly important implications for the study of ecological dynamics through time, and have addressed questions pertaining to species co-occurrence, invasive species, and resistance and resilience to disturbance (e.g., Quinlan *et al*., 2005; Hamrová *et al*., 2010; De Laender *et al*., 2012; and Thienpont *et al*., 2015). Paleolimnologists using traditional and proven approaches have an extensive range of tools and proxies available to them. Aquatic organisms leave a broad range of detectable remains in lake sediment records and these are frequently used in paleolimnological studies, including, diatom valves (e.g., Dixit *et al*., 1992), chrysophyte cysts (e.g. Douglas & Smol, 1995), cladoceran zooplankton subfossils (e.g., Korosi & Smol, 2012), zooplankton diapause resting eggs (e.g., Brendonck & De Meester, 2003) and chironomid head capsules (e.g., Quinlan & Smol, 2001), amongst others. Cyanobacteria and other phytoplankton can be tracked using pigment concentrations in sediments (Leavitt & Hodgson, 2001). Furthermore, exogenous materials such as pollen particles from watershed vegetation enter lake systems and preserve in sediments, and are also widely used in paleolimnological studies (e.g., Schwark *et al*., 2002). The use of these approaches have been very useful for historical reconstructions of climate, water quality and watershed conditions. In addition, paleoecologists have addressed more ecological questions such as: Have there been species extinctions or invasions? Or have organisms displayed phenotypic plasticity through time?

Even as paleolimnological approaches are increasingly adopted for the study of aquatic biodiversity (e.g. Gregory-Eaves & Beisner, 2011), DNA-based approaches are also gaining prominence as a way to expand the detection of biodiversity available through a paleolimnological lens. Indeed, DNA-based approaches may be the best solution for several of

the key themes for paleolimnological research outlined by Seddon *et al*., (2014). The application of DNA-based approaches for characterizing biodiversity contained within lake sediments can focus on the use of either (or both) intracellular and extracellular DNA. For example, investigators may choose to extract DNA from diapause resting stages and identify individuals using DNA barcoding (*sensu* Montero-Pau *et al*., 2008), or they may use of high-throughput sequencing (HTS) methods to isolate, amplify and sequence environmental DNA from sediment slurries (Coolen *et al*., 2013). Indeed, DNA-based methods have the potential to considerably increase the range of taxa and functional groups that we are able to identify from lake sediments, by adding piscivorous and planktivorous fish, bacterivorous protists, unpigmented bacteria, fungi and eukaryotic parasites and viruses to our repertoire of organisms accessible for study (Gregory-Eaves & Domaizon, 2014).

The addition of organismal groups to lake sediment core analyses is not the only reason for the integration of DNA-based approaches into paleolimnology. Using DNA can also help to identify potential bias within traditional methods as well. For example, the identification and study of zooplankton diapause eggs deposited in sediments has largely relied on "resurrection ecology", wherein eggs are extracted from sediments and then exposed to conditions favourable to hatching (e.g. Kerfoot *et al*., 1999; Kerfoot & Weider, 2004; Angeler, 2007; and Derry *et al*., 2010). The organisms that hatch from these eggs can then be identified morphologically. However, this method is not only inherently difficult but it also means that only eggs that can be induced to hatch can be identified. Consequently, eggs that have already hatched or have been damaged while in the sediment egg bank are not readily identifiable (except for scant literature focusing on the morphological identification of cladoceran ephippial cases, e.g., Vandekerkhove *et al*., 2004).

Metabarcoding is the DNA-based approach used to identify multiple species or groups of species from a bulk sample or slurry of water, soil or sediment (Taberlet *et al*., 2012), providing an opportunity for very comprehensive study of freshwater biodiversity. Because the use of metabarcoding to infer environmental influences on aquatic communities is still fairly novel in paleolimnology, we chose to conduct this study in a region where a history of disturbance has been well established, documented and studied. We focused on microbial eukaryotes because of their prevalence in aquatic food webs and large diversity. As such, we used metabarcoding approaches to characterize microbial eukaryote assemblages through time in two lakes in the iron-ore mining region of Schefferville, Québec. Iron-ore exploitation began in this region ~ 1939, exposing our two focal lakes to considerable anthropogenic stress. Lake Dauriat is a highly disturbed lake on the downstream side of the town of Schefferville and its historical community shifts have been examined in a number of previous paleolimnological studies (e.g., Choulik & Moore, 1991; Laperrière *et al*., 2008; Aebischer *et al*., 2015; Winegardner *et al*., (In prep., Chapter 3)), as well as in ethnographic studies of the region (Boutet, 2012). Lake Knob is located on the upstream side of the town and has historically served as the town's drinking water source. A more thorough description of both of these lakes, as well as the geochemical and metal contamination history of the region are found in Winegardner *et al*., (In prep); Chapter 3.

Our questions of interest using high-throughput sequencing approaches in this system were: How do these two lakes with contrasting histories differ in terms of community dynamics? And, are there ecological explanations for detected differences in eukaryote assemblages? We hypothesized that there would be a reduction in eukaryote richness (operational taxonomic unit (OTU) richness) during the mining period as specialist (non-metal tolerant), as well as susceptible generalist species were pushed towards their upper bound of tolerance to metal

loading, or eliminated entirely, and metal-tolerant taxa came to dominate the assemblages. We also expected greater turnover in community composition (temporal beta diversity) between time points during the mining period owing to increased disturbance, relative to the pre- and post-mining periods. Finally, we expected that metal loading would select for metal-tolerant specialists with such replacement observable through time, potentially negating a trend of declining taxon richness.

## Methods

*Description of the study system*

The region surrounding Schefferville, Québec, Canada (54°48'N; 66°50'W) is considered to have been exposed to three anthropogenically-altered time periods: "Pre-mining" up to 1938; "Mining" from 1939 to 1977 (with the most intense development c. 1950s); and "Post-mining" starting in 1978 (mine decommissioning in 1982). It is important to note however that there has been a resumption of some iron-ore extraction in the 2000s, but this activity has recently taken a downturn. Lake Dauriat is located downstream of the town and receives considerable run-off and pollution from the town site due to its de-vegetated shoreline and lack of a buffer zone. This lake also received raw sewage from the town until wastewater treatment facilities were installed in 1975. The second lake, Lake Knob is located upstream of the town, has a partially vegetated shoreline and functions as the town's drinking water source. The two lakes show considerable metal enrichment through time (See Winegardner *et al*., (In prep); Chapter 3).

*Sediment core collection*

Several sediment cores were collected from the two focal lakes in the summer (July) of 2012 and used for this study as well as the work outlined in Chapter 3. The longest core from

each lake was used for radiometric dating, and geochemical analyses as well as for providing samples for enumerating cladoceran subfossils and extracting environmental DNA. The remainder of the cores were used as supplementary material and for exploratory studies. Cores were stored in a dark fridge within ~ 30 minutes of collection. Each sediment core was extruded at 0.25 cm increments, and the intervals stored in a freezer at the McGill Subarctic Research Station. Frozen samples were subsequently shipped cold to their final destination over a one-day period and returned to a freezer (-20°C) upon arrival. While there is potential for contamination between samples by smearing from sides of the cores during the extruding process, previous work by has demonstrated by Pal *et al*., (2015) that this had a negligible effect. The alternative option of cross-sectionally splitting the cores and extracting samples from the centre of the cores was not used due to facility limitations and concerns regarding transporting full cores to the necessary laboratory.

*Radiometric dating and geochemical analyses*

We established core chronologies by first measuring several radiometric isotopes ($^{210}$Pb, $^{226}$Ra, and $^{137}$Cs) in 15 sediment intervals that spanned the length of each core. Using these data, we applied a Constant Initial Concentration (CIC) model to develop an age-depth relationship for each lake (radiometric dating and model selection is outlined in a Technical Appendix; see Ch. 3, S2). We also measured geochemical profiles for a suite of heavy metals down these two reference cores: Ag, Al, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, Li, Mg, Mn, Mo, Na, Ni, P, S, Sb, Sc, Sr, Te, Ti, U, V, W, Y, Zn, and Zr. These methods are described in Chapter 3 (Winegardner *et al*., (In prep)) and were completed at a commercial laboratory, Actlabs (Ancaster, Ontario, Canada). Both lakes show metal enrichment from historical (pre-1850 CE in Lake Knob and pre-1920 in Lake Dauriat) to recent times (see Ch. 3 S3), with metal loading for

many heavy metals higher than accepted guidelines for aquatic system health (International Quality Guidelines for the Protection of Aquatic Life; ISQG) and Probable Effect Levels (PEL) (Canadian Council of Ministers of the Environment, 2014).

*DNA extraction and sequencing*

We selected ten samples of frozen (extruded) sediment from the Lake Dauriat core and six samples of frozen sediment from the Lake Knob core; selecting the samples so that they were from (approximately) evenly spaced points down the core profiles (core heights were ~ 40 cm for Dauriat and ~ 30 cm for Knob), resulting in a total of16 samples for subsequent DNA extractions and analyses (one sample was later removed post sequencing, see below) (Table 1). We aimed for each sample to have a wet mass of between 0.5 and 1.5 g and as such, combined adjacent 0.25 cm intervals in order to achieve the desired weight. Samples were then placed in sterile Falcon tubes and subsequently refrozen at -20°C.

We extracted DNA from each sediment sample (one extraction per sample) using a PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories Inc.; see Appendix S1 for complete extraction details and Table 1 for concentrations of DNA extracts prior to amplification). The DNA extracts were shipped to the Biodiversity Institute of Ontario, Guelph, Ontario, Canada for high-throughput Roche 454 Pyrosequencing. Three amplicons were prepared of each sample using three different primer sets designed to target bacterial DNA, diatom DNA or eukaryotic DNA. We report the full set of primers here in order to understand our Methods, however we only describe the statistical analyses and subsequent Discussion for the eukaryote-targeted sequences. The diatom specific primers were D512/D978rev for 18S (amplicon size ~ 466 bp, described in Zimmerman *et al*., 2011). The bacteria specific primers were 563F/907rM for 16S (amplicon size ~ 344 bp, described in Schauer *et al*., (2003) and Claesson *et al*., (2010)). For the

eukaryotes, we used the primer set NSF573F/NSR951R for 18S rRNA gene (amplicon size ~ 378 bp, described in Mangot *et al*., (2013)). All amplicons of each sample were normalized and pooled prior to sequencing.

We assembled PCRs into 25 µl volumes for each sample/master mix/primer combination: 2 µL DNA, 17.5 µL molecular biology grade water, 2.5 µL 10x Invitrogen buffer, 1 µL 50x $MgCl_2$, 0.5 µL dNTP mix, 1 µL forward primer (10uM), 1 µL reverse primer (10 uM), and 0,5 µL Invitrogen Platinum Taq polymerase (5 U/ul). Thermocycler regimes were performed as per Zimmerman *et al*., (2011), Schauer *et al*., (2003) and Mangot *et al*., (2013), with a negative control (no DNA) included for each amplification. Amplicon success was visualised using 2% E-gels 96 Agarose (Invitrogen, Burlington ON, Canada) as well as fluorometer readings (S2).The PCR products of each sample were then sequenced in 1/16 lanes of a 70X75 Picotitre plate using a Roche 454 FLX sequencer following the procedure outlined in Shokralla *et al*., (2014), where the group-specific amplicons were then sequenced in the same run. Detailed information regarding amplicon concentrations post-PCR are shown in Appendix S2. The sequences used in the analyses for this study will be deposited in GenBank. Table 2 shows that sequencing success varied across the organismal groups targeted as well as through depth in the two sediment cores, with the deepest sample in Lake Dauriat showing particularly low sequencing success for bacteria and eukaryotes (but a relatively high number of diatom DNA sequences). This may be an effect of the high-throughput where extensive amplification of one primer group results in decreased success of another. Since we chose to focus on the eukaryotic assemblages for this manuscript, we removed the deepest sample from Lake Dauriat from subsequent analyses because of low sequencing success for eukaryotic organisms (as such, the remainder of the analyses had n = 15).

*Sequence analysis and bioinformatics of eukaryotic reads*

The 454 sequencing using the eukaryotic primer set produced approximately 200,000 DNA sequences across the 15 samples with an average read length of 343 bp. We used the Phylogenetic Analysis of Next-generation Amplicons pipeline (PANAM V3: https://github.com/panammeb/;Taib *et al*., 2013a) described in Taib *et al*., (2013b) to clean and process the raw sequence data as well as provide phylogenetic affiliations for the sequences. The PANAM pipeline is a Perl program designed for a Linux/Unix environment that includes other open source programs and databases (PANGEA (Giongo *et al*., 2010); USEARCH (Edgar, 2010); SILVA (Pruesse *et al*., 2007); HMMER (Eddy, 1998); FASTTREE (Price *et al*., 2010)) and contains reference sequence databases, taxonomy files and the ability to do reference profile alignments (Taib *et al*., 2013), hence we used this pipeline for the processing of the raw DNA sequences as well as clustering into Operational Taxonomic Units (OTUs) and phylogenetic affiliation of these OTUs. We used options in the PANAM pipeline to remove DNA sequences less than 200 bp in length as well as those with a quality score (Phred score) less than 23. DNA sequences were then clustered into OTUs using a threshold of 95% similarity. The 95% threshold was chosen based on Mangot *et al*., (2013), where they demonstrated 95% as a relevant cut-off for OTU delineation. OTUs were then compared to the reference databases within PANAM and assigned to phyletic groups (Taib *et al*., 2013). Taxonomy assignments were given using Nearest Neighbour taxonomy based on a blast of DNA sequences against the SILVA database. The parameters and thresholds used for the processing of the sequences in the PANAM pipeline are shown in Appendix S3.

Once the raw DNA sequences had been processed through the PANAM pipeline, we removed the Embryophyta (green plants) and Metazoa (animals) OTUs from the analysis. Then,

the OTUs retrieved with only one DNA sequence in the whole dataset (i.e., singletons) were also removed from the further analysis. We then rarefied the number of OTUs for each sample to the minimum number of DNA sequences across the (now 15) samples (741 DNA sequence) using *rrarefy*() in the R package `vegan` (Oksanen, 2015).

*Statistical analyses*

Statistical analyses were done in R version 3.1.2 (R Core Team, 2014) and Past3 (Hammer, 2001). After rarefying the number of DNA sequences across the samples for the two lakes we computed alpha diversity metrics for the OTUs in each of the samples including number of OTUs, Shannon diversity, Simpson diversity, Pielou's evenness, and Chao1 in order to track patterns in alpha diversity through time.

To explore the trajectories of change in the eukaryote assemblages preserved in the cores of both Lake Dauriat and Lake Knob, we used Non-metric multidimensional scaling (NMDS) using *metaMDS*() (`vegan`). We used hierarchical clustering analysis (*hclust*() from `vegan`) to identify groups of assemblages based on similarity and then tested for significant differences between groups of samples (by lake and time period) using Analyses of Similarity (ANOSIM) (*anosim*() from `vegan`). We then performed Similarity Percentages (SIMPER) tests (using PAST3) on the samples from each lake to assess the importance of the various OTUs and taxonomic groups in their contribution of the similarity of groups of samples (e.g. by lake).

To examine changes in eukaryote assemblages through time, and to detect specific "breakpoints" at which there were significant changes in eukaryote assemblages in each lake we used multivariate regression tree analyses (*sensu* De'ath, 2002) with the eukaryote OTUs as the response variables and age of sediment as the explanatory variable using *mvpart*() in the R package `mvpart` (https://cran.r-project.org/src/contrib/Archive/mvpart/) for each lake

separately. To determine whether our genetic approaches lead to different conclusions than traditional paleolimnological ones, we also did similar analyses using the diatom assemblage data as reported in Laperrière *et al*., (2008) (this was to compare overall trends between the two datasets, not to compare composition). Laperrière *et al*., (2008) included diatom assemblages characterized using traditional paleolimnological techniques and taxonomy from 37 evenly spaced samples from a Lake Dauriat core retrieved in 1999. Using their diatom data, the response variables in the multivariate regression tree were the diatom species and the explanatory variable was estimated age of the sediment sample. Performing the multivariate regression tree analyses on the paleolimnological diatom data (from Laperrière *et al*., 2008) also allowed us to investigate a potential problem regarding the relatively low temporal density of HTS eukaryote data. The data from the HTS analyses consists of a relatively small sample size (n = 9 for Dauriat and n = 6 for Knob) and using intervals that were not evenly spaced in terms of the years between the samples (as related to time as the explanatory variable). Therefore, comparing the breakpoints obtained from the multivariate regression trees performed on the HTS data with the breakpoints from the regression trees using the paleolimnological diatom data also allowed us to consider the effect of small sample size and uneven sampling intervals.

We assessed temporal beta diversity (turnover through time) in each lake for our HTS eukaryote data by comparing the eukaryote assemblage present in each sample to the next sample (temporally, progressing from the oldest sample to the most recent) using the R function *TBI*() (Legendre, 2015). This function computed the total temporal beta diversity as well as the proportion of total temporal beta diversity explained by the loss and gain of OTUs (Legendre & Salvat, 2015).

**Results**

Alpha diversity indices showed pronounced variation in each lake and different trajectories between lakes, with the most sensitive metrics being Shannon diversity and the number of OTUS (Fig. 1; Appendix S4). The most striking change in OTU richness and diversity in Lake Dauriat was the increase starting circa the 1930s during the start of the iron-ore mining period, which was followed by a decrease during peak iron-ore production in the 1960s. This was subsequently followed by an increase in ~1980s that continued to present times (Fig. 1a). Alpha diversity and OTU richness in Lake Knob showed a generally increasing trend over time (Fig. 1b).

Taxonomic composition of the eukaryotic assemblages, based on Nearest Neighbour (NN) taxonomic classifications varied through time in both Lake Dauriat (Fig. 2) and Lake Knob (Fig. 3). In Lake Dauriat, the Bacillariophyta (diatoms) and Chlorophyta dominated community composition through time, with the Fungi also as prominent member of the assemblages. Between the 1930s and 1970s, Chlorophyta and diatoms appeared to alternate in dominance, with peaks in diatoms around 1945 and 1963. In Lake Knob, the diatoms made up the majority of the community assemblage (upwards of 60%) prior to the 1920s. In the most recent intervals (1980s – 2012), the diatoms did not dominate the assemblages, but remained the richest group of organisms.

The eukaryote assemblages showed marked trajectories in the NMDS (Fig. 4), but with some convergence more recently. Hierarchical clustering analyses of the samples showed clustering of samples prior to the 1970s from both lakes (Fig. 4), and was related to differences in diatom and Chlorophyte taxa (Table 3). However, further testing with ANOSIM showed that these temporal groups were not significantly different. We also found no significant differences

between the assemblages as grouped by time period (three pre-defined time periods: "Pre-mining", pre-1938; "Mining", 1939 to 1977; and "Post-mining", after 1977) or as grouped by lake identity (Dauriat or Knob) using an ANOSIM test ($R \sim 0$ indicating greater differences within groups than between).

Using multivariate regression tree analyses (MRT), breakpoints were identified in the HTS eukaryote assemblages in 1968 and 1987 for Lake Dauriat (note: these years represent the median year between the two samples for which the significant breakpoints were identified). The variation explained by this MRT model was 0.25 (adjusted $R^2$). For Lake Knob, a single assemblage breakpoint was identified at 1955 for the HTS data, with an adjusted $R^2$ of 0.89. Using the (taxonomically identified) diatom assemblage data from Laperrière *et al*., (2008), breakpoints in these assemblages occurred in 1935 and 1977 (Fig. 5), corresponding to the clustering groups as reported in Laperrière *et al*., (2008).

Given that we were interested in biodiversity dynamics, we also compared temporal beta diversity trends for both of the lakes that were calculated using the HTS eukaryotic data and the subfossil assemblages. With the HTS data, we found that turnover peaked between 1963 and 1973 in Lake Dauriat (Fig. 6a), and between 1928 and the early 1980s in Lake Knob (Fig. 6b). The extent to which total temporal beta diversity was explained by the loss or gain of abundances (relative abundance in terms of number of DNA sequences) on an OTU-by-OTU basis varied through time for both lakes, sometimes reaching equivalent proportions. However, in Dauriat, the peak in temporal beta diversity was also associated with a peak in the loss component of the turnover. These peaks in beta diversity during the main mining period were comparable to those observed using traditional taxonomic cladoceran data in Chapter 3 as well as the diatoms for Lake Dauriat from Laperrière et al. (2008) (Fig. 6a&b).

## Discussion

High-throughput sequencing approaches are still in active development. As a result, in the absence of a complete set of calibration studies, they are most powerful when used in conjunction with other methods to study biodiversity. To this end, we employed both methods in a well-studied system so that we could evaluate the robustness of our conclusions. Our ecological question revolved around whether there were differences in the eukaryote assemblages between the two lakes and how these assemblages were affected by disturbances related to mining projects. Overall, we found that there were temporal changes in the eukaryote communities, and as expected, there were differences in each lake. Additionally, both lakes showed a high prevalence of diatom taxa, but Lake Dauriat was unique in its alternating dominance of Chlorophyta with diatom taxa. As phytoplankton are sensitive to changes in nutrient inputs (Watson *et al*., 1997), these results are consistent with the history of anthropogenic disturbance in the region, with Lake Dauriat experiencing significant eutrophication between the 1950s and 1970s (prior to the installation of wastewater treatment facilities) associated with sewage discharge into the lake (Laperrière *et al*., 2008) and both lakes receiving metal inputs.

Given the documented history of Lake Dauriat and earlier paleolimnological work (Laperrière *et al*., (2008) showing changes in diatom communities and Winegardner *et al*., (In prep.; Chapter 3) showing reduced cladoceran richness, we expected a significant reduction in eukaryote richness (number of OTUs in this case) during the main mining period (1939-1977, with town construction occurring ~ 1954), as we predicted that susceptible taxa would be eliminated from the lakes during intense metal loading (as well as eutrophication in the case of Lake Dauriat). There was some evidence for this hypothesis in Lake Dauriat, where there was a

reduction in both the number of OTUs and Shannon Diversity (Appendix S4 and Fig. 1a) in the mid-1960s, corresponding with peak iron-ore production. During the initial mining period, we also observed diversity and the number of eukaryote OTUs in Lake Dauriat increased and this may be as a result of initial nutrient enrichment during initial exploration and development activities. However, once eutrophication and metal enrichment reached a certain threshold (as occurred in the 1950s/1960s), taxa losses were observed. After this taxa reduction, OTU richness and alpha diversity began to increase (with some fluctuation), throughout the remaining period of iron-ore extraction and town development. This increase in the number of eukaryotic OTUs resulted in a peak of OTU richness and diversity in the most recent sample analyzed. In Lake Knob, OTU richness and alpha diversity increased consistently from the late 1800s to present times (Appendix S4 and Fig. 1b). Unfortunately we were only able to refine our core chronology for this lake after the DNA extractions were performed, and thus do not have an adequate sample resolution to quantify the eukaryotic response during the mining period at this site based on HTS methods alone.

Our second set of hypotheses focused on temporal beta diversity. While we expected some degree of temporal beta diversity over the sediment records, we predicted the highest turnover to be during the mining period, with metal loading as a strong driver of assemblage change given that heavy metals are known to cause local extinctions and select for metal-tolerant specialists (replacement through time in addition to simply taxa loss; as was shown for the green alga *Chlorella vulgaris* and copper pollution for example (Foster, 1977). The magnitude of temporal beta diversity peaked at both lakes during the mining period. The peak in temporal beta diversity occurred between 1963 and 1973 in Lake Dauriat (error estimation for years at that depth in the core is between 4 and 6 years), and between 1928 and 1982 in Lake Knob (error estimation = 2-

10 years; although admittedly we only have pre and post-mining samples analysed for Knob). The peak in temporal beta diversity was also higher in terms of magnitude in Lake Dauriat (>90% turnover) relative to Lake Knob (~70% turnover).

There are a number of aspects of our sequencing results that could have been influenced by our methods or by an amplification bias of different groups and taxa; highlighting considerations in future studies of this nature (see Appendix S5 for further details). These considerations include: low sequencing depth, primer specificity, and the process of bioinformatics analysis. However, we believe trends reported herein are robust as the observed changes in the biodiversity of the eukaryote assemblages as characterized by the high-throughput sequencing used in this study were consistent with the trends documented by applying the same analyses to traditional taxonomic data from this system (e.g., Laperrière *et al*., 2008; Winegardner *et al*., (In prep.; Chapter 3).

We recognize that the number of samples for each lake available for the MRT analyses was low, and hence we sought to verify the assemblage breakpoints identified with the MRTs by using a (traditional paleolimnological) diatom-only dataset with a larger sample size and even temporal increments between intervals (the Laperrière *et al*., 2008 dataset; available only for Lake Dauriat). The assemblage break points identified for the (HTS) eukaryotic samples from Lake Dauriat were 1968 and 1987 (approximate years based on $^{210}$Pb dating). The first year, 1968, corresponds approximately with a peak in iron-ore extraction in the 1960s (Bradbury 1984). The second breakpoint, 1987, occurs after wastewater treatment installation was installed in the town (Aebischer *et al*., 2015) and reduced the flow of untreated sewage into Lake Dauriat as well as after final decommissioning of the majority of mine sites. These Lake Dauriat breakpoints were similar but also differed to those identified with the larger diatom-only dataset

from Laperrière *et al*., (2008) of, 1935 and 1977 (again, approximate based on $^{210}$Pb dating). As such, there appears to be agreement in the later breakpoint in that both the eukaryotic HTS assemblages and the taxonomic diatom assemblages appear to respond to a water-quality change brought on in Lake Dauriat around the mid-1970s. It is important to note that this breakpoint was identified in the eukaryotic data with a much smaller number of samples; perhaps pointing to the relative power of using 'whole' biodiversity studies. The first breakpoint identified in the Laperrière et al. (2008) diatom dataset was earlier than that identified with the eukaryote data (~1935) and corresponds roughly to the start of early iron-ore extraction activities in the region. This may be an effect of the data itself, which were more abundant and more evenly spaced for the diatoms, or could reflect the greater sensitivity of algal assemblages to disturbance as opposed to eukaryotic assemblages on the whole (or at least additional information provided by algal assemblages) (Dixit *et al*., 1992; McCormick & Cairns, 1994). For Lake Knob, the only breakpoint identified in the eukaryotic assemblages was in 1955, corresponding with town construction, reflective of the fact that Lake Knob did not experience the eutrophication observed in Lake Dauriat from the 1950s-1970s that was alleviated by the installation of wastewater treatment.

Microbial eukaryotic biodiversity in the Schefferville region has been dominated by diatoms in a lake with a history of ambient metal loading (Lake Knob) and a lake with a history of both ambient metal loading and run-off of construction materials and sewage discharge (Lake Dauriat). In the latter, the Chlorophyta as well as the Fungi and unclassified Stramenopiles have also been important groups in making up the total eukaryotic assemblage. Variation in taxon richness, composition and temporal turnover can all be related to the known anthropogenic history of the area and these results are verified when comparing to a traditional (diatom-only)

paleolimnogical study, thus adding to the increasing volume of work that has documented the influence of mining and development activities on aquatic biodiversity in this region. However, the pilot nature of this study meant that we worked with a low number of study sites and a low sample size within the two lakes of interest. This, combined with issues regarding DNA purity, primer specificity and potential for amplification bias means that we have identified a number of directions for future research on the use of metabarcoding and paleolimnology. These challenges are further combined with active areas of research in the field of metabarcoding and aquatic biomonitoring, namely distinguishing between inactive and active cells, definition of ecologically relevant molecular units and the interpretation of quantitative data from high-throughput sequencing (Pawlowski *et al*., (In press)).

