Hierarchical Spatial Structuring of Stream Insect Diversity through DNA Barcoding

by

Trevor T. Bringloe

A Thesis

presented to

The University of Guelph

In partial fulfillment of requirements

for the degree of

Master of Science

in

Integrative Biology

Guelph, Ontario, Canada

ABSTRACT

Hierarchical Spatial Structuring of Stream Insect Diversity through DNA Barcoding

Trevor T. Bringloe

University of Guelph 2014

Advisors:

Dr. Sarah J. Adamowicz

Dr. Karl Cottenie

Biodiversity is often studied in the context of species distributions across spatial scales. Diversity components analysis—the partitioning of total diversity into local diversity and distributional heterogeneity measures—assesses the spatial structure of biodiversity. While previous works have relied on morphological specimen identifications, here, DNA barcoding is coupled with additive diversity partitioning to assess stream larval Trichoptera (caddisfly) species diversity across spatial scales ranging from m^2 to Canadian sub-arctic vs. temperate USA regions, and is used in conjunction with checkerboard analyses at a small spatial extent to investigate the importance of biotic interactions. I found that taxonomic resolution influenced the interpretation of results. In addition, Trichoptera diversity was similarly structured at two disparate regions, suggesting similar underlying mechanisms govern how regional diversity is distributed. Interspecific competition was important at small spatial scales. My thesis illustrates the utility of DNA-based species identification coupled with diversity partitioning in the study of biodiversity.

Acknowledgements

I thank everyone who provided help and hospitality during field collections: Bernard Sweeney, John Jackson, Jan Battle, Mike Broomall, and the Stroud Water Research Center staff in Pennsylvania, USA; Rory Eckenswiller and the Algonquin Wildlife Research Station staff; LeeAnn Fishback and the Churchill Northern Studies Center staff. I also thank those who helped with sample collection and processing: Vivian Harvey (funded by the Hart Research Assistantship in Aquatic Biodiversity), Emma Lorraine, Krisselle Rondolo, Susannah Ripley, Fatima Mitterboeck, Graham Ansell, Omar Zaheer, Randelee MacDonald, and Nicholas Jeffery. I especially want to thank Gillian Martin for her constant support and insight during the course of the project. I thank the staff of the Canadian Centre for DNA Barcoding at the Biodiversity Institute of Ontario, University of Guelph, for molecular analysis.

I also extend gratitude to my supervisors Dr. Sarah Adamowicz and Dr. Karl Cottenie for hearing out my many project ideas, constantly providing insight and feedback, and always expecting the best possible work. I also thank my committee member Dr. Hafiz Maherali for ensuring this was a manageable, carefully considered project.

I am also forever indebted to my newly wedded wife Lauren Murdock for her unwavering love and support during these past two years of our lives.

This project was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant and an NSERC Post-Graduate Scholarship, the Ontario Ministry of Research and Innovation through the Global Leadership Round in Genomics and Life Sciences program, a grant from the Government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project, an Ontario Graduate Scholarship, the O'Brien Fellowship, the Northern Research Fund (NRF) of the Churchill Northern Studies Centre, the government of Canada through the Northern Scientific Training Program (NSTP), and by the University of Guelph through the Work-Study program and an International Collaborative Research travel grant from the College of Biological Sciences.

Table of Contents

List of Tables

List of Figures

Chapter 1: The study of spatial patterns in species diversity

Why do we study biodiversity?

Humanity has long been fascinated by the diversity of life on earth. Today, our fascination with biodiversity is coupled with an urgent need to understand the abundance and distribution of species on earth, as human activity continues to drive species extinction (Pimm and Raven 2000). But what is biodiversity and why is it important? Biological diversity is defined by the International Convention on Biodiversity (2003) as "the variability among living organisms from all sources including, *inter alia*, terrestrial, marine and other aquatic ecosystems, and the ecological complexes of which they are part; this includes diversity within species, between species, and of ecosystems." In this respect, species diversity is a facet of biodiversity, but tends to be the focal point of efforts to conserve biological diversity on earth. The value of biodiversity can be justified from anthropocentric and non-anthropocentric viewpoints (Oksanen 1997). Anthropocentric viewpoints will attempt to pin an economic value on the maintenance of biodiversity; if biodiversity is vital to ecosystem function, then its global short-term value has been calculated at US $33 trillion, twice the global gross product (Costanza et al. 1997). The economic value of biodiversity ranges from raw materials for building houses to pharmaceuticals; current-day extinction rates are particularly alarming given that the potential benefits of these lost species remains unknown (Randall 1991). Many, however, prefer the intrinsic value of biodiversity given its uniqueness and irreplaceability; humans therefore have a moral obligation to understand and protect biodiversity since they are responsible for its rapid decline (Oksanen 1997).

From an ecological standpoint, biodiversity provides functional characteristics that determine processes such as food-webs and nutrient cycling, and in doing so, promotes ecosystem stability; not surprisingly, ecosystems respond to changes in species diversity (Bengtsson et al. 1997, Hooper et al. 2005). Biodiversity is also interesting from an evolutionary standpoint given at least 8.7 million species are estimated to currently occupy the earth (Mora et al. 2011). Modern-day estimates also indicate 86% of total species and 91% of ocean species still await description (Mora et al. 2011). Evolutionary biologists seek to understand how so many species came into existence, how they co-exist, and how this biodiversity is maintained through time. Biologists also struggle with

the notion of describing such an incredible number of species and are developing new methods for detecting and documenting biodiversity (Hebert 2003a, b).

Understanding the distribution of species diversity serves as critical basic knowledge for protecting biodiversity (Willis and Whittaker 2002). In particular, understanding the distribution of species according to spatial scale has become a central issue in ecology (Weins 1989). Spatial scale is important when considering biodiversity because different processes are expected to structure species diversity according to the spatial scale under consideration. Long-term biogeographic and evolutionary processes are important at continental to global spatial scales, shorter-term environmental and dispersal processes are important at regional scales, while short-term biotic interactions are likely important at very localized spatial scales (Willis and Whittaker 2002, Cavender-Bares et al. 2009). Several theoretical frameworks are now used to evaluate the relative importance of these processes structuring species diversity, including metacommunity theory (Leibold et al. 2004, Cottenie 2005) and phylogenetic community structure (Webb 2000, Webb et al. 2002); analyses such as checkerboard (Stone and Roberts 1990) and diversity partitioning (Crist et al. 2003) may also be used to assess species assemblages according to spatial scale. If our understanding of biodiversity is to be comprehensive, studies into the assemblage of species diversity must incorporate very small and truly large spatial scales.

The study of gamma (γ), beta (β), and alpha (α) diversity components

Before a theoretical framework can be used to determine processes that structure biodiversity, one must first understand at which spatial scales variation in species diversity occurs. Studies of this nature are exploratory and observational rather than hypothesis-driven experiments. Though not employing the hypothetico-deductive model of scientific enquiry, exploratory observational studies are integral to establishing baseline data from a natural or "real" world setting. In particular, such studies inform future hypotheses that can be tested to elucidate the mechanism structuring observed patterns. Diversity components analysis explores species diversity according to spatial scale by partitioning total diversity (gamma, γ) into constituent local diversity (alpha, α) and distributional heterogeneity in diversity units (beta, β) (Crist et al. 2003, Veech 2005). The recent surge in diversity components analysis no doubt owes much of its

success to the introduction of null models that allow testing for non-random distributions in these components (Crist et al. 2003). Studies using null models as a point of comparison can now explore the distribution of biodiversity according to space (Beck et al. 2012, Francisco-Ramos and Arias-González 2013), time (Martínez-Falcón et al. 2011, Pech-Canche et al. 2011), habitat changes or human impacts (Flohre et al. 2011, Caners et al. 2013), and even with regards to specialized species-host relationships (Morais et al. 2011). Diversity partitioning is also readily applied as a conservation tool, informing efforts on how to maximize the protection of diversity by determining crucial locations and spatial scales contributing to γ (Muller and Gossner 2010, Wu et al. 2010, Paknia and Pfeiffer 2011).

Diversity components can be partitioned in a number of ways depending on the objective of a given study, which has resulted in several ways to measure and interpret β diversity. The advent of diversity components can be traced back to Whittaker (1960, 1972), who initially proposed the multiplicative breakdown of regional diversity (γ) into local diversity (α) and community dissimilarity coefficient (β) components, where $\gamma=\alpha*\beta$. Multiplicative β (where $\beta=\gamma/\alpha$) is therefore a ratio of regional to local diversity and can be defined as the number of times as rich an entire (regional) set of species is compared to its constituent (local) units (Tuomisto 2010). The multiplicative method of partitioning diversity stands in contrast to the additive approach first proposed by MacArthur (1965) and reintroduced by Lande (1996) and Veech et al. (2002). In the additive sense, γ is partitioned as $\alpha+\beta$ (MacArthur 1965). Additive β (where $\beta=\gamma-\alpha$) is therefore defined as the amount of diversity (typically number of species) by which the regional data set (γ) exceeds the average amount of diversity in a single sampling unit (α) (Tuomisto 2010). Additive β has become especially prevalent in diversity components analyses given additive β has the same units as γ and α (number of species), making it a more tangible value, particularly for conservation efforts. A major contention between multiplicative and additive diversity partitioning is the relative independence of β; while β is never truly independent in both frameworks (given it is derived from measurements of γ and α), absolute values of β are highly correlated with values of γ and α when using additive partitioning, whereas in the multiplicative framework β behaves more independently for comparisons across spatial scales, site, or studies (Wilsey 2010). As such, additive β is

used to assess which spatial scales contribute most to γ, whereas the multiplicative framework assesses which spatial scales show the greatest turnover among sampled units (a spatial scale with high turnover may not necessarily contribute most to γ). During the course of this project, I define β as additive in order to identify important spatial scales contributing to total diversity. Because alternative definitions of β exist, the most appropriate definition will depend on the objectives of a given study (Jost 2006, 2007; Tuomisto 2010, Veech and Crist 2010, Wilsey 2010, Anderson et al. 2011, Melo et al. 2011, Jurasinski et al. 2012).

Studying biotic impacts on biodiversity distributions

While diversity partitioning is often successfully used to show non-random structure in biodiversity at local to regional spatial extents, it is far more difficult to infer the importance of very localized patterns and to determine how much of that variation is attributable to competitive interactions. This is especially true of observational data; while the observational approach collects data directly from a natural setting and provides the capacity for large-scale studies, determining exact mechanisms driving patterns in species distributions post hoc remains challenging. For instance, sampling over local to regional spatial extents necessarily conflates regional-scale processes, such as dispersal limitation and environmental filtering, with inferences of very localized biotic interactions because they may produce the same patterns (Willis and Whittaker 2002, Cavender-Bares et al. 2009). Observational frameworks used to infer the importance of competition have therefore met opposition, and while they are useful in the detection of non-random community structure, their utility for detecting interspecific interactions remains in doubt (Schoener and Adler 1991, Ulrich 2004, Mayfield and Levine 2010).

Checkerboard analysis, in particular, is used to assess the importance of competition in species assemblages. A checkerboard represents a case where two species from two sites do not occur together, producing a checkerboard pattern in a presence/absence site-by-species matrix (Diamond 1975, Stone and Roberts 1990). Presumably, if the incidence of checkerboards is greater than expected by chance in a full site-by-species matrix, then this pattern would reflect non-random species segregation driven by competitive interactions. Habitat heterogeneity and dispersal limitation between sites, however, may also generate the checkerboard pattern, muddling

interpretation of the metric (Schoener and Adler 1991). Over the course of this project, I explore the distribution of species diversity over a small spatial extent where environmental filtering and dispersal limitations are relatively unimportant compared to competitive interactions. I also use additive diversity components to first establish at which spatial scales variation in species distributions occur; checkerboards are then analysed at these spatial scales to infer if biotic interactions occur between specific taxa. This novel combination of methods is a much-needed approach for elucidating the role of biotic interactions in structuring the distribution of biodiversity.

Species identification through DNA barcoding

While our understanding of species diversity and distribution continues to develop, these advances come at a time when biologists are increasingly aware of severe limitations in our understanding of how many species occupy this earth and our ability to detect them (Hutchinson 1959, Hebert et al. 2003a, b, Smith et al. 2008). Molecular techniques for specimen identification, particularly DNA sequencing, has helped provide a more accurate picture of existing species diversity. DNA barcoding, in particular, was formally introduced more than a decade ago as an alternative way to assign species names to specimens, addressing concerns and limitations with traditional morphological identifications (Hebert et al., 2003a, b). The premise is to use DNA sequence data as the determinant of species membership, using a marker that is standardized across most animal taxa, a portion of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene. COI generally exhibits low genetic variation within a species, but suitable genetic divergence among species to enable species-level identification. DNA barcoding has been especially useful in revealing previously undetected diversity. For instance, Sweeney et al. (2011) demonstrated barcoding increased stream insect species inventories by almost 6-fold that of amateur identifiers and twice that of expert taxonomists, attributed in part to DNA barcoding's ability to identify immature or damaged specimens as well as to the incompleteness of taxonomic understanding of certain taxa. In addition, DNA barcoding can facilitate identification of sexually dimorphic organisms (Blagoev et al. 2013) and is also revealing hidden diversity by delineating cryptic species complexes in invertebrate (Hebert et al. 2004, Witt et al. 2006, Smith et al. 2008) and vertebrate taxa (Fouquet et al. 2007, Ward et al. 2008, Vieites et al. 2009). Though DNA barcoding

relies heavily on refined molecular techniques, studies of species diversity nonetheless stand to gain accuracy from specimen identification using DNA sequence data.

Stream insects as model system for biodiversity studies

Stream insects are well suited for assessing diversity patterns across spatial scales. Stream insects facilitate sampling in that they are ubiquitous and diverse, and freshwater systems can be clearly delineated for scaling up spatial extents. Freshwater insects also possess DNA-based species-level reference libraries through various biodiversity assessments at select regions, including Churchill (Manitoba, Canada; Zhou et al. 2009, 2010; Ruiter et al. 2013), Algonquin Provincial Park (Ontario, Canada; Martin 2013), and some areas of the Delaware Drainage in Philadelphia (USA; Sweeney et al. 2011). Caddisflies (order Trichoptera), in particular, are well suited for a large-scale study of diversity components; the diversity of larval caddisfly life histories, ranging from case-building filter feeders to free-living predators, ensures these insects inhabit all types of freshwater environments. Caddisflies are notable for their larval case building, which is achieved using silk and material ranging from woody debris to vacated shells (Marshall 2006). Caddisflies disperse as flighted adults along the periphery of streams, typically not leaving more than 60 m from the stream bank (Bilton et al. 2001). As a group, caddisflies can be found from the Canadian subarctic through to boreal and temperate forests. Few studies of diversity components have hierarchically sampled species diversity in such disparate regions. If general rules governing the distribution of species are to be identified, then spatial scales contributing to variation in species diversity must be determined in variable regions; these efforts inform further explicit analyses that aim to identify processes structuring diversity and to test whether these processes differ according to different habitats.

While caddisflies are particularly well suited for a large-scale analysis of species diversity, several insect orders are good candidates for a localized small-scale study into the importance of biotic interactions; these include caddisflies, beetles (order Coleoptera), and blackflies (family Simuliidae). Aquatic beetles are diverse, inhabiting nearly any habitat type (benthos to pelagic, intertidal to stream bank species) as aquatic larvae and adults, though the adult stage is terrestrial in some species (Morse and Holzenthal 2008). Signatures of competitive interactions may be maintained in fully aquatic species of

beetle as they are always present in the aquatic environment. Blackflies belong to the hyper-diverse insect order Diptera and are exclusively lotic freshwater filter feeders as larvae, attaching to the substrate with silk and hooks on the posterior proleg, catching particles with labral fans. Blackfly individuals will disperse relatively large distances (typically 15 km) as flighted adults in search of a blood meal (Morse and Holzenthal 2008). Blackfly larvae and caddisflies are known to undergo competitive interactions. Members of the caddisfly family Hydropsychidae are aggressively territorial for filter-feeding locations, attacking other hydropsychids and Simuliidae (behavioural observation: Jansson and Vuoristo 1979; field experiments: Hemphill 1988, Georgian and Thorp 1992). In a system where competitive interactions are known to be important, a small-scale study would be beneficial to establish at which spatial scale(s) species diversity is structured by these interactions, which in turn informs new hypotheses testing for interactions between specific taxa.

Thesis objectives

The objectives of my thesis were twofold, addressing both large and small-scale patterns in stream insect species diversity. My Chapter 2 objective was to assess small (m^2) to large-scale (19^o latitude) spatial patterns in species diversity of stream caddisflies using diversity components analyses. This objective included using DNA barcoding to explore limitations with taxonomic resolution and missed species, issues not addressed in previous studies investigating species diversity components with morphological identification. This study also entailed comparing patterns in additive diversity components at 3 disparate regions at differing latitudes, which might exhibit different patterns given the large gradient in habitat types (sub-Arctic, low-land boreal, and temperate habitats). My Chapter 3 objective was to test for signatures of competitive interactions at a single stream (White Clay Creek, PA, USA). Diversity partitioning was first used to establish if α diversity within White Clay Creek was less than expected by chance and to identify which within-stream spatial scales (i.e. β components) contribute most to the total species pool. Checkerboard analyses were then used at the spatial scale contributing most to total diversity to test hypotheses regarding competitive interactions between specific taxa. Chapter 4 addresses limitations with my thesis and proposes avenues for future research.

My thesis forwards novel methods for studying diversity components and checkerboard analyses. I expect to contribute to the advancement of biodiversity science by showcasing the utility of molecular-based species identification. My study of regional diversity components should help inform future hypotheses regarding how biodiversity becomes distributed through space. As well, I expect my localized study of diversity components and checkerboards will successfully showcase a novel observational approach to investigating the importance of biotic interactions in the assembly of species. In particular, my study will further our understanding of freshwater distributional patterns in species diversity, which in turn may inform stream conservation and restoration efforts.

Chapter 2: Diversity components of stream insects through DNA barcoding: small to large-scale patterns and the importance of taxonomic resolution

Abstract

The partitioning of species diversity (α, β, and γ) is commonly used to study diversity according to spatial scales but often faces difficulties associated with species-level morphological identification. The objective of this Chapter was to partition caddisfly species diversity across small (m^2) to large (19^o latitude) spatial extents using DNA barcoding for species identification. Caddisfly larvae were sampled in a hierarchical spatial design (m^2 kicks [5 m extent of replication], transects [20-50 m], streams [10 km], sub-regions [25-100 km], and regions [2500 km]) from streams in Churchill (MB, Canada), Algonquin Provincial Park (ON, Canada), and southern Pennsylvania (USA). Caddisfly diversity was additively partitioned for each region and for the full data set according to species, genus, and family taxonomic levels. The number of species missed within each region was also estimated by quantifying species historically sampled in each region but not sampled in the present study and by using diversity estimators (Chao1, Chao2). Sub-regional excess (β sub-region) of species was greater than expected by chance for all regions, while Churchill and Pennsylvania had significant excess of species at the stream level (β stream). Diversity components were similar for Churchill and Pennsylvania despite the large gradient in habitat parameters, suggesting similar processes drive the distribution of caddisfly species at disparate regions. The distribution of diversity across regions differed according to taxonomic level analyzed, particularly between species and family levels. Caddisfly families showed smaller regional and sub-regional excess of diversity compared to genera or species but yielded relatively larger estimates of diversity components at the stream and lower levels. Diversity components studies based on mixed taxonomic-level datasets are likely biasing their estimates of β depending on the spatial scale considered. Future studies may employ molecular techniques to ensure species-level resolution and reduce undersampling, in turn minimizing biases in the estimate of β diversity.

Introduction

Diversity components analysis, the partitioning of total diversity (gamma, γ) into constituent local diversity (alpha, α) and distributional heterogeneity (beta, β), can be used to determine the relative contribution of different spatial scales to overall diversity, which in turn informs efforts to understand processes driving the distribution of species diversity. Additive partitioning in particular has become common place, where $\beta=\gamma-\alpha$ and β is defined as the amount of diversity (typically number of species) by which the regional data set (γ) exceeds the average amount of diversity in a single sampling unit (α) (Tuomisto 2011). Studies of additive partitioning have covered a variety of taxa and spatial scales ranging from macro-fauna in deep-sea systems (Cunha 2013) to dung beetles in the Italian Alps (Negro et al 2011). Not surprisingly, these studies tend to find higher β diversity with increasing spatial scale, typically attributed to key habitat differences such as between ecoregions (Muller and Gossner 2010, Marques and Schoereder 2014), although marine environments may show low levels of β diversity due to increased connectivity among regions (Francisco-Ramos and Arias-González 2013). Relatively few studies cover spatial expanses greater than 1000 km, and only a few analyze diversity across a large latitudinal gradient (Hof et al. 2008, Thieltges et al. 2011, Kraft et al. 2011). Of these studies, two indicate coarse-grained β diversity decreases with increasing latitude for freshwater animal species (Hof et al. 2008, Thieltges et al. 2011), while Kraft et al. (2011) suggest the magnitude of difference between observed and expected fine-grained β diversity is actually consistent across latitudes in woody plant systems. It is thus not clear whether structure in diversity components generally differs according to latitude; more studies with hierarchical sampling of disparate regions covering a broad latitudinal band and more types of taxa are needed to resolve this question. I attempt to add insight on this issue by sampling caddisfly larvae across a 19^{o} latitudinal extent and adding a high degree of spatial resolution (i.e. β diversity is measured across small to large spatial scales).

The partitioning of species diversity across spatial scales, however, must contend with limitations in our ability to detect species (Hamilton et al. 2010, Mora et al. 2011). Diversity components studies, to my knowledge, have so far relied solely on morphological identifications when assigning species and are therefore subject to the

inherent potential inaccuracies with morphological identifications, particularly for hard-to-identify taxa such as invertebrates (Fig. 2.1). I analyzed studies of additive diversity components studies citing Crist et al. 2003, which introduced the statistical framework for analyzing additive partitions and is therefore cited by most studies of diversity components. Of 118 studies, 33% contain some issues with species identifications, with a clear taxonomic tendency towards difficulties identifying invertebrate taxa (Fig. 2.1). Solutions presented in these papers include identifying to the lowest possible taxonomic unit (15% of studies had mixed taxonomic level data, typically species and genera, occasionally family), the exclusion of specimens too difficult to identify (5% of articles; excluded specimens were typically invertebrate juveniles), and the use of morpho-species (15% of articles; some specimens could not be assigned to a recognized species name). 9% of the 118 articles did not appear to provide any context for how specimens were identified, while 58% simply indicated they identified to species according to a given taxonomic key (i.e. no problems indicated; Fig. 2.1). Even if no problems are indicated in a study, taxonomic resolution is still likely above the species level, especially for taxa in which molecular methods often reveal additional species (i.e. invertebrates, Hebert et al. 2004, Witt et al. 2006, Smith et al. 2008). The degree to which morphological identification impacts the taxonomic resolution of studies is difficult to estimate; articles almost never indicate the incidence rate of difficulties with specimen identification (though Summerville and Crist 2005 make a commendable effort by indicating <20% of their species total were morpho-species, which were verified by expert taxonomists). Also unclear is the incidence rate of undetected species in diversity components analysis; if species are not recorded due to insufficient sampling or difficulties in morphological identification, then values of diversity components will necessarily be impacted (generally leading to higher estimations of β, Beck et al. 2013). Here, I address the issue of taxonomic resolution by analyzing additive partitions of caddisfly diversity according to taxonomic level and estimate the missed species component contributing to γ.

Diversity components studies could resolve issues with morphological species identification using molecular data, in particular, DNA barcoding (Hebert et al. 2003a, b). DNA-based species identification facilitates the study of taxa for which traditional identification is challenging and those for which discriminating morphological features

are missing from many life phases or one sex. DNA-based specimen identification would also help with the detection of unknown species (for the same reasons above) while ensuring consistent identifications of known species. In addition, because DNA barcoding involves uploading specimen and sequence information to a public database (Barcode of Life Data System, BOLD, Ratnasingham and Hebert 2007; see Appendix I), consistent species-level occurrence data are available worldwide. BOLD may therefore be used to quantify missed diversity in the calculation of γ; the use of sequence data for specimen identification also facilitates comparisons of specimens across studies and sampled regions, ensuring identification of species remains consistent.

Chapter objectives

For Chapter 2, I chose to define β additively given this approach is used to quantify the relative contributions of multiple spatial scales to total species diversity. The additive approach in turn pinpoints spatial scales of interest for further study, quantifying in units of species richness those spatial scales that contribute most to total diversity and pinpointing those that deviate significantly from expected values. Defining β additively is also relevant given most diversity components studies employ this method; as such, my research questions address knowledge gaps in most of the primary literature on diversity components. The diversity of stream caddisflies (order Trichoptera) is used to address my research questions given their abundance and ubiquity across a large latitudinal gradient (see Chapter 1: stream insects as model system).

My main research question is "how is stream caddisfly species diversity structured across small (m^2) to large (19^o latitude) spatial extents?" I first assessed whether or not a large percentage of total diversity γ would remain undetected in my sampling despite intense sampling efforts and the use of molecular techniques to assist species identification. Studies of additive diversity partitioning, to my knowledge, have never directly quantified this "missed component," though sampling effort is routinely assessed with species rarefaction curves. If a large component of γ remains undetected as "missed species," then the exact structure of the diversity components will remain uncertain, and studies of additive diversity partitioning should consider the implications of undersampling bias.

My second objective was to determine if diversity components for each region along the sampled latitudinal gradient were similar or different. A few studies suggest coarse-grained β diversity decreases with latitude while fine-grained β diversity remains consistent. More information is needed about the structure of biodiversity in disparate regions, as characterized using consistent methods, in order to generate informed hypotheses about mechanisms driving large-scale patterns of biodiversity.

My last objective was to determine if taxonomic resolution impacts results of diversity components when expressed as a percentage of total diversity. The importance of taxonomic resolution was explored by comparing the partitioning of diversity at the species, genus, and family levels. Determining if taxonomic resolution impacts relative values of diversity components will shed light on current literature, in which species-level datasets are not always feasible and mixed taxonomic datasets are common.

Methods

Sampling of regions

In order to assess species diversity of stream caddisflies across multiple spatial scales, a hierarchically nested sampling design was implemented at 3 regions (Table 2.1, Figs. 2.2 and 2.3). For sampling of the 3 regions (Churchill, MB; Algonquin, ON; Southern Pennsylvania, USA) individual samples (m^2 kicks) were nested in transects running perpendicular to the flow direction (with the scale of replication being approximately 5 m); transects were nested within a stream reach (spanning approximately 20-50 m); streams were nested within a sub-region (a 10 km extent); sub-regions were nested in regions (a 25-100 km extent depending on region); regions span a latitudinal gradient from sub-arctic to temperate North America (19^o latitude, 2500 km extent; Fig. 2.2, Table 2.1). Transects within stream reaches consisted of 2 riffles and 1 pool as per the Ontario Benthos Biomonitoring Network: Protocol Manual to maximize the sampled diversity at a stream (OBBN, Jones et al. 2007). The 3 regions were selected on the basis of historical sampling (i.e. known stream locations in Churchill, MB, and Pennsylvania, USA), DNA barcode reference libraries (Zhou et al. 2009, 2010, Sweeney et al. 2011, Ruiter et al. 2013), and availability of existing data (Algonquin Provincial Park, ON; Martin 2013). The 3 regions also span several major habitat types. Sub-arctic Churchill is

a crossroads of different habitats including the Hudson Bay coast, tundra plains, and boreal forest, featuring harsh abiotic conditions for much of the year. Algonquin Provincial Park features over 7500 km^2 of lowland boreal forest, marking a mix of northern coniferous trees and southerly deciduous vegetation, and 1200 km of streams and rivers. Pennsylvania features temperate forests and a more "hilly" terrain. Algonquin Provincial Park data were obtained from pre-existing summer data (June 26-July 13, 2011; Martin 2013); consequently, this data set lacks within-stream spatial scales, but otherwise met the criteria of the above sampling design. The full sampling design was implemented at 9 streams from 3 sub-regions in Churchill, MB (July 15-23, 2012, Fig. 2.2) and at 9 streams from 3 sub-regions in southern Pennsylvania, USA, (June 17-19, 2013, Fig. 2.3). These dates were intended to represent comparable sampling times (summer) at each of the regions; overlap in species composition was assessed to test this assumption, ensuring between-region differences in species composition was not due to variation in phenology among the years sampled (Appendix II). The streams sampled represent the equivalent of 1^{st}-2^{nd} order streams (mean wetted width+SE=5.64+0.44 m, n=117 transects), though Pennsylvanian streams were the only ones to meet dendritic criteria (Churchill and Algonquin followed a less predictable drainage pattern, characteristic of the Canadian Shield).

Sampling effort was standardized by kicking a 1 m^2 of substrate for 2 minutes for each kick event (e.g. each sample) and allowing material to flow into the kick net (OBBN, Jones et al. 2007). A m^2 kick net (500 μm) was used for most streams; a d-net (500 μm) was used for one stream in Churchill, MB, which was too narrow for the kick net (R3-3, Fig. 2.2). Transects were sampled sequentially from downstream to upstream transects to avoid disturbing stream locations before collection occurred. The insects/material caught during each kick were washed into a bucket, and insects were picked off the net using forceps and included in the sample. The net was washed in the stream to remove noticeable particulates in between each kick. Three kicks/transect were taken in Churchill, while 2 kicks/transect were taken at Pennsylvania to facilitate sample processing due to the high abundance of insects. Each kick was subsampled in the field by stirring the material in the bucket (to homogenize the sample) and transferring a ladle (~200 ml) of material into a mason jar (236 or 473 ml depending on size of sample).

Additional ladles of sample were moved into the jar until more than 100 insects were estimated to be in the subsample, or until the sample was transferred in its entirety, for Algonquin and Churchill; this was later adapted to a 100 caddisfly rule for sampling in Pennsylvania. The number of ladles sampled from the field and total number of ladles in the sample bucket were recorded in order to extrapolate total abundance after sorting procedures. Samples were transported to a field lab and preserved using 95% ethanol, then stored in a freezer (-20 ^{o}C). Ethanol was changed at least once within 24 hours. Samples for the Algonquin data set were live sorted in the field; specimens were picked from the samples and placed in ethanol rather than storing the material to be studied under a microscope at a later date. Sample processing of the bulk samples consisted of removing all the invertebrates from the preserved material using a 4x dissecting microscope and sorting caddisflies into 20 ml vials according to family (CABIN, McDermott et al. 2010). Ethanol was changed regularly in the 20 ml vials until it remained clear in order to promote DNA preservation.

DNA barcoding of specimens

Specimens were selected from each m^2 kick and DNA barcoded to obtain species-level identifications. Twenty individuals/family/m^2 kick, where possible, were randomly selected for barcoding; if fewer than 20 specimens occurred for a given family in a given m^2 kick, all specimens were selected for barcoding. Twenty specimens/family/kick was deemed sufficient on the basis that barcoding 20 specimens/family/stream will return 70% of the species diversity/stream for caddisfly larvae (Martin, 2013). Intensive specimen sampling at the kick level was combined with multiple kicks at the stream level (n=6-9) in order to capture close to the full species diversity at each stream. Note that 20 specimens/family/stream were selected for barcoding in the Algonquin dataset (Martin, 2013). Caddisfly specimens were sequenced for the barcode region of the mitochondrial cytochrome *c* oxidase subunit I gene (COI) following protocols established by the Canadian Centre for DNA Barcoding at the Biodiversity Institute of Ontario (BIO), University of Guelph. Most of the molecular work was performed at the Biodiversity Institute of Ontario through a pipeline specifically designed for insects (see Appendix I for more details). Photographs and specimen information will be made publically available in several online projects on the Barcode of Life Data System (BOLD,

Ratnasingham and Hebert 2007). Projects: Lotic Trichoptera Larvae of Churchill-Summer 2012, LEPTO; dataset of Gillian Martin Trichoptera of Algonquin Park, DS-GMTAP; Lotic Trichoptera Larvae of Southern Pennsylvania-2013, LTLSP). Species were defined according to molecular operational taxonomic units (MOTUs) based on the sequence data. Delineations were done on the basis of Barcode Index Numbers (BINs; Ratnasingham and Hebert 2013), which are assigned to specimens using an algorithm that clusters similar sequences within the Barcode of Life Data System.

