STREAM BIODIVERSITY AND CO-OCCURRENCE NETWORK TOPOLOGY ALONG SUB-CONTINENTAL TO GLOBAL NUTRIENT AND LAND USE GRADIENTS

by

WILLIAM ROBERT BUDNICK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at
the University of Texas at Arlington
August, 2020
Arlington, Texas

Supervising Committee:

Sophia I. Passy, Supervising Professor
Jonathan Campbell
Paul Chippindale
Eric Smith
Matthew Walsh

Abstract

Freshwaters are some of the most vulnerable ecosystems to global change forces, such as land use and eutrophication, implicated in the loss of biodiversity and biotic homogenization (increased similarity) of species communities. However, much less is known about the impact of these forces on the mechanisms underlying biodiversity and the patterns of species co-occurrence. In this dissertation, I use sub-continental to global stream community datasets to investigate the effects of land use and nutrients on algal, macroinvertebrate, and fish biodiversity, abundance, and complexity of co-occurrence patterns to pursue three objectives. I first investigate how agriculture contributes to biotic homogenization, and questioned if homogenization processes operate at local and regional scales (Chapter 2). Using US, French, and Canadian datasets, I determined that generally agriculture homogenizes communities, concurrent with a loss of regional biodiversity and an increase in local diversity, and that relative abundances of common species, not spatial distributions, contribute greatly to homogenization. I also observed that diatoms and insects diverged from fish in terms of biodiversity, abundance, and assembly patterns, emphasizing the importance of body size and/or dispersal capacity over trophic position. In my second investigation, I combine network theory and metacommunity approaches to explore i) whether nutrient supply and nutrient ratio constrain the complexity of sub-continental co-occurrence networks in stream algae and ii) the relative influence of climate and dispersal on co-occurrence relationships vs. metacommunity composition (Chapter 3). I found that nutrient supply and ratio are both important drivers of the size and complexity of algal co-occurrence networks. Further, climate and space (surrogate for dispersal processes) were major influences on co-occurrence relationships across networks but nutrient context determined their relative contributions. Notably, climate and space had differential effect on co-occurrence

network topology compared to metacommunity composition, which indicated that processes driving individual species relationships are detached from those driving metacommunity-level patterns. In my third investigation, I examined the impact of spatial scale and body size on the shape of the node degree distribution, which is probability distribution describing how links (co-occurrences) between nodes (species) are distributed in a co-occurrence network (Chapter 4). Using diatom and fish datasets, my results are the first to show an explicit correlation between the observation scale (sub-regional to sub-continental) from which the network is constructed and the shape of the node degree distribution, which further depended upon body size. My findings advance our understanding of biodiversity and co-occurrence patterns in streams and their response to global change, scale and organismal biology. My work also demonstrates a successful integration of mathematical, metacommunity, and network approaches, which provides a solid analytical foundation to improve conservation recommendations and develop more holistic solutions in the face of the on-going biodiversity crisis.

Dedication

I am forever indebted to my department, which is basically everyone, who stood behind me and supported me during my bout of cancer treatment. You have all shown me life-changing kindness and eased a burden that no graduate student should be forced to bear while following their dreams. I will not forget what you have done for me. I hope you will continue to do the same for anyone else in as destitute and trying a circumstance as I then represented.

For Christina, for putting up with me and keeping my head straight during my most stressful times. I promise that the best years are now in front of us.

Acknowledgements

I would like to thank my supervisor, Sophia Passy, for the help and guidance and for her unlimited patience and wisdom during my years at UTA. I am grateful for the work we have produced and look forward to more contributions to science down the road. I am thankful also for my committee: Jonathan Campbell, Paul Chippindale, Eric Smith, and Matthew Walsh who patiently and gratefully help supervise the production of this exhaustive work. I thank my French (Thibault Leboucher, Aurélien Jamoneau, Juiliette Tison-Rosebery, Jérôme Belliard and Evelyne Tales), Finnish (Janne Soininen and Virpi Pajunen) and Canadian collaborators (Isabelle Lavoie and Stéphane Campeau) who assisted intellectually and materially with my manuscripts as well as my first co-first authored paper. Joe Mruzek and Katrina Pound were essential lab mates during our many projects and I thank them for their help in fleshing out many of the ideas I pursued. Chad Larson was essential in helping me organize my datasets. I finally thank my former advisor, Mike Kaller, and friend Chris Bonvillain, who continue to inspire me to be the best crayfish expert I can be (even though this dissertation has no crayfish in it). Chelae up!

Table of Contents

Table of Figures

List of Tables

Chapter 1
General Introduction

Global biodiversity patterns are changing, and generally not for the better. Losses of species and increasingly similar community compositions are slowly becoming the norm (Finderup Nielsen *et al.*, 2019) and the alarm is being loudly sounded by many researchers to organize, collaborate, and act (Driscoll *et al.*, 2018; Eriksson & Hillebrand, 2019). Human influences, such as climate change, nutrient modification, and land cover alterations, are among the principal drivers of global biodiversity loss. This loss is particularly pronounced in aquatic ecosystems owing to their disproportionate representation of species richness and vulnerability to landscape-scale hydrological influences and modifications (Sala *et al.*, 2000; Porter *et al.*, 2013). Further complicating the crisis is that aquatic systems are generally understudied relative to terrestrial systems, which inhibits our ability to synthesize general conservation strategies. Therefore, the goal of this dissertation is to improve our understanding of the consequences of global change on aquatic biota.

Ecological communities vary in their constituent species through time and space, a pattern known as β-diversity. However, terrestrial and aquatic communities have begun to experience biotic homogenization, or the process of communities becoming more similar in species composition (decreased beta diversity), as a result of global change, especially through agricultural land use (McKinney, 2006; Petsch, 2016; Finderup Nielsen *et al.*, 2019). Local and regional scale assembly processes drive the spatial distribution and relative abundance of species, which in turn determine β-diversity. However, it is unclear how agriculture affects these across taxa differing in body size and dispersal capacity. Therefore, in my first chapter, I explored whether agricultural land cover was associated with biotic homogenization of benthic

algae, macroinvertebrate, and fish in streams from the United States, Canada, and France. This comprehensive intercontinental study examined if stream communities in watersheds dominated by agriculture vs. forest were driven by different community assembly processes, and to what extent two local mechanisms (relative abundance vs. spatial aggregation) played a role. I found that agricultural effects were generally associated with homogenization of all three taxa, and such effects were mainly through eutrophication. I also found that the relative contributions of local and regional assembly processes were not strongly influenced by land cover across all taxa, but there were clear differences in the magnitudes of these processes across taxa. I finally found that relative abundance patterns generally characterize community variability over spatial aggregation, and the role of relative abundance generally increases with agricultural cover. This work broke major ground in finding that i) agriculturally-driven homogenization is a general problem for aquatic taxa, ii) local and regional homogenization processes differ across taxa, which emphasizes the need for organism-specific considerations, and iii) homogenization is very strongly characterized by relative abundance differences across sites, not necessarily spatial differences.

Aquatic biodiversity patterns also reflect how species co-occur, or not, in time and space, and these patterns are a product of shared ecological niches or biotic interactions (Araújo *et al.*, 2011; Poisot *et al.*, 2015; Morueta-Holme *et al.*, 2016; Freilich *et al.*, 2018). In stream ecosystems, nutrient effects, such as supply and ratio, are one of the most important environmental contexts for species coexistence and response to climate and dispersal processes (Elser *et al.*, 2007; Cardinale *et al.*, 2009; Harpole *et al.*, 2011). However, it is not known how species co-occurrence patterns are influenced by these nutrient dimensions, which is a concern, given the ongoing global change of freshwaters, experiencing anthropogenic nutrient inputs and

removal (Allan, 2004). Therefore, in my second chapter, I used a subcontinental benthic algal dataset to investigate whether subcontinental co-occurrence networks, which represent species as nodes and co-occurrence relationships among them as edges, differ in their topologies between two nutrient supply contexts (eutrophic and oligotrophic) and two nutrient ratio contexts (N-limited and P-limited). I further explored if nutrient context determine how strongly climate and dispersal structure co-occurrence relationships vs. metacommunity compositions. I found that nutrient supply and ratio were important drivers of network properties, metacommunity composition, and their response to climate and dispersal, which were differential. These findings emphasize the differences between individual-level vs. metacommunity-level patterns across climate and dispersal gradients.

The node degree distribution (NDD) of ecological networks, i.e. the frequency distribution describing how links are distributed among nodes, is of wide interest in science and technology because different shapes of the NDD are associated with different information transfer rates and susceptibility to targeted or random species removal (Albert *et al.*, 2000; Dunne *et al.*, 2002). NDD shapes are often described using statistical models (Amaral *et al.*, 2000), but there is general controversy as to which statistical models best fit NDDs (Broido & Clauset, 2019; Holme, 2019). Surprisingly, no comprehensive analyses to date have been performed to compare NDD models for their plausibility along spatial gradients and organismal groups. Therefore, in my third chapter, I used diatom and fish datasets and examined how the variety and quality of statistical models of NDDs depended upon spatial extent (subregional to subcontinental) and if the patterns were consistent between taxa differing in body size. I found that there was a strong spatial dependency in the type and number of best fitting statistical models. As spatial extent increased, diatom networks became more inhomogeneous (broad-scale)

and predictable, explained by fewer plausible models, while the fish networks became more homogeneous (single-scale) and less predictable, explained by more plausible models. Further, I found little evidence for power law models as a plausible fit of NDD shapes, shedding light on a long-standing debate for the applicability of these models in ecological networks. This work was the first to comprehensively demonstrate explicit spatial and taxonomic constraints on the co-occurrence network NDDs, deepening our knowledge of the origins of ecological network variability.

References

Albert, R., Jeong, H. & Barabási, A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.

Allan, J.D. (2004) Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 257–284.

Amaral, L.A.N., Scala, A., Barthelemy, M. & Stanley, H.E. (2000) Classes of small-world networks. *Proceedings of the national academy of sciences*, **97**, 11149–11152.

Araújo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, **34**, 897–908.

Barberán, A., Bates, S.T., Casamayor, E.O. & Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal*, **6**, 343–351.

Broido, A.D. & Clauset, A. (2019) Scale-free networks are rare. *Nature Communications*, **10**, 1–10.

Cardinale, B.J., Hillebrand, H., Harpole, W.S., Gross, K. & Ptacnik, R. (2009) Separating the influence of resource "availability" from resource "imbalance" on productivity-diversity relationships. *Ecology Letters*, **12**, 475–487.

Connor, N., Barberán, A. & Clauset, A. (2017) Using null models to infer microbial cooccurrence networks. *PLoS ONE*, **12**.

Driscoll, D.A., Bland, L.M., Bryan, B.A., Newsome, T.M., Nicholson, E., Ritchie, E.G. & Doherty, T.S. (2018) A biodiversity-crisis hierarchy to evaluate and refine conservation indicators. *Nature Ecology and Evolution*, **2**, 775–781.

Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002) Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, **5**, 558–567.

Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B. & Smith, J.E. (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, **10**, 1135–1142.

Eriksson, B.K. & Hillebrand, H. (2019) Rapid reorganization of global biodiversity Marine systems outpace terrestrial habitats in biodiversity erosion. *Science*, **366**, 308–309.

Finderup Nielsen, T., Sand-Jensen, K., Dornelas, M. & Bruun, H.H. (2019) More is less: net gain in species richness, but biotic homogenization over 140 years. *Ecology Letters*, **22**, 1650–1657.

Freilich, M.A., Wieters, E., Broitman, B.R., Marquet, P.A. & Navarrete, S.A. (2018) Species co-

occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology*, **99**, 690–699.

Harpole, W.S., Ngai, J.T., Cleland, E.E., Seabloom, E.W., Borer, E.T., Bracken, M.E.S., Elser, J.J., Gruner, D.S., Hillebrand, H., Shurin, J.B. & Smith, J.E. (2011) Nutrient co-limitation of primary producer communities. *Ecology Letters*, **14**, 852–862.

Holme, P. (2019) Rare and everywhere: Perspectives on scale-free networks. *Nature communications*, **10**, 1–3.

Li, D., Poisot, T., Waller, D.M. & Baiser, B. (2018) Homogenization of species composition and species association networks are decoupled. *Global Ecology and Biogeography*, **27**, 1481–1491.

McKinney, M.L. (2006) Urbanization as a major cause of biotic homogenization. *Biological Conservation*, **127**, 247–260.

Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B.J., Peet, R.K., Ott, J.E., Violle, C., Enquist, B.J., Jørgensen, P.M. & Svenning, J.C. (2016) A network approach for inferring species associations from co-occurrence data. *Ecography*, **39**, 1139–1150.

Petsch, D.K. (2016) Causes and consequences of biotic homogenization in freshwater ecosystems. *International Review of Hydrobiology*, **101**, 113–122.

Poisot, T., Stouffer, D.B. & Gravel, D. (2015) Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, **124**, 243–251.

Porter, E.M., Bowman, W.D., Clark, C.M., Compton, J.E., Pardo, L.H. & Soong, J.L. (2013) Interactive effects of anthropogenic nitrogen enrichment and climate change on terrestrial and aquatic biodiversity. *Biogeochemistry*, **114**, 93–120.

Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L.R., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Chapter 2
Local and regional drivers of taxonomic homogenization in stream communities along a land use gradient

*Published in Global Ecology and Biogeography
DOI Link: https://doi.org/10.1111/geb.12976
Published with permission (Clearance License: 4832500817711)

William R. Budnick[1], Thibault Leboucher[2], Jerome Beilard[3], Janne Soininen[4], Isabelle Lavoie[5], Katrina Pound[1], Aurélien Jamoneau[2], Juliette Tison-Rosebery[2], Evelyne Tales[3], Virpi Pajunen[4], Stephane Campeau[6], and Sophia I. Passy[1*]

[1]Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA

[2]National Research Institute of Science and Technology for Environment and Agriculture (Irstea), Aquatic Ecosystems and Global Changes Research Unit, Bordeaux, France

[3]Irstea, Hydrosystems under Change (HYCAR) Research Unit, Antony, France

[4]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

[5]Institut national de la recherche scientifique, centre Eau Terre Environnement, 490 rue de la Couronne, Québec City, Québec G1K 9A9, Canada

[6]Université du Québec à Trois-Rivières, 3351 boul. des Forges, Trois-Rivières, Québec G9A 5H7, Canada

[*]Corresponding author: sophia.passy@uta.edu

Introduction

Landscape transformations from continuous undeveloped expanses to agricultural fields and urban sprawls have accelerated the global biodiversity decline (Newbold, Hudson, Hill, Contu, Lysenko et al., 2015). Human land use (hereafter land use) underlies declines in both regional richness, i.e. γ-diversity (Barlow, Lennox, Ferreira, Berenguer, Lees et al., 2016), and dissimilarity among biological communities, i.e. β-diversity, resulting in taxonomic homogenization across space and time (Petsch, 2016). Biodiversity losses from land use stem from habitat loss, fragmentation, eutrophication, and physicochemical stress, altogether considered among the primary threats facing global biodiversity (Sala, Stuart Chapin, Iii, Armesto, Berlow et al., 2000; Devictor, Julliard, Clavel, Jiguet, Lee et al., 2008). Preventing biodiversity losses and mitigating subsequent homogenization remain a top priority because both can translate into decreased biological integrity and ecosystem resilience (de Juan, Thrush & Hewitt, 2013; Socolar, Gilroy, Kunin & Edwards, 2016). Therefore, it is critical from a conservation planning standpoint to continue investigating how land use affects ecological processes underlying global diversity in order to mitigate the ongoing biodiversity crisis.

Land use effects on biodiversity occur across scales, operating either in a top-down or bottom-up fashion or both (Flohre, Fischer, Aavik, Bengtsson, Berendse et al., 2011). Top-down mechanisms function through the regional species pool (γ-diversity), which is a product of speciation and extinction, large-scale dispersal, climate, and evolutionary, geological, and land use history (Zobel, 2016). Bottom-up mechanisms include local-level assembly processes, e.g. environmental filtering, interspecific interactions, and small-scale dispersal (Márquez & Kolasa, 2013), which constrain local (α) diversity and subsequently affect site-to-site community

dissimilarity. Studies across terrestrial and freshwater systems have reported a general decline in γ-diversity because of land use, but divergent patterns of α-diversity, including decreased α-diversity, owing to losses of sensitive and endemic species, and stable, or even increased α-diversity, owing to greater rates of species invasion and colonization (Vellend, Baeten, Myers-Smith, Elmendorf, Beauséjour et al., 2013; Newbold *et al.*, 2015; Gonzalez, Cardinale, Allington, Byrnes, Arthur Endsley et al., 2016). Thus, land use likely exerts differential impact on the species pool and local assembly processes that may cause γ- and α-diversity, respectively, to vary at different rates, which in turn influences β-diversity response (Kraft, Comita, Chase, Sanders, Swenson et al., 2011).

β-diversity is usually treated as a scalar linking average α-diversity with γ-diversity, thus reflecting spatial or temporal differences among localities. One can then measure the influence of α- and γ-diversity as proxies of local and regional drivers of β-diversity, respectively. Specifically, null models that constrain the observed species pool variation (i.e., γ-diversity) can assess the role of local assembly by calculating a β-diversity measure (β_{DEV}) corresponding solely to α-diversity variation (e.g., Kraft *et al.*, 2011) (Fig. 2.1a). β_{DEV} can be further decomposed into fractions reflecting roles of intraspecific spatial aggregation (i.e., the spatial clumping pattern of individuals within species) and the regional species abundance distribution (SAD, vector of species abundances) (Xu, Chen, Liu & Ma, 2015) (Fig. 2.1b). Intraspecific spatial aggregation results from dispersal, competitive, and environmental mechanisms that cluster individuals of species across fewer sites, thus bolstering β-diversity (Veech, 2005). However, regional SADs influence β-diversity because rare species are less likely to be locally sampled due to low regional abundance (He & Legendre, 2002). Although examined across latitudes, the two fractions of local assembly have not been studied in other contexts and it is

unknown whether these components are responsive to strong ecological influences (e.g., land use).

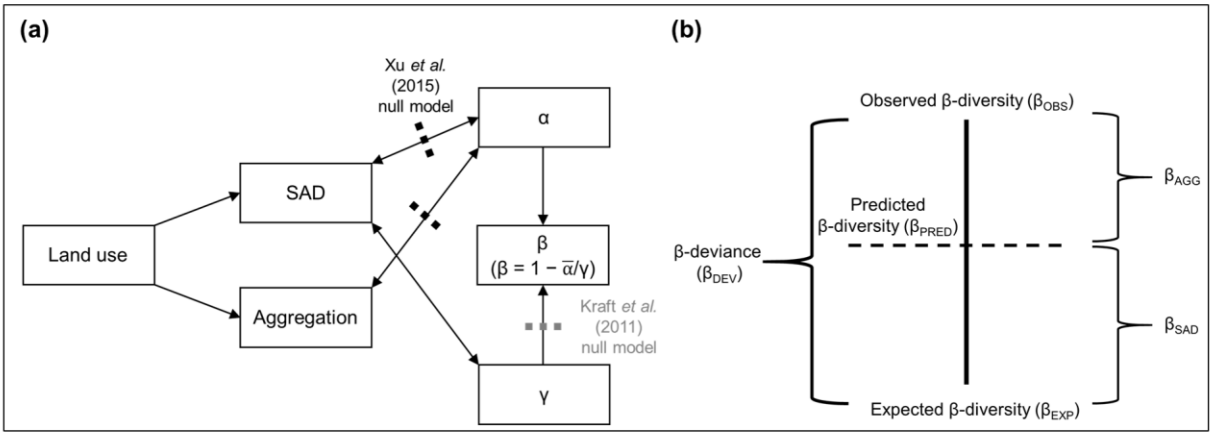

Fig. 2.1 a. Conceptual model depicting the land use effect on the species abundance distribution (SAD) and intraspecific spatial aggregation, which in turn interact with local (α) and regional (γ) diversity. β-diversity is calculated as a function of average α-diversity and γ-diversity. Interactions that were controlled for by the null models of Kraft *et al.* (2011) and Xu *et al.* (2015) are marked with a thick dotted line. **b.** Diagram summarizing the Xu *et al.* (2015) partition of β_{DEV} into fractions explained by the SAD and intraspecific spatial aggregation using an occupancy-abundance based null model procedure. The null model β_{DEV} is taken as the raw difference between expected β-diversity (β_{EXP}) and observed β-diversity (β_{OBS}). The fraction of β_{DEV} explained by the SAD, β_{SAD}, is the difference between predicted β-diversity (β_{PRED}) and expected β-diversity (β_{EXP}), whereas the fraction of β_{DEV} explained by intraspecific aggregation (β_{AGG}) represents the difference between β_{OBS} and β_{PRED}.

Studying how local assembly and regional species pool processes interplay is an ongoing area of research in terrestrial systems because it may explain how β-diversity varies with land use (Socolar *et al.*, 2016). Surprisingly, little attention is focused on freshwater systems, even though freshwater biodiversity is more vulnerable to land use relative to terrestrial systems, particularly through habitat modification (Sala *et al.*, 2000; Wiens, 2016) and eutrophication from agriculture (Withers, Neal, Jarvie & Doody, 2014). Although primary productivity in agricultural streams could increase with eutrophication, forest streams, which are usually low in nutrients and have more shading, tend to harbor higher biodiversity stemming from greater physical and environmental heterogeneity that translates into greater ecosystem complexity (Penaluna, Olson, Flitcroft, Weber, Bellmore et al., 2017). Agriculture probably causes changes in physicochemical heterogeneity as well, but this subject is poorly explored. Thus, the scarcity of data, especially for aquatic taxa, has inhibited general understanding of how land use influences local and regional processes driving β-diversity.

Impacts of agricultural eutrophication on β_{DEV} are not understood, although null models have been used to assess environmental disturbance (e.g., Myers, Chase, Crandall & Jiménez, 2015). We hypothesize β-diversity response to eutrophication, including variation in β_{DEV}, depends on trophic level, body size, and dispersal capacity. For example, many unicellular producers, like diatoms, have high nutrient demands and may benefit from increased nutrients (Passy, 2008; Soininen, Jamoneau, Rosebery & Passy, 2016). Diatom microscopic size, high local abundance, and broad geographic distributions, allowing both in-stream and overland passive dispersal (Finlay, 2002), may result in weak β-diversity and β_{DEV} response to agriculture. Smaller bodied macroscopic organisms, such as aquatic insects, may be more constrained in active dispersal capacity during larval stages but exhibit greater overland mobility during winged

adult life stages, which could offset some harmful agricultural effects. In contrast, larger consumers with more limited geographic dispersal capacity, such as fish, may be negatively affected by eutrophication due to ammonia toxicity, loss of suitable habitat, and lower quality food sources (Allan, 2004).

In this study, we compared spatial patterns of biodiversity and abundance in streams with watersheds dominated by agriculture vs. forest. Our objectives were to determine: i) how β-diversity and related biodiversity properties respond to agriculture (through nutrient enrichment or physicochemical heterogeneity), ii) if agriculture alters the relative contribution of local assembly effects to β-diversity, iii) whether agriculture differentially constrains the fractions of local assembly explained by spatial aggregation vs. the SAD, and iv) if the relationships outlined in i) to iii) vary across organismal groups (Table 2.1).

Materials and Methods

Data sources and site selection

Our datasets (six in total) comprise stream organisms sampled from the US, France, and Canada (Fig. 2.2). Each dataset included community data and physicochemistry from watersheds dominated by either forest or agriculture. Only streams with ≥ 50% of their upstream watershed belonging to one of the two land cover categories were included in our analyses. We examined biodiversity patterns across three US datasets (diatoms, insects, and fish), two French datasets (diatoms and fish), and one Canadian dataset (diatoms), constructed as follows.

Table 2.1. Summary of procedures and analyses performed with corresponding expectations and observations.

Procedures	Analyses	Expectations	Observations
1) Determine the differences in physicochemistry and physicochemical heterogeneity between land covers.	PCA, PERMDISP, MANOVA	Land cover would be characterized effectively by physicochemical parameters and potentially by physicochemical heterogeneity.	1) All analyses clearly separated streams into two groups, corresponding to forest and agriculture; 2) Agricultural streams had elevated nutrient levels, suggestive of eutrophication; 3) Physicochemical heterogeneity was greater among agricultural streams except in the Canadian diatom dataset.
2) Reveal the responses of $\overline{\alpha}$-, γ-, and β-diversity, SAD evenness, and SAD skewness to physicochemistry and physicochemical heterogeneity.	MANOVA, Variance partitioning	The responses of biodiversity components to physicochemistry and physicochemical heterogeneity may differ depending on body size, dispersal capacity, and trophic level (autotroph vs. heterotroph).	1) In general, β- and γ-diversity were negatively related to eutrophication, whereas $\overline{\alpha}$-diversity increased. SADs tended to be more uneven in agricultural streams due to buildup of common species and/or increased dominance; 2) Covariance of land use with physiochemistry explained most of the diversity variation across datasets, whereas environmental heterogeneity poorly explained diversity; 3) Pure land cover and pure physicochemistry generally explained some additional variation in the diversity components.