**Acknowledgments**

We wish to thank Dr. Lyle White for his insight during project design, Tanya Tran and Hana Trigui for their help with DNA extractions, and Stéphanie Boilard at the University of Guelph for extensive help and guidance while sequencing samples at the Biodiversity Institute of Ontario. Julien Gogot and Bassam Ghaleb from UQÀM – GEOTOP completed the radiometric dating and Jean-Philippe Jenny provided assistance developing age models. Thank you to Paul del Giorgio and the CarBBAS lab for help during the 2012 field season, and to Oksana Choulik at the McGill Subarctic Research Station for her hospitality and insight into Schefferville's history. This research was supported by an NSERC postgraduate scholarship to AKW, the Northern Scientific Training program (grants to AKW and NS), a QCBS Excellence Award to AKW, a McGill Interdisciplinary Research grant awarded to IGE, Lyle White & BW.

**Literature Cited**

Aebischer S., Cloquet C., Carignan J., Maurice C., & Pienitz R. (2015). Disruption of the geochemical metal cycle during mining: Multiple isotope studies of lake sediments from Schefferville, subarctic Québec. *Chemical Geology,* **412**, 167-178.

Angeler D.G. (2007). Resurrection ecology and global climate change research in freshwater ecosystems. *Journal of the North American Benthological Society*, **26**, 12-22.

Boutet J.S. (2012). An Innu-Naskapi ethnohistorical geography of industrial mining development at Schefferville, Québec. MA thesis, Department of Geography, Memorial University.

Bradbury J. (1984). The impact of industrial cycles in the mining sector: the case of the Québec-Labrador region in Canada. *International Journal of Urban and Regional Research*, **8**, 311-331.

Brendonck L., & De Meester L. (2003). Egg banks in freshwater zooplankton: evolutionary and ecological archives in the sediment. *Hydrobiologia*, **491**, 65-84.

Canadian Council of Ministers of the Environment. (2014). Canadian Environmental Quality Guidelines. Available: http://ceqg-rcqe.ccme.ca/en/index.html#void. Accessed 01 November 2015.

Choulik O., & Moore T.R. (1991). Response of a subarctic lake chain to reduced sewage loading. *Canadian Journal of Fisheries and Aquatic Sciences,* **49**, 1237-1245.

Claesson M.J., Wang Q., O'Sullivan O., Greene-Diniz R., Cole J.R., Ross R.P., & O'Toole P.W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, **38**, e200.

Coolen M.J.L., Orsi W.D., Balkema C., Quince C., Harris K., Sylva S.P., Filipova-Marinova M., Giosan L. (2013). Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. *Proceedings of the National Academy of Sciences,* **110**, 8609-8614.

De'ath G. (2002). Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology*, **83**, 1105-1117.

De Laender F., Verschuren D., Bindler R., Thas O., & Janssen C.R. (2012). Biodiversity of freshwater diatom communities during 1000 years of metal mining, land use, and climate change in Central Sweden. *Environmental Science and Technology*, **46**, 9097-9105.

Derry A.M., Arnott S.E., & Boag P.T. (2010). Evolutionary shifts in copepod acid tolerance in an acid-recovering lake indicated by resurrected resting eggs. *Evolutionary Ecology,* **24**, 133-145.

Dixit S.S., Smol J.P., Kingston J.C., & Charles D.F. (1992). Diatoms: powerful indicators of environmental change. *Environmental Science and Technology,* **26**, 22-23.

Douglas M.S.V., & Smol J.P. (1995). Paleolimnological significance of observed distribution patterns of chrysophyte cysts in arctic pond environments. *Journal of Paleolimnology*, **13**, 79-83.

Eddy S.R. (1998). Profile hidden Markov models. *Bioinformatics,* **14**, 755-763.

Edgar R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* **26**, 2460-2461.

Foster P.L. (1977). Copper exclusion as a mechanism of heavy metal tolerance in a green alga. *Nature,* **269**, 322-323.

Giongo A., Crabb D.B., Davis-Richardson A.G., Chauliac D., Mobberley J.M. *et al*. (2010). PANGEA: pipeline for analysis of next generation amplicons. *Journal of the International Society for Microbial Ecology*, **4**, 852-861.

Gregory-Eaves I., & Beisner B.E. (2011). Paleolimnological insights for biodiversity science: an emerging field. *Freshwater Biology*, **56**, 2653-2661.

Gregory-Eaves I., & Domaizon I. (2014). Analysis of DNA archived in lake sediments. ASLO e-lecture. doi: 10.4319/lol.2014.igregory-eaves_idomaizon.7.

Hammer Ø., Harper D.A.T., & Ryan P.D. (2001). PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4: 9pp. Available online: http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Hamrová E., Goliáš V., & Petrusek A. (2010). Identifying century-old long-spined *Daphnia*: species replacement in a mountain lake characterised by paleogenetic methods. *Hydrobiologia,* **643**, 97-106.

Kerfoot W.C., Robbins J.A., & Weider L.J. (1999). A new approach to historical reconstruction: combining descriptive and experimental paleolimnology. *Limnology and Oceanography,* **44**, 1232-1247.

Kerfoot W.C., & Weider L.J. (2004). Experimental paleoecology (resurrection ecology): chasing Van Valen's Red Queen hypothesis. *Limnology and Oceanography,* **49**, 1300-1316.

Korosi J.B., & Smol J.P. (2012). An illustrated guide to the identification of cladoceran subfossils from lake sediments in northeasterm North America: part 1- the Daphniidae, Leptodoridae, Bosminidae, Polyphemidae, Holopedidae, Sididae, and Macrothricidae. *Journal of Paleolimnology,* 10.1007/s10933-012-9632-3.

Laperrière L., Fallu M-A., Hausmann S., Pienitz R., & Muir D. (2008). Paleolimnological evidence of mining and demographic impacts on Lac Dauriat, Schefferville (subarctic Québec, Canada). *Journal of Paleolimnology*, **40**, 309-324.

Leavitt P.R., & Hodgson D.A. (2001). Sedimentary Pigments in Smol, J.P., Birks, H.J.B., and Last, W.M. (eds) *Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Siliceous Indicators*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Legendre P. (2015). R function. TBI(). Available online:http://adn.biol.umontreal.ca/~numericalecology/FonctionsR/

Legendre P., & Salvat B. (2015). Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.

Mangot J-F., Domaizon I., Taib N., Marouni N., Duffaud E., Bronner G., & Debroas D. (2013). Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental Microbiology*, **15**, 1745-1758.

McCormick P.V., & Cairns Jr., J. (1994). Algae as indicators of environmental change. *Journal of Applied Phycology,* **6**, 509-526.

Montero-Pau J., Gómez A., & Muñoz J. (2008). Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography: Methods,* **6**, 218-222.

NanoDrop. (2007). Technical Support Bulletin: 260/280 and 260/230 Ratios, NanoDrop® ND-1000 and ND-8000 8-Sample Spectrophotometers. Available online: http://www.bio.davidson.edu/gcat/protocols/NanoDrop_tip.pdf.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., & Wagner H. (2015). Package 'vegan'. Available: http://cran.r-project.org/web/packages/vegan/vegan.pdf

Pal S., Gregory-Eaves I., & Pick F.R. (2015). Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. *Journal of Paleolimnology*, **54**, 87-101.

Pawlowski J., Lejzerowicz F., Apotheloz-Perret-Gentil L., Visco J., & Esling, P. (In press). Protist metabarcoding and environmental biomonitoring: time for change. *European Journal of Protistology*, 10.1016/j.ejop.2016.02.003

Price M.N., Dehal P.S., & Arkin A.P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE,* **5**, e9490. doi: 10.1371/journal.pone.0009490.

Pruesse E., Quast C., Knittel K., Fuchs B.M., Ludwig W. *et al*. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research,* **35**, 7188-7196.

Quinlan R., Douglas M.S.V., & Smol J.P. (2005). Food web changes in arctic ecosystems related to climate warming. *Global Change Biology,* **11**, 1381-1386.

Quinlan R., & Smol J.P. (2001). Chironomid-based inference models for estimating end-of-summer hypolimnetic oxygen from south-central Ontario shield lakes. *Freshwater Biology,* **46**, 1529-1551.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org/

Schauer M., Balagué V., Pedrós-Alió C., & Massana R. (2003). Seasonal changes in taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquatic Microbial Ecology*, **31**, 163-174.

Schwark L., Zink K., & Lechterbeck J. (2002). Reconstruction of postglacial to early Holocene vegetation history in terrestrial Central Europe via cuticular lipid biomarkers and pollen records from lake sediments. *Geology*, **30**, 463-466.

Seddon A.W.R., Mackay A.W., Baker A.G. *et al*. (2014). Looking forward through the past: identification of 50 priority research questions in paleoecology. *Journal of Ecology,* **102**, 256-267.

Shokralla S., Gibson J.F., Nikbakht H., Janzen D.H., Hallwachs W., & Hajibabaei M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources,* **14**, 892-901.

Taberlet P., Coissac E., Pompanon F., Brochmann C., & Willerslev E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology,* **21**, 2045-2050.

Taib N., Bronner G., & Debroas D. (2013a). PANAM: Phylogenetic Analysis of Next generation AMplicons. Read Me. Available online: https://github.com/panammeb/

Taib N., Mangot J-F., Domaizon I., Bronner G., & Debroas D. (2013b). Phylogenetic affilitation of SSU rRNA genes generated by massively parallel sequencing: New insights into the freshwater protist diversity. *PLoS ONE,* **8**, e58950, 10.1371/journal.pone.0058950.

Thienpont J.R., Korosi J.B., Cheng E.S., Deasley K., Pisaric M.F.J., & Smol J.P. (2015). Recent climate warming favours more specialized cladoceran taxa in western Canadian Arctic lakes. *Journal of Biogeography,* **42**, 1553-1565.

Vandekerkhove J., Declerck S., Vanhove M., Brendonck L., Jeppesen E., Conde Porcuna J-M., & De Meester J. (2004). Use of ephippial morphology to assess richness of anomopods: potentials pitfalls. *Journal of Limnology*, **63**, 75-84.

Watson S.B., McCauley E., & Downing J.A. (1997). Patterns in the phytoplankton composition across temperate lakes of differing nutrient status. *Limnology and Oceanography,* **42**, 487-495.

Winegardner A.K., Salter N., Aebischer S., Pienitz R., Derry A.M., Beisner B.E., & Gregory-Eaves I. (In prep; Thesis Ch. 3).Cladoceran zooplankton diversity in lakes from a northern mining region: responses to multiple stressors characterized by alpha and beta diversity.

Zimmerman J., Jahn R., & Gemeinholzer B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution,*10.1007/s13127-011-0050-6.

**Tables**

**Table 1: Depth of sediment samples from the Schefferville cores and DNA quality information as measured by NanoDrop ®.** 'Concentration' refers to DNA concentration; 'OD260/280' refers to the ratio of absorbance at 260 nm and 280 nm and is a measure of the purity of DNA and RNA where ~1.8 is generally accepted as pure for DNA (~ 2 for RNA); 'OD260/230' refers to a measure of nucleic acid purity (generally 2.0-2.2) (NanoDrop, 2007).

| Lake | Median interval depth (cm) | Estimated median year (CIC model) | Estimated year error | Concentration (ng $\mu L^{-1}$) | OD260/280 | OD260/230 |
|---|---|---|---|---|---|---|
| **Dauriat** | 0.5 | 2011 | ~0 | 102.6 | 2.052 | 2.08 |
| **Dauriat** | 3.25 | 2004 | 1 | 167.1 | 1.87 | 2.09 |
| **Dauriat** | 7.25 | 1994 | 2.2 | 40.7 | 1.75 | 1.63 |
| **Dauriat** | 12.5 | 1981 | 2.6 | 20.9 | 1.70 | 1.41 |
| **Dauriat** | 16 | 1973 | 4 | 14.6 | 1.64 | 1.46 |
| **Dauriat** | 20.25 | 1963 | 6 | 23.6 | 1.80 | 1.41 |
| **Dauriat** | 25.5 | 1950 | 12.1 | 44.9 | 1.72 | 1.64 |
| **Dauriat** | 27.5 | 1945 | 15 | 21.7 | 1.55 | 0.91 |
| **Dauriat** | 33.5 | 1930 | 15 | 8.8 | 1.47 | 0.43 |
| **Dauriat** | 37.5 | 1920 | 15 | 22.5 | 1.13 | 0.91 |
| **Knob** | 0.5 | 2012 | ~0 | 122.6 | 1.84 | 1.84 |
| **Knob** | 2.75 | 1982 | 2.4 | 136.6 | 1.82 | 1.89 |
| **Knob** | 8 | 1928 | 10 | 78.1 | 1.74 | 1.41 |
| **Knob** | 14 | 1866 | 15 | 35.1 | 1.57 | 1.00 |
| **Knob** | 17.25 | 1833 | 15 | 38.3 | 1.50 | 0.84 |
| **Knob** | 26.75 | 1736 | 15 | 24.5 | 1.54 | 0.89 |

**Table 2: Sequencing success represented as number of DNA sequences for the two core profiles (singleton sequences removed).** The analyses in this study focus on eukaryotic sequences (as amplified by the general eukaryotic primer), however sequencing success for the diatom-specific is also shown. The deepest Dauriat sample* was removed from analyses because of low sequencing success in terms of the number of eukaryotic sequences.

| Sample | # of eukaryotic sequences | # of diatom sequences |
|---|---|---|
| **Dauriat- 0.5** | **741** | 235 |
| **Dauriat- 3.25** | **1070** | 425 |
| **Dauriat- 7.25** | **1267** | 359 |
| **Dauriat- 12.5** | **1199** | 55 |
| **Dauriat- 16** | **1366** | 14 |
| **Dauriat- 20.25** | **3525** | 129 |
| **Dauriat- 25.5** | **2924** | 147 |
| **Dauriat- 27.5** | **2987** | 118 |
| **Dauriat- 33.5** | **4941** | 56 |
| **Dauriat- 37.5*** | **592** | 4062 |
| **Knob- 0.5** | **2239** | 1172 |
| **Knob- 2.75** | **4601** | 2780 |
| **Knob- 8** | **4149** | 859 |
| **Knob- 14** | **3920** | 537 |
| **Knob- 17.25** | **6288** | 626 |
| **Knob- 26.75** | **3207** | 459 |

**Table 3: OTUs contributing to dissimilarity between groups of samples as defined by hierarchical clustering.**

| OTU | Avg. dissimilarity | Contr. (%) | Cumulative contr. (%) | Taxonomic group | Functional group | Pigmentation |
|---|---|---|---|---|---|---|
| **OTU1430** | 22.8 | 27.3 | 27.3 | Bacillariophyta | Autotroph | Pigmented |
| **OTU2325** | 4.7 | 5.7 | 32.9 | Chlorophyta | Autotroph | Pigmented |
| **OTU1470** | 3.3 | 3.9 | 36.8 | Chlorophyta | Autotroph | Pigmented |
| **OTU1770** | 2.7 | 3.3 | 40.1 | Fungi | Unknown | Non-pigmented |
| **OTU1395** | 1.2 | 1.4 | 41.5 | Chlorophyta | Autotroph | Pigmented |
| **OTU105** | 1.1 | 1.3 | 42.8 | Bacillariophyta | Unknown | Pigmented |
| **OTU1266** | 1.0 | 1.2 | 44.0 | Bacillariophyta | Autotroph | Pigmented |
| **OTU1788** | 0.9 | 1.3 | 45.1 | Chlorophyta | Autotroph | Pigmented |
| **OTU864** | 0.9 | 1.1 | 46.2 | Dinophyceae | Autotroph | Pigmented |
| **OTU2288** | 0.9 | 1.0 | 47.2 | Labyrinthulida | Parasite | Non-pigmented |

**Figures**

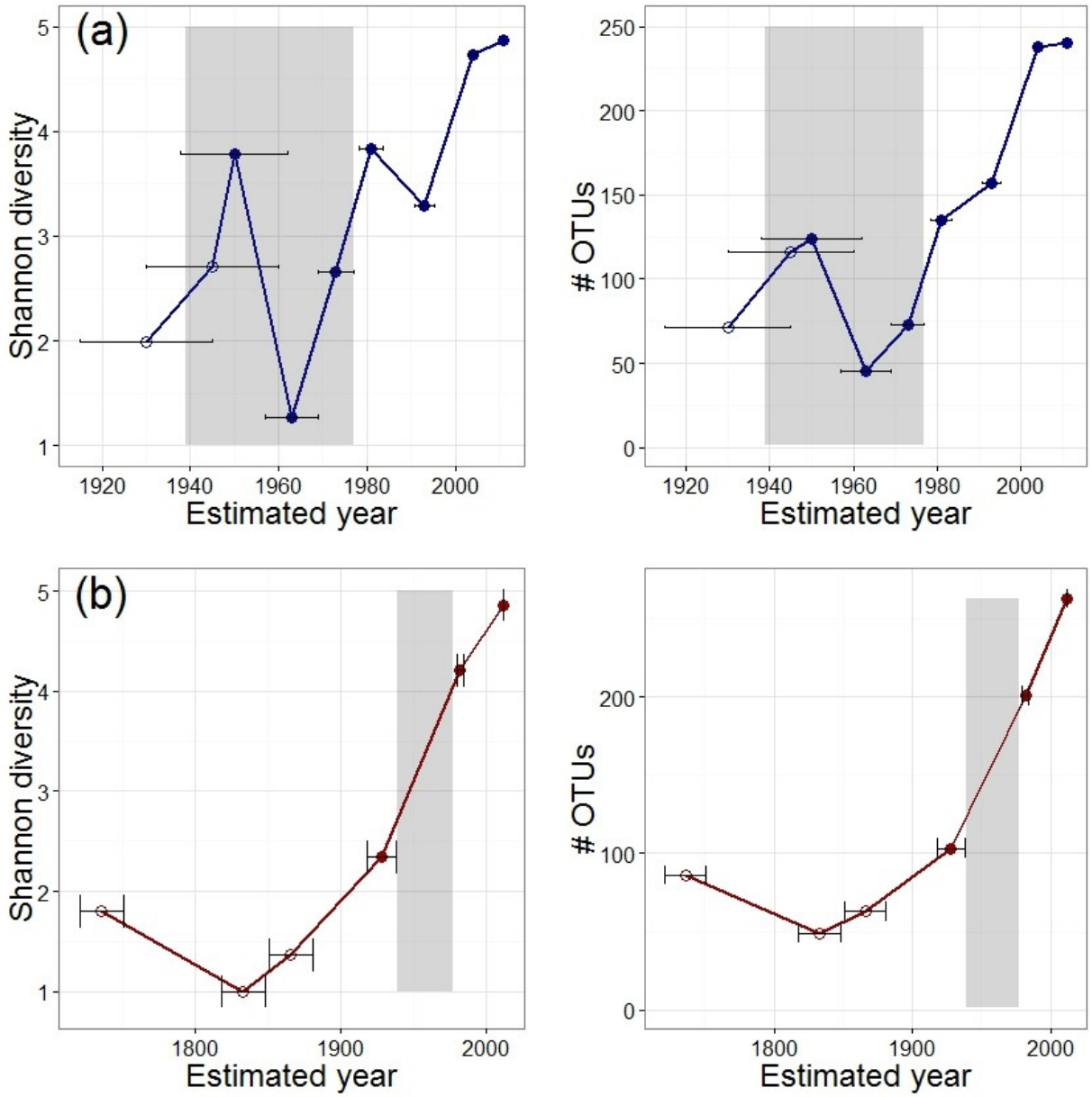



**Figure 1: Selected alpha diversity metrics for (a) Lake Dauriat and (b) Lake Knob through time.** Both rarefied number of Operational Taxonomic Units (OTUs) and Shannon-Weiner diversity are shown. Horizontal error bars represent error associated with radiometric dating. Filled circles are samples where the age estimation has been made directly from $^{210}$Pb decay and the associated age model (see Winegardner *et al*., Chapter 3 for more details). Open circles are

samples where the age estimation has been extrapolated from the age-specific sedimentation rates estimated from the selected age models.

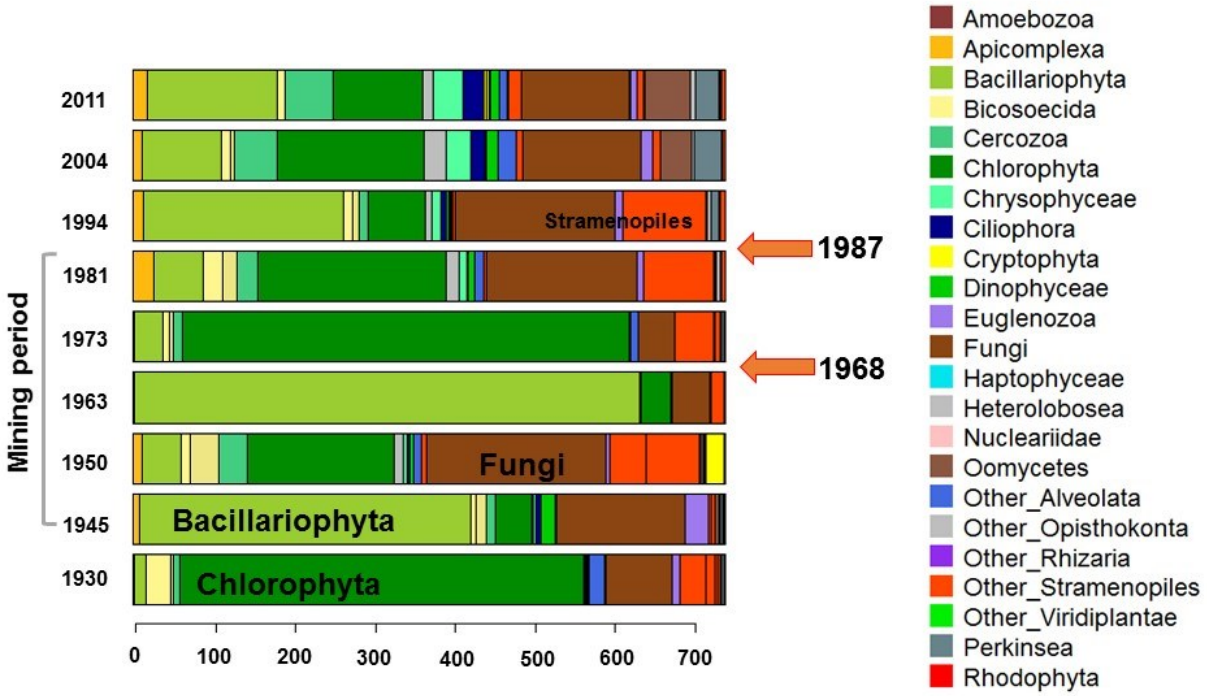



**Figure 2: Eukaryote assemblages in Lake Dauriat through time.** Assemblages are characterized by the number of sequences in a taxonomic group in each sample (OTUs classified to taxonomic groups), to the rarefied level of 741 sequences. Arrows indicate the assemblage breakpoints identified by a multivariate regression tree (Adj. $R^2 = 0.25$): 1955 and 1979.

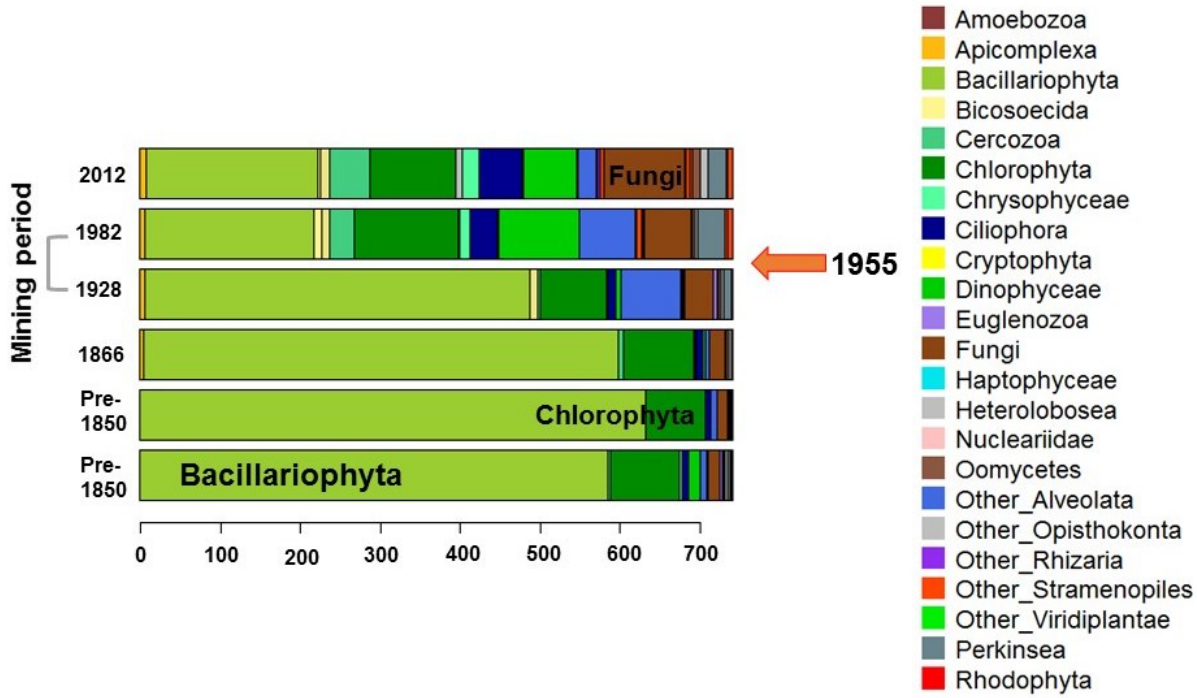



**Figure 3: Eukaryote assemblages in Lake Knob through time.** Assemblages are characterized by the number of sequences in a taxonomic group in each sample (OTUs classified to taxonomic groups), to the rarefied level of 741 sequences. The arrow indicates the assemblage breakpoint identified by a multivariate regression tree (Adj. $R^2 = 0.89$): 1955.