Though sequencing success was generally very high (94% of 2312 attempted specimens), morphological IDs were used in cases where sequencing success was <85% for a given m^2 kick. A small level of sequence failure/m^2 kick was accepted in order to address higher incidences of sequence failure, which was facilitated by establishing an arbitrary lower acceptable limit of 85% success/m^2 kick. Whenever success was above the acceptable level, densities were calculated with the available data. Cases of excessive failures (<85% success) were addressed according to the amount of diversity sampled with the barcoding data. Failed specimens belonging to widespread single-species families were morphologically examined to confirm membership and recorded as that single species. Another possibility was the occurrence of widespread low-diversity families with distinguishable single-species genera. Failed specimens in these cases were assigned to species according to genus membership (Morse and Holzenthal 2008). For cases of failures in species-rich families (more than 1 species/genus), barcoding was attempted a second time using the same procedure as before, which generally brought sequence success up to acceptable levels ($\geq$85% return).

Missed species estimations

To estimate undersampling (i.e. missed species), historical records were examined for species occurring in the sampled regions but not encountered during sampling for this project. For Churchill, BOLD projects were searched for samples from stream environments (1735 records for Trichoptera larvae throughout Churchill 2007-2010; Ruiter et al. 2013). Missing species in Churchill were considered species caught as larvae from streams during the months of July and early August (within the first week, near the times sampling was performed for this project). For Algonquin, because more streams were sampled than were used in the present analysis (see Data Analysis: additive

diversity components), species excluded from the analysis were counted as missing species; the regional checklist for Algonquin caddisfly species is rather limited, with approximately 23 species recorded in the literature prior to 2000, and so the estimated regional species pool was based entirely on the 2011 Algonquin dataset (Martin 2013). For Pennsylvania, no relevant summer records were found on BOLD; however, 700 historical records were accessed through the Pennsylvania Aquatic Insects webpage (Wilcox and Hagenbuch 2010). Historical records were searched for the 3 counties sampled (Chester, Berks and Schuylkill) and 10 neighbouring counties (Delaware, Montgomery, Lehigh, Carbon, Luzerne, Columbia, Northumberland, Dauphin, Lebanon and Lancaster); a species was considered missing if caught in one of these counties any time after 1960 (most records date to before 2000) during the months of June, July, and August. The timeframe for Pennsylvania species was widened towards later summer (i.e. July, August) given the historical records were for adults caught with light traps deployed near streams (not larvae as in Churchill and Algonquin) and would therefore have been caught as adults later in the summer compared to the larvae. Note that progression through the summer season differs in the sampled regions. Seasonal conditions (i.e. temperature) for July in Churchill are comparable to June in Pennsylvania; presumably species succession therefore follows a month behind in Churchill compared to Pennsylvania. The full species lists are presented as an appendix (Appendix II).

EstimateS 9.1 (Colwell 2013) was also used to estimate the missed species component based on specimens sampled in this study using diversity estimators for Churchill and Pennsylvania; the estimation of regional diversity for Algonquin was provided by Martin (2013). Individual (Chao1) and sample-based (i.e. stream; Chao2) estimators of total diversity were calculated based on 100 randomizations of the input abundance data (Colwell et al. 2012). The mean number of species estimated after all individuals or samples were resampled in the analysis was considered the total species pool for sampled streams within regions and compared with the historical values. The estimation curves are presented in Appendix III.

Data analysis: additive diversity components

In order to assess the distribution of species diversity over the sampled spatial scales, additive partition analyses were performed on species richness using PARTITION

3.0 (Crist et al. 2003, Veech and Crist 2009). Total diversity for the 3 regions was defined as:

North American caddisfly diversity $\gamma=\alpha_{kick} + \beta_{kick} + \beta_{transect} + \beta_{stream} + \beta_{sub\text{-}region} + \beta_{region}$

In addition to the species-level analyses, the North American caddisfly data were pooled at the genus and family levels for further partitions at coarser taxonomic levels. Partitions according to taxonomic resolution were performed for both Churchill and Pennsylvania data individually and combined; Algonquin data were not used in the analysis of taxonomic resolution due to missing kick and transect spatial levels. Analyses performed using Algonquin data, however, are presented in Appendix IV (exclusion of Algonquin did not change the results or interpretation). In order to assess departures from null distributions, PARTITION 3.0 runs iterations of the input data, randomly allocating individuals to each m^2 kick (individual-based randomization), which simulates distributions of the sampled specimens according to random chance. Observed abundances of species are therefore retained in the null iterations and can be used to partition diversity according to different indices (such as Shannons or Simpsons diversity); here, species richness is used to estimate diversity at each spatial level. The null distribution is used to assess significance of departures (greater or less) in the observed values (represented as a p-value). Five thousand iterations were run for each partition in the calculation of the null distributions. Standard errors for observed α component estimations were also calculated from the raw data.

Due to Algonquin having a large number of streams unevenly distributed among sub-regions (Fig. 2.2), diversity components were calculated for Algonquin data using 6 randomized combinations of 8 Algonquin streams (n=3, 3, and 2 streams for the 3 sub-regions, Table 2.1) and taking the average values of diversity for each spatial scale; all streams were used in at least 2 of the 6 partitions. The balance of the partition analysis (i.e. number and evenness of streams included in the analysis) had negligible impacts on results (See Appendix V for details). Six kick samples (out of 81) with no recorded caddisfly species were excluded from the analysis of Churchill species, causing a small level of unbalanced sample design at the kick level.

Goodness of fit tests, or G-tests, were used to determine if the distributions in the diversity components according to taxonomic resolution or region sampled were

significantly different. The test statistic G was calculated as 2*(Σ observed richness*ln[observed richness/expected richness]), and significance was determined using the chi-square distribution (Sokal and Rohlf 1981). In the analyses comparing regions, Churchill and Pennsylvania were arbitrarily selected to serve as the observed values to determine if Algonquin was significantly different from the other regions, while Pennsylvania was used as the observed value to determine if Churchill was significantly different from Pennsylvania. The opposite analyses (where Algonquin and Churchill are used as observed values) were also performed; this did not change the result. As well, α kick, β kick, and β transect were pooled for G-tests comparing Churchill and Pennsylvania to Algonquin, given these components correspond to the lowest spatial component (α stream) for Algonquin. In the analysis of taxonomic resolution, the lowest taxonomic level was used as the observed value (i.e. species in the comparison of species and genera distributions), given increased taxonomic resolution would provide more accurate estimates of the distribution of species-level diversity.

The proportion of environmental variation occurring at each spatial scale was quantified for each region in order to inform mechanisms structuring diversity components. Detailed methods and results for this analysis are presented in Appendix VI.

Results

DNA barcoding success

A total of 4 out of 129 m^2 kicks containing caddisfly larvae did not meet the criteria for acceptable specimen ID coverage (≥85% success): a single Churchill sample (family Hydroptilidae, R3-1 transect 1 kick 2, 11/16 specimens [69% success], Fig. 2) and 3 samples in Pennsylvania (family Glossosomatidae, R2-3 transect 2 kick 1, 3/5 specimens [60% success]; Glossosomatidae, R3-3 transect 1 kick 1, 4/5 specimens [80% success]; family Uenoidae, R3-3 transect 3 kick 2, 3/4 specimens [75% success], Fig 2.3A). Difficulties were also experienced when barcoding the caddisfly family Rhyacophilidae from Pennsylvania; sequence success was 26% (n=87 attempted). The issue was scrutinized and the failures were attributed to a single species, *Rhyacophila carolina;* these specimens were morphologically identified to species using Oliver and

Flint (1962), further corroborated by the single specimen that did return a sequence matching *R. carolina*.

Missed species components and diversity components at each region

A large percentage of the total caddisfly diversity historically recorded for each region was missed during sampling. This percentage was similar for Churchill and Algonquin (39 and 46%, respectively), while most of the diversity in Pennsylvania was recorded as missed species (64%). The recalculation of β region based on the full species list (61% with inclusion of missed species) shows a modest decrease in the estimate as a percentage compared to β region in the undersampled dataset (64% with exclusion of missed species) though the absolute values are very different (131.4 and 63.2 respectively; Appendix II, Table A2.1). Individual-based (Chao1) and sample-based (i.e. stream; Chao2) diversity estimators indicated the regional species pool at Churchill, Algonquin, and Pennsylvania was undersampled. Individual-based estimation at Churchill indicated γ is 29 (max of 37, min of 28) based on 1247 individuals, whereas sample-based estimator indicated γ is 40 (max of 78, min of 31) based on 9 streams (i.e. 30% missed, Fig. 2.4); neither of the estimator curves plateaued, indicating estimations of γ based on resampling of the dataset may themselves be underestimated. The sample-based estimator was closer to the estimated regional species pool based on historical data (n=46). For Pennsylvania, individual-based estimation indicated γ is 57 (max of 69, min of 54) based on 7027 individuals, whereas sample-based estimator indicated γ is 70 (max of 99, min of 60) based on 9 streams (i.e. 23% missed, Fig. 2.4); the sample-based estimator did appear to plateau. Algonquin regional diversity was estimated to be 67 species (i.e. 60% missed; Fig. 2.4; Martin 2013). See appendix III for Churchill and Pennsylvania estimation curves.

Excluding the missed species diversity components, the partitioning of diversity at Churchill and Pennsylvania produced similar observations (G_4=1.67, p=0.795), but Algonquin was different from these regions (G_2=9.02, p=0.011). All the regions had greater than expected sub-regional excess of species (β sub-region), corresponding to 43 and 45% of total diversity for Churchill and Pennsylvania and 57% of diversity at Algonquin (Fig. 2.5). Churchill and Pennsylvania also had greater than expected stream excess of species (β stream), corresponding to 23-27% of total respective diversity (Fig.

2.5); Algonquin had similar stream excess of species (24%), though this value was not different from the null expectation. The within-stream spatial levels had a significant shortage of species (Fig. 2.5), with the α stream components (α stream for Algonquin, α kick+β kick+β transect for Churchill and Pennsylvania) corresponding to 30-32% of total respective diversity for Churchill and Pennsylvania and 19% for Algonquin. Values for diversity components according to region are presented in Appendix VII.

Diversity components according to taxonomic level

The distribution of diversity components differed according to taxonomic resolution. In the large-scale analysis (combining Churchill and Pennsylvania), the distribution in diversity components appeared to be different for species and genus taxonomic levels, though the differences only approached significance (G_5=10.84, p=0.055; Fig. 2.6). The distribution in diversity components was different between species and family levels (G_5=50.96, p=<0.001) and different between genus and family levels (G_5=16.99, p=<0.005; Fig. 2.6). Species, genus, and family-level partitions all had greater than expected regional excess of respective taxonomic units, but the percentage of regional-level taxon excess decreased with coarser taxonomic resolution: β region for species is 49%, 42% for genera, and 23% for families (Fig. 2.6). Each taxonomic level also had significant sub-regional excess of species, but the decrease in β with coarser taxonomic resolution was not as pronounced as β region (β sub-region is 22% for species, 16% for genera, and 17% for families; Fig. 2.6). β stream components marked a transition from significantly small to significantly large values; β stream for species was significantly small at 12%, β stream for genera was 16% and not different from the null expectation, while β stream for families was significantly large at 19%. Diversity components at the α stream level and lower (e.g. α and β kick and transect levels) were smaller than expected by chance, though these components increased with coarser taxonomic resolution (as in the β stream component, Fig. 2.6).

The importance of taxonomic resolution differed at the regional level. The distribution of diversity components at Churchill differed between species and genus levels (G_4=11.2, p=0.024) and species and family levels (G_4=36.7, p=<0.001), but did not differ between genus and family levels (G_4=7.7, p=0.103; Fig. 2.7). Similar to Churchill, the distribution of diversity components at Pennsylvania differed between species and

genus levels (G_4=16.8, p=0.002) and species and family levels (G_4=24.7, p=<0.001), but did not differ between genus and family levels (G_4=0.95, p=0.917; Fig. 2.7). Diversity components for the two regions also did not differ at the genus level (G_4=3.14, p=0.535) or at the family level (G_4=4.0, p=0.406). Sub-regional excess of species was always greater than expected by chance and decreased with coarser taxonomic levels in both Churchill and Pennsylvania. Churchill β sub-region was 43% for species, 29% for genera, and 19% for families; Pennsylvania β sub-region was 45% for species, 27% for genera, and 24% for families (Fig. 2.7). Stream excess of species was also greater than expected for Churchill and Pennsylvania for each taxonomic level, but did not appear to decrease with coarser taxonomic resolution; β stream for Churchill was 27-29%, while Pennsylvania β stream was 23-25% (Fig. 2.7). Within-stream β components (β transect and β kick) were less than expected by chance for both Churchill and Pennsylvania and appeared to increase modestly with coarser taxonomic resolution. Absolute values for diversity components according to taxonomy are presented in Appendix VII.

Environmental variation according to spatial scale

Environmental variation generally occurred at the transect and stream spatial levels, but not at a sub-regional level; however, Pennsylvania differed slightly in that environmental variation did occur at the sub-regional level. See appendix VI for more details.

Discussion

Additive partitioning of species richness is routinely used to study the distribution of biodiversity according to spatial scale. Unclear, however, is whether or not diversity components are structured differently in disparate habitats, particularly along a large latitudinal gradient. These assessments of species diversity, however, must be reconciled with the growing recognition that a large portion of species diversity remains undetected with traditional morphological identifications (Hebert et al. 2003a, b, Hamilton et al. 2010, Mora et al. 2011). Consequently, many estimations of γ, α, and β diversity are likely inaccurate. Here, DNA barcoding has been employed for species identification of caddisflies given its ability to make consistent species identifications, including the

detection of unknown species. Caddisfly species diversity was also partitioned across small (m^2) to large spatial extents (19^o of latitude).

Missed species in diversity components analysis

My first objective was to assess whether or not a large percentage of total diversity γ will remain undetected in my study despite intense sampling efforts and the use of molecular techniques to assist species identification. Here I attempted to define the regional species pool as those species available at the time of sampling and that had the potential to be included in my sampling scheme (e.g. 1^{st}-2^{nd} order streams during mid-summer). I found that 39-64% of the total diversity based on historical records from each of the sampled regions remained undetected. In addition, sample-based diversity estimators indicated a large percentage of species were missed in the sampling of 9 streams for each region; 30% of species were missed in Churchill, 23% of species were missed in Pennsylvania and 60% of species were missed in Algonquin (Fig. 2.4). The varied estimations of missed diversity for each region reflects differences in sampling effort; Pennsylvania was the only region where caddisfly larvae were specifically targeted. As such, characterization of the regional species pool is likely less accurate and underestimated for Algonquin and Churchill where caddisfly larvae were not specifically targeted during sampling, particularly for Algonquin where sampling was not replicated within streams. Individual-based estimation of diversity, however, indicated only a small percentage of diversity was missed within the sampled streams for Churchill and Pennsylvania (3% and 5%, respectively). The individual-based estimations of the missed species indicate species within the sampled streams were comprehensively sampled. These results highlight an important trade-off in the sampling of species; habitats may be thoroughly sampled, but at the expense of larger-scale sub-regional replication of sites. In the case of stream invertebrates, variation in species assemblages among riffles and stream reaches has been shown to be low, meaning regional diversity is characterized most efficiently by sampling multiple stream sites and a variety of substratum types (Ligiero et al. 2010).

Temporal constraints to sampling are also an important consideration when defining the regional species pool. Estimations of the missed species component based on historical records was considerably larger than sample-based estimations; this was

especially true for Pennsylvania where historical records indicated 64% of species were missed while resampling of the dataset estimated this figure to be 23%. The discrepancy between the two values may be due to temporal variation in stream insect community structure, which is high between years (Lenat and Resh 2001). Because historical records incorporate several years of sampling, this may overestimate the actual number of available species at the time of sampling; on the other hand, historical records serve to better characterize the full species pool through time, which may be useful if regions are repeatedly sampled. Diversity components analyzed across regional scales need to incorporate sampling efforts adjusted to optimize thorough detection of species diversity; temporal constraints to sampling will also have important implications for how the regional species pool is explicitly defined.

The missing species component in the partitions of the 3 sampled regions poses the questions of whether the missed species should be included in the observed distribution and how this redistribution would occur. Beck et al. (2013), for instance, show that severe undersampling in assessments of β diversity leads to larger estimates of β and decreased precision in measurements (the results are more likely to be affected by random chance). The β diversity components measured for caddisflies in this study were overestimated, particularly at the larger spatial scales where replication was limited and species were most likely to be missed, but the overestimation was marginal when considered as a percentage of total diversity; β region was 64% for the sampled species list and 61% when missed species were included, but the absolute values were 63.2 and 131.4, respectively. Beck et al. (2013) do note that while absolute values of diversity components are heavily impacted by undersampling, the interpretation of components using null models remains robust. The results of this study therefore show that undersampling can be severe when regional sampling effort is not optimized and not sustained over time, and that absolute values of additive diversity components should be treated with skepticism.

Several measures may be taken to alleviate the overestimation of β diversity when species diversity is undersampled. Since rare species are most likely to be missed during sampling, de-emphasizing rare species in the calculation of diversity components may decrease overestimation of β (Beck et al. 2013). Shannons or Simpsons diversity indices

do this by factoring in abundance data, such that common species (i.e. species most likely to be sampled) contribute most to the diversity index. These metrics, however, fundamentally assume the importance of a species is approximated by its abundance relative to other species. Rare species could promote functional redundancy and therefore ecosystem stability (Hooper et al. 2005), and, depending on the context and biological question, are important since they are products of evolution over vast geological time (Oksanen 1997). In addition, density or abundance data are not always available or easily obtainable in a given study. Though not providing a complete solution yet for all taxa, as observed here in the taxonomically patterned sequencing failures, DNA barcoding provides a tangible step in the right direction for analyses of diversity components by (1) detecting more unknown species and decreasing the missed species component and (2) explicitly quantifying the regional species pool (γ) independent of sampling intensity within a single study through the Barcode of Life Data System (BOLD, Ratnasingham and Hebert 2007, Smith et al. 2008, Sweeney et al. 2011).

Large-scale diversity components of stream caddisflies

My second objective was to determine if diversity components for each region along the sampled latitudinal gradient were similar or different. I found caddisfly species richness was partitioned similarly for the sub-arctic (Churchill) and temperate North America (Pennsylvania), but was different for Algonquin Provincial Park (Fig. 2.5). One reason why Algonquin was significantly different may be due to differences in sampling effort as compared to Churchill and Pennsylvania, which were more thoroughly sampled at the within-stream levels. The larger estimates of β diversity at the sub-regional level and lower α stream components at Algonquin may be partially explained by undersampling; species missed in some sub-regions or streams but detected in others would artificially inflate estimates of β diversity. If the values at Algonquin were not obscured by differences in sampling effort compared to the other regions, then it is possible sub-regional or stream-level processes differ in Algonquin; environmental variation or dispersal limitations are possible avenues for further investigation. Note that environmental variation in Algonquin was relatively low at the sub-regional level compared to the other spatial scales, but high at the stream level, indicating the possibility

that environmental variation may have driven β stream diversity (Appendix VI Fig. A6.2).

In contrast to the differences at Algonquin, Churchill and Pennsylvania showed similar distributions of diversity components (Fig. 2.5). These results suggest that β diversity can be structured similarly at very different latitudes. Evidently, this conclusion must be considered in the context of minimal regional and latitudinal replication, and significant differences at Algonquin, and is therefore difficult to generalize beyond Churchill and Pennsylvania. It should be noted that nearly all studies of diversity components report smaller than expected α diversity and larger than expected β diversity, particularly at large spatial scales; this will almost always be the case due to the additive nature of diversity components (Wilsey 2010). Additive diversity components from different latitudes will follow the same tendency; however, I have demonstrated that the β components sampled at Churchill (n=28 species) and Pennsylvania (n=54 species) exceed α components by the same percentages. Two possibilities exist to explain the similarities between Churchill and Pennsylvania: either the caddisfly species diversity at both regions is structured by similar processes, or different processes produced the same outcome. According to results for the importance of environmental variation driving β diversity from both regions, it would seem environmental variation played a larger role at the sub-regional level in Pennsylvania compared to Churchill, likely due to the sampling of a northern hilly region in Pennsylvania (Appendix VI, Fig. A6.2). Further analyses seeking to explain variation in species diversity at the stream and sub-regional level are needed to confirm if Churchill and Pennsylvania are similarly structured due to similar or different processes.

An interesting implication of the consistent structure in diversity components is the application of the observed caddisfly species diversity structure to new regions; if the total diversity is roughly known in a new location (with an approximately 70 km spatial extent), one may be able to reasonably estimate the species diversity at various spatial levels. If generalizations such as this do exist, then conservation efforts could readily standardize protocols for very different regions, protecting diversity according to spatial levels that promote heterogeneity in the distribution of species and maximize total diversity. These generalizations should emerge as additive partitioning expands to more

taxa and more studies that include large spatial extents spanning many degrees of latitude.

The importance of taxonomic resolution

My third objective was to investigate the importance of taxonomic resolution in the additive partitioning of species diversity, which to my knowledge has not been addressed. Taxonomic resolution was important; in the full-scale analysis (combining Churchill and Pennsylvania) the distribution in diversity components according to species and genus-level data was significantly different from family-level data, though the differences between species and genus-level results only approached significance (Fig. 2.6). As well, the individual analyses of Churchill and Pennsylvania indicated that species-level data differed from genus and family-level data, which produced similar distributions in diversity components (Fig. 2.7). The results for Churchill and Pennsylvania analysed independently are particularly relevant to studies of additive diversity components given many cover regional spatial extents (<250 km), and a large portion experience difficulties achieving full species resolution datasets for difficult to ID taxa (Fig. 2.1). This means biases in reported distributions in diversity components due to taxonomic resolution are likely frequent in the current literature. For regional datasets, diversity components expressed as percentages of total diversity are likely underestimated at larger spatial extents (i.e. sub-regional) while small spatial scale components are overestimated (i.e. kick and transect components; Fig. 2.7).

Interestingly, while β diversity at large spatial scales (region and sub-region) decreased with coarser taxonomic resolution, undersampling likely led to some overestimation of these values. The direction and magnitude of the actual bias in diversity components is therefore less clear when both issues are present. Undersampling would, however, be somewhat relieved with decreasing taxonomic resolution given the better chance of sampling and successfully identifying taxa. Absolute values for diversity components could, nonetheless, be highly inaccurate when studies are marred by unintended coarse taxonomic resolution and undersampling. DNA-based species identification helps alleviate coarse taxonomic resolution and undersampling simultaneously by delineating species in an objective and consistent manner because morphological information is not needed to make the identification. I forward DNA-

barcoding, in particular, as a viable stepping-stone towards truer estimates of diversity components.

Mechanisms structuring diversity components

Once patterns in the distribution of biodiversity are established using diversity components, informed hypotheses can be forwarded about mechanisms driving species distributions. As indicated above, diversity components at each region sampled appeared to be driven by environmental variation at the stream level within each region and at the sub-regional level in Pennsylvania (Appendix VI, Fig.A6.2). The importance of dispersal limitations, which were not quantified for the sampled caddisflies, should also be investigated at the sub-regional level given the large sub-regional β component generally does not correspond with important environmental variation. The partitioning of species diversity according to the full dataset gave opportunity to assess β at a very large spatial scale (i.e. β region). The partitioning of species diversity at the largest spatial scale (~2500 km between Churchill and Pennsylvania), not surprisingly, indicated a large excess of species at the regional level (i.e. greater than expected β region component, Fig. 2.6). The excessively large β region diversity can be attributed to several factors, including latitudinal restrictions in species ranges due to large environmental differences and dispersal limitations between regions (i.e. sub-arctic vs. temperate; Appendix VI, Fig.A6.1) and evolutionary processes operating over large spatial scales such as speciation, which would further differentiate the regional species pools. Because the regions were sampled in different years, between-year differences in species occurrences may have biased β region upward if some species that inhabit both regions were present in only one year (Lenat and Resh 2001). Overlap in species between regions, however, suggests they were sampled at comparable times; species overlap was the same for both sampled and historical species lists, at 13% overlap between Algonquin and Pennsylvania, 2% between Algonquin and Churchill and 1% between all regions (a single species, *Hydropsyche bronta*; Appendix II).

The analyses according to taxonomic resolution also showed interesting trends in how diversity components are structured at regional spatial extents. Increased taxonomic resolution clearly decreased the percentage of total diversity found at small spatial scales

(α stream, α and β transect and kick) while increasing estimates of these components at large spatial scales (β sub-region and region). In the full-scale analysis (combining Churchill and Pennsylvania) the β stream component also showed a clear transition from a significantly small value at the species level to a significantly large value at the family level. Alpha components may have decreased with increasing taxonomic resolution due to the fact that species geographic ranges will be restricted relative to families; this means the same families will tend to consistently reoccur in a set of samples, whereas a single species may only appear in a subset of geographically close samples. Niche overlap may also play an important role in explaining these trends if niche similarity shows a phylogenetic signal (Webb 2000, Webb et al. 2002); because closely related species are functionally similar, competitive interactions may prevent co-existence in the same habitat. In the context of regional diversity, this may result in low α diversity for species compared to families, which may co-exist due to functional differentiation. Along the same lines, values of β diversity at spatial scales with important environmental variation may switch from being significantly small to significantly large going from species to family-level data. This switch in significance occurred at the stream level in the full-scale partition analysis, possibly due to stream insect families being more spatially structured due to environmental variation than species; stream insect species may be less structured by environmental variation due to redundancy in environmental preferences (Martin 2013).

On a final note, the spatial extent of the species assemblages analysed was important in the interpretation of diversity components. For instance, while the β sub-region component is significantly small in the partitioning of diversity using the full data set (Fig. 2.6), it is significantly large when considering the regions individually (Fig. 2.5 and 2.7). This difference illustrates an important point; whenever a diversity component is considered larger or smaller, one must remember the point of comparison. In the context of large-scale distributions of biodiversity, β sub-region is relatively small and unimportant, but upon considering a region on its own, it becomes the main driver of β diversity. The interpretation of additive partitions is highly sensitive to the spatial context of the analysis. As well, implications for results regarding the importance of taxonomic resolution differed according to spatial extent of analysis; the differences between species

and coarser taxonomic resolution were more pronounced at regional spatial extents. While additive partitioning of species diversity is a versatile tool for describing distributions in diversity components and informing hypotheses about mechanisms driving biodiversity, the spatial extent under consideration is a critical component to the interpretation of these patterns.

Conclusions

Diversity components facilitate the study of the distribution of biodiversity across spatial scales, informing hypotheses about mechanisms driving species diversity by first establishing the patterns and spatial scales of interest. This study has demonstrated that diversity components can be structured very similarly at disparate regions, particularly from vastly different latitudes; further investigations may reveal a level of consistency that could be used in conservation efforts that seek to maximize the protection of diversity in new regions. Future studies should investigate whether or not this consistency, if true, is taxon specific. In addition, DNA-based species identification (i.e. DNA barcoding) can help move diversity components studies forward in a constructive manner by alleviating biases due to undersampling and coarse taxonomic resolution, given DNA barcoding can appropriately handle understudied and difficult to ID taxa. DNA-based specimen identification will, however, be contingent on refined molecular methods, particularly developing more reliable primer cocktails for DNA amplification and sequencing across taxa. The cost to sequence specimens, though becoming more economic (Stein et al. 2014), may also present limitations; next-generation sequencing is a promising next step towards the cost and time effective generation of species richness data without identifying specimens individually (Hajibabaei et al. 2011, Shokralla et al. 2012). Datasets for diversity components should be made available; as more data are produced, meta-analyses will become necessary if we are to elucidate general rules for how biodiversity is maintained and distributed.

Table 2.1. Spatial scales used to assess diversity components of stream caddisflies (order: Trichoptera). Note: number of replicates represents the numbers used in analyses. Algonquin has 1 fewer stream in its analysis and does not have transect or kick-level data. 75 kicks were taken at Churchill (3/transect, 6 excluded due to no caddisflies), and 54 kicks were taken at Pennsylvania (2/transect).

Spatial scale	**Spatial extent of replicates**	**# of replicates**	**Biological significance**
Region	2500 km	3	Habitat zones spanning 19^{o} of latitude, i.e. biogeographic processes.
Sub-region	25-100 km	9 (3/region)	Within-region habitat differences and dispersal limitation
Stream	10 km	26 (2-3/sub-region)	Between-stream environmental differences
Transect	20-50 m	54 (3/stream)	Within-stream habitat differences (i.e. riffles and pools)
m^2 kick	5 m	129 (2-3/transect)	Smallest spatial scale, potential for biotic interactions

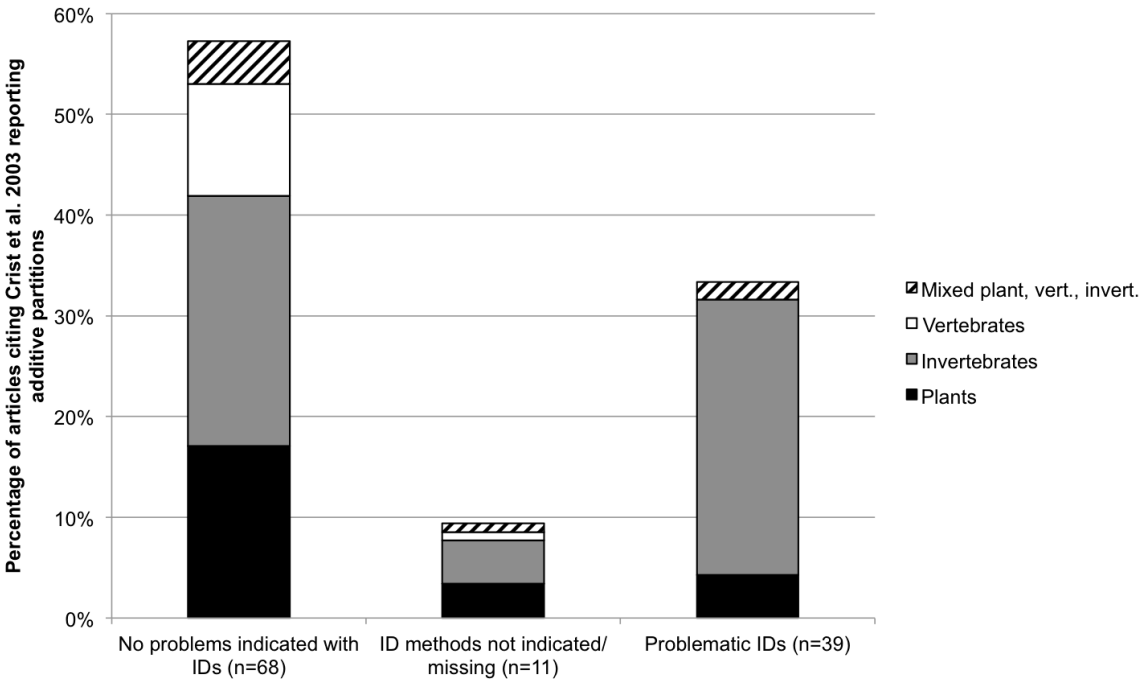

Figure 2.1. Summary of 118 articles citing Crist et al. 2003 and reporting estimates of additive diversity components. Articles with problematic identification of specimens include those reporting mixed taxonomic level datasets (typically mixed species and genus), specimens excluded from the dataset due to inability to morphologically ID (typically invertebrate juveniles), and the assignment of morpho-species (specimens that could not be assigned to a recognized species name).