Table 2.1. Continued:

Procedures	Analyses	Expectations	Observations
3) Determine if land use influences the relative roles of local assembly and the regional species pool in driving β-diversity.	Null models, Permutational ANOVA, Variance partitioning	The contribution of local assembly should be responsive to agricultural land use, however, the magnitude and direction of the response may vary across organismal groups.	1) The role of local assembly was generally weakly affected by land use, and not in a consistent way across datasets, suggesting a potential influence of organismal type and biogeography.
4) Determine if β-deviation (β_{DEV}) is explained by the species abundance distribution (SAD) or intraspecific spatial aggregation.	Null models, Permutational ANOVA	It is unknown how land use may influence the fractions of β_{DEV} explained by the SAD and intraspecific spatial aggregation.	1) The SAD was the dominant fraction of β_{DEV} and this pattern was independent of land use and organismal group. However, the SAD fraction was significantly higher in agriculture across all datasets, which may be the underlying factor of taxonomic homogenization; 2) Intraspecific spatial aggregation fraction was negative for agricultural streams and positive for forest streams, indicating that intraspecific aggregation was lower than expected across disturbed streams

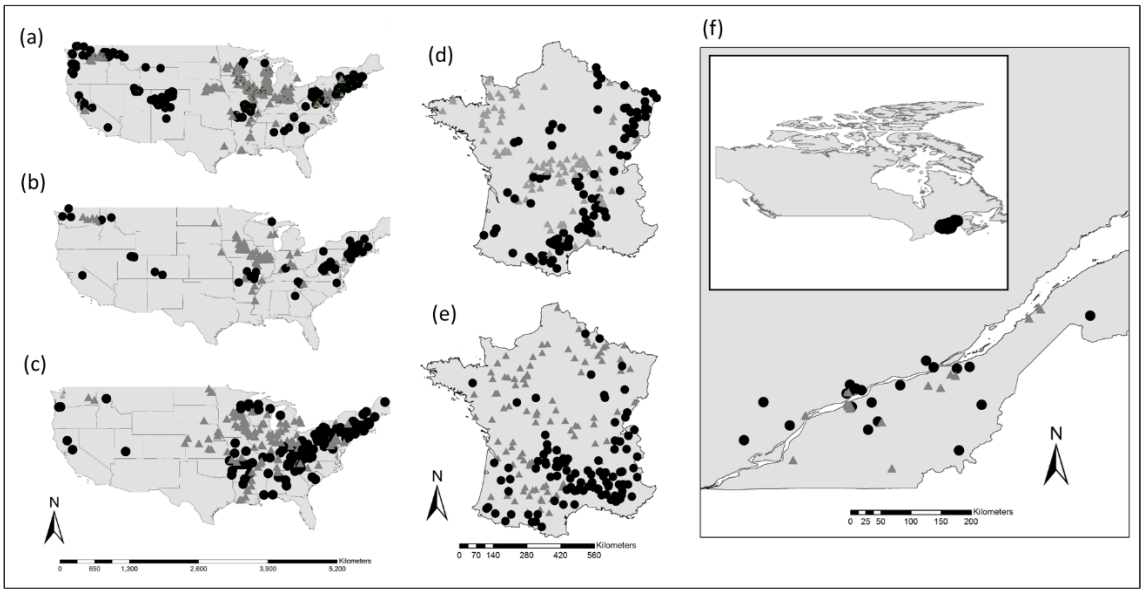

Fig. 2.2: Maps of diatom, macroinvertebrate, and fish sampling localities in the US, France, and Canada. Grey triangles represent agriculture samples, whereas black circles represent forest samples. a = US diatoms, b = US insects, c = US fish, d = French diatoms, e = French fish, f = Canadian diatoms.

United States – US community data, spanning 19 latitudinal degrees and 55 longitudinal degrees, were obtained from the National Water-Quality Assessment (NAWQA) Program of the United States Geological Survey and the National Rivers and Streams Assessment (NRSA) of the United States Environmental Protection Agency. Communities were collected in the warm months during low flow conditions (July through September) to constrain seasonal succession and variation in temperature and flow. NAWQA communities (diatoms, insects, and fish) were sampled between 1993-2010, whereas NRSA communities (fish), between 2011 and 2012. Diatoms were collected from the richest-targeted habitats, encompassing hard substrates or macrophytes. Depending on available substrate, a defined area of 25 cobbles, 5 woody snags or 5

macrophyte beds was sampled within a stream reach and the samples were composited. Diatoms were identified generally to species in counts of 400-800 cells. Benthic insects (class Insecta) were composed of combined sieved samples taken from the richest-targeted habitats (i.e., riffles, main-channel, and natural-bed instream habitats). Insects were identified to the lowest possible category (order to species) in counts of 400-800 individuals. Both NAWQA and NRSA fish were generally identified to species in counts of 400-950 individuals taken from riffle, pool, and run habitats using electrofishing equipment with seines.

Land use and cover data were generated by the NAWQA and NRSA using National Land Cover Datasets 1992 and 2006, 30 m resolution. We selected 400 streams for diatoms and 126 streams for insects split equally between both land cover categories. Since fish communities and environmental data in both the NAWQA and NRSA data were sampled with similar methods, we combined both fish datasets into a single dataset comprising 231 streams (116 agricultural and 115 forested streams).

France – French diatom data were sourced from a national dataset including field collections of 200 streams from 2011. Algae were collected from stones during the low flow period in June through September with a standardized sampling method (Afnor, 2007). Diatoms were identified generally to species in counts of about 320-475 cells. The French fish dataset was collected by the French National Agency for Water and Aquatic Environments (ONEMA) during low flow periods between May and October 2011. The dataset comprised 200 streams with fish identified to species in counts of 10-3300 individuals sampled with electrofishers. For both French datasets, we used 100 agricultural and 100 forest streams, spanning 8 latitudinal and 14

longitudinal degrees. Land use cover data were obtained from the CORINE land cover database (European Environment Agency, 2013)

Canada – Canadian diatom data included 46 stream samples (23 streams in both land cover categories) collected in August to September during the low flow period between 2002 and 2009 (Lavoie, Campeau, Zugic-Drakulic, Winter & Fortin, 2014) spanning 3 latitudinal and 6 longitudinal degrees. Samples were composites of rock scrapes (5-10 rocks) per stream reach, targeting riffles and runs. Diatoms were mainly identified to species in counts of at least 400 valves. Land use cover data were compiled from government GIS databases, including the Ecoforestry Information System, Annual Crop Inventory, and the Insured Crop Database.

Environmental data – All datasets had associated physicochemical and coordinate data (i.e., GCS coordinates re-projected with Lambert Conformal Conic). Environmental variables in our analyses included water temperature, air temperature, nitrite + nitrate (or total nitrogen when absent), ammonia, orthophosphate, total phosphorus, specific conductance, and pH (Appendix 1, Table S1.1 in Supplemental Information). Environmental data for the US datasets consisted of the average for the month of sample collection. Environmental data for French diatoms included the median of measurements obtained 30 days before and 15 days after the diatom sample date. The French fish environmental data represented the average of 12 monthly measurements prior to fish sampling. Air temperature for French diatom data were not recorded at the time of sampling and were obtained from the WorldClim database (Hijmans, Cameron, Parra, Jones & Jarvis, 2005), whereas air temperatures for French fish streams were measured at the stream. Canadian environmental data were seasonal averages calculated from water samples collected from July to September.

Diversity, spatial aggregation, and species abundance distribution

We calculated $\overline{\alpha}$-diversity (average richness across samples), γ-diversity (total richness per land use), and β-diversity of stream samples for both land cover categories for each dataset. We used equation (1) to calculate the observed β-diversity (β_{OBS}),

$$\beta_{OBS} = 1 - \frac{\overline{\alpha}}{\gamma} \quad (1)$$

which indicated the average proportion of the species pool absent from a stream.

We used the null model framework developed by Xu *et al.* (2015) to quantify i) the magnitude of the local assembly effect on β-diversity after controlling for γ-diversity and ii) the contributions of the SAD vs. intraspecific spatial aggregation to local assembly (Fig. 1b). First, the difference (i.e., β-deviance, β_{DEV}) between β_{OBS} and the expected β-diversity (β_{EXP}, i.e., β-diversity expected assuming completely random sampling, see Appendix S2) was taken to quantify local assembly absent the effect of γ-diversity. β_{DEV} is bounded between 0 and 1, with larger β_{DEV} corresponding to greater local control. Secondly, we calculated β-diversity predicted when intraspecific spatial aggregation is constant across all species (β_{PRED}). Then, the difference between β_{PRED} and β_{EXP} reveals what fraction of β_{DEV} is contributed by the SAD (β_{SAD}), while the remaining fraction of β_{DEV} is attributed to spatial aggregation (β_{AGG}). In this model, β_{SAD} can exceed β_{DEV} if β_{PRED} exceeds β_{OBS}. The corresponding aggregation fraction will in turn be negative because the sum of the two fractions, β_{SAD} and β_{AGG}, must equal 1, thus meaning that the pattern is less aggregated than expected by the null model. To test whether the two land covers differ in their magnitude of intraspecific spatial aggregation, we used maximum likelihood methods and calculated the aggregation parameter, *k*, across samples within each land cover (Appendix S3). Because smaller *k* corresponded to greater aggregation, we analyzed the

reciprocal of the parameter for easier interpretation. In summary, the procedure yielded six measurements: β_{EXP}, β_{PRED}, β_{DEV}, β_{SAD}, β_{AGG}, and $1/k$.

Regional SADs for both land cover categories was analyzed by summing abundances of each species across all stream samples and calculating the standard deviation (parameter σ) of the Poisson-lognormal distribution fit of the abundance data using the `sads´ R package (Prado, Mirands & Chalom, 2017). Parameter σ reflects SAD evenness with greater σ values corresponding to increased unevenness. To determine if changes in σ were associated with prevalence of rare vs. common species, we also examined the relationship of σ with the skewness (`skewness´ function from R package `moments´, Komsta & Novomestky, 2015) of the log-transformed regional species abundances for each land cover category. Skewness was significant if skewness divided by the standard error of the skewness (i.e., $(6/n)^{0.5}$, where n = number of species) was greater than 2. Significant positive skew indicates greater prevalence of abundant species, while significant negative skew reveals higher number of rare species compared to the lognormal distribution.

Statistical analyses

Resampling scheme – The described procedures in our study typically produced a single value without any estimate of error, which prohibits statistical comparisons between datasets. Therefore, to test for abiotic and biotic differences between land covers, we conducted a resampling procedure by randomly selecting 50% of the streams within each land cover category for each dataset without replacement 999 times. Each loop calculated the median of each physicochemical variable, an estimate of physicochemical heterogeneity, biodiversity ($\overline{\alpha}$-, β-, and γ-diversity), SAD, and null model measures including the null model β-diversity values, and

the within group intraspecific aggregation (1/*k*). This procedure generated six new datasets that contained resampled physiochemistry data and biotic measures, which were used further statistical analyses. R scripts are available as supplementary material for online publication only (see Appendices S3 and S4).

Eutrophication and physicochemical heterogeneity – We employed principal components analysis with all resampled, standardized median physicochemical variables (mean = 0, standard deviation = 1) to create a synthetic variable corresponding to the major physicochemical trend. The first PCA axis represented a eutrophication gradient and explained between 53.1% (French diatom samples) and 94.3% (Canadian diatom samples) of the variation among samples (Appendix 2.2, Fig. S2.1).

To estimate physicochemical heterogeneity within each land cover, we used permutational analysis of multivariate dispersion on standardized physiochemical data with the `betadisper´ function from R package `vegan´ (Anderson, Ellingsen & McArdle, 2006; Oksanen, Blanchet, Friendly, Kindt, Legendre et al., 2017). In this procedure, physicochemical heterogeneity is calculated as the average distance from a multivariate group median (group = land cover) with larger distances corresponding to greater within-group heterogeneity.

Environmental effects – We determined how land use-driven eutrophication and physicochemical heterogeneity affected diversity components using a combination of univariate and multivariate techniques and variance partitioning. For each dataset, we used permutational MANOVA function `adonis´ from package `vegan´ to test for differences in the multivariate mean of the $\overline{\alpha}$-, β-, and γ-diversity between land covers. If the permutational MANOVA was significant, we

followed with permutational ANOVA using the `perm.anova´ function provided in `RVAideMemore´ (999 permutations; Herve, 2018) for each dependent variable. We then used RDA-based variance partitioning models (`vegan´ function `varpart´) on each dataset to identify major explanatory factors underlying diversity patterns, with eutrophication, physicochemical heterogeneity, and land cover (coded as dummy variables) as predictors and the diversity measures ($\overline{\alpha}$-, β-, and γ-diversity) as response variables.

We employed permutational MANOVA and permutational ANOVA to determine if the resampled β_{DEV}, β_{SAD}, β_{AGG}, 1/*k*, σ, and skewness differed between land covers. Because total abundance and γ-diversity influence the shape of the regional SAD, we controlled their influences by regressing parameter σ against total abundance and γ-diversity of the resample and obtaining the residuals, which were then used in subsequent analyses. To further explore if β_{DEV} was sensitive to variation in SAD unevenness (residual σ) and intraspecific spatial aggregation (1/*k*), we calculated Pearson correlations within both land cover categories for all datasets. Pearson correlations were also used to assess whether residual σ correlated with skewness and 1/*k*. We then implemented variance partitioning to determine if eutrophication, physicochemical heterogeneity, land cover, or their covariance explained the variation in β_{DEV}.

Results

Eutrophication and Environmental Heterogeneity Effects on Diversity and the SAD

Permutational MANOVA and permutational ANOVAs of environmental data showed that all physiochemistry levels were significantly elevated ($P < 0.05$) in agricultural land use across all datasets. Permutational ANOVAs also indicated greater physicochemical heterogeneity among agricultural streams in all but the Canadian diatom dataset (higher in forest land cover) and the

US Fish dataset (no differences, Fig. 2.3). MANOVA of $\overline{\alpha}$-, β-, and γ-diversity against land use revealed that land use significantly affected the diversity measures across all datasets. Following our first objective, we demonstrated that β-diversity declined with agriculture across all datasets. Gamma diversity usually decreased, whereas $\overline{\alpha}$-diversity often increased with agriculture (Table 2.2). Except for French fish, SADs were generally significantly more uneven for agricultural land use than forest (higher residual σ), although the differences were mainly small (Fig. 2.4, columns 1-3). Intraspecific aggregation (1/*k*) was always greater in forest than in agriculture and was negatively correlated with residual σ, meaning more even SADs were always associated with higher aggregation (Appendix 2.2, Table S2.2). Skewness was significantly positive in the insect and all three diatom datasets, but non-significant in the two fish datasets. When positive, skewness correlated positively with residual σ regardless of land cover (although weakly for diatoms), indicating that SAD unevenness was generally characterized by greater abundances of more common species.

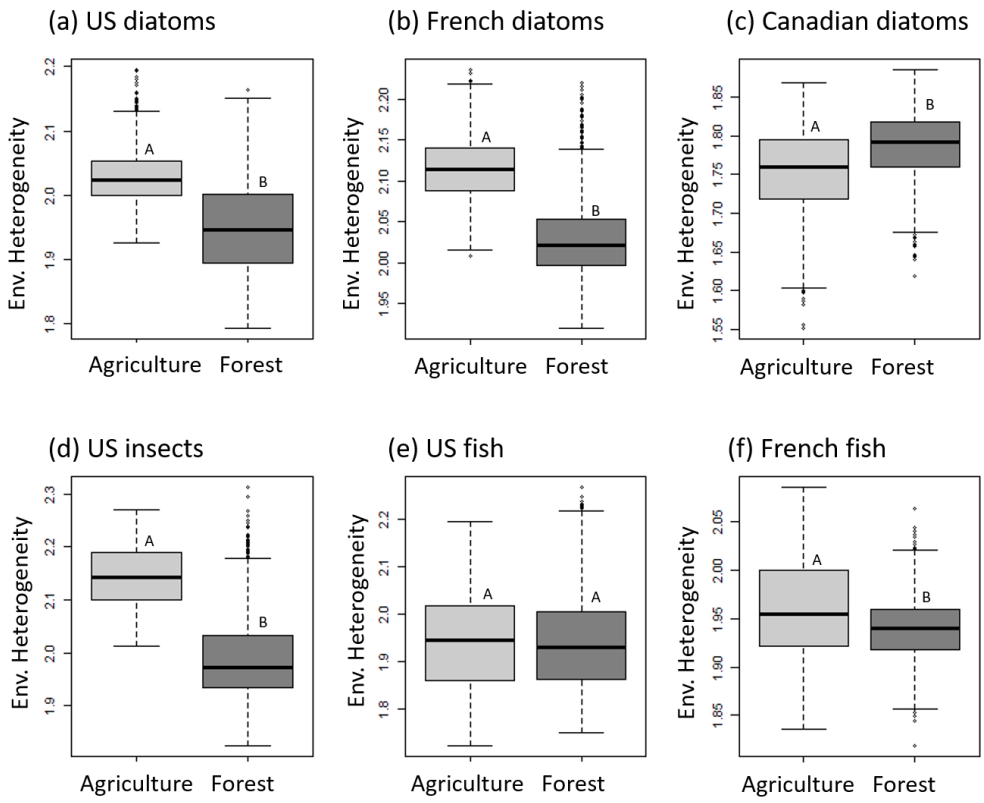

Fig. 2.3 a-f. Boxplots showing differences in resampled physicochemical heterogeneity between land covers for each dataset. a = US diatoms, b = French diatoms, c = Canadian diatoms, d = US insects, e = US fish, f = French fish. Points indicate resamples that fall outside the interquartile range. Different letters denote significant differences in mean heterogeneity (permutational ANOVA, $P < 0.05$).

Table 2.2. Summary of the impact of agricultural land use on resampled diversity measures as positive or negative percent change relative to forest cover. Significant differences between land covers were detected in all comparisons (permutational MANOVA and ANOVA, $P < 0.05$).

		% Change from agriculture		
Taxonomic group	Country	$\overline{\alpha}$	γ	$^{1}\beta_{OBS}$
Diatoms	US	+20.71	−3.54	−1.09
	France	+13.33	−7.46	−2.22
	Canada	−12.15	−23.14	−4.64
Insects	US	−20.42	−22.98	−0.59
Fish	US	+9.55	+12.97	−2.29
	France	+54.99	+26.91	−6.41

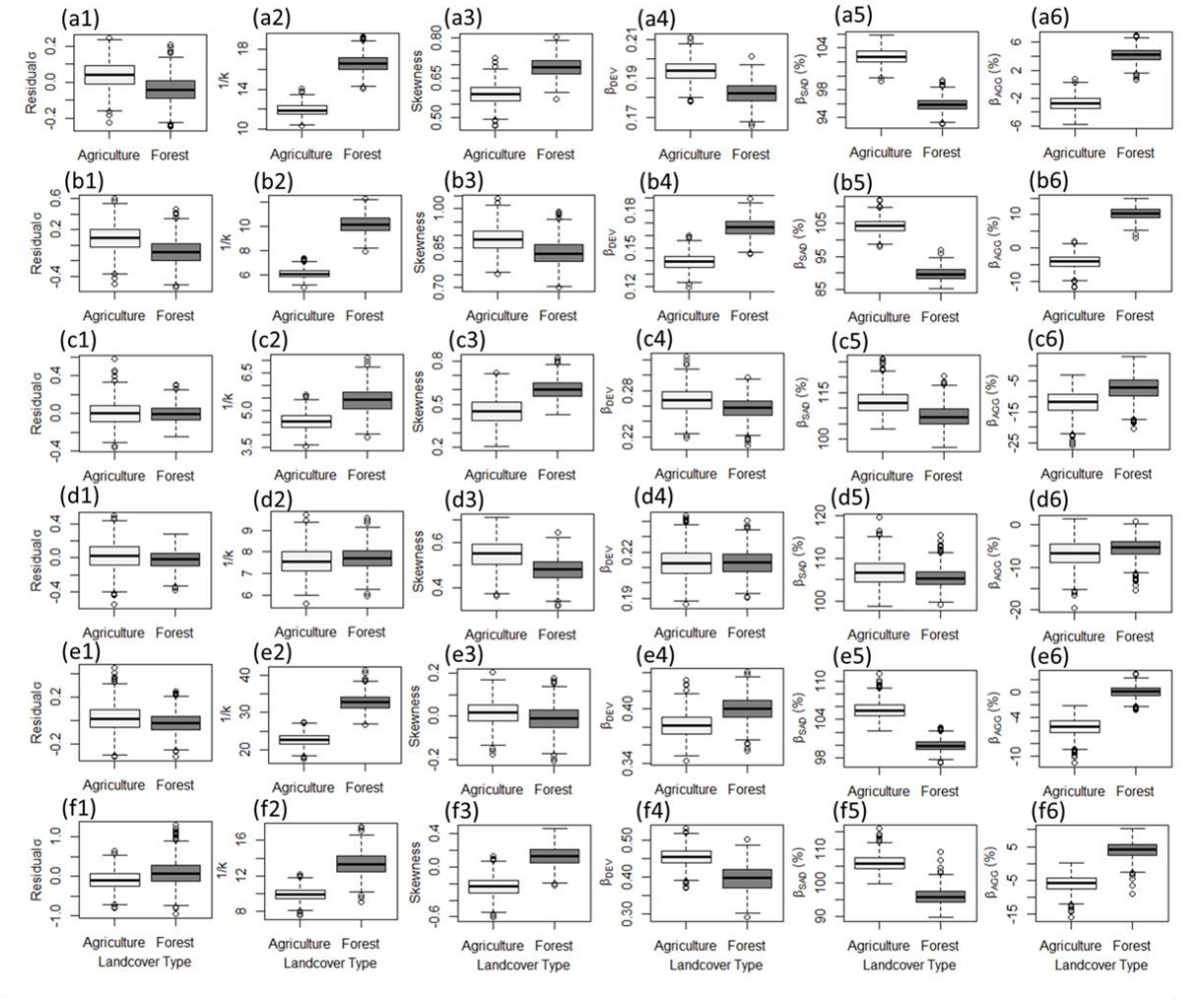

Fig. 2.4, a-f. Boxplots of resampled SAD and null model metrics showing the differences between land covers for each dataset. a = US diatoms, b = French diatoms, c = Canadian diatoms, d = US insects, e = US fish, f = French fish. Significant differences were observed in all comparisons (permutational ANOVA, $P < 0.05$) except β_{DEV} for US insects (panel d3).

Our first objective was to determine how biodiversity explained by land use, eutrophication, and physicochemical heterogeneity. Variation in all diversity measures was primarily explained by covariance effects, while pure land cover, pure eutrophication, and pure physicochemical heterogeneity contributed minorly (Fig. 2.5). In general, covariance of eutrophication with land cover explained most of the variation, indicating that land use constrained biotic variability mainly through eutrophication rather than physicochemical heterogeneity. However, the insect dataset differed from the rest in that the covariance fraction of land cover, eutrophication, and physicochemical heterogeneity captured most of the variation.

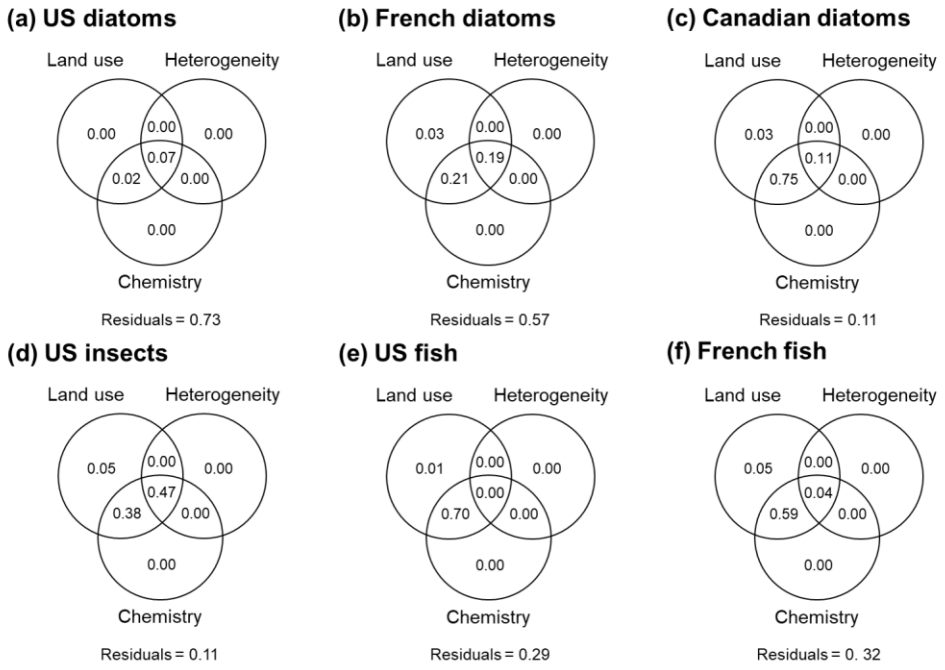

Fig. 2.5, a-f. Venn diagrams showing output of redundancy analysis-based variance partitioning of diversity measures ($\overline{\alpha}$-, β-, and γ-diversity). Values represent model adjusted R^2 values. Values in intersections represent covariance fractions, whereas values in circles represent pure fractions.

Eutrophication-associated shifts in local assembly across organismal groups

For our second objective, we found local assembly weakly drove diatom and insect β-diversity (β_{DEV} generally less than 0.26 across land covers) but had a relatively greater influence on fish β-diversity (β_{DEV} between 0.38-0.45). β_{DEV} differed significantly between forest and agriculture (permutational ANOVA) in all datasets except insects (no difference). However, the magnitude of the difference in β_{DEV} was usually small (3.49 to 16.04%), with the direction of the difference depending on organismal group and biogeographic region (Fig. 2.4, column 4).

Contribution of the SAD vs. intraspecific spatial aggregation to β_{DEV}

For objective three, the partitioning of β_{DEV} revealed that β_{SAD} generally exceeded 100% and β_{AGG} was negative, regardless of land cover except for the US and French diatom datasets, which showed $\beta_{SAD} < 100\%$ and positive β_{AGG} for forest land use (Fig. 2.4, columns 5-6). As changes in β_{SAD} correspond to equal and opposite changes in β_{AGG}, we focus on β_{SAD} for brevity. β_{SAD} represented nearly all of β_{DEV} regardless of dataset and land cover type (~ 90-110% of total deviance) and was significantly (although marginally) larger in agricultural land use than in forest cover. Further, β_{DEV} was generally negatively correlated with residual σ, regardless of land cover or organismal group, implying that increased SAD unevenness was usually associated with greater contribution of the regional species pool (Appendix 2.3, Table S2.2). Variance partitioning of β_{DEV} across datasets showed mixed patterns among and within organismal groups over what effects best explained β_{DEV} (Fig. 2.6).