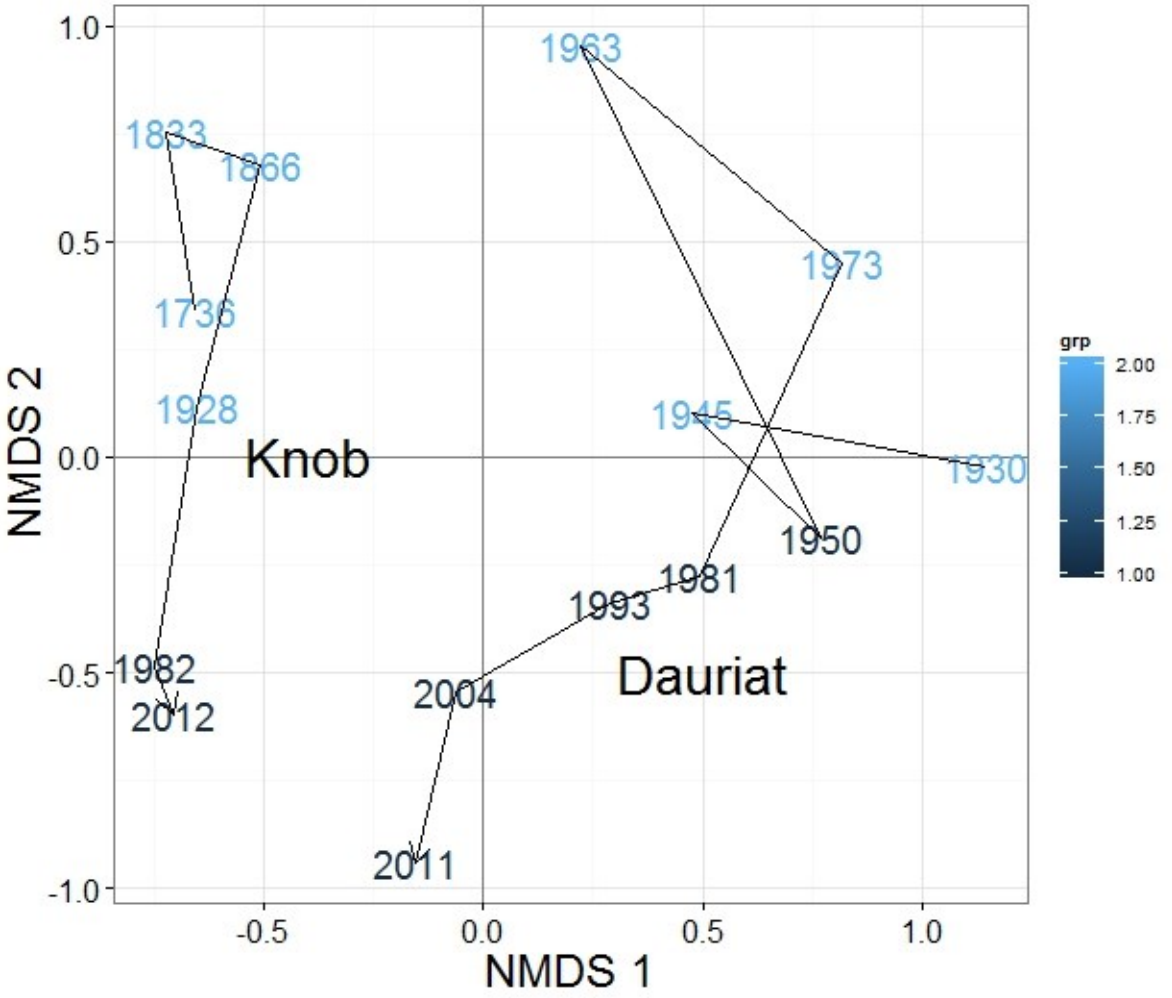



**Figure 4: NMDS biplot of the eukaryote assemblages from the HTS data for Lake Dauriat (labelled) and Lake Knob.** Two distinct groups identified by hierarchical cluster analysis are identified by shading as denoted by the legend 'grp'. While these groups were delineated by the hierarchical clustering, they were not found to be significantly different using ANOSIM analysis.

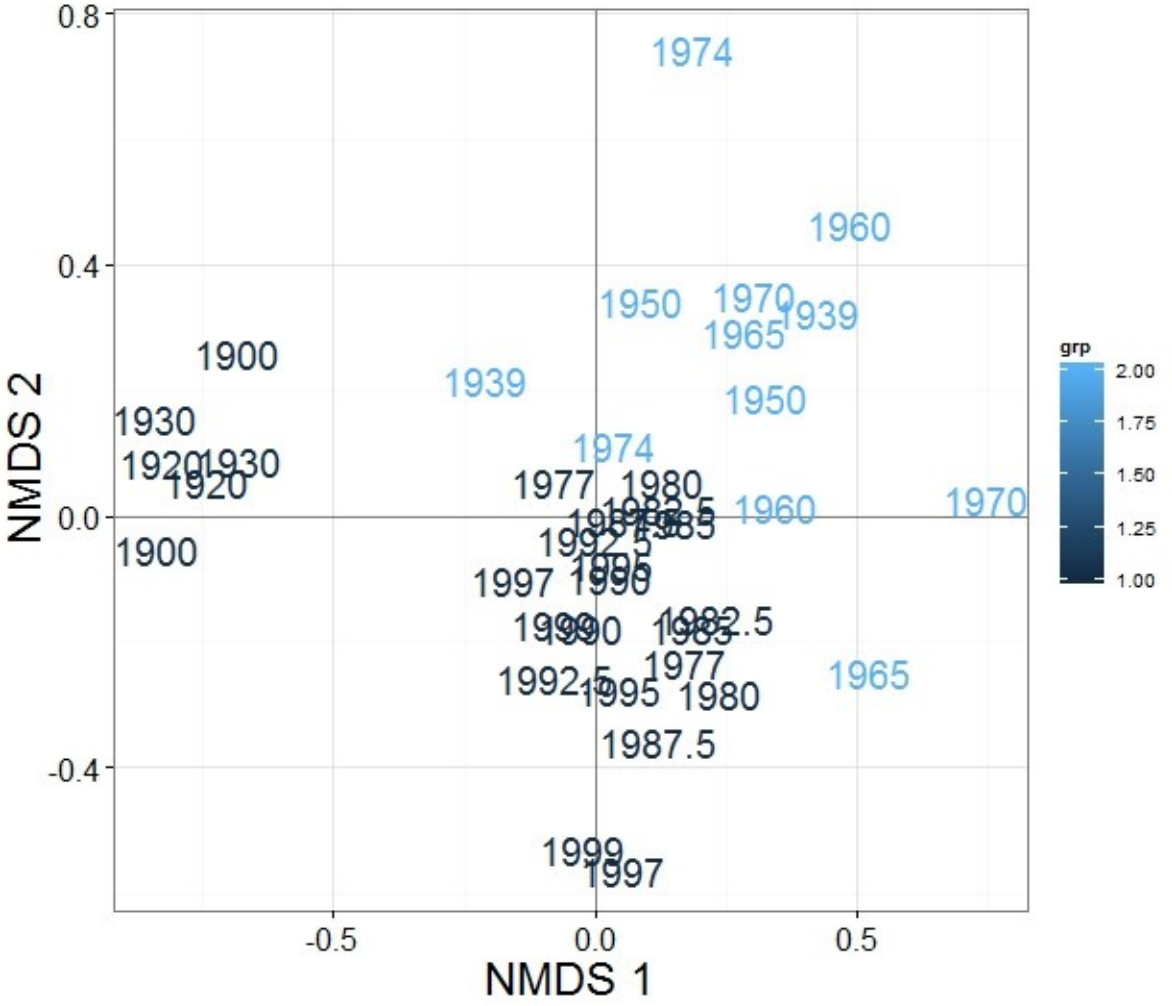



**Figure 5: NMDS biplot of the diatom assemblages (from traditional taxonomy) for Lake Dauriat as obtained from Laperrière *et al.*, (2008).** Groups identified by hierarchical clustering are identified by shading ('grp'). These clusters are also delineated by assemblage breakpoints identified by a multivariate regression tree (1935 and 1977), Adj. $R^2 = 0.7$.

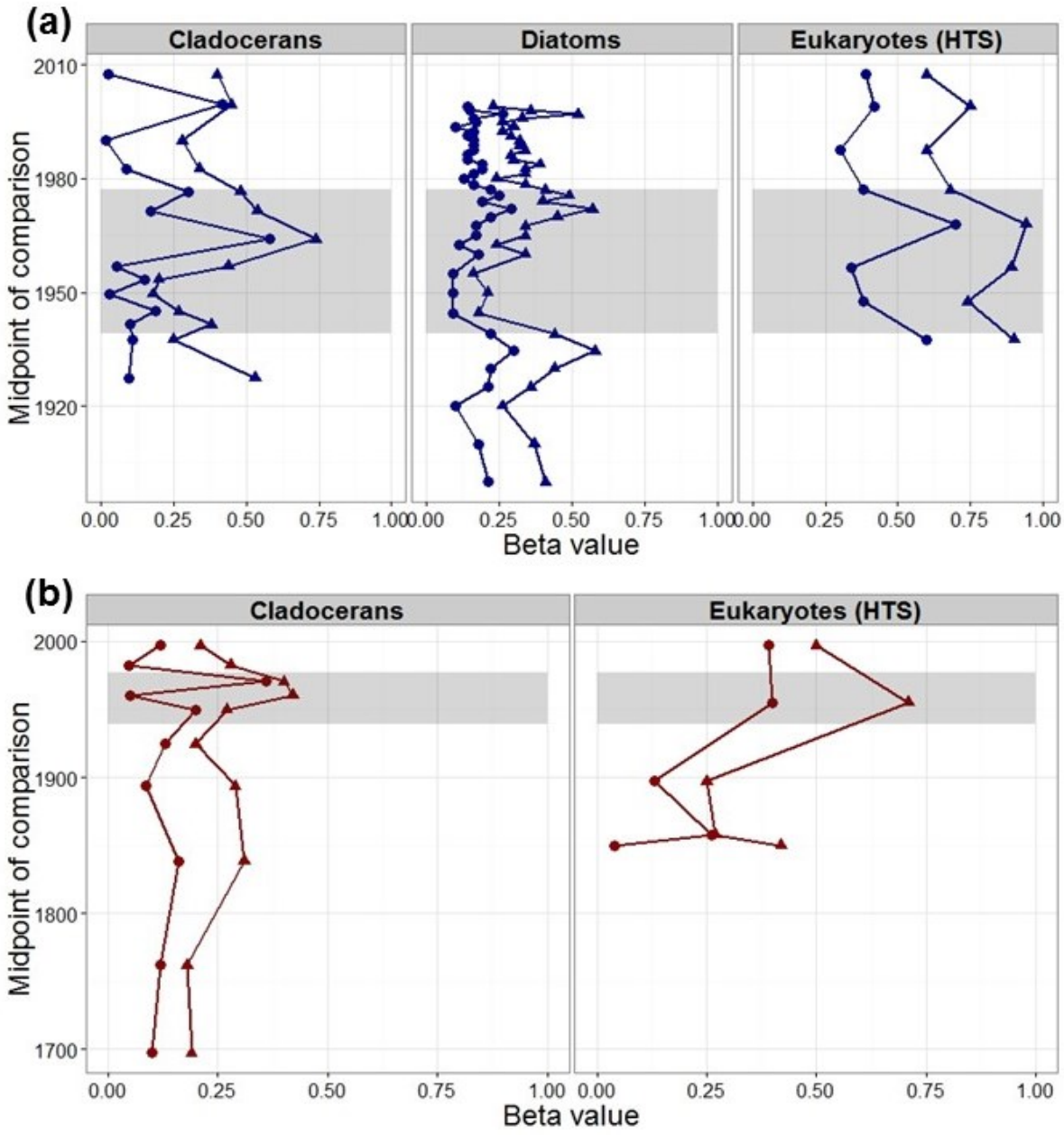



**Figure 6: Temporal beta diversity for (a) Lake Dauriat's and (b) Lake Knob.** Data shown for Lake Dauriat includes cladoceran assemblages as characterized by traditional taxonomy (Thesis, Chapter 3), diatom assemblages as characterized by traditional taxonomy (Laperrière *et al*., 2008), and eukaryotic assemblages characterized by HTS. Data shown for Lake Knob includes cladoceran communities and eukaryotic communities (HTS). Total temporal beta diversity for both (a) and (b) is shown by the triangle points, whereas the circle points represent

the proportion of the total beta diversity explained by the loss of taxa (individual OTUs in for eukaryotes and taxa for cladocerans, also called the “loss component”. The “gain component” is the inverse of the loss component. The main mining period is shaded.

# GENERAL CONCLUSION

This thesis combines four separate manuscripts that inform biodiversity study of the Anthropocene. I chose to make paleolimnology a central part of my work, in order to investigate temporal scales inaccessible via other types of data. The incorporation of these relatively long temporal scales is important because it allows one to more fully characterize biodiversity dynamics in the Anthropocene. Long temporal scales and larger spatial scales are important in biodiversity study because many stressors to biodiversity are occurring globally (for example, see Steffen *et al*., (2011) for a full list of stressors increasing on the global scale). Many of these global stressors also show variation in their magnitudes and effects at local levels (e.g. climate change, MacDonald, 2010). And so, the global nature of many stressors to biodiversity, does not diminish the importance of considering local scales as well. Because of this local variation, and the need to examine biodiversity over long time periods, I used this thesis to progress through various types of temporal and spatial scaling. I started with the calibration of paleolimnological and water-column samples to understand how environmental and spatial drivers vary across these different types of samples (Chapter 1). I then looked at diatom biodiversity across a fairly large spatial scale (conterminous United States), coupling these spatial analyses with studies of temporal beta diversity over the last 150 years (longer than many studies included in the meta-analyses highlighted earlier in this thesis) at a historical and modern time point (Chapter 2).

In later chapters (Chapters 3 and 4), I examined temporal beta diversity at a higher resolution (multiple time points) and finally, I used metabarcoding methods to examine microbial eukaryote biodiversity (i.e., fungi and protists) as a whole in a metal-contaminated region. Exploring biodiversity, and especially beta diversity at these different scales helped to illuminate some key findings related to the McGill *et al*., (2015) conceptual framework for predicting biodiversity trends in the Anthropocene.

I found that diatom assemblages reconstructed from surface sediments were structured similarly across environmental and spatial gradients to those sampled from the water-column. This means that there is great potential for integration of data from both paleolimnological and neo-ecological sources, not just in studies where sampling using both of these methodologies is paired together, but also in studies that integrate data from both sources (such as for meta-analyses). For example, Battarbee *et al*., (2007) compiled a database of paleolimnological studies across Europe, noting where time-series data for the sampled sites were also available. Databases such as this could be used to complete further studies calibrating monitoring versus paleolimnological studies, but could also be used to answer questions across large environmental gradients, using data from the various sources to increase statistical power.

In addition to finding similarities across environmental and spatial gradients for water-column versus paleolimnological diatom samples, I also quantified the extent to which diatom beta diversity across the conterminous US varies both spatially and temporally. I found that the magnitude of diversity change across space and time depended on specific factors. Sites with exceptional contributions to spatial diversity were characterized as being those with low conductivity whereas sites that had exceptional temporal turnover were those found in areas with low forest cover. This latter finding is important because it shows that beta diversity is not always directional in nature, i.e. it is not a given that spatial beta diversity across the US landscape is greater in modern times than in historical (at least for diatoms), but that many factors can be found to influence turnover, both historically and in the present. This is slightly different to studies in Arctic ecosystems, such as Smol *et al*., (2005), where beta diversity of algae and invertebrate communities in circumpolar lakes continually increased (across many time points in the sediment record) from historical times to present. This is likely because

climate warming has had an omnibus effect on diversity in the Arctic, to the extent that has not been experienced yet in temperate systems.

At a local scale, I found metal loading from mining contamination in Schefferville, Québec to be a key driver of low zooplankton richness in combination with high temporal change in community composition. In this case, trends in taxa richness were congruent with those of beta diversity in that both pointed towards increased change in zooplankton communities during metal and sewage contamination. The use of both of these different indices (taxa richness and temporal beta diversity), were key in identifying the effect of mining associated activities on zooplankton communities, characterizing the impact on taxa richness as well as community composition. In addition to the utility of using different biodiversity indices to understand effects on diversity in the Schefferville system, being able to characterize the biodiversity of the Schefferville system in a more complete way, via metabarcoding also proved useful and showed that several groups of microbial eukaryotes had similar trends in beta diversity to that of cladoceran zooplankton. With more work focusing on the calibration of DNA-based approaches with sediment records, these methods have considerable potential to expand the use of paleolimnology in biodiversity research.

*Significance of findings and future directions*

This thesis has made strong contributions to the study of beta diversity as well as the understanding of drivers that affect aquatic biodiversity across the conterminous U.S. and at local scales in historically mined regions. Chapter 2 lays out an interesting analytical framework for assessing beta diversity both spatially and temporally that could be employed for more robust datasets. I showed proof of concept that temporal beta diversity can be partitioned into

explanatory comments and that the significance of individual site contributions to temporal beta diversity can be quantified (Temporal Beta Diversity Index) in similar ways to established methods for spatial beta diversity (Local Contributions to Beta Diversity). These approaches are ripe for expansion and could lead to very unique insight into mechanisms structuring both spatial and temporal beta diversity in many different systems. In order to build on this foundation, future work should focus on utilizing these methods on data from additional organismal groups, as well as higher resolution taxonomic data and time-series data.

Chapters 1 and 2 are both based on diatom data from the U.S. EPA National Lakes Assessment. While diatoms are accepted as strong paleolimnological proxies to reconstruct changes in water quality and environmental conditions, some researchers consider them to be ubiquitous across landscapes and therefore argue that spatial variables will always trump environmental drivers when explaining biodiversity change (Heino *et al*., 2009). As such, expanding the beta diversity analyses developed in this work to other aquatic groups, would allow for the examination of differences in how various aquatic groups exhibit beta diversity and begin to develop an understanding of indicator groups for beta diversity. For example, are there some organismal groups affected by specific land use changes more than other in terms of beta diversity? Are there specific groups that show temporal or spatial turnover along environmental gradients prior to other groups? Candidate data for this could include data already available from the National Lakes Assessment, including soft phytoplankton and rotifers (the latter of which is well-characterized to species level for the 2007 assessment). The use of pre-existing NLA data would continue to develop a picture of aquatic beta diversity across the conterminous United States, but it's important to note that the methods developed in this work would be of equal use to one working in terrestrial systems.

Datasets with higher or more certain taxonomic resolution, or those with long and well populated time series (such as the North American Breeding Bird Survey), would also be ideal jumping off points for future beta diversity with the analyses shown in Chapters 2-4. Time-series data would allow for an in-depth study of temporal beta diversity components and the potential for switching through time between niche and neutral processes (as discussed in Chapter 3). Datasets where historical abiotic data is available in addition to biological data would also build on the work presented in this thesis because it would mean that historical spatial diversity could be related to historical land use and environmental conditions, and that temporal beta diversity could be related to land use change as opposed to modern land use measurements.

In addition to growing the body of knowledge related to beta diversity, this thesis explored the use of DNA-based methods in paleolimnology. These applications were exploratory in nature, and could be significantly improved in order to better apply the analyses developed in this thesis to robust genetic data. The DNA barcoding of resting stages mentioned in Chapter 3 provided some background and context to consider when thinking about cladoceran dynamics between the two study lakes. Further development of DNA extraction procedures for diapause resting eggs and cases could lead to an in-depth characterization of the resting egg bank present in these lakes, and allow for a more thorough examination how multiple stressors might cause the deposition or emergence of cladoceran zooplankton from these eggs, and how this might relate to the observed patterns of temporal beta diversity. In Chapter 4, the characterization of microbial eukaryotes through time using high-throughput sequencing (HTS) showed patterns consistent with taxonomically identified cladocerans and diatoms, however this study initially also included a more in depth study of both diatom and bacterial assemblages using HTS. In the

end, this work was out of scope of the current thesis, but future work should endeavour to pair HTS of multiple aquatic groups with taxonomic data.

*Raw data and further information on data sources*

Information on online data archives as well as raw data for this thesis are found in Appendix E.

## Literature Cited


Battarbee R.W., Morley D., Bennion H., & Simpson G.L. (2007). A meta-database for recent paleolimnological studies. *PAGES News,* **15**, 23-24.

Heino J., Bini L.M., Karjalainen S.M., Mykrä H., Soininen J., Vieira L.C.G., & Diniz-Filho J.A.F. (2009). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos,* **119**, 129-137.

McGill B.J., Dornelas M., Gotelli N.J., & Magurran A.E. (2015). Fifteen forms of biodiversity trend in the Antropocene. *Trends in Ecology and Evolution,* **30**, 104-113.

Smol J.P., Wolfe A.P., Birks H.J.B., Douglas M.S.V., *et al*. (2005). Climate-driven shifts in the biological communities of arctic lakes. *Proceedings of the National Academy of Sciences,* **102**, 4397-4402.

**APPENDIX**

**Appendix A: Supporting information for Chapter 1**

**Ch.1 S1: Metadata available as a separate file in online Supporting Information (*Freshwatre Biology*).** Data matrices from metadata used to create the following data matrices for statistical analyses.

*Metadata*

- Metadata for archived data: nla2007_Winegardner_metadata.txt

| Matrix name | Type | Sample size (N) | Description |
|---|---|---|---|
| **Sites** | Main | 468 | Site names of all lakes, corresponding latitude and longitude coordinates, state, ecoregion, lake-origin. |
| **Environment** | Main | 468 | Water quality variables measured from water-column (contemporary sampling) and Zmax. |
| **Surface sediment diatoms** | Main | 468 | Relative abundance of diatom species from surface sediment samples. Surface sediment generally collected from first visit to a site (May to June). |
| **Water-column diatoms** | Main | 468 | Relative abundance of diatom species from water-column samples. Collected from first visit to a site (May to June). |
| **Water-column diatoms (subset V1)** | Subset | 51 | Relative abundance of diatom species from water-column samples. Subset of main "Contemporary diatoms" data matrix. Only sites with a second round of |

| | | | |
|---|---|---|---|
| | | | diatom sampling included. |
| **Water-column diatoms (subset V2)** | Subset | 51 | Relative abundance of diatom species from water-column samples. Collected from a second visit to site (generally August to September). |
| **Water-column diatoms (subset mean)** | Subset (manipulated) | 51 | Relative abundance of diatom species from water-column samples. Average between first and second visit. |
| **Surface sediment diatoms (subset)** | Subset | 51 | Relative abundance of diatom species from surface sediment samples. Subset of main "Surface sediment diatoms" data matrix. 51 sites to match subsets of contemporary diatoms. |

**Ch. 1 S2: Classification of diatom taxa as benthic, planktontic and tychoplanktonic.** Classification of the diatom species found in this study was made using the following online databases: "Diatoms of the United States": http://westerndiatoms.colorado.edu/, "ANSP Algae Image Database": http://diatom.ansp.org/algae_image/SearchCriteria.aspx.

| Species | Classification |
|---|---|
| ***Achnanthes* cf. *levanderi*** | benthic |
| ***Achnanthes curtissima*** | benthic |
| ***Achnanthes gracillima*** | benthic |
| ***Achnanthes hauckiana*** | benthic |
| ***Achnanthes minutissima*** | benthic |
| ***Achnanthes pseudoswazi*** | benthic |
| ***Achnanthes* spp.** | benthic |
| ***Achnanthidium altergracillima*** | benthic |
| ***Achnanthidium caledonicum*** | benthic |
| ***Achnanthidium catenatum*** | benthic |
| ***Achnanthidium* cf. *gracillimum*** | benthic |
| ***Achnanthidium* cf. *strictum*** | benthic |
| ***Achnanthidium deflexum*** | benthic |
| ***Achnanthidium eutrophilum*** | benthic |
| ***Achnanthidium exiguum*** | benthic |
| ***Achnanthidium jackii*** | benthic |
| ***Achnanthidium minutissimum*** | benthic |
| ***Achnanthidium rivulare*** | benthic |
| ***Actinella punctata*** | benthic |
| ***Actinocyclus normanii*** | tychoplanktonic |
| ***Adlafia bryophila*** | benthic |
| ***Amphicampa eruca*** | benthic |
| ***Amphipleura pellucida*** | benthic |
| ***Amphora* cf. *libyca*** | benthic |
| ***Amphora copulata*** | benthic |
| ***Amphora fogediana*** | benthic |
| ***Amphora inariensis*** | benthic |
| ***Amphora montana*** | benthic |
| ***Amphora ovalis*** | benthic |
| ***Amphora pediculus*** | benthic |
| ***Amphora* spp.** | benthic |
| ***Amphora thumensis*** | benthic |
| ***Amphora veneta*** | benthic |
| ***Aneumastus tuscula*** | benthic |
| ***Anomoeoneis brachysira*** | benthic |
| ***Anomoeoneis sphaerophora*** | benthic |

| | |
|---|---|
| ***Asterionella formosa*** | planktonic |
| ***Asterionella ralfsii*** | planktonic |
| ***Asterionella* spp.** | planktonic |
| ***Aulacoseira alpigena*** | tychoplanktonic |
| ***Aulacoseira ambigua*** | tychoplanktonic |
| ***Aulacoseira crassipunctata*** | tychoplanktonic |
| ***Aulacoseira crenulata*** | tychoplanktonic |
| ***Aulacoseira distans*** | tychoplanktonic |
| ***Aulacoseira granulata*** | tychoplanktonic |
| ***Aulacoseira herzogii*** | tychoplanktonic |
| ***Aulacoseira islandica*** | tychoplanktonic |
| ***Aulacoseira italica*** | tychoplanktonic |
| ***Aulacoseira lacustris*** | tychoplanktonic |
| ***Aulacoseira laevissima*** | tychoplanktonic |
| ***Aulacoseira muzzanensis*** | tychoplanktonic |
| ***Aulacoseira nygaardii*** | tychoplanktonic |
| ***Aulacoseira perglabra*** | tychoplanktonic |
| ***Aulacoseira pfaffiana*** | tychoplanktonic |
| ***Aulacoseira pusilla*** | tychoplanktonic |
| ***Aulacoseira* spp.** | tychoplanktonic |
| ***Aulacoseira subarctica*** | tychoplanktonic |
| ***Aulacoseira subborealis*** | tychoplanktonic |
| ***Aulacoseira tenella*** | tychoplanktonic |
| ***Aulacoseira valida*** | tychoplanktonic |
| ***Bacillaria paradoxa*** | benthic |
| ***Biremis circumtexta*** | benthic |
| ***Brachysira apiculata*** | benthic |
| ***Brachysira brebissonii*** | benthic |
| ***Brachysira microcephala*** | benthic |
| ***Brachysira vitrea*** | benthic |
| ***Caloneis amphisbaena*** | benthic |
| ***Caloneis silicula*** | benthic |
| ***Campylodiscus clypeus*** | tychoplanktonic |
| ***Chaetoceros* spp.** | planktonic |
| ***Chamaepinnularia begeri*** | benthic |
| ***Chamaepinnularia bremensis*** | benthic |
| ***Cocconeis pediculus*** | benthic |
| ***Cocconeis placentula*** | benthic |
| ***Coscinodiscus* spp.** | benthic |
| ***Craticula ambigua*** | benthic |
| ***Craticula cuspidata*** | benthic |