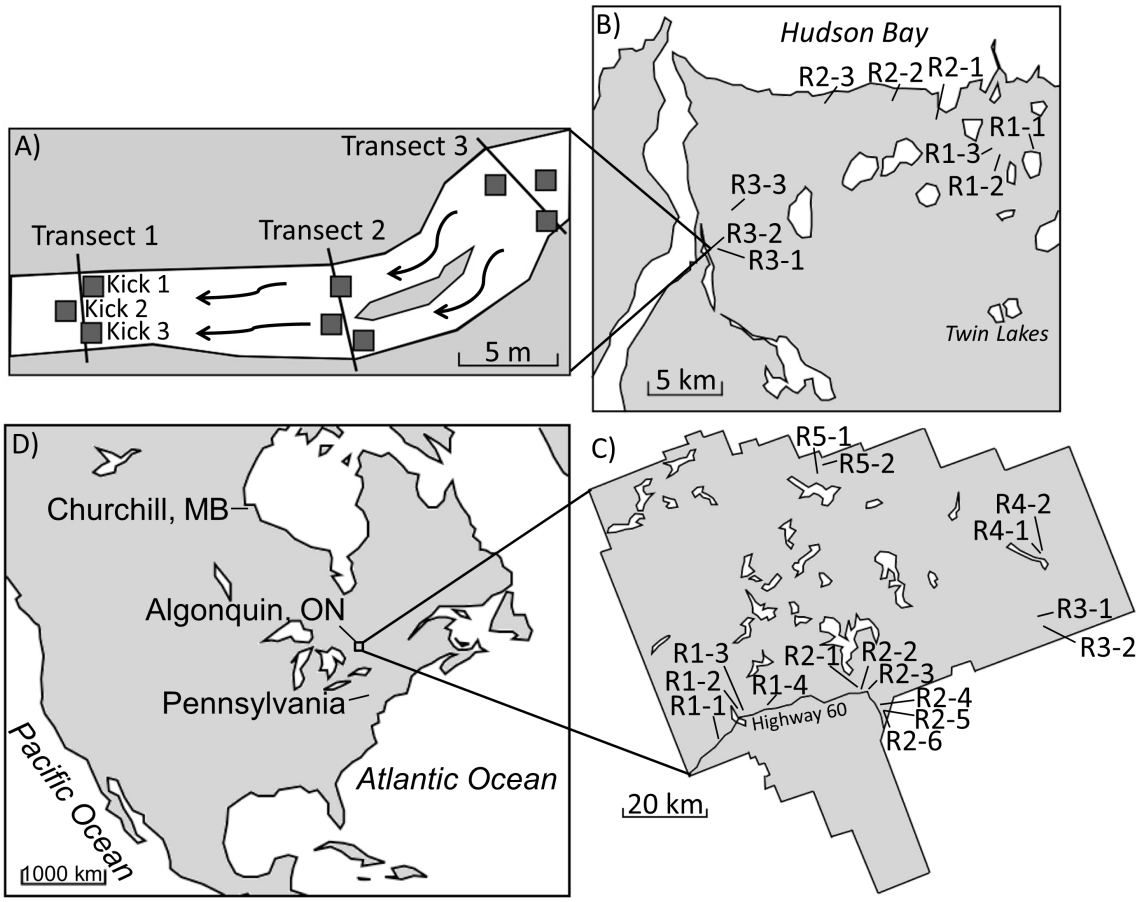

Figure 2.2. Field sampling design, from large to small spatial scales. A) Sampling design within streams, where each square represents a m2 kick. The arrows represent the direction of stream flow. B) Churchill (MB, Canada), where 9 streams were sampled (July 15-23, 2012). C) Algonquin Provincial Park (ON, Canada), where 16 streams were sampled (June 26-July 13, 2011). D) 3 major regions where the nested sampling design was implemented. R=sub-region (1, 2, 3), followed by stream number sampled within sub-region (1, 2, 3). Note: Algonquin has an unbalanced design.

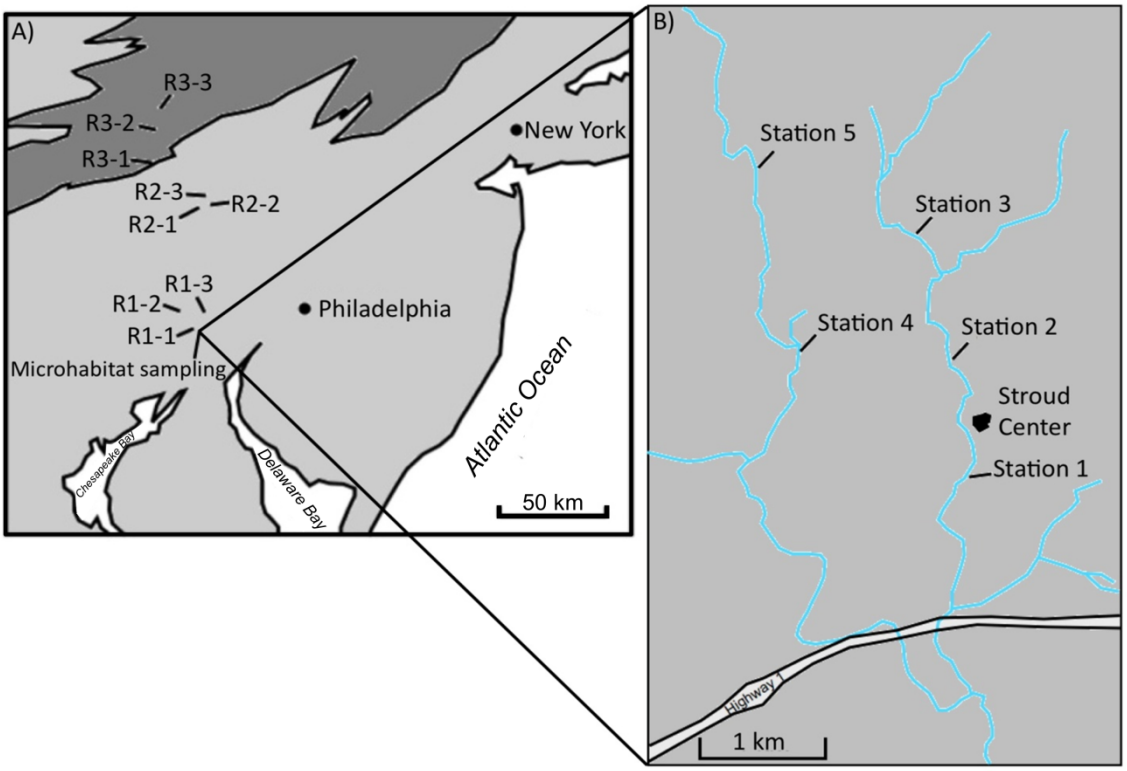

Figure 2.3. Field sampling design implemented in Pennsylvania, USA (June 17-20, 2013). A) Stream locations where R=sub-region (1, 2, 3), followed by stream number sampled within region (1, 2, 3). The darker grey represents increased elevation, i.e. hilly regions. B) White Clay Creek watershed and stations where microhabitats (i.e. rocks) were sampled (June 20, 2013).

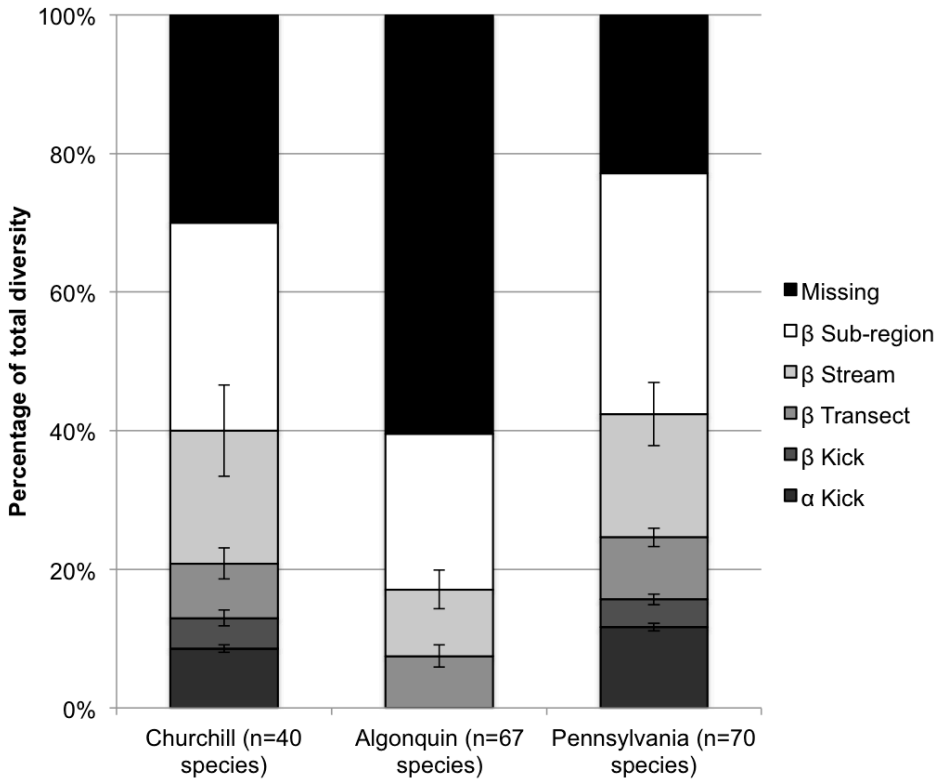

Figure 2.4. Additive partitions of summer caddisfly (order: Trichoptera) species at 3 regions, including species missed during sampling. Sampled regions are Churchill (MB Canada; July 15-23, 2012), Algonquin Provincial Park (ON, Canada; June 26-July 13, 2011), and Pennsylvania (USA; June 17-19, 2013). The missed component is based on sample-based (e.g. streams) diversity estimators. Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Note: the Algonquin partition does not include within-stream spatial levels (i.e. transect and kick levels).

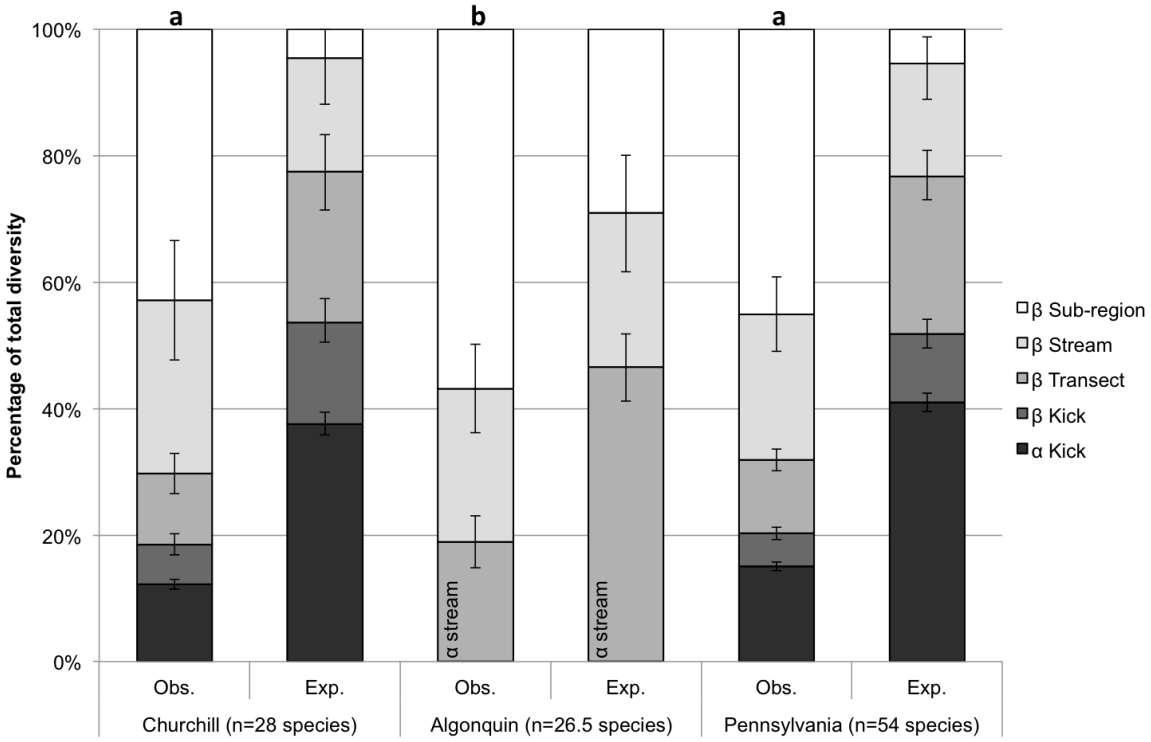

Figure 2.5. Additive partitions of sampled summer caddisfly (order: Trichoptera) species at 3 regions. Sampled regions are Churchill (MB Canada; July 15-23, 2012), Algonquin Provincial Park (ON, Canada; June 26-July 13, 2011), and Pennsylvania (USA; June 17-19, 2013). Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Error bars for expected values are max and min estimations from the null model. Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. Note: the Algonquin partition does not include within-stream spatial levels (i.e. transect and kick levels). Letters above observed diversity components indicate statistically similar or different distributions (omitting within stream spatial scales for comparisons with Algonquin). See supplementary material (Appendix VII) for raw values of α and β.

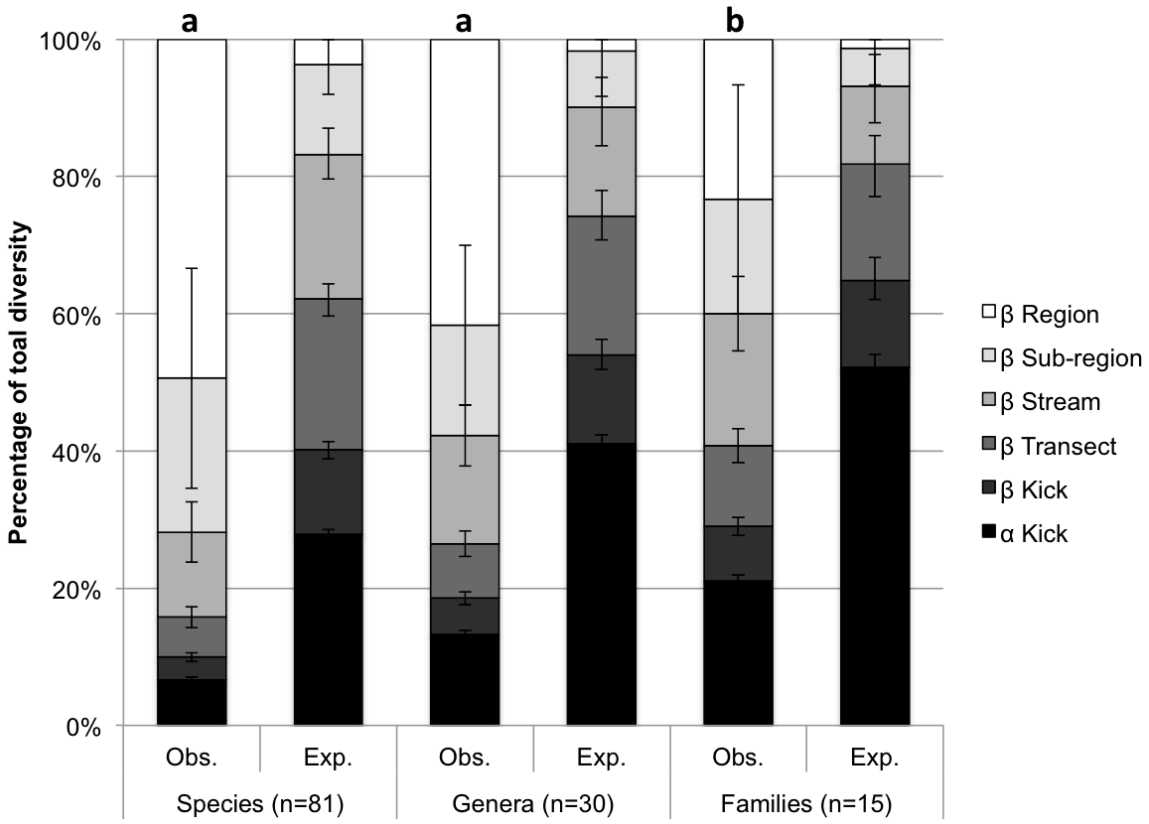

Figure 2.6. Additive partitions of sampled summer caddisflies (order: Trichoptera) according to taxonomic level. Regions are Churchill (MB, Canada) and Pennsylvania (USA; 2012-13). Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Error bars for expected values are based on max and min estimations from the null model. Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. Letters above observed diversity components indicate statistically similar or different distributions. See supplementary material (Appendix VII) for raw values of α and β.

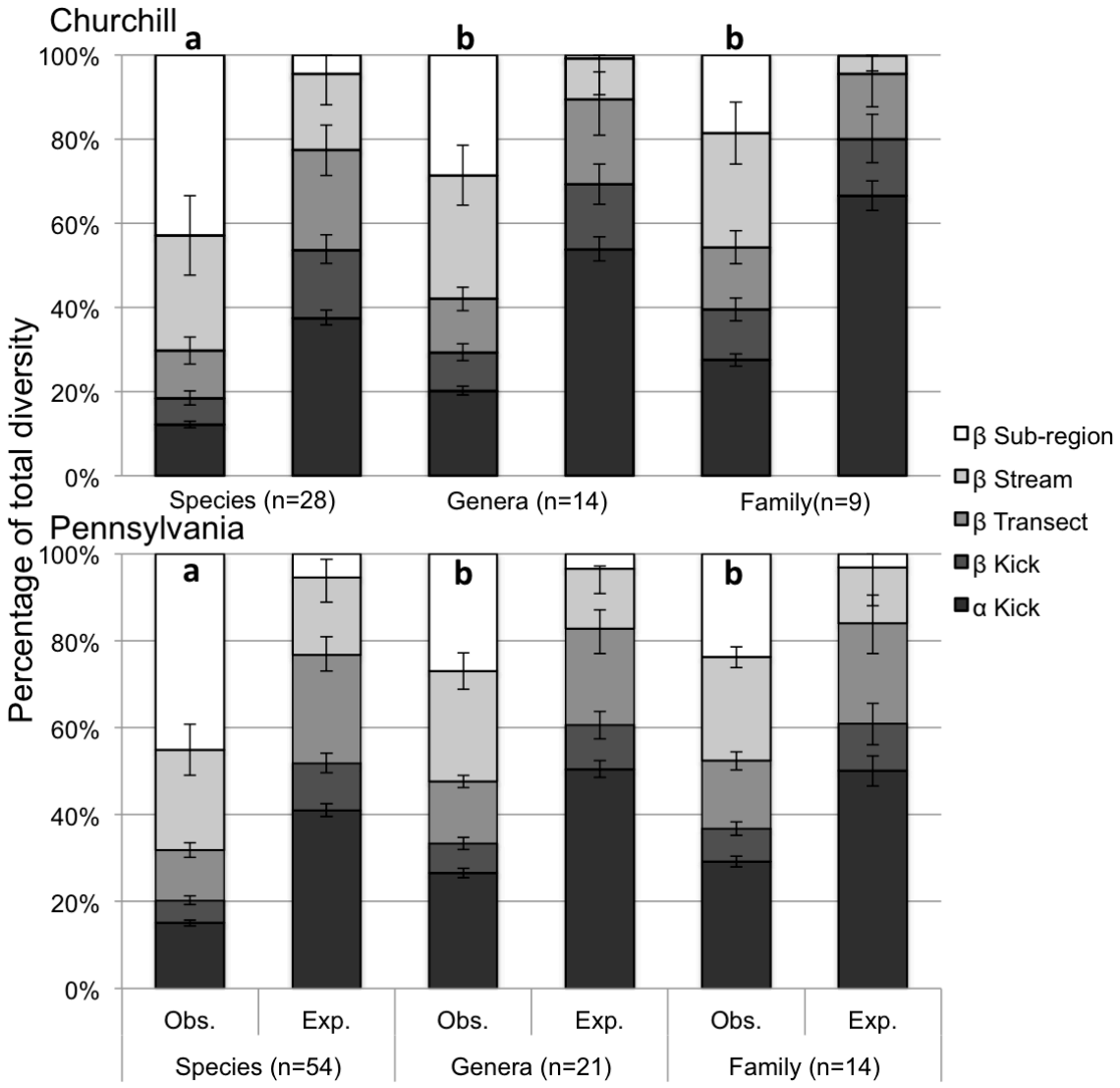

Figure 2.7. Additive partitions of sampled summer caddisflies (order: Trichoptera) according to taxonomic level for Churchill (MB Canada; July 15-23, 2012) and Pennsylvania (USA; June 17-19, 2013) individually. Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Error bars for expected values are based on max and min estimations from the null model. Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. Letters above observed diversity components indicate statistically similar or different distributions (within the regions). See supplementary material (Appendix VII) for raw values of α and β.

Chapter 3: Investigating the importance of competition in assemblages of stream insect species

Abstract

Competitive interactions are assumed to play a role in governing species distributions, but are difficult to infer using large-scale observational data. The objective of this chapter was to determine if stream insect species diversity is non-randomly distributed at small spatial extents, where environmental filtering and dispersal limitation become relatively unimportant, and in turn test for competitive interactions amongst specific taxa. In order to do this, rocks from riffles were sampled for stream insects, in particular caddisfly (order Trichoptera), beetle (order Coleoptera), and black fly (family Simuliidae) larvae at White Clay Creek (PA, USA; June 20, 2013). Additive partitioning of species diversity was used to assess variation in the distribution of species according to spatial scale, which informed checkerboard analyses that tested for signatures of competitive interactions among specific taxa. Only a small portion of the total species diversity occurred at the smallest spatial level (α rock of 29-39%), indicating species are unevenly distributed within White Clay Creek. The sampled taxa were also distributed differently, with species of Trichoptera having increased distributional variation of species at the scale of kilometers (β station and arm was 46% of total diversity) compared to species of Coleoptera which were primarily structured at the scale of meters (β rock and transect was 39% of total diversity). For sites spanning White Clay Creek (kilometer scale), significant segregation of species (i.e. high C-score) occurred within members of Trichoptera and Hydropsychidae, with the strongest segregation occurring between Hydropsychidae and Simuliidae. This study demonstrates stream insect species are non-randomly distributed at small spatial scales and that the distribution of species can be linked to competitive interactions.

Introduction

Diversity components analysis is used to assess the distribution of biodiversity across spatial scales, informing further investigations that seek to explain patterns in the distribution of species (Crist et al. 2003, Veech 2005; see Chapter 2). While diversity components are often used to assess variation in the distribution of species at regional scales, very localized assessments are far less common. Of the few studies that additively partition diversity over a very localized spatial extent (<5 km), β diversity at various spatial levels is often greater than expected by chance, meaning diversity remains non-randomly distributed even at small spatial extents (marine studies: De Troch et al. 2008, Gheerardyn et al. 2010, Ligeiro et al. 2010, Rodríguez-Zaragoza et al. 2011; terrestrial: Negro et al. 2011). If species diversity remains structured at small spatial extents where environmental variation and dispersal limitations are minimized, then I propose here that the potential for biotic interactions becomes a viable hypothesis. Such small-scale approaches to studying the role of competition in species assemblages are needed; though ecologists are confident biotic interactions play some role in governing how species assemble, detecting these interactions using large-scale observational data has proven difficult.

The role of competition in the assemblage of species has a long history of conceptualization. Charles Darwin (1895) first recognized competition was likely to be stronger between congeneric species sharing similar habitats and adaptations as compared to more distantly related (and functionally dissimilar) species. Elton (1946) formalized Darwin's hypothesis by showing that, in a given habitat, a high percentage of genera over a variety of taxa are represented by a single species, suggesting competitive exclusion between closely related species indeed occurs. Hardin (1960) dubbed this the "competitive exclusion principle," stating that if (i) 2 non-interbreeding populations occupy the same ecological niche and (ii) the 2 populations are sympatric and (iii) one population multiplies slightly faster than the other, then the population that multiplies the fastest ultimately drives the other to extinction. The competitive exclusion principle has since served as the basis to more modern studies and analyses that investigate the role of competitive interactions driving patterns in biodiversity. Checkerboard analyses, in particular, infer the importance of competition in species assemblages based on the

prediction that certain combinations of species are "forbidden" in a site due to interspecific competition (Diamond 1975, Gotelli and McCabe 2002). If species tend to not co-occur, i.e. "segregate" according to Diamond (1975), this segregation should be reflected as a mutual tendency for two species not to co-occur, or a "checkerboard pattern," in a site-by-species matrix when sampling multiple sites (Stone and Roberts 1990). The inference of competitive interactions is based on a higher than expected average number of checkerboards across species pairs, or the "C-score," in comparison to a null model that simulates random structure based on the input data.

Though checkerboard analyses are a promising tool for detecting signatures of competition, their use must be optimized in order to ensure the post hoc inference is not conflated with other processes. In particular, Diamond's (1975) community assembly rule has been contrasted with other explanations for species segregations such as random chance (Connor and Simberloff 1979, Ulritch 2004) and habitat heterogeneity (Schoener and Adler 1991). Today, researchers recognize checkerboards can be used to detect non-random assemblages of species (Gotelli and McCabe 2002), but the interpretation of non-random C-scores remains compromised by the possibility of species segregating due to environmental filtering and dispersal limitations; consequently, uncertainty in the interpretation of checkerboards remains high (Boschilia et al. 2008, Kamilar and Ledogar 2011).

I propose a novel approach to reduce the potential confounding effects of environmental filtering due to habitat heterogeneity and dispersal limitation when using observational data. Because the influence of both processes increases with spatial scale, sampling over a small spatial extent will reduce the likelihood of detecting their confounding effects (Willis and Whittaker 2002, Cavender-Bares et al. 2009). With these carefully considered sampling designs, environmental and dispersal factors will become relatively weak in driving species distributions compared to local competitive interactions, which facilitates the interpretation of the C-score. Note that the scale at which competitive interactions govern the assemblage of species will depend on the system (i.e. habitat, taxa) under investigation. The importance of controlling for these factors by sampling in consistent microhabitats has been recognized (e.g. sampling chironomids

from rocks in riffles; Heino 2005) and has been implemented at least once to detect diurnal patterns in competitive interactions for flying insects in a meadow (D'Amen et al. 2012).

Of particular interest for studying the importance of competition in the assemblage of species are stream insects. The families Hydropsychidae (net-spinning caddisflies) and Simuliidae (blackflies) are abundant filter feeders, attaching to stream rocks as aquatic larvae in fast-flowing streams where competition for filter-feeding locations is intense (Thorp 1983, Georgian and Thorp 1992). Hydropsychidae, however, are territorial with each other and other taxa for these feeding locations, exhibiting a range of aggressive behaviour ranging from stridulation as a warning mechanism to sometimes fighting to the death for retreats (Jansson and Vuoristo 1979). Simuliidae, on the other hand, will actively avoid filter-feeding sites occupied by Hydropsychidae (based on in situ experiments; Hempill 1988). Given Simuliidae are subordinate to the larger, more aggressive Hydropsychidae, these competitive interactions should result in exclusion of simuliid species within the filter-feeding locations (i.e. on rocks) when hydropsychids are present. Also of interest are Coleoptera, given many members have fully aquatic lifestyles that may maintain patterns in species distributions due to competition (Morse and Holzenthal 2008). In particular, diving beetles have been shown to displace other predators such as damselflies in mesocosm experiments, altering trophic cascades and ecosystem function (Atwood et al. 2014). Because previous studies have relied on lab based and in situ experimental evidence, an observational approach is needed to determine if these competitive interactions result in similar patterns in nature.

Before competitive interactions can be elucidated between specific stream insect taxa, however, spatial scales exhibiting variation in the distribution of species must be established. Additive diversity partitioning has been used previously to investigate small-scale variation in the distribution of stream insect families; β diversity amongst Surber samples has been shown to be greater than expected in gravel substrate, while β diversity among riffles and stream segments has also been shown to be greater than expected contributors to total diversity (Ligeiro et al. 2010). A species-level investigation of additive diversity components of specific taxa (i.e. Trichoptera, Coleoptera and

Simuliidae) is needed to reaffirm these patterns at higher taxonomic resolution, which can be followed up with checkerboard analyses at the appropriate spatial scale for testing for competitive interactions amongst stream insect species.

Objective and hypotheses

For Chapter 3, I first sought to establish at which spatial scales stream insect species diversity is structured using additive diversity components at White Clay Creek (PA, USA). This investigation in turn informed at which spatial scales and between which taxa checkerboard analyses should be performed in order to elucidate the details of competitive interactions at White Clay Creek. Because insects were sampled over a small spatial extent, environmental variation and dispersal limitations were assumed to not be important. In order to support this assumption, Table 3.1 presents maximum and minimum values for several environmental parameters at White Clay Creek compared to larger-scale sampling efforts (see Appendix VI for details on environmental data collection). Environmental parameters at White Clay Creek exhibited 1% to 67% of the total variability observed across the total number of streams sampled; hydraulic head (mm) and substrate 2 exhibited considerably more variation than other parameters at White Clay Creek (62 and 67% of total variation across 2500 km, respectively; Table 3.1).

My main research question is "are stream insect species at White Clay Creek spatially structured by competitive interactions, and if so, between which taxa do these interactions occur?" My first objective was to use additive diversity partitioning to establish at which spatial scale important variation in species diversity occurs for taxa anticipated to undergo competitive interactions (Trichoptera, Coleoptera and Simuliidae). The scale at which greatest variation in species distributions occurred (meters vs. kilometers) was then used as the spatial extent for checkerboard analyses testing for taxon-specific interactions.

My second objective was to investigate competitive interactions between specific taxa using checkerboards based on the results from additive partitioning of species diversity. I hypothesize that if competitive interactions structure species distributions at White Clay Creek, this will be reflected as significantly large C-scores for specific sets of species. In particular, the tendency for segregation of Hydropsychidae and Simuliidae

should be especially strong compared to segregation among species within either taxon; this is because, in a checkerboard analysis of Hydropsychidae and Simuliidae, hydropsychids are the dominant competitor, which will result in more competitive exclusion and generate a more consistent checkerboard pattern. If true, then after standardizing the C-scores from each analysis, I would expect a higher C-score for Hydropsychidae and Simuliidae together compared to the C-scores from the separate analyses of both taxa.

My Chapter 3 presents a novel combination of methods to studying the importance of competitive interactions in the distribution of biodiversity: sampling over small spatial extents, establishing patterns in the distribution of biodiversity using diversity components analysis, followed by testing for explicit taxon specific interactions using checkerboard analysis. Such an approach may be adapted for use in studies of other systems and taxa. As well, my thesis will add insight into the importance of competition in the assemblage of stream insect communities.

Methods

Sampling of White Clay Creek (PA, USA)

In order to assess the importance of competition in stream insect community assembly, rocks were sampled throughout the White Clay Creek watershed in Chester County, Pennsylvania (USA, June 20, 2013; Fig. 2.3B). Five stations were sampled spanning 2 arms of White Clay Creek (Fig 2.3B). At each station, 3 riffles were sampled running perpendicular to the stream flow (n=15 total transects). At each transect, 2 rocks were sampled (n=30 total rocks); a rock sample sometimes consisted of 2-3 rocks if the selected rock was part of a tight cluster (given Hydropsychidae build their net case in crevices). Sampling consisted of gently moving rocks from the stream into a water-filled bucket where the rock(s) were hand scrubbed to remove insects. Once clean and inspected, each rock was wrapped in tinfoil before being discarded; the weight of the tinfoil needed to cover the rock(s) was converted into a measurement of the area sampled (where weight[g]/0.004=area [cm^2], conversion calibrated in the lab), which was then used in the conversion of species abundances per rock into density per m^2 of surface area of substrate. Insects washed into the bucket were filtered through a 200 μm hand net and

moved into a 118-237 ml jar and preserved with 95% ethanol. Samples were stored in a freezer (-20 ^{o}C), and ethanol was changed at least once within 24 hours. Sample processing consisted of removing all the invertebrates from the bulk sample using a 4x dissecting microscope and sorting insects according to order or family (for Trichoptera, Coleoptera, and Simuliidae) into 20 ml vials (CABIN, McDermott et al. 2010).