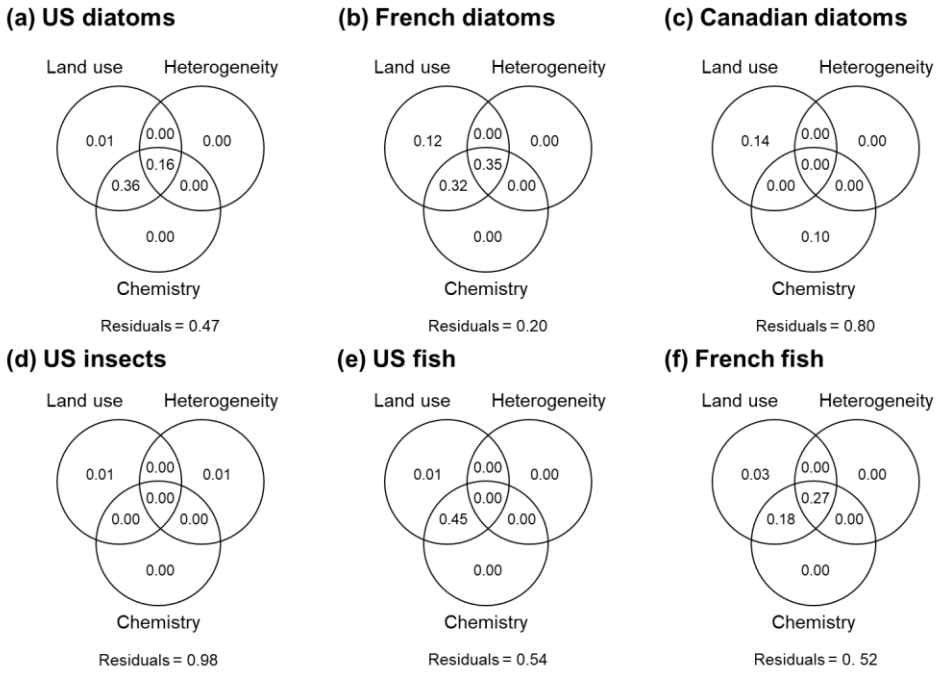

Fig. 2.6, a-f. Venn diagrams showing output of regression-based variance partitioning of β_{DEV}. Values represent model adjusted R^2 values. Values in intersections represent covariance fractions whereas values in circles represent pure fractions.

Variability across organismal groups

Consistent with our fourth objective, we demonstrated that smaller organisms (diatoms and insects) with greater dispersal capacity were more similar in terms of SAD and β_{DEV} patterns, but diverged from fish. However, we also observed divergence in some ecological patterns between datasets within organismal groups (i.e., diatoms and fish) in that $\overline{\alpha}$-diversity, γ-diversity, SAD

skewness, and β_{DEV} responses varied between country of origin, which indicated context dependency of our results.

Discussion

In this comprehensive study of stream organisms from two continents, agriculture and subsequent eutrophication were generally associated with reduced β- and γ-diversity and increased $\overline{\alpha}$-diversity. First, covariance of land use with physicochemical gradients, rather than with physicochemical heterogeneity, characterized regional biodiversity loss with land use. Second, all datasets showed significant shifts in magnitude of β_{DEV} with eutrophication but the direction (i.e., stronger or weaker local assembly effects) depended on organismal group and potentially biogeographical factors. Third, the regional SAD overrode intraspecific spatial aggregation in explaining β_{DEV} and its influence and unevenness increased with agriculture.

Eutrophication and Environmental Heterogeneity Effects on Diversity and the SAD

With respect to objective one, regional biodiversity loss, local diversity gains, and increased community similarity in aquatic taxa were correlated with agricultural land use, consistent with patterns expected for taxonomic homogenization (Petsch, 2016). Recently, Ribiero *et al*. (2015) explored the generality of floral homogenization consequential of agricultural land use and noted that too many studies focus on a single spatial scale or a single taxon. For aquatic taxa, agriculturally-associated changes in β-diversity have been reported, however, we have only begun to examine these changes at broader spatial scales. For example, Winegardner *et al*. (2017) attributed greater temporal β-diversity of diatoms across modified US landscapes to richness gains and losses stemming from disproportionate influence of contemporary vs. past

land use, yet observed no changes in spatial β-diversity. In contrast, diatom spatial β-diversity declined with eutrophication in French streams (Jamoneau, Passy, Soininen, Leboucher & Tison-Rosebery, 2018). Our investigation, exploring diatoms, insects, and fish across regional to subcontinental scales, demonstrates that the detrimental effects of agriculture on the regional biodiversity in stream ecosystems are independent of species biology or scale.

We further revealed that biodiversity variation between forest and agriculture was mainly driven by land use differences in physicochemistries rather than physicochemical heterogeneity, a result contrary to conventional wisdom that higher environmental heterogeneity brings greater turnover. While agriculture may homogenize the landscape, we show that it tended to lead to greater stream physicochemical heterogeneity, possibly due to variability in fertilization and landscape management regimes. Heterogeneity is an important mechanism of co-existence because it offsets competitive exclusion (Tilman & Pacala, 1993). However, we observed that physicochemical heterogeneity poorly explained β-diversity, because eutrophication in agricultural streams may have exceeded the physiological thresholds of sensitive species and decoupled compositional and environmental variability (Bini, Landeiro, Padial, Siqueira & Heino, 2014). The lack of a relationship may also be due to our measure of heterogeneity, which did not incorporate other aspects of heterogeneity, such as variability in substrate size, known to diminish with agriculture (Allan, 2004).

Increased prevalence of common species over spatial and temporal scales is a hallmark of taxonomic homogenization (Olden & Rooney, 2006), but our findings are restricted to the spatial dimension. Notably, while across datasets SADs were generally more uneven in agriculture, they were more positively skewed compared to forest only in two datasets, i.e. US insects and French diatoms. In these datasets, homogenization in agriculture was characterized by greater

prevalence of common relative to rare species, which has also been observed in terrestrial arthropods (Simons, Gossner, Lewinsohn, Lange, Türke et al., 2015; Komonen & Elo, 2017). However, SADs were more positively skewed in forest cover than in agriculture for two datasets (US and Canadian diatoms), and not skewed for both fish datasets. This suggested that stronger SAD unevenness in agriculture resulted from either buildup of common species or greater regional dominance by a relatively few species. Like recent terrestrial and tropical studies (Vázquez & Gaston, 2004; Lohbeck, Bongers, Martinez-Ramos & Poorter, 2016), we showed that SAD unevenness was associated with agriculturally-driven homogenization. Future research on homogenization should incorporate novel methods and procedures, like we employed, to elucidate how habitat modification and trait distribution contribute to the two forms of unevenness, i.e. asymmetry vs. dominance.

Land use-associated shifts in local assembly across organismal groups

Following objective two, we examined how local assembly (β_{DEV}) varied between forested and agricultural streams. In general, β_{DEV} marginally differed between land covers, suggesting that the strength of local vs. regional mechanisms was relatively unaffected by physicochemical stressors, consistent with prior work, reporting that fire disturbance altered β-diversity but not its causes (Myers *et al.*, 2015). Community comparisons revealed that the magnitude of β_{DEV} usually increased with body size, which here was linked with dispersal capacity. Smaller β_{DEV} values in diatoms and insects indicated that the observed species pool exerted greater influence on β-diversity relative to local assembly. These results are corroborated by earlier research showing that diatom and insect communities are unsaturated, whereby local richness is limited by the size of the regional pool as opposed to local interactions (Passy, 2009; Al-Shami, Heino,

Che Salmah, Abu Hassan, Suhaila et al., 2013; but see Thornhill, Batty, Death, Friberg & Ledger, 2017). Therefore, it is possible that regional effects play a greater role in structuring local richness and β-diversity of smaller and more dispersive organisms than of larger and less dispersive organisms, and these relationships are not consistently affected by eutrophication.

In contrast, β_{DEV} in both fish datasets approaching 0.50 suggested relatively similar local and regional control of β-diversity, in agreement with prior observations of comparable contributions of regional and local factors to fish richness (Angermeier & Winston, 1998). Taxonomic homogenization is a particularly prevalent phenomenon among freshwater fish (Petsch, 2016) and our study elucidated that the possible causes include both local and regional processes.

Other nearly uniform patterns, independent of land use, were the negative correlation of residual σ of the regional SAD and the positive correlation of intraspecific aggregation (1/*k*) with β_{DEV}. These correlations indicated that more even regional SADs and increased intraspecific spatial aggregation were associated with stronger local constraints on β-diversity. Recent work has only begun to explore the relationship of SAD evenness with taxonomic homogenization, showing clear links between the two with implications for conservation (e.g., Simons *et al.*, 2015; Komonen & Elo, 2017). Our study is novel in that it demonstrates that local and regional processes controlling β-diversity are dependent on SAD evenness—a finding that could guide future stream conservation and management decisions, which need to be scale-explicit. For example, if preserving β-diversity, then adopting practices promoting abundance of less common species may be beneficial, given that SAD evenness is positively correlated with β-diversity.

The contribution of the SAD vs. intraspecific spatial aggregation to β_{DEV}

To our knowledge, we are the first to explore how land use affects partitioning of β_{DEV} into SAD vs. spatial aggregation fractions, i.e. β_{SAD} vs. β_{AGG} (objective three). β_{SAD} accounted for most of β_{DEV}, similar to observations for global tree communities (Xu *et al.*, 2015), but opposite to findings, with a different null model, for Czech forests (Sabatini, Jiménez-Alfaro, Burrascano, Lora & Chytrý, 2017). We further discovered that β_{SAD} largely exceeded β_{AGG} across organismal groups, datasets, and land cover types. However, β_{SAD} was significantly higher in agriculture compared to forest in all datasets. The two land covers also diverged in β_{AGG}—less spatial aggregation than predicted by the null model ($\beta_{AGG} < 0$) was detected in agriculture across all datasets, while some aggregation ($\beta_{AGG} > 0$) was observed in forest streams in four out of six datasets. These results suggest that although land use did not constrain the magnitude of local assembly effects (β_{DEV}), it did control the mechanisms of local assembly, i.e. land use increased the role of the SAD, but diminished the influence of aggregation.

Organismal and geographic dependencies in biodiversity response to homogenization

In pursuit of our fourth objective, we found that organismal groups responded differently to land use, as reported by other studies (e.g., Angermeier & Winston, 1998; Thornhill *et al.*, 2017). Insects resembled diatoms in biodiversity, SAD shape, and β_{DEV} patterns, which suggested that body size and dispersal capacity may be more important than trophic position (autotroph vs. heterotroph) in predicting ecological responses to agricultural eutrophication. We generally expected consistent responses of these metrics to agriculture, regardless of country of origin (i.e., diatoms and fish). We reasoned that agriculture, being a major habitat alteration, will override all other influences, yet within both groups, there was divergence depending on region. We ensured

that variation in individual counts and mean counts among samples and differences in geographic spread across datasets did not contribute to their dissimilarity (data not shown). Thus, our findings of within-taxon variability with respect to biodiversity and the SAD highlighted the importance of considering context dependency. Histories of land use disturbance among geographic regions can set biodiversity and relative abundance patterns on different trajectories by affecting processes underlying β-diversity (Cramer, Hobbs & Standish, 2008). For example, European fish diversity has been historically depauperate relative to North American fauna owing particularly to differences in glacial influence (Oberdorff, Hugueny & Guégan, 1997). Furthermore, French aquatic communities have been impacted by agricultural activities far longer than their North American counterparts (Hahn & Orrock, 2015).

In summary, we determined eutrophication is a major driver of β-diversity losses among stream taxa, although the importance of geographic context was shown through the varied biodiversity responses within taxonomic groups. Local assembly generally was weakly affected by agriculture. However, in agriculture the regional SAD became significantly more uneven and its effect on local assembly significantly increased compared to forest, which may be the underlying causes of taxonomic homogenization. Biodiversity, SAD shape, and β_{DEV} depended more strongly on body size and/or dispersal than trophic position. Future research should explore how local and regional processes operate in tandem with the SAD to uncover whether homogenization drivers are specific to organismal groups and the regions from which they were sampled. Here we examined β-diversity loss from a taxonomic perspective. We recommend future investigations on agriculture-driven homogenization to be conducted across space and time and on taxonomic, phylogenetic, and functional diversity for more holistic understanding of its causes and patterns.

References

Afnor, N.F. (2007) T90-354, Qualite de l'eau. Determination de l'Indice Biologique Diatomees (IBD). 1-79.

Al-Shami, S.A., Heino, J., Che Salmah, M.R., Abu Hassan, A., Suhaila, A.H. & Madrus, M.R. (2013) Drivers of beta diversity of macroinvertebrate communities in tropical forest streams. *Freshwater Biology*, **58**, 1126-1137.

Allan, J.D. (2004) Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 257-284.

Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**, 683-693.

Angermeier, P.L. & Winston, M.R. (1998) Local vs. regional influences on local diversity in stream fish communities of Virginia. *Ecology*, **79**, 911-927.

Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Mac Nally, R., . . . Oliveira, V.H.F. (2016) Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature*, **535**, 144.

Bini, L.M., Landeiro, V.L., Padial, A.A., Siqueira, T. & Heino, J. (2014) Nutrient enrichment is related to two facets of beta diversity for stream invertebrates across the United States. *Ecology*, **95**, 1569-1578.

Cramer, V.A., Hobbs, R.J. & Standish, R.J. (2008) What's new about old fields? Land abandonment and ecosystem assembly. *Trends in Ecology & Evolution*, **23**, 104-112.

de Juan, S., Thrush, S.F. & Hewitt, J.E. (2013) Counting on β-diversity to safeguard the resilience of estuaries. *Plos One*, **8**, e65575.

Devictor, V., Julliard, R., Clavel, J., Jiguet, F., Lee, A. & Couvet, D. (2008) Functional biotic homogenization of bird communities in disturbed landscapes. *Global Ecology and Biogeography*, **17**, 252-261.

European Environment Agency (2013) Corine Land Cover 2006 seamless vector data (Version 17). In. European Environment Agency, Kopenhagen, Denmark.

Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061-1063.

Flohre, A., Fischer, C., Aavik, T., Bengtsson, J., Berendse, F., Bommarco, R., . . . Eggers, S. (2011) Agricultural intensification and biodiversity partitioning in European landscapes comparing plants, carabids, and birds. *Ecological Applications*, **21**, 1772-1781.

Gonzalez, A., Cardinale, B.J., Allington, G.R.H., Byrnes, J., Arthur Endsley, K., Brown, D.G., . . . Loreau, M. (2016) Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology*, **97**, 1949-1960.

Hahn, P.G. & Orrock, J.L. (2015) Land-use history alters contemporary insect herbivore community composition and decouples plant–herbivore relationships. *Journal of Animal Ecology*, **84**, 745-754.

He, F. & Legendre, P. (2002) Species diversity patterns derived from species–area models. *Ecology*, **83**, 1185-1198.

Herve, M. (2018) RVAideMemoire: Testing and plotting for biostatistics. R Package version 0.9-69-3.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Jamoneau, A., Passy, S.I., Soininen, J., Leboucher, T. & Tison-Rosebery, J. (2018) Beta diversity of diatom species and ecological guilds: Response to environmental and spatial mechanisms along the stream watercourse. *Freshwater Biology*, **63**, 62-73.

Komonen, A. & Elo, M. (2017) Ecological response hides behind the species abundance distribution: Community response to low-intensity disturbance in managed grasslands. *Ecology and Evolution*, **7**, 8558-8566.

Komsta, L. & Novomestky, F. (2015) moments: Moments, Cumulants, skewness, kurtosis, and related tests. R package version 0.14.

Kraft, N.J.B., Comita, L.S., Chase, J.M., Sanders, N.J., Swenson, N.G., Crist, T.O., . . . Myers, J.A. (2011) Disentangling the Drivers of β Diversity Along Latitudinal and Elevational Gradients. *Science*, **333**, 1755.

Lavoie, I., Campeau, S., Zugic-Drakulic, N., Winter, J.G. & Fortin, C. (2014) Using diatoms to monitor stream biological integrity in Eastern Canada: An overview of 10 years of index development and ongoing challenges. *Science of The Total Environment*, **475**, 187-200.

Lohbeck, M., Bongers, F., Martinez-Ramos, M. & Poorter, L. (2016) The importance of biodiversity and dominance for multiple ecosystem functions in a human-modified tropical landscape. *Ecology*, **97**, 2772-2779.

Márquez, J.C. & Kolasa, J. (2013) Local and regional processes in community assembly. *Plos One*, **8**, e54580.

Myers, J.A., Chase, J.M., Crandall, R.M. & Jiménez, I. (2015) Disturbance alters beta-diversity but not the relative importance of community assembly mechanisms. *Journal of Ecology*, **103**, 1291-1299.

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

Oberdorff, T., Hugueny, B. & Guégan, J.F. (1997) Is there an influence of historical events on contemporary fish species richness in rivers? Comparisons between Western Europe and North America. *Journal of Biogeography*, **24**, 461-467.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2017) vegan: Community Ecology Package. R Package Version 2.4-2.

Olden, J.D. & Rooney, T.P. (2006) On defining and quantifying biotic homogenization. *Global Ecology and Biogeography*, **15**, 113-120.

Passy, S.I. (2008) Continental diatom biodiversity in stream benthos declines as more nutrients become limiting. *Proceedings of the National Academy of Sciences*, **105**, 9663-9667.

Passy, S.I. (2009) The relationship between local and regional diatom richness is mediated by the local and regional environment. *Global Ecology and Biogeography*, **18**, 383-391.

Penaluna, B.E., Olson, D.H., Flitcroft, R.L., Weber, M.A., Bellmore, J.R., Wondzell, S.M., . . . Reeves, G.H. (2017) Aquatic biodiversity in forests: a weak link in ecosystem services resilience. *Biodiversity and Conservation*, **26**, 3125-3155.

Petsch, D.K. (2016) Causes and consequences of biotic homogenization in freshwater ecosystems. *International Review of Hydrobiology*, **101**, 113-122.

Prado, P.I., Mirands, M.D. & Chalom, A. (2017) "sads": Maximum Likelihood Models for Species Abundance Distributions. R package Version 0.4-1.

Sabatini, F.M., Jiménez-Alfaro, B., Burrascano, S., Lora, A. & Chytrý, M. (2017) Beta-diversity of central European forests decreases along an elevational gradient due to the variation in local community assembly processes. *Ecography*, **41**, 1038-1048.

Sala, O.E., Stuart Chapin, F., Iii, Armesto, J.J., Berlow, E., Bloomfield, J., . . . Wall, D.H. (2000) Global Biodiversity Scenarios for the Year 2100. *Science*, **287**, 1770-1774.

Simons, N.K., Gossner, M.M., Lewinsohn, T.M., Lange, M., Türke, M. & Weisser, W.W. (2015) Effects of land-use intensity on arthropod species abundance distributions in grasslands. *Journal of Animal Ecology*, **84**, 143-154.

Socolar, J.B., Gilroy, J.J., Kunin, W.E. & Edwards, D.P. (2016) How Should Beta-Diversity Inform Biodiversity Conservation? *Trends in Ecology & Evolution*, **31**, 67-80.

Soininen, J., Jamoneau, A., Rosebery, J. & Passy, S.I. (2016) Global patterns and drivers of species and trait composition in diatoms. *Global Ecology and Biogeography*, **25**, 940-950.

Solar, R.R.d.C., Barlow, J., Ferreira, J., Berenguer, E., Lees, A.C., Thomson, J.R., . . . Oliveira, V.H.F. (2015) How pervasive is biotic homogenization in human-modified tropical forest landscapes? *Ecology Letters*, **18**, 1108-1118.

Thornhill, I., Batty, L., Death, R.G., Friberg, N.R. & Ledger, M.E. (2017) Local and landscape scale determinants of macroinvertebrate assemblages and their conservation value in ponds across an urban land-use gradient. *Biodiversity and Conservation*, **26**, 1065-1086.

Tilman, D. & Pacala, S. (1993) The maintenance of species richness in plant communities. In R.E. Ricklefs & D. Schulter (Eds.), *Species Diversity in Ecological Communities* (pp. 13-25). Chicago, IL: Chicago Press.

Vázquez, L.B. & Gaston, K.J. (2004) Rarity, commonness, and patterns of species richness: the mammals of Mexico. *Global Ecology and Biogeography*, **13**, 535-542.

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

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

Wiens, J.J. (2016) Climate-Related Local Extinctions Are Already Widespread among Plant and Animal Species. *PLOS Biology*, **14**, e2001104.

Winegardner, A.K., Legendre, P., Beisner, B.E. & Gregory-Eaves, I. (2017) Diatom diversity patterns over the past c. 150 years across the conterminous United States of America: Identifying mechanisms behind beta diversity. *Global Ecology and Biogeography*, **26**, 1303-1315.

Withers, P.J.A., Neal, C., Jarvie, H.P. & Doody, D.G. (2014) Agriculture and eutrophication: where do we go from here? *Sustainability*, **6**, 5853-5875.

Xu, W., Chen, G., Liu, C. & Ma, K. (2015) Latitudinal differences in species abundance distributions, rather than spatial aggregation, explain beta-diversity along latitudinal gradients. *Global Ecology and Biogeography*, **24**, 1170-1180.

Zobel, M. (2016) The species pool concept as a framework for studying patterns of plant diversity. *Journal of Vegetation Science*, **27**, 8-18.

Appendix 2.1. Expanded description of environmental data and null model correlation results.

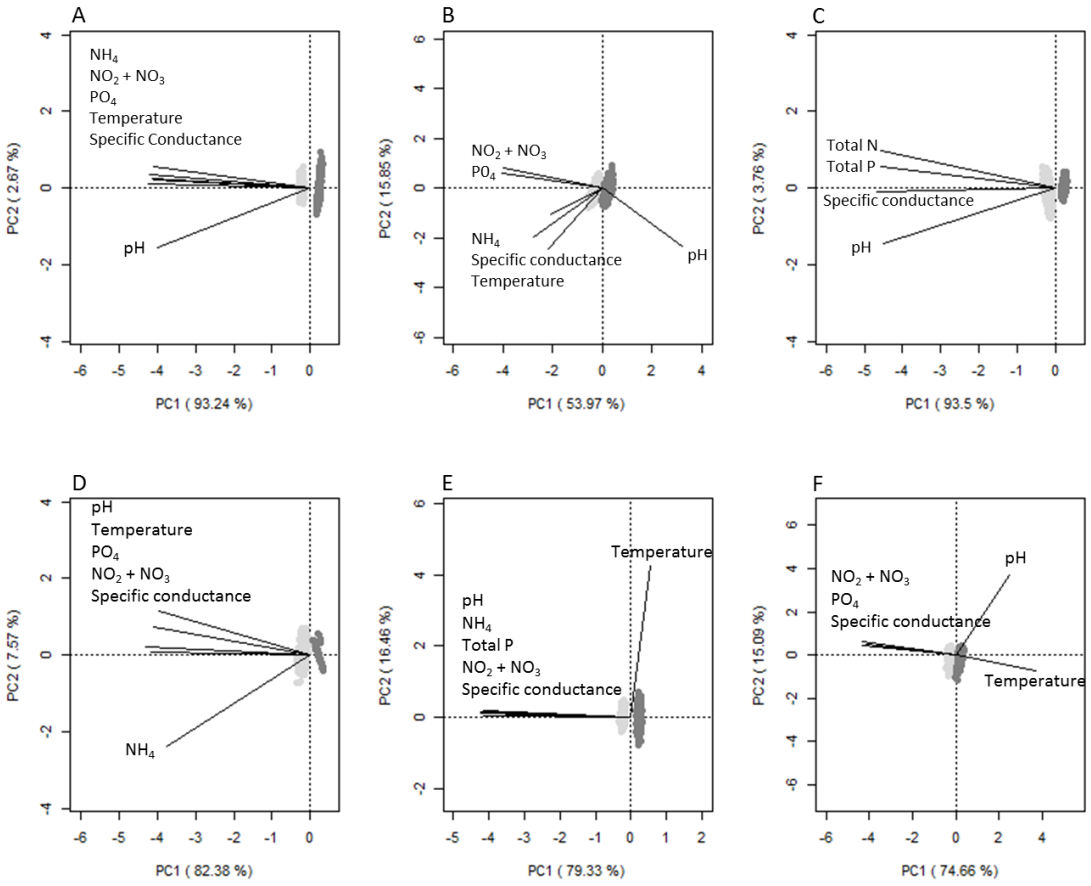

Fig. S2.1 Principal component analysis of resampled environmental data for each dataset. A = US diatoms, B = French diatoms, C = Canadian diatoms, D = US insects, E = US fish, F = French fish. Light gray = agricultural stream resamples, dark gray = forest stream resamples.

Table S2.1 Environmental and landscape variables used for analysis of environmental heterogeneity in land covers within each species dataset.

Organismal group	Country	Environmental variables	Landscape variables
Diatoms	US	pH Water Temperature (°C) Specific conductance (µS/cm) Ammonia (µg/L) Nitrite + Nitrate (µg/L) Orthophosphate (µg/L)	Drainage area (km^2) Site elevation (m)
	France	pH *Air Temperature (°C) Specific conductance (µS/cm) Ammonia (µg/L) Nitrite + Nitrate (µg/L) Orthophosphate (µg/L)	Drainage area (km^2)
	Canada	pH Specific conductance (µS/cm) Total nitrogen Total phosphorus	Drainage area (km^2)

Table S2.1 Continued

Organismal group	Country	Environmental variables	Landscape variables
Insects	US	pH Water Temperature (°C) Specific conductance (µS/cm) Ammonia (µg/L) Nitrite + Nitrate (µg/L) Orthophosphate (µg/L)	Drainage area (km^2) Site elevation (m)
Fish	US	pH	Drainage area (km^2)
		Water Temperature (°C)	Site elevation (m)
		Specific conductance (µS/cm)	
		Ammonia (µg/L)	
		Nitrite + Nitrate (µg/L)	
		Total phosphorus (µg/L)	
	France	pH	Drainage area (km^2)
		Air temperature (°C)	Site elevation (m)
		Ammonia (µg/L)	
		Nitrite + Nitrate (µg/L)	
		Orthophosphate (µg/L)	

Table S2.2. Table of Pearson correlations of null model and SAD parameters. All correlations were significant ($P < 0.05$) unless noted by asterisk.