| | |
|---|---|
| ***Craticula halophila*** | benthic |
| ***Craticula* spp.** | benthic |
| ***Craticula submolesta*** | benthic |
| ***Ctenophora pulchella*** | benthic |
| ***Cyclostephanos costatilimbus*** | planktonic |
| ***Cyclostephanos damasii*** | planktonic |
| ***Cyclostephanos dubius*** | planktonic |
| ***Cyclostephanos invisitatus*** | planktonic |
| ***Cyclostephanos* spp.** | planktonic |
| ***Cyclostephanos tholiformis*** | planktonic |
| ***Cyclotella atomus*** | planktonic |
| ***Cyclotella bodanica*** | planktonic |
| ***Cyclotella* cf. *meduanae*** | planktonic |
| ***Cyclotella* cf. *polymorpha*** | planktonic |
| ***Cyclotella comensis*** | planktonic |
| ***Cyclotella cryptica*** | planktonic |
| ***Cyclotella distinguenda*** | planktonic |
| ***Cyclotella gamma*** | planktonic |
| ***Cyclotella glabriuscula*** | planktonic |
| ***Cyclotella hakanssoniae*** | planktonic |
| ***Cyclotella meduanae*** | planktonic |
| ***Cyclotella meneghiniana*** | planktonic |
| ***Cyclotella michiganiana*** | planktonic |
| ***Cyclotella ocellata*** | planktonic |
| ***Cyclotella quillensis*** | planktonic |
| ***Cyclotella rossii*** | planktonic |
| ***Cyclotella schumannii*** | planktonic |
| ***Cyclotella* spp.** | planktonic |
| ***Cyclotella striata*** | planktonic |
| ***Cyclotella tripartita*** | planktonic |
| ***Cymbella affinis*** | benthic |
| ***Cymbella angustata*** | benthic |
| ***Cymbella cistula*** | benthic |
| ***Cymbella cymbiformis*** | benthic |
| ***Cymbella diluviana*** | benthic |
| ***Cymbella gracilis*** | benthic |
| ***Cymbella helvetica*** | benthic |
| ***Cymbella pusilla*** | benthic |
| ***Denticula elegans*** | benthic |
| ***Denticula kuetzingii*** | benthic |
| ***Diadesmis confervacea*** | benthic |

| | |
|---|---|
| ***Diatoma mesodon*** | benthic |
| ***Diatoma* spp.** | benthic |
| ***Diatoma tenuis*** | benthic |
| ***Diatoma vulgaris*** | benthic |
| ***Diploneis marginestriata*** | benthic |
| ***Diploneis oblongella*** | benthic |
| ***Diploneis oculata*** | benthic |
| ***Diploneis pseudovalis*** | benthic |
| ***Diploneis puella*** | benthic |
| ***Diploneis subovalis*** | benthic |
| ***Discostella asterocostata*** | unclassed |
| ***Discostella glomerata*** | planktonic |
| ***Discostella pseudostelligera*** | planktonic |
| ***Discostella* spp.** | planktonic |
| ***Discostella stelligera*** | planktonic |
| ***Discostella stelligeroides*** | planktonic |
| ***Discostella woltereckii*** | planktonic |
| ***Ellerbeckia arenaria*** | benthic |
| ***Encyonema gracile*** | benthic |
| ***Encyonema minutum*** | benthic |
| ***Encyonema perpusillum*** | benthic |
| ***Encyonema prostrata*** | benthic |
| ***Encyonema silesiacum*** | benthic |
| ***Encyonopsis cesatii*** | benthic |
| ***Encyonopsis falaisensis*** | benthic |
| ***Encyonopsis krammeri*** | benthic |
| ***Encyonopsis microcephala*** | benthic |
| ***Entomoneis alata*** | benthic |
| ***Epithemia adnata*** | benthic |
| ***Epithemia sorex*** | benthic |
| ***Epithemia* spp.** | benthic |
| ***Epithemia turgida*** | benthic |
| ***Eunotia bilunaris*** | benthic |
| ***Eunotia* cf. *carolina*** | benthic |
| ***Eunotia* cf. *pirla*** | benthic |
| ***Eunotia croatana*** | benthic |
| ***Eunotia exigua*** | benthic |
| ***Eunotia flexuosa*** | benthic |
| ***Eunotia implicata*** | benthic |
| ***Eunotia incisa*** | benthic |
| ***Eunotia intermedia*** | benthic |

| | |
|---|---|
| ***Eunotia naegelii*** | benthic |
| ***Eunotia paludosa*** | benthic |
| ***Eunotia pectinalis*** | benthic |
| ***Eunotia rhomboidea*** | benthic |
| ***Eunotia* spp.** | benthic |
| ***Eunotia zasuminensis*** | benthic |
| ***Fallacia omissa*** | benthic |
| ***Fallacia pygmaea*** | benthic |
| ***Fallacia* spp.** | benthic |
| ***Fallacia tenera*** | benthic |
| ***Fistulifera pelliculosa*** | benthic |
| ***Fragilaria acidobiontica*** | benthic |
| ***Fragilaria bidens*** | benthic |
| ***Fragilaria brevistriata*** | benthic |
| ***Fragilaria capucina*** | benthic |
| ***Fragilaria crotonensis*** | benthic |
| ***Fragilaria cyclopum*** | benthic |
| ***Fragilaria distans*** | benthic |
| ***Fragilaria longifusiformis*** | benthic |
| ***Fragilaria microstriata*** | benthic |
| ***Fragilaria nitzschioides*** | benthic |
| ***Fragilaria pinnata*** | benthic |
| ***Fragilaria rhabdosoma*** | benthic |
| ***Fragilaria sepes*** | benthic |
| ***Fragilaria* spp.** | benthic |
| ***Fragilaria tenera*** | benthic |
| ***Fragilaria vaucheriae*** | benthic |
| ***Fragilaria zeilleri*** | benthic |
| ***Fragilariaceae* spp.** | benthic |
| ***Fragilariforma virescens*** | benthic |
| ***Frustulia amphipleuroides*** | benthic |
| ***Frustulia crassinervia*** | benthic |
| ***Frustulia krammeri*** | benthic |
| ***Frustulia.rhomboides*** | benthic |
| ***Frustulia saxonica*** | benthic |
| ***Frustulia* spp.** | benthic |
| ***Geissleria acceptata*** | benthic |
| ***Geissleria ignota*** | benthic |
| ***Gomphoneis transsilvanica*** | benthic |
| ***Gomphonema affine*** | benthic |
| ***Gomphonema angustatum*** | benthic |

| | |
|---|---|
| ***Gomphonema augur*** | benthic |
| ***Gomphonema clavatum*** | benthic |
| ***Gomphonema gracile*** | benthic |
| ***Gomphonema insigne*** | benthic |
| ***Gomphonema intricatum*** | benthic |
| ***Gomphonema minutum*** | benthic |
| ***Gomphonema olivaceum*** | benthic |
| ***Gomphonema parvulum*** | benthic |
| ***Gomphonema pseudoaugur*** | benthic |
| ***Gomphonema pseudotenellum*** | benthic |
| ***Gomphonema pumilum*** | benthic |
| ***Gomphonema.rhombicum*** | benthic |
| ***Gomphonema* spp.** | benthic |
| ***Gyrosigma acuminatum*** | benthic |
| ***Gyrosigma attenuatum*** | benthic |
| ***Gyrosigma macrum*** | benthic |
| ***Gyrosigma obtusatum*** | benthic |
| ***Gyrosigma spencerii*** | benthic |
| ***Gyrosigma* spp.** | benthic |
| ***Hannaea arcus*** | planktonic |
| ***Hantzschia amphioxys*** | benthic |
| ***Hippodonta capitata*** | benthic |
| ***Hippodonta hungarica*** | benthic |
| ***Hippodonta* spp.** | benthic |
| ***Karayevia clevei*** | benthic |
| ***Karayevia laterostrata*** | benthic |
| ***Karayevia suchlandtii*** | benthic |
| ***Kobayasiella subtilissima*** | benthic |
| ***Kobayasiella venezuelensis*** | benthic |
| ***Luticola cohnii*** | benthic |
| ***Luticola mutica*** | benthic |
| ***Mastogloia elliptica*** | benthic |
| ***Mastogloia smithii*** | benthic |
| ***Mayamaea agrestis*** | benthic |
| ***Mayamaea atomus*** | benthic |
| ***Mayamaea recondita*** | benthic |
| ***Mayamaea* spp.** | benthic |
| ***Melosira arentii*** | planktonic |
| ***Melosira varians*** | tychoplanktonic |
| ***Meridion circulare*** | benthic |
| ***Navicula angusta*** | benthic |

| | |
|---|---|
| ***Navicula angustata*** | benthic |
| ***Navicula aquaedurae*** | benthic |
| ***Navicula arvensis*** | benthic |
| ***Navicula canalis*** | benthic |
| ***Navicula capitatoradiata*** | benthic |
| ***Navicula* cf. *graciloides*** | benthic |
| ***Navicula* cf. *menisculus*** | benthic |
| ***Navicula* cf. *minima*** | benthic |
| ***Navicula* cf. *recens*** | benthic |
| ***Navicula* cf. *veneta*** | benthic |
| ***Navicula cincta*** | benthic |
| ***Navicula cryptocephala*** | benthic |
| ***Navicula cryptotenella*** | benthic |
| ***Navicula erifuga*** | benthic |
| ***Navicula germainii*** | benthic |
| ***Navicula gregaria*** | benthic |
| ***Navicula heimansioides*** | benthic |
| ***Navicula jaagii*** | benthic |
| ***Navicula kotschyi*** | benthic |
| ***Navicula lanceolata*** | benthic |
| ***Navicula laterostrata*** | benthic |
| ***Navicula leptostriata*** | benthic |
| ***Navicula libonensis*** | benthic |
| ***Navicula medioconvexa*** | benthic |
| ***Navicula menisculus*** | benthic |
| ***Navicula minima*** | benthic |
| ***Navicula notha*** | benthic |
| ***Navicula oblonga*** | benthic |
| ***Navicula obsoleta*** | benthic |
| ***Navicula peregrina*** | benthic |
| ***Navicula praeterita*** | benthic |
| ***Navicula pseudoventralis*** | benthic |
| ***Navicula radiosa*** | benthic |
| ***Navicula reichardtiana*** | benthic |
| ***Navicula rhynchocephala*** | benthic |
| ***Navicula* spp.** | benthic |
| ***Navicula submuralis*** | benthic |
| ***Navicula symmetrica*** | benthic |
| ***Navicula tripunctata*** | benthic |
| ***Navicula trivialis*** | benthic |
| ***Navicula utermoehlii*** | benthic |

| | |
|---|---|
| ***Navicula veneta*** | benthic |
| ***Navicula viridula*** | benthic |
| ***Navicula viridulacalcis*** | benthic |
| ***Navicula vitabunda*** | benthic |
| ***Navicula vulpina*** | benthic |
| ***Navicula wildii*** | benthic |
| ***Neidium ampliatum*** | benthic |
| ***Nitzschia acicularis*** | benthic |
| ***Nitzschia agnita*** | benthic |
| ***Nitzschia amphibia*** | benthic |
| ***Nitzschia amphibioides*** | benthic |
| ***Nitzschia. angustata*** | benthic |
| ***Nitzschia angustatula*** | benthic |
| ***Nitzschia archibaldii*** | benthic |
| ***Nitzschia bulnheimiana*** | benthic |
| ***Nitzschia capitellata*** | benthic |
| ***Nitzschia compressa*** | benthic |
| ***Nitzschia constricta*** | benthic |
| ***Nitzschia dissipata*** | benthic |
| ***Nitzschia.diversa*** | benthic |
| ***Nitzschia filiformis*** | benthic |
| ***Nitzschia fonticola*** | benthic |
| ***Nitzschia frustulum*** | benthic |
| ***Nitzschia gessneri*** | benthic |
| ***Nitzschia gracilis*** | benthic |
| ***Nitzschia gracilliformis*** | benthic |
| ***Nitzschia homburgienis*** | benthic |
| ***Nitzschia incognita*** | benthic |
| ***Nitzschia inconspicua*** | benthic |
| ***Nitzschia intermedia*** | benthic |
| ***Nitzschia lacunarum*** | benthic |
| ***Nitzschia lacuum*** | benthic |
| ***Nitzschia liebethruthii*** | benthic |
| ***Nitzschia linearis*** | benthic |
| ***Nitzschia macilenta*** | benthic |
| ***Nitzschia obtusa*** | benthic |
| ***Nitzschia palea*** | benthic |
| ***Nitzschia paleacea*** | benthic |
| ***Nitzschia perminuta*** | benthic |
| ***Nitzschia pumila*** | benthic |
| ***Nitzschia pura*** | benthic |

| | |
|---|---|
| ***Nitzschia radicula*** | benthic |
| ***Nitzschia recta*** | benthic |
| ***Nitzschia sigma*** | benthic |
| ***Nitzschia silicula*** | benthic |
| ***Nitzschia sinuata*** | benthic |
| ***Nitzschia sociabilis*** | benthic |
| ***Nitzschia solita*** | benthic |
| ***Nitzschia* spp.** | benthic |
| ***Nitzschia subacicularis*** | benthic |
| ***Nitzschia subtilis*** | benthic |
| ***Nitzschia suchlandtii*** | benthic |
| ***Nitzschia supralitorea*** | benthic |
| ***Nitzschia tropica*** | benthic |
| ***Nitzschia valdecostata*** | benthic |
| ***Nupela lapidosa*** | benthic |
| ***Nupela neotropica*** | benthic |
| ***Nupela* spp.** | benthic |
| ***Opephora martyi*** | benthic |
| ***Opephora olsenii*** | benthic |
| ***Orthoseira roeseana*** | planktonic |
| ***Pinnularia borealis*** | benthic |
| ***Pinnularia braunii*** | benthic |
| ***Pinnularia gibba*** | benthic |
| ***Pinnularia interrupta*** | benthic |
| ***Pinnularia microstauron*** | benthic |
| ***Pinnularia* spp.** | benthic |
| ***Pinnularia viridis*** | benthic |
| ***Placoneis elginensis*** | benthic |
| ***Plagiotropis lepidoptera*** | benthic |
| ***Planothidium delicatulum*** | benthic |
| ***Planothidium frequentissimum*** | benthic |
| ***Planothidium joursacense*** | benthic |
| ***Planothidium lanceolatum*** | benthic |
| ***Planothidium rostratum*** | benthic |
| ***Platessa conspicua*** | benthic |
| ***Pleurosigma elongatum*** | benthic |
| ***Pleurosigma salinarum*** | benthic |
| ***Psammodictyon constrictum*** | benthic |
| ***Psammothidium bioretii*** | benthic |
| ***Psammothidium grischunum*** | benthic |
| ***Psammothidium helveticum*** | benthic |

| | |
|---|---|
| ***Psammothidium marginulatum*** | benthic |
| ***Psammothidium rossii*** | benthic |
| ***Psammothidium sacculum*** | benthic |
| ***Psammothidium scoticum*** | benthic |
| ***Psammothidium subatomoides*** | benthic |
| ***Psammothidium ventralis*** | benthic |
| ***Pseudostaurosira brevistriata*** | tychoplanktonic |
| ***Pseudostaurosira neoelliptica*** | benthic |
| ***Pseudostaurosira parasitica*** | benthic |
| ***Pseudostaurosira polonica*** | benthic |
| ***Pseudostaurosira pseudoconstruens*** | benthic |
| ***Pseudostaurosira subsalina*** | benthic |
| ***Pseudostaurosira trainorii*** | benthic |
| ***Pseudostaurosiropsis connecticutensis*** | planktonic |
| ***Pseudostaurosiropsis geocollegarum*** | planktonic |
| ***Pseudostaurosiropsis* spp.** | planktonic |
| ***Puncticulata bodanica*** | planktonic |
| ***Puncticulata comta*** | planktonic |
| ***Puncticulata radiosa*** | planktonic |
| ***Reimeria sinuata*** | benthic |
| ***Rhoicosphenia abbreviata*** | benthic |
| ***Rhopalodia gibba*** | benthic |
| ***Rossithidium petersennii*** | benthic |
| ***Rossithidium pusillum*** | benthic |
| ***Sellaphora hustedtii*** | benthic |
| ***Sellaphora laevissima*** | benthic |
| ***Sellaphora mutata*** | benthic |
| ***Sellaphora pupula*** | benthic |
| ***Sellaphora seminulum*** | benthic |
| ***Skeletonema potamos*** | planktonic |
| ***Stauroforma exiguiformis*** | benthic |
| ***Stauroneis anceps*** | benthic |
| ***Staurosira construens*** | tychoplanktonic |
| ***Staurosira elliptica*** | tychoplanktonic |
| ***Staurosirella berolinensis*** | tychoplanktonic |
| ***Staurosirella lapponica*** | tychoplanktonic |
| ***Staurosirella leptostauron*** | tychoplanktonic |
| ***Staurosirella oldenburgiana*** | tychoplanktonic |
| ***Staurosirella pinnata*** | tychoplanktonic |

| | |
|---|---|
| ***Stenopterobia delicatissma*** | benthic |
| ***Stephanodiscus agassizensis*** | planktonic |
| ***Stephanodiscus alpinus*** | planktonic |
| ***Stephanodiscus binderanus*** | planktonic |
| ***Stephanodiscus* cf. *minutus*** | planktonic |
| ***Stephanodiscus hantzschii*** | planktonic |
| ***Stephanodiscus medius*** | planktonic |
| ***Stephanodiscus minutulus*** | planktonic |
| ***Stephanodiscus neoastraea*** | planktonic |
| ***Stephanodiscus niagarae*** | planktonic |
| ***Stephanodiscus parvus*** | planktonic |
| ***Stephanodiscus rotula*** | planktonic |
| ***Stephanodiscus* spp.** | planktonic |
| ***Stephanodiscus vestibulis*** | planktonic |
| ***Surirella amphioxys*** | benthic |
| ***Surirella angusta*** | benthic |
| ***Surirella biseriata*** | benthic |
| ***Surirella brebissonii*** | benthic |
| ***Surirella brightwellii*** | benthic |
| ***Surirella minuta*** | benthic |
| ***Surirella ovalis*** | benthic |
| ***Surirella* spp.** | benthic |
| ***Synedra acus*** | planktonic |
| ***Synedra biceps*** | planktonic |
| ***Synedra capitata*** | planktonic |
| ***Synedra cyclopum*** | planktonic |
| ***Synedra delicatissima*** | planktonic |
| ***Synedra demerarae*** | planktonic |
| ***Synedra mazamaensis*** | planktonic |
| ***Synedra radians*** | planktonic |
| ***Synedra rumpens*** | planktonic |
| ***Synedra* spp.** | planktonic |
| ***Synedra subrhombica*** | planktonic |
| ***Synedra ulna*** | planktonic |
| ***Tabellaria fenestrata*** | planktonic |
| ***Tabellaria flocculosa*** | planktonic |
| ***Tabellaria quadriseptata*** | planktonic |
| ***Tabellaria* spp.** | planktonic |
| ***Tabellaria ventricosa*** | planktonic |
| ***Tabularia fasciculata*** | planktonic |
| ***Tabularia tabulata*** | planktonic |

| | |
|---|---|
| ***Thalassiosira baltica*** | planktonic |
| ***Thalassiosira pseudonana*** | planktonic |
| ***Thalassiosira* spp.** | planktonic |
| ***Thalassiosira visurgis*** | planktonic |
| ***Thalassiosira weissflogii*** | planktonic |
| ***Tryblionella calida*** | benthic |
| ***Tryblionella gracilis*** | benthic |
| ***Tryblionella hungarica*** | benthic |
| ***Tryblionella* spp.** | benthic |

**Ch 1 S3: RV coefficients from comparisons of diatom assemblage matrices and lake positions within RDAs, for genus-level data.** "WC" refers to water-column samples and "SSed" to surface sediment samples. The "Ordination overlay" column lists the two matrices, of which their structure is compared *symmetrically* using an RV coefficient. The column for the RV coefficient of the full set of fitted scores refers to the comparison of scores from all of the axes within an ordination versus the RV coefficient of the 1st axis scores, which is the correlation between the main axis of variation in one matrix and the main axis of variation in another (synonymous with Ordinary Least Squares Regression).

| Ordination overlay | RV coefficient of 1st axis of fitted scores (*P*- value) | RV coefficient of (full set) of fitted scores (P-value) |
|---|---|---|
| **Matrix A:** Site scores from WC environmental RDA **Matrix B:** Site scores from SSed environmental RDA **(all genera)** | 0.50 (*P*<0.001) | 0.20 (*P*<0.001) |
| **Matrix A:** Site scores from WC spatial RDA **Matrix B:** Site scores from SSed spatial RDA **(all genera)** | 0.51 (*P*<0.001) | 0.17 (*P*<0.001) |

**Ch. 1 S4: Transformation parameters for environmental variables transformed using Box-Cox transformation, rounded to two decimal places.** Normality of the transformed variables was tested with a Shapiro-Wilks test which uses an approximate p-value cut off of 0.1 to test for normality ($p<0.1$ not normal). The formula for the transformation is: $\mathbf{x'_\lambda = (x^\lambda - 1) / \lambda}$. *denotes variables where $p<0.1$ but near 0.05 and histogram of the variable distribution appeared close to normal, hence the Box-Cox transformation of the variable was retained.

| Variable | Time step(s) | λ | ß | $\delta^2$ | *P*-value (Shapiro-Wilks) |
|---|---|---|---|---|---|
| **Total Nitrogen** | Visit 1 (n=468) | -0.2 | 3.55 | 0.07 | 0.296 |
| **Secchi depth** | Visit 1 (n=468) | 0.2 | 0.41 | 1.1 | 0.07825* |
| **Cl** | Visit 1 (n=468) | 0.05 | 6.21 | 5.21 | 0.04137* |
| **Chl *a*** | Visit 1 (n=468) | -0.09 | 1.88 | 1.4 | 0.7996 |
| **$SiO_2$** | Visit 1 (n=468) | 0.18 | 1.93 | 2.62 | 0.8821 |
| **Total Phosphorus** | Visit 1 (n=51) | -0.1 | 2.63 | 1.11 | 0.688 |
| **Conductivity** | Visit 1 (n=51) | -0.16 | 3.52 | 0.2 | 0.8335 |
| **Turbidity** | Visit 1 (n=51) | 0.008 | 1.22 | 1.59 | 0.7213 |
| **Total phosphorous** | Visit 2 (n=51) | -0.06 | 2.91 | 1.7 | 0.6893 |
| **Conductivity** | Visit 2 (n=51) | -0.17 | 3.47 | 0.19 | 0.7164 |
| **Turbidity** | Visit 2 (n=51) | -0.09 | 1.3 | 1.24 | 0.3609 |
| **Total phosphorous** | Mean of 1 and 2 (n=51) | -0.1 | 2.71 | 1.07 | 0.9437 |
| **Conductivity** | Mean of 1 and 2 (n=51) | -0.17 | 3.49 | 0.19 | 0.7829 |
| **Turbidity** | Mean of 1 and 2 (n=51) | -0.03 | 1.34 | 1.38 | 0.317 |

## Appendix B: Supporting information for Chapter 2

### Ch2 S1: Description of sediment core screening with respect to length and predicted age of bottom sediment core samples.

Lakes sampled in the 2007 NLA were assigned a weight based on their surface area to reflect the portion of all U.S. lakes represented in the survey (see U.S. EPA, 2011-2012 for details). Thus, the 1000+ lakes sampled in 2007 were estimated to represent ~50000 lakes across the U.S. (U.S. EPA, 2009). However, not all of these lakes could be included in this paleolimnological study, firstly because not all lakes were cored and secondly because of differences in sediment age. The need to accurately identify the age of bottom core sediment samples from the 2007 NLA is exemplified in the debate around Bachman *et al*., (2013), which also used surface and bottom core sediment data from the 2007 NLA, but for the purpose of quantifying the extent of eutrophication across the U.S.A. The main criticisms of their work (Smith, 2014; and McDonald *et al*., 2014) revolved around the fact that Bachman *et al*., (2013) had relied on descriptions from the various field teams collecting the sediment cores, as well as (mostly) qualitative criteria from the EPA to classify whether a lake where was one where the bottom of the core was sufficiently deep to have reached sediment from pre-European settlement conditions (hereafter referred to as a "high confidence" (HC) cores, e.g. see U.S. EPA (2010), p.32-33). Indeed, Bachman *et al*., (2013) and later Bachman *et al*., (2014) identified 233 lakes with cores deemed HC by the U.S. EPA based on a number of factors (U.S. EPA, 2010) and used data from bottom samples of these cores in subsequent analyses. However, there is still considerable variation in core length and comments by both Smith (2014) and McDonald *et al*., (2014) suggest that the criteria used to identify these cores preferentially selected samples from relatively short cores (likely not pre-European settlement, or even pre-industrial conditions).

While Bachman *et al*., (2013) and (2014) used the information available on the sediment cores to the greatest extent possible, we utilized a more extensive, three-fold approach to increase the accuracy in identifying cores where the bottom samples likely represent pre-industrial conditions. We also used a pre-1850 CE cut-off to refer to pre-industrial conditions and are not attempting to determine whether the bottoms of cores date back to pre-European settlement conditions. First, by using the length of the cores collected, we estimated the approximate age of the core bottoms using regression equations of latitude and sedimentation rate developed by Brothers *et al*., (2008). We determined that the majority of cores longer than 30 cm in length had bottom sediments estimated to date back to at least 1850 CE. Secondly, we cross-referenced the list of cores greater than 30 cm in length with the list of designated HC cores. This produced a list of sites with core bottom samples likely older than 1850 CE. Finally, after procuring leftover sediment from a set of bottom core samples (archived at The Academy of Natural Sciences of Drexel University), we randomly selected 35 bottom core samples and further selected an additional 15 bottom core samples from the shortest cores where material was available to undergo radiometric dating. Radiometric dating produced ratios of $^{214}$Bi and $^{210}$Pb, where activity of $^{210}$Pb within two standard errors of $^{214}$Bi indicates sediment older than 1850 CE (Dixit *et al*., 1999; Vermaire *et al*., 2012). These radiometric estimates are not as accurate as measuring the decay of unsupported $^{210}$Pb throughout a full core, but was the best approximation available since intervals between the top and bottom samples were not kept from the collected cores.

We used chi-square tests to test two hypotheses; first, that age assignment by radiometric dating is independent of core length and second, that age assignment by radiometric dating is independent of the age estimate using the Brothers *et al*., (2008) equation which accounts for the

variation in sedimentation rates in lakes across latitudes. The chi-square test of age assignment and core length (n = 35), transformed into binary variables, resulted in a *P*-value of 0.9, meaning that we cannot reject the null hypothesis that core length is independent of whether a sample is determined to be pre-1850 CE. While we could not use this radiometric dating method on all the candidate bottom samples (due to sample availability and costs), the Chi-square test of age assignment by radiometric dating (for the 35 samples that could be radiometrically dated) and age assignment by regression equations (n = 35) resulted in a *P*- value of 0.6, meaning that we cannot reject the null hypothesis that these two age assignments are independent. As such, both core length and age based on regression estimates appear to be reasonable indicators of core age, especially as the regression estimates factor in the effect of latitude on sedimentation rate. However, age based on regression estimates (i.e. the Brothers *et al*., 2008) provide more robust predictions of ages when compared to radiometric dating.