DNA barcoding of specimens

DNA barcoding followed the same procedure as in Chapter 2 (see Appendix I). When available, twenty individuals/family/rock were randomly selected for DNA barcoding. As well, species were defined according to molecular operational taxonomic units (MOTUs) based on the sequence data. Delineations were done on the basis of barcode index numbers (BINs), which are assigned to specimens using an algorithm that clusters similar sequences through the Barcode of Life Data System (Ratnasingham and Hebert 2007, 2013). Low sequencing success for a given insect family/sample was also treated as in Chapter 2.

Additive diversity components

Additive diversity partitioning was used to assess if species diversity was non-randomly distributed at White Clay Creek. Additive partition analyses were performed on species richness using PARTITION 3.0 (Crist et al. 2003, Veech and Crist 2009). Total diversity for White Clay Creek was defined as:

White Clay Creek insect diversity $\gamma=\alpha_{rock} + \beta_{rock} + \beta_{transect} + \beta_{station} + \beta_{arm}$

Partitions were performed for the families Hydropsychidae and Simuliidae and the orders Trichoptera and Coleoptera. In order to assess departures from null distributions, expected values of diversity components were determined using individual-based randomization procedures (simulating diversity components if individual specimens are randomly allocated to each rock). Significance of departures (greater or less) in the observed values was determined using two-tailed p-values (>0.975 if less than expected value, <0.025 if greater than expected value). Five thousand iterations were run for each partition in the calculation of the null distributions. Standard errors for observed α component estimations were also calculated from the raw data. The spatial design was balanced most of the time, except for 2 rocks (for both Coleoptera and Simuliidae) and 1

transect (Coleoptera) with no recorded species; these samples are necessarily removed from the analysis given they cannot provide an index of diversity.

Goodness-of-fit tests, or G-tests, were also used to determine if the distribution in diversity components (expressed as a percentage of total diversity) differed among the analysed taxa. The test statistic G was calculated as $2*(\Sigma$ observed richness*ln[observed richness/expected richness]), and significance was determined using the chi-square distribution (Sokal and Rohlf 1981). Trichoptera was used as the expected distribution testing for departures in the other taxa; because Hydropsychidae is a subset of Trichoptera, a comparison was not made between these two distributions in diversity components. Hydropsychidae was used as the expected value to determine if Simuliidae and Coleoptera were significantly different from Hydropsychidae, while Simuliidae was used as the expected distribution to determine if Coleoptera was significantly different from Simuliidae. The opposite analyses (where Simuliidae and Coloeptera were used as expected distributions) were also performed; results remained the same.

Checkerboard analyses

Partition analysis was used to establish at which spatial scale(s) variation in the distribution of stream insect species occurs, which in turn informed at which spatial scale checkerboards should be analysed to detect signatures of competition between specific taxa. Checkerboards were calculated using the co-occurrence module in EcoSim7.0 (Gotelli and Entsminger 2001). The number of checkerboard units for a species pair is calculated as $CU=(r_i-S)*(r_j-S)$, where r_i and r_j are the number of cases where species i and j occur without the other, and S is the number of sites inhabited by both species (Stone and Roberts 1990). The C-score is the average of checkerboard units across all species pairs. To test for significant departures from random chance, observed C-scores were compared to the null distribution of C-scores based on random iterations of the input matrix; significance was determined by the two-tail probability that values greater than the average null C-score were due to chance (e.g. 0.025 for significantly greater observed values, 0.975 for significantly lesser values). The C-score was evaluated for species presence/absence matrices for each taxon separately (Trichoptera, Hydropsychidae, Coleoptera and Simuliidae) for sites spanning the entirety of White Clay Creek (scale of 3 km; station and arm levels from partition analyses). Additional analyses were

performed for site-by-species matrices with all species sampled (Trichoptera, Coleoptera, and Simuliidae) and for Hydropsychidae and Simuliidae together.

Because it was not clear at which sample grain competitive exclusion occurs (e.g. are species excluded from rocks, transects or entire stream reaches), checkerboards were performed defining samples (or sites) at different spatial extents. Note that sites (however defined) were always analysed spanning the entire sampled spatial extent of White Clay Creek (approximately 3 km). Analyses were performed at the rock (meter; n=30), transect (~3 meters; n=15), and station (~20 meters; n=5) sample grains; a correlation between C-scores and site grain was performed, defining rock, transect and station site grains as 1, 3, and 20 m. Given only 2 arms of White Clay Creek were sampled, this site grain was excluded due to insufficient sample size. Note that the shift in sample grain (i.e. site size) is an entirely different issue from the question of which spatial scale exhibits variation in species distributions (which is addressed through additive partition analyses).

For the null distribution in C-scores, species incidences were randomly shuffled into 50,000 iterations of the input matrix, as per Fayle et al.'s (2011) recommendation for reducing Type I errors. A fixed row and column null model was used in the analysis, as opposed to fully randomized null matrices. When the number of times a species occurred (column totals) is fixed in the analysis, the rarity or commonness of each species is retained in the null matrices; similarly, fixed row totals retain the species richness/site. The fixed model is most appropriate for "island lists" where the full species list is known for well-defined habitat patches, as in the comprehensive sampling of rocks at White Clay Creek (Gotelli 2000). Fixed column approaches are also recommended because they are robust to Type I errors (Gotelli 2000).

C-scores were standardized for comparisons between analyses, where the standardized C-score=C_{obs}-C_{sim}/standard deviation$_{sim}$. Given the C-score is standardized by the number of standard deviations above or below the expected C-score, values greater than 2 or less than -2 represent non-random values. Degenerate sites (sites without any species) were kept in the analyses, as degenerate matrices have little effect for well-behaved null models (i.e. fixed species models; Gotelli 2000). Sites without species in the

input matrix consisted of 2 rocks in the separate analyses of Simuliidae and Coleoptera, and a single transect in the analysis of Coleoptera.

Results

DNA barcoding success

Sequencing success was high for White Clay Creek specimens (1492/1562=95.5%). Two samples from White Clay Creek had <85% sequencing success for a given family (Uenoidae, Station 2 riffle 1 rock 1, 1/2 specimens [50% success]; Uenoidae, station 2 riffle 3 rock 2, 3/4 specimens [75% success], Fig. 2.3B).

Additive diversity components

The distribution in diversity components differed according to the taxa analysed. Diversity components for Coleoptera and Simuliidae differed from that of Trichoptera (Coleoptera: G_4=12.31, p=0.015; Simuliidae: G_4=16.5, p=0.002; Fig. 3.1). Diversity components for Coloeoptera was also different from Hydropsychidae (G_4=13.4, p=0.010) and from Simuliidae (G_4=11.5, p=0.021), while the distribution for Simuliidae was similar to that of Hydropsychidae (G_4=5.80, p=0.215; Fig. 3.1). α diversity at the smallest scale (rock) was considerably smaller than expected by chance for all taxa, accounting for 29-39% of the total diversity. β rock was not different from expected values for all taxa, though it was notably larger (approaching significance) for Coleoptera, accounting for 13% of total diversity. Likewise, while β transect was significantly greater than expected for all taxa, it was slightly larger in Simuliidae and Coleoptera, accounting for 22 and 25% of total diversity (as opposed to 16% and 17% in Trichoptera and Hydropsychidae, respectively). β station and β arm was also significantly large in all taxa, accounting for a large proportion of total diversity in Trichoptera and Hydropsychidae (a combined 46 and 42% of total diversity, respectively), while Coleoptera and Simuliidae showed comparatively smaller β values at the larger spatial scales (30 and 31% of total diversity, respectively).

Checkerboards

Significantly high C-scores were detected in the analyses of Trichoptera; Hydropsychidae; Hydropsychidae and Simuliidae; and Trichoptera, Coleoptera, and Simuliidae (Fig. 3.2), indicating important segregation of species. These analyses showed

significantly high C-scores at the rock and transect sample grain levels, but not when sites were defined at the station spatial scale. Standardized C-scores were negatively correlated with the spatial grain of sites (r=-0.53, p=0.024, n=18). The analysis of Trichoptera and Simuliidae had the highest standardized C-score (3.46, p=<0.001), which is greater than the standardized C-score for the separate analyses of Hydropsychidae (2.10, p=0.027) and Simuliidae (0.36, p=0.334). Coleoptera and Simuliidae on their own showed random structure at all sample grains.

Discussion

Competition is assumed to be important in governing species distributions at small spatial scales, but detecting signatures of competition remains difficult with observational data. Here, additive diversity partitioning was used to inform these efforts by establishing the spatial scales at which diversity is structured within White Clay Creek (PA, USA); these patterns in turn informed at which spatial extent checkerboards should be conducted in order to infer competitive interactions between specific taxa.

Non-random distributions in species diversity at White Clay Creek

Species diversity was non-randomly distributed within White Clay Creek, and patterns in additive diversity components differed according to taxa (Fig. 3.1). This was first evidenced by the significant shortage of species at the smallest spatial level (α rock), where on average only 29-39% of the total species pool occurred. Though Ligeiro et al. (2010) analysed stream insect diversity at the family level using a similar small scale design (<5 km), they also found alpha diversity at the level of centimeters was only 30% of the total species pool. These results indicate that within-stream processes, and thus potentially very local competition, significantly affect the distribution of insect species. In terms of the distribution of species at larger within-stream spatial levels, each taxa analysed showed important variation in the distribution of species at the stream arm and station levels (e.g. kilometers extent), however Trichoptera and Hydropsychidae showed comparatively larger β values at these spatial levels, which accounted for 46 and 42% of the total variation. Ligeiro et al. (2010) report β diversity for stream insect families at comparable spatial levels to be 30% of total diversity for stone substrate.

These results indicate most of the differences in Trichopteran species occurrences at White Clay Creek occurred at the scale of kilometers; checkerboards were therefore performed for sites spanning White Clay Creek to determine if this variation could be attributed to specific competitive interactions. It should be noted, however, that while this spatial level was deemed optimal for species of Trichoptera, including Hydropsychidae, Coleoptera showed more variation in species distributions at the rock and transect levels (where β was 39% of total diversity; Fig. 3.1), while Simuliidae showed equal variation at the meter and kilometer scales (β rock and transect was 29% of total diversity, β station and arm was 31%; Fig. 3.1). Checkerboard analyses investigating competitive interactions involving Coleoptera therefore ought to be executed across a smaller spatial extent (10s of meters) with more rock and transect sites; additional sampling would be required to amass an appropriate number of samples across the smaller spatial extent. The differences in additive partition results for each taxa also indicate within stream processes structure diversity differently according to taxa. If these patterns are attributed to competitive interactions, then competition appears to structure Coleoptera across smaller spatial extents compared to Trichoptera.

Competitive interactions at White Clay Creek

Once it was established that the majority of variation in species diversity at White Clay Creek was structured at the scale of kilometers, checkerboard analyses were used to investigate if specific taxa tended to segregate; patterns of segregation would, in turn, be indicative of important competitive interactions. Note that the environment was assumed to have not impacted species distributions at White Clay Creek. Given recorded variation in environmental parameters at White Clay Creek tended to be considerably less than variation recorded across a much larger spatial extent (1-66% of variation across 2500 km), it was likely environmental filtering played a weak role in structuring the distribution of species compared to competitive interactions (Table 3.1). However, future studies using checkerboards should devise formal analyses to prove environmental heterogeneity and dispersal limitations did not restrict species distributions.

Checkerboard analyses detected significant segregation of species at White Clay Creek, particularly within Trichoptera, Hydropsychidae, and between members of

Hydropsychidae and Simuliidae (Fig. 3.2). Because the distribution of species at White Clay Creek is not likely due to environmental filtering or dispersal limitation, these patterns of segregation are attributed to competitive interactions. Competition amongst Simuliidae and Hydropsychidae are not surprising given previous experimental studies describing interactions between the two taxa; Simuliidae are often excluded from filter-feeding locations through avoidance behavior of Hydropsychidae, or otherwise though physical attacks (Hemphill 1988). This polarized competitive relationship was further evidenced by the greater standardized C-score in the analysis of Hydropsychidae and Simuliidae together compared to either taxa on its own (Fig. 3.2). The significantly high C-score for Trichoptera, including Hydropsychidae, also indicates important interactions occur between species within these taxa; Hydropsychidae are also known to be aggressive with each other, competing for retreat locations on rocks (Jansson and Vuoristo 1979). Besides detecting signatures of these interactions, the checkerboard analyses also suggested segregation amongst species of Trichoptera produced a larger C-score when site grain was defined as riffles (as opposed to rocks; Fig. 3.2). Though formal testing is needed, this suggests the competitive "reach," the scale at which competitive interactions unfold amongst species, occurs at the level of a few meters, as opposed to centimeters (on a single rock).

According to the checkerboard analyses, Simuliidae and Coleoptera did not undergo important competitive interactions with members of their respective taxa. While this may be the case, important variation in the distribution of species also occurred at smaller spatial scales (i.e. β rocks and transects), especially for Coleoptera (Fig. 3.1). The spatial extent of the checkerboard analysis (kilometers) may not have been optimal for detecting patterns of segregation within these taxa. In addition, because important variation also occurred at the larger spatial extent analysed for these two taxa, it is possible the checkerboard analyses simply did not include the correct taxa. For instance, Coleoptera could be analysed with other functionally similar insects such as scrapers within Trichoptera or predatory guilds such as Odonata (Atwood et al. 2014), or other abundant insect orders not identified to species in this study (i.e. Ephemeroptera, Plecoptera, Chironomidae).

On a final note, several limitations with the checkerboard analysis and the detection of competitive interactions must be addressed. For one, common species present an inherent problem with the analysis; if certain species occur at every site and drive other species to exclusion, this will not produce a checkerboard as expected of competitive interactions. This occurs because there is no way to assess if the excluded species would exist without the presence of the common one. In this case, important competitive interactions occur within the community but will go undetected by the checkerboard analysis. Similarly, checkerboards rely solely on presence absence data, or competitive exclusion, as a proxy for competitive interactions. If competitive exclusion is a rare outcome in competitive interactions, then density data are needed to investigate the effect two species have on one another. While diversity partitioning cannot infer the importance of competition between specific taxa, it can be used to determine if diversity is non-randomly structured according to different orders of diversity (weighting the diversity index to emphasize rare or common species; see Appendix VIII). Finally, while checkerboards can determine if species tend to segregate within specific taxa, they cannot elucidate if specific species pairs segregate more than expected by chance. The number of checkerboard units may be investigated for species pairs, but is complicated by the lack of a null model for comparison of values.

Conclusions

The role of competitive interactions has long been assumed to play a role in governing species distributions (Darwin 1895, Elton 1946, Hardin 1960), but has remained difficult to infer using observational data. By sampling over a small spatial extent, however, researchers can minimize the importance of other processes such as environmental filtering and dispersal limitations, and begin attributing non-random distributions of species to competitive interactions. This was done for stream insect assemblages at White Clay Creek (PA, USA) using additive diversity partitioning. Once non-random structure in species distributions was confirmed at White Clay Creek, checkerboard analyses were used to identify important interactions between specific taxa, in particular the filter-feeding Hydropsychidae and Simuliidae larvae. Researchers seeking to investigate the role of competitive interactions in community structure might employ a similar approach, but should consider resampling taxa at spatial scales deemed

to be important through additive partitioning analyses. As well, researchers may consider the possibility of using other analyses that utilize observational data such as phylogenetic community structure (Webb 2000, Webb et al. 2002). If researchers can supplement sampling of species diversity at regional spatial extents with more localized efforts, then we will achieve a clearer picture of the full-scale of processes governing the distribution of biodiversity.

Table 3.1. Environmental variable maxima and minima for White Clay Creek vs. values from large-scale sampling. Large scale values include sampling from Churchill, MB (July 15-23, 2012); Algonquin Provincial Park, ON (Canada; June 26-July 13, 2011); and Southern Pennsylvania (USA; June 17-19, 2013). Values at the 2500 km extent include a single reading at White Clay Creek independent of the values taken at the 3 km extent. The last column represents variation at White Clay Creek expressed as a percentage of variation at the 2500 km extent. Note that substrate type is a categorical variable (see appendix VI).

	White Clay Creek (3 km)		2500 km Extent		% Var.
Variable	Min	Max	Min	Max	
Water Temp (Celsius)	15.73	18.16	13.5	25.2	20.8
DO (mg/L)	8.98	10.87	5.25	14.2	21.1
Conductivity (uS/cm)	194	336	9	6437	2.2
pH	8.29	8.45	6.23	9.22	5.4
Chlorophyll (mg/L)	1.1	2.7	0.2	30.4	5.3
Salinity	0.09	0.13	0.01	3.54	1.1
Substrate 1 (mm)	Cobble (65-250)	Boulder (>250)	Silt (<0.06)	Bed Rock	20.0
Substrate 2 (mm)	Sand (0.06-2)	Bed Rock	Clay	Bed Rock	66.7
Max depth (m)	0.04	0.2	0.01	2	8.0
Wetted width (m)	1.4	7.71	0.65	25.05	25.9
Max hydraulic head (mm)	0	80	0	130	61.5

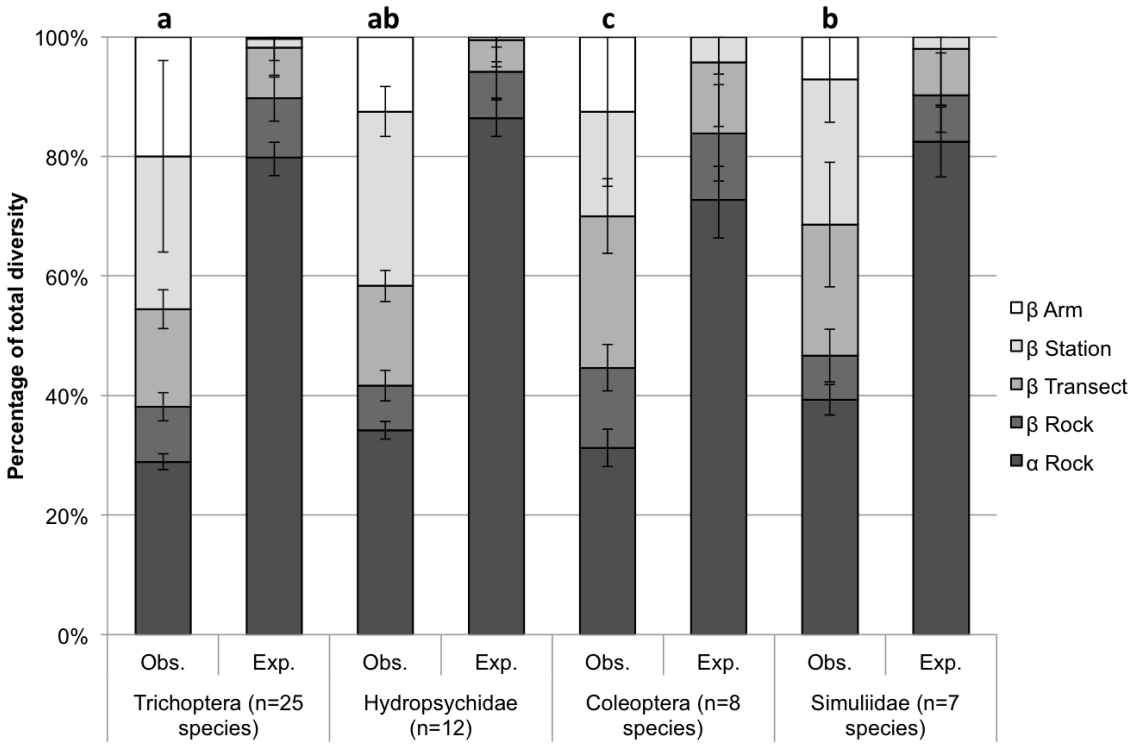

Figure 3.1. Additive partitions of sampled stream insect species diversity at White Clay Creek (PA, USA; June 20, 2013). Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Error bars for expected values are max and min estimations from the null model. Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. Letters above observed diversity components indicate statistically similar or different distributions. See supplementary material (Appendix VII) for raw values of α and β.

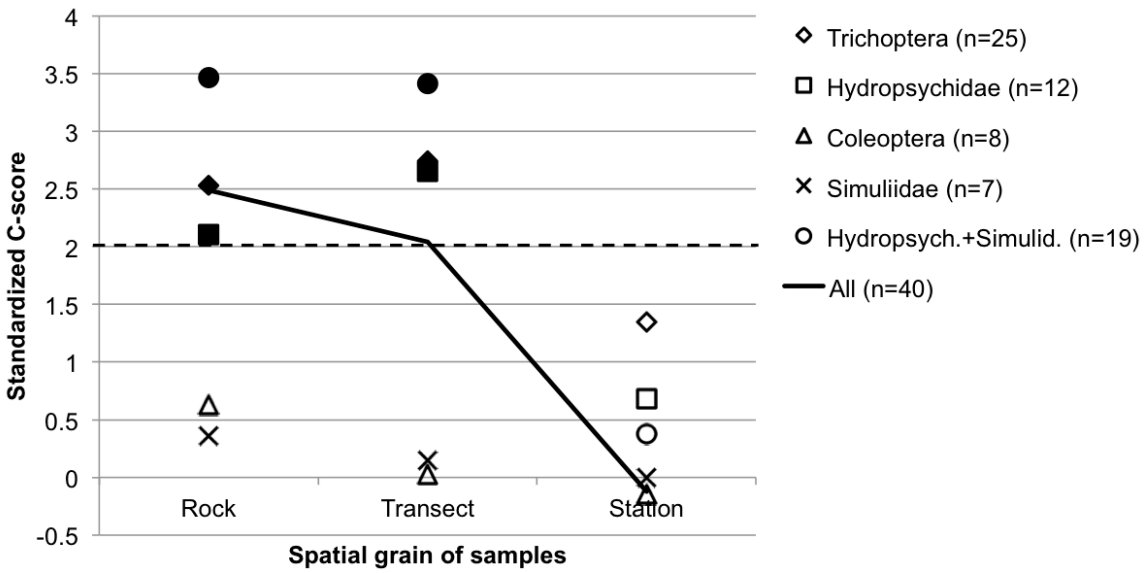

Fig. 3.2. Standardized C-score according to taxon and spatial scale of sites at White Clay Creek (PA, USA; June 20, 2013). Values of standardized C-score representing random community structuring are bound by the dashed line; values above the line represent segregation of species. Filled markers also indicate a C-score significantly different than expected by chance, while non-filled markers indicate C-scores that do not deviate from expectations from random structuring.

Chapter 4: General Conclusions

Biodiversity is important on a number of levels, from its ecological and evolutionary complexity to the wealth of economic benefit we stand to gain from its continued preservation. Our efforts to conserve biodiversity, particularly species diversity, begin with understanding processes driving its distribution. Before we can do this, however, we must establish at which spatial scales variation in biodiversity occurs. Diversity components analyses do this by partitioning species diversity into constituent regional (γ) and local (α) diversity and the distributional heterogeneity among local units (β) (Crist et al. 2003, Veech 2005). My thesis addressed methodological issues with the detection of species and how this affects the partitioning of diversity, while adding to our understanding of how diversity components are structured in disparate regions. These efforts in turn inform future hypotheses regarding processes that structure biodiversity.

My objectives with regards to methodology were to estimate the amount of undetected species diversity in my own sampling and determine the impact of taxonomic resolution on results. While the impact of taxonomic resolution has not yet been addressed in studies of diversity components, undersampling ought to result in overestimation of β diversity (Beck et al. 2013). I sampled stream caddisfly larvae from sub-arctic Churchill (MB, Canada), mixed-boreal forest Algonquin Provincial Park (ON, Canada), and temperate southern Pennsylvania (USA) and used DNA sequence data to make species-level identifications. I found that a large percentage of regional diversity remained undetected in my sampling and that taxonomic resolution, particularly moving from species to genus or family-level identifications, significantly impacted the distribution in diversity components. Studies of diversity components that use morphological identifications likely compound these two issues by failing to detect cryptic species and/or failing to identify specimens consistently to the species level. Consequently, values of diversity components are potentially highly inaccurate in reported studies, particularly if absolute values are presented.

I also investigated whether or not regions from different latitudes have similar structure in diversity components. I found that while diversity components were structured differently at Algonquin, diversity components were nearly identical for

Churchill and Pennsylvania. While previous studies have demonstrated that coarse-grained β diversity changes with latitude (Hof et al. 2008, Thieltges et al. 2011), and that fine-grained β remains relatively consistent (Kraft et al. 2011), I demonstrate that β diversity from very fine to relatively coarse-grained spatial levels can be remarkably similar in two regions, despite a large latitudinal gradient. The similar structure at Churchill and Pennsylvania indicates factors driving the distribution of species amongst vastly different habitat types may be more universal than anticipated.

My investigation of stream caddisfly diversity components has several limitations. For one, the replication at the regional level was minimal, making inferences of processes structuring diversity along the latitudinal gradient difficult. Therefore, my study informs future hypotheses rather than making any definitive assertions. More studies along a latitudinal gradient incorporating longitudinal replication are needed before conclusions are reached. The sampling effort to complete such a large-scale study, however, is also problematic; future research will have to carefully consider the number of spatial scales and replicates needed to attain a reasonable understanding of diversity components according to latitude. A meta-analysis will likely be needed to reach a conclusion based on the combined efforts of many studies. Complicating these efforts, however, is the question of how taxonomy and time scales sampled affect estimates of diversity components. My study was limited in that only a single order was considered, with no temporal replication. Future studies might consider a more thorough investigation of whether or not diversity components shift within and between years and how differences between taxa affects how species are distributed across spatial scales (i.e. differences in dispersal capabilities, phenotypic plasticity, trophic levels, etc.). Another limitation of my study was that dispersal capabilities were not assessed for stream caddisflies, and as such, any inferences of dispersal limitations, particularly at the sub-regional level, only serve to inform future studies that might confirm this suspicion. Finally, I investigated the impacts of coarse taxonomic resolution on diversity components as a proxy for biases associated with datasets based on morphological identifications of species. The merits of DNA-based specimen identification could be more directly assessed by comparing results from two versions of a single dataset, one based on morphological and the other on molecular approaches.

While processes driving distributions in biodiversity largely operate at local (environmental) to regional spatial scales (biogeographic), interspecific interactions may be an important driver of biodiversity at extremely localized scales. My thesis also investigated localized patterns in insect species distributions at a single stream (White Clay Creek, PA, USA) using diversity components, which in turn informed hypotheses regarding competitive interactions between specific taxa that were tested using checkerboard analyses (Stone and Roberts 1990). Detecting competitive interactions using observational data is often met with skepticism because regional processes may generate similar patterns expected of biotic interactions; I addressed these concerns by sampling over a small spatial extent where environmental variation and dispersal limitations are minimized. Altogether, my study represents a novel approach to studying how species diversity is structured at small spatial scales and the importance of biotic interactions. I found that species diversity at the smallest spatial scale (e.g. on rocks) was significantly less than expected by chance, and that the distribution of biodiversity at larger spatial scales (m to km scale) differed for different insect taxa; Trichoptera were most structured at larger spatial extents (scale of kilometers) while Coleopetera and Simuliidae showed important structuring at smaller spatial scales (scale of meters). Checkerboards indicated important competitive interactions within Trichoptera, within Hydropsychidae and between Hydropsychidae and Simuliidae likely structured species distributions across the scale of White Clay Creek; intensive sampling at smaller spatial scales (meters) is needed to determine if competitive interactions occur within Coleoptera and Simuliidae.

Several limitations must be addressed regarding my study at White Clay Creek. For one, the sampling design at White Clay Creek was intended for additive diversity partitioning across multiple spatial scales, meaning sampling for checkerboards analyses was not optimized. Sampling rocks evenly across White Clay Creek would have been more appropriate for a checkerboard analysis, and partition analysis indicated intensive sampling at a smaller spatial extent was needed for Coleoptera and possibly Simuliidae; as such, the dataset compiled may not have been appropriate for assessing if competitive interactions occur within these taxa. The appropriate spatial grain with which to analyse species distributions using checkerboards was also an unknown going into sampling of

White Clay Creek. The checkerboard analyses indicated strong patterns of segregation at the transect level; transect level data could be used to optimize sampling efficiency in future studies. In addition, the inference of competition structuring species diversity would benefit immensely from a temporal component to sampling. If species assemblages shift towards segregated patterns over time, then the inference of competition becomes much more convincing because the changes in community structure are observed, rather than inferred post hoc.

To conclude, the partitioning of species diversity across spatial scales can be used to comprehensively establish patterns in the distribution of biodiversity. In particular, future studies could investigate the notion that species diversity can be consistently structured regardless of major habitat differences. DNA-based species identification should be considered in these studies in order to alleviate biases due to undersampling and coarse taxonomic level datasets. Another major contribution of my thesis is in showing that a localized spatial approach to diversity components and checkerboard analyses is valuable for revealing the importance of biotic interactions in structuring the distribution of species diversity using observational data. With continued efforts to study biodiversity using diversity components, researchers will move forward with new ideas and hypotheses regarding the distribution of diversity on earth.

Literature Cited

Anderson, M. J., et al. 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecology Letters 14:19-28.

Atwood, T. B., E. Hammill, D. S. Srivastava and J. S. Richardson. 2014. Competitive displacement alters top-down effects on carbon dioxide concentrations in a freshwater ecosystem. Oecologia 175:353-361.

Beck, J., J. D. Holloway, C. Vun Khen and I. J. Kitching. 2012. Diversity partitioning confirms the importance of beta components in tropical rainforest Lepidoptera. The American Naturalist 180(3):64-74.

Beck, J., J. D. Holloway and W. Schwanghart. 2013. Undersampling and the measurement of beta diversity. Methods in Ecology and Evolution 4:370-382.

Bengtsson J., H. Jones and H. Setälä. 1997. The value of biodiversity. Trends in Ecology and Evolution 12(9):334-336

Bilton, D. T., J. R. Freeland and B. Okamura. 2001. Dispersal in freshwater invertebrates. Annual Review of Ecology and Systematics 32:159-81.

Blagoev, G. A., N. I. Nikolova, C. N. Sobel, P. D. N. Hebert and S. J. Adamowicz. 2013. Spiders (Araneae) of Churchill, Manitoba: DNA barcodes and morphology reveal high species diversity and new Canadian records. BMC Ecology 44(13):1-17.

Boschilia S. M., E. F. Oliveira and S. M. Thomaz. 2008. Do aquatic macrophytes co-occur randomly? An analysis of null models in a tropical floodplain. Oecologia 156:203-214.

Caners, R. T., S. E. Macdonald and R. J, Belland. 2013. Bryophyte assemblage structure after partial harvesting in boreal mixedwood forest depends on residual canopy abundance and composition. Forest Ecology and Management 289:489-500.

Cavender-Bares, J., K. H. Kozak, P. V. A. Fine and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. Ecology Letters 12:693-715.

Colwell, R. K. 2013. EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples. Version 9.1. Persistent URL[http://purl.oclc.org/estimates]

Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. Journal of Plant Ecology 5:3-21.

Connor, E. F. and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology 60(6):1132-1140.

Costanza, R. et al. 1997. The value of the world's ecosystem services and natural capital. Nature 387:253-260.

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

Crist, T. O., J. A. Veech, J. C. Gering and K. S. Summerville. 2003. Partitioning species diversity across landscapes and regions: a hierarchical analysis of α ß and γ diversity. The American Naturalist 162(6):734-743.