Land cover	Dataset	σ vs. 1/*k*	σ vs. skew	σ vs. β_{DEV}	1/*k* vs. β_{DEV}
Agriculture	US diatoms	−0.101	0.141	−0.425	0.568
	French diatoms	−0.047*	0.331	−0.426	0.729
	Canadian diatoms	−0.287	0.610	−0.831	0.671
	US insects	−0.024*	0.577	−0.629	0.458
	US fish	−0.242	----	−0.752	0.487
	French fish	−0.067	----	−0.464	0.637
Forest	US diatoms	−0.155	0.102	−0.380	0.529
	French diatoms	−0.352	0.530	−0.703	0.607
	Canadian diatoms	−0.023*	0.388	−0.622	0.704
	US insects	−0.166	0.431	−0.581	0.506
	US fish	0.149	----	−0.532	0.265
	French fish	−0.449	----	−0.708	0.741

---- Skewness was not significant for this dataset

Appendix 2.2. Description of null model machinery

Although the null model we employed is well explained by Xu *et al.* (2015), we provide a convenient summary of the model here. This null model is a series of equations that calculates β-diversity values based on occupancy-abundance probability distributions (He and Gaston 2003). The null model procedure calculates three β-diversity values: observed β-diversity, expected β-diversity, and predicted β-diversity. In this study, observed β-diversity for each land use category was originally calculated as the fraction of the species pool not observed at a given site:

$$\beta_{OBS} = 1 - \frac{\overline{\alpha}}{\gamma} \quad (1).$$

However, this relationship can be equivalently re-expressed as the sum total of the proportion of sites occupied by the i^{th} species:

$$\beta_{OBS} = 1 - \frac{\sum_{i=1}^{\gamma} p_i}{\gamma} \quad (2).$$

The null model uses this alternate form of equation 1 as the base equation for this procedure. The p_i variable of equation 2 is treated as the focal parameter for the model procedure as two additional beta diversity values (expected and predicted β-diversity) are generated by substituting in occupancy-abundance models for p_i. Expected β-diversity (β_{EXP}) is first derived using an occupancy-abundance model that assumes the probability of sampling the i^{th} species with n individuals across m sites is a Bernoulli trial:

$$p_i = 1 - \left(1 - \frac{1}{m}\right)^{n_i} \quad (3).$$

Substitution of equation (3) for p_i in equation (2) calculates expected β-diversity, or the β-diversity expected under a completely random distribution of individuals across sites:

$$\beta_{EXP} = \frac{\left(\sum_{i=1}^{y}\left(1-\frac{1}{m}\right)^{n_i}\right)}{\gamma} \quad (4).$$

Predicted β-diversity (β_{PRED}) is calculated with a non-random occupancy-abundance relationship that assumes the probability sampling the i^{th} species with *n* individuals across *m* sites follows a negative binomial distribution:

$$p_i = 1 - \left(1 + \frac{n_i}{mk}\right)^{-k} \quad (5),$$

where *m* is the total number of sites and *k* is a parameter empirically estimated from the negative binomial fit and represents the average magnitude of intraspecific spatial aggregation across all samples (He and Gaston, 2000). Equation 5 then substitutes in for p_i in equation 2 to yield the predicted β-diversity model, which estimates β-diversity when intraspecific spatial aggregation is equal across species:

$$\beta_{PRED} = \frac{\left(\sum_{i=1}^{y}\left(1-\frac{n_i}{mk}\right)^{-k}\right)}{\gamma} \quad (6).$$

The negative-binomial *k* parameter in equations 5 and 6 is usually unknown, and therefore it must be empirically estimated using maximum likelihood methods. This is done by first pooling together all species and samples across datasets into a single metacommunity dataset and then maximizing the log-likelihood function:

$$l = \sum_{i=1}^{\gamma}[o_i log(p_i) + (m - o_i) log(1 - p_i)] \quad (7),$$

where o_i is the proportion of streams occupied by the i^{th} species, m is the total number of streams, and p_i is the predicted occupancy model (equation 5). This likelihood-estimated value of k is what is used for equations 5 and 6 to generate β_{PRED}.

Chapter 3
The impacts of nutrient supply and imbalance on subcontinental co-occurrence and interaction networks of stream algae

William R. Budnick[1], Joseph L. Mruzek[1], Chad A. Larson[2], and Sophia I. Passy[1*]

[1]Department of Biology, University of Texas at Arlington, 501 S. Nedderman Dr., Arlington Texas, 76019, USA

[2]Washington State Department of Ecology, Environmental Assessment Program, 300 Desmond Dr SE, Lacey, Washington, 98503, USA

Author Emails and ORCID IDs:
William R. Budnick: William.budnick@mavs.uta.edu; https://orcid.org/0000-0001-9288-6782
Joseph Mruzek: Joseph.mruzek@mavs.uta.edu; https://orcid.org/0000-0002-8067-3824
Chad A. Larson: clar461@ecy.wa.gov; https://orcid.org/0000-0002-0329-8979
Sophia I. Passy: Sophia.passy@uta.edu; https://orcid.org/0000-0002-8230-9380

*Corresponding Author

Introduction

The world's freshwater ecosystems are undergoing global changes in nutrient inputs, which are either increasing, causing eutrophication and deterioration of water quality (Dodds & Smith, 2016; Stoddard *et al.*, 2016) or decreasing, leading to oligotrophication and potentially restoration of the original state (Flaim *et al.*, 2016; Verbeek *et al.*, 2018). As anthropogenic factors continue to force the pendulum to shift among nutrient extremes, how species and communities respond is a topic of great research interest and environmental concern (Tilman & Isbell, 2015; Wang *et al.*, 2016).

Numerous investigations have shown that both availability and balance (i.e., relative proportions) of major nutrients, such as nitrogen and phosphorus, constrain biodiversity, composition, and biomass production across ecosystems (Elser *et al.*, 2007; Cardinale *et al.*, 2009; Harpole *et al.*, 2011; Lewandowska *et al.*, 2016; Cook *et al.*, 2018). However, we are just beginning to understand the impacts of nutrient supply and balance on the topology of species co-occurrence networks. Co-occurrence networks graphically represent the pairwise species relationships within a metacommunity, and the topological properties of these networks are sensitive to ecological gradients (Poisot *et al*., 2015). For example, eutrophication was shown to increase species inter-connectedness (e.g., the mean number of neighbors) and decrease network subdivision into modules (Rocha *et al*., 2015, Cao *et al*., 2018, Wang *et al*., 2019) with potentially negative consequences for network resilience and susceptibility to disturbance. Nutrient balance effects (e.g., N-limitation vs. P-limitation) on network parameters are much less studied, yet modularity and connectance in heterotrophic communities were demonstrated to respond to nutrient ratios (Larsen *et al.*, 2019).

Notably, our knowledge of nutrient effects on the topology of co-occurrence networks comes primarily from terrestrial systems (but see Qu *et al*. 2019). This is concerning as aquatic systems are disproportionately susceptible to human influences because of their relative isolation within the surrounding landscape matrix (Woodward *et al.*, 2010). Although we noted that past studies have compared co-occurrence network topologies across nutrient categories, e.g. oligotrophic, eutrophic, N-, and P-limited, none have explored to what extent they are driven by shared niches, dispersal, or actual interspecific interactions, given that all of these factors could underlie species co-occurrence relationships (Morueta-Holme *et al*. 2016).

The goal of this investigation was to determine how nutrient supply and imbalance affect properties of co-occurrence networks in stream benthic algae, growing under oligotrophic (low N and P concentrations), eutrophic (high N and P concentrations), and N- or P-limited conditions (intermediate N and P concentrations but Redfield ratio below or above 16:1, respectively). We focused on network size, i.e. number of nodes (species) and edges (species co-occurrences), edge characteristics (mean proportion of positive links), connectivity (mean shortest path length and connectance), clustering (local and global), and modularity. Edges can be positive or negative in value and their relative proportions indicate the prevalence of mutualism and facilitation vs. competition, which has implications for network stability (Mougi & Kondoh, 2012; Suweis *et al.*, 2014). Interconnectedness among nodes (Dunne *et al.*, 2002; Estrada, 2007; Borrett *et al.*, 2010) and their tendency to form cliques (local or global) (Shirley & Rushton, 2005) can influence network susceptibility to disturbance and propagation of effects. Modularity reflects the overall subdivision of the network into discrete groups (modules), having the potential to represent functional groups or niche partitioning (Montoya *et al.*, 2015).

We provide the first study of how species co-occurrence networks respond to each of four nutrient contexts (oligotrophic vs. eutrophic and N-limited vs. P-limited) at a subcontinental scale, while controlling for climate and dispersal. We used as a model system benthic algae because their metacommunities are sensitive to nutrient, climatic and dispersal effects (Soininen, 2007; Verleyen *et al.*, 2009; Soininen *et al.*, 2016; Leboucher *et al.*, 2019) but it still remains unknown whether this sensitivity transcends to the level of network topology. We thus pursued the following two objectives: 1) quantify network topology under each nutrient context and test for differences and 2) determine the relative contributions of climate and space (surrogate of dispersal) to network topology vs. community composition and whether these factors differ between nutrient contexts.

In addressing objective 1, we test the hypothesis that nutrient contexts (supply and ratio) influence co-occurrence network topology. Greater algal biodiversity with nutrient addition (Hillebrand *et al.*, 2007; Passy & Larson, 2019) should translate into overall larger co-occurrence networks with more nodes and edges. Eutrophic conditions promote speciose overstory guilds (high profile and motile) that are sensitive to nutrient limitation, whereas understory low profile species dominate under nutrient limitation (Passy, 2007; Marcel *et al.*, 2017; Wu *et al.*, 2017). Eutrophication is, therefore, likely to produce networks with greater local clustering, modularity, and path lengths due to coexistence of different functional groups, while oligotrophic conditions may lead to greater network connectivity and global clustering among fewer and comparatively functionally uniform species. Algal functional responses to N:P ratios are varied and system-specific, making it difficult to predict how network topology may respond to N:P ratio, but some studies have reported a tendency of P-limitation to stimulate high profile diatoms, whereas N-

limitation to favor cyanobacteria and/or motile diatoms (Smith 1983, Stelzer & Lamberti 2001, Qu *et al*., 2019).

In addressing objective 2, we used a recently developed null model (Morueta-Holme *et al.*, 2016) to test hypotheses that gradients in climate, dispersal, or both structure metacommunity composition, which in turn constrains topologies of algal co-occurrence networks. In aquatic systems, climate drives local nutrient availability and ratio by affecting precipitation runoff and flows (Jeppesen *et al.*, 2009; Özen *et al.*, 2010) and metabolic requirements (Woodward *et al.*, 2010). Climate-nutrient relationships in turn impact competitive and facilitative mechanisms in algae (Bestion *et al.*, 2018; Marañón *et al.*, 2018), which can have consequences for algal co-occurrence patterns, although this topic has not been studied with network methods. At the spatial scale of this study, we expected climatic factors to substantially constrain network topology across nutrient contexts due to their pronounced influence on algal species composition and distribution (Pajunen *et al.*, 2016; Jyrkänkallio-Mikkola *et al.*, 2017). However, stochastic processes, which dominate in productive environments (Steiner & Leibold, 2004; Chase, 2010; Leboucher *et al.*, 2019), could more strongly control eutrophic than oligotrophic network topology, but this question has not been explored so far. It is also unknown if or how networks differing in nutrient ratio would respond to climatic and spatial control. We therefore examined network topology as a function of climatic and spatial factors across nutrient contexts and tested whether these factors exercised a similar effect on metacommunity structure.

Methods

Datasets

We used data from 740 stream sites, where samples were taken from June to September between 1993-2011 by the National Water-Quality Assessment Program of the US Geological Survey. Algae were collected from a defined area in the richest-targeted habitats, which encompass hard substrates or macrophytes in faster currents. Generally, 25 cobbles, 5 woody snags or 5 macrophyte beds were sampled within a stream reach and the material was composited into a single sample. All taxa were identified mainly to species and their densities were measured as cells per cm^2. Species were categorized by ecological guild, following Passy (2007) and Passy & Larson (2011), including low profile (growing close to the substratum), high profile (extending into the biofilm matrix), and motile (fast moving).

Each sample was classified based on nutrient supply (eutrophic or oligotrophic) and ratio (N-limited or P-limited). Total nitrogen (TN) and total phosphorus (TP) concentrations were below 0.669 and 0.025 mg/L, respectively, in oligotrophic samples but above 1.499 and 0.075 mg/L, respectively, in eutrophic samples (Dodds & Smith, 2016). Nitrogen-limited samples were identified as those with Redfield N:P ratios less than 16:1, whereas P-limited samples had ratios above 16:1 (Redfield, 1934). Nitrogen and phosphorus-limited sites had mesotrophic conditions with TN and TP between those in the oligotrophic and eutrophic sites. Each nutrient context dataset consisted of 185 sites. All samples were unique to each nutrient context and samples were as similarly geographically distributed as possible (Appendix 3.1, Fig. S3.1). Climatic predictors (bioclim variables 1-19, Fick & Hijmans, 2017) were downloaded for each site.

Co-occurrence networks

Prior to network construction (Fig. 3.1), we filtered the species by site metacommunity matrix for each nutrient context by removing rare species, which occurred in less than 10% of samples. We created weighted co-occurrence networks for each nutrient context by calculating a partial Spearman correlation matrix on relative species density, standardized with mean = 0 and standard deviation = 1. Partial Spearman correlations were used over raw correlations to account for the influence of indirect species associations and because the null model we employed requires their use (see Gradient Effects section). We used random matrix theory (RMT) methods from R-package "RMThreshold" (Menzel, 2016) to objectively determine thresholds below which correlations were likely spurious and should be removed from further analyses (Appendix 3.2).

After thresholding the partial correlation matrix with the RMT-selected correlation value, species with all correlations below the threshold were removed. The resulting correlation adjacency matrix was used to generate a weighted network using R package "igraph" (Csardi and Nepusz 2020). We then used a fast-greedy clustering algorithm to identify the modules in the network. After assigning the nodes into their modules, we calculated network topology measures, including number of nodes, number of edges, mean node degree, mean shortest path length, clustering coefficients (a.k.a transitivities), modularity, and number of modules. Mean shortest path length, measuring on average how many nodes lie between any two nodes in the network, was calculated with a harmonic mean.

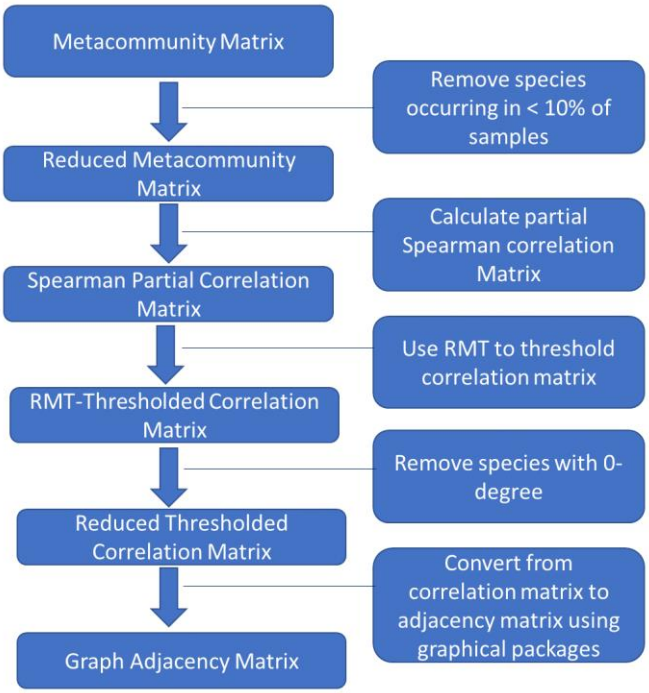

Fig. 3.1. Flow chart showing the construction of nutrient supply and ratio networks.

Gradient effects on co-occurrence networks

To explore whether species co-occurrence networks are driven by niche overlap along climatic gradients, small scale environmental filtering and dispersal (space), or both, we used the null model outlined by Morueta-Holme et al. (2016), which we briefly describe. This procedure compares observed partial correlation values with a distribution of predicted partial correlation values and removes observed partial correlations that are less extreme than predicted. The predicted correlation distributions are produced by calculating partial correlations on a predicted metacommunity matrix that contains the same species as the observed species-by-site matrix, but with abundances predicted using species distribution models (SDM). The SDM-predicted abundances represent those expected owing to climatic, spatial, or both gradients. The null model

procedure then randomly samples individuals using a lottery model for each site in the predicted metacommunity matrix and calculates a partial correlation matrix after randomization. The observed value is then statistically compared with the null expectation, and if found significantly more extreme, the observed value is retained in the matrix or otherwise deleted as it represents a shared response to a gradient.

Here, we modified the null model procedure to use Spearman partial correlation values rather than the Pearson values to account for non-linear but monotonic correlations. We used the partial correlation matrix created by the RMT procedure as the observed correlation matrix. Densities for each species, predicted by climate or spatial variables, or both, were calculated separately using optimally-parameterized boosted regression trees (BRT) with a Bernoulli link-function and trained using 10-fold cross-validation (Elith *et al.*, 2008). Spatial variables were generated using Moran's eigenvector maps (MEM) with functions in R package "adespatial" (Dray *et al.*, 2016). MEMs are orthogonal variables whose values represent spatial autocorrelation patterns. They are generated as the eigenvectors of a spatial weighting matrix describing the strength of the spatial relationship between pairs of samples. Using function "listw.select", we selected the best fitting combination of spatial weighting matrix and weighting definitions that generated a minimum of 10 significant MEMs (identified with permutational methods during the selection procedure) and best explained species densities (Bauman *et al.*, 2018).

BRTs for each species were optimized by iteratively finding the combination of tuning parameters (bag fractions ranging from 0.5 to 0.75, tree complexity parameters ranging from 1 to 5, and numbers of trees fit), which maximized area under the curve (AUC) values. Then, we refitted the models using the parameters corresponding to maximum AUC. For the BRT

procedure with the combined climate and spatial predictors, we used all climatic and spatial predictors. The BRTs produced probability estimates owing to the specified Bernoulli link function, which we transformed to densities by multiplying each species' observed density by the species' probability for the sample. When a species' probabilities could not be predicted with any combination of tuning parameters, we used its observed abundances in the predicted matrix because the model fits indicated that their observed presence patterns were no better predicted than a random expectation given the predictor variables used. We then combined the predicted abundances for each species into a single predicted metacommunity matrix and subsequently used it in the null model. We set the null model to run for 1000 iterations, transforming the randomized predicted densities to relative densities for each iteration as described for network construction, and identified insignificant correlations, which were then removed from the observed partial correlation matrix. We reconstructed the networks and calculated network parameters using these newly thresholded correlation matrices, which represented interaction networks free of co-occurrences due to climatic niche overlap and/or environmental filtering and dispersal.

Statistical analyses

We tested our first hypothesis by a resampling protocol to examine whether the ecological networks differed in their topology between nutrient contexts. This was done by subsampling 75% of the metacommunity (141 streams) and recalculating all network parameters as described previously for 9,999 iterations. We verified that the resampling procedure did not bias our comparisons by ensuring that 1) the observed topological measurement fell within the 95% range

of the resampled values, and 2) the medians of the resampled distributions followed the same ranking patterns across the nutrient contexts as the observed values (data not shown).

We note the oligotrophic and eutrophic data consisted of samples that were either N- or P-limited, therefore we could not make fair pairwise comparisons of these networks with the N- and P-limited networks. Therefore, we separated our statistical comparisons and assessed differences between oligotrophic vs. eutrophic networks and N-limited vs. P-limited networks. For both sets of comparisons, we used R package "RVAideMemoire" to first perform permutational MANOVA (function "pairwise.perm.manova", 999 permutations, Herve et al., 2018), which examined if the topological parameters differed between nutrient contexts in the nutrient supply and ratio networks. We followed significant results with permutational pairwise t-tests on each network parameter (function "perm.pairwise.t.test" with 999 permutations).
We used redundancy analysis-based variation partitioning (Oksanen *et al.*, 2019) on the metacommunity matrix of each nutrient context to determine if algal species densities responded to climatic and spatial factors, which we compared with the null model output for the respective networks to assess objective 2. On each metacommunity, climate variables were reduced to principal components axes to represent climatic factors, with the number of axes selected based on Kaiser-Guttman criterion (eigenvalues > 1, Peres-Neto *et al.*, 2005). Spatial factors were created using MEM, as described above for the BRT procedure, with significant MEMs retained using forward selection. The pure fractions of climate and space were tested using permutational analysis of variance (R package `vegan function "anova.rda"). After the null model analysis, we qualitatively assessed how climatic, spatial, and climatic + spatial factors influenced overall network topologies when compared with the original networks by calculating the absolute proportional change in each measured network parameter.

Results

All four networks varied in size, ranging from 66 nodes in the oligotrophic network to 97 nodes in the N-limited network (Appendix 3.1, Table S3.1). Examination of the edge counts and connectance revealed that all networks were sparse and poorly connected. Positive network edges generally constituted the majority of edges in the networks, although their proportion was the highest in the P-limited network. All networks consisted primarily of diatoms (86.3% – 92.4% of nodes) with minor representation by cyanobacteria (4.8% – 7.6%). The remaining algal groups generally comprised <5% of the nodes. Motile taxa dominated all networks (46.4-50.5% of all classified nodes) except for the oligotrophic network, where high profile taxa prevailed instead (39.4% of all classified nodes).

Topological Comparisons Across Contexts

Permutational MANOVA on subsampled topological measures (Table 3.1) indicated that the compared nutrient supply and ratio networks significantly differed from each other. Network size was significantly larger (greater numbers of nodes and edges) in the eutrophic and N-limited networks than the oligotrophic and P-limited networks, respectively. Other notable differences included higher clustering, number of modules and mean node degree in the eutrophic than the oligotrophic networks and greater proportions of positive edges in P-limited than N-limited networks. The remaining network parameters varied less between network contexts.

Table 3.1. Raw network parameters of algal co-occurrence networks differing in nutrient availability (oligotrophic vs. eutrophic) and nutrient ratio (N-limited vs. P-limited). All values were significantly different from each other (tested with permutational t-tests, $P < 0.05$, 999 permutations after significant permutational MANOVA).

Parameter	Nutrient supply Networks		Nutrient ratio Networks	
	Oligotrophic	Eutrophic	N-limited	P-limited
No. Nodes	66	93	97	83
No. Edges	84	144	164	124
% Positive Edges	0.60	0.56	0.55	0.66
Local Clustering	0.07	0.12	0.16	0.11
Global Clustering	0.09	0.16	0.15	0.09
Modularity	0.64	0.58	0.57	0.60
No. Modules	8	11	10	9
Mean Path Length	3.36	3.48	3.25	3.41
Mean Degree	2.55	3.10	3.38	2.99
Connectance	0.04	0.03	0.04	0.04

Null Model Analysis and Variance Partitioning

Qualitative examination of the networks produced by the null model procedure indicated that controlling for climatic, spatial, and combined climatic and spatial effects changed network topology, i.e. mean absolute % change in network parameters ranged between 38% and 72% on average (Fig. 3.2). Notably, controlling for spatial variables produced the greatest proportional change in network topology in the oligotrophic network, while climate + space control generated the greatest changes in the other three networks.

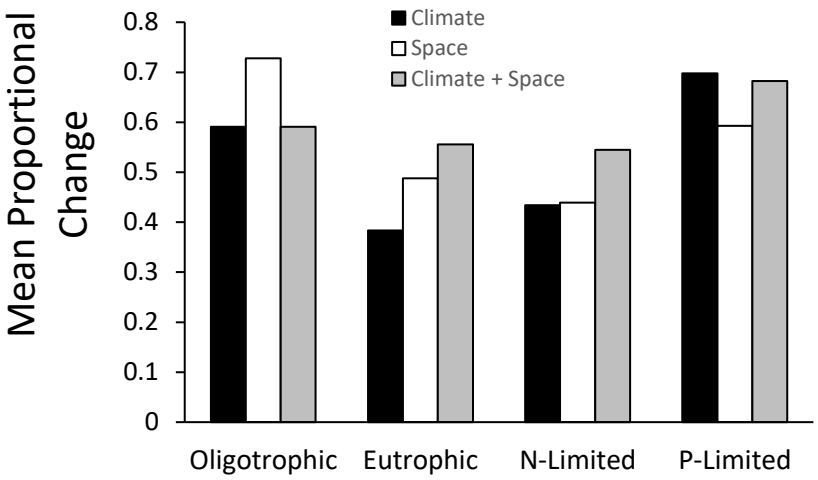

Fig 3.2. Mean proportional change in network topology from the original network across all network parameters after controlling for climatic, spatial, and climatic + spatial variables. For each controlled network, the absolute value of the proportional change in each network parameter was calculated and then averaged across all values to obtain the mean proportional change.