**Literature Cited**

Bachman R.W., Hoyer M.V., & Canfield, Jr., D.E. (2013). The extent that natural lakes in the United States of America have been changed by cultural eutrophication. *Limnology and Oceanography*, **58**, 945-950.

Bachman R.W., Hoyer M.V., & Canfield, Jr., D.E. (2014). Response to comments: Quantification of the extent of cultural eutrophication of natural lakes in the United States. *Limnology and Oceanography*, **59**, 2231-2239.

Brothers S., Vermaire J.C., & Gregory-Eaves I. (2008). Empirical models for describing recent sedimentation rates in lakes distributed across broad spatial scales. *Journal of Paleolimnology*, **40**, 1003-1019.

Dixit S.S., Smol J.P., Charles D.F., Hughes R.M., Paulsen S.G., & Collins G.B. (1999). Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1860-1869.

McDonald C.P., Lottig N.R., Stoddard J.L., Herlihy A.T., Lehmann S., Paulsen S.G., Peck D.V., Pollard A.I., & Stevenson R.J. (2014). Comment on Bachman et al. (2013): A nonrepresentative sample cannot describe the extent of cultural eutrophication of natural lakes in the United States. *Limnology and Oceanography*, **59**, 2226-2230.

Smith V.H. (2014). Comment: Cultural eutrophication of natural lakes in the United States is real and widespread. *Limnology and Oceanography*, **59**, 2217-2225.

U.S. EPA. (2011-2012). National Lakes Assessment. Field Operations Manual. EPA 841-B-11-003. U.S. Environmental Protection Agency, Washington, DC, USA.

Vermaire J.C., Prairie Y.T., & Gregory-Eaves I. (2012). Diatom-inferred decline of macrophyte abundance in lakes of southern Quebec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 511-524.

**Ch 2 S2: Description of the Temporal Beta diversity Indices (TBI) performed using *TBI().***

The *TBI()* (Legendre, 2015) function outputs four components:

A: Species (or genera that both time points have in common

B: Abundance loss component, i.e. loss of abundances on a species-by-species (or genera-by-genera) basis from the first time point compared to the second time point.

C: Abundance gain component, i.e. gain of abundances on a species-by-species (or genera-by-genera) basis in the second time point compared to the first time point.

D: Total temporal beta diversity (B + C)

Each component is divided by a denominator, which is (2A+B+C) for computations based on the percentage difference index. Each component is then in the [0,1] interval and D = B + C.

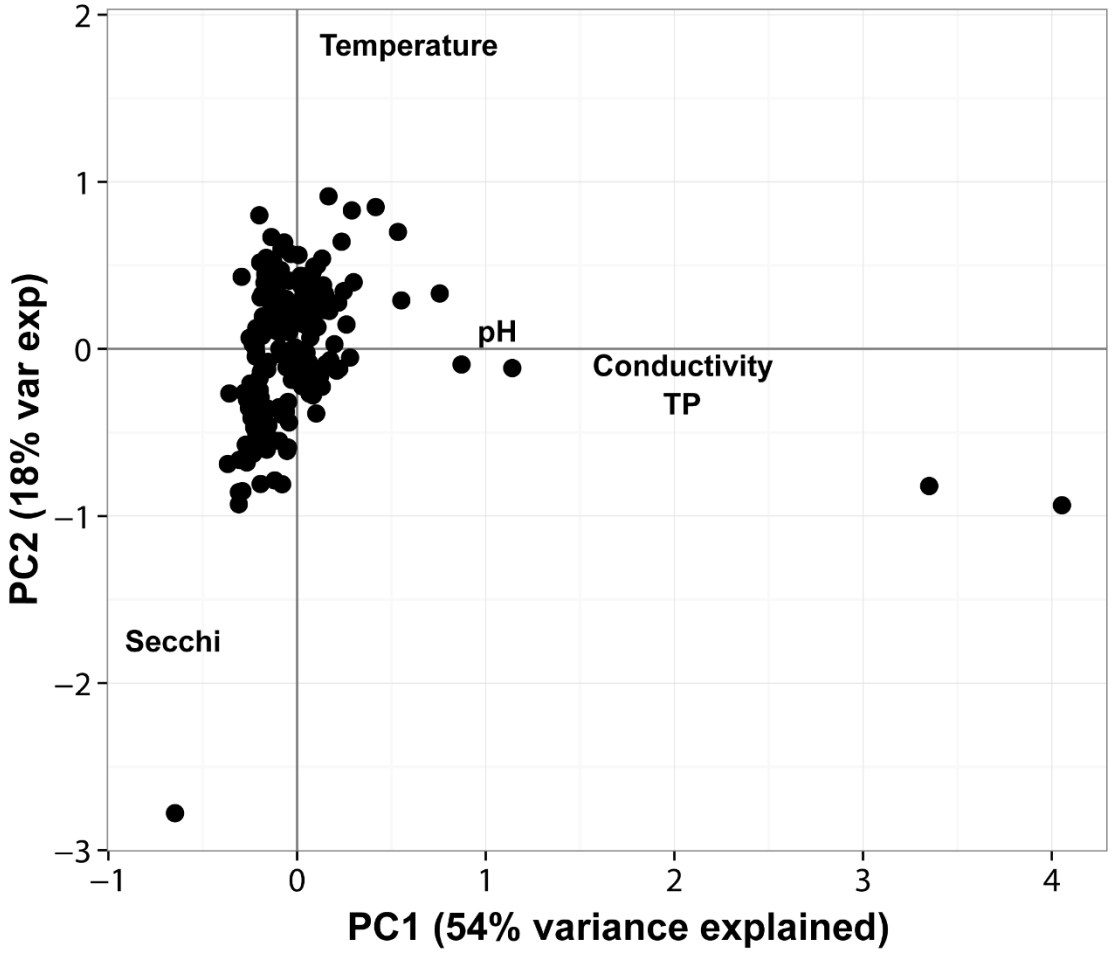
We computed temporal beta diversity between the historical diatom assemblage and the 2007 diatom assemblage for each lake, recording total beta diversity (D) as well as the species/genus loss (B) and gain components (C). We summarized these values for each ecoregion and examined the relationship between each of these components and latitude and longitude, using OLS regression. To use the function *decompose.D()*, we needed to combine the 2007 and historical diatom data sets consisting of 927 and 1002 unique species, variants and morphotypes respectively, resulting in a total of 1212 unique operational taxonomic units for use in the temporal beta diversity analysis (and same procedure when using genus-level data). This "stacking" of datasets results in a number of zeroes being included in the resultant data frame; the changes in gamma diversity between time periods were small (see article Table 2). As such, we assessed the effect of this inclusion of zeros by examining subsets of the data with more or less zeros, but did not detect patterns in the magnitude of beta diversity.

**Literature Cited**

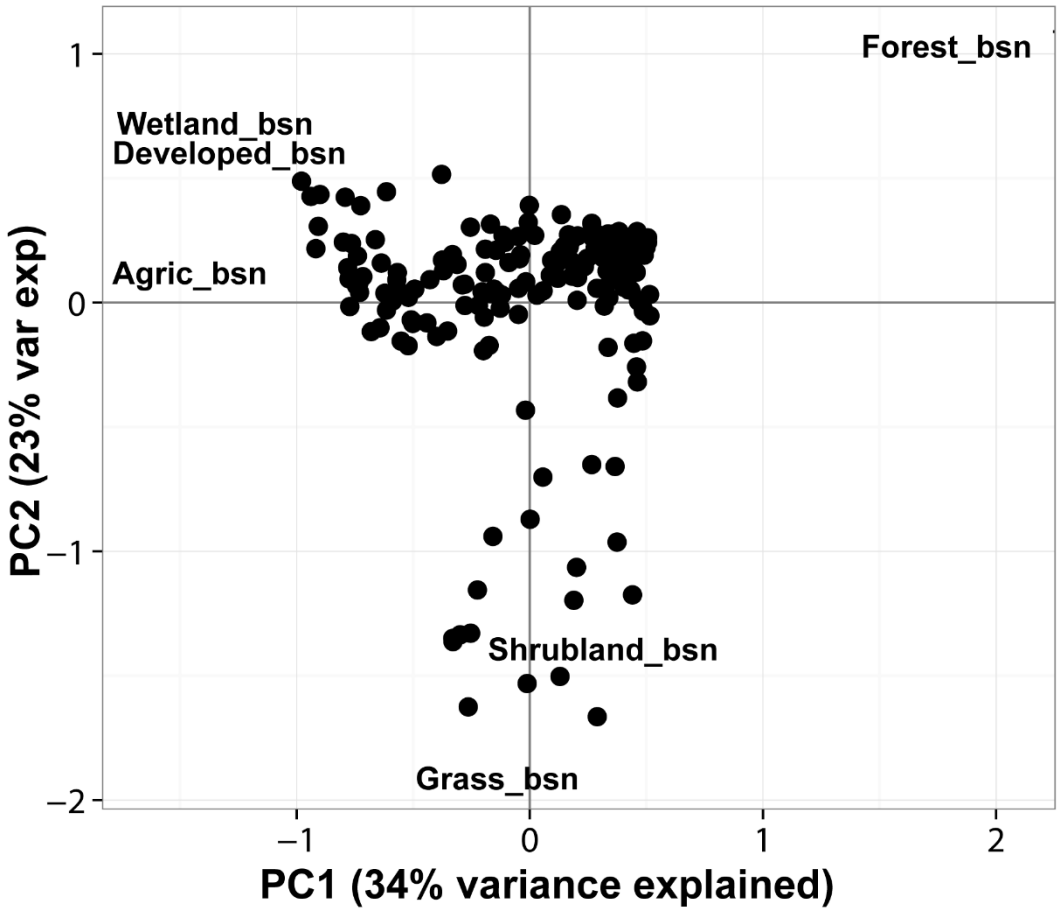
Legendre P. (2015). R function: TBI(). Available online: http://adn.biol.umontreal.ca/~numericalecology/FonctionsR/.

**Ch 2 S3: Principal Component Analyses (PCAs) of (a) standardized 2007 water quality variables and (b) land cover variables.**

(a) Water quality variables – Chlorophyll *a*, TN and TP were correlated and shown here by TP. Conductivity and TP were log transformed.



(b) Land cover variables (NLCD 1992) – "Forest_bsn" refers to percent forest including all forest types; "Wetland_bsn" refers to percent wetland; "Developed_bsn" refers to residentially developed area in a basin, including low, medium and high; "Agric_bsn" refers to percent agriculture including all agriculture types (crops, pasture etc.); "Shrubland_bsn" refers to percent shrubland.



**Ch 2 S4: Tables and figures for species-level analyses (n = 59).**

**S4, Table 1: Spatial beta (ß) diversity, mean rarefied species richness, mean alpha (α) diversity and gamma (γ) diversity for each ecoregion (Shannon diversity of species' sums).** Beta diversity was calculated using total variance computed using *beta.div()* based on percentage difference matrices. "Hx" refers to the historical sediments. Genus richness was rarefied after rare genera (<2% relative abundance) had been excluded. The Xeric, Southern Plains, Temperate Plains, Western Mountains and Upper Midwest ecoregions were excluded in these mean values because there were no sites in those regions for this reduced sample size.

| | ß-diversity | | Rarefied S | | α-diversity (Shannon) | | α-diversity (Simpson) | | γ-diversity | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Ecoregion** | Hx | 2007 | Hx | 2007 | Hx | 2007 | Hx | 2007 | Hx | 2007 |
| **All** | 0.40 | 0.37 | 56.4 | 52.8 | 2.9 | 2.9 | 0.9 | 0.9 | 4.5 | 4.4 |
| **Coastal Plains** | 0.45 | 0.40 | 32.7 | 37.2 | 2.0 | 2.5 | 0.7 | 0.8 | 3.4 | 3.7 |
| **Northern Appalachians** | 0.39 | 0.36 | 59.8 | 54.9 | 3.0 | 2.9 | 0.9 | 0.9 | 4.5 | 4.4 |

**S4, Table 2: Explanatory components for historical and 2007 spatial beta diversity,** as computed using *beta.div.comp()*. 'Repl' refers to the replacement component; 'AbDiff' refers to the abundance difference component; and 'Repl/Total' and 'AbDiff/Total' are these two components with total beta diversity as the denominator.

| Ecoregion | Repl | AbDiff | Repl/Total | AbDiff/Total |
|---|---|---|---|---|
| ***Historical*** | | | | |
| **All** | 0.40 | 0 | 1 | 0 |
| **Coastal Plains** | 0.45 | 0 | 1 | 0 |
| **Northern Appalachians** | 0.39 | 0 | 1 | 0 |
| ***2007*** | | | | |
| **All** | 0.37 | 0 | 1 | 0 |
| **Coastal Plains** | 0.40 | 0 | 1 | 0 |
| **Northern Appalachians** | 0.36 | 0 | 1 | 0 |

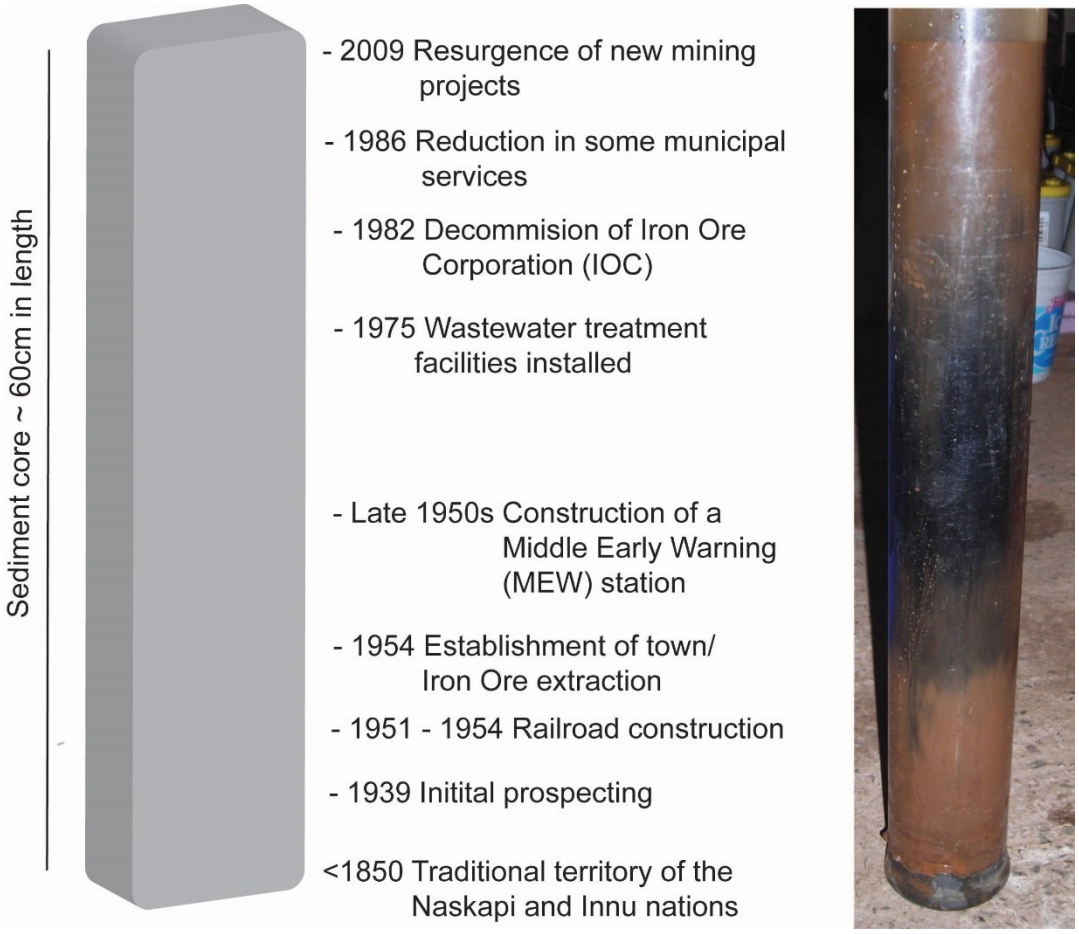
**S4, Table 3: Mean (and standard deviation) temporal beta diversity components for the ecoregions.** 'Total beta' refers to the mean value of the temporal beta diversity in each region region (mean value of the D column in the 'BCD' table provided by the function *TBI()*. It was computed using the percentage difference index applied to the diatom abundance data; values are in the [0,1] range. Total beta is the sum of 'Species loss' and 'Species gain'. Species loss refers to the component representing loss of abundances on a species by species basis between the historical and 2007 time points. Species gain refers to the component representing gain of abundances on a species by species basis between the historical and 2007 time points. These components were computed on a lake-by-lake basis and then averaged for each ecoregion.

| Ecoregion | Species loss | Species gain | Total beta |
|---|---|---|---|
| **Coastal Plains** | 0.3 (0.07) | 0.3 (0.07) | 0.6 (0.15) |
| **Northern Appalachians** | 0.2 (0.08) | 0.2 (0.08) | 0.5 (0.17) |


**S4, Figure 1: LCBD values for (a) historical spatial beta diversity and (b) 2007 spatial beta diversity and exceptional sites (c) for temporal beta diversity (TBI).** A lake has a significant LCBD value if $P < 0.05$, and is coded as "True" (open circles). LCBD values across all sites (in either historical or 2007) sum to 1. Note that none of the TBI values were significant after correction for multiple testing in (c)

**Appendix C: Supporting information for Chapter 3**

**Ch 3 S1: Timeline of major development activities in the Schefferville, Québec region.**



## Ch 3 S2: Technical appendix for radiometric dating

### Descriptions of radiometric dating models for the two Schefferville lakes:

Raw activity was measured using a pure germanium EGG-ORTEC® (Ametek, Inc.) gamma spectrometer (volumetric capacity ~10cm$^3$).

### Explanation of age models

Four age-depth models were considered for each sediment core: the $^{210}$Pb Limit model, the Constant Initial Concentration model (CIC), the Constant Flux Constant Sedimentation model (CFCS) and the Constant Rate of Supply model (CRS). Models not considered independently were the Constant Sedimentation model, Constant Flux model, or the Periodic Flux model (described in Sanchez-Cabeza & Ruiz-Fernández, 2012).

#### *$^{210}$Pb limit model*

The $^{210}$Pb limit model is based on the assumption that unsupported $^{210}$Pb (excess) is not found in the sediment column approximately 120 years before present (Ditchburn *et al*., 2011). As such, this limit in the sediment profile can be located downcore, and the depth at which the excess $^{210}$Pb signal is not detectable corresponds to approximately 120 years minus the date of coring, and then can provide an estimate of the sediment accumulation rate.

#### *Constant Initial Concentration model (CIC)*

The CIC model, sometimes called the Constant Activity (CA) model (Sanchez-Cabeza & Ruiz-Fernández, 2012), is a straight-forward model that makes the assumption that there is a constant sediment accumulation rate between years because the input of detritus to the lake does

not change (Ghaleb, 2007). Sedimentation rate is ultimately calculated using the decay constant of $^{210}$Pb ($\lambda$) and the slope of the linear relationship between depth within a core (x-axis) and the natural logarithm of unsupported $^{210}$Pb (Ghaleb, 2007). As this model assumes that the sedimentation rate does not change, any deviations from an exponential decrease of the profile of $^{210}$Pb are attributed to error (Ghaleb, 2007).

*Constant Flux Constant Sedimentation model (CFCS)*

The CFCS model combines assumptions and parameters from both the Constant Flux (CF) model and the Constant Sedimentation model (CS) (Sanchez-Cabeza & Ruiz-Fernández, 2012). In this way, the model assumes that mass accumulation rates are constant between layers (CS model) and also that excess $^{210}$Pb flux to the sediment surface is constant (CF model) (Sanchez-Cabeza & Ruiz-Fernández, 2012). This means that, like the CRS model (described after), excess $^{210}$Pb fluxes to the sediment may appear higher or be diluted by a change in sedimentation rate, but there may also be a variable source of $^{210}$Pb.

*Constant Rate of Supply model (CRS)*

Contrary to other models, the CRS model accommodates for changes in erosion and sedimentation rate (Appleby, 2001). The CRS model does not assume that deviation from exponential decrease of $^{210}$Pb is due solely to measurement or statistical errors (Ghaleb, 2007). In this model, unsupported $^{210}$Pb is constantly supplied, but its activity is diluted during time periods of high sedimentation (e.g. during an erosion event), and conversely, appears higher during time periods of low sedimentation (Ghaleb, 2007). The model integrates the total amount of unsupported $^{210}$Pb throughout the core with the amount in each individual layer.

*Summary*

| Model | Inputs | Key characteristics |
|---|---|---|
| **$^{210}$Pb limit** | Unsupported $^{210}$Pb through core (depths). | Assumes limit of unsupported $^{210}$Pb is at ~120 years ago. |
| **CIC** | Slope of the linear relationship between depth in core and natural logarithm of unsupported $^{210}$Pb. | Assumes that input of detrital material to lake does not vary. |
| **CFCS** | Utilizes the same slope relationship as CIC, but can have variable slopes at different points in the core. | $^{210}$Pb supply may be variable and may also appear changed due to sedimentation rate changes (see CRS). |
| **CRS** | Requires integration of total inventory $^{210}$Pb (unsupported) with inventories from the individual layers of the core. Generally requires continuous sampling down the core profile. | Supply of $^{210}$Pb does not change, but activity may be increased or diluted by sedimentation rate changes during different time periods. |

### (A) Dauriat (Latitude: 54.806117 , Longitude: -66.823076 )

*Activity data*

| Core interval (cm) | Mean depth (cm) | Depth error (cm) | $^{210}$Pb (dpm* g$^{-1}$) | $^{210}$Pb error | $^{226}$Ra (dpm g$^{-1}$) | $^{226}$Ra error | $^{137}$Cs (dpm g$^{-1}$) | $^{137}$Cs error | $^{210}$Pb excess | ln(excess Pb) |
|---|---|---|---|---|---|---|---|---|---|---|
| **0-1.5** | 0.75 | 0.75 | 37.34 | 2.41 | 4.20 | 1.03 | 2.56 | 0.39 | 23.14 | 3.14 |
| **2.25-3.75** | 3 | 0.75 | 22.35 | 1.25 | 2.83 | 0.32 | 1.87 | 0.14 | 19.51 | 2.97 |
| **7-8** | 7.5 | 0.5 | 10.59 | 1.03 | 2.14 | 0.43 | 1.72 | 0.17 | 8.45 | 2.13 |
| **10-11.25** | 10.625 | 0.625 | 9.96 | 0.69 | 2.00 | 0.26 | 1.45 | 0.11 | 7.96 | 2.07 |
| **13-14.5** | 13.75 | 0.75 | 7.31 | 1.18 | 2.15 | 0.53 | 1.19 | 0.21 | 5.16 | 1.64 |
| **15-16** | 15.5 | 0.5 | 6.48 | 1.43 | 2.48 | 0.68 | 1.29 | 0.27 | 3.99 | 1.38 |
| **21-22.25** | 21.625 | 0.625 | 6.97 | 0.71 | 1.97 | 0.32 | 2.03 | 0.15 | 5.00 | 1.61 |
| **23-24** | 23.5 | 0.5 | 6.31 | 0.58 | 2.81 | 0.26 | 4.12 | 0.17 | 3.50 | 1.25 |
| **26-27** | 26.5 | 0.5 | 7.42 | 0.67 | 4.03 | 0.32 | 3.29 | 0.16 | 3.39 | 1.22 |
| **28-28.75** | 28.375 | 0.375 | 4.70 | 0.47 | 2.79 | 0.22 | 1.29 | 0.09 | 1.90 | 0.64 |
| **31-31.75** | 31.375 | 0.375 | 2.77 | 0.29 | 2.81 | 0.18 | 0.16 | 0.04 | 0 | NA |

*dpm = disintegrations per minute

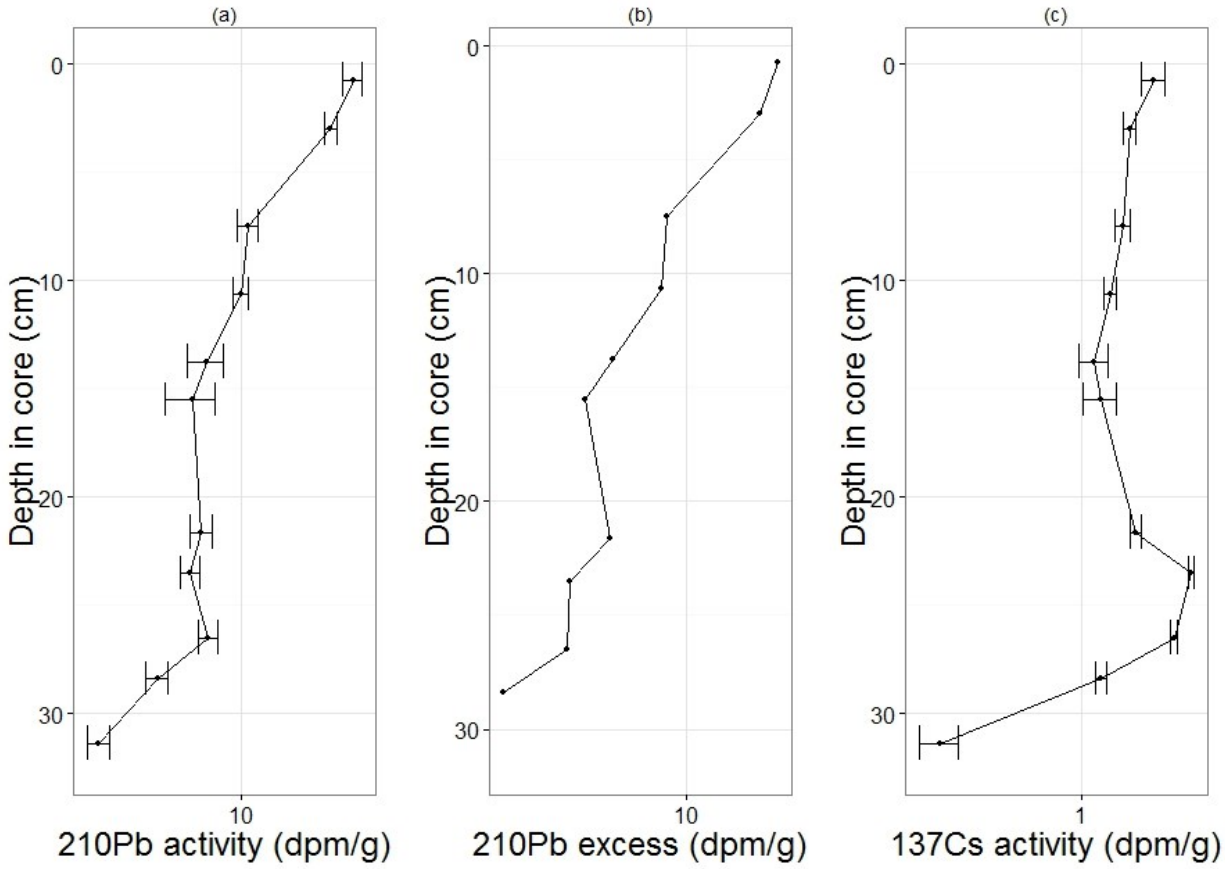



**Figure S2, 1: $^{210}$Pb (a), unsupported (excess) $^{210}$Pb (b) and $^{137}$Cs (c) activity on a log-scale as a function of depth (cm) in the Dauriat core.** Gaps in unsupported $^{210}$Pb come from points where estimate was less than zero.