Cunha M. R., F. L. Matos, L. Génio, A. Hilário, C. J. Moura, A. Ravara and C. F. Rodrigues. 2013. Are organic falls bridging reduced environments in the deep sea? - results from colonization experiments in the Gulf of Cádiz. PLoS ONE 8(10):e76688.

Darwin C. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.

D'Amen M., D. Birtele, L. Zapponi and S. Hardersen. 2013. Patterns in diurnal co-occurrence in an assemblage of hoverflies (Diptera: Syrphidae). European Journal of Entomology 110(4):649-656.

De Troch, M., J. L. Melgo-Ebarle, L. Angsinco-Jimenez, H. Gheerardyn and M. Vincx. 2008. Diversity and habitat selectivity of harpacticoid copepods from sea grass beds in Pujada Bay, the Philippines. Journal of the Marine Biological Association of the United Kingdom 88(3):515-526.

Diamond J. M. 1975. Assembly of species communities. In: Cody M. L., Diamond J. M. (eds) Ecology and Evolution of Communities. Belknap Press, Harvard University:342-444.

Elton C. 1946. Competition and the structure of ecological communities. Journal of Animal Ecology 15(1):54-68.

Fayle, T. M. and A. Manica. 2010. Reducing over-reporting of deterministic co-occurrence patterns in biotic communities. Ecological Modelling 221:2237-2242.

Flohre, A. et al. 2011. Agriculture intensification and biodiversity partitioning in European landscapes comparing plants, carabids and birds. Ecological Applications 21(5):1772-1781.

Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3:294-299.

Fouquet, A., A. Gilles, M. Vences, C. Marty, M. Blanc, N. J. Gemmell. 2007. Underestimation of species richness in neotropical frogs revealed by mtDNA analyses. PloS ONE 2(10):1-10.

Francisco-Ramos, V. and J. E. Arias-González. 2013. Additive partitioning of coral reef fish diversity across hierarchical spatial scales throughout the Caribbean. PLoS ONE 8(10):1-14.

Georgian, T. and J. H. Thorp. 1992. Effects of microhabitat selection on feeding rates of net-spinning caddisfly. Ecology 73(1):229-240.

Gheerardyn, H., M. De Troch, M. Vincx and A. Vanreusel. 2010. Diversity and community structure of harpacticoid copepods associated with cold-water coral substrates in the Porcupine Seabight (North-East Atlantic). Helgoland and Marine Research 64:53-62.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81(9):2606-2621.

Gotelli, N. J. and G. L. Entsminger. 2001. EcoSim: Null models software for ecology. Version 7.0. Acquired Intelligence Inc. & Kesey-Bear.

Gotelli, N. J. and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology 83(8): 2091-2096.

Hajibabaei, M. et al. 2005. Critical factors for assembling a high volume of DNA barcodes. Philosophical Transactions of the Royal Society B: Biological Sciences 360:1959–1967.

Hajibabaei M., S. Shokralla, X. Zhou, G. A. C. Singer and D. J. Baird. 2011. Environmental barcoding: a next generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6(4):1-7.

Hamilton, A. J., et al. 2011. Quantifying uncertainty in estimation of tropical arthropod species richness. The American Naturalist 176(1):90-95.

Hardin G. 1960. The competitive exclusion principle. Science 131(3409):1292-1297

Hebert P. D. N., A. Cywinska, S. L. Ball, J. R. deWaard. 2003a. Biological identifications through DNA barcodes. Proceedings of The Royal Society 270:313-321.

Hebert P. D. N., S. Ratnasingham, J. R. deWaard. 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London 270:96-99.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences. 101(41):14812-14817.

Heino, J. 2005. Metacomminity patterns of highly diverse stream midges: gradients, checkerboards, and nestedness, or is there only randomness? Ecological Entomology 30:590-599.

Hemphill, N. 1988. Competition between two stream dwelling filter-feeders Hydrophyche oslari and Simulium virgatum. Oecologia 77(1):73-80.

Hof, C., M. Brandle and R. Brandl. 2008. Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. Global Ecology and Biogeography 17:539-546.

Hooper, D. U. et al. 2005. Effect of biodiversity on ecosystem functioning: a consensus of current knowledge 75(1):3-35.

Hutchinson G. E. 1959. Homage to Santa Rosalia or why are there so many kinds of animals? The American Naturalist 93(870):145-159.

Ivanova, N. V., J. R. DeWaard and P. D. N. Hebert. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes 6:998–1002.

Ivanova, N. V. and C. Grainger. 2007a. COI Amplification. CCDB Protocols. http://www.ccdb.ca/docs/CCDB_Amplification.pdf

Ivanova, N. V. and C. Grainger. 2007b. Sequencing. CCDB Protocols. http://www.ccdb.ca/docs/CCDB_Sequencing.pdf

Jansson, A. and T. Vuoristo. 1979. Significance of stridulation in larval Hydropsychidae (Trichoptera). Behaviour 71(1/2):167-186.

Jones C., K. M. Somers, B. Craig and T. B. Reynoldson. 2007. Ontario Benthos Biomonitoring Network: Protocol Manual. Queen's Printer for Ontario.

Jost, L. 2006. Entropy and diversity. Oikos 113:363-375.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. Ecology 88: 2427-2439.

Jurasinski, G., V. Retzer and C. Beierkuhnlein. 2012. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. Oecologia 159:15-26.

Kamilar J. M. and J. A. Ledogar. 2011. Species co-occurrence patterns and dietary resource competition in primates. American Journal of Physical Anthropology 144:131-139.

Kraft, N. J. B. et al. 2011. Disentangling the drivers of β diversity along latitudinal and elevational gradients. Science 333:1755-1758.

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76:5-13.

Leibold M. A. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters. 7: 601-613.

Ligeiro R., A. S. Melo and M. Callisto. 2010. Spatial scale and the diversity of stream macroinvertebrates in a Neotropical catchment. Freshwater Biology 55:424-435.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology: the benefits of genus- and species-level identifications. Journal of the North American Benthological Society 20:287–298.

MacArthur, R. H. 1965. Patterns of species diversity. Biological Reviews of the Cambridge Philosophical Society 40:510-533.

Marques, T. and J. H. Schoereder. 2014. Ant diversity partitioning across spatial scales: ecological processes and implications for conserving tropical dry forests. Austral Ecology 39:72-82.

Marshall, S. A. 2006. Caddisflies. In: Insects: Their Natural History and Diversity. Firefly Book, pp. 237-246.

Martin G. 2013. Impact of taxonomic resolution on the detection of metacommunity patterns in freshwater invertebrates using DNA barcoding. MSc Thesis. University of Guelph. Ontario Canada.

Martínez-Falcón, A. P., M. A. Marcos-García and C. E. Moreno. 2011. Temporal shifts and niche overlapping in *Copestylum* (Diptera, Syrphidae) communities reared in cactus species in a central Mexican scrubland. Ecological Research 26:341-350.

Mayfield M. M. and J. M. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. Ecology Letters 13:1085-1093.

McDermott H., T. Paull and S. Strachan. 2010. Canadian Aquatic Biomonitoring Network: Laboratory Manuals, Processing, Taxonomy, and Quality Control of Benthic Macroinvertebrate Samples. Environment Canada.

Melo, A. S., F. Schneck, L. U. Hepp, N. R. Simões, S. Siqueira and L. M. Bini. 2011. Focusing on variation: methods and applications of the concept of beta diversity in aquatic ecosystems. Acta Limnologica Brasiliensia 23(3):318-331.

Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson and B. Worm. 2011. How many species are there on earth and in the ocean. PLoS ONE 9(8):e1001127.

Morais, H. C., E. R. Sujii, M. Almeida-Neto, P. S. De-Carvalho, J. D. Hay and I. R. Diniz. 2011. Host plant specialization and species turnover of caterpillars among hosts in the Brazilian Cerrado. Biotropica 43(4):467-472.

Morse, J. C. and R. W. Holzenthal. 2008. An Introduction to the Aquatic Insects of North America, 4th Ed. Kendall Hunt.

Muller J. and M. M. Gossner. 2010. Three-dimensional partitioning of diversity informs state-wide strategies for the conservation of saproxylic beetles. Biological Conservation 143:625-633.

Negro M., A. Rolando and C. Palestrini. 2011. The impact of overgrazing on dung beetle diversity in the Italian maritime alps. Environmental Entomology 104(5):1081-1092.

Oliver, S. and Jr. Flint. 1962. Larvae of the caddisfly genus Rhyacophila in eastern North America (Trichoptera: Rhyacophilidae). Proceedings of the United States National Museum 113(3464):465-493.

Oksanen M. 1997. The moral value of biodiversity. Ambio 26(8):541-545.

Paknia, O. and M. Pfeiffer. 2011. Hierarchical partitioning of ant diversity: implications for conservation of biogeographical diversity in arid and semi-arid areas. Diversity and Distributions 17:122-131.

Pech-Canche, J. M., C. E. Moreno and G. Halffter. 2011. Additive partitioning of Phyllostomid bat richness at fine and coarse spatial and temporal scales in Yucatan, Mexico. Écoscience 18(1):42-51.

Pimm S. L. and P. Raven. 2000. Extinction by numbers. Nature 403:843-845.

Randall A. 1991. The value of biodiversity. Ambio 20(2):64-68.

Ratnasingham, S. and P. D. N. Hebert. 2007. BOLD database. http://www.barcodinglife.com/. Date accessed: March 14, 2014.

Ratnasingham, S. and P. D. N. Hebert. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE 8(8):e66213.

Rodríguez-Zaragoza, F. A., A. L. Cupul-Magaña, C. M. Galván-Villa, E. Ríos-Jara, M. Ortiz, E. G. Robles-Jarero, E. López-Uriarte, J. E. Arias-González. 2011. Additive partitioning of reef fish diversity variation: a promising marine biodiversity research tool. Biodiversity Conservation 20:1655-1675.

Ruiter, D. E., E. E. Boyle and X. Zhou. 2013. DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area. BMC Ecology 13(5):1-39.

Schoener T. W. and G. H. Adler. 1991. Greater resolution of distributional complementarities by controlling for habitat affinities: a study with Bahaman lizards and birds. The American Naturalist 137(5):669-692.

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

Smith, A. M., J. R. Rodriguez, J. B. Whitfield, A. R. Deans, D. H. Janzen. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences 105(34):12359-12364.

Sokal R. R. and J. F. Rohlf. 1981. Biometry. San Francisco: Freeman: 859.

Stein E. D., M. C. Martinez, S. Stiles, P. E. Miller and E. V. Zakharov. Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? PLoS ONE 9(4):e95525

Stone L. and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85:74-79.

Summerville, K. S. and T. O. Crist. 2005. Temporal patterns of species accumulation in a survey of Lepidoptera in a beech-maple forest. Biodiversity and Conservation 14:3393-3406.

Sweeney B. W., J. M. Battle, J. K. Jackson and T. Dapkey. 2011. Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? Journal of the North American Benthological Society 30(1):195-216.

The International Convention on Biological Diversity. 2003. Convention on Biological Diversity. http://www.biodiv.org/convention/

Thieltges, D. W., C. Hof, D. M. Dehling, M. Brändle, R. Brandl, and R. Poulin. 2011. Host diversity and latitude drive trematode diversity patterns in European freshwater fauna. Global Ecology and Biogeography 20:675-682.

Thorp J. H. 1983. An evaluation of hypotheses on the evolutionary differentiation of catchnets in net-spinning caddisflies (Hydropsychidae). Oikos 40:308-312.

Tuomisto, H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography 33:2-22.

Ulritch W. 2004. Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. Oikos 107:603-609.

Underwood A. J. 1997. Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance. England: Cambridge University Press.

Underwood A. J. and D. J. Fletcher. 2002. How to cope with negative estimates of components of variance in ecological field studies. Journal of Experimental Marine Biology and Ecology 273:89-95.

Veech, J. A., K. S. Summerville, T. O. Crist, J. C. Gering. 2002. The additive partitioning of species diversity: recent revival of an old idea. Oikos 99:3-9.

Veech, J. A. 2005. Analyzing patterns of species diversity as departures from random expectations. Oikos 108:149-155.

Veech, J. A. and T. O. Crist. 2009. PARTITION: software for hierarchical partitioning of species diversity, version 3.0. http://www.users.muohio.edu/cristto/partition.htm. Date accessed: Jan 10, 2014.

Veech J. A. and T. O. Crist. 2010. Toward a unified view of diversity partitioning. Ecology 91(7):1988-1992.

Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Köhler, F. Glaw and M. Vences. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proceedings of the National Academy of Sciences 106(20):8267-8272.

Ward, R. D., B. H. Holmes and G. K. Yearsley. 2008. Brief communications: DNA barcoding reveals a likely second species of Asian sea bass (barramundi) (*Lates calcarifer*). Journal of Fish Biology 72:458-463.

Webb C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. The American Naturalist 156(2):145-155.

Webb C. O., D. D. Ackerly, M. A. McPeek and M. J. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33:475-505.

Weins J. A. 1989. Spatial scaling in ecology. Functional Ecology 3(4):385-397.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30(3):279-338.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21(2/3):213-251.

Wilcox, R. and I. Hagenbuch. 2010. Pennsylvania Aquatic Insects. http://paaquaticfliesrus.bd.psu.edu/webroot/index.aspx. Date accessed: Feb 12, 2014.

Willis, K. J. and R. J. Whittaker. 2002. Species diversity-scale matters. Science 295:1245-1248.

Wilsey, B. J. 2010. An empirical comparison of beta diversity indices in establishing prairies. Ecology 91(7):1984-1988.

Witt, J. D. S., D. L. Threloff and P. D. N. Hebert. 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Molecular Ecology 15:3073-3082.

Wu, F., X. J. Yang and J. X. Yang. Additive diversity partitioning as a guide to regional montane reserve design in Asia: an example from Yuccan Province, China. Diversity and Distributions 16:1022-1033.

Zhou X., S. J. Adamowicz, L. M. Jacobus, R. E. Dewalt, P. D. N. Hebert. 2009. Towards a comprehensive barcode library for arctic life - Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. Frontiers in Zoology 6(30):1-9.

Zhou X., L. M. Jacobus, R. E. DeWalt, S. J. Adamowicz, P. D. N. Hebert. 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. Journal of the North American Benthological Society 29(3):814-837.

Appendix I: DNA barcoding pipeline

DNA barcoding was used to identify species of stream caddisflies (order Trichoptera), beetles (order Coleoptera), and blackflies (family Simuliidae). The pipeline for DNA barcoding consists of several steps, including DNA extraction, amplification, sequencing, and editing. Prior to DNA extraction, 1-2 photographs of each specimen (lateral, ventral, or dorsal profiles) were taken using Leica® imaging software and uploaded to respective BOLD projects (Lotic Trichoptera Larvae of Churchill-Summer 2012, LEPTO; Lotic Trichoptera Larvae of Southern Pennsylvania-Summer 2013, LTLSP; White Clay Creek Microhabitat Sampling for Insect Larvae-Summer 2013 [PA, USA], WCCMS). DNA extraction began by removing a leg from each specimen and organizing them in 96-well Eppendorf plates containing 50 µl of ethanol, which was evaporated prior to DNA extraction. Extracts were then prepared using 5 ml of invertebrate lysis buffer and 0.5 ml Proteinase K per plate and a glass-fiber protocol (Ivanova et al. 2006); extracts were re-suspended in 50 µl of molecular grade water. The COI barcode region of DNA extracts were amplified using 6.25 µl of 10% trelahose, 2 µl of ddH_2O, 1.25 µl 10X buffer, 0.625 µl of 50 mM $MgCl_2$, 0.0625 µl of 10mM dNTP, 0.06 µl Taq polymerase (Invitrogen®), 0.125 µl of 10µM of LepFol primer cocktails (equal parts of both primers for forward and reverse), and 2 µl of DNA template (i.e. the extracts; Folmer et al. 1994, Hebert et al. 2004, Ivanova et al. 2006). Eight plates were processed through to sequencing in the Adamowicz lab, while 32 were submitted to the Canadian Center for DNA Barcoding for barcoding procedures. See Table A1.1 for primer sequences and thermocycling regime times.

PCR products were checked for successful amplification using E-gels. 3.5 µl of PCR products were added to 14.5µl of ddH_2O and entered into the E-gel wells; E-gels ran for approximately 6 minutes before being inspected under UV light (Ivanova & Grainger 2007a). Successfully amplified products moved onto the DNA sequencing step. Only the reverse COI barcode primer was used for unidirectional sequencing; because species identification and delineation was the main goal of barcoding for this project, I did not seek the "barcode standard" compliance that is used for reference library building, which requires at least two trace files. For sequencing reactions, PCR products were diluted with 40 µl of dH_2O, then 2 µl of diluted product was added to 5µl of 10% trehalose,

0.875 µl of ddH2O, 0.250 µl of BigDye®, 1.875 µl of 5x buffer, and 1 µl of 10 µM reverse primer LepFol cocktail (Ivanova & Grainger 2007b; see Table A1.1 for thermocycler regime times). Sequencing products then underwent cleanup and Sanger sequencing through the Canadian Center for DNA Barcoding (CCDB; Hajibabaei et al. 2005). The reverse sequence chromatograms were then aligned and edited using Codon Code Aligner (CodonCode Corporation v. 4.2.1) and uploaded to respective BOLD projects. Alignments were verified to be free of gaps and stop codons. Final sequences were also submitted to GenBank.

Table A1.1. Primers and regime times used to sequence the COI barcode regions of stream insects.

Primer	**Forward or Reverse**	**5' to 3' sequence**	**Reference**
LepF1	F	ATTCAACCAATCATAAAGATATTGG	Hebert et al. 2004
LepR1	R	TAAACTTCTGGATGTCCAAAAAATCA	Hebert et al. 2004
LCO1490	F	GGTCAACAAATCATAAAGATATTGG	Folmer et al. 1994
HCO2198	R	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. 1994
Regime Times			
PCR	94^{o}C 1 min; 5x[94^{o}C 40 s, 45^{o}C 40 s, 72^{o}C 1 min]; 35x[94^{o}C 40 s, 51^{o}C 40 s, 72^{o}C 1 min]; 72^{o}C 5 min; maintain at 4^{o}C		
Sequencing	96^{o}C 2 min; 30x[96^{o}C 30 s, 55^{o}C 15 s]; 60^{o} 15 s; 15x[96^{o}C 10 s, 55^{o}C 5 s, 60^{o}C 2 mins]; 60^{o}C 4 min; maintain at 4^{o}C		

Appendix II: Species lists

A full caddisfly species list according to region was constructed in order to assess the assumption each region was sampled at comparable times (i.e. differences in species occurrences between regions not due to the different years sampled). I tested the null hypothesis that regional overlap in all species recorded previous to sampling and species sampled during the course of this project will be the same; if true, differences between regions for sampled species occurrences are not due to the different years sampled (because the sampled overlap is consistent with expectations based on historical sampling). The sampled species list indicated 15/114 (13.16%) species sampled overlapped between Algonquin and Pennsylvania, 2/114 (1.75%) species sampled overlapped between Churchill and Algonquin, and a single species (0.88%, *Hydropsyche bronta*) overlapped between all regions (Table A2.1). The full species list indicated nearly identical percentages (13.49, 1.86, 0.93%, respectively; x^2=0.54, p=0.91; Table A2.1). Because the null hypothesis is not rejected, the sampled regions are deemed to have been sampled at comparable times, and patterns detected in the distribution of biodiversity are not likely due to yearly variation in species occurrences.

Table A2.1. Caddisfly (order: Trichoptera) species recorded in sampled regions. Bolded species represent new BINs to BOLD. CH=Churchill (MB, Canada), APP=Algonquin Provincial Park (ON, Canada), PA=Pennsylvania (USA). Filled circles represent species sampled in the present study, empty circles represent species not sampled but that have historically been found in a given region at similar times and habitats (i.e. missed species). This list represents 215 summer species (CH=46 species, APP=49 species, PA=155).

Taxonomy				**Occurrence**		
Family	**Genus**	**Species (or interim)**	**BIN(s)**	**CH**	**APP**	**PA**
Apataniidae	*Apatania*	***A. sp.*BOLD:ACK6518**	**BOLD:ACK6516**	○		●
Beraeidae	*Beraea*	*B. fontana*				○
		B. nigritta				○
Brachycentridae	*Brachycentrus*	*B. americanus*	BOLD:AAA4295	●		
	Micrasema	*M. sp.*BOLD:AAB2409	BOLD:AAB2409		●	●
		M. charonis	BOLD:AAC9355			●
		M. wataga	BOLD:AAB2410			●
Calamoceratidae	*Heteroplectron*	*H. americanum*				○
Dipseudopsidae	*Phylocentropus*	*P. lucidus*				○
Glossosomatidae	*Agapetus*	*A. minutus*	BOLD:AAD6616			●
		A. pinatus	BOLD:AAJ7261			●
	Glossosoma	*G. sp.*BOLD:ACF5706	BOLD:ACF5706	●		
		G. intermedium	BOLD:AAA9475	●		

		G. nigrior	BOLD:AAB5627			●
	Protoptila	*P. maculata*				○
		P. palina				○
Goeridae	*Goera*	*G. calcarata*				○
		G. fuscula	BOLD:AAA6219			●
		G. stylata				○
Helicopsychidae	*Helicopsyche*	*H. borealis*				○
Hydropsychidae	*Arctopsyche*	*A. ladogensis*		○		
	Cheumatopsyche	*C. sp.*BOLD:AAA5695	BOLD:AAA5695		●	●
		*C. sp.*BOLD:AAA8252	BOLD:AAA8252		●	
		*C. sp.*BOLD:AAA3891	BOLD:AAA3891		●	
		*C. sp.*BOLD:ACE5263	BOLD:ACE5263		●	
		***C. sp.*BOLD:ACJ2073**	**BOLD:ACJ2073**			●
		C. analis	BOLD:ABZ3099			●
		C. campyla	BOLD:ACE5262		●	●
		C. ela				○
		C. enigma	BOLD:AAC1007			●
		C. gracilis				○
		C. halima	BOLD:AAA3956			●
		C. harwoodi	BOLD:ACJ1066			●
		C. minuscula	BOLD:ACE5144		●	○
		C. oxa	BOLD:ACF3382		●	●
		C. sordina				○
		C. vannotei				○
	Diplectrona	*D. sp.*BOLD:AAA1609	BOLD:AAA1609		●	●
		*D. sp.*BOLD:AAA1610	BOLD:AAA1610		●	●
		*D. sp.*BOLD:AAA1611	BOLD:AAA1611			●
	Hydropsyche	*H. sp.*BOLD:ACC8104	BOLD:ACC8104		●	
		H. alternans	BOLD:AAA3236	○	●	
		H. betteni	BOLD:AAA1669		●	●
		H. bronta	BOLD:AAA3450	●	●	●
		H. hageni				○
		H. impula				○
		H. morosa	BOLD:AAA3680		●	●
		H. opthalmica	BOLD:AAI2269			●
		H. phalerata				○
		H. scalaris				○
		H. slossonae	BOLD:AAA2527		●	●
		H. sparna	BOLD:AAA2528		●	●
		H. valanis				○
		H. ventura	BOLD:AAA5028		●	●
		H. vexa		○		
		H. walkeri				○
	Macrostemum	*M. zabratum*				○
	Parapsyche	*P. apicalis*	BOLD:AAA8586		●	○
Hydroptilidae	*Hydroptila*	*H. alabama*				○
		H. ampoda				○
		H. armata				○
		H. callia				○
		H. consimilis	BOLD:ABX5069	●		○
		H. delineata				○
		H. grandiosa	BOLD:AAA9834			●
		H. gunda				○
		H. perdita	BOLD:AAE5187			●
		H. spatulata				○
		H. strepha				○

	Ithytrichia	*I. sp.*BOLD:AAQ3188	BOLD:AAQ3188		●	
	Leucotrichia	*L. pictipes*	BOLD:AAC6744			●
	Mayatrichia	*M. ayama*	BOLD:AAF0842		●	
	Ochrotrichia	***O. sp.*BOLD:ACI5903**	**BOLD:ACI5903**	●		
		O. aegerfasciella				○
		O. eliaga	BOLD:AAX0103	●		
		O. anabola				○
	Oxyethira	*O. coercens*	BOLD:AAF2530	●		
		O. forcipata				○
		O. grisea				○
		O. rivicola				○
		O. sida	BOLD:ABA5467		●	○
		O. zeronia				○
	Palaeagapetus	*P. celcus*				○
	Stactobiella	*S. palmata*				○
Lepidostomatidea	*Lepidostoma*	*L. americanum*	BOLD:AAC8896			●
		L. bryanti				○
		L. griseum	BOLD:AAE3970		●	●
		L. latipenne	BOLD:AAC4102			●
		L. ontario				○
		L. pictile	BOLD:AAA2390		●	○
		L. sommermanae				○
		L. togatum	BOLD:AAA2325	○	●	○
		L. vernale				○
Leptoceridae	*Ceraclea*	***C. sp.*BOLD:ACG9707**	**BOLD:ACG9707**	●		
		C. alabamae				○
		C. ancylus				○
		C. annulicornis	BOLD:AAA5429	●		
		C. cancellata				○
		C. diluta	BOLD:AAC2090		●	○
		C. excisa	BOLD:AAB1628	●		
		C. flava				○
		C. maculata				○
		C. mentiea				○
		C. nigronervosa	BOLD:AAC3781	●		
		C. punctata				○
		C. resurgens	BOLD:AAB8249	●		
		C. tarsipunctata	BOLD:ACE3230		●	○
		C. transversa				○
		C. wetzeli				○
	Leptocerus	*L. americanus*				○
	Mystacides	*M. sp.*BOLD:ACF0896	BOLD:ACF0896		●	
		M. interjectus	BOLD:ACJ9907	●		
		M. sepulchralis	BOLD:AAA8765			●
	Nectopsyche	*N. pavida*	BOLD:AAC6606		●	
		N. candida				○
		N. exquisita				○
Leptoceridea	*Oecetis*	*O. avara*				○
		O. cinerascens				○
		O. inconspicua				○
		O. ochracea		○		
		O. persimilis	BOLD:AAB9989		●	○
		*O. sp.*BOLD:AAI3634	BOLD:AAI3634		●	
		S. guttatus				○
	Setodes	*S. incertus*	BOLD:AAC0284			●
	Triaenodes	*T. aba*				○
		T. flavescens				○

		T. ignitus	BOLD:AAB3345		●	○
		T. injustus				○
		T. marginatus				○
		T. reuteri		○		
		T. tardus	BOLD:ABY4173		●	○
Limnephilidae	*Anabolia*	*A. bimaculata*		○		
	Asynarchus	*A. montanus*		○		
		A. rossi	BOLD:ABX4942	●		
	Frenesia	*F. difficilis*	BOLD:AAE6279		●	
	Glyphopsyche	*G. irrorata*	BOLD:AAE0748		●	
	Grammotaulius	*G. interrogationis*		○		
	Hydatophylax	*H. argus*				○
	Limnephilus	*L. argenteus*		○		
		L. canadensis		○		
		L. externus	BOLD:AAA2803	●		
		L. extractus	BOLD:ABX5731	●		
		L. hageni	BOLD:AAA2635	●		
		L. indivisus				○
		L. infernalis		○		
		L. nigriceps	BOLD:ABX5272	●		
		L. partitus	BOLD:AAA6553	●		
		L. perpusillus	BOLD:AAB8644	●		
		L. picturatus		○		
		L. sericeus		○		
		L. submonilifer				○
	Philarctus	*P. bergrothi*	BOLD:AAA2068	●		
	Pycnopsyche	*P. gentilis*				○
		P. guttifera	BOLD:AAB0591		●	
		P. lepida				○
		P. luculenta	BOLD:AAD3761			●
		P. scabripennis	BOLD:AAH7074		●	○
Mollanidae	*Molanna*	*M. blenda*				○
		M. flavicornis		○		
		M. tryphena				○
Odontoceridae	*Psilotreta*	*P. frontalis*				○
		P. labida				○
		P. rufa				○
Philopotamidae	*Chimarra*	*C. sp.*BOLD:AAA8554	BOLD:AAA8554		●	
		C. aterrima	BOLD:ACE4956			●
		C. obscura	BOLD:AAA1545		●	○
		*C. sp.*BOLD:AAA1546	BOLD:AAA1546		●	
	Dolophilodes	*D. distincta*	BOLD:AAA2305		●	●
Phryganeidae	*Agrypnia*	*A. improba*	BOLD:ACK0044	●		
		A .pagetana	BOLD:AAB4401	●		
		A. straminea	BOLD:AAA9377	●		
	Banksiola	*B. dossuaria*				○
	Oligostomis	*O. sp.*BOLD:ACE7466	BOLD:ACE7466		●	
		O. pardalis	BOLD:ACE7465		●	
	Ptilostomis	*P. ocellifera*				○
Polycentropodidae	*Neureclipsis*	*N. crepuscularis*		○		○
		N. valida	BOLD:AAD1447	●	●	
	Nyctiophylax	*N. affinis*				○
		N. celta				○
		N. denningi				○
		N. moestus	BOLD:AAC3569		●	○
	Polycentropus	*P. barri*				○
		P. carolinensis				○
		P. centralis				○

		P. cinereus	BOLD:AAA3438			●
		P. confusus	BOLD:AAB7267			●
		P .elarus	BOLD:AAF1032			●
		P. maculatus	BOLD:AAE1803			●
		P. pentus	BOLD:AAF0274		●	
		*P. sp.*BOLD:AAA3439	BOLD:AAA3439			●
		*P. sp.*BOLD:AAA3442	BOLD:AAA3442			●
Psychomiidae	*Lype*	*L. diversa*	BOLD:AAA9500		●	●
	Psychomyia	*P. flavida*	BOLD:ABZ2387			●
Rhyacophilidae	*Rhyacophila*	*R. sp.*BOLD:AAG9021	BOLD:AAG9021		●	
		R. angelita	BOLD:AAA6495	●		
		R. atrata				○
		R. banksi				○
		R. brunnea				○
		R. carolina	BOLD:AAB9620			●
		R. fuscula	BOLD:AAB1607			●
		R. glaberrima				○
		R. invaria				○
		R. minora	BOLD:AAB5700			●
		R. mongolita	BOLD:AAI0891	●		
		R. nigrita	BOLD:AAD1559			●
		R. torva	BOLD:AAB8940			●
		R. vibox	BOLD:AAB0827		●	○
Sericostomatidae	*Agarodes*	*A. griseus*				○
Uenoidae	*Neophylax*	*N. sp.*BOLD:AAB7629	BOLD:AAB7629			●
		*N. sp.*BOLD:AAB8195	BOLD:AAB8195			●
		N. aniqua				○
		N. mitchelli	BOLD:ACE4360			●
		N. oligius	BOLD:AAB7627			●
		N. ornatus				○

Table A2.2. Sampled insect species list for White Clay Creek (PA, USA; June 20, 2013). Bolded species represent new BINs to BOLD. Note: *Psychomyia flavida and Rhyacophila carolina* were morphologically identified, so no BIN is indicated. The list represents 40 species found at White Clay Creek.