General patterns were observed in some parameters after climate and spatial control, owing in part to substantial increases in 0-degree nodes (i.e., nodes without any connections, Figs. 3.3, 3.4). Edge counts, connectance, and mean node degree decreased in a strongly correlated fashion across all controlled networks between 35% to 52.5%, whereas modularity values increased by 12.4% to 25.8%. Interestingly, some of the largest changes after control occurred in the P-limited networks, which underwent the greatest declines in mean degree and

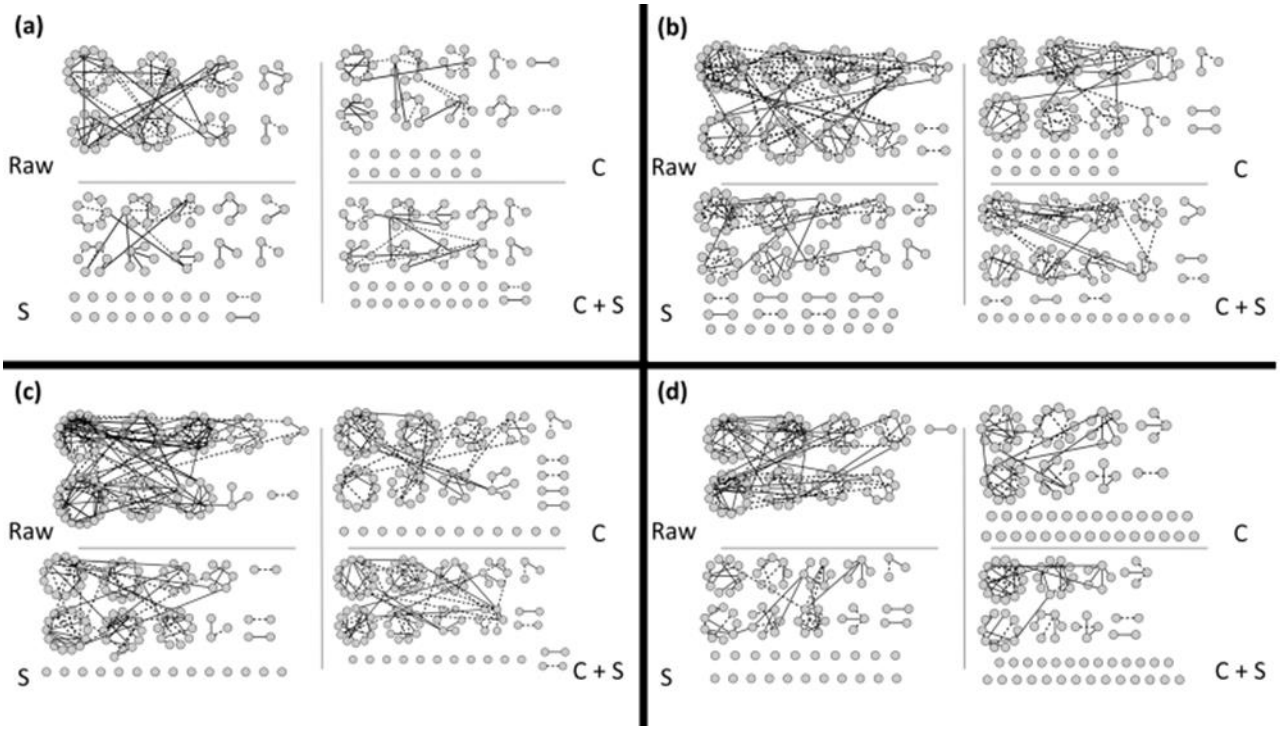

Figure 3.3. Raw networks and networks reconstructed after controlling for climate = C, space = S, and Climate + Space = C + S in (a) oligotrophic, (b) eutrophic, (c) N-limited, and (d) P-limited streams. Solid lines = positive co-occurrences, dashed lines = negative co-occurrences.

connectance and the greatest increases in modularity and number of modules. Responses of the remaining network parameters to control effects differed greatly among the networks. The oligotrophic network showed strong positive changes in mean local and global clustering in response to all control effects, whereas changes were relatively small and negative in the eutrophic network. Declines in clustering were similar in magnitude for both nutrient ratio networks. Responses in mean path length to controlling variables differed across the networks, however mean path length generally declined in the oligotrophic network, but increased in the

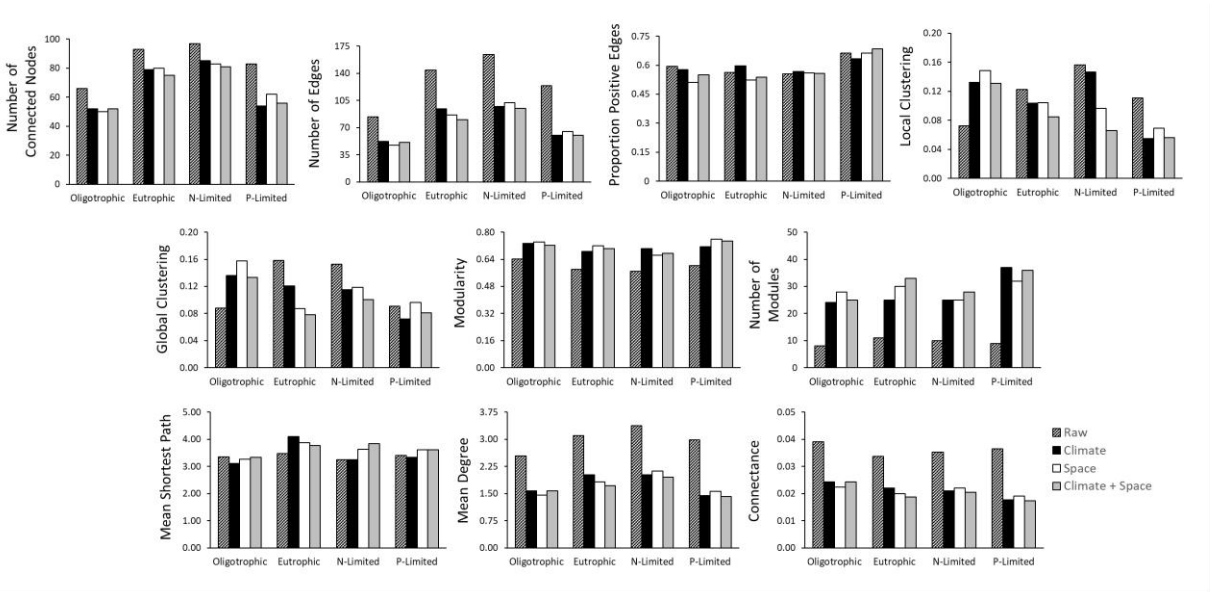

Fig. 3.4. Network parameter values across nutrient contexts in the raw data and after controlling for climate, space and climate + space.

remaining networks. Finally, changes in the fractions of positive edges were strongly dependent on nutrient context and showed no clear general trends with respect to any controlling variable.

Variance partitioning analysis on the metacommunities revealed that climatic and spatial variables generally explained significant fractions of algal relative abundances, with spatial variables predominating regardless of nutrient context (Fig. 3.5). These results did not remarkably change when examining the metacommunity composed of species only found in the networks, with the exception of climatic effects, which became insignificant in the reduced metacommunities. The analysis further revealed that climatic effects (both pure and spatially structured climate) generally explained < 6% of the total variability in composition in contrast to pure spatial factors, which explained usually between 14% to 30% of metacommunity variation.

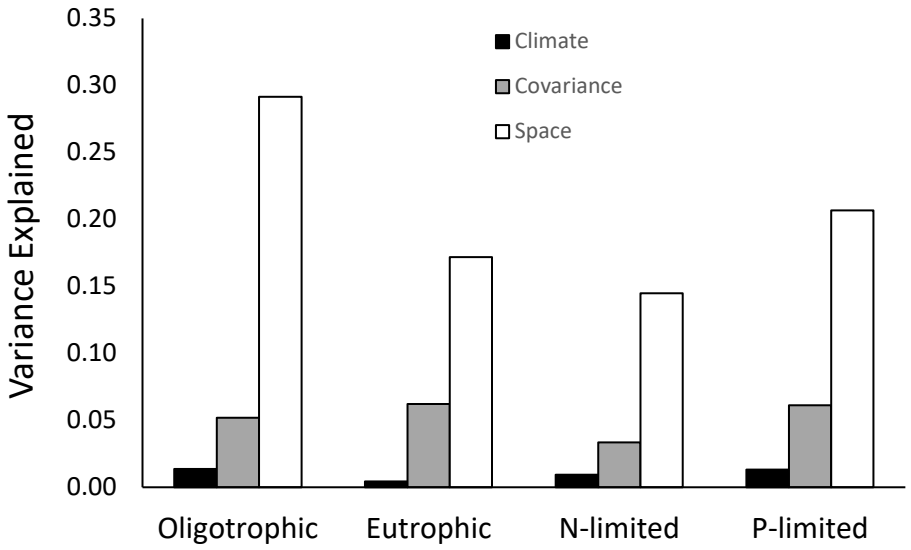

Fig. 3.5. Variance partitioning of algal metacommunity composition into fractions (as adjusted R^2) explained by pure climatic, pure spatial, and covariance effects across nutrient contexts. Pure climate and space fractions within each nutrient context were significant sources of explained variance (each fraction was tested for significance using permutational methods).

Discussion

The dependence of algal composition and biomass production on nutrient supply and balance has been studied for decades and is generally well recognized (Hecky & Kilham, 1988; Bergström, 2010; Hayes *et al.*, 2015). Here we examined for the first time how the topology of algal co-occurrence and interaction networks is structured by nutrient context, consistent with our first objective. Networks assembled under various nutrient supply and ratio scenarios differed in size, i.e. number of nodes and edges in nutrient supply networks but only number of nodes in nutrient ratio networks. Nutrient supply further affected network clustering, complexity and subdivision

into modules, while nutrient ratio underlined the type of interactions and clustering. The remaining topological differences, although significant, were weak. Following objective 2, we found strong support for the prediction that climatic and spatial gradients constrain the topology of the networks and that the magnitude of this influence depends on nutrient context. We further demonstrated that different mechanisms control metacommunity composition vs. network topology.

Nutrient supply context influenced the size, complexity and modular structure of the networks, agreeing with our first hypothesis. Eutrophic networks were dominated by the most speciose guild, the motile guild, while the high profile guild prevailed in the oligotrophic network but the overall guild distribution was more equitable. Higher richness and dominance under eutrophic conditions translated into larger networks that were more clustered and complex (higher mean node degree), encompassing more modules than the smaller oligotrophic networks. In the nutrient imbalanced networks, higher diversity and greater clustering were observed under N-limitation, while P-limitation resulted in proportionately more positive edges. However, these differences cannot be attributed to guild composition, which was similar.

In general, more diverse networks showed proportionately greater clustering consistent with similar observations for mutualistic pollinator networks (Gómez *et al.*, 2011). Clustering in networks reflects the presence of redundant pathways (Karimi *et al.*, 2017) and stronger organization of nodes into distinct subgroups (Girvan and Newman 2002). Much of our knowledge on the environmental dependence of clustering comes from research on soil microbiota, where clustering decreased with CO_2 concentrations (Zhou *et al.*, 2011; Sauvadet *et al.*, 2016) but showed no response to anthropogenic disturbance Zapellini et al. (2015). Our findings add to this knowledge by identifying nutrient supply and ratio as drivers of network clustering patterns in stream ecosystems. Potential avenues for future research include

determining whether network clustering affects ecosystem functions and whether it could be used for monitoring nutrient impacts in streams.

An increase in topological complexity (mean node degree) with node diversity was detected only in the nutrient supply context. This was probably a consequence of the stronger guild dominance in the eutrophic compared to the oligotrophic network, given that nodes from the same guild are likely to co-occur due to similar environmental preference. Topological complexity was decoupled from network diversity in the nutrient ratio context as guild frequencies were comparable. Thus, we showed that the dependency of algal network size and topology presented itself mainly within nutrient supply context, which is consistent with other studies that found that network complexity is dependent on trophic conditions (Dai et al. (2017) and correlates, albeit not always positively, with biodiversity (Shi *et al.*, 2016; Li *et al.*, 2018). Further, our study highlights that differences in guild compositions likely underlay this relationship.

Relationships of network size to certain topological parameters, particularly mean path length and modularity, and proportions of positive links cannot be explained easily by guild compositions and appear to be more complicated. Much work has shown that network topologies are strongly driven by the presence of highly connected nodes (Proulx *et al.*, 2005; Poisot *et al.*, 2015), which were generally absent in our networks. As a result, these networks were poorly connected with connectance of 0.03-0.04, falling below the commonly reported values of 0.05 to 0.3 (Thompson et al. 2012). Thus, a major conclusion here is that, in subcontinental-scale co-occurrence networks, higher degree nodes are likely very rare and their absence explains why our networks display a poor relationship of biodiversity and some aspects of network topology—because species lost or gained between nutrient networks were by and large poorly connected.

All of our networks were dominated by positive relationships, as expected at large scales, due to spatial scaling of mutualisms at the expense of competitive relationships (Araujo and Rozenfeld 2014). Proportions of positive connections are typically examined in ecological networks because of their role in conferring stability (Muogi and Kondoh 2012, Lurgi et al., 2015, Garcia-Callejas et al., 2015). However, with our analyses, we found that increased positive co-occurrences may not necessarily impart stability. Specifically, the P-limited network, which had the greatest proportion of positive edges before and after control, was drastically reduced in topological complexity after the null model analysis compared with any other network. We thus conclude that positive co-occurrences do not necessarily relate to network topological stability, but further work in this area is necessary to determine the generality of this pattern.

With respect to objective 2, accounting for climatic and/or spatial effects substantially altered network topology, consistent with results reported for North American tree communities (Morueta-Holme et al. 2016). It has been argued that co-occurrence relationships contain much extraneous information (e.g., shared ecological responses) that should be controlled for to identify relevant biological interactions (Berlow, 1999; Bairey *et al.*, 2016). Our results support this notion and reveal that the two nutrient supply and the N-limited networks controlled for climate, space, and both generally retained 56%-66% of the co-occurrences in the original networks, while the P-limited network retained only 50% to 53%. These percentages could be attributed in part to biotic interactions. The removal of large fractions of connected nodes and edges in our networks by the null model has important ramifications for co-occurrence network analyses because many network properties are highly sensitive to network size and sparseness of edges (Poisot and Gravel 2014). For our study, this means that variation in subcontinental co-occurrence network properties reflects how nutrient contexts influenced climate preferences and

spatial relationships across the metacommunity. Our findings also highlight an opportunity worth exploring, as null model methods can reveal whether other variables, such as disturbance or competition/predation dynamics, underlie co-occurrence patterns, potentially indicating their importance in driving community organization.

As predicted, climatic effects significantly constrained network topology across nutrient contexts. Inconsistent with our expectation for a more prominent impact of space on eutrophic than oligotrophic networks, we found the opposite—co-occurrences in oligotrophic streams were more strongly affected by space. In fact, all three groups of variables, climatic, spatial and climatic + spatial, contributed more to the oligotrophic than eutrophic networks. A similar, but more striking contrast was observed with the ratio networks, which showed that all three groups of variables produced a much greater change in the P-limited relative to the N-limited network. Collectively, these results suggest that co-occurrence networks under eutrophic and N-limited conditions, with greater diversity and local clustering, were more strongly driven by biotic interactions, offsetting the influence of climate and dispersal. This implies that nutrient context through its effect on species/guild composition and biotic interactions (Carrick *et al.*, 1988; Law *et al.*, 2014), may modulate network response to climate change. We advocate that future investigations experimentally test the importance of biotic interactions under different nutrient contexts and assess to what extent these interactions provide resistance to environmental fluctuation.

Across nutrient contexts, climate and spatial drivers had different effects on species composition vs. network properties. Prior reports as to whether ecological drivers of network topology and community composition coincide have been mixed, with some studies showing correspondence (Mokross et al. 2014), while others, divergence (Li et al., 2018). In this

investigation, network topology was controlled by both climate and space, whereas species composition was primarily explained by spatial factors with climate (i.e., pure and spatially structured) having only a modest contribution. The discrepant influence of climate on network topology vs. species composition could be explained with a scale mismatch in species vs. metacommunity response. Individual species distributions, and by extension, pair-wise species co-occurrences in the networks, are more spatially constrained and thus may be strongly driven by climate, as shown for stream diatoms (Pajunen *et al.*, 2016). Conversely, at the subcontinental scale of our metacommunities, geographic barriers among the discrete hydrologic systems that comprised their habitats, may have contributed to dispersal limitation subsuming the climatic effect. Although environmental factors predict better the variance in diatom metacommunities compared to space (Soininen and Teittinen 2019), these factors are generally associated with nutrient levels (Soininen 2007), which in our case was accounted for by nutrient context. Thus, within nutrient context, spatial predictors outperformed environmental forces at a metacommunity level.

Our study lays groundwork for connecting novel methodology with classic ecological research to generate new information explaining why species co-occur. Nutrient supply and ratios affected subcontinental co-occurrence network properties and the magnitude of climatic and dispersal effects. The differential response of network topology and species composition to climate and dispersal calls for broader multi-level ecological approaches to better understand how the environment structures biological communities.

References

Bairey, E., Kelsic, E.D. & Kishony, R. (2016) High-order species interactions shape ecosystem diversity. *Nature Communications*, **7**, 12285.

Bauman, D., Drouet, T., Dray, S. & Vleminckx, J. (2018) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. *Ecography*, **41**, 1638–1649.

Bergström, A.K. (2010) The use of TN:TP and DIN:TP ratios as indicators for phytoplankton nutrient limitation in oligotrophic lakes affected by N deposition. *Aquatic Sciences*, **72**, 277–281.

Berlow, E.L. (1999) Strong effects of weak interactions in ecological communities. *Nature*, **398**, 330–334.

Bestion, E., Schaum, C.-E. & Yvon-Durocher, G. (2018) Nutrient limitation constrains thermal tolerance in freshwater phytoplankton. *Limnology and Oceanography Letters*, **3**, 436–443.

Borrett, S.R., Whipple, S.J. & Patten, B.C. (2010) Rapid development of indirect effects in ecological networks. *Oikos*, **119**, 1136–1148.

Cardinale, B.J., Hillebrand, H., Harpole, W.S., Gross, K. & Ptacnik, R. (2009) Separating the influence of resource "availability" from resource "imbalance" on productivity-diversity relationships. *Ecology Letters*, **12**, 475–487.

Carrick, H.J., Lowe, R.L. & Rotenberry, J.T. (1988) Guilds of Benthic Algae along Nutrient Gradients: Relationships to Algal Community Diversity. *Journal of the North American Benthological Society*, **7**, 117–128.

Chase, J.M. (2010) Stochastic community assembly causes higher biodiversity in more productive environments. *Science*, **328**, 1388–1391.

Cook, S.C., Housley, L., Back, J.A. & King, R.S. (2018) Freshwater eutrophication drives sharp reductions in temporal beta diversity. *Ecology*, **99**, 47–56.

Csardi, G. & Nepusz, T. (2019) igraph: Network Analysis and Visualization.

Dai, W., Zhang, J., Tu, Q., Deng, Y., Qiu, Q. & Xiong, J. (2017) Bacterioplankton assembly and interspecies interaction indicating increasing coastal eutrophication. *Chemosphere*, **177**, 317–325.

Dodds, W.K. & Smith, V.H. (2016) Nitrogen, phosphorus, and eutrophication in streams. *Inland Waters*, **6**, 155–164.

Dray, A.S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N. & Wagner, H.H. (2016) adespatial: Multivariate Multiscale Spatial Analysis. R package, Version: 0.0-7.

Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002) Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, **5**, 558–567.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B. & Smith, J.E. (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, **10**, 1135–1142.

Estrada, E. (2007) Food webs robustness to biodiversity loss: The roles of connectance, expansibility and degree distribution. *Journal of Theoretical Biology*, **244**, 296–307.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.

Flaim, G., Eccel, E., Zeileis, A., Toller, G., Cerasino, L. & Obertegger, U. (2016) Effects of re-oligotrophication and climate change on lake thermal structure. *Freshwater Biology*, **61**, 1802–1814.

Gómez, J.M., Perfectti, F. & Jordano, P. (2011) The functional consequences of mutualistic network architecture. *PLoS ONE*, **6**, e16143.

Harpole, W.S., Ngai, J.T., Cleland, E.E., Seabloom, E.W., Borer, E.T., Bracken, M.E.S., Elser, J.J., Gruner, D.S., Hillebrand, H., Shurin, J.B. & Smith, J.E. (2011) Nutrient co-limitation of primary producer communities. *Ecology Letters*, **14**, 852–862.

Hayes, N.M., Vanni, M.J., Horgan, M.J. & Renwick, W.H. (2015) Climate and land use interactively affect lake phytoplankton nutrient limitation status. *Ecology*, **96**, 392–402.

Hecky, R.E. & Kilham, P. (1988) Nutrient limitation of phytoplankton in freshwater and marine environments: A review of recent evidence on the effects of enrichment. *Limnology and Oceanography*, **33**, 796–822.

Herve, M. (2019) Testing and Plotting Procedures for Biostatistics "RVAideMemoire"'. *Cran*.

Hillebrand, H., Gruner, D.S., Borer, E.T., Bracken, M.E.S., Cleland, E.E., Elser, J.J., Harpole, W.S., Ngai, J.T., Seabloom, E.W., Shurin, J.B. & Smith, J.E. (2007) Consumer versus resource control of producer diversity depends on ecosystem type and producer community structure. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10904–10909.

Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K.M., Andersen, H.E., Lauridsen, T.L., Liboriussen, L., Beklioglu, M., Özen, A. & Olesen, J.E. (2009) Climate Change Effects on Runoff, Catchment Phosphorus Loading and Lake Ecological State, and Potential Adaptations. *Journal of Environmental Quality*, **38**, 1930–1941.

Jyrkänkallio-Mikkola, J., Meier, S., Heino, J., Laamanen, T., Pajunen, V., Tolonen, K.T., Tolkkinen, M. & Soininen, J. (2017) Disentangling multi-scale environmental effects on stream microbial communities. *Journal of Biogeography*, **44**, 1512–1523.

Karimi, B., Maron, P.A., Chemidlin-Prevost Boure, N., Bernard, N., Gilbert, D. & Ranjard, L. (2017) Microbial diversity and ecological networks as indicators of environmental quality. *Environmental Chemistry Letters*, **15**, 265–281.

Larsen, M.L., Wilhelm, S.W. & Lennon, J.T. (2019) Nutrient stoichiometry shapes microbial coevolution. *Ecology Letters*, **22**, 1009–1018.

Law, R.J., Elliott, J.A., Jones, I.D. & Page, T. (2014) The influence of different environmental conditions upon the initial development and ecological dynamics of phytobenthic communities. *Fundamental and Applied Limnology*, **185**, 139–153.

Leboucher, T., Budnick, W.R., Passy, S.I., Boutry, S., Jamoneau, A., Soininen, J., Vyverman, W. & Tison-Rosebery, J. (2019) Diatom β-diversity in streams increases with spatial scale and decreases with nutrient enrichment across regional to sub-continental scales. *Journal of Biogeography*, **46**, 734–744.

Lewandowska, A.M., Biermann, A., Borer, E.T., Cebrián-Piqueras, M.A., Declerck, S.A.J., De Meester, L., Van Donk, E., Gamfeldt, L., Gruner, D.S., Hagenah, N., Harpole, W.S., Kirkman, K.P., Klausmeier, C.A., Kleyer, M., Knops, J.M.H., Lemmens, P., Lind, E.M., Litchman, E., Mantilla-Contreras, J., Martens, K., Meier, S., Minden, V., Moore, J.L., Venterink, H.O., Seabloom, E.W., Sommer, U., Striebel, M., Trenkamp, A., Trinogga, J., Urabe, J., Vyverman, W., Van de Waal, D.B., Widdicombe, C.E. & Hillebrand, H. (2016) The influence of balanced and imbalanced resource supply on biodiversity-functioning relationship across ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**, 20150283.

Li, D., Poisot, T., Waller, D.M. & Baiser, B. (2018) Homogenization of species composition and species association networks are decoupled. *Global Ecology and Biogeography*, **27**, 1481–1491.

Marañón, E., Lorenzo, M.P., Cermeño, P. & Mouriño-Carballido, B. (2018) Nutrient limitation suppresses the temperature dependence of phytoplankton metabolic rates. *ISME Journal*, **12**, 1836–1845.

Marcel, R., Berthon, V., Castets, V., Rimet, F., Thiers, A., Labat, F. & Fontan, B. (2017) Modelling diatom life forms and ecological guilds for river biomonitoring. *Knowledge and Management of Aquatic Ecosystems*, **2017-Janua**, 1.

Menzel, U. (2016) RMThreshold: Signal-Noise Separation in Random Matrices by using Eigenvalue Spectrum Analysis.

Mokross, K., Ryder, T.B., Côrtes, M.C., Wolfe, J.D. & Stouffer, P.C. (2013) Decay of interspecific avian flock networks along a disturbance gradient in Amazonia. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132599.

Montoya, D., Yallop, M.L. & Memmott, J. (2015) Functional group diversity increases with modularity in complex food webs. *Nature Communications*, **6**, 1–9.

Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B.J., Peet, R.K., Ott, J.E., Violle, C., Enquist, B.J., Jørgensen, P.M. & Svenning, J.C. (2016) A network approach for inferring species associations from co-occurrence data. *Ecography*, **39**, 1139–1150.

Mougi, A. & Kondoh, M. (2012) Diversity of interaction types and ecological community stability. *Science*, **337**, 349–351.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2019) vegan: Community Ecology Package.

Özen, A., Karapinar, B., Kucuk, I., Jeppesen, E. & Beklioglu, M. (2010) Drought-induced changes in nutrient concentrations and retention in two shallow Mediterranean lakes subjected to different degrees of management. *Hydrobiologia*, **646**, 61–72.

Pajunen, V., Luoto, M. & Soininen, J. (2016) Climate is an important driver for stream diatom distributions. *Global Ecology and Biogeography*, **25**, 198–206.

Passy, S.I. (2007) Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany*, **86**, 171–178.

Passy, S.I. & Larson, C.A. (2019) Niche dimensionality and herbivory control stream algal biomass via shifts in guild composition, richness, and evenness. *Ecology*, **100**, e02831.

Passy, S.I. & Larson, C.A. (2011) Succession in Stream Biofilms is an Environmentally Driven Gradient of Stress Tolerance. *Microbial Ecology*, **62**, 414.

Peres-Neto, P.R., Jackson, D.A. & Somers, K.M. (2005) How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, **49**, 974–997.

Poisot, T., Stouffer, D.B. & Gravel, D. (2015) Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, **124**, 243–251.

Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005) Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, **20**, 345–353.

Qu, X., Peng, W., Liu, Y., Zhang, M., Ren, Z., Wu, N. & Liu, X. (2019) Networks and ordination analyses reveal the stream community structures of fish, macroinvertebrate and benthic algae, and their responses to nutrient enrichment. *Ecological Indicators*, **101**, 501–511.

Redfield, A.C. (1934) On the Proportions of Organic Derivatives in Sea Water and Their Relation to the Composition of Plankton. *University Press of Liverpool, James Johnstone Memorial Volume*, 1767–192.