*$^{210}$Pb limit model*

The limit of excess $^{210}$Pb was at approximately 31.375cm. Based on an estimated $^{210}$Pb limit at 120 years ago, the sedimentation rate down the core is estimated to be 0.26 cm yr$^{-1}$ (31.375/120).

| Interval | Mean depth (cm) | Estimated sedimentation rate (cm yr$^{-1}$) | Cumulative years based on $^{210}$Pb limit | Age based on $^{210}$Pb limit |
|---|---|---|---|---|
| **0-1.5** | 0.75 | 0.26 | 2.88 | 2009.12 |
| **2.25-3.75** | 3 | 0.26 | 11.54 | 2000.46 |
| **7-8** | 7.5 | 0.26 | 28.85 | 1983.15 |
| **10-11.25** | 10.625 | 0.26 | 40.87 | 1971.13 |
| **13-14.5** | 13.75 | 0.26 | 52.88 | 1959.12 |
| **15-16** | 15.5 | 0.26 | 59.62 | 1952.38 |
| **21-22.25** | 21.625 | 0.26 | 83.17 | 1928.83 |
| **23-24** | 23.5 | 0.26 | 90.38 | 1921.62 |
| **26-27** | 26.5 | 0.26 | 101.92 | 1910.08 |
| **28-28.75** | 28.375 | 0.26 | 109.13 | 1902.87 |

*CIC model*

Linear equation for the relationship between depth (x) and natural logarithm of unsupported $^{210}$Pb (y): -0.07626x + 2.95741.

Estimated sedimentation rate: ln(Pb half life) / slope = (0.03/0.076) = 0.41

| Interval | Mean depth (cm) | ln(unsupported Pb) | Estimated sedimentation rate (cm yr$^{-1}$) | Cumulative years based on CIC | Age based on CIC |
|---|---|---|---|---|---|
| **0-1.5** | 0.75 | 3.14 | 0.41 | 1.83 | 2010.17 |
| **2.25-3.75** | 3 | 2.97 | 0.41 | 7.34 | 2004.66 |
| **7-8** | 7.5 | 2.13 | 0.41 | 18.34 | 1993.66 |
| **10-11.25** | 10.625 | 2.07 | 0.41 | 25.98 | 1986.02 |
| **13-14.5** | 13.75 | 1.64 | 0.41 | 33.63 | 1978.37 |
| **15-16** | 15.5 | 1.38 | 0.41 | 37.91 | 1974.09 |
| **21-22.25** | 21.625 | 1.61 | 0.41 | 52.89 | 1959.11 |
| **23-24** | 23.5 | 1.25 | 0.41 | 57.47 | 1954.53 |
| **26-27** | 26.5 | 1.22 | 0.41 | 64.81 | 1947.19 |
| **28-28.75** | 28.375 | 0.64 | 0.41 | 69.39 | 1942.61 |

*CFCS model*

CFCS model can be considered equivalent to the CIC model, but multiple slopes can be used to calculate mean sedimentation rates for different time periods.

Slope zones: Slope 1 (0-15cm), Slope 2 (15-21cm), Slope 3 (>21cm)

| Slope zone | Linear equation between depth and ln(unsupported Pb) | Estimated sedimentation rate (ln Pb half life)/slope |
|---|---|---|
| **0-15cm** | -0.11731x + 3.2255 | 0.27 |
| **15-21cm** | 0.03755x + 0.79796 | 0.83 |
| **>21cm** | -0.1144x + 4.0259 | 0.27 |

| Interval | $^{210}$Pb excess | ln(unsupported Pb) | Slope zone | Estimated sedimentation rate (cm yr$^{-1}$) | Cumulative years based on CSCF | Age based on CSCF |
|---|---|---|---|---|---|---|
| **0-1.5** | 23.14 | 3.14 | 1 | 0.27 | 2.83 | 2009.17 |
| **2.25-3.75** | 19.51 | 2.97 | 1 | 0.27 | 11.32 | 2000.68 |
| **7-8** | 8.45 | 2.13 | 1 | 0.27 | 38.31 | 1983.69 |
| **10-11.25** | 7.96 | 2.07 | 1 | 0.27 | 40.10 | 1971.90 |
| **13-14.5** | 5.16 | 1.64 | 1 | 0.27 | 51.89 | 1960.11 |
| **15-16** | 3.99 | 1.38 | 2 | 0.83 | 58.50 | 1953.50 |
| **21-22.25** | 5.00 | 1.61 | 2 | 0.83 | 26.12 | 1985.88 |
| **23-24** | 3.50 | 1.25 | 3 | 0.27 | 28.39 | 1983.61 |
| **26-27** | 3.39 | 1.22 | 3 | 0.27 | 97.53 | 1914.47 |
| **28-28.75** | 1.90 | 0.64 | 3 | 0.27 | 104.43 | 1907.57 |

*CRS model*

Because samples extracted downcore for activity measurements were not continuous, this model was not considered further as the total inventory of $^{210}$Pb would need to be interpolated in order to use this method.

*Model selection and comparisons*

The $^{210}$Pb limit and CFCS models showed absolute congruence until~1950s. At that point, an apparently erroneous age estimate under the CFCS model showed a sharp departure from the other age models. This age estimate was likely due to one of the three slopes used to

estimate sedimentation rates for the CFCS model. The slope used for the observations between 15 and 22cm contains a point where the excess $^{210}$Pb increases as opposed to continuing to decrease exponentially. This apparent increase in excess $^{210}$Pb at these intervals is most likely due to increased erosional processes during that time period, which can be further verified by observed peaks in heavy metals (e.g. such as Al). This means that the deposition of $^{210}$Pb did not change, rather $^{210}$Pb activity was artificially lowered due to increased land erosion, resulting in an increased excess $^{210}$Pb value. The three slopes used to calculate the CFCS model, therefore, were not appropriate for estimating ages in this core. The CIC model was selected as the best age model for Dauriat because of the timing of the peak in $^{137}$Cs.

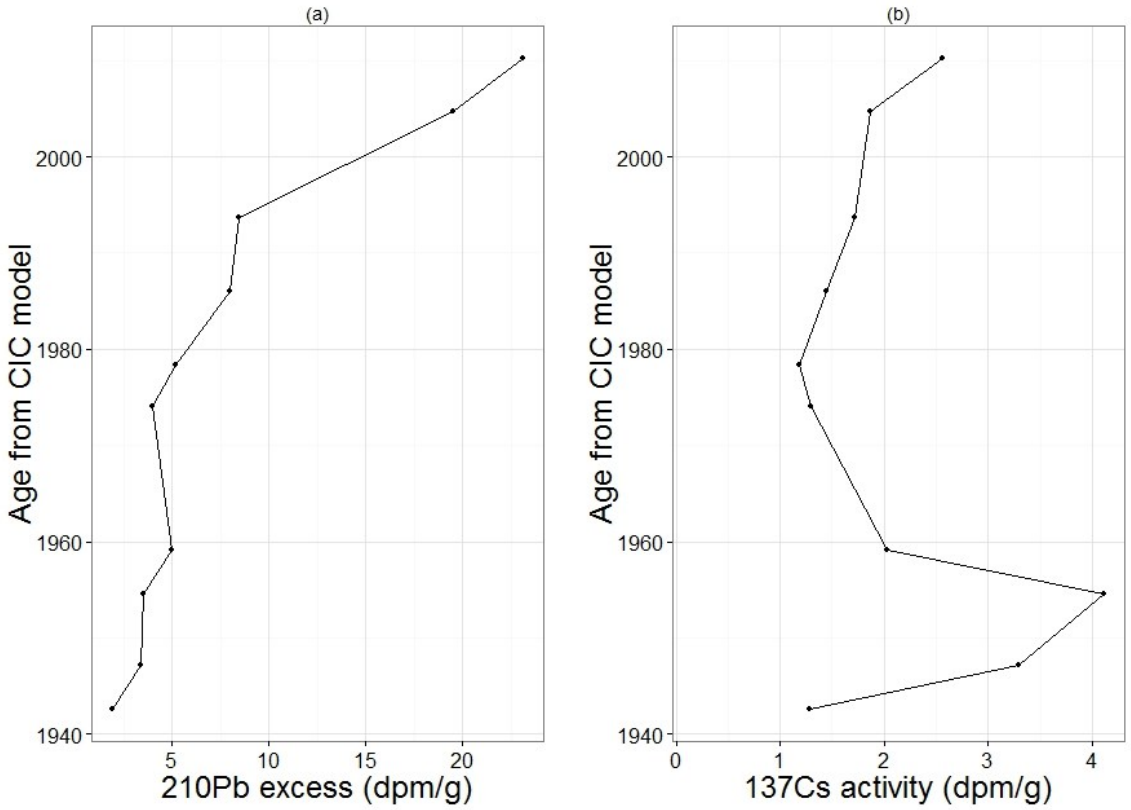



**Figure S2, 2: Age estimates from the CIC model in relation to unsupported (excess) $^{210}$Pb (a) and $^{137}$Cs activity (b) in the Dauriat core.** Gaps in unsupported $^{210}$Pb come from points where estimate was less than zero.

**(B) Knob (Latitude: 54.792068, Longitude: -66.808004 )**

*Activity data*

| Core interval (cm) | Mean depth (cm) | Depth error (cm) | $^{210}$Pb (dpm g$^{-1}$) | $^{210}$Pb error | $^{226}$Ra (dpm g$^{-1}$) | $^{226}$Ra error | $^{137}$Cs (dpm g$^{-1}$) | $^{137}$Cs error | $^{210}$Pb excess | ln(excess Pb) |
|---|---|---|---|---|---|---|---|---|---|---|
| **0-1.75** | 0.875 | 0.875 | 15.86 | 0.75 | 2.01 | 0.15 | 4.08 | 0.14 | 13.85 | 2.63 |
| **1.75-2.25** | 2 | 0.25 | 11.40 | 0.72 | 1.66 | 0.22 | 4.20 | 0.17 | 9.73 | 2.28 |
| **4-4.75** | 4.375 | 0.25 | 11.45 | 0.67 | 2.46 | 0.21 | 6.35 | 0.22 | 8.99 | 2.20 |
| **6-7** | 6.5 | 0.5 | 6.45 | 0.42 | 2.28 | 0.18 | 2.84 | 0.11 | 4.17 | 1.43 |
| **8.25-9** | 8.675 | 0.325 | 4.25 | 0.39 | 2.45 | 0.19 | 0.81 | 0.06 | 1.80 | 0.59 |
| **10-11** | 10.5 | 0.5 | 2.81 | 0.38 | 2.31 | 0.19 | 0.28 | 0.06 | 0.50 | NA |
| **12-13** | 12.5 | 0.5 | 2.58 | 0.31 | 2.75 | 0.19 | 0.18 | 0.05 | 0.00 | NA |
| **15-16** | 15.5 | 0.5 | 2.70 | 0.33 | 2.44 | 0.18 | 0.08 | 0.05 | 0.26 | NA |
| **18-18.75** | 18.375 | 0.375 | 2.70 | 0.40 | 2.50 | 0.21 | 0.04 | 0.06 | 0.21 | NA |
| **21-21.25** | 21.125 | 0.125 | 2.59 | 0.30 | 2.54 | 0.17 | 0.04 | 0.04 | 0.04 | NA |
| **24-24.25** | 24.125 | 0.125 | 3.33 | 0.35 | 2.31 | 0.18 | 0.00 | 0.00 | 1.01 | NA |
| **27.25-27.5** | 37.375 | 0.125 | 2.27 | 0.30 | 2.48 | 0.17 | 0.01 | 0.43 | 0.00 | NA |
| **30-30.25** | 30.125 | 0.125 | 3.06 | 0.38 | 3.01 | 0.22 | 0.05 | 0.06 | 0.05 | NA |
| **33-33.25** | 33.125 | 0.125 | 3.12 | 0.32 | 2.47 | 0.18 | 0.00 | 0.00 | 0.65 | NA |
| **33.5-33.75** | 33.675 | 0.075 | 2.32 | 0.27 | 2.37 | 0.16 | 0.00 | 0.00 | 0.00 | NA |

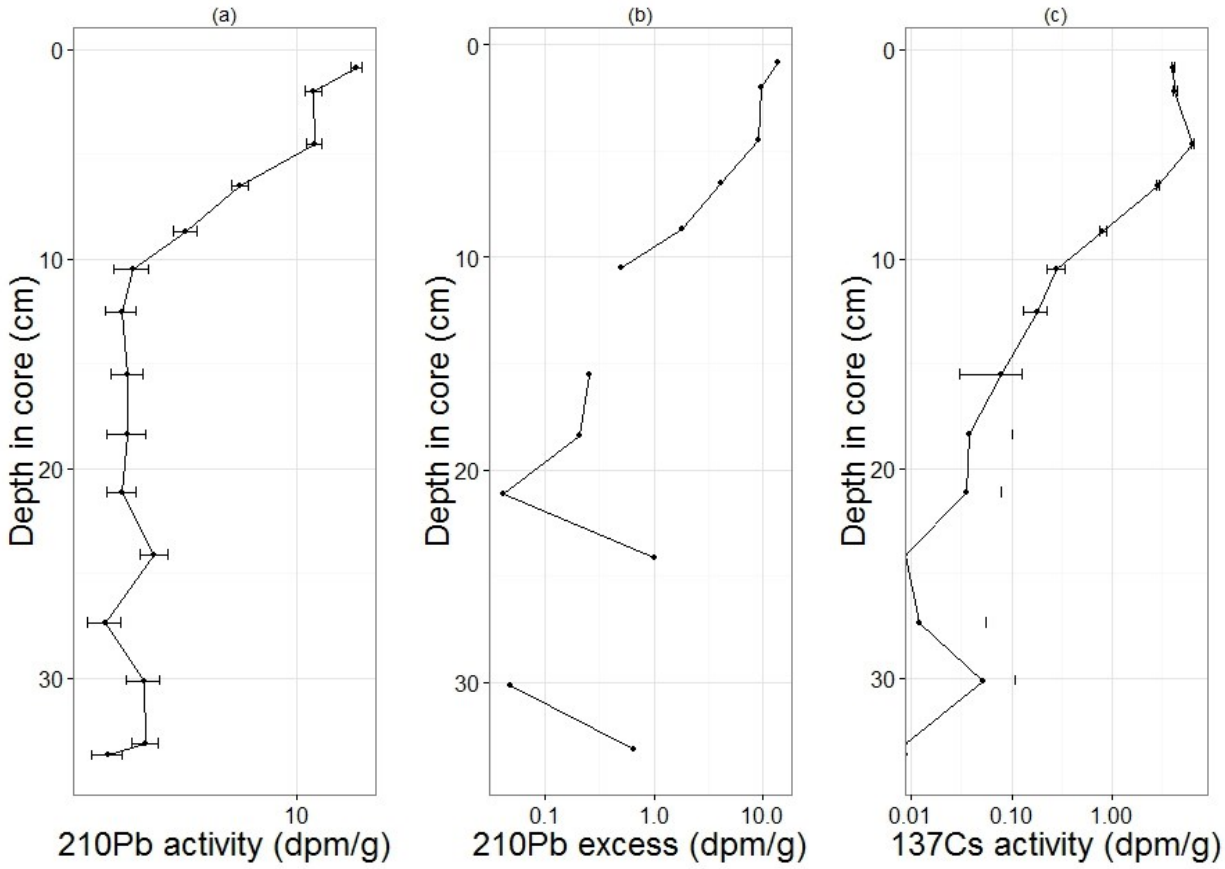



**Figure S2, 3: $^{210}$Pb (a), unsupported (excess) $^{210}$Pb (b) and $^{137}$Cs (c) activity on a log-scale as a function of depth (cm) in the Knob core.** Gaps in unsupported $^{210}$Pb come from points where estimate was less than zero.

*$^{210}$Pb limit model*

The limit of excess $^{210}$Pb was at approximately 10.5cm. Based on an estimated $^{210}$Pb limit at 120 years ago, the sedimentation rate down the core is estimated to be 0.0875 cm yr$^{-1}$ (10.5/120).

| Interval | Mean depth (cm) | Estimated sedimentation rate (cm yr$^{-1}$) | Cumulative years based on $^{210}$Pb limit | Age based on $^{210}$Pb limit |
|---|---|---|---|---|
| **0-1.75** | 0.875 | 0.09 | 10.00 | 2002.00 |
| **1.75-2.25** | 2 | 0.09 | 22.86 | 1989.14 |
| **4-4.75** | 4.5 | 0.09 | 51.43 | 1960.57 |
| **6-7** | 6.5 | 0.09 | 74.29 | 1937.71 |
| **8.25-9** | 8.675 | 0.09 | 99.14 | 1912.86 |
| **10-11** | 10.5 | 0.09 | 120.00 | 1892.00 |
| **12-13** | 12.5 | 0.09 | 142.86 | 1869.14 |
| **15-16** | 15.5 | 0.09 | 177.14 | 1834.86 |
| **18-18.75** | 18.375 | 0.09 | 210.00 | 1802.00 |
| **21-21.25** | 21.125 | 0.09 | 241.43 | 1770.57 |
| **24-24.25** | 24.125 | 0.09 | 275.71 | 1736.29 |
| **27.25-27.5** | 27.375 | 0.09 | 312.86 | 1699.14 |
| **30-30.25** | 30.125 | 0.09 | 344.29 | 1667.71 |
| **33-33.25** | 33.125 | 0.09 | 378.57 | 1633.43 |
| **33.5-33.75** | 33.675 | 0.09 | 384.86 | 1627.14 |

*CIC model*

Linear equation from the relationship between depth (x) and natural logarithm of unsupported $^{210}$Pb (y): -0.3193x + 3.1626.

Estimated sedimentation rate: ln(Pb half life) / slope = (0.03/0.3193) = 0.0973

| Interval | Mean depth (cm) | ln(unsupported Pb) | Estimated sedimentation rate (cm yr$^{-1}$) | Cumulative years based on CIC | Age based on CIC |
|---|---|---|---|---|---|
| **0-1.75** | 0.875 | 2.63 | 0.10 | 8.99 | 2003.01 |
| **1.75-2.25** | 2 | 2.28 | 0.10 | 20.55 | 1991.45 |
| **4-4.75** | 4.5 | 2.20 | 0.10 | 46.25 | 1965.75 |
| **6-7** | 6.5 | 1.43 | 0.10 | 66.80 | 1945.20 |
| **8.25-9** | 8.675 | 0.59 | 0.10 | 89.16 | 1922.84 |
| **10-11** | 10.5 | 0.00 | 0.10 | 107.91 | 1904.09 |
| **12-13** | 12.5 | NA | 0.10 | 128.47 | 1883.53 |
| **15-16** | 15.5 | NA | 0.10 | 159.30 | 1852.70 |
| **18-18.75** | 18.375 | NA | 0.10 | 188.85 | 1823.15 |
| **21-21.25** | 21.125 | NA | 0.10 | 217.11 | 1794.90 |
| **24-24.25** | 24.125 | NA | 0.10 | 247.94 | 1764.06 |
| **27.25-27.5** | 27.375 | NA | 0.10 | 281.35 | 1730.65 |
| **30-30.25** | 30.125 | NA | 0.10 | 309.61 | 1702.39 |
| **33-33.25** | 33.125 | NA | 0.10 | 340.44 | 1671.56 |
| **33.5-33.75** | 33.675 | NA | 0.10 | 346.09 | 1665.91 |

*CFCS model*

Not applicable, matches the CIC model.

*CRS model*

Because samples extracted downcore for activity measurements were not continuous, this model was not considered further as the total inventory of $^{210}$Pb would need to be interpolated in order to use this method.

*Model selection and comparisons*

The CFCS model provided the same results as the CIC model because there was only a single slope observable when examining the plot of depth within the core and the natural logarithm of excess $^{210}$Pb. The single slope used in the CIC model resulted in age estimates that were congruent with those from the $^{210}$Pb limit model. Age estimates from a CRS model could not be obtained because there was not continuous sampling of activity downcore. However, as historical information (and our own geochemical data) does not suggest changes in erosional processes around this lake, the CIC model was used for subsequent analyses.

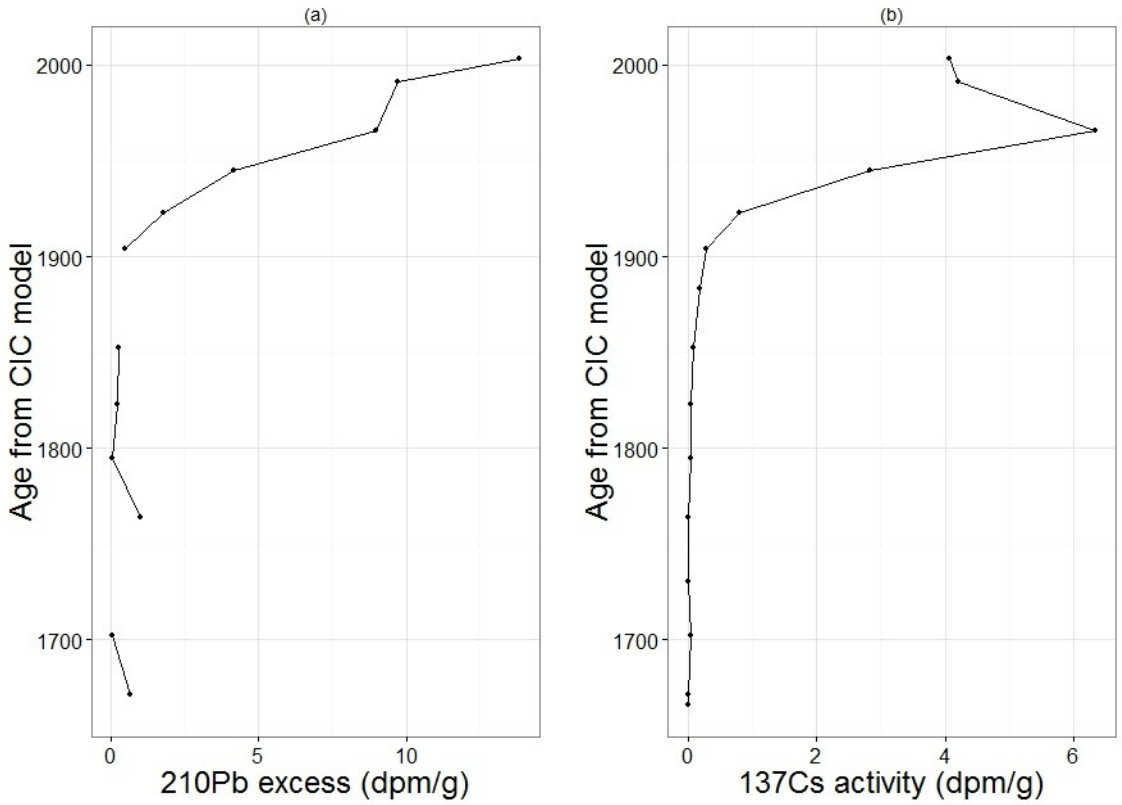



**Figure S2, 4: Age estimates from the CIC model in relation to unsupported (excess) $^{210}$Pb (a) and $^{137}$Cs activity (b) in the Knob core.** Gaps in unsupported $^{210}$Pb come from points where estimate was less than zero.

**Literature Cited**

Appleby P.G. (2001). Chronostratigraphic techniques in recent sediments. In *Tracking Environmental Change Using Lake Sediments. Volume 1: Basin Analysis, Coring, and Chronological Techniques*. Eds: Last, W.M., and Smol, J.P. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Ditchburn R.G., Barry B.J., Graham I.J., Levy R., Vandergoes M., & Zondervan A. (2011). Radiometric dating of sediment from lakes Ohau and Tekapo, New Zeland. The Geological Society of America Annual Meeting, 9-12 October 2011, Minneapolis, Minnesota, USA. Available online: https://gsa.confex.com/gsa/2011AM/finalprogram/abstract_197615.htm

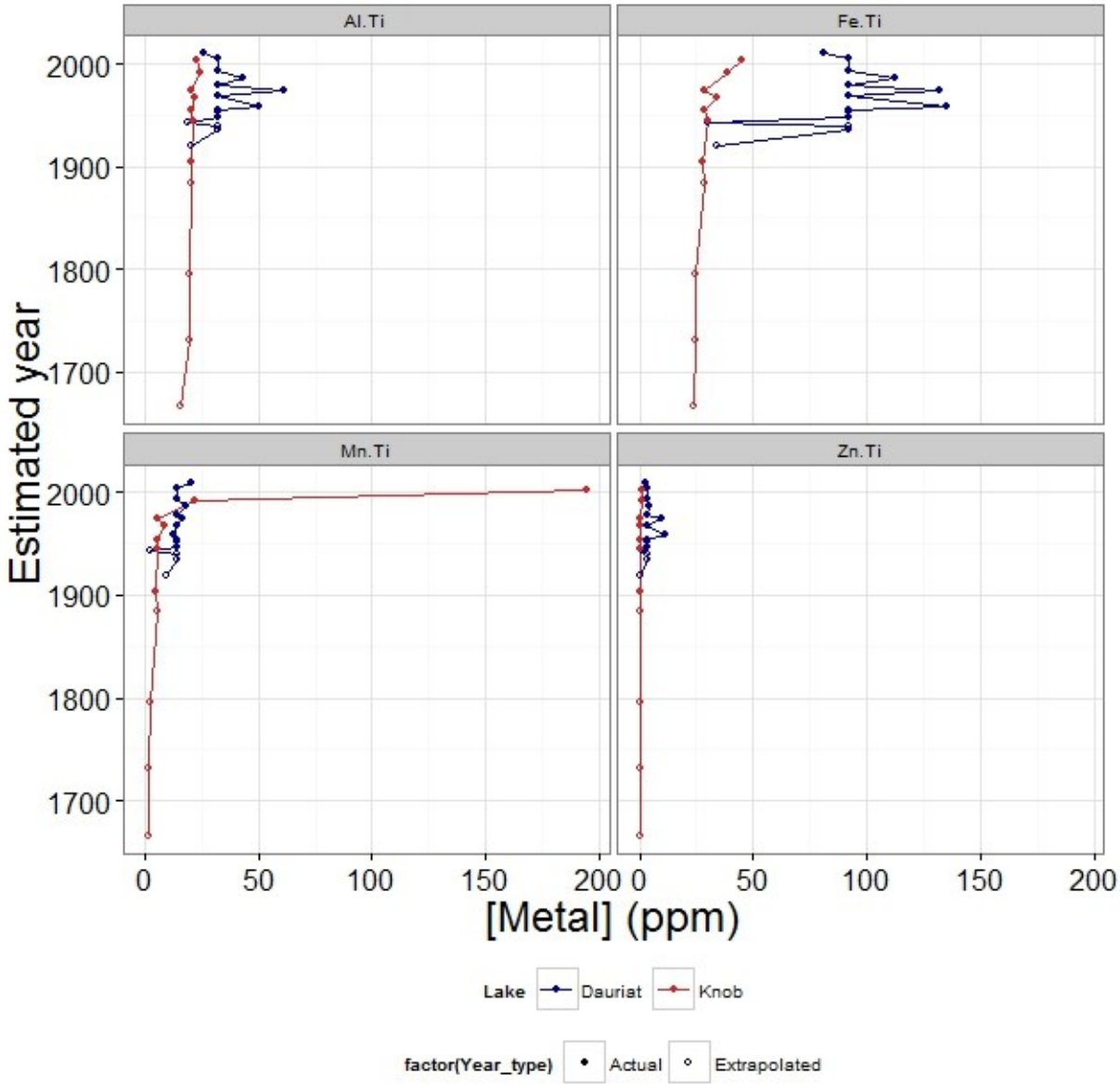
Ghaleb, B. (2007). Overview of the methods for the measurement and interpretation of short-lived radioisotopes and their limits. *Earth and Environmental Science,* **5**, 012007.