Order	**Family**	**Genus**	**Species**	**BIN(s)**
Coleoptera	Elmidae	*Oulimnius*	*O. latiusculus*	BOLD:AAB2007
		Optioservus	*O. sp.*BOLD:AAA9073	BOLD:AAA9073
		Stenelmis	*S. sp.*BOLD:AAC4909	BOLD:AAC4909
			*S. sp.*BOLD:AAH2196	BOLD:AAH2196
			*S. sp.*BOLD:AAH6674	BOLD:AAH6674
		Unknown	***sp.*BOLD:AJC5785**	**BOLD:ACJ5785**
	Psephenidae	*Psephenus*	*P. herricki*	BOLD:AAB6713
		Ectopria	*E. thoracica*	BOLD:AAE8162
Diptera	Simuliidae	*Simulium*	*S. bracteatum*	BOLD:AAB1114
			S. decorum	BOLD:AAB7749
			***S. sp.*BOLD:ACI0635**	**BOLD:ACI0635**
			S. tuberosum	BOLD:AAA9710
			S. venustrum	BOLD:AAA4265
			S. verecundum	BOLD:AAA1697
			S. vittatum	BOLD:AAA4121
Trichoptera	Glossosomatidae	*Glossosoma*	*G. nigrior*	BOLD:AAB5627
	Goeridae	*Goera*	*G. calcarata*	BOLD:AAA3306
	Hydropsychidae	*Cheumatopsyche*	*C. analis*	BOLD:ABZ3099
			*C.sp.*BOLD:AAA5695	BOLD:AAA5695
			C. oxa	BOLD:ACF3382
			*C.sp.*BOLD:ACJ2073	BOLD:ACJ2073
			C. pinaca	BOLD:AAD4262
		Diplectrona	*D. sp.*BOLD:AAA1609	BOLD:AAA1609
			*D. sp.*BOLD:AAA1610	BOLD:AAA1610
			*D. sp.*BOLD:AAA1611	BOLD:AAA1611
		Hydropsyche	*H. betteni*	BOLD:AAA1669
			H. bronta	BOLD:AAA3450
			H. slossonae	BOLD:AAA2527
			H. sparna	BOLD:AAA2528
	Hydroptilidae	*Leucotrichia*	*L. pictipes*	BOLD:AAC6744
	Psychomiidae	*Psychomyia*	*P. flavida*	
	Philopotamidae	*Chimarra*	*C. aterrima*	BOLD:ACE4956
		Dolophilodes	*D. distincta*	BOLD:AAA2305
			*D. sp.*BOLD:AAA2304	BOLD:AAA2304
	Polycentropodidae	*Polycentropus*	*P. cinereus*	BOLD:AAA3438
	Rhyacophilidae	*Rhyacophila*	*R. carolina*	
	Uenoidae	*Neophylax*	*N. concinnus*	BOLD:AAC3072
			N. mitchelli	BOLD:ACE4360
			*N.sp.*BOLD:AAB5566	BOLD:AAB5566
			N. oligius	BOLD:AAB7627

Appendix III: Individual and sample-based rarefaction curves

Diversity estimation curves were calculated using EstimateS 9.0 (Colwell 2013) for Churchill and Pennsylvania to estimate the missed regional diversity component in diversity partitioning analysis. Individual and sample-based (e.g. streams) rarefactions curves (Chao1 and Chao2; Colwell et al. 2012) were calculated based on 100 randomizations of the input abundance data. This appendix presents the estimator curves.

The individual-based rarefaction curves suggest the within stream diversity was generally sufficiently sampled. Sampled diversity closely approximated the estimated diversity based on individual-based rarefaction curves, however, the lack of a plateau in the curves (particularly for Churchill) also indicates undetected species potentially remain (Figs. A3.1 and A3.3). Sample-based estimations (e.g. streams) indicate the regional diversity at Churchill and Pennsylvania was not sufficiently characterized with the 9 streams sampled (Figs. A3.2 and A3.4); future studies evidently need to sample more streams (likely closer to 20) in order to accurately sample the regional species pool. Also notable are the less steep curves in Pennsylvania compared to Churchill; this is likely due to differences in sampling effort. Churchill sampling focused on characterizing the entire stream insect community, while Pennsylvania sampling focused exclusively on the Trichoptera. Thus, more Trichoptera specimens were sampled and barcoded in Pennsylvania and the regional species pool was more accurately characterized compared to Churchill.

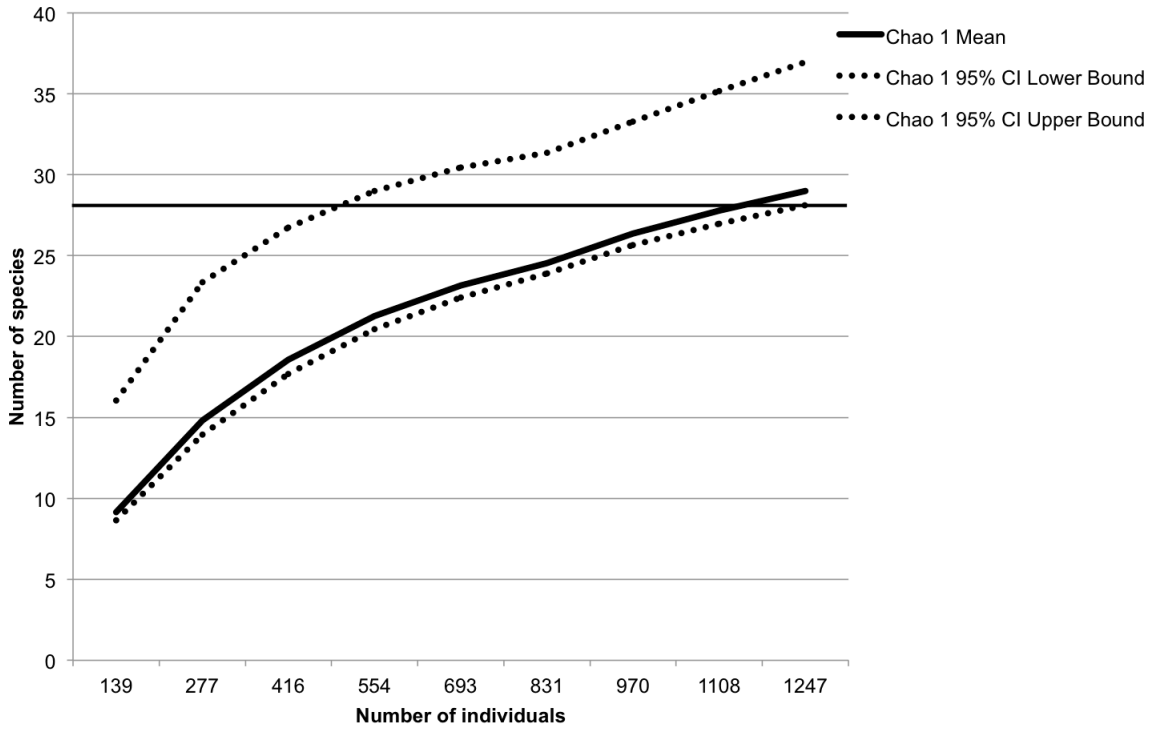

Figure A3.1. Individual-based diversity estimator curve for Churchill specimens (MB, Canada; July 15-23, 2012). 28 species were recorded in Churchill (represented as the solid horizontal line).

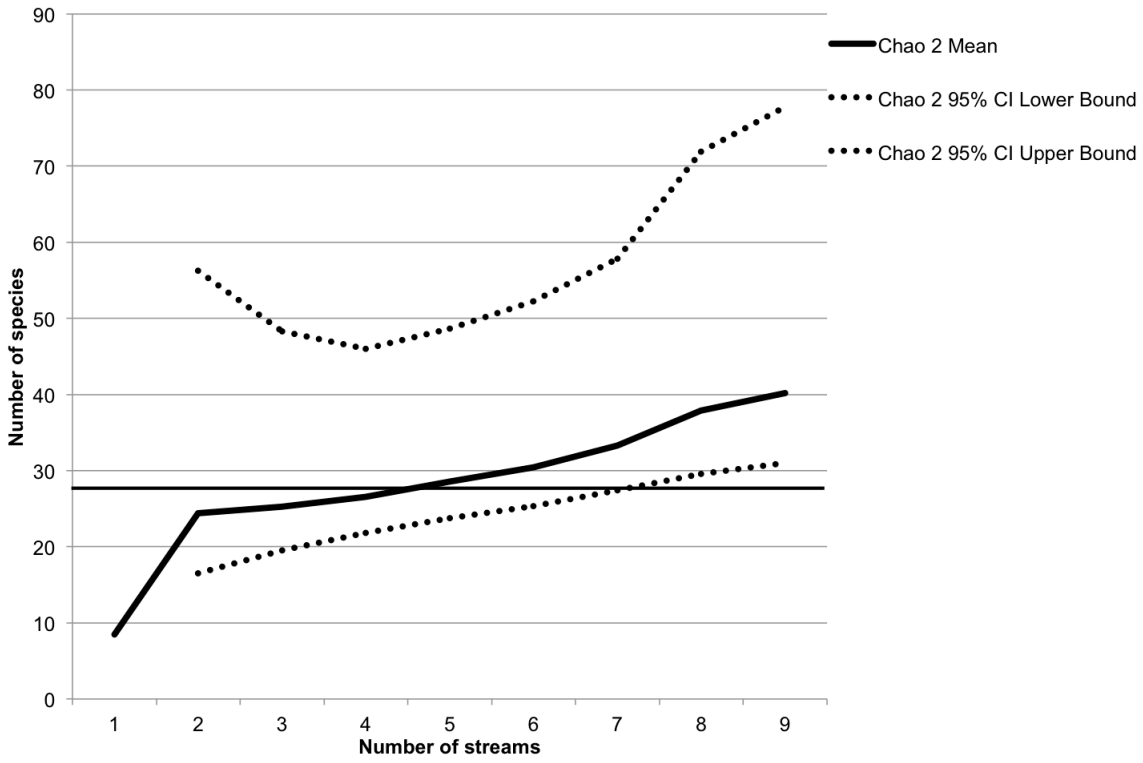

Figure A3.2. Sample-based diversity estimator curve for Churchill specimens (MB, Canada; July 15-23, 2012). 28 species were recorded in Churchill (represented as the solid horizontal line).

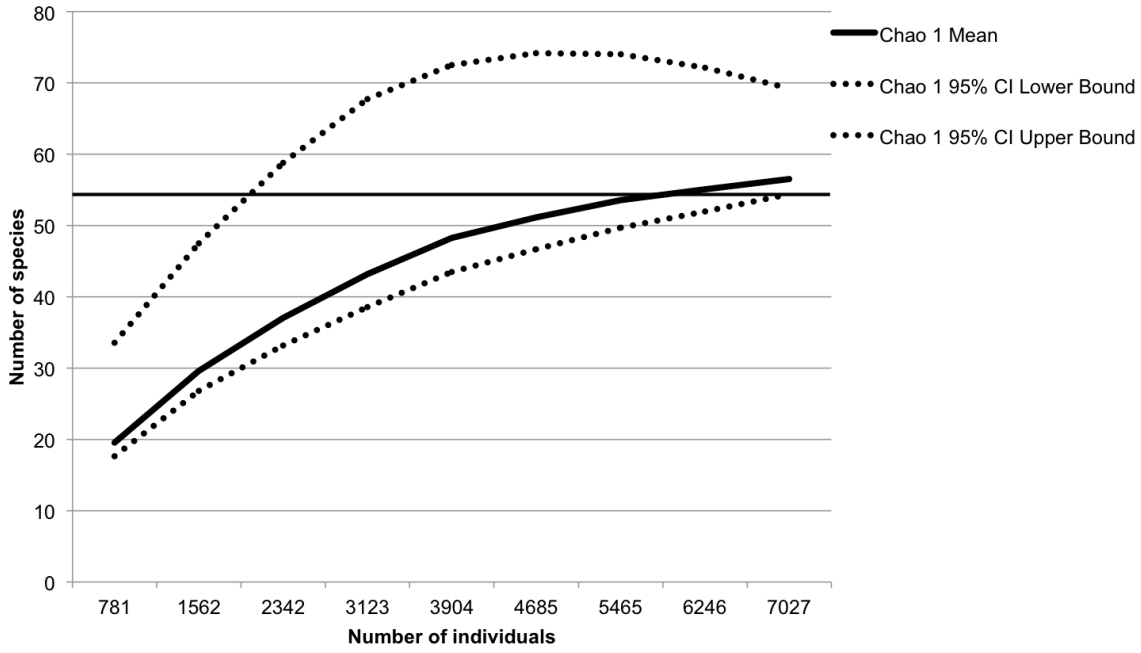

Figure A3.3. Individual-based diversity estimator curve for Pennsylvania specimens (USA, June 17-19, 2013). 54 species were recorded in Pennsylvania (represented as the solid horizontal line).

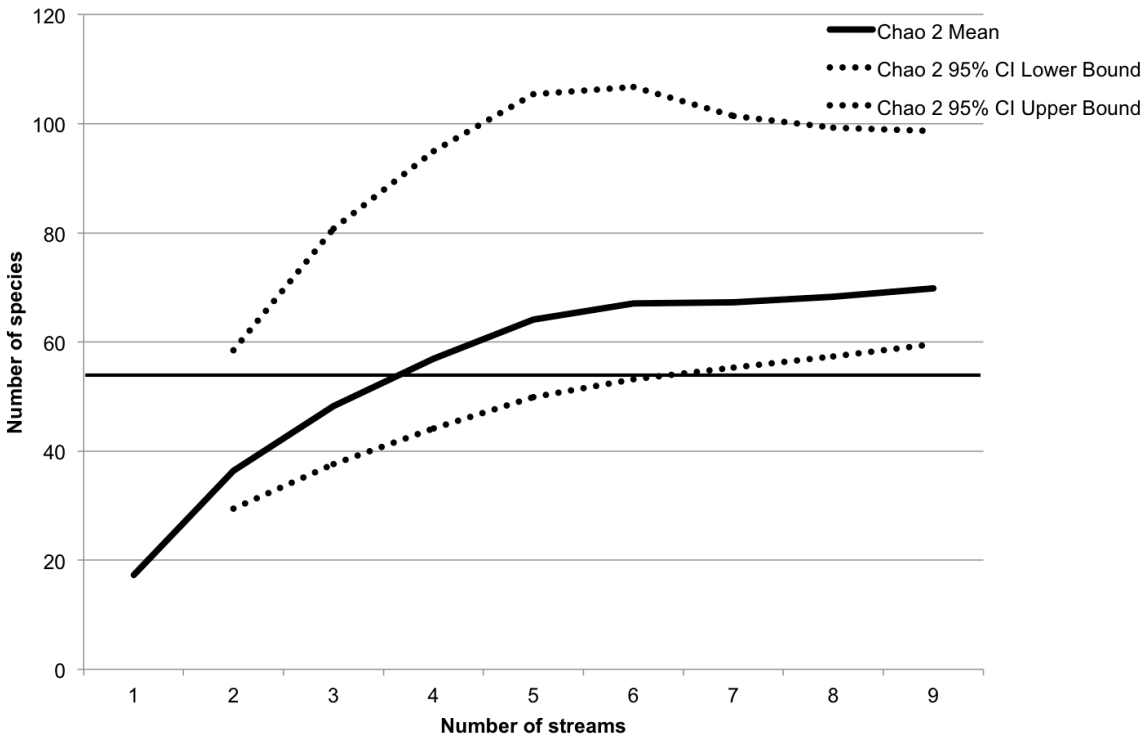

Figure A3.4. Sample-based diversity estimator curve for Pennsylvania specimens (USA, June 17-19, 2013). 54 species were recorded in Pennsylvania (represented as the solid horizontal line).

Appendix IV: Taxonomic level partitions using Algonquin data

In the analyses of diversity components according to taxonomic level, Algonquin Provincial Park was excluded due to missing kick and transect-level data. Partitions according to taxonomic level excluding kick and transect-level data are presented here in order to include Algonquin in the analysis. Stream-level data for Churchill and Pennsylvania were calculated by averaging the densities/m^2 kick (n=6-9). The distribution of diversity components including the three regions differed according to taxonomic resolution, as was found for the analysis including just two regions (Table A4.1). The distribution in diversity components was the same for species and genus taxonomic levels, though the differences did approach significance (G_3=6.95, p=0.073). The distribution in diversity components was different between species and family levels (G_3=63.33, p=<0.001) and different between genus and family levels (G_3=27.98, p=<0.001; Fig. A4.1).

Table A4.1. Additive richness components for larval summer caddisflies (order: Trichoptera) across regions, according to taxonomic level. Sampled regions are Churchill (MB, Canada), Algonquin Provincial Park (ON, Canada), and Pennsylvania (USA; summers 2011-13). Upper and lower expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. P-values less than 0.025 indicate an observed value significantly greater than the expected value.

Analysis (n=species)	Spatial scale	α or β	Obs	SE	Exp	Lower limit	Upper limit	*p*(obs>exp)
Species (n=99)	Stream	α	10.4	1.17	28.0	26.7	29.2	>0.999
	(n=26)	β	8.6		17.5	15.3	19.6	>0.999
	Sub-region	α	19.0	3.05	45.5	42.9	48.0	>0.999
	(n=9)	β	17.1		23.8	19.6	28.3	>0.999
	Region	α	36.2	8.94	69.3	64.1	74.1	>0.999
	(n=3)	β	63.2		30.0	25.2	35.2	<0.001
Genus (n=36.5)	Stream	α	6.6	0.59	13.7	12.9	14.5	>0.999
	(n=26)	β	4.2		6.7	5.3	8.1	>0.999
	Sub-region	α	10.8	1.33	20.4	18.7	22.0	>0.999
	(n=9)	β	6.1		8.0	5.26	10.7	0.989
	Region	α	16.8	2.13	28.4	25.3	31.4	>0.999
	(n=3)	β	19.7		8.1	5.1	11.2	<0.001
Family (n=15)	Stream	α	5.1	0.44	8.3	7.6	8.9	>0.999
	(n=26)	β	2.6		2.7	1.9	3.6	0.621
	Sub-region	α	7.7	0.87	11.0	10.0	11.9	>0.999
	(n=9)	β	3.0		2.2	0.9	3.5	0.040
	Region	α	10.7	1.68	13.2	11.7	14.0	>0.999
	(n=3)	β	4.3		1.8	1.0	3.3	<0.001

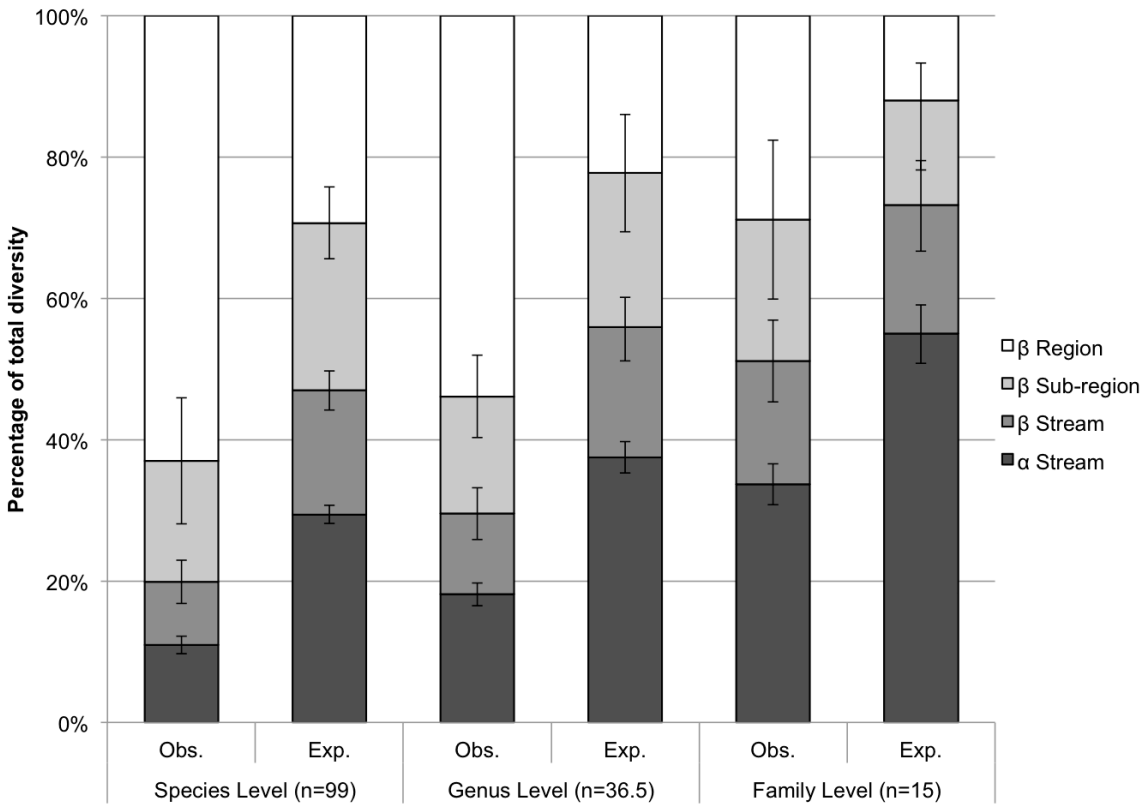

Figure A4.1. Additive partitions of sampled summer caddisfly (order: Trichoptera) diversity spanning North America, according to taxonomic level. Sampled regions are Churchill (MB, Canada), Algonquin Provincial Park (ON, Canada) and Pennsylvania (USA; 2011-13). Error bars for observed values represent 1 SE and apply to the α component at a given spatial level. Error bars for expected values are max and min estimations from the null model. Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data.

Appendix V: Investigation of sampling design balance of diversity components analyses

Several streams were excluded from the analysis of Algonquin stream caddisfly species diversity in order to maintain a balanced design in the partition analyses. The impact of design balance on partition results was therefore assessed by comparing the design used in the study (mostly balanced, missing one stream) to a completely balanced design (2 streams/region) and completely unbalanced design (5 regions, n=4, 6, 2, 2, 2 streams). The impacts of design balance to the full-scale analysis of species additive and multiplicative partitioning were assessed at the species, genus and family levels, and at the orders of diversity q=0, 0.999 and 2 for multiplicative diversity partitioning.

Results

The impact of design balance on the full-scale partition analysis was minimal. Values for diversity components were nearly identical for the balanced and nearly-balanced designs, while the β sub-region level was slightly greater and α stream was slightly less in the analysis of all streams and regions (unbalanced design; Fig. A5.1). The upward biased β sub-region level must be due to the inclusion of several extra sub-regions from Algonquin, most of which only had 2 streams. The lower α stream value indicates the full inclusion of streams must increase the relative abundance of low-richness sites; such undersampling may have contributed to artificial turnover between sub-regions, increasing the β sub-region value. All things considered, Algonquin had a slight impact on the values for diversity components when it is disproportionately represented (relative to the other regions sampled).

The impact of design balance was minimal in multiplicative partitioning, regardless of taxonomic level and order of q (Fig. A5.2). If anything, the mostly balanced approach tended to result in slightly lower values of species turnover at the lower spatial levels, generally making it the most conservative design. The unbalanced design also tended to lead to slightly larger values for species turnover at the stream level, which is again explained by the relatively larger incidence of low-richness streams from the Algonquin dataset. Generally speaking, however, the range of design balance scenarios explored here did not impact interpretation of the partition results. These results suggest

partition analyses are quite robust to uneven sampling designs, over the range of balance scenarios explored here, and cases of low sample numbers.

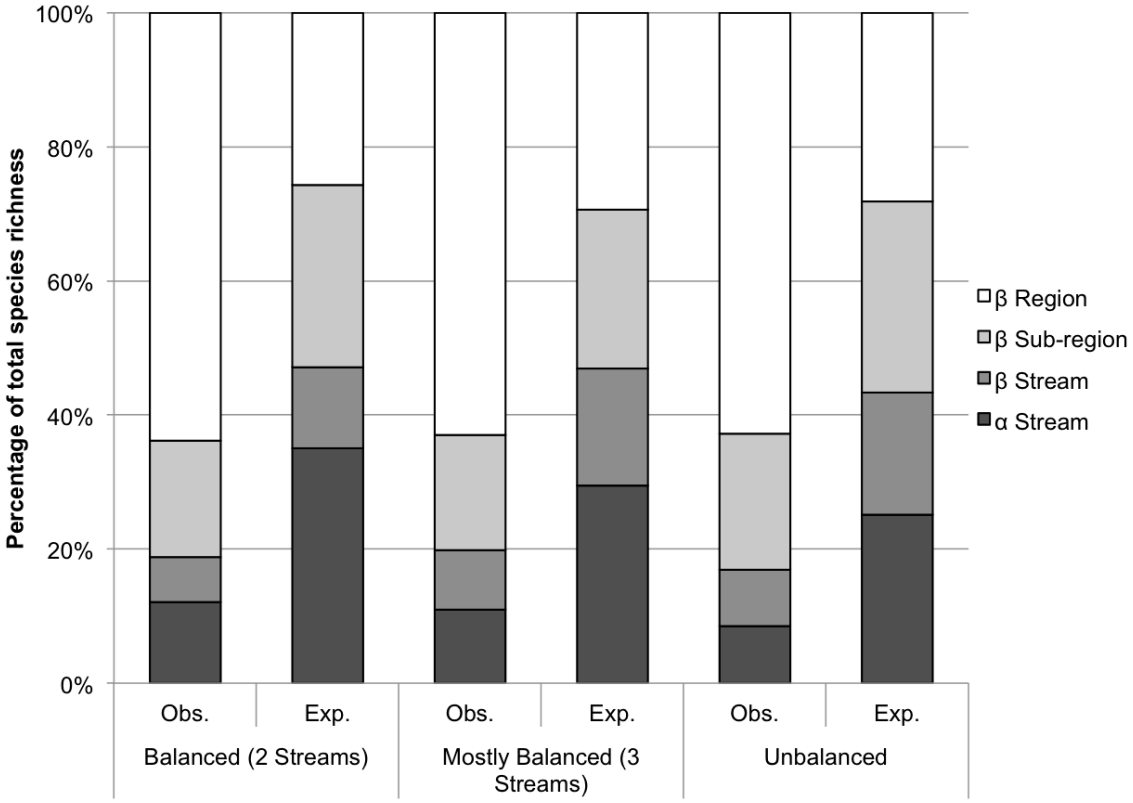

Fig. A5.1. Additive partitions of sampled summer caddisfly (order: Trichoptera) species diversity spanning North America according to design balance of the analysis. Sampled regions are Churchill (MB Canada), Algonquin Provincial Park (ON, Canada), and Pennsylvania (USA; 2011-13). Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. The mostly balanced design has 3 streams/sub-region, except for 1 sub-region in Algonquin (26 streams in total), while the unbalanced design has 5 sub-regions in Algonquin with very uneven numbers of streams (n=4, 6, 2, 2, 2 streams). The balanced design has 18 streams in the analysis, while the unbalanced design has 34.

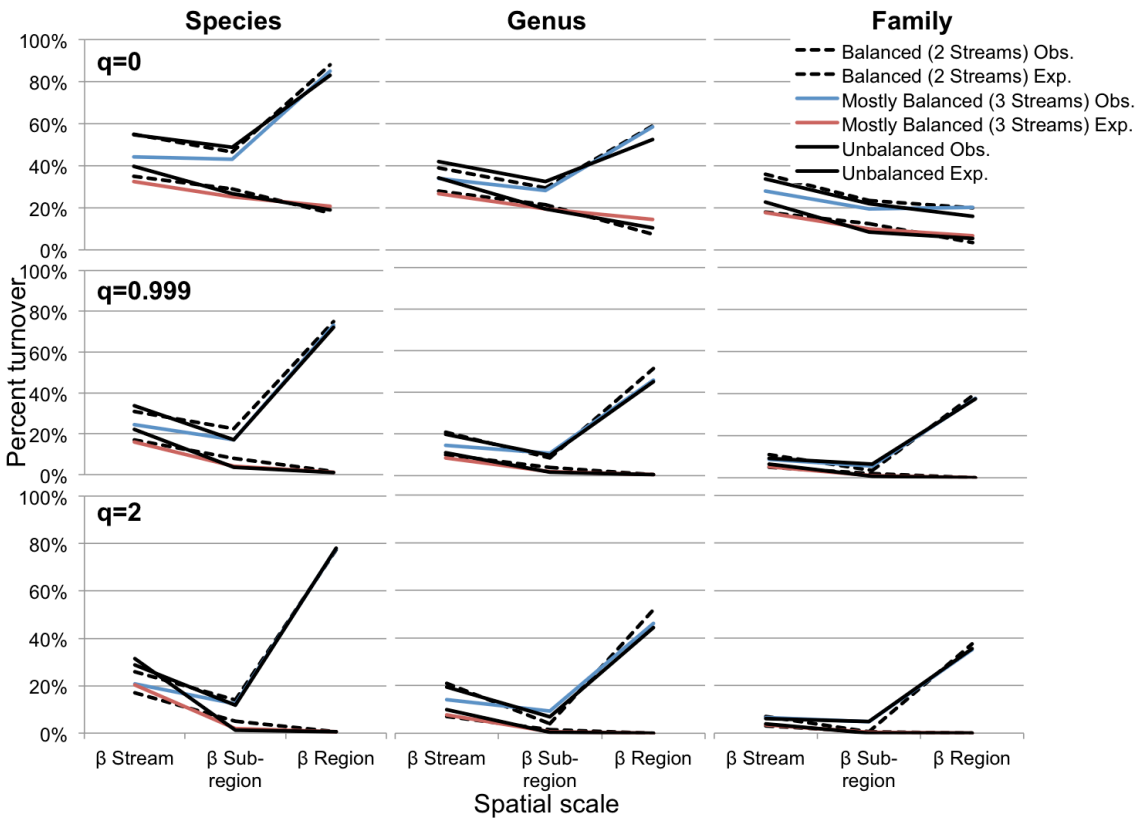

Fig. A5.2. Percent turnover of sampled summer caddisfly (order: Trichoptera) diversity spanning North America according to design balance of the analysis. Sampled regions are Churchill (MB, Canada), Algonquin Provincial Park (ON, Canada), and Pennsylvania (USA; 2011-13). Expected values of diversity were calculated based on 5000 individual-based randomization iterations of the input data. The mostly balanced design has 3 streams/sub-region, except for 1 sub-region in Algonquin (26 streams in total), while the unbalanced design has 5 sub-regions in Algonquin with very uneven numbers of streams (n=4, 6, 2, 2, 2 streams). The balanced design has 18 streams in the analysis, while the unbalanced design has 34.

Appendix VI: Analysis of environmental variation according to spatial scale

The importance of environmental variation structuring the distribution of species diversity was considered during the course of the project. Environmental parameters were measured during sampling, and the relative contributions of sampled spatial scales to environmental variation was assessed and compared to species turnover within and between regions.

Sampling methods

Habitat variables were characterized at each transect sampled for insects. Macrophytes, including emergent, rooted floating, submergent and free floating plants; floating, filamentous, attached, and slime or crust algae; and woody debris and detritus were characterized as (1) abundant, (2) present, or (3) absent. Primary and secondary substrates were also characterized using a modified Wetworth scale. Seven categories were used: (1) clay (hard pan), (2) silt (gritty, <0.06 mm), (3) sand (grainy, 0.06-2 mm), (4) gravel (2-65 mm), (5) cobble (65-250 mm), (6) boulder (>250 mm), and (7) bedrock. In addition, max stream depth (m), max hydraulic head (mm) and wetted width (m) were also recorded.

Several habitat variables were also recorded for entire stream sites, typically in the middle of the center transect (unless water level was too shallow for the probe). Water chemistry measurements included temperature (^{o}C), dissolved oxygen (DO, mg/L), conductivity (µs/cm), turbidity (NTU), pH, chlorophyll (mg/L) and salinity (ppm). Percent canopy cover was approximated at each site as (1) 0-24%, (2) 25-49%, (3) 50-74%, or (4) 75-100%. GPS coordinates, site drawings, site photographs and general comments were also recorded at each stream.