Rocha, J.C., Peterson, G.D. & Biggs, R. (2015) Regime shifts in the anthropocene: Drivers, risks, and resilience. *PLoS ONE*, **10**, e0134639.

Sauvadet, M., Chauvat, M., Cluzeau, D., Maron, P.A., Villenave, C. & Bertrand, I. (2016) The dynamics of soil micro-food web structure and functions vary according to litter quality. *Soil Biology and Biochemistry*, **95**, 262–274.

Shi, S., Nuccio, E.E., Shi, Z.J., He, Z., Zhou, J. & Firestone, M.K. (2016) The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecology Letters*, **19**, 926–936.

Shirley, M.D.F. & Rushton, S.P. (2005) The impacts of network topology on disease spread. *Ecological Complexity*, **2**, 287–299.

Soininen, J. (2007) Environmental and spatial control of freshwater diatoms—a review. *Diatom Research*, **22**, 473–490.

Soininen, J., Jamoneau, A., Rosebery, J. & Passy, S.I. (2016) Global patterns and drivers of

species and trait composition in diatoms. *Global Ecology and Biogeography*, **25**, 940–950.

Soininen, J. & Teittinen, A. (2019) Fifteen important questions in the spatial ecology of diatoms. *Freshwater Biology*, **64**, 2071–2083.

Steiner, C.F. & Leibold, M.A. (2004) Cyclic assembly trajectories and scale-dependent productivity-diversity relationships. *Ecology*, **85**, 107–113.

Stoddard, J.L., Van Sickle, J., Herlihy, A.T., Brahney, J., Paulsen, S., Peck, D. V., Mitchell, R. & Pollard, A.I. (2016) Continental-scale increase in lake and stream phosphorus: are oligotrophic systems disappearing in the United States? *Environmental Science and Technology*, **50**, 3409–3415.

Suweis, S., Grilli, J. & Maritan, A. (2014) Disentangling the effect of hybrid interactions and of the constant effort hypothesis on ecological community stability. *Oikos*, **123**, 525–532.

Tilman, D. & Isbell, F. (2015) Biodiversity: recovery as nitrogen declines. *Nature*, **528**, 336.

Verbeek, L., Gall, A., Hillebrand, H. & Striebel, M. (2018) Warming and oligotrophication cause shifts in freshwater phytoplankton communities. *Global Change Biology*, **24**, 4532–4543.

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

Wang, J., Pan, F., Soininen, J., Heino, J. & Shen, J. (2016) Nutrient enrichment modifies temperature-biodiversity relationships in large-scale field experiments. *Nature Communications*, **7**, 13960.

Wang, R., Dearing, J.A., Doncaster, C.P., Yang, X., Zhang, E., Langdon, P.G., Yang, H., Dong, X., Hu, Z. & Xu, M. (2019) Network parameters quantify loss of assemblage structure in human-impacted lake ecosystems. *Global change biology*, **25**, 3871–3882.

Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L.E., Cross, W.F., Friberg, N., Ings, T.C. & Jacob, U.T.E. (2010) *Ecological networks in a changing climate*. *Advances in ecological research*, pp. 71–138. Elsevier.

Wu, N., Dong, X., Liu, Y., Wang, C., Baattrup-Pedersen, A. & Riis, T. (2017) Using river microalgae as indicators for freshwater biomonitoring: Review of published research and future directions. *Ecological Indicators*, **81**, 124–131.

Zappelini, C., Karimi, B., Foulon, J., Lacercat-Didier, L., Maillard, F., Valot, B., Blaudez, D., Cazaux, D., Gilbert, D. & Yergeau, E. (2015) Diversity and complexity of microbial communities from a chlor-alkali tailings dump. *Soil Biology and Biochemistry*, **90**, 101–110.

Zhou, J., Deng, Y., Luo, F., He, Z. & Yang, Y. (2011) Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. *MBio*, **2**, e00122-11.

Appendix 3.1: Background for constructing networks with RMT and integrating regression trees to select an objective correlation threshold.

Networks displaying species co-occurrence relationships as pair-wise correlations use an adjacency matrix format known as a weighted adjacency matrix. Weighted adjacency matrices provide the same information as binary adjacency matrices in that a non-zero value indicates the existence of an edge between two nodes, but the edges in these matrices are also numerically valued and may or may not be signed (positive or negative). However, the problem associated with the use of correlation matrices as a weighted network adjacency matrix is that, unless the adjacency matrix is appropriately thresholded, all nodes will have connections with all other nodes, resulting in a completely connected graph that is unsuitable for network analysis (i.e., topological comparison). Threshold selection in correlation-based weighted adjacency matrices is a crucial step because threshold selection has major consequences on nearly every topological parameter of interest (Perkins and Langston 2009). Therefore, selection of thresholds, regardless of empirical or theoretical justification, introduces bias into network analysis that can impact the interpretation of relationships represented in the network.

Random matrix theory (RMT) has been integrated into the study of networks because it may allow for objective selection of a threshold value based solely on statistical patterns within the adjacency matrix. Particularly, the RMT approach is attractive as it optimally reveals clustering and modular structure in the network adjacency matrix, while removing values most likely originating from random Gaussian sampling. We do not discuss the mathematical machinery of the method here and instead point the reader to other resources for more in-depth proofs of the method (Bandyopadhyay and Jalan 2007, Rai and Jalan 2015). Instead, we

summarize the major points here and describe our modification of the commonly used RMT approach.

RMT-based matrix threshold selection is an iterative procedure with the goal of identifying a threshold value given a set of candidate thresholds. For each candidate value, usually differing from each other in intervals of 0.01 unit (although other intervals can be used), the network adjacency matrix is zeroed-out below the candidate value, the adjacency matrix eigenvalues are calculated, and the pairwise differences between the eigenvalues taken. Each candidate threshold value thus generates a unique frequency distribution of pairwise differences, which are termed eigenvalue spacings in RMT parlance. These spacings are then fit with two statistical distributions, the Wigner-Surmise and the exponential distribution. Generally, eigenvalue spacings are usually fit best by the Wigner-Surmise, a unimodal Gaussian-like distribution, at low candidate threshold values and exponential distribution at high candidate threshold values. The Wigner-Surmise distribution encompasses correlated eigenvalue spacings, i.e. the frequency distribution of spacings is clustered around the mean. The exponential distribution, comprises uncorrelated eigenvalue spacings with an exponential decay in frequency.

With RMT-thresholding, the optimal candidate threshold value is the one that corresponds to the transition point at which the eigenvalue spacings are fit better by the exponential distribution over the Wigner-Surmise distribution. This transition point has been assessed using chi-squared methods (Deng et al. 2012). However, in this paper, the optimal threshold value indicating the transition point is not automatically chosen in the used R-package, but instead requires manual selection based on the divergence in behavior of fitted model loglikelihoods. Therefore, we incorporated a regression tree procedure to objectively select the threshold value. In this procedure, the response variable was the difference between the log-

likelihoods of a Wigner-Surmise fit vs. exponential fit of the distribution of eigenvalue spacings and the predictor values were the potential candidate correlation thresholds. We then used the predictor value at the first tree split as the correlation threshold at which the eigenvalue spacings become uncorrelated from each other. Visual observation of the plots of fitted model log-likelihoods against candidate thresholds supported this approach.

Chapter 4
The scale of stream co-occurrence networks depends on spatial scale

William R. Budnick[1], Joseph L. Mruzek[1], Chad A. Larson[2], T. Leboucher[3], and S. I. Passy[1*]

[1]Department of Biology, University of Texas at Arlington, 501 S. Nedderman Dr., Arlington Texas, 76019, USA

[2]Washington State Department of Ecology, Environmental Assessment Program, 300 Desmond Dr SE, Lacey, Washington, 98503, USA

[3]Aquatic Ecosystems and Global Changes Research Unit, INRAE, 50 avenue de Verdun, 33612 Cestas, France.

Author Emails and ORCID IDs:
William R. Budnick: William.budnick@mavs.uta.edu; https://orcid.org/0000-0001-9288-6782
Joseph L. Mruzek: Joseph.mruzek@mavs.uta.edu; https://orcid.org/0000-0002-8067-3824
Chad A. Larson: clar461@ecy.wa.gov; https://orcid.org/0000-0002-0329-8979
Thibault Leboucher: Thibault.leboucher@inrae.fr; https://orcid.org/0000-0003-0353-8896
Sophia I. Passy: Sophia.passy@uta.edu; https://orcid.org/0000-0002-8230-9380

*Corresponding Author

Introduction

How and why species occur together continue to challenge community ecologists. In response, a number of methods ranging from simultaneous differential equations (Gause, 1932) to elaborate computer algorithms designed to shuffle community tables (Gotelli, 2000) have been developed to study this intriguing problem. However, any pairwise methods ignore the complex interwoven nature of a community (Bascompte, 2009), where co-occurrence is controlled by direct and indirect relationships among species (Barberán *et al.*, 2012). The relatively recent application of network theory has moved community ecology forward by allowing description of the origin, structure, and consequences of these complex biotic networks. Specifically, co-occurrence network analysis treats species as nodes, draws co-occurrence relationships between species as edges, and applies graph theory to analyze the complexity of the resulting web. Co-occurrence network analysis has shown considerable promise in unravelling the potential ecological underpinnings of species relationships, including dispersal limitation, competition, and environmental constraint. Nevertheless, much of our current knowledge is reliant on networks constructed at relatively small spatial scales. As global change is not inherently local, studies examining how co-occurrence networks vary across spatial scales (hereafter referred to in the sense of spatial "extents"), and whether scale-dependency is in turn taxonomically dependent, are sorely needed.

The formation and topological properties of a network are strongly correlated with the node degree distribution (NDD), a frequency distribution that quantifies the number of nodes with *k* connections to other nodes (Proulx *et al.*, 2005; Newman, 2007; Fortuna *et al.*, 2010). Real-world networks generally belong to a class called "small-world", characterized by short path lengths and non-random clustering patterns (Watts & Strogatz, 1998). The NDD of

networks with small-world structure generally fall into one of three classes: exponentially-bounded, power law, and cross-over between the two (Amaral *et al.*, 2000; Lewis *et al.*, 2010; Barabási, 2016; Holme, 2019) (Fig. 4.1). Networks with exponentially-bounded NDDs (e.g., exponential and Poisson distributions), commonly termed "single-scale," are homogeneous and thin-tailed. Plainly, this means that most nodes have a similar number of edges and the NDD rapidly and asymptotically approaches 0 at large *k* values. The single-scale terminology does not refer to a spatial aspect of the NDD, but instead to how the standard deviation (the scale parameter) of the NDD is set by the mean (Albert *et al.*, 2000). Power law NDDs, on the other hand, are inhomogeneous and fat-tailed — the majority of nodes have only a few edges, but some nodes have very high degrees. Power laws are termed scale-free distributions when the absolute value of the power law exponent, g, is between 2 and 3, because only the mean node degree is defined, while the standard deviation tends to infinity (i.e., there is no "defined" scale). Notably, power law models can generally fit only part of the distribution, therefore, it is common to test for prevalence of power laws after discarding the lower tail of the distribution (Gillespie, 2015; Broido & Clauset, 2019). Cross-over NDDs, also termed "broad scale," including exponentially-truncated power law (referred to as truncated power law henceforth), log-normal, and stretched exponential (Weibull) distributions, fall between exponentially-bounded and power law NDDs and possess properties of both (Amaral *et al.*, 2000). All three classes of NDD have been reported for ecological networks (Dunne *et al.*, 2002; Guimarães *et al.*, 2007; Dormann *et al.*, 2009) but recent evidence suggests that the power law NDD may be rare (Broido & Clauset, 2019).

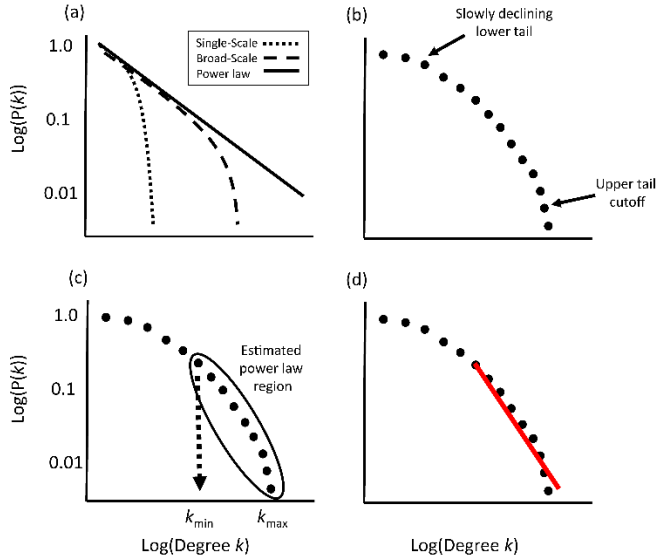

Fig. 4.1. Conceptual diagram showing node degree distributions (NDDs), drawn as cumulative frequency, $P(k)$, against node degree, k, on log scales. (a) An ideal power law fit over the entire NDD conforms to a linear pattern, while single-scale fits decline more rapidly. Broad-scale fits present with a power-law like region before strongly showing exponential declines. (b) More often, real networks tend to be characterized by a slowly declining lower tail and an exponentially-truncated upper tail, causing the power law model to generate poor fits. (c) To account for these features of real networks and examine the power law-like nature of the distribution, procedures commonly use maximum likelihood estimation and Komolgorov-Smirnov methods to examine a region of the NDD most closely conforming to a power law (oval). This region is identified by estimating a minimum degree value, k_{min}, below which the lower tail of the distribution is ignored, and the remainder of the distribution most closely corresponds to the power law. model up to the maximum degree value, k_{max}. (d) The power law model is then fit over the distribution for $k \geq k_{min}$ and assessed for power law properties (e.g., parameters).

The shape of the NDD has been a subject of great interest across numerous fields of science and technology, including biology, epidemiology, sociology, transportation, and communications (Clauset *et al.*, 2009) because it determines susceptibility to targeted node removals and information transmission rates within networks of molecules, individuals, species, power supplies, etc. (Albert *et al.*, 2000; Dunne *et al.*, 2002; Estrada, 2007; Gillespie, 2015). In ecology, understanding what models generally best describe the NDD patterns and whether ecological factors influence the model fits can provide crucial information on species network structure and function. Most ecological network analyses have focused primarily on terrestrial taxa, with less attention given towards aquatic networks, although recent efforts have begun to remedy this deficit (Thompson *et al.*, 2012; Echevarria & Rodriguez, 2017). Given that aquatic communities are disproportionately susceptible to anthropogenic and global change factors (Sala *et al.*, 2000; Dudgeon *et al.*, 2006), it is especially important to derive knowledge for this class of biota while simultaneously advancing our general understanding of co-occurrence network origins and patterns.

In this study, we examined what models provide optimal fits for NDD in major aquatic organisms and whether the frequencies of these models are organism- and spatial extent-dependent. We used metacommunity datasets of stream diatoms and fish and examined the topologies of their co-occurrence networks from subregional to subcontinental extents with a recently developed "windows" methodology (Leboucher *et al.*, 2019). To our knowledge, this is the first investigation to explore empirical NDDs at several spatially-explicit extents and provide an up to date, and strongly needed, critical analysis of statistical distribution fits to NDDs. Additionally, we tested the generality of these patterns between two taxonomic groups that differ greatly in body size (small vs. large) and dispersal capacity (passive vs. active).

Here we test two hypotheses. We first hypothesized that the NDD fit depends on spatial extent. Important ecological phenomena, such as species-area relationship (Palmer & White, 1994), abundance distributions (McGill *et al.*, 2007), and biotic interactions (Araújo & Rozenfeld, 2014), are not conserved across spatial extents. Previous studies have reported that networks of increasing spatial extents tend to exhibit properties of larger size (Lafferty & Dunne, 2010; Galiana *et al.*, 2018) and lower connectance (Sugiura, 2010). The explanations ranged from increased sampling of rare species to biological and ecological processes and mechanisms preventing links between species (i.e., forbidden links, Jordano *et al.*, 2003), rendering these networks sparsely connected. Network size and connectance are among the most important determinants of the NDD (Poisot & Gravel, 2014), which we predict constrain the types and variety of models observed as spatial extent increases. Consequently, we expected NDDs are generally more strongly skewed toward rare species at larger spatial extents, which would thus exhibit greater prevalence of single-scaled NDDs.

We second hypothesized that body size through its effect on dispersal capacity will constrain the prevalence of particular best model fits across spatial extents. How body size and dispersal differences between taxonomic groups contribute to network variability is not well studied. Given the impact of these organismal properties on metacommunity composition (De Bie *et al.*, 2012), they may underlie co-occurrence patterns, and consequently, a variation in NDD. Stronger dispersal limitation in large organisms, such as fish (Shurin *et al.*, 2009), may result in prevalence of exponentially-bounded networks with very few well connected nodes across scales, whereas in small and better dispersing diatoms, this tendency may be observed mostly at larger spatial scales. Therefore, to address these two hypotheses, we pursued four objectives: 1) assess the dependency of network size and connectance on spatial scale, 2)

determine if prevalence of certain NDD classes (i.e., single-scale, broad-scale, and power law) varies with spatial extent, 3) establish if network size, connectance, and spatial extent predict the classes of NDD, and 4) evaluate if the above relationships differ between taxonomic groups.

Materials and Methods

Data

Periphytic diatom and fish data were sourced from the US Geological Survey's National Water-Quality Assessment (NAWQA) Program and the US Environmental Protection Agency's National Rivers and Streams Assessment (NRSA). Periphyton and fish were collected from 3314 and 2964 stream sites, respectively, sampled across the conterminous United States from 1993-2011 (NAWQA) and 2012-2013 (NRSA, Appendix 4.1, Fig. S4.1). Because community data from both government datasets were sampled with similar methods, we compiled them together for our analysis. Samples targeted at least 600 individuals for the diatoms. Individual counts in fish samples were highly variable. Therefore, we rarified sample abundances to 300 individuals and used rarified abundances for each species in the fish dataset. Diatom samples were collected largely from hard substrates or macrophytes. Fish samples were collected from wadeable streams using backpack electrofishing and seining of stream reaches ranging in length from 100 to 550 m.

Landscape window construction

We used a method outlined by Leboucher et al. (2019) to create "landscape windows" of fixed spatial extents (Appendix 4.1, Fig. S4.2). Briefly, the purpose of the method is to place square windows of specific spatial extent at fixed spatial distances,and examine ecological patterns

among sites that lie within that window. We created square landscape windows of sizes 400, 800, 1200, and 1600 km, which respectively corresponded to spatial extents of 160,000 km^2, 640,000 km^2, 1,440,000 km^2, and 2,560,000 km^2. We then specified that valid windows at each size must contain samples that occupy at least 77% of the total area of the window, which was determined by partitioning the window into a 3 × 3 grid and verifying that samples occurred in at least 7/9 of the cells within the window. We also specified that the origins of the windows must be placed at 1/3 the window size in distance from each other, so that no window contained the exact same distribution of sites as another window. Finally, we specified that the minimum number of samples for a valid window should increase with the window size to ensure that sample densities were similar across window sizes.

Network construction

For each landscape window, we removed species that occurred in fewer than 10 sites in the window and then calculated a Spearman correlation matrix on standardized relative abundances. We then used random matrix theory methods (Menzel, 2016) to identify a correlation threshold value, below which correlations are likely spurious (see Appendix 3.1). Those correlations were eliminated from the correlation matrix (i.e., replaced with a 0). Species that lost all correlations with this method were subsequently removed from the correlation matrix. We then used R package "igraph" (Csardi & Nepusz, 2019) to transform the thresholded correlation matrix into a weighted adjacency matrix format from which we constructed the co-occurrence network. For each network, we determined the network size (number of nodes and number of edges), connectance ($\frac{2*\#\ edges\ observed}{(\#nodes^2-\#nodes)}$), and the node degree distribution, i.e. the frequency of each node degree.

Model Fits and Comparisons

For model fitting, we extracted the degree sequence for each network and fit it to 6 statistical distributions describing the shape of the NDD (Alstott *et al.*, 2014; Broido & Clauset, 2019) (Table 4.1). We performed two sets of model fits on each window, whereby we fit all 6 models on 1) the entire degree distribution and 2) the region of the NDD most likely to conform to a power law (Fig 1). For the former, we specified that the model be over the range of the NDD from degree of 1 to the maximum network degree (k_{max}). For the latter, we used functions from R package 'poweRlaw' (Gillespie, 2015) which identify using Komolgorov-Smirnov and maximum likelihood methods the minimum degree value, k_{min}, below which the distribution is discarded, and above which the distribution most likely corresponds to a power law. We also examined the power law exponent, γ, for all power law and truncated power law models fit over the whole NDD and the estimated power law region, which informed us if scale-free patterns were present (scale-free if $2 \leq \gamma \leq 3$).

In both model fitting schemes (whole NDD and pawer law region of the NDD), we calculated AIC values to assess quality of fit. We ranked all model fits according to the AIC values with lower AIC indicating better quality of fit, and recorded which model produced the lowest AIC. For each pair of models (15 total contrasts), we calculated their ΔAIC to determine if they differed from each other significantly in quality of fit. Models that had $\Delta AIC \geq 2$ were considered to be significantly poorer in fit relative to the best model, and $\Delta AIC < 2$ were considered similar in quality of fit (Broido & Clauset, 2019).

Table 4.1. Probability density and mass functions commonly fit to describe node degree distributions. Across models, P(*k*) is probability for a node degree, *k*, λ = rate parameter (exponential, exponentially-truncated power law, and Weibull distributions), μ = mean (Poisson and lognormal distributions), β = scale parameter (Weibull distribution only), and γ = exponent of the power law function (power law and exponentially-truncated power law distribution).

Class	Model	Probability functions
Exponentially-bounded thin-tailed (single scale)	Exponential	$P(k) = \lambda e^{-\lambda k}$
	Poisson	$P(k) = \frac{e^{-\mu}\mu^{k}}{k!}$
Cross-over (broad scale)	Exponentially-truncated power law	$P(k) = ak^{-\gamma}e^{-\lambda k}$
	Lognormal	$P(k) = \frac{1}{\sqrt{2\pi}\sigma k} e^{\frac{(\ln k - \mu)^2}{2\sigma^2}}$
	Weibull	$P(k) = ak^{\beta-1}e^{-\lambda k^{\beta}}$
Fat-tailed (scale-free if γ = 2-3)	Power law	$P(k) = ak^{-\gamma}$

Statistical analyses

For each taxonomic group, we used Welch's ANOVA tests to assess if network size (number of nodes and edges) and connectance varied across window sizes (i.e., spatial extents). Post-hoc Games-Howell pairwise tests (Dag *et al.*, 2018) identified spatial extents where differences arose.

For each window's NDD and its power law region, we determined which of the 6 tested models was the best fitting by raw AIC value, i.e. lowest is best, and by ΔAIC criterion, i.e. the difference in AIC between the best fit and all alternative models. Using ΔAIC, we determined for

each spatial scale, given the raw AIC, how many models were considered equally plausible (i.e., ΔAIC < 2). We used Kruskall-Wallis tests coupled with post-hoc Mann-Whitney pairwise tests with Bonferroni P-value correction to examine if the number of plausibly fitting models was spatially dependent. We then recorded the scale classes (single-scale, broad-scale, power law, or some combination) of these models observed among the plausible models and tabulated their frequencies (% of windows). We assessed if the classes of models fit over the NDD and power law regions generally depended on spatial extent using Fisher's exact tests. We followed these exact tests with logistic and multinomial logistic regressions to test whether spatial extent, network size, and network connectance each predicted the classes of models observed among the best fitting models. Stepwise model selection was performed to reduce the predictors in all logistic models. All analyses were conducted in R version 3.5.1.

Results

Network properties vary with scale

A total of 359 and 171 windows were generated for the diatoms and fish datasets, respectively (Appendix 4.1, Table S4.1), with 57% to 69% of the windows in both datasets corresponding to the 400 km scale. We generally found in both taxonomic groups that node count, edge count, and connectance depended on window size in agreement with our initial expectations (Fig. 4.2). Node counts significantly increased with window size, whereas connectance strongly decreased, although in the fish dataset, node counts were not significantly different between the 1200 and 1600 km windows. In both diatoms and fish, edge counts were significantly lower at 400 km but did not differ among 800, 1200 and 1600 km windows.

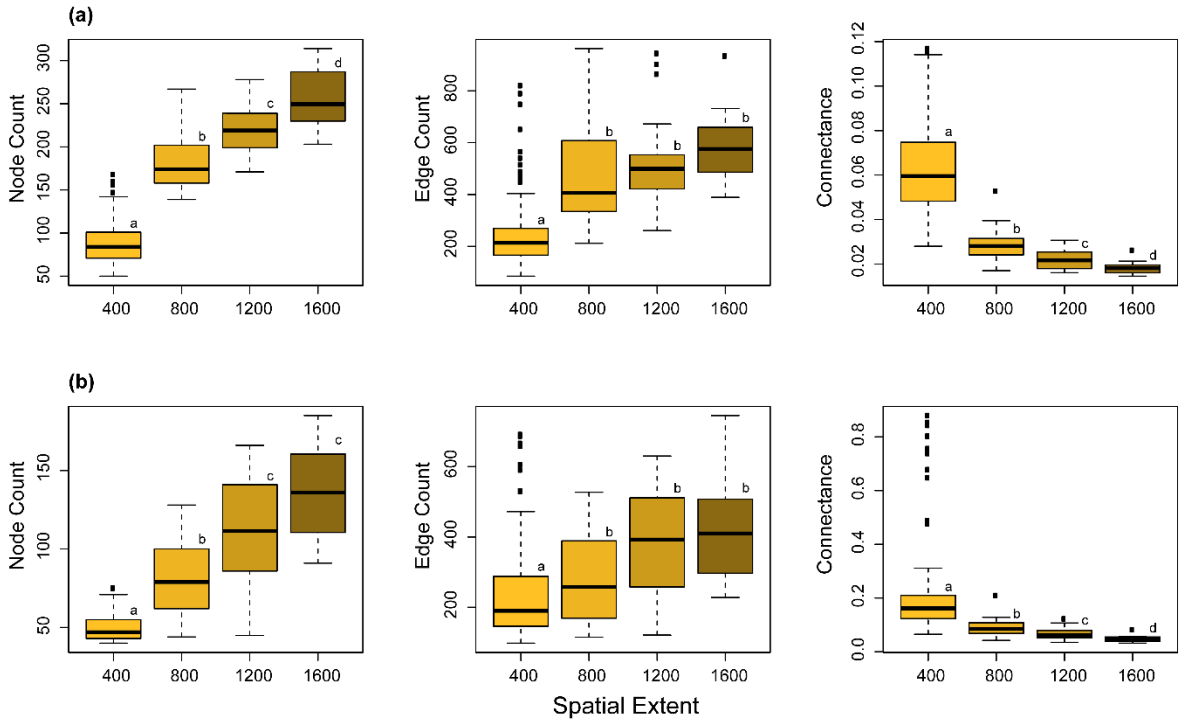

Figure 4.2. Comparisons of network size (node and edge counts) and connectance across spatial extents (km) for (a) diatoms and (b) fish. Different lowercase letters indicate statistical differences ($P < 0.05$) as determined by Welch's ANOVA and post-hoc Games-Howell pairwise comparisons. Points indicate values greater than 1.5 × (Interquartile range).