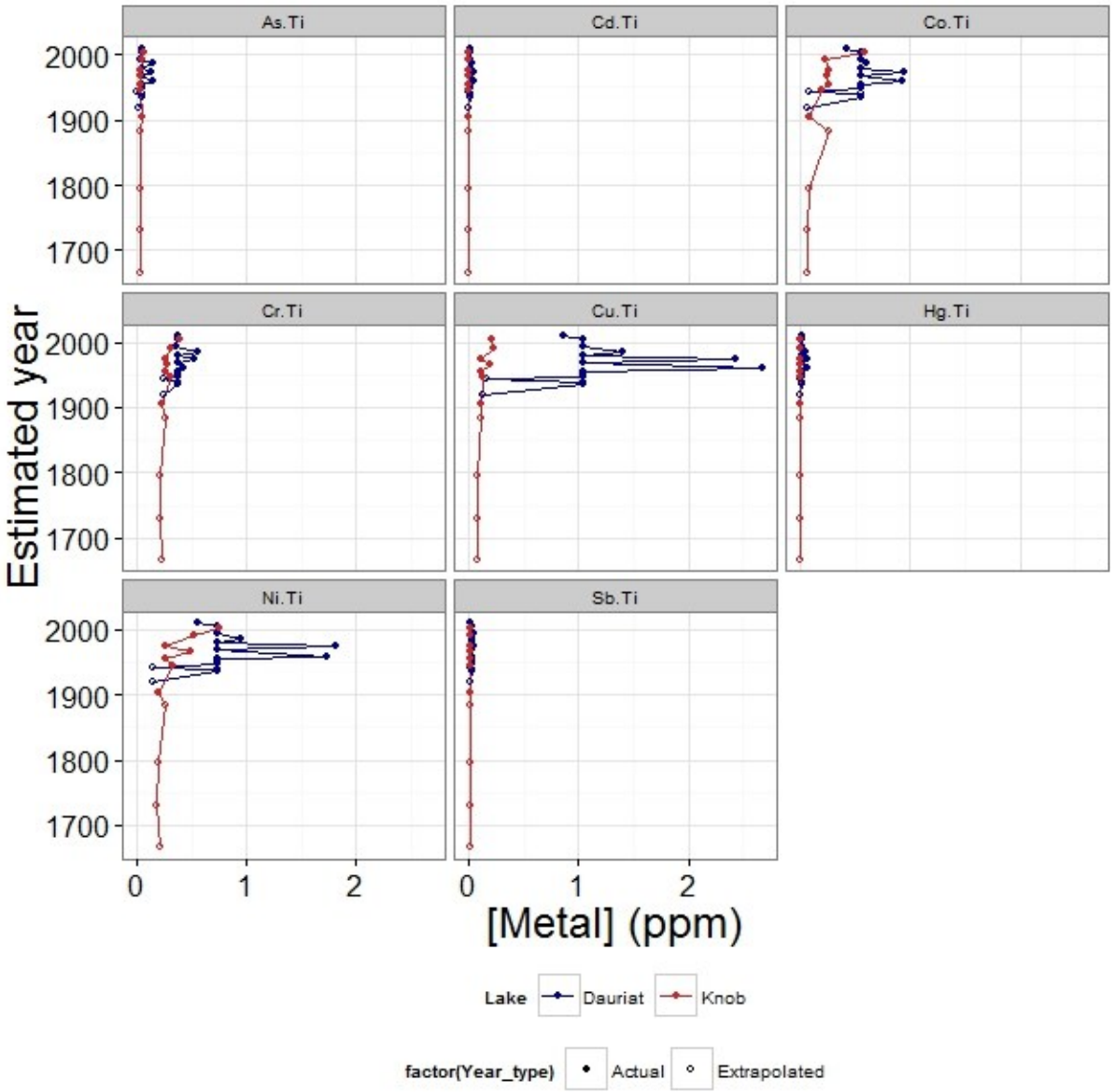
Sanchez-Cabeza J.A., & Ruiz-Fernández A.C. (2012). $^{210}$Pb sediment radiochronology: An integrated formulation and classification of dating models. *Geochimica et Cosmochimica Acta,* **82**, 183-200.

**Ch 3 S3: Models used to assess the relationship between cladoceran taxa richness and metal enrichment.** In each model, the response variable was either 'Clad_S', referring to rarefied cladoceran taxa richness or 'Clad_PC1', referring to the 1st axis scores from a PCA of cladoceran taxa. The explanatory variable was either 'Metal_EF', referring to cumulative enrichment factor or 'Metal_PC1', referring to the 1st axis scores from a PCA of heavy metal concentrations. The Lake parameter refers to lake identity, either 'Dauriat' or 'Knob'. Models were run using REML = TRUE and then REML = FALSE when assessed via AIC.

| Base model | Formula in *lmer*() | Description |
|---|---|---|
| $\mathbf{Clad\_S_{ij} \sim Metal\_EF_i + Lake_j + \varepsilon}$ | Clad_S ~ Metal_EF + (1\|Lake) | Varying intercept for lake (random factor) |
| | Clad_S ~ Metal_EF + (1 + Metal_EF\|Lake) | Varying intercept and slope for lake (varying intercept and slope with respect to metal EF) |
| | Clad_S ~ Metal_EF (*lm*() used) | **Null linear model** |
| $\mathbf{Clad\_S_{ij} \sim Metal\_PC1_i + Lake_j + \varepsilon}$ | Clad_S ~ Metal_PC1 + (1\|Lake) | Varying intercept for lake (random factor) |
| | Clad_S ~ Metal_PC1 + (1 + Metal_PC1\|Lake) | Varying intercept and slope for lake (varying intercept and slope with respect to metal PC1) |
| | Clad_S ~ Metal_PC1 (*lm*() used) | **Null linear model** |
| $\mathbf{Clad\_PC1_{ij} \sim Metal\_EF_i + Lake_j + \varepsilon}$ | Clad_PC1 ~ Metal_EF + (1\|Lake) | Varying intercept for lake (random factor) |
| | Clad_PC1 ~ Metal_EF + (1 + Metal_EF\|Lake) | Varying intercept and slope for lake |
| | Clad_PC1 ~ Metal_EF (*lm*() used) | **Null linear model** |
| $\mathbf{Clad\_PC1_{ij} \sim Metal\_PC1_i + Lake_j + \varepsilon}$ | Clad_PC1 ~ Metal_PC1 + (1\|Lake) | Varying intercept for lake (random factor) |
| | Clad_PC1 ~ Metal_PC1 + (1 + Metal_PC1\|Lake) | Varying intercept and slope for lake |
| | Clad_PC1 ~ Metal_PC1 (*lm*() used) | **Null linear model** |

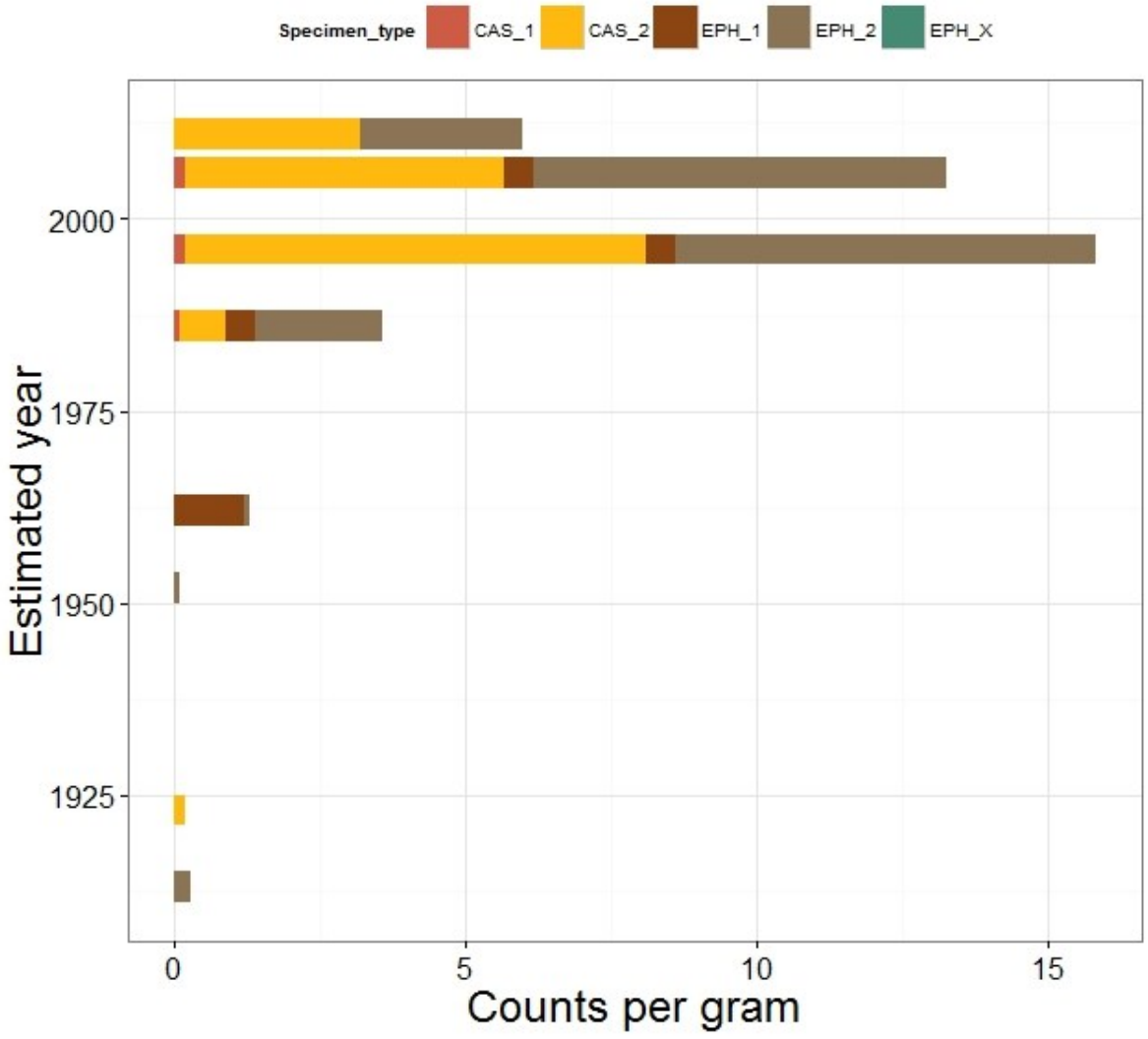
**Ch 3 S4: Full metal profiles for selected metals.** The top panel depicts Al, Fe, Mn and Zn (higher concentration metals). The bottom panel depicts As, Cd, Co, Cr, Cu, Hg, Ni, and Sb (lower concentration metals). Metals are expressed as ppm concentrations standardized by Titanium ppm concentrations.



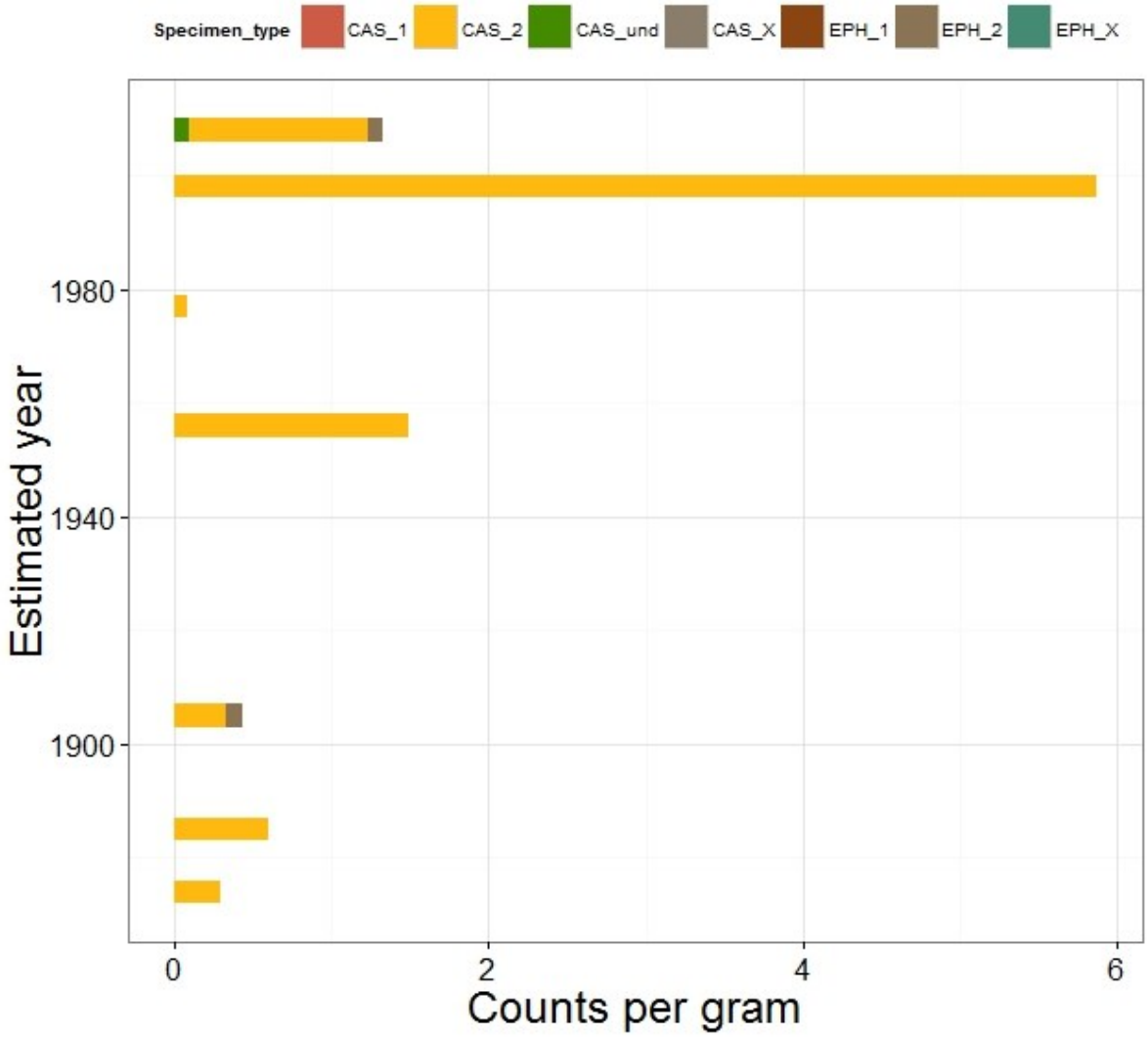

As.Ti
Cd.Ti
Co.Ti
Cr.Ti
Cu.Ti
Hg.Ti
Ni.Ti
Sb.Ti
Estimated year
2000
1900
1800
1700
0
1
2
[Metal] (ppm)
Lake
Dauriat
Knob
factor(Year_type)
Actual
Extrapolated

**Ch 3 S5: Cladoceran resting stages extracted from sediment intervals from Lake Dauriat (core 2 from 2013) and Knob (core 1 from 2013).** The sediment core number is specified at the top of each panel. 'Specimen_type' refers to the type of resting stage material observed in the samples: 'CAS_1' refers to an empty casing that would have held a single diapause egg; 'CAS_2' refers to an empty casing that would have held two diapause egg; 'CAS_X' refers to an empty casing that would have held more than two eggs; 'CAS_und' refers to casing where it could not be determined whether it would have held one or two eggs; 'EPH_1' refers to a casing with one egg present; 'EPH_2' refers to a casing with two eggs present; 'EPH_X' refers to a casing with multiple eggs present. No comment is made on the viability of the observed eggs. Gaps between histogram bars represent intervals where no resting stages were found in the sediment cores.

**(A) Dauriat**



**(B) Knob**


Specimen_type
CAS_1
CAS_2
CAS_und
CAS_X
EPH_1
EPH_2
EPH_X
Estimated year
1980
1940
1900
0
2
4
6
Counts per gram

## Appendix D: Supporting information for Chapter 4

### Ch 4 S1: DNA extraction protocol for the PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories Inc.).

As per the kit user protocol, we added 0.5-1.5 g of wet sediment for each sample, to PowerBead tubes with 15 mL of PowerBead solution and vortexing for one minute, then adding 1.2 mL of PowerMax solution C1 (cell lysis solution) and vortexing for an additional 30 seconds. Sample tubes were then shaken in a 65°C water bath for 30 minutes, followed by centrifugation for three minutes at room temperature (2500 x *g*). The supernatant from each sample tube was transferred to a clean tube, augmented with 5 mL of PowerMax solution C2 (inhibitor removal solution), inverted twice and incubated for 10 minutes at 4°C. Tubes were then centrifuged again for four minutes at 2500 x *g*. The supernatant removal, incubation and centrifuge steps were then repeated but with PowerMax solution C3 (inhibitor removal solution). We then added 30 mL of PowerMax solution C4 (high concentration salt solution) was added to the supernatant and centrifuged three more times using a the PowerMax spin filters, discarding the flow through each time (all three times at 1500 x *g* for two minutes at room temperature). Then, we added 10 mL of PowerMax solution C5 (wash solution) to the provided spin filters and centrifuged the samples at 2500 x *g* for three minutes at room temperature. The flow-through was discarded and the spin filter was centrifuged for an additional five minutes on the same settings. The spin filters were then placed into new collection tubes where 5 mL of PowerMax solution C6 (sterile elution buffer) was added, followed by further centrifugation at 2500 x *g* for three minutes at room temperature. Finally, the spin filters were discarded and DNA concentration for each sample was measured using a NanoDrop ® (Table 1). DNA samples were then frozen at -20°C. DNA concentrations of the extracts prior to amplification via PCR are shown in Table 1 (main text).

**Ch 4 S2: DNA concentrations from amplicons post-PCR step as measured using fluorometer (ng μL$^{-1}$).**

| Lake | Median interval depth (cm) | Estimated median year | Estimated year error | Eukaryote-primer | Diatom-primer | Bacteria-primer |
|---|---|---|---|---|---|---|
| **Dauriat** | 0.5 | 2011 | ~0 | 209.1 | 243.7 | 102.1 |
| **Dauriat** | 3.25 | 2004 | 1 | 235.0 | 278.3 | 141.8 |
| **Dauriat** | 7.25 | 1994 | 2.2 | 259.9 | 346.7 | 91.5 |
| **Dauriat** | 12.5 | 1981 | 2.6 | 258.0 | 303.9 | 62.9 |
| **Dauriat** | 16 | 1973 | 4 | 332.2 | 489.6 | 58.3 |
| **Dauriat** | 20.25 | 1963 | 6 | 291.3 | 405.9 | 89.5 |
| **Dauriat** | 25.5 | 1950 | 12.1 | 191.4 | 373.3 | 107.0 |
| **Dauriat** | 27.5 | 1945 | 15 | 177.3 | 353.7 | 53.2 |
| **Dauriat** | 33.5 | 1930 | 15 | 218.3 | 531.9 | 93.5 |
| **Dauriat** | 37.5 | 1920 | 15 | 6.8 | 112.7 | 99 |
| **Knob** | 0.5 | 2012 | ~0 | 136.9 | 326.0 | 88.5 |
| **Knob** | 2.75 | 1982 | 2.4 | 113.0 | 367.4 | 94.0 |
| **Knob** | 8 | 1928 | 10 | 146.0 | 358.2 | 98.5 |
| **Knob** | 14 | 1866 | 15 | 197.3 | 236.7 | 78.6 |
| **Knob** | 17.25 | 1833 | 15 | 98.3 | 426.7 | 102.3 |
| **Knob** | 26.75 | 1736 | 15 | 175.9 | 420.6 | 68.9 |

**Ch 4 S3: Parameters used in the PANAM pipeline.**

The analyses completed in the PANAM V3 pipeline were completed for each of the bacteria, diatom and eukaryote-targeted sequences using the following general parameters:

- Minimum read length allowed: 200 bp
- Minimum average quality score: 23
- Sequences were retained if they matched the forward primer.
- % mismatch of primers allowed: 0 (default)
- Clustering threshold: 0.95

*For sequences amplified by the general eukaryote primers*, the following information was used in the pipeline:

- Forward primer sequence: CGCGGTAATTCCAGCTCCA
- Reverse primer sequence: TTGGYRAATGCTTTCGC
- Maximum read length: 394 bp

*For sequences amplified by the bacterial primers*, the following information was used in the pipeline:

- Forward primer sequence: AYTGGGYDTAAAGNG
- Reverse primer sequence: CCGTCAATTCMTTTGAGTTT
- Maximum read length: 400 bp

*For sequences amplified by the diatom-specific primers*, the following information was used in the pipeline:

- Forward primer sequence: ATTCCAGCTCCAATAGCG
- Reverse primer sequence: GACTACGATGGTATCTAATC
- Maximum read length: 466 bp

**Ch 4 S4: Summary of diversity metrics for the eukaryotic assemblages** (n = 15, one Dauriat sample excluded).

| Lake | Estimated median year | # of OTUs | Shannon | Simpson | Pielou | Chao1 |
|---|---|---|---|---|---|---|
| **Dauriat** | 2011 | 240 | 4.9 | 0.99 | 0.89 | 378.5 |
| **Dauriat** | 2004 | 238 | 4.7 | 0.98 | 0.86 | 405.2 |
| **Dauriat** | 1994 | 157 | 3.3 | 0.87 | 0.65 | 378.8 |
| **Dauriat** | 1981 | 135 | 3.8 | 0.96 | 0.78 | 228.5 |
| **Dauriat** | 1973 | 73 | 2.7 | 0.84 | 0.62 | 142.1 |
| **Dauriat** | 1963 | 45 | 1.3 | 0.42 | 0.33 | 195.0 |
| **Dauriat** | 1950 | 124 | 3.8 | 0.95 | 0.79 | 219.6 |
| **Dauriat** | 1945 | 116 | 2.7 | 0.76 | 0.57 | 304.5 |
| **Dauriat** | 1930 | 71 | 2.0 | 0.61 | 0.47 | 155.1 |
| **Knob** | 2012 | 263 | 4.9 | 0.98 | 0.87 | 587.0 |
| **Knob** | 1982 | 201 | 4.2 | 0.96 | 0.79 | 395.6 |
| **Knob** | 1928 | 103 | 2.3 | 0.70 | 0.51 | 252.6 |
| **Knob** | 1866 | 63 | 1.4 | 0.41 | 0.33 | 104.3 |
| **Knob** | 1833 | 49 | 1.0 | 0.31 | 0.26 | 89.9 |
| **Knob** | 1736 | 86 | 1.8 | 0.55 | 0.40 | 171.8 |

**Ch 4 S5: Summary of considerations related to DNA quality and primer choice in this study.**

DNA purity and concentration as well as PCR procedures may have played a role in determining the sequencing success we obtained and thus influenced the final assemblages identified. While we did not find a relationship between DNA concentration of the DNA extracts from the sediments and depth in the sediment cores, measures of the purity of DNA, RNA and nucleic acids were generally below optimal amounts. DNA and RNA purity are typically assessed by the 260:280 ratio, where 1.8 is often accepted as representing 'pure' DNA and 2.0 as depicting 'pure' RNA (NanoDrop, 2007), Table 1. The 260:280 values for the samples used in this study ranged from 1.13 (the deepest Dauriat sample, that was eventually removed from the analyses) to 2 (surface sediments), Table 1 (main text). In terms of nucleic acid purity, for which an optimal range of the 260:230 ratio is 2.0-2.2, we found a decline over time (depth of the core) with values for Dauriat ranging from 2.09 to 0.43 and from 1.89 to 0.89 Lake Knob. As such, purity of the DNA extracts was not optimal for all samples and were somewhat below the optimal range for Lake Knob. It is not clear whether DNA/RNA/Nucleic acid purity compromised amplification and sequencing success. However, sequencing success across the sediment cores was also somewhat low compared to the potential output of various high-throughput sequencing platforms (e.g. Shokralla *et al*., 2012).

Cleaning procedures such as the one employed in the PANAM pipeline typically remove about 15-30% of DNA sequences (Debroas, Pers. Comm.), and the removal of singletons from DNA sequences can further reduce sequencing success. As such, we had low to moderate sequencing success for our samples. However we have chosen to mostly focus on comparisons of beta diversity between samples in relation to their depth within the cores and specifically for dominant groups only. So while the 741 rarefied DNA sequences for our eukaryotic DNA

sequences is too low for investigating the total diversity of eukaryotes, we can still compare the OTU distribution or number of OTUs within the identified dominant groups (i.e. Diatoms, Chlorophyta, Fungi).

We chose our three primer sets to target a broad consensus of aquatic biodiversity. While we employed both a diatom-specific primer and a general eukaryotic-primer, we had higher sequencing success for the diatoms with the general eukaryote primer. Indeed, the number of DNA sequences produced from the diatom-specific primer were very low for the majority of samples in Dauriat (except for the deepest and excluded sample) (Table 3; main text), even though the DNA sequences from the general eukaryote primer captured a considerable amount of diatom OTU diversity. Optimal primers/barcodes for diatom diversity is still an active field of research (Domaizon, Pers. Comm.), though it is generally accepted that the 18S marker is (with rbcl) the most appropriate marker for biomonitoring type studies (Mann *et al*., 2010). While diatom primers are well developed for water-column samples (e.g. Zimmerman *et al*., 2011), they are still not well characterized for sediment DNA, though a recent study by Capo *et al*., (2015), using the Mangot *et al*., (2013) primers had similar amplification success between both water-column and sediment samples dating back 60 years. That being said, there are many potential reasons for low primer success, e.g.: the targeting of a non-optimal barcode region or the length of the region amongst others. Additionally, both primer specificity and the taxonomic reference libraries used with high-throughput sequencing approaches are important; taxonomic coverage of reference libraries being particularly important for diatom communities in avoiding apparent amplification bias (Kermarrec *et al*., 2013).

**Literature Cited**

Capo E., Debroas D., Arnaud F., & Domaizon I. (2015). Is planktonic diversity well recorded in sedimentary DNA? Toward the reconstruction of past protistan diversity. *Microbiology of Aquatic Systems,* **70**, 865-875.