Variance components analysis of environmental variables

Variance components analysis was used in order to assess the relative contributions of each spatial scale to environmental variation and relate this to species distributions. Variance components for each spatial scale were calculated as percentages of total environmental variation using results from Model II ANOVAs of the environmental variables measured. The structure of the ANOVAs and calculations for variance components are given in Table A6.1. To correct for heterogeneity of variances, dissolved oxygen (mg/L), conductivity (µs/cm), and wetted width (m) were log10

transformed, while hydraulic head (mm) and max depth (m) were square root transformed (Underwood 1997). ANOVAs with Algonquin data were run 6 times, with the same combinations of 8 streams as in the analysis of additive partitioning of Algonquin caddisfly species; the percentages for variation at each spatial scale were averaged for the 6 ANOVAs. Subsequently, because Algonquin had one fewer stream in the analysis, a dummy value(s) was used in the ANOVAs, calculated as the average value across two streams for the third sub-region (Underwood 1997); the degrees of freedom reflected the missing stream and the harmonic mean for unbalanced replicates was used in the denominator to variance components. The variance components occasionally gave negative values; though not an uncommon problem in this analysis, negative values of variance components are non-intuitive and require correcting. While it is common practice to set the negative component to 0, this creates a bias in the calculation of the other values. Rather, a better procedure is to pool the sum of squares for the violating factor with the next lowest factor in the hierarchy (i.e. negative values for regions were dissolved and pooled into stream variation) as described by Fletcher and Underwood (2002).

To simplify interpretation and comparison of the results with turnover across spatial scales, percentages for variation at each of the spatial levels for all the environmental factors were combined using principal components analysis using dimension reduction in SPSS Statistics Standard 21.

Results and discussion

Considering all the regions sampled, environmental variation was broken into 3 major components that showed important variation peaking at the regional, stream and transect levels (Fig. A6.1). The peak in transect-level environmental variation did not correspond to increased turnover of species, though the peak in stream-level environmental variation did. The peak in regional-level environmental variation also corresponded to a substantial increase in species turnover. Interestingly, the sub-region spatial level showed almost no important environmental variation in the context of the whole dataset (Fig. A6.1), though species turnover did increase at this level. The multi-region analysis suggests environmental variation appears to be an important driver of species turnover at the stream and regional levels, but that other processes must be

driving species turnover at the other spatial levels (except at the transect level where turnover was minimal). Dispersal limitations at larger scales and biotic interactions at the kick level are plausible explanations.

In the analysis of variation within each region, important environmental variation peaked at the transect and stream levels (as in the multi-region analysis) for Churchill and Algonquin (Fig. A6.2). Species turnover was also similar as in the multi-region analysis; species turnover peaked at the stream level for Churchill and Algonquin, and important species turnover occurred at the sub-region level, which did not correspond with a peak in environmental variation. These patterns are congruent with results from the multi-region analysis; environmental variation between streams was an important driver of species turnover while other processes such as dispersal limitations at the sub-region level or biotic interactions at the kick level better explain turnover at the respective levels. Pennsylvania on the other hand, showed important environmental variation peaking at the stream and sub-regional levels, which also corresponded to increased levels of species turnover (Fig. A6.2). In the case of Pennsylvania, it would seem environmental variation at the sub-regional level drove species turnover to some degree; the most northern sites sampled represent a hilly sub-region, which likely explains the important environmental variation measured at this level in Pennsylvania.

Table A6.1. Structure for model II ANOVAs investigating environmental variation over spatial scales of sampling. A) represents the model for the dataset including transect-level data, B) represents the model for data with stream level data (i.e. does not have transect level). For variance components, *a* is the number of sub-regions within a region, *b* is the harmonic mean number of streams within a sub-region, and *n* is the number of replicate transects per stream. MS=mean square.

	Source of variation	**df**	**Denominator to F-ratio**	**Variance component**
A)	Region R_i	2	$MS_{S(R)}$	$(MS_R-MS_{SR(R)})/abn$
	Sub-region(Region) $SR_j(R_i)$	6	$MS_{S(SB(R))}$	$(MS_{SR(R)}-MS_{S(SR(R))})/bn$
	Stream(Region((Locale)) $S_k(SR_j(R_i))$	17	MS_e	$(MS_{S(SR(R))}-MS_e)/n$
	Error (transect) $e_{l(kji)}$	51		MS_e
	Total	76		
B)	Region R_i	2	$MS_{SR(R)}$	$(MS_R-MS_{SR(R)})/ab$
	Sub-region(Region) $SR_j(R_i)$	6	MS_e	$(MS_{SR(R)}-MS_e)/b$
	Error (Stream) $e_{l(kji)}$	17		MS_e
	Total	25		

Table A6.2. ANOVAs for environmental variables according to spatial scale. Regions sampled include Churchill (MB, Canada: July 15-23, 2012), Algonquin Provincial Park (ON, Canada: June 26-July 13, 2011), and Pennsylvania (USA: June 17-19, 2013). 6 ANOVAs were run with different combinations of streams and regions from Algonquin. A single dummy stream is used in all analyses with Algonquin data. Bolded mean squares (MS) indicate a significant F-ratio.

Variable	Source of variation	df	MS1	MS2	MS3	MS4	MS5	MS6	Average % variation
SQRT(Max depth[m])	Region	2	**0.2616**	0.0635	0.0924	0.1403	0.0135	0.0282	5.22
	Sub-region(Region)	6	0.0312	0.0858	0.0639	0.0447	0.0636	0.0489	0.94
	Stream(Sub-region(Region))	17	**0.0641**	**0.0642**	**0.0703**	**0.0850**	**0.0851**	**0.0912**	50.86
	error (transect)	51	0.0150	0.0149	0.0163	0.0145	0.0143	0.0157	42.97
Log10(Wetted width[m])	Region	2	**0.6484**	0.2179	0.1320	0.1666	0.0764	0.1140	4.46
	Sub-region(Region)	6	0.1133	0.0649	0.1176	0.2917	0.0732	0.0642	1.67
	Stream(Sub-region(Region))	17	**0.1821**	**0.1947**	**0.2180**	**0.1927**	**0.2053**	**0.2286**	78.83
	error (transect)	51	0.0103	0.0103	0.0103	0.0106	0.0105	0.0105	15.04
SQRT(Max hydraulic head[mm])	Region	2	27.6394	6.6991	6.4132	11.485	8.8204	14.025	0.42
	Sub-region(Region)	6	**27.2849**	11.692	13.013	**37.007**	12.276	10.852	17.90
	Stream(Sub-region(Region))	17	7.1344	7.1335	6.9959	**6.6761**	**6.6752**	6.5376	17.04
	error (transect)	51	3.9567	4.1462	4.4173	3.2466	3.4360	3.7071	64.64
Substrate 1	Region	2	10.2809	**11.318**	**16.901**	10.528	**11.861**	**18.111**	28.42
	Sub-region(Region)	6	**3.4352**	2.1759	2.5926	3.5463	2.1389	2.2222	11.43
	Stream(Sub-region(Region))	17	**1.1209**	**1.1209**	**1.1111**	**1.6176**	**1.6176**	**1.6078**	17.71
	error (transect)	51	0.4183	0.5229	0.4967	0.6797	0.7843	0.7582	42.44
Substrate 2	Region	2	4.6327	3.3735	4.6327	3.0278	3.2500	3.0278	2.15
	Sub-region(Region)	6	3.8981	4.4907	3.8981	1.6389	1.4907	1.6389	4.69
	Stream(Sub-region(Region))	17	**2.5065**	**2.5065**	**3.2124**	**2.1144**	**2.1144**	**2.8203**	40.39
	error (transect)	51	0.7582	0.7843	0.8366	0.6405	0.6667	0.7190	52.77
Woody debris	Region	2	0.1481	1.0370	0.7315	0.6420	2.1235	1.6698	2.63
	Sub-region(Region)	6	0.9136	1.0617	0.9414	1.0988	0.9506	0.9043	3.10
	Stream(Sub-region(Region))	17	**0.7320**	**0.7320**	**0.9379**	**0.7974**	**0.7974**	**1.0033**	49.57
	error (transect)	51	0.2092	0.1699	0.1699	0.2222	0.1830	0.1830	44.69
Detritus	Region	2	1.4444	1.4444	1.3611	1.3333	1.3333	1.3611	16.66
	Sub-region(Region)	6	0.3704	0.3704	0.2870	0.5926	0.5926	0.4537	8.82
	Stream(Sub-region(Region))	17	**0.2222**	**0.3399**	**0.1928**	**0.2614**	**0.3791**	0.2320	22.21
	error (transect)	51	0.1046	0.1046	0.1046	0.1307	0.1307	0.1307	52.31
Emergent macrophytes	Region	2	5.3364	5.5309	5.5309	5.7068	6.5679	6.5679	23.88
	Sub-region(Region)	6	1.7006	**2.2840**	**2.2840**	1.5895	**1.8395**	**1.8395**	22.01
	Stream(Sub-region(Region))	17	**0.7680**	**0.6797**	**0.6797**	**0.7680**	**0.6797**	**0.6797**	27.81
	error (transect)	51	0.1961	0.1569	0.1569	0.1961	0.1569	0.1569	26.30
Submergent free floating	Region	2	2.9753	4.9660	4.4568	2.9753	4.9660	4.4568	14.30
	Sub-region(Region)	6	**2.3580**	**2.0340**	**1.9877**	**2.3580**	**2.0340**	**1.9877**	36.50
	Stream(Sub-region(Region))	17	**0.3399**	**0.3497**	**0.3791**	**0.5752**	**0.5850**	**0.6144**	19.42
	error (transect)	51	0.1699	0.1830	0.1830	0.1307	0.1438	0.1438	29.78
Filamentous algae	Region	2	2.6512	3.1975	2.6512	3.0093	3.5926	3.0093	7.56
	Sub-region(Region)	6	1.7253	1.7531	1.7253	1.7562	1.7654	1.7562	4.58
	Stream(Sub-region(Region))	17	**1.5000**	**1.4902**	**1.5000**	**1.4869**	**1.4771**	**1.4869**	69.59
	error (transect)	51	0.1307	0.1176	0.1307	0.1176	0.1046	0.1176	18.27
Attached algae	Region	2	5.4198	4.4938	**5.9660**	**7.0000**	5.7778	**7.6944**	34.71
	Sub-region(Region)	6	1.0988	1.2099	1.1265	1.1481	1.4074	1.1019	8.85
	Stream(Sub-region(Region))	17	**0.7582**	**0.9150**	**0.7680**	**0.6667**	**0.8235**	**0.6765**	42.40
	error (transect)	51	0.0915	0.0784	0.0784	0.0784	0.0654	0.0654	14.03
Slime or crust algae	Region	2	**11.8920**	**12.827**	**12.827**	**10.732**	**11.593**	**11.593**	60.33
	Sub-region(Region)	6	0.4969	0.5062	0.5062	0.6080	0.6543	0.6543	1.13
	Stream(Sub-region(Region))	17	**0.4281**	**0.4183**	**0.4183**	**0.5850**	**0.5752**	**0.5752**	14.79
	error (transect)	51	0.1830	0.1699	0.1699	0.1830	0.1699	0.1699	23.75
Canopy cover	Region	2	1.5926	1.0093	1.1481	1.0000	1.7500	2.3333	2.01
	Sub-region(Region)	6	1.7778	1.8611	2.2222	1.8889	1.3056	1.4444	16.16
	error (stream)	17	0.7843	1.0490	0.9020	1.0196	1.2843	1.1373	81.83

Water temperature (°C)	Region	2	19.3157	10.901	7.2829	11.028	4.9772	2.6661	0.06
	Sub-region(Region)	6	**19.1702**	**14.885**	**13.549**	**21.721**	**16.254**	**14.264**	50.73
	error (stream)	17	2.9585	2.3891	2.3894	4.3181	3.7487	3.7490	49.22
pH	Region	2	**8.3359**	**10.136**	**11.584**	**7.8928**	**9.6452**	**11.059**	94.03
	Sub-region(Region)	6	**0.3742**	**0.1737**	0.1414	**0.2879**	0.1111	0.0963	1.58
	error (stream)	17	0.0479	0.0578	0.0654	0.0388	0.0487	0.0563	4.39
Log10(DO[mg/L])	Region	2	0.0213	0.0305	0.0279	0.0342	**0.0458**	**0.0426**	45.06
	Sub-region(Region)	6	**0.0076**	**0.0077**	**0.0075**	0.0069	0.0059	0.0060	7.48
	error (stream)	17	0.0023	0.0027	0.0025	0.0038	0.0042	0.0039	47.45
Log10(Conductivity[µs/cm])	Region	2	**4.8980**	**5.1683**	**5.7580**	**4.6432**	**4.9025**	**5.4694**	85.32
	Sub-region(Region)	6	**0.3155**	**0.2998**	**0.3029**	**0.3245**	**0.3143**	**0.3288**	4.38
	error (stream)	17	0.0720	0.0735	0.0720	0.0631	0.0646	0.0632	10.30

Table A6.3. ANOVAs for environmental variables according to spatial scale at Churchill, MB (July 15-23, 2012). % variation is, in some cases, adjusted to correct for negative values for variance components (where the mean squares [MS] for the F-ratio denominator at a given spatial level is greater than the numerator).

Variable	Source of variation	df	MS	F-ratio	*p*	% variation
SQRT(Max depth[m])	Sub-region	2	0.0028	0.1113	0.8965	0.00
	Stream(Sub-region)	6	0.0254	4.1484	0.0086	42.60
	error (transect)	18	0.0061			57.40
Log10(Wetted width[m])	Sub-region	2	0.0837	0.4447	0.6605	0.00
	Stream(Sub-region)	6	0.1881	11.1964	<0.001	74.23
	error (transect)	18	0.0168			25.77
SQRT(Max hydraulic head[mm])	Sub-region	2	4.4413	0.9734	0.4304	0.00
	Stream(Sub-region)	6	4.5624	1.0474	0.4284	1.33
	error (transect)	18	4.3561			98.67
Substrate 1	Sub-region	2	2.8148	5.4286	0.0451	35.23
	Stream(Sub-region)	6	0.5185	1.7500	0.1667	23.86
	error (transect)	18	0.2963			40.91
Substrate 2	Sub-region	2	3.5926	1.4478	0.3069	7.46
	Stream(Sub-region)	6	2.4815	3.5263	0.0174	50.00
	error (transect)	18	0.7037			42.54
Woody debris	Sub-region	2	1.4444	4.8750	0.0553	42.47
	Stream(Sub-region)	6	0.2963	4.0000	0.0102	32.88
	error (transect)	18	0.0741			24.66
Detritus	Sub-region	2	0.4444	1.0909	0.3944	1.92
	Stream(Sub-region)	6	0.4074	5.5000	0.0022	63.46
	error (transect)	18	0.0741			34.62
Emergent macrophytes	Sub-region	2	3.3704	2.5278	0.1599	27.64
	Stream(Sub-region)	6	1.3333	9.0000	<0.001	54.27
	error (transect)	18	0.1481			18.09
Submergent free floating	Sub-region	2	1.8148	4.4545	0.0652	32.76
	Stream(Sub-region)	6	0.4074	2.2000	0.0911	28.45
	error (transect)	18	0.1852			38.79
Filamentous algae	Sub-region	2	2.3333	0.8400	0.4768	0.00
	Stream(Sub-region)	6	2.7778	37.5000	<0.001	92.11
	error (transect)	18	0.0741			7.89
Attached algae	Sub-region	2	2.2593	1.5250	0.2914	13.21
	Stream(Sub-region)	6	1.4815	20.0000	<0.001	75.47
	error (transect)	18	0.0741			11.32
Slime or crust algae	Sub-region	2	1.0370	2.1538	0.1972	15.15
	Stream(Sub-region)	6	0.4815	2.6000	0.0541	39.39
	error (transect)	18	0.1852			45.45
Canopy cover	Sub-region	2	0.4444	0.3636	0.7095	0.00
	error (Stream)	6	1.2222			100.00
Water temperature (^{o}C)	Sub-region	2	25.175	6.1320	0.0355	63.11
	error (Stream)	6	4.1055			36.89
pH	Sub-region	2	0.1459	6.3624	0.0329	64.13

	error (Stream)	6	0.0229			35.87
Log10(DO[mg/L])	Sub-region	2	0.0171	4.7060	0.0590	55.26
	error (Stream)	6	0.0036			44.74
Log10(Conductivity[μs/cm])	Sub-region	2	0.4483	5.4956	0.0440	59.98
	error (Stream)	6	0.0816			40.02

Table A6.4. ANOVAs for environmental variables according to spatial scale at Algonquin Provincial Park, ON (June 26-July 13, 2011). 6 ANOVAs were run with different combinations of streams and regions. A single dummy stream was used. Bolded mean squares (MS) indicate a significant F-ratio.

Variable	Source of variation	df	MS1	MS2	MS3	MS4	MS5	MS6	Average % variation
SQRT(Max depth[m])	Sub-region	2	0.0244	0.5159	0.3191	0.1139	0.2842	0.1520	7.10
	Stream(Sub-region)	5	**0.1786**	**0.1788**	**0.1997**	**0.2498**	**0.2500**	**0.2708**	57.53
	error (transect)	15	0.0356	0.0350	0.0398	0.0339	0.0333	0.0380	35.36
Log10(Wetted width[m])	Sub-region	2	0.6241	0.1883	0.6627	2.1492	0.1821	0.1010	14.85
	Stream(Sub-region)	5	**0.3514**	**0.3943**	**0.4735**	**0.3874**	**0.4302**	**0.5094**	79.29
	error (transect)	15	0.0089	0.0087	0.0087	0.0097	0.0095	0.0095	5.86
SQRT(Max hydraulic head[mm])	Sub-region	2	**143.45**	3.1112	14.998	**235.99**	13.410	0.5876	22.37
	Stream(Sub-region)	5	**16.735**	**16.732**	**16.264**	**15.177**	**15.174**	**14.706**	32.52
	error (transect)	15	4.8101	5.4542	6.3760	2.3955	3.0397	3.9615	45.11
Substrate 1	Sub-region	2	**15.694**	4.3611	8.1111	17.806	5.1389	5.8889	20.38
	Stream(Sub-region)	5	**2.5222**	2.5222	2.4889	4.2111	4.2111	4.1778	34.60
	error (transect)	15	0.8000	1.1556	1.0667	1.6889	2.0444	1.9556	45.02
Substrate 2	Sub-region	2	8.1389	13.472	8.1389	2.4722	1.1389	2.4722	13.42
	Stream(Sub-region)	5	**3.1444**	**3.1444**	**5.5444**	**1.8111**	**1.8111**	**4.2111**	54.44
	error (transect)	15	0.6667	0.7556	0.9333	0.2667	0.3556	0.5333	32.15
Woody debris	Sub-region	2	0.4444	1.7778	0.6944	1.8889	0.5556	0.1389	7.22
	Stream(Sub-region)	5	0.3556	0.3556	**1.0556**	0.5778	0.5778	**1.2778**	23.32
	error (transect)	15	0.4889	0.3556	0.3556	0.5333	0.4000	0.4000	69.46
Detritus	Sub-region	2	**1.0000**	1.0000	**0.2500**	**2.5000**	2.5000	**1.2500**	38.21
	Stream(Sub-region)	5	**0.1333**	**0.5333**	0.0333	0.2667	**0.6667**	0.1667	29.90
	error (transect)	15	0.0444	0.0444	0.0444	0.1333	0.1333	0.1333	31.89
Emergent macrophytes	Sub-region	2	1.6944	**6.9444**	**6.9444**	0.6944	**2.9444**	**2.9444**	40.92
	Stream(Sub-region)	5	0.6556	0.3556	0.3556	0.6556	0.3556	0.3556	21.55
	error (transect)	15	0.3556	0.2222	0.2222	0.3556	0.2222	0.2222	37.53
Submergent free floating	Sub-region	2	**4.0556**	1.1389	0.7222	4.0556	1.1389	0.7222	22.25
	Stream(Sub-region)	5	0.1778	0.2111	0.3111	**0.9778**	**1.0111**	**1.1111**	34.99
	error (transect)	15	0.2667	0.3111	0.3111	0.1333	0.1778	0.1778	42.77
Filamentous algae	Sub-region	2	0.0278	0.2778	0.0278	0.1389	0.2222	0.1389	0.68
	Stream(Sub-region)	5	0.2556	0.2222	0.2556	0.2111	0.1778	0.2111	8.16
	error (transect)	15	0.2667	0.2222	0.2667	0.2222	0.1778	0.2222	91.16
Attached algae	Sub-region	2	0.1111	1.1111	0.3611	0.5556	2.8889	0.1389	10.42
	Stream(Sub-region)	5	0.4889	**1.0222**	**0.5222**	0.1778	**0.7111**	0.2111	39.79
	error (transect)	15	0.1778	0.1333	0.1333	0.1333	0.0889	0.0889	49.79
Slime or crust algae	Sub-region	2	0.1389	0.2222	0.2222	0.4722	0.8889	0.8889	3.70
	Stream(Sub-region)	5	0.2111	**0.1778**	**0.1778**	**0.7444**	**0.7111**	**0.7111**	60.41
	error (transect)	15	0.0889	0.0444	0.0444	0.0889	0.0444	0.0444	35.89
Canopy cover	Sub-region	2	1.7778	2.0278	3.1111	2.1111	0.3611	0.7778	19.32
	error (Stream)	5	0.5333	1.4333	0.9333	1.3333	2.2333	1.7333	80.68
Water temperature (^{o}C)	Sub-region	2	20.248	7.3911	3.3836	27.901	11.498	5.5303	38.35
	error (Stream)	5	3.8427	1.9067	1.9077	8.4653	6.5293	6.5303	68.04
pH	Sub-region	2	**0.8547**	0.2532	0.1563	**0.5956**	0.0652	0.0209	21.10
	error (Stream)	5	0.1011	0.1347	0.1604	0.0701	0.1038	0.1294	78.90
Log10(DO[mg/L])	Sub-region	2	0.0056	0.0060	0.0054	0.0036	0.0006	0.0009	6.81
	error (Stream)	5	0.0033	0.0046	0.0038	0.0084	0.0096	0.0089	93.19
Log10(Conductivity[µs/cm])	Sub-region	2	0.1694	0.1224	0.1317	0.1964	0.1660	0.2094	39.56
	error (Stream)	5	0.0693	0.0744	0.0695	0.0391	0.0442	0.0393	60.44

Table A6.5. ANOVAs for environmental variables according to spatial scale at Pennsylvania, USA (June 17-19, 2013). % variation is, in some cases, adjusted to correct for negative values for variance components (where the mean squares [MS] for the F-ratio denominator at a given spatial level is greater than the numerator).

Variable	Source of variation	df	MS	F-ratio	*p*	% variation
SQRT(Max depth[m])	Sub-region	2	0.0376	2.2858	0.1828	15.22
	Stream(Sub-region)	6	0.0165	2.1624	0.0957	35.51
	error (transect)	18	0.0076			49.27
Log10(Wetted width[m])	Sub-region	2	0.0065	0.1700	0.8476	0.00
	Stream(Sub-region)	6	0.0381	7.5670	<0.001	62.49
	error (transect)	18	0.0050			37.51
SQRT(Max hydraulic head[mm])	Sub-region	2	63.0908	37.5619	<0.001	66.64
	Stream(Sub-region)	6	1.6797	0.5881	0.7356	5.47
	error (transect)	18	2.8561			27.89
Substrate 1	Sub-region	2	2.7778	1.8293	0.2397	14.85
	Stream(Sub-region)	6	1.5185	5.1250	0.0031	53.71
	error (transect)	18	0.2963			31.44
Substrate 2	Sub-region	2	0.2778	0.1339	0.8772	0.00
	Stream(Sub-region)	6	2.0741	2.3333	0.0764	21.63
	error (transect)	18	0.8889			78.37
Woody debris	Sub-region	2	4.0556	1.4698	0.3023	11.53
	Stream(Sub-region)	6	2.7593	14.9000	<0.001	73.64
	error (transect)	18	0.1852			14.83
Detritus	Sub-region	2	6.5000	9.0000	0.0156	59.09
	Stream(Sub-region)	6	0.7222	3.5455	0.0170	22.16
	error (transect)	18	0.2037			18.75
Emergent macrophytes	Sub-region	2	1.6111	5.4375	0.0449	41.04
	Stream(Sub-region)	6	0.2963	2.6667	0.0497	27.75
	error (transect)	18	0.1111			31.21
Submergent free floating	Sub-region	2	7.6111	18.6818	0.0026	79.23
	Stream(Sub-region)	6	0.4074	5.5000	0.0022	13.44
	error (transect)	18	0.0741			7.33
Filamentous algae	Sub-region	2	4.5556	3.1948	0.1136	38.76
	Stream(Sub-region)	6	1.4259	19.2500	<0.001	52.98
	error (transect)	18	0.0741			8.26
Attached algae	Sub-region	2	1.3889	5.3571	0.0463	50.41
	Stream(Sub-region)	6	0.2593	7.0000	<0.001	34.71
	error (transect)	18	0.0370			14.88
Slime or crust algae	Sub-region	2	2.5000	2.5000	0.1623	20.00
	Stream(Sub-region)	6	1.0000	3.0000	0.0327	40.00
	error (transect)	18	0.3333			40.00
Canopy cover	Sub-region	2	3.1111	5.6000	0.0424	60.53
	error (Stream)	6	0.5556			39.47
Water temperature (^{o}C)	Sub-region	2	12.0883	11.2467	0.0093	77.35
	error (Stream)	6	1.0748			22.65
pH	Sub-region	2	0.1221	4.2569	0.0707	52.05

	error (Stream)	6	0.0287			47.95
Log10(DO[mg/L])	Sub-region	2	0.0000	0.2863	0.7608	0.00
	error (Stream)	6	0.0001			100.00
Log10(Conductivity[µs/cm])	Sub-region	2	0.3287	5.0857	0.0511	57.66
	error (Stream)	6	0.0646			42.34

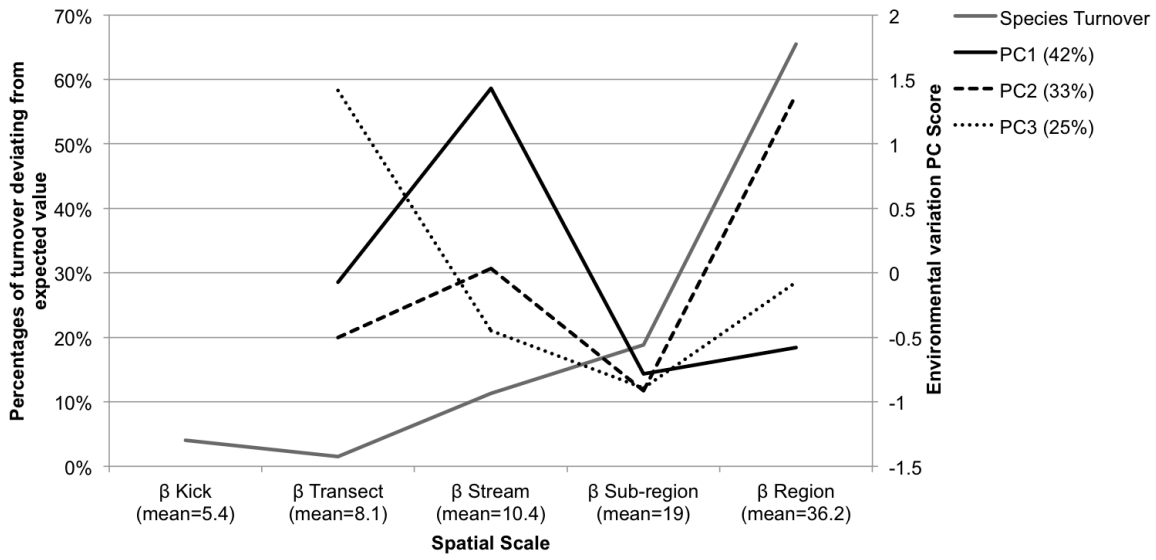

Figure A6.1. Turnover of stream larval caddisfly species and environmental variation according to sampled spatial scales. Regions sampled include Churchill (MB, Canada: July 15-23, 2012), Algonquin Provincial Park (ON, Canada: June 26-July 13, 2011) and Pennsylvania (USA: June 17-19, 2013). Mean number of species is indicated in parentheses for each spatial scale.

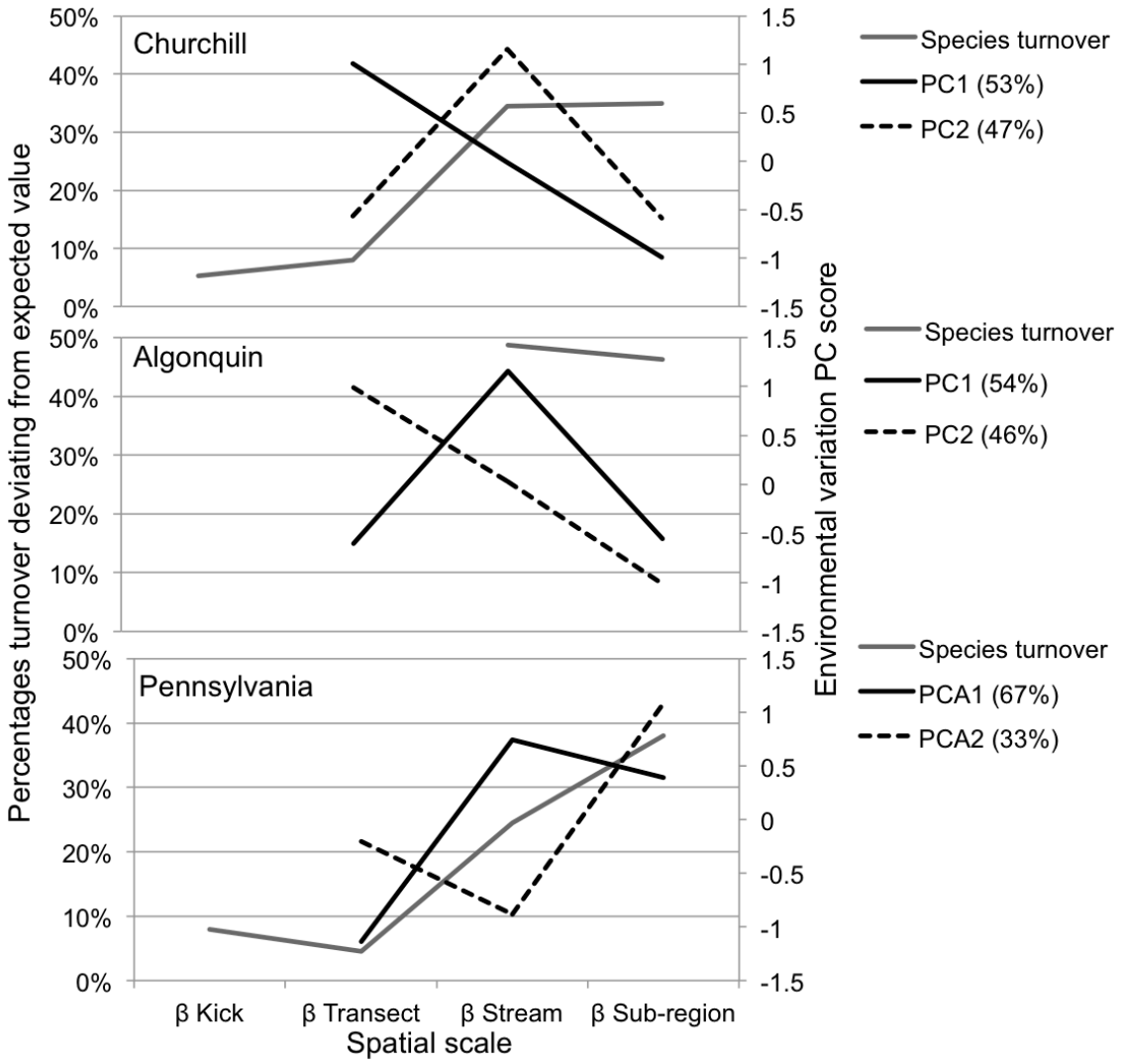

Figure A6.2. Turnover of stream larval caddisfly species and environmental variation according to sampled spatial scales at each of the regions. Dates for sampled regions are: Churchill (MB, Canada), July 15-23, 2012; Algonquin Provincial Park (ON, Canada), June 26-July 13, 2011; Pennsylvania (USA), June 17-19, 2013.