Model Fits over the whole NDD and power law region of the NDD

When examining raw AIC values, there were several prominent differences with respect to best NDD fits between diatom and fish networks across spatial extents (Table 4.2). Noticeably, diatom network NDDs were best fit by thin-tailed exponential distributions in the smallest

Table 4.2. Frequency table presenting the top model fits of network node degree distributions (NDD) across taxonomic groups and window sizes as ranked by raw AIC. Models were fit over both the entire NDD and the estimated power law region of the NDD. Values indicate number of windows and values in parentheses, % of all windows for a given window size.

		Best fitting model of the entire NDD by AIC (% of windows)					
Taxonomic Group	Window size (km)	Exponential	Poisson	Truncated power law	Lognormal	Weibull	Power law
Diatoms	400	144 (57.4)	1 (0.4)	29 (11.6)	8 (3.2)	69 (27.5)	0
	800	27 (41.5)	0	32 (49.2)	1 (1.5)	5 (7.7)	0
	1200	6 (20.7)	0	21 (72.4)	0	2 (6.9)	0
	1600	1 (7.1)	0	10 (71.4)	0	3 (21.4)	0
Fish	400	26 (26.8)	1 (1.0)	3 (3.1)	1 (1.0)	66 (68.0)	0
	800	28 (68.3)	0	3 (7.3)	2 (4.9)	8 (19.5)	0
	1200	13 (59.1)	0	6 (27.3)	0	3 (13.6)	0
	1600	8 (72.7)	0	3 (27.3)	0	0	0
		Best fitting model of the power law region of the NDD by AIC (% of windows)					
Diatoms	400	189 (75.3)	48 (19.1)	5 (2.0)	2 (0.8)	2 (0.8)	5 (2.0)
	800	51 (78.5)	19 (10.8)	6 (9.2)	0	1 (1.5)	0
	1200	23 (79.3)	3 (10.3)	3 (10.3)	0	0	0
	1600	12 (85.7)	0	2 (14.3)	0	0	0
Fish	400	37 (38.1)	47 (48.5)	0	4 (4.1)	9 (9.3)	0
	800	12 (29.3)	20 (48.8)	0	1 (2.4)	8 (19.5)	0
	1200	13 (59.1)	6 (27.3)	1 (4.5)	0	2 (9.1)	0
	1600	5 (45.5)	2 (18.2)	1 (9.1)	1 (9.1)	2 (18.2)	0

windows (57.4% of windows). This fit became rarer at larger spatial extents, where truncated power law fits increased (from 11.6% of 400 km windows to 71.4% of 1600 km windows). Fish NDDs showed different trends. The cross-over Weibull distribution was the most frequent best fit among the smallest windows (68% of 400km windows), becoming rarer in favor of exponential and truncated power law distributions, measuring 72.7% and 27.3% of the fits, respectively, in the larger windows. For both taxonomic groups, the pure power law distribution never ranked as the best fitting model over the entire NDD, and γ-values never exceeded 2 to indicate scale-fee

behavior of the NDD (Appendix 4.1, Fig. S4.3). Log-normal and Poisson best fits of the NDDs were very few and only observed among smaller windows.

We then estimated the region of the NDD most likely corresponding to a power law. Interestingly, we found that best fitting power law fits of this region were very rare, i.e. only 5 instances and only in diatoms among the smallest windows (Table 4.2). Further, of all power laws models fit over this region, 71 diatom windows (19.8%) and 18 fish windows (5%) had estimated γ-values within the scale-free region and none were found among the 5 instances where the power law was the best fitting model (smallest γ observed was 3.78). Instead, in diatoms, the exponential distribution characterized most frequently the power law regions across all window sizes, followed by the Poisson distribution (at all but the 1600 km windows) and the truncated power law (at all but the 400 km windows). In fish, Poisson fits were more prevalent for 400 and 800 km windows and exponential fits, for 1200 and 1600 km windows. Best fits of the power law region by the Weibull distribution were observed across all window sizes in fish but not in diatoms.

AIC comparison of models describing the whole NDD and the power law region of the NDD

A comparison of the NDD models using ΔAIC revealed that a single best fitting model was only common for the entire NDD, where such model prevailed at the largest window sizes for diatoms (1600 km) but the smallest (400 km) for fish (Table 4.3). Conversely, at smaller window sizes in diatoms and larger window sizes in fish, three best fitting models were most common. Kruskall-Wallis tests indicated that window size affected the number of best fitting models over the whole NDD of both taxonomic datasets, but not for their power law regions (Fig. 4.3). Post-hoc Mann-Whitney pairwise comparisons revealed that the number of plausible models generally declined

Table 4.3. Frequency table (% of windows) of number of plausible alternative best model fits of the entire NDD and the power law region of the NDD. The number of plausible alternative best model fits corresponded to the number of models with DAIC < 2. Values indicate number of windows and values in parentheses, % of all windows for a given window size.

		Number of best fitting models of the whole NDD (% of windows)					
Taxonomic Group	Window Scale (km)	1 Best Model	2 Best Models	3 Best Models	4 Best Models	5 Best Models	6 Best Models
Diatoms	400	17 (6.8)	42 (16.7)	175 (69.7)	17 (6.8)	0	0
	800	13 (20.0)	15 (23.1)	33 (50.8)	4 (6.2)	0	0
	1200	10 (34.5)	9 (31.0)	10 (34.5)	0	0	0
	1600	7 (50.0)	5 (35.7)	2 (14.3)	0	0	0
Fish	400	38 (39.2)	20 (20.6)	37 (38.1)	2 (2.1)	0	0
	800	8 (19.5)	0	29 (70.7)	4 (9.8)	0	0
	1200	4 (18.2)	2 (9.1)	16 (72.7)	0	0	0
	1600	0	0	11 (100)	0	0	0
		Number of best fitting models of the power law region of the NDD (% of windows)					
Diatoms	400	1 (0.4)	3 (1.2)	103 (41.0)	96 (38.2)	40 (15.9)	8 (3.2)
	800	0	3 (4.6)	34 (52.3)	17 (26.2)	11 (16.9)	0
	1200	1 (3.4)	0	14 (48.3)	10 (34.5)	4 (13.8)	0
	1600	0	1 (7.1)	7 (50.0)	5 (35.7)	1 (7.1)	0
Fish	400	4 (4.1)	6 (6.2)	34 (35.1)	36 (37.1)	13 (13.4)	4 (4.1)
	800	3 (7.3)	3 (7.3)	17 (41.5)	16 (39.0)	1 (2.4)	1 (2.4)
	1200	0	1 (4.5)	10 (45.5)	8 (36.4)	3 (13.6)	0
	1600	1 (9.1)	0	6 (54.5)	3 (27.3)	1 (9.1)	0

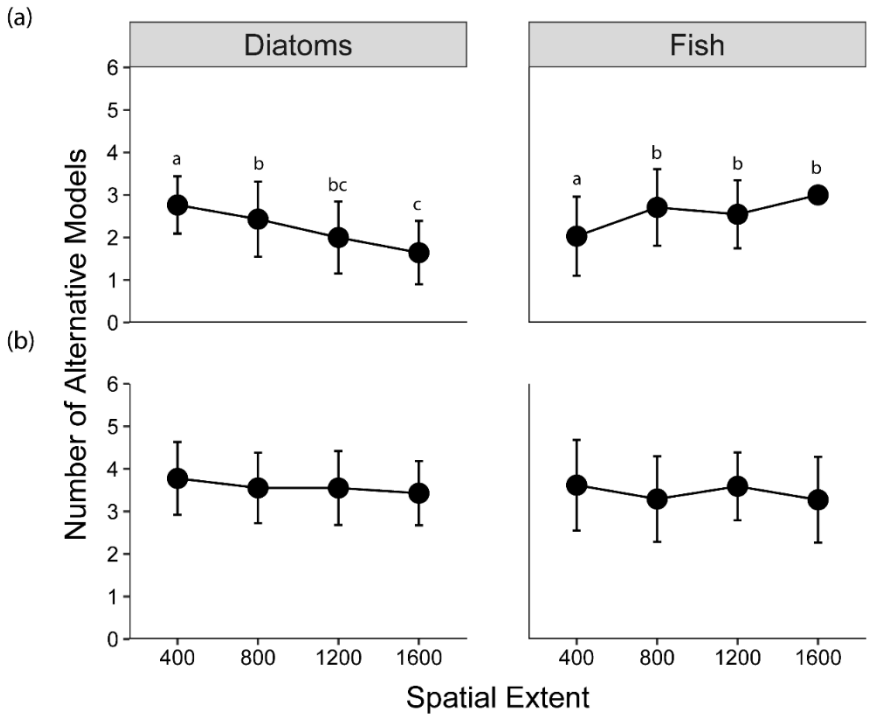

Fig. 4.3. Results of Kruskall-Wallis tests examining mean number of plausible alternative models of (a) whole NDD and (b) power law region across spatial extents. Letters, if present, indicate significant differences between means as indicated by post-hoc Mann-Whitney tests. Error bars denote standard deviation.

with window size for diatoms, but increased in fish, albeit up to 800 km. When observed, the single best fitting models usually were broad-scale, i.e. either truncated power law or Weibull (Appendix 4.1. Fig. S4.5). The whole NDD tended to be best explained by up to 4 alternative models (most often 3 models), whereas the power law region was best explained by up to 6 alternative models (although usually between 3 and 4 models depending on window size and taxon). Notably, the set of 3 models that most frequently fit the whole NDD and the power law

region of the NDD in both diatoms (58.2% and 34.3% of all windows, respectively) and fish 54.4% and 15.2% of all, respectively), contained the exponential distribution, the truncated power law distribution, and the Weibull distribution.

Model fits as a function of scale and network properties

Only three distinct groupings of model classes were observed among the plausible sets of best fitting models: broad-scale only (B), single + broad-scale models (SB), and single-scale + broad-scale + power law models (SBP). Fisher's exact tests on both the diatom and fish data indicated that spatial extent influenced the frequency of particular classes of models among the best models fit over the NDD, with broad-scale models generally decreasing in frequency and single + broad-scale models increasing in frequency (Fig. 4.4a, both $P < 0.05$). However, this association was insignificant when considering models fit over the power law region of the NDD (Fig. 4.4b). Stepwise logistic regressions tested which variables, i.e. window size, node count, edge count, or connectance, predicted the probability of best fitting model class/classes (Appendix 4.1, Table S4.2). Important predictors of model class over the whole NDD included window size, node count, and edge count for diatoms, but node count and connectance for fish. Although window size was a significant predictor for fish, it did not enter the logistic model, likely because its effect was subsumed by other variables already in the model. Multinomial logistic regressions performed on the power law region of the NDDs indicated that window size was a significant predictor of model class only in the diatom dataset and only for distinguishing SBP (single-scale + broad-scale + power law) from B (broad-scale). Other difference between the diatom and fish multinomial logistic regression results was the greater number of network parameters in the fish model.

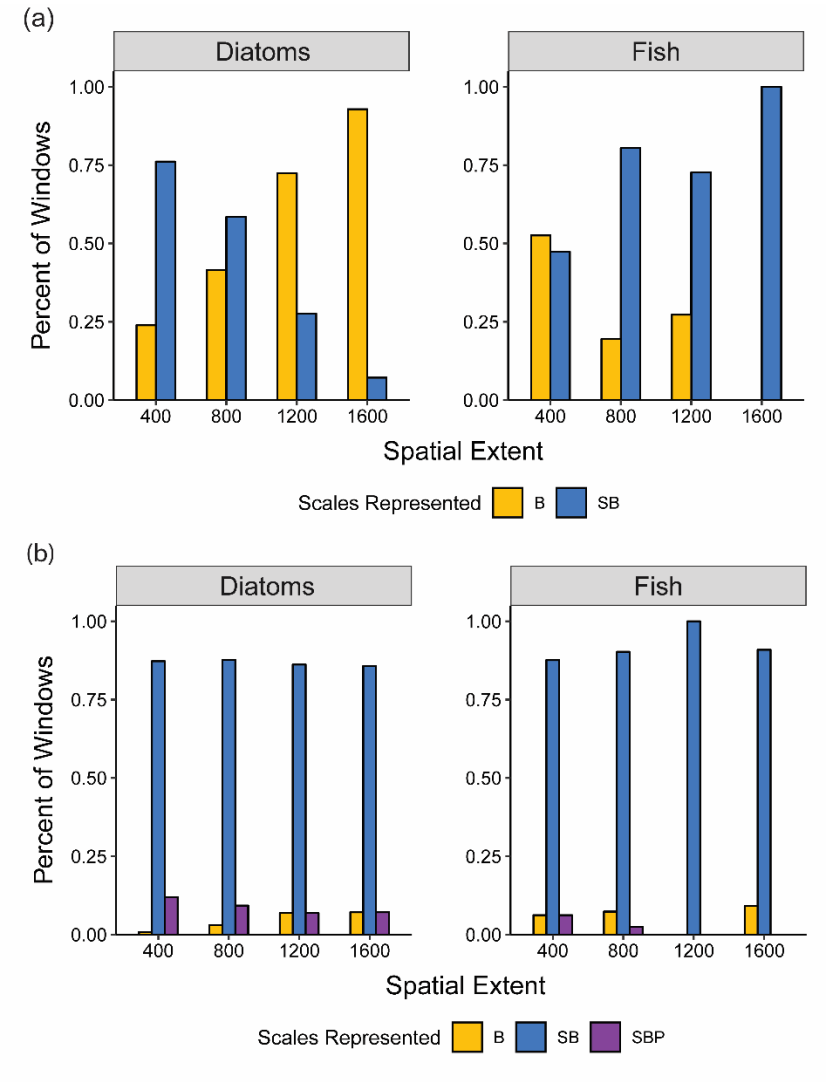

Fig. 4.4. Frequency chart of the classes of models (single - scale, S, broad-scale, B, or power law, P) represented among the best fitting models for (a) the whole node degree distribution (NDD) and (b) the estimated power law region of the NDD across spatial scales. Multiple simultaneous letters indicate the combinations of classes that were observed among the best fitting models.

Discussion

Our subcontinental study reveals that spatial extent along with basic network properties are important determinants of the variety of node degree distributions in co-occurrence networks. We further demonstrate that taxonomic considerations are crucial as biological differences underlie the shape of the NDD and its response to spatial extents and/or network properties.

Spatial considerations on network size and connectance

In pursuit of our first objective, we found that along the spatial gradient, co-occurrence networks sharply increased in node counts, while edge counts were not affected by windows beyond 400 km. This led to a corresponding decrease in connectance. Our findings for co-occurrence networks are thus consistent with past work on other ecological networks, such as food webs and pollinator networks (Fonseca & John, 1996; Thompson & Townsend, 2005; Olesen *et al.*, 2007; Ings *et al.*, 2009; Wood *et al.*, 2015; Galiana *et al.*, 2018). Two explanations are generally offered to explain network size patterns across spatial extents (Lafferty & Dunne, 2010). First, networks from larger spatial extents not only sample larger species pools, but are also more likely to include rare species by incorporating more sites. Second, larger areas are more heterogeneous and may allow species to coexist via niche partitioning, which may not be possible at smaller spatial extents. Given the subcontinental extent of our investigation and the difference in size of our windows, which sampled progressively larger species pools and heterogeneity, both explanations for the increase in network size with spatial extent are plausible. The negative response of connectance to spatial extent is attributable to the much slower increase in edge counts compared with node counts.

This finding appears to be in agreement with the constant species-link scaling hypothesis (Cohen & Newman, 1985). According to this hypothesis, connectance becomes sparser at larger network sizes because the number of links per species does not increase with network size but remains constant. This pattern has been observed historically in food webs (Cohen & Briand, 1984). Here, we explored whether this pattern could have occurred in our data by using piecewise regressions to test if there were any relationships between node counts and edges per node (Appendix 4.1, Fig. S6). In general, we found that indeed edge counts per node do not scale effectively with network size (with the exception of a few very small degrees in fish), consistent with the hypothesis. Taken together, this indicates that connectance of co-occurrence networks are partly constrained by constant link scaling effects as spatial scale increases, however this relationship may be less likely to be observed if networks are too small.

Spatial considerations on NDD and its power law region

Power law nearly always poorly described the NDD regardless of spatial extent. Instead, single-scale and broad-scale models provided better fits, consistent with a large body of literature (Dunne *et al.*, 2002; Williams, 2011; Wood *et al.*, 2015). However, we found that the frequency of these fits was determined by spatial extent, addressing our second objective and by taxon, addressing our fourth objective. As the spatial extent increased, the best fitting models of the whole NDD transitioned from single-scale to broad-scale in diatoms. In contrast, in fish, single scale models became more plausible at larger extents. Divergence in NDD shape has been previously reported but in a spatially-implicit manner, e.g. in plant-pollinator networks along altitudinal transects (single-scale to power law and vice-versa, Ramos-Jiliberto *et al.*, 2010) and

for mutualistic networks along global latitudinal gradients (albeit using evenness indices as proxies for NDDs, Sakai *et al.*, 2016). Transition of NDD shape with spatial extent, on the other hand, has not been explored. Thus, to our knowledge, this is the first investigation on co-occurrence networks to demonstrate in a spatially-explicit way that NDD shape shifts with spatial extent and that this shift is also taxon-specific.

The increase in broad-scale models with spatial extent (at the expense of single-scale models) in diatoms suggests that at larger extents some power law-like properties of the NDD emerge, namely increased probabilities of high degree nodes. Such nodes are typically treated as keystone species and/or generalists owing to their disproportionate degree values relative to the rest of the network (Freilich *et al.*, 2018; Galiana *et al.*, 2018). The prevalence of broad-scale NDDs at larger spatial scales in diatoms indicates that the process generating power law regions, i.e. preferential attachment (Barabási 2016), is scale-dependent. As diatoms are species-rich, highly dispersive and patchily distributed due to environmental filtering, constructing co-occurrence networks at smaller scales may eliminate large proportions of the species pool and with this, many potential interactions. Therefore, larger scales are necessary to adequately capture the species co-occurrence patterns of small-bodied organisms, such as diatoms. In contrast, in fish, the increase in single-scale best fits at larger spatial extents implies that the processes leading to preferential attachment (e.g., species interactions and sharing of niche space) operate at small spatial extents. Studies have correlated single-scale NDDs with strong spatial-constraints on species relationships (Barthélemy, 2003; Wright, 2010). This would suggest that stronger dispersal limitation in larger-bodied fish (Shurin *et al.*, 2009) may be an important mechanism that contributes to the increased abundance of single-scale NDDs as spatial extent increases.

Opposite to the whole NDDs, the classes of best model fits of the power law NDD region showed little spatial dependency for both taxa. This region was mainly characterized by both single-scale and broad-scale models in the two organismal groups, regardless of spatial scale, while the power law distribution rarely, if ever, was a good fit. Regrettably, studies explicitly comparing model fits of the power law NDD region have been ignored in the ecological literature, even though this is a very active area of research in general network theory (Broido & Clauset, 2019; Holme, 2019). The main impetus for assessing this region is that, after discarding "noisy" degree information contained in the lower tail, power law patterns are more clearly revealed (Barabási, 2016). Power law regions more strongly concordant with the power law are to be expected in networks with non-random association patterns (Tylianakis *et al.*, 2018). However, our findings that the power law was generally a poor fit of this region and instead was adequately explained by broad-scale models indicate that non-random association patterns, may be severely restricted within the estimated power-law region The probable explanation for this general pattern is that the power law region itself may be actually too small to be adequately fit by the model (i.e., estimated k_{min} too high) (Clauset *et al.*, 2009). However, more consequentially, these findings also indicate that scale-free NDDs are unlikely to be observed within co-occurrence networks, and if power law fits do produce a scale-free exponent, the actual model will generally be outperformed by single and broad scale model fits. Although it is possible to assess the truncated power law model for scale-free patterns (Barabási, 2016), which accounts directly for the exponential-cutoff behavior in the upper tail of the NDD, we observed only one instance of a scale-free model among the best fitting models in only the diatom data. Consequently, our findings are strongly consistent with Broido and Clauset (2019) who reported that, for a set of 495 biological networks comprising mainly food webs and protein interaction

networks and no co-occurrence networks, NDDs that conform to power laws in general, and to the scale-free condition in particular, are probably extremely rare. Our analysis complements this conclusion by showing that it applies also to co-occurrence networks across spatial scales and organismal groups.

In further support of our hypotheses, we observed spatial patterns in the number of best plausible models of the whole NDD, which decreased with scale in diatoms but increased in fish. Importantly, this result implies that as scale increases, co-occurrence networks converge on a smaller set of topographies and thus likely become more predictable in their whole NDD shape, but only in well-dispersing species, such as diatoms. Conversely, in more dispersal limited organisms, such as fish, predictability of the NDD was the highest at the lowest scale, where the number of best fitting models was the lowest. It is well established that large network sizes, coupled with very low connectance, such as those observed in this study, impose very strong constraints on the possible alternative configurations (i.e., NDD) of the network (Poisot & Gravel, 2014). However, as we see here, network size and connectance alone cannot explain the variability in NDD but organismal biology may also play a role. Specifically, network size increased and connectance decreased with window size in both diatoms and fish, but the trends in number of best fitting models diverged between the two groups, possibly as a result of their differential dispersal capacities. The number of best plausible models over the power law region was independent of scale for both taxa, which may be due to the comparatively limited size of this region, allowing multiple models to generate comparable fits.

To address our objectives, we examined how spatial extent and network properties correlated with the class of best fitting models for the whole NDD and the power law region of the NDD (third objective) and whether this correlation depended on organismal group (fourth

objective). Spatial extent and network size were found to be important determinants of model classes in both NDDs, but only for diatoms, whereas network properties alone were sufficient to predict the fish model classes. Thus, it is noteworthy that trait differences between species (here, body size) can potentially determine whether spatial constraints are direct (in the case of diatoms) or indirect through network properties (in the case of fish). Consistent with previous work (Thompson & Townsend, 2005; Wood *et al.*, 2015), we demonstrate that ecological interpretations of the origins of NDD variability cannot ignore spatial context, but add that organismal traits may further contribute to the direct importance of network properties vs. spatial extent. As ecological network analyses continue to evolve and find further applications, our research here shows that conclusions about species co-occurrence networks should be framed with respect to the spatial extent at which the network is constructed and the ecology of the focal group.

References

Albert, R., Jeong, H. & Barabási, A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.

Alstott, J., Bullmore, E. & Plenz, D. (2014) Powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, **9**, e85777.

Amaral, L.A.N., Scala, A., Barthelemy, M. & Stanley, H.E. (2000) Classes of small-world networks. *Proceedings of the national academy of sciences*, **97**, 11149–11152.

Araújo, M.B. & Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography*, **37**, 406–415.

Barabási, A.-L. (2016) *Network science*, Cambridge university press.

Barberán, A., Bates, S.T., Casamayor, E.O. & Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal*, **6**, 343–351.

Barthélemy, M. (2003) Crossover from scale-free to spatial networks. *EPL (Europhysics Letters)*, **63**, 915.

Bascompte, J. (2009) Disentangling the web of life. *Science*, **325**, 416–419.

De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, H., Denys, L., Vanhecke, L. & Van Der Gucht, K. (2012) Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology letters*, **15**, 740–747.

Broido, A.D. & Clauset, A. (2019) Scale-free networks are rare. *Nature Communications*, **10**, 1–10.

Clauset, A., Shalizi, C.R. & Newman, M.E.J. (2009) Power-law distributions in empirical data. *SIAM review*, **51**, 661–703.

Cohen, J.E. & Briand, Fredeiri. (1984) Trophic links of community food webs. *Proceedings of the National Academy of Sciences*, **81**, 4105–4109.

Cohen, J.E. & Newman, C.M. (1985) A stochastic theory of community food webs I. Models and aggregated data. *Proceedings of the Royal society of London. Series B. Biological sciences*, **224**, 421–448.

Csardi, G. & Nepusz, T. (2019) igraph: Network Analysis and Visualization.

Dag, O., Dolgun, A. & Konar, N.M. (2018) onewaytests: An R Package for One-Way Tests in Independent Groups Designs. *The R Journal*, **10**, 175–199.

Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, Graphs and Null Models:

Analyzing Bipartite Ecological Networks. *The Open Ecology Journal*, **2**, 7–24.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.H., Soto, D., Stiassny, M.L.J. & Sullivan, C.A. (2006) Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, **81**, 163–182.

Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002) Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, **5**, 558–567.

Echevarria, G. & Rodriguez, J.P. (2017) Co-occurrence patterns of fish species in two aquatic habitats of the Arauca River floodplain, Venezuela. *Community Ecology*, **18**, 137–148.

Estrada, E. (2007) Food webs robustness to biodiversity loss: The roles of connectance, expansibility and degree distribution. *Journal of Theoretical Biology*, **244**, 296–307.

Fonseca, C.R. & John, J.L. (1996) Connectance: a role for community allometry. *Oikos*, 353–358.

Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., Poulin, R. & Bascompte, J. (2010) Nestedness versus modularity in ecological networks: Two sides of the same coin? *Journal of Animal Ecology*, **79**, 811–817.

Freilich, M.A., Wieters, E., Broitman, B.R., Marquet, P.A. & Navarrete, S.A. (2018) Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology*, **99**, 690–699.

Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D. & Montoya, J.M. (2018) The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, **2**, 782–790.

Gause, G.F. (1932) Experimental studies on the struggle for existence: I. Mixed population of two species of yeast. *Journal of experimental biology*, **9**, 389–402.

Gillespie, C.S. (2015) Fitting heavy tailed distributions: The powerlaw package. *Journal of Statistical Software*, **64**, 1–16.

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

Guimarães, P.R., Machado, G., de Aguiar, M.A.M., Jordano, P., Bascompte, J., Pinheiro, A. & dos Reis, S.F. (2007) Build-up mechanisms determining the topology of mutualistic networks. *Journal of Theoretical Biology*, **249**, 181–189.

Holme, P. (2019) Rare and everywhere: Perspectives on scale-free networks. *Nature communications*, **10**, 1–3.

Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M.,

Van Veen, F.J.F., Warren, P.H. & Woodward, G. (2009) Ecological networks - Beyond food webs. *Journal of Animal Ecology*, **78**, 253–269.

Jordano, P., Bascompte, J. & Olesen, J.M. (2003) Invariant properties in coevolutionary networks of plant-animal interactions. *Ecology Letters*, **6**, 69–81.

Lafferty, K.D. & Dunne, J.A. (2010) Stochastic ecological network occupancy (SENO) models: a new tool for modeling ecological networks across spatial scales. *Theoretical Ecology*, **3**, 123–135.

Leboucher, T., Budnick, W.R., Passy, S.I., Boutry, S., Jamoneau, A., Soininen, J., Vyverman, W. & Tison-Rosebery, J. (2019) Diatom β-diversity in streams increases with spatial scale and decreases with nutrient enrichment across regional to sub-continental scales. *Journal of Biogeography*, **46**, 734–744.

Lewis, T., Pickl, S., Peek, B. & Xu, G. (2010) *Network science*, Cambridge university press.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.

Menzel, U. (2016) Package 'RMThreshold.'

Montoya, J.M. & Solé, R. V (2003) Topological properties of food webs: from real data to community assembly models. *Oikos*, **102**, 614–622.

Newman, M.E.J. (2007) Component sizes in networks with arbitrary degree distributions. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **76**, 45101.

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007) The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19891–19896.

Palmer, M.W. & White, P.S. (1994) Scale dependence and the species-area relationship. *The American Naturalist*, **144**, 717–740.

Poisot, T. & Gravel, D. (2014) When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ*, **2014**, e251.

Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005) Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, **20**, 345–353.

Ramos-Jiliberto, R., Domínguez, D., Espinoza, C., López, G., Valdovinos, F.S., Bustamante, R.O. & Medel, R. (2010) Topological change of Andean plant–pollinator networks along an altitudinal gradient. *Ecological Complexity*, **7**, 86–90.

Sakai, S., Metelmann, S., Toquenaga, Y. & Telschow, A. (2016) Geographical variation in the heterogeneity of mutualistic networks. *Royal Society open science*, **3**, 150630.

Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L.R., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Shurin, J.B., Cottenie, K. & Hillebrand, H. (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia*, **159**, 151–159.

Sugiura, S. (2010) Species interactions–area relationships: biological invasions and network structure in relation to island area. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 1807–1815.

Thompson, R.M., Dunne, J.A. & Woodward, G.U.Y. (2012) Freshwater food webs: towards a more fundamental understanding of biodiversity and community dynamics. *Freshwater Biology*, **57**, 1329–1341.

Thompson, R.M. & Townsend, C.R. (2005) Food-web topology varies with spatial scale in a patchy environment. *Ecology*, **86**, 1916–1925.

Tylianakis, J.M., Martínez-García, L.B., Richardson, S.J., Peltzer, D.A. & Dickie, I.A. (2018) Symmetric assembly and disassembly processes in an ecological network. *Ecology letters*, **21**, 896–904.

Watts, D.J. & Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.

Williams, R.J. (2011) Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological Networks. *PLOS ONE*, **6**, e17645.

Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015) Effects of spatial scale of sampling on food web structure. *Ecology and evolution*, **5**, 3769–3782.

Wright, C.K. (2010) Spatiotemporal dynamics of prairie wetland networks: power-law scaling and implications for conservation planning. *Ecology*, **91**, 1924–1930.

Appendix 4.1. Ancillary Figures and Table Output

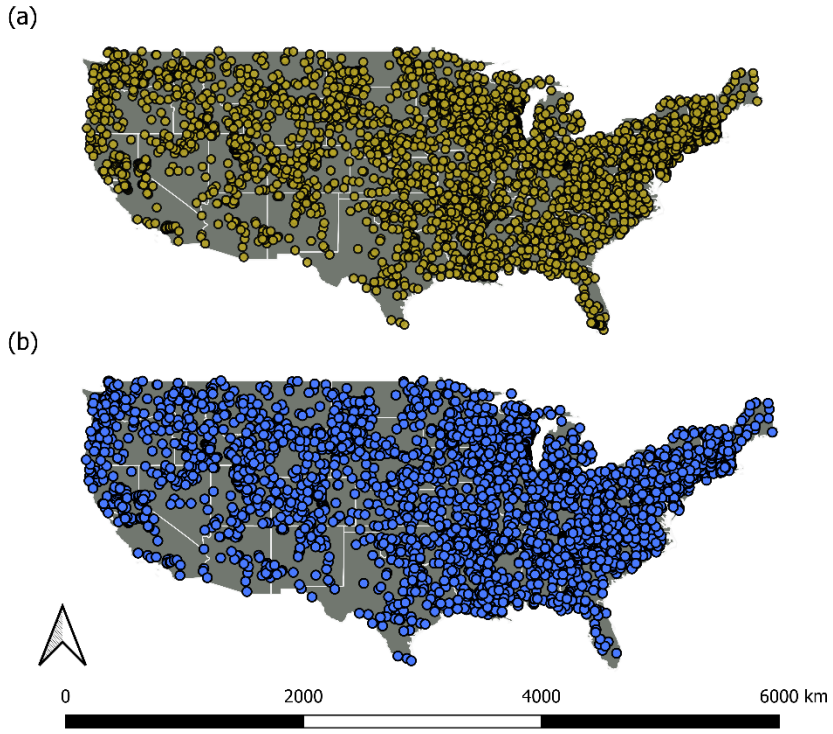

Fig. S4.1. Continental distributions of sampled sites for (a) diatoms, and (b) fish. Each circle represents a single stream sample.

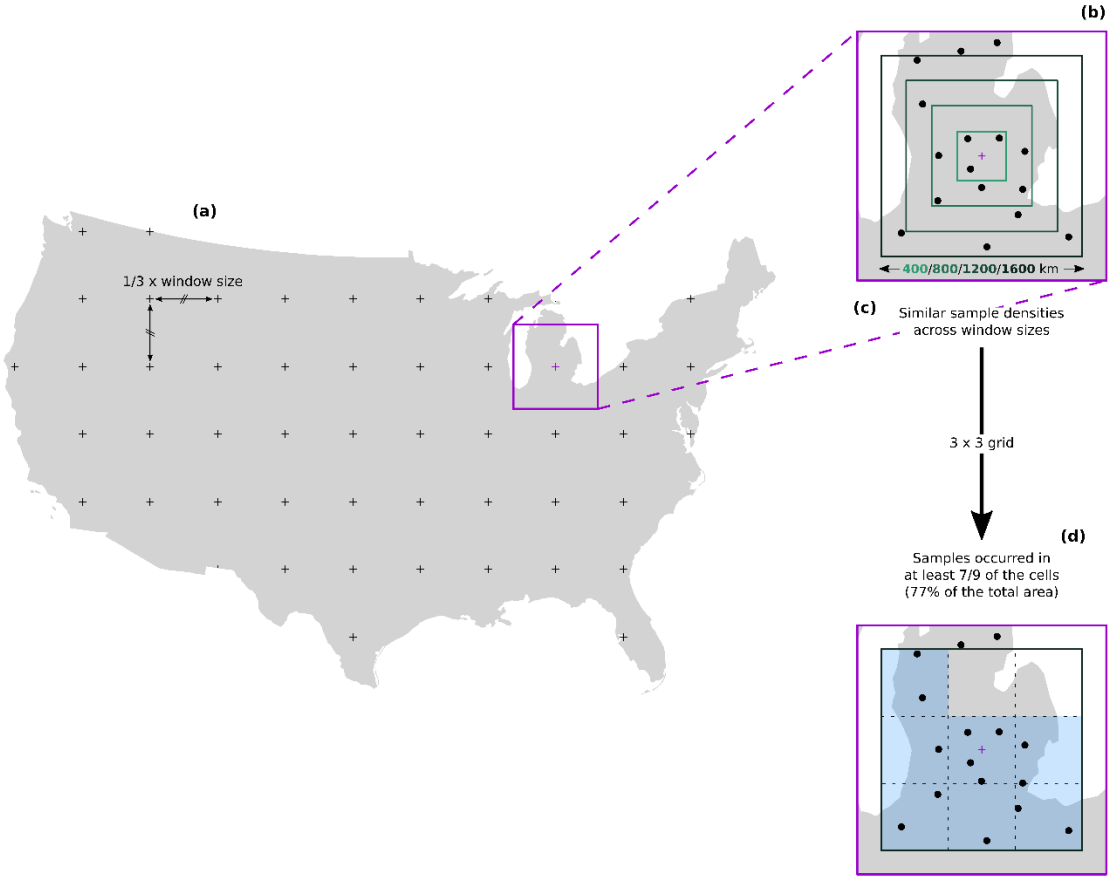

Fig. S4.2. Schematic of landscape windows construction. (a) Origin coordinate points are placed across a continental map layer, spaced apart at a distance (in km) of 1/3 the desired window size. Thus, for example, for a 400 × 400 km window, the origin spacing will be 133km. (b) For each origin, a landscape window of the desired size is constructed and centered around the origin point. (c) The window is then examined for validity by dividing the window into a 3 × 3 grid and determining if samples are present in at least 7/9 cells.

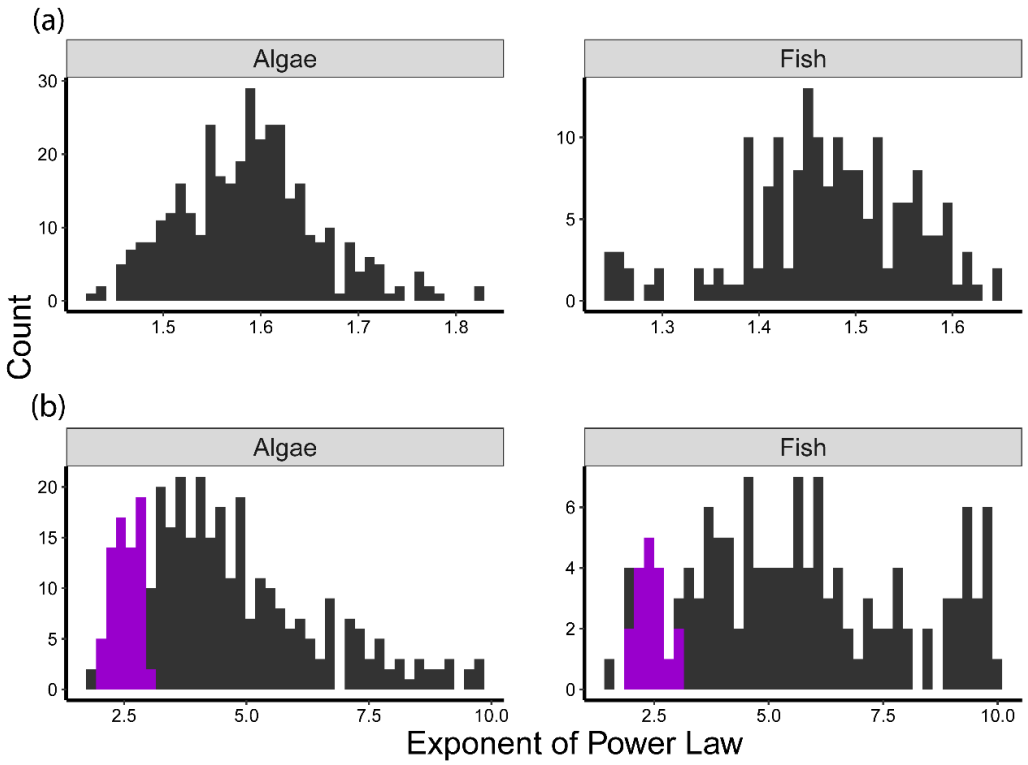

Fig. S4.3. Histograms of power law exponents estimated from power law models fit over (a) the whole node degree distribution, and (b) the power law region of the NDD. Purple bars indicate those power law fits whose estimated exponents fell within the scale-free range, $2 \leq \gamma \leq 3$.

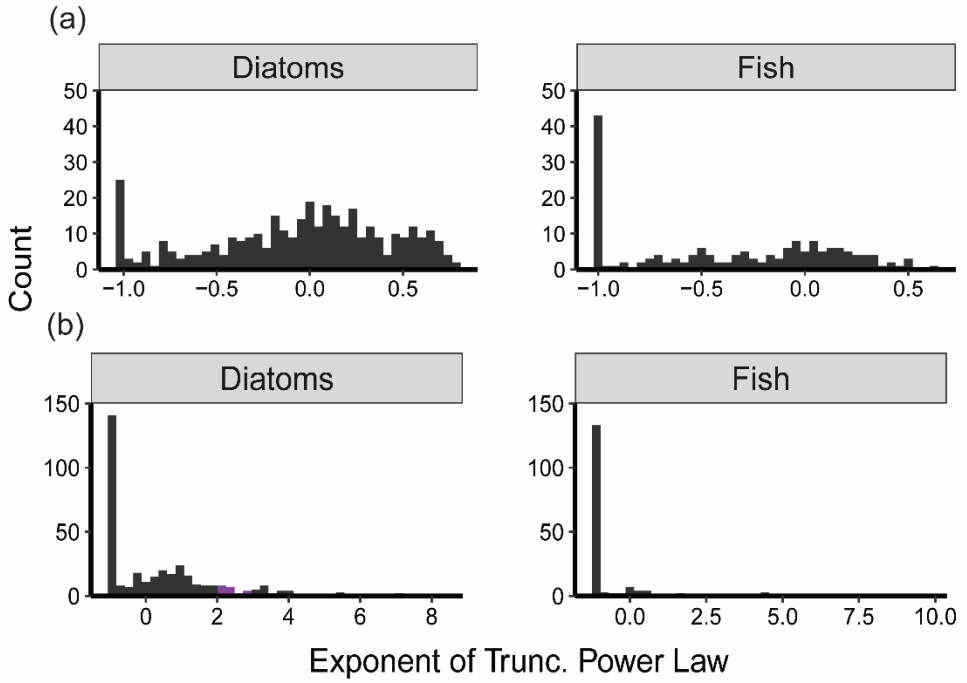

Fig. S4.4. Histograms of truncated power law exponents estimated from truncated power law models fit over (a) the whole node degree distribution, and (b) the power law region of the NDD. Purple bars indicate those power law fits whose estimated exponents fell within the scale-free range, $2 \leq \gamma \leq 3$.

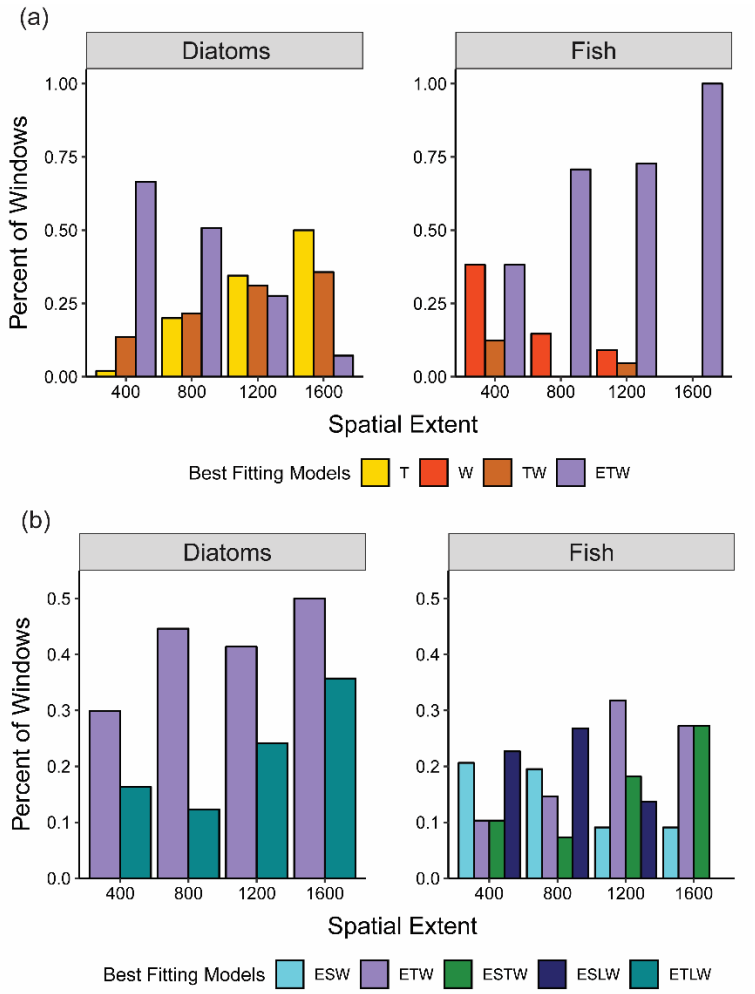

Figure S4.5. Frequency chart of the types of plausible models of (a) the whole node degree distribution (NDD) and (b) the estimated power law region of the NDD. Within each row, an individual letter corresponds to one of 6 models: E = exponential distribution, S = Poisson distribution, T = truncated power law distribution, L = lognormal distribution, W = Weibull distribution, P = power law distribution. Multiple simultaneous letters indicate where several models had $\Delta AIC < 2$. Only models and model combinations comprising > 10% of total windows for at least one spatial scale (for the whole NDD) and >20% (for models fit over the power law region) are shown.

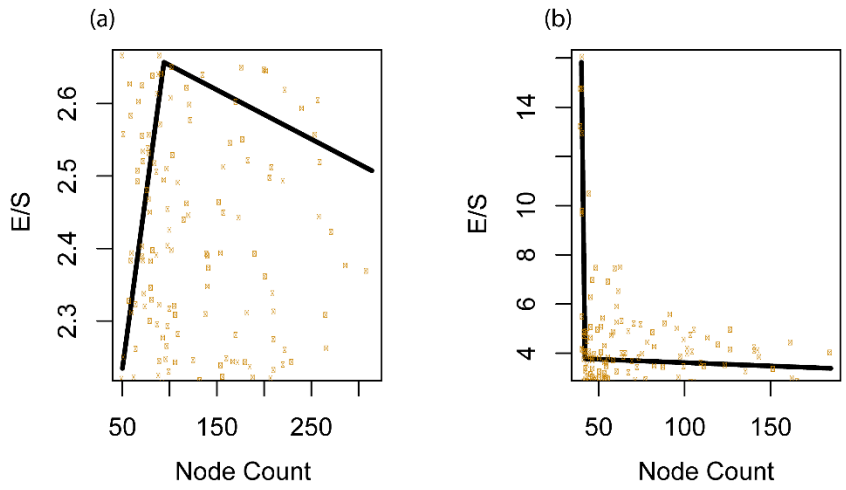

Fig. S4.6. Plots of piecewise regressions examining the relationship between network size (node count) and the number of edges per species node (E/S) for (a) diatom networks and (b) fish networks. Yellow dots represent individual networks. Black lines represent the regression segments. A single breakpoint existed for the diatom data (indicating the presence of non-zero slopes), and the model was generally poorly fit ($R^2 = 0.02$), with slope coefficients above and below the breakpoint very close to 0. Further, below the estimated breakpoint (node count = 91), the model almost encompassed nearly half the data (44% of windows) network node. Fish also showed a breakpoint in their data, and the model indicated the relationship between edges per node and node count was strongly negative ($R^2 = 0.65$), at node counts < 42 before becoming virtually non-existent at larger node counts. However, the data below the breakpoint only comprised 13 windows, indicating that this relationship existed in a very narrow range of small network sizes.

Table S4.1. Number of landscape windows generated and summaries of site counts within each window scale.

Taxonomic Group	Window Scale (km)	No. Windows	Minimum Site Count	Mean Site Count (SD)	Maximum Site Count
Diatoms	400	251	46	86 (37.56)	278
	800	65	184	292 (89.07)	501
	1200	29	420	603 (135.41)	849
	1600	14	755	998 (183.67)	1314
Fish	400	97	62	116 (32.34)	219
	800	41	182	333 (98.73)	569
	1200	22	411	633 (170.89)	948
	1600	11	750	1024 (251.12)	1461

Table S4.2. Results of binary and multinomial logistic regression tests examining whether network properties (node and edge counts and connectance) and spatial scale (window size) predict the classes of models (single-scale, broad-scale, power law, or a combination) observed among the best fitting models. Binary logistic regressions were fit for models describing the whole NDD (there were only two dependent variable states, B and SB). Multinomial logistical regressions were used for testing NDD models fit over the power law region (there were three dependent variable states, B, SB, and SBP). Letters in the outcome variable column represent model classes observed among the best fitting models (SB = single-scale + broad-scale; SBP = single-scale + broad-scale + power law). Reference categorical outcome for both binary and multinomial logistic regression is "broad-scale".

Logistic Regression:	Taxa	Outcome Variable	Variable	Coefficient	Standard Error	*Z*	P
Binary	Diatoms	SB	Node Count	0.019	0.008	2.447	0.014
			Edge Count	-0.005	0.002	-3.480	< 0.001
			Scale	-1.507	0.398	-3.785	< 0.001
	Fish		Node Count	0.023	0.008	2.765	0.006
			Connectance	-4.032	1.846	-2.185	0.029
Multinomial	Diatoms	SB	Node Count	-0.0016	0.012	-0.135	0.893
			Scale	-0.0018	0.0019	-0.922	0.357
		SBP	Node Count	0.01409	0.013	1.078	0.284
			Scale	-0.005	0.002	-2.275	0.023
	Fish	SB	Node Count	0.094	0.041	2.321	0.020
			Edge Count	-0.024	0.008	-3.048	0.002
			Connectance	11.145	5.515	2.020	0.043
		SBP	Node Count	0.029	0.041	0.697	0.486
			Edge Count	-0.022	0.008	-2.640	0.008
			Connectance	-2.031	0.564	-3.598	< 0.001

Chapter 5
General Conclusions

The goal of my research was to understand the patterns, the causes, and the consequences of biodiversity change and species co-occurrence. I investigated these community properties from subregional to global scales and along natural and anthropogenic gradients, including climate, eutrophication and land use.

In chapter 2, I explored biodiversity responses to land cover (agriculture vs. forest) in major stream organisms, including diatoms, insects and fish. The main research questions I investigated in this chapter surrounded whether homogenization of aquatic communities (i.e., increased similarity in species compositions) was a general consequence of agriculture, if so, what were the driving mechanisms (local vs. regional), and if there were taxonomic differences in responses owing to differences in body size and dispersal capacity. By examining the α (local diversity), γ (regional diversity) and β (community variation) components of biodiversity, I found that streams with substantial agricultural land cover contain communities that were generally homogenized in their constituent species as a result of decreased γ-diversity, increased α-diversity, and increased unevenness in relative abundances. By using a null model, I found that agriculture did not affect the contributions of local vs. regional processes to β-diversity, which however varied among studied groups, but controlled the balance between different local mechanisms. The major takeaway with this study was that agriculture represents a major detriment to global aquatic biodiversity, but the homogenization mechanisms may vary across organismal groups differing in body size and dispersal capacity but not trophic position.

In Chapter 3, I examined how the amounts and ratios of nutrients (nitrogen and phosphorus) determine algal community co-occurrence network structure, metacommunity

compositions, and their responses to climate and dispersal gradients. The main motivation for this investigation was that nutrient effects on co-occurrence network topology, particularly in freshwater communities, are understudied. I investigated i) whether nutrient supply (eutrophic vs. oligotrophic) and ratio (N-limited vs. P-limited) constrain topological properties of algal co-occurrence networks and ii) to what extent climate and space (a surrogate for dispersal) affect co-occurrence network topology vs. metacommunity composition across nutrient supply and ratio contexts. I found that nutrient supply was positively related to network size (node and edge counts), which was associated with observable increases in motile algal species, while N-limitation was linked to higher node counts. Node clustering patterns in the network varied within both nutrient contexts, while other topological differences were generally smaller. Climatic and spatial factors had pronounced effects on network topology that further depended on nutrient context. Thus, the oligotrophic and P-limited networks exhibited much greater change when climate and/or space were controlled for compared to respectively eutrophic and N-limited networks. A comparative assessment of network vs. compositional responses to climate and space identified an important distinction—while climate and space contributed to network topology, space was the dominant factor behind compositional variability, regardless of nutrient context. My findings highlight the need for developing integrative multi-level approaches (from metacommunities to co-occurrence networks) to fully understand biological responses to complex and interactive abiotic forces.

In chapter 4, I examined co-occurrence networks of stream diatoms and fish from the perspective of the node degree distribution (NDD), which statistically represents the connection patterns observed in the network. The shape of NDDs is a very active area of research in network theory, but comparatively neglected in ecology. As spatial extent dictates the shapes of

many ecological patterns, I predicted that it would also constrain the shape of the NDD in conjunction with organism-specific traits, such as body size. Therefore, the goal of this chapter was to describe the shape of NDD, using single-scale (exponential and Poisson distributions), power law or broad-scale/cross-over models (truncated power law, lognormal, and Weibull distributions), and whether the model fits depend on spatial extent and organismal group. I found