Debroas, D. Pers. Comm.

Domaizon, I. Pers. Comm.

Kermarrec L., Franc A., Rimet F., Chaumeil P., Humbert J.F., & Bouchez, A. (2013). Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources*, **13**, 607-619.

Mangot J-F., Domaizon I., Taib N., Marouni N., Duffaud E., Bronner G., & Debroas D. (2013). Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental Microbiology,* **15**, 1745-1758.

Mann D.G., Sato S., Trobajo R., Vanormelingen P., & Souffreau C., (2010). DNA barcoding for species identification and discovery in diatoms. *Cryptogram Algol*., **31**, 557-577.

NanoDrop. (2007). Technical Support Bulletin: 260/280 and 260/230 Ratios, NanoDrop® ND-1000 and ND-8000 8-Sample Spectrophotometers. Available online: http://www.bio.davidson.edu/gcat/protocols/NanoDrop_tip.pdf.

Shokralla S., Spall J.L., Gibson J.F., & Hajibabaei M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**, 1794-1805.

Zimmerman J., Jahn R., & Gemeinholzer B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, 10.1007/s13127-011-0050-6.

**Appendix E: Information on raw data sources and locations of data**

*Chapter 1*

All data used in Chapter 1 is available as Supplementary Information from Winegardner A.W., Beisner B.E., Legendre P., & Gregory-Eaves I. (2015). Are the landscape-level drivers of water column and surface sediment diatoms different? *Freshwater Biology,* **60**, 267-281: http://onlinelibrary.wiley.com/doi/10.1111/fwb.12478/full

*Chapter 2*

All data used in Chapter 2 is available from the U.S. EPA as part of the data release for the 2007 National Lakes Assessment: https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

The following .csv files were accessed along with their associated metadata:

- NLA 2007 Basin Landuse Metrics- Data 20061022
- NLA 2007 Site Information – Data 20091113
- National Lakes Assessment 2007 Final Data Notes
- NLA 2007 Buffer Landuse Metrics- Data 20091022
- NLA 2007 Chemical Conditions Estimates- Data 20091123
- NLA 2007 Phytoplankton Diatom Count- Data 20091023
- NLA 2007 Profile- Data 20091008
- NLA 2007 Secchi- Data 20091009

*Chapter 3*

The following raw data was used in Chapter 3:

**Table 1: Sample information for Lake Dauriat (DAR) and Lake Knob (KB).**

| Sample | Estimated year | Depth (cm) |
|---|---|---|
| DAR_1 | 2010 | 0.75 |
| DAR_2 | 2005 | 3 |
| DAR_3 | 1994 | 7.5 |
| DAR_4 | 1986 | 10.625 |
| DAR_5 | 1979 | 13.5 |
| DAR_6 | 1974 | 15.5 |
| DAR_7 | 1969 | 17.75 |
| DAR_8 | 1959 | 21.625 |
| DAR_9 | 1955 | 23.5 |
| DAR_10 | 1952 | 24.375 |
| DAR_11 | 1952 | 24.375 |
| DAR_12 | 1947 | 26.5 |
| DAR_13 | 1943 | 28.375 |
| DAR_14 | 1940 | 29.625 |
| DAR_15 | 1935 | 31.375 |
| DAR_16 | 1920 | 37.75 |
| KB_1 | 2003 | 0.875 |
| KB_2 | 1991 | 2 |
| KB_3 | 1975 | 3.625 |
| KB_4 | 1967 | 4.375 |
| KB_5 | 1954 | 5.625 |
| KB_6 | 1945 | 6.5 |
| KB_7 | 1904 | 10.5 |
| KB_8 | 1884 | 12.5 |
| KB_9 | 1795 | 21.125 |
| KB_10 | 1731 | 27.375 |
| KB_11 | 1666 | 33.675 |

**Table 2a: Cladoceran relative abundance data (all species, first 12 alpabetically).**

| Sample | *Acroperus harpae* | *Alona affinis* | *Alona barbulata* | *Alona circumfimbria guttata* | *Alona costata* | *Alona intermedia* | *Alona quadrangularis* | *Alona spp* | *Alonella excisa* | *Alonella nana* | *Bosmina longirostris* | *Eubosmina longispina* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0 | 0.008 | 0.017 | 0.215 | 0 | 0 | 0.008 | 0.157 | 0.008 | 0.041 | 0.248 | 0.123 |
| **DAR_2** | 0 | 0.020 | 0 | 0.408 | 0.041 | 0.041 | 0.020 | 0 | 0 | 0.061 | 0 | 0.327 |
| **DAR_3** | 0 | 0 | 0.015 | 0.200 | 0 | 0.007 | 0.015 | 0.081 | 0 | 0.015 | 0.178 | 0.385 |
| **DAR_4** | 0.012 | 0 | 0.036 | 0.181 | 0 | 0 | 0.012 | 0.145 | 0 | 0.024 | 0.205 | 0.313 |
| **DAR_5** | 0.019 | 0 | 0 | 0.340 | 0.038 | 0 | 0.038 | 0 | 0 | 0.057 | 0 | 0.302 |
| **DAR_6** | 0 | 0.013 | 0 | 0.067 | 0 | 0 | 0 | 0.173 | 0 | 0.013 | 0.187 | 0.413 |
| **DAR_7** | 0.023 | 0.116 | 0 | 0.372 | 0.023 | 0 | 0 | 0 | 0 | 0 | 0 | 0.186 |
| **DAR_8** | 0 | 0 | 0 | 0 | 0 | 0 | 0.008 | 0.017 | 0.008 | 0.008 | 0.471 | 0.372 |
| **DAR_9** | 0 | 0.018 | 0 | 0.074 | 0 | 0.019 | 0 | 0 | 0 | 0.037 | 0 | 0.630 |
| **DAR_10** | 0 | 0 | 0 | 0.097 | 0 | 0 | 0 | 0 | 0.032 | 0 | 0 | 0.581 |
| **DAR_11** | 0 | 0 | 0 | 0.097 | 0 | 0 | 0 | 0 | 0.032 | 0 | 0 | 0.581 |
| **DAR_12** | 0.048 | 0 | 0 | 0.095 | 0 | 0 | 0 | 0 | 0 | 0.048 | 0 | 0.643 |
| **DAR_13** | 0.027 | 0.054 | 0.041 | 0.041 | 0 | 0.014 | 0.014 | 0.041 | 0 | 0.014 | 0.149 | 0.365 |
| **DAR_14** | 0.075 | 0.050 | 0 | 0.050 | 0 | 0.050 | 0 | 0 | 0 | 0.075 | 0 | 0.275 |
| **DAR_15** | 0.120 | 0.049 | 0 | 0.195 | 0.024 | 0.049 | 0 | 0 | 0 | 0.049 | 0 | 0.195 |
| **DAR_16** | 0.083 | 0 | 0 | 0.042 | 0 | 0 | 0 | 0.250 | 0 | 0.042 | 0.292 | 0.125 |
| **KB_1** | 0.083 | 0.028 | 0 | 0.042 | 0 | 0.014 | 0.028 | 0 | 0 | 0 | 0.375 | 0.222 |
| **KB_2** | 0.065 | 0.116 | 0 | 0.026 | 0 | 0 | 0.026 | 0.013 | 0.013 | 0.026 | 0.091 | 0.390 |
| **KB_3** | 0.130 | 0.026 | 0 | 0 | 0 | 0 | 0.105 | 0 | 0 | 0.026 | 0 | 0.368 |
| **KB_4** | 0.037 | 0.037 | 0.009 | 0.009 | 0 | 0 | 0.037 | 0.065 | 0.009 | 0.009 | 0.374 | 0.140 |
| **KB_5** | 0.130 | 0.102 | 0 | 0.077 | 0 | 0 | 0.052 | 0 | 0 | 0.077 | 0 | 0.179 |
| **KB_6** | 0.028 | 0.084 | 0.028 | 0.014 | 0 | 0 | 0.042 | 0.014 | 0.014 | 0.014 | 0.127 | 0.338 |
| **KB_7** | 0.063 | 0.063 | 0.038 | 0.013 | 0 | 0 | 0.051 | 0 | 0 | 0.013 | 0.342 | 0.114 |
| **KB_8** | 0.040 | 0.100 | 0 | 0.120 | 0 | 0.020 | 0.060 | 0 | 0 | 0.040 | 0 | 0.360 |
| **KB_9** | 0.063 | 0.063 | 0.048 | 0.016 | 0 | 0 | 0.032 | 0 | 0.016 | 0.032 | 0.175 | 0.238 |
| **KB_10** | 0.029 | 0.044 | 0 | 0 | 0 | 0 | 0 | 0.029 | 0.015 | 0.029 | 0.265 | 0.206 |
| **KB_11** | 0.029 | 0.057 | 0 | 0.014 | 0 | 0 | 0.043 | 0.043 | 0 | 0.014 | 0.357 | 0.157 |

**Table 2b: Cladoceran relative abundance data (all species, second 12 alpabetically).**

| Sample | *Campto cercus spp* | *Chydorus. cf sphaericus* | *Chydorus. gibbus* | *Chydorus piger* | *Daphnia longispina* | *Daphnia pulex* | *Diasparalona rostrata* | *Dunhevedia crassa* | *Eurycercus spp* | *Graptolebris testudinaria* | *Holopedium. gibberum* | *Ilyocryptus spp* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0 | 0.066 | 0.025 | 0 | 0.050 | 0.008 | 0 | 0 | 0 | 0 | 0 | 0.008 |
| **DAR_2** | 0 | 0.061 | 0 | 0 | 0 | 0 | 0.020 | 0 | 0 | 0 | 0 | 0 |
| **DAR_3** | 0 | 0.044 | 0.015 | 0 | 0.007 | 0 | 0 | 0 | 0.007 | 0.015 | 0 | 0 |
| **DAR_4** | 0 | 0.048 | 0 | 0 | 0 | 0.012 | 0 | 0 | 0.012 | 0 | 0 | 0 |
| **DAR_5** | 0 | 0.151 | 0 | 0.019 | 0 | 0.037 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_6** | 0.013 | 0.080 | 0 | 0 | 0.013 | 0 | 0 | 0.013 | 0.013 | 0 | 0 | 0 |
| **DAR_7** | 0.047 | 0.140 | 0 | 0 | 0 | 0.047 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_8** | 0 | 0.091 | 0 | 0 | 0.008 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_9** | 0 | 0.167 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_10** | 0.016 | 0.226 | 0 | 0 | 0 | 0.016 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_11** | 0.016 | 0.226 | 0 | 0 | 0 | 0.016 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_12** | 0 | 0.143 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_13** | 0 | 0.068 | 0.014 | 0 | 0 | 0.014 | 0 | 0 | 0.014 | 0.041 | 0 | 0.041 |
| **DAR_14** | 0 | 0.250 | 0.050 | 0.025 | 0 | 0.025 | 0 | 0 | 0.050 | 0 | 0 | 0 |
| **DAR_15** | 0 | 0.244 | 0.024 | 0 | 0 | 0 | 0 | 0 | 0 | 0.024 | 0 | 0 |
| **DAR_16** | 0 | 0.083 | 0.042 | 0 | 0 | 0 | 0 | 0 | 0.042 | 0 | 0 | 0 |
| **KB_1** | 0.013 | 0.056 | 0.028 | 0 | 0.028 | 0 | 0 | 0 | 0.028 | 0 | 0 | 0 |
| **KB_2** | 0 | 0.078 | 0.026 | 0 | 0.052 | 0 | 0 | 0 | 0.039 | 0.013 | 0 | 0 |
| **KB_3** | 0.026 | 0.184 | 0.026 | 0 | 0 | 0.026 | 0 | 0 | 0 | 0.053 | 0 | 0 |
| **KB_4** | 0.009 | 0.131 | 0.065 | 0 | 0.009 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 |
| **KB_5** | 0 | 0.205 | 0.051 | 0 | 0 | 0 | 0 | 0 | 0.077 | 0.026 | 0 | 0 |
| **KB_6** | 0 | 0.127 | 0.056 | 0 | 0 | 0.014 | 0 | 0 | 0.029 | 0.014 | 0.014 | 0 |
| **KB_7** | 0.025 | 0.063 | 0.101 | 0 | 0 | 0 | 0 | 0 | 0.063 | 0 | 0 | 0 |
| **KB_8** | 0 | 0.180 | 0 | 0 | 0 | 0 | 0 | 0 | 0.060 | 0 | 0 | 0 |
| **KB_9** | 0 | 0.048 | 0.158 | 0 | 0 | 0.016 | 0 | 0 | 0.064 | 0.016 | 0 | 0 |
| **KB_10** | 0.015 | 0.088 | 0.147 | 0 | 0 | 0.015 | 0 | 0 | 0.074 | 0.015 | 0 | 0 |
| **KB_11** | 0 | 0.071 | 0.086 | 0 | 0.014 | 0 | 0 | 0 | 0.043 | 0 | 0 | 0 |

**Table 2c: Cladoceran relative abundance data (all species, third 12 alpabetically).**

| Sample | *Leydigia leydigi* | *Monospilus dispar* | *Ophryoxus gracilis* | *Paralona. piger* | *Pleuroxus denticulatus* | *Pleuroxus laevis* | *Pleuroxus .procurvus* | *Pleuroxus trigonellus* | *Pleuroxus. truncatus* | *Pleuroxus spp* | *Polyphemus pediculus* | *Sida crystalina* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0.008 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.008 |
| **DAR_2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.007 | 0 | 0 | 0 | 0.007 |
| **DAR_4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_5** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_6** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_7** | 0 | 0.023 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.023 |
| **DAR_8** | 0 | 0 | 0 | 0.008 | 0 | 0 | 0 | 0.008 | 0 | 0 | 0 | 0 |
| **DAR_9** | 0.037 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 |
| **DAR_10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.016 | 0 | 0 | 0.016 |
| **DAR_11** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.016 | 0 | 0 | 0.016 |
| **DAR_12** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.024 |
| **DAR_13** | 0 | 0 | 0 | 0.054 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_14** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.025 |
| **DAR_15** | 0 | 0.024 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **DAR_16** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **KB_1** | 0 | 0 | 0.014 | 0.014 | 0 | 0 | 0 | 0 | 0 | 0.014 | 0 | 0.014 |
| **KB_2** | 0 | 0 | 0 | 0 | 0 | 0.013 | 0 | 0 | 0 | 0 | 0 | 0.013 |
| **KB_3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.026 |
| **KB_4** | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0.009 | 0 | 0.009 | 0 | 0 |
| **KB_5** | 0 | 0.026 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **KB_6** | 0 | 0 | 0.014 | 0.014 | 0 | 0 | 0 | 0 | 0 | 0 | 0.014 | 0 |
| **KB_7** | 0 | 0 | 0.013 | 0.025 | 0 | 0 | 0 | 0 | 0 | 0.0127 | 0 | 0 |
| **KB_8** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.020 | 0 | 0 | 0 |
| **KB_9** | 0 | 0 | 0 | 0.016 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **KB_10** | 0 | 0 | 0 | 0.029 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **KB_11** | 0 | 0 | 0 | 0.029 | 0.014 | 0.014 | 0 | 0.014 | 0 | 0 | 0 | 0 |

**Table 3a: Geochemical data, standardized by Titanium (first 12).**

| Sample | Ag:Ti | Al:Ti | As:Ti | Ba:Ti | Be:Ti | Bi:Ti | Ca:Ti | Cd:Ti | Co:Ti | Cr:Ti | Cu:Ti | Fe:Ti |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0.026 | 26.048 | 0.048 | 1.186 | 0.029 | 0.009 | 1.762 | 0.016 | 0.424 | 0.376 | 0.857 | 80.952 |
| **DAR_2** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_3** | 0.051 | 32.000 | 0.026 | 1.164 | 0.036 | 0.017 | 1.727 | 0.024 | 0.555 | 0.364 | 1.045 | 91.818 |
| **DAR_4** | 0.063 | 43.143 | 0.143 | 0.814 | 0.057 | 0.036 | 2.357 | 0.029 | 0.607 | 0.550 | 1.400 | 112.143 |
| **DAR_5** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_6** | 0.079 | 61.333 | 0.125 | 0.567 | 0.083 | 0.016 | 3.500 | 0.042 | 0.942 | 0.525 | 2.425 | 132.500 |
| **DAR_7** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_8** | 0.063 | 50.462 | 0.146 | 0.631 | 0.062 | 0.015 | 2.923 | 0.044 | 0.931 | 0.423 | 2.677 | 135.385 |
| **DAR_9** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_10** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_11** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_12** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_13** | 0.002 | 18.968 | 0.009 | 1.306 | 0.010 | 0.006 | 1.065 | 0.004 | 0.087 | 0.242 | 0.168 | 30.355 |
| **DAR_14** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_15** | 0.051 | 32.000 | 0.048 | 1.164 | 0.036 | 0.015 | 1.762 | 0.024 | 0.555 | 0.376 | 1.045 | 91.818 |
| **DAR_16** | 0.001 | 20.265 | 0.012 | 1.315 | 0.009 | 0.006 | 0.794 | 0.002 | 0.062 | 0.250 | 0.135 | 34.118 |
| **KB_1** | 0.003 | 22.476 | 0.062 | 2.343 | 0.014 | 0.009 | 3.095 | 0.007 | 0.590 | 0.395 | 0.219 | 45.238 |
| **KB_2** | 0.001 | 24.130 | 0.048 | 1.570 | 0.013 | 0.008 | 1.696 | 0.007 | 0.235 | 0.313 | 0.226 | 39.130 |
| **KB_3** | 0.001 | 20.581 | 0.038 | 1.422 | 0.010 | 0.007 | 1.165 | 0.003 | 0.260 | 0.260 | 0.122 | 28.756 |
| **KB_4** | 0.002 | 22.269 | 0.035 | 1.435 | 0.012 | 0.007 | 1.385 | 0.006 | 0.246 | 0.285 | 0.196 | 33.808 |
| **KB_5** | 0.002 | 20.581 | 0.038 | 1.422 | 0.010 | 0.007 | 1.165 | 0.003 | 0.260 | 0.260 | 0.122 | 28.756 |
| **KB_6** | 0.002 | 21.033 | 0.033 | 1.407 | 0.010 | 0.006 | 1.233 | 0.004 | 0.197 | 0.310 | 0.137 | 29.900 |
| **KB_7** | 0.002 | 20.129 | 0.045 | 1.365 | 0.010 | 0.006 | 1.097 | 0.002 | 0.074 | 0.223 | 0.106 | 27.613 |
| **KB_8** | 0.002 | 20.581 | 0.038 | 1.422 | 0.010 | 0.007 | 1.165 | 0.003 | 0.260 | 0.260 | 0.122 | 28.756 |
| **KB_9** | 0.002 | 19.323 | 0.039 | 1.410 | 0.010 | 0.006 | 1.032 | 0.002 | 0.074 | 0.212 | 0.081 | 25.065 |
| **KB_10** | 0.002 | 19.967 | 0.037 | 1.470 | 0.007 | 0.006 | 1.033 | 0.001 | 0.067 | 0.213 | 0.0800 | 24.733 |
| **KB_11** | 0.001 | 16.107 | 0.036 | 1.296 | 0.007 | 0.007 | 0.714 | 0.002 | 0.072 | 0.236 | 0.089 | 24.000 |

**Table 3b: Geochemical data, standardized by Titanium (second 12).**

| Sample | Ga:Ti | Hg:Ti | K:Ti | Li:Ti | Mg:Ti | Mn:Ti | Mo:Ti | Na:Ti | Ni:Ti | P:Ti | Pb:Ti | S:Ti |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0.029 | 0.010 | 6.333 | 0.129 | 3.286 | 20.333 | 0.005 | 1.238 | 0.562 | 1.157 | 0.433 | 1.619 |
| **DAR_2** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_3** | 0.027 | 0.009 | 6.364 | 0.127 | 3.909 | 14.545 | 0.018 | 1.181 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_4** | 0.021 | 0.050 | 6.857 | 0.150 | 5.071 | 18.357 | 0.029 | 1.214 | 0.943 | 2.479 | 0.879 | 21.64 |
| **DAR_5** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_6** | 0.025 | 0.058 | 6.750 | 0.208 | 4.667 | 16.917 | 0.042 | 1.083 | 1.817 | 3.425 | 0.833 | 38.250 |
| **DAR_7** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_8** | 0.031 | 0.062 | 6.923 | 0.131 | 3.615 | 12.462 | 0.046 | 1.077 | 1.738 | 4.700 | 0.808 | 27.615 |
| **DAR_9** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_10** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_11** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_12** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_13** | 0.035 | 0.003 | 6.968 | 0.084 | 2.000 | 2.4645 | 0.016 | 1.710 | 0.152 | 0.326 | 0.084 | 0.806 |
| **DAR_14** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_15** | 0.029 | 0.010 | 6.857 | 0.129 | 3.615 | 14.545 | 0.018 | 1.214 | 0.736 | 1.673 | 0.664 | 9.818 |
| **DAR_16** | 0.041 | 0.003 | 7.676 | 0.094 | 2.324 | 9.735 | 0.015 | 1.500 | 0.144 | 0.156 | 0.044 | 0.118 |
| **KB_1** | 0.081 | 0.004 | 6.429 | 0.148 | 2.762 | 194.762 | 0.033 | 1.381 | 0.752 | 0.410 | 0.152 | 0.762 |
| **KB_2** | 0.091 | 0.004 | 7.130 | 0.130 | 2.696 | 22.261 | 0.017 | 1.435 | 0.522 | 0.391 | 0.170 | 0.652 |
| **KB_3** | 0.077 | 0.003 | 6.757 | 0.107 | 2.394 | 5.289 | 0.011 | 1.427 | 0.265 | 0.301 | 0.056 | 0.756 |
| **KB_4** | 0.085 | 0.003 | 6.808 | 0.119 | 2.500 | 8.731 | 0.012 | 1.346 | 0.488 | 0.385 | 0.138 | 0.692 |
| **KB_5** | 0.077 | 0.003 | 6.759 | 0.107 | 2.394 | 5.289 | 0.011 | 1.427 | 0.265 | 0.301 | 0.056 | 0.756 |
| **KB_6** | 0.077 | 0.003 | 6.867 | 0.110 | 2.433 | 5.933 | 0.010 | 1.333 | 0.320 | 0.343 | 0.080 | 1.067 |
| **KB_7** | 0.077 | 0.003 | 6.645 | 0.103 | 2.355 | 4.645 | 0.013 | 1.419 | 0.203 | 0.258 | 0.029 | 1.710 |
| **KB_8** | 0.077 | 0.003 | 6.759 | 0.107 | 2.394 | 5.289 | 0.011 | 1.427 | 0.265 | 0.301 | 0.056 | 0.756 |
| **KB_9** | 0.071 | 0.003 | 6.710 | 0.1.000 | 2.323 | 2.590 | 0.006 | 1.677 | 0.197 | 0.194 | 0.029 | 0.774 |
| **KB_10** | 0.073 | 0.003 | 7.000 | 0.1.000 | 2.333 | 1.910 | 0.003 | 1.933 | 0.187 | 0.1700 | 0.030 | 0.733 |
| **KB_11** | 0.075 | 0.003 | 4.893 | 0.104 | 2.250 | 1.818 | 0.007 | 1.857 | 0.211 | 0.176 | 0.032 | 0.750 |

**Table 3c: Geochemical data, standardized by Titanium (last 8).**

| Sample | Sb:Ti | Sc:Ti | Sr:Ti | Te:Ti | V:Ti | Y:Ti | Zn:Ti | Zr:Ti |
|---|---|---|---|---|---|---|---|---|
| **DAR_1** | 0.023 | 0.033 | 0.205 | 0.009 | 0.395 | 0.205 | 2.352 | 0.229 |
| **DAR_2** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_3** | 0.045 | 0.036 | 0.200 | 0.091 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_4** | 0.035 | 0.050 | 0.236 | 0.043 | 0.450 | 0.450 | 4.429 | 0.350 |
| **DAR_5** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_6** | 0.041 | 0.092 | 0.292 | 0.016 | 0.475 | 0.667 | 9.750 | 0.383 |
| **DAR_7** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.1363 | 0.327 |
| **DAR_8** | 0.038 | 0.092 | 0.277 | 0.015 | 0.485 | 0.592 | 11.000 | 0.231 |
| **DAR_9** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_10** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_11** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_12** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_13** | 0.016 | 0.032 | 0.261 | 0.010 | 0.310 | 0.081 | 0.832 | 0.239 |
| **DAR_14** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_15** | 0.035 | 0.036 | 0.236 | 0.015 | 0.427 | 0.309 | 3.136 | 0.327 |
| **DAR_16** | 0.014 | 0.035 | 0.218 | 0.009 | 0.344 | 0.074 | 0.250 | 0.415 |
| **KB_1** | 0.023 | 0.033 | 0.248 | 0.009 | 0.348 | 0.148 | 0.862 | 0.219 |
| **KB_2** | 0.021 | 0.039 | 0.222 | 0.008 | 0.387 | 0.148 | 0.870 | 0.0826 |
| **KB_3** | 0.017 | 0.037 | 0.209 | 0.013 | 0.346 | 0.090 | 0.499 | 0.450 |
| **KB_4** | 0.019 | 0.038 | 0.196 | 0.007 | 0.373 | 0.127 | 0.750 | 0.269 |
| **KB_5** | 0.017 | 0.037 | 0.209 | 0.013 | 0.346 | 0.090 | 0.499 | 0.450 |
| **KB_6** | 0.016 | 0.040 | 0.187 | 0.023 | 0.363 | 0.097 | 0.557 | 0.487 |
| **KB_7** | 0.016 | 0.039 | 0.187 | 0.013 | 0.345 | 0.084 | 0.442 | 0.468 |
| **KB_8** | 0.012 | 0.037 | 0.209 | 0.013 | 0.346 | 0.090 | 0.499 | 0.450 |
| **KB_9** | 0.016 | 0.035 | 0.226 | 0.039 | 0.323 | 0.071 | 0.390 | 0.465 |
| **KB_10** | 0.016 | 0.033 | 0.267 | 0.013 | 0.320 | 0.067 | 0.367 | 0.450 |
| **KB_11** | 0.018 | 0.025 | 0.179 | 0.018 | 0.329 | 0.057 | 0.400 | 0.450 |

*Chapter 4*

The same radiometric dating and geochemical data used in Chapter 3 was also used for Chapter 4. Traditional diatom taxonomy data from Lake Dauriat was obtained from:
Laperrière L., Fallu M-A., Hausmann S., Pienitz R., & Muir D. (2008). Paleolimnological evidence of mining and demographic impacts on Lac Dauriat, Schefferville (subarctic Québec, Canada). *Journal of Paleolimnology*, **40**, 309-324.

Raw sequence data for the eurkaryotic sequences will be available on GenBank and the accession numbers reported:
Winegardner A.K., Capo E., Domaizon I., Debroas D., Hajibabei M., Shokralla S., Wing B., Beisner B.E., & Gregory-Eaves, I. (In preparation). Microbial eukaryotic biodiversity dynamics during the Anthropocene from a northern mining region: an exploration using High-Throughput Sequencing.