Appendix VII: Values from partition and checkerboard analyses

Result values for all analyses, including diversity components and checkerboards, are presented here as an appendix.

Table A7.1. Additive richness components for larval summer caddisfly (order: Trichoptera) species at each region. Regions are Churchill (MB, Canada; July 15-23, 2012), Algonquin Provincial Park (ON, Canada; June 26-July 13, 2011) and Pennsylvania (USA; June 17-19, 2013). Upper and lower expected values of diversity components were calculated based on 5000 individual-based randomization iterations of the input data. P-values less than 0.025 indicate an observed value significantly greater than the expected value.

Region (n=species)	Spatial scale	α or β	Obs	SE	Exp	Lower limit	Upper limit	*p*(obs> exp)
Churchill (n=28)	Kick (n=75)	α	3.43	0.22	10.52	10.04	11.05	>0.999
		β	1.76		4.49	3.75	5.41	>0.999
	Transect (n=27)	α	5.19	0.46	15.01	14.15	16.07	>0.999
		β	3.14		6.69	5.4	8.22	>0.999
	Stream (n=9)	α	8.33	0.90	21.7	20	23.33	>0.999
		β	7.67		5.04	3.33	6.89	<0.001
	Sub-region (n=3)	α	16.00	2.65	26.73	24.67	28	>0.999
		β	12.00		1.27	0	3.33	<0.001
Algonquin (n=26.5)	Stream (n=8)	α	5.02	1.09	12.35	10.92	13.73	>0.999
		β	6.43		6.44	4.37	8.64	0.449
	Sub-region (n=3)	α	11.44	1.85	18.79	16.34	21.22	>0.999
		β	15.06		7.71	5.28	10.17	<0.001
Pennsylvania (n=54)	Kick (n=54)	α	8.15	0.37	22.16	21.35	22.93	>0.999
		β	2.81		5.83	4.78	6.74	>0.999
	Transect (n=27)	α	10.96	0.54	27.99	26.78	29.22	>0.999
		β	6.26		13.45	11.78	15.55	>0.999
	Stream (n=9)	α	17.22	0.92	41.44	39.44	43.67	>0.999
		β	12.45		9.61	7.0	12.23	<0.001
	Sub-region (n=3)	α	29.67	3.18	51.05	48	53.33	>0.999
		β	24.33		2.95	0.67	6.0	<0.001

Table A7.2. Additive richness components for larval summer caddisflies (order: Trichoptera) across 2 regions, according to taxonomic level. Regions are Churchill (MB, Canada) and Pennsylvania (USA; Summers 2011-13). Upper and lower expected values of diversity components were calculated based on 5000 individual-based randomization iterations of the input data. P-values less than 0.025 indicate an observed value significantly greater than the expected value.

Analysis (n=species)	**Spatial scale**	**α or β**	**Obs**	**SE**	**Exp**	**Lower limit**	**Upper limit**	***p*(obs>exp)**
Churchill and Pennsylvania								
Species (n=81)	Kick	α	5.40	0.29	22.58	22.03	23.11	>0.999
	(n=129)	β	2.67		9.96	9.19	10.86	>0.999
	Transect	α	8.07	0.53	32.54	31.46	33.50	>0.999
	(n=54)	β	4.71		17.85	16.37	19.27	>0.999
	Stream	α	12.78	1.25	50.39	48.33	52.11	>0.999
	(n=18)	β	10.05		16.99	14.50	19.72	>0.999
	Sub-region	α	22.83	3.57	67.38	64.50	70.50	>0.999
	(n=6)	β	18.17		10.68	6.83	14.17	<0.001
	Region	α	41.00	13.0	78.06	74.50	81.00	>0.999
	(n=2)	β	40.00		2.94	0.00	6.50	<0.001
Genera (n=30)	Kick	α	3.98	0.18	12.32	11.98	12.71	>0.999
	(n=129)	β	1.58		3.87	3.31	4.37	>0.999
	Transect	α	5.56	0.28	16.19	15.56	16.89	>0.999
	(n=54)	β	2.38		6.07	5.22	6.98	>0.999
	Stream	α	7.94	0.55	22.26	21.22	23.39	>0.999
	(n=18)	β	4.73		4.78	3.39	6.11	>0.540
	Sub-region	α	12.67	1.33	27.04	25.33	28.33	>0.999
	(n=6)	β	4.83		2.44	0.83	4.00	<0.001
	Region	α	17.50	3.50	29.48	27.50	30.00	>0.999
	(n=2)	β	12.50		0.52	0.00	2.50	<0.001
Family (n=15)	Kick	α	3.16	0.13	7.83	7.53	8.11	>0.999
	(n=129)	β	1.19		1.89	1.56	2.27	>0.999
	Transect	α	4.35	0.20	9.71	9.30	10.22	>0.999
	(n=54)	β	1.76		2.56	1.97	3.22	>0.999
	Stream	α	6.11	0.37	12.28	11.56	12.89	>0.999
	(n=18)	β	2.89		1.70	0.73	2.67	>0.001
	Sub-region	α	9.00	0.82	13.98	13.17	14.67	>0.999
	(n=6)	β	2.50		0.83	0.00	1.67	<0.001
	Region	α	11.50	2.50	14.81	14.00	15.00	>0.999
	(n=2)	β	3.50		0.19	0.00	1.00	<0.001
Churchill								
Species (n=28)	Kick	α	3.43	0.22	10.52	10.04	11.05	>0.999
	(n=75)	β	1.76		4.49	3.75	5.41	>0.999
	Transect	α	5.19	0.46	15.01	14.15	16.07	>0.999
	(n=27)	β	3.14		6.69	5.4	8.22	>0.999
	Stream	α	8.33	0.90	21.7	20	23.33	>0.999
	(n=9)	β	7.67		5.04	3.33	6.89	<0.001
	Sub-region	α	16.00	2.65	26.73	24.67	28	>0.999
	(n=3)	β	12.00		1.27	0	3.33	<0.001
Genera (n=14)	Kick	α	2.84	0.15	7.54	7.15	7.96	>0.999
	(n=75)	β	1.27		2.16	1.61	2.75	>0.999

	Transect	α	4.11	0.28	9.7	9.04	10.37	>0.999
	(n=27)	β	1.78		2.83	2.08	3.63	>0.999
	Stream	α	5.89	0.39	12.53	11.33	13.44	>0.999
	(n=9)	β	4.11		1.35	0.44	2.67	<0.001
	Sub-region	α	10	1.0	13.88	12.67	14	>0.999
	(n=3)	β	4		0.12	0	1.33	<0.001
Family (n=9)	Kick	α	2.48	0.13	5.99	5.67	6.31	>0.999
	(n=75)	β	1.08		1.21	0.84	1.62	0.860
	Transect	α	3.56	0.24	7.2	6.7	7.74	>0.999
	(n=27)	β	1.33		1.4	0.81	1.93	0.681
	Stream	α	4.89	0.35	8.6	7.89	9.0	>0.999
	(n=9)	β	2.44		0.39	0	1.0	<0.001
	Sub-region	α	7.33	0.67	9.0	8.67	9.0	>0.999
	(n=3)	β	1.67		0	0	0.33	<0.001
Pennsylvania								
Species (n=54)	Kick	α	8.15	0.37	22.16	21.35	22.93	>0.999
	(n=54)	β	2.81		5.83	4.78	6.74	>0.999
	Transect	α	10.96	0.54	27.99	26.78	29.22	>0.999
	(n=27)	β	6.26		13.45	11.78	15.55	>0.999
	Stream	α	17.22	0.92	41.44	39.44	43.67	>0.999
	(n=9)	β	12.45		9.61	7.0	12.23	<0.001
	Sub-region	α	29.67	3.18	51.05	48	53.33	>0.999
	(n=3)	β	24.33		2.95	0.67	6.0	<0.001
Genera (n=21)	Kick	α	5.57	0.23	10.59	10.13	11.11	>0.999
	(n=54)	β	1.43		2.13	1.68	2.68	>0.999
	Transect	α	7.0	0.30	12.72	11.96	13.52	>0.999
	(n=27)	β	3.0		4.65	3.67	5.63	>0.999
	Stream	α	10.0	0.29	17.37	16.11	18.44	>0.999
	(n=9)	β	5.33		2.92	1.44	4.34	<0.001
	Sub-region	α	15.33	0.88	10.29	19.0	21.0	>0.999
	(n=3)	β	5.67		0.71	0	2.0	<0.001
Family (n=14)	Kick	α	4.09	0.18	7.02	6.52	7.5	>0.999
	(n=54)	β	1.06		1.51	1.11	1.95	>0.999
	Transect	α	5.15	0.22	8.53	7.85	9.19	>0.999
	(n=27)	β	2.18		3.23	2.3	4.19	>0.999
	Stream	α	7.33	0.29	11.76	10.78	12.67	>0.999
	(n=9)	β	3.34		1.8	0.56	3.11	<0.001
	Sub-region	α	10.67	0.33	13.56	12.33	14.0	>0.999
	(n=3)	β	3.33		0.44	0	1.67	<0.001

Table A7.3. Additive richness components for larval summer stream insect species at White Clay Creek (PA, USA; June 20, 2013). Upper and lower expected values of diversity components were calculated based on 5000 individual-based randomization iterations of the input data. P-values less than 0.025 indicate an observed value significantly greater than the expected value.

Analysis	Spatial scale	α or β	Obs	SE	Exp	Lower limit	Upper limit	*p*(obs>exp)
Trichoptera (n=25)	Rock	α	7.23	0.34	19.95	19.20	20.60	>0.999
	(n=30)	β	2.30		2.49	1.77	3.17	0.820
	Transect	α	9.53	0.59	22.44	21.47	23.33	>0.999
	(n=15)	β	4.07		2.12	1.07	3.07	<0.001
	Station	α	13.60	0.81	24.57	23.40	25.00	>0.999
	(n=5)	β	6.40		0.35	0.00	1.60	<0.001
	Arm	α	20.00	4.00	24.91	24.00	25.00	>0.999
	(n=2)	β	5.00		0.09	0.00	1.00	<0.001
Hydropsychidae (n=12)	Rock	α	4.10	0.18	10.37	10.0	10.77	>0.999
	(n=30)	β	0.90		0.94	0.46	1.44	0.602
	Transect	α	5.00	0.31	11.30	10.73	11.8	>0.999
	(n=15)	β	2.00		0.64	0.20	1.20	<0.001
	Station	α	7.00	0.32	11.94	11.4	12.0	>0.999
	(n=5)	β	3.50		0.06	0.00	0.60	<0.001
	Arm	α	10.50	0.50	12.0	11.5	12.0	>0.999
	(n=2)	β	1.50		0.00	0.00	0.50	<0.001
Coleoptera (n=8)	Rock	α	2.50	0.25	5.82	5.31	6.27	>0.999
	(n=28)	β	1.07		0.89	0.40	1.33	0.084
	Transect	α	3.57	0.31	6.71	6.07	7.36	>0.999
	(n=14)	β	2.03		0.95	0.20	1.64	<0.001
	Station	α	5.60	0.50	7.66	6.80	8.00	>0.999
	(n=5)	β	1.40		0.34	0.00	1.20	<0.001
	Arm	α	7.00	1.00	8.00	7.5	8.00	>0.999
	(n=2)	β	1.00		0.00	0.00	0.5	<0.001
Simuliidae (n=7)	Rock	α	2.75	0.18	5.77	5.36	6.18	>0.999
	(n=29)	β	0.52		0.55	0.23	0.99	0.5728
	Transect	α	3.27	0.31	6.31	5.87	6.80	>0.999
	(n=15)	β	1.53		0.54	0.07	1.07	<0.001
	Station	α	4.80	0.73	6.86	6.20	7.00	>0.999
	(n=5)	β	1.70		0.14	0.00	0.80	<0.001
	Arm	α	6.50	0.50	7.00	7.00	7.00	>0.999
	(n=2)	β	0.50		0.00	0.00	0.00	<0.001

Table A7.4. Checkerboard results for summer larval species of orders Trichoptera and Coleoptera and family Simuliidae at White Clay Creek (PA, USA; June 20, 2013). Note, for Simuliidae in the sites and species fixed analysis, sample size for stations and species was insufficient to produce expected values, and therefore these scales are excluded.

Analysis (n=species)	**Spatial scale**	**n**	**Obs. C-score**	**Exp. C-score**	**SD**	**Standardiz-ed C-score**	***p*(obs< exp)**	***p*(obs> exp)**
Null model: Sites and species equiprobable								
Trichoptera	Rock	30	10.46	25.41	0.74	-20.20	<0.001	>0.999
(n=25)	Transect	15	14.17	33.54	1.67	-11.57	<0.001	>0.999
	Station	5	25.50	37.51	5.20	-2.31	0.014	0.987
Hydropsychidae	Rock	30	2.48	6.79	0.36	-11.95	<0.001	>0.999
(n=12)	Transect	15	3.25	7.90	0.67	-6.88	<0.001	>0.999
	Station	5	6.40	8.05	1.76	-0.94	0.184	0.830
Coleoptera (n=8)	Rock	30	1.02	2.58	0.21	-7.56	<0.001	>0.999
	Transect	15	1.42	3.44	0.40	-5.12	<0.001	>0.999
	Station	5	0.70	2.55	0.76	-2.42	0.013	0.991
Simuliidae (n=7)	Rock	30	0.844	2.47	0.19	-8.41	<0.001	>0.999
	Transect	15	1.00	2.71	0.34	-4.97	<0.001	>0.999
	Station	5	0.00	2.03	0.67	-3.02	0.001	>0.999
Hydro+Sim	Rock	30	6.70	17.8	0.63	-17.25	<0.001	>0.999
(n=19)	Transect	15	8.15	20.81	1.25	-10.10	<0.001	>0.999
	Station	5	10.2	19.28	3.16	-2.88	0.004	0.997
All (n=40)	Rock	30	28.63	68.64	1.4	-28.58	<0.001	>0.999
	Transect	15	36.26	90.62	3.30	-16.48	<0.001	>0.999
	Station	5	45.00	90.65	9.61	-4.75	<0.001	>0.999
Null model: Sites fixed, species equiprobable								
Trichoptera	Rock	30	10.46	26.08	0.51	-30.40	<0.001	>0.999
(n=25)	Transect	15	14.17	33.83	1.36	-14.45	<0.001	>0.999
	Station	5	25.50	38.67	4.96	-2.66	0.006	0.995
Hydropsychidae	Rock	30	2.48	7.51	0.21	-24.31	<0.001	>0.999
(n=12)	Transect	15	3.25	8.58	0.49	-10.84	<0.001	>0.999
	Station	5	6.40	9.06	1.63	-1.63	0.066	0.942
Coleoptera (n=8)	Rock	30	1.02	2.65	0.12	-13.81	<0.001	>0.999
	Transect	15	1.42	3.79	0.29	-8.06	<0.001	>0.999
	Station	5	0.70	2.88	0.65	-3.36	0.001	0.999
Simuliidae (n=7)	Rock	30	0.844	2.77	0.11	-17.07	<0.001	>0.999
	Transect	15	1.00	3.13	0.24	-8.99	<0.001	>0.999
	Station	5	0.00	1.54	0.53	-2.92	0.001	>0.999
Hydro+Sim	Rock	30	6.79	18.8	0.42	-28.66	<0.001	>0.999
(n=19)	Transect	15	8.152	21.89	0.99	-13.82	<0.001	>0.999
	Station	5	10.2	19.18	2.92	-3.08	0.002	0.998
All (n=40)	Rock	30	28.63	68.58	1.07	-37.25	<0.001	>0.999
	Transect	15	36.26	89.68	2.86	-18.68	<0.001	>0.999
	Station	5	45.00	91.07	9.27	-4.97	<0.001	>0.999
Null model: Sites equiprobable, species fixed								
Trichoptera	Rock	30	10.46	10.21	0.39	0.64	0.732	0.271
(n=25)	Transect	15	14.17	14.31	0.72	-0.19	0.392	0.613
	Station	5	25.50	23.76	1.84	0.95	0.864	0.155
Hydropsychidae	Rock	30	2.48	2.10	0.18	2.19	0.993	0.007
(n=12)	Transect	15	3.25	2.88	0.31	1.18	0.894	0.114

	Station	5	6.40	5.15	0.88	1.42	0.968	0.046
Coleoptera (n=8)	Rock	30	1.02	1.29	0.12	-2.29	0.015	0.985
	Transect	15	1.42	1.60	0.25	-0.75	0.217	0.783
	Station	5	0.70	0.60	0.31	0.32	0.638	0.503
Simuliidae (n=7)	Rock	30	0.844	0.91	0.12	-0.54	0.289	0.711
	Transect	15	1.00	0.85	0.19	0.78	0.776	0.240
	Station	5	0.00	0.74	0.33	-2.24	0.023	>0.999
Hydro+Sim (n=19)	Rock	30	6.79	6.24	0.32	1.71	0.970	0.031
	Transect	15	8.152	7.45	0.53	1.33	0.926	0.077
	Station	5	10.2	10.71	1.24	-0.40	0.300	0.726
All (n=40)	Rock	30	28.63	29.53	0.70	-1.29	0.105	0.896
	Transect	15	36.26	38.22	1.20	-1.63	0.064	0.936
	Station	5	45.00	45.91	2.53	-0.36	0.306	0.706
Null model: Sites and species fixed								
Trichoptera (n=25)	Rock	30	10.46	10.15	0.12	2.53	0.989	0.011
	Transect	15	14.17	13.61	0.20	2.74	0.991	0.009
	Station	5	25.50	24.63	0.65	1.34	0.910	0.109
Hydropsychidae (n=12)	Rock	30	2.48	2.39	0.05	2.10	0.976	0.027
	Transect	15	3.25	3.06	0.07	2.65	0.989	0.014
	Station	5	6.40	6.18	0.32	0.68	0.827	0.242
Coleoptera (n=8)	Rock	30	1.02	0.99	0.04	0.63	0.776	0.245
	Transect	15	1.42	1.42	0.07	0.03	0.604	0.459
	Station	5	0.70	0.72	0.13	-0.15	0.785	0.717
Simuliidae (n=7)	Rock	30	0.844	0.83	0.03	0.36	0.699	0.334
	Transect	15	1.00	0.99	0.06	0.15	0.655	0.402
	Station	5	0.00	N/A	N/A	N/A	N/A	N/A
Hydro+Sim (n=19)	Rock	30	6.79	6.49	0.08	3.46	0.996	0.004
	Transect	15	8.152	7.69	0.14	3.41	0.997	0.004
	Station	5	10.2	10.15	0.40	0.38	0.737	0.334
All (n=40)	Rock	30	28.63	28.11	0.21	2.49	0.986	0.015
	Transect	15	36.26	35.55	0.35	2.04	0.963	0.039
	Station	5	45.00	45.11	0.90	-0.12	0.551	0.497

Appendix VIII: Multiplicative diversity partitioning at White Clay Creek

In addition to additive diversity partitioning of species diversity at White Clay Creek (PA, USA; June 20, 2013), multiplicative diversity partitioning was used to investigate variation in the distribution of biodiversity. Multiplicative partitioning of species diversity was used to calculate % turnover at each of the sampled spatial scales. The total species pool for White Clay Creek was defined according to the following equation:

White Clay Creek assemblages $\gamma=[[[\alpha_{rock} \times \beta_{rock}] \times \beta_{transect}] \times \beta_{station}] \times \beta_{Arm}$

In order to assess departures from random structure, null distributions in the diversity components were calculated using PARTITION3.0. Individual specimens based on the input data were randomly allocated to each rock (individual-based randomization), simulating distributions in the data according to random chance. The null distribution was used to assess significance of departures in the observed values, represented as a two-tailed p-value for whether or not the observed value was greater than the expected value (e.g. p=0.025; significantly small values are represented by p=0.975). Five thousand randomization iterations of the input data were performed for each partition; partition analyses typically perform between 1000 and 10,000 randomizations, though no basis has been established for an ideal number of iterations. Separate analyses were performed for several taxonomic groups: Trichoptera, Hydropsychidae, Coleoptera, and Simuliidae.

In order to assess variation in the distribution of species according to rarity or commonness, values of α and β were calculated to different orders of diversity (q), which determines the sensitivity of the diversity value to rare (low abundance) and common (high abundance) species. α and γ diversity to the order q were calculated according to the equation ${}_qD=(\Sigma_{(s,\, i=1)} p_i^q)^{1/(1-q)}$, where D is the diversity index and p is the proportional abundance (i.e. density/m^2) of a species in a sample for all species i=1 to s. Multiplicative β was subsequently calculated as ${}^qD_\gamma/{}^qD_\alpha$. The value of q was set to 0 (species richness), 0.999 (approaching 1, given 1 creates a division by 0 in the above equation; 0.999 is equivalent to Shannon's diversity index), and 2 (Simpson's diversity index, Jost 2006), emphasizing rare species (richness), common species (Simpson), and neither (Shannon).

Values of β were converted into percentages of turnover deviating from expected values, isolating the amount of turnover that must be due to biotic interactions. Values of multiplicative β represent the number of compositionally unique samples in a set of samples (Whittaker 1972). Subtracting 1 from multiplicative β changes this value to the number of turnovers amongst a set of replicates. Because the number of replicates defines the upper limit to the number of turnovers, the metric can further be presented as a percentage for across-scale comparisons.

Results and discussion

Significant turnover of species was detected at White Clay Creek; however, the degree of turnover varied according to taxon, spatial level, and order of diversity. Species turnover was consistently significant for all taxa at all spatial levels at q=0; Coleoptera had the greatest turnover at the rock level (37%), with decreasing turnover with increasing spatial scale (Fig. A8.1). Trichoptera had the greatest turnover at the station level (33%), with modest turnover at the rock and transect levels (20 and 17% respectively), with Hydropsychidae showing a similar pattern (Fig. A8.1). Simuliidae also had the greatest deterministic turnover at the station level (24%) and lowest turnover at the stream arm level (8%, Fig. A8.1).

The same trends occurred for Trichoptera, Hydropsychidae, and Coleoptera at q=0.999 and 2, but not for Simuliidae. At q=0.999 and 2, species turnover was generally less than at q=0, and any turnover that was 2% or less was not significantly greater than expected by chance. Coleoptera again had the greatest turnover at the rock level (31 and 27% for q=0.999 and 2 respectively), which tended to decrease with increasing spatial scale (except for a modest increase from 0 to 3% at the station to arm level for q=2; Fig. A8.1). Trichoptera showed the greatest turnover at the station and stream arm levels (15 and 16% respectively at q=0.999, 13 and 11% respectively at q=2), with modest turnover at the rock and transect levels (8 and 6% respectively at q=0.999, 6 and 5% respectively at q=2; Fig. A8.1). Hydropsychidae also showed greatest turnover at the station and stream arm levels (18 and 16% respectively at q=0.999, 17 and 11% respectively at q=2), with low turnover at the rock and transect levels (3 and 5% respectively at q=0.999, 4 and 1% respectively at q=2; Fig. A8.1). Simuliidae behaved oppositely at q=0.999 and 2 compared to q=0. Turnover was not significantly greater than expected by chance at the

station level for q=0.999 and 2, while high levels of turnover were detected at the stream arm level (29 and 37% respectively; Fig. A8.1). Turnover decreased from the rock to station level for Simuliidae at q=0.999 and 2, with 11% and 9% respective turnover at the rock level (Fig. A8.1).

The results indicate important turnover of species occurs at all within stream spatial scales sampled, but that the degree of turnover among sampling units differs according to scale and taxa. For instance, Coleoptera showed the greatest turnover at the rock level, indicating important differences in species composition among rocks close to one another (e.g. within transects; Fig. A8.1). Trichoptera on the other hand, demonstrated important turnover among units at larger spatial extents, particularly between stations (e.g. kilometer scale; Fig. A8.1). Simuliidae species were structured at the station level, however, more common species (q=0.999 and 2) were structured at larger spatial extents (e.g. arm-level; Fig. A8.1). Because the influence of environmental filtering and dispersal limitations are minimized by sampling at small spatial scales (i.e. within White Clay Creek), it is likely biotic interactions are structuring these taxa at differing spatial scales. These results demonstrate that if biotic interactions are important in structuring the distribution of species at small spatial scales, the resulting structure in biodiversity can nonetheless range from very fine-scale (i.e. rocks) to coarser (i.e. kilometers) spatial levels.

Table A8.1. Multiplicative partition results for summer larval species of orders Trichoptera and Coleoptera and family Simuliidae at White Clay Creek (PA, USA; June 20, 2013). Presented are values of α and β to the diversity order of 0, 0.999 and 2. Observed values are presented as absolute values of β/percentages turnover deviating from expected value. Expected values of diversity components were calculated based on 5000 individual-randomization iterations of the input data. Observed values significantly greater than expected values are bolded (*p*-value of ≤0.025).

Analysis (n=species)	**Spatial scale**	**α or β**	**q=0**		**q=0.999**		**q=2**	
			Obs.	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**
Trichoptera (n=25)	Rock (n=30)	α	7.23	19.95	4.30	6.83	3.29	4.81
		β	**1.32/20**	1.12	**1.09/8**	1.01	**1.07/6**	1.01
	Transect (n=15)	α	9.53	22.44	4.70	6.92	3.53	4.84
		β	**1.43/17**	1.09	**1.12/6**	1.01	**1.10/5**	1.00
	Station (n=5)	α	13.60	24.57	5.27	6.96	3.89	4.85
		β	**1.47/33**	1.01	**1.21/15**	1.00	**1.18/13**	1.00
	Arm (n=2)	α	20.00	24.91	6.36	6.97	4.60	4.85
		β	**1.25/25**	1.00	**1.16/16**	1.00	**1.11/11**	1.00
Hydropsychidae (n=12)	Rock (n=30)	α	4.10	10.37	3.06	4.59	2.56	3.54
		β	**1.22/13**	1.09	**1.04/3**	1.01	1.02/1	1.01
	Transect (n=15)	α	5.00	11.3	3.18	4.64	2.60	3.57
		β	**1.4/17**	1.06	**1.1/5**	1.00	**1.08/4**	1.00
	Station (n=5)	α	7.00	11.94	3.50	4.66	2.81	3.57
		β	**1.5/36**	1.00	**1.25/18**	1.00	**1.24/17**	1.00
	Arm (n=2)	α	10.5	12	4.37	4.66	3.49	3.57
		β	**1.14/14**	1.00	**1.16/16**	1.00	**1.11/11**	1.00
Coleoptera (n=8)	Rock (n=28)	α	2.50	5.82	1.87	3.03	1.60	2.20
		β	**1.43/37**	1.15	**1.28/31**	1.05	**1.22/27**	1.02
	Transect (n=14)	α	3.57	6.71	2.40	3.19	1.95	2.25
		β	**1.57/25**	1.14	**1.30/16**	1.03	**1.17/9**	1.01
	Station (n=5)	α	5.60	7.66	3.11	3.28	2.29	2.27
		β	**1.25/15**	1.04	**1.05/3**	1.01	1.00/0	1.00
	Arm (n=2)	α	7.00	8.00	3.28	3.30	2.29	2.28
		β	**1.14/14**	1.00	**1.03/3**	1.00	**1.03/3**	1.00
Simuliidae (n=7)	Rock (n=28)	α	2.75	5.77	2.09	3.32	1.77	2.72
		β	**1.19/12**	1.10	**1.11/11**	1.03	**1.09/9**	1.02
	Transect (n=15)	α	3.27	6.31	2.33	3.42	1.93	2.79
		β	**1.47/19**	1.09	**1.15/7**	1.02	**1.11/5**	1.01
	Station (n=5)	α	4.80	6.86	2.69	3.51	2.14	2.83
		β	**1.35/24**	1.02	1.02/1	1.01	0.93/0	1.00
	Arm (n=2)	α	6.50	7.00	2.74	3.53	2.00	2.84
		β	**1.08/8**	1.00	**1.29/29**	1.00	**1.37/37**	1.00

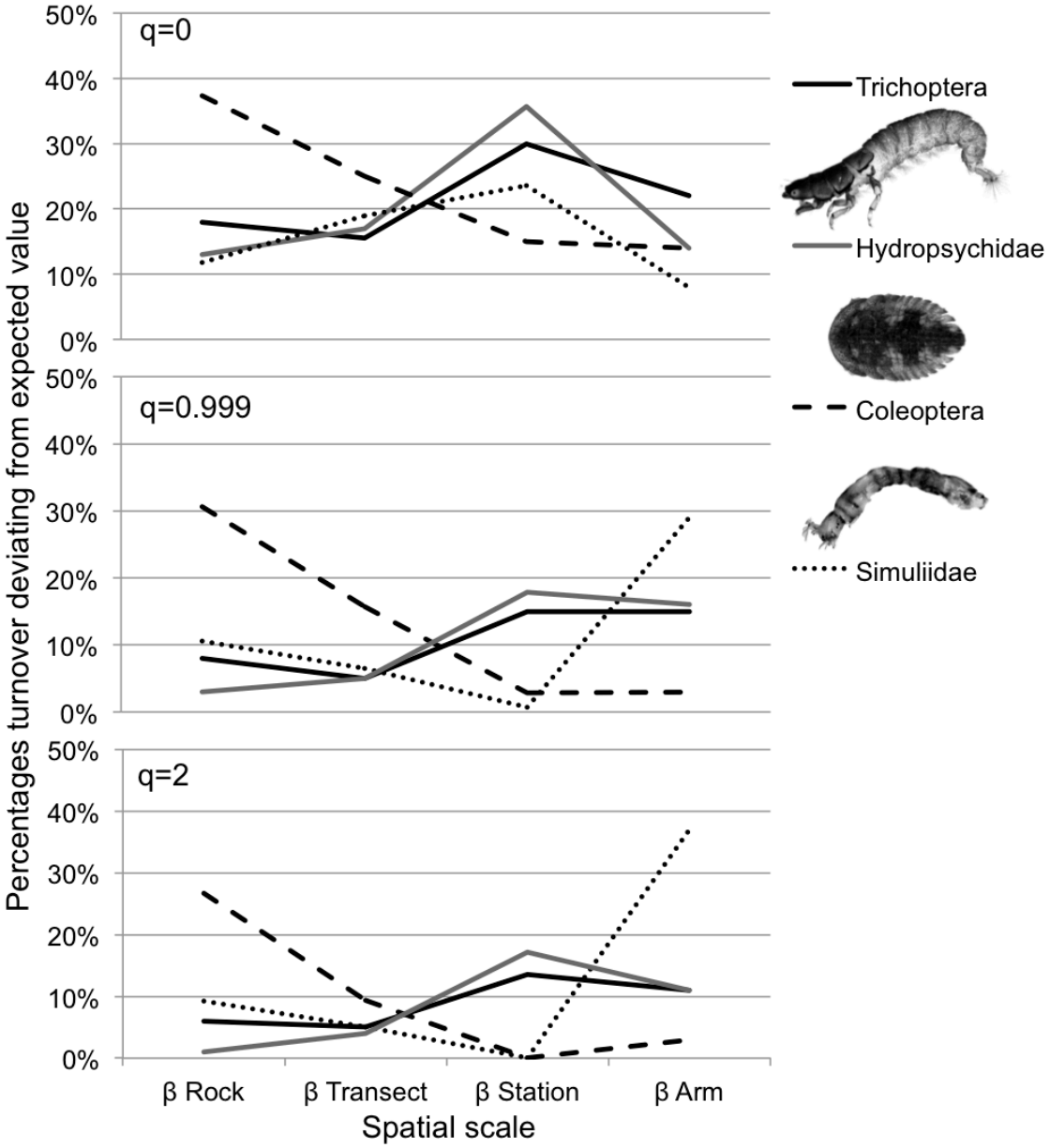

Figure A8.1. Multiplicative partitions of sampled summer aquatic insect species (orders Trichoptera, Coleoptera and family Simuliidae) at White Clay Creek (PA, USA; June 20, 2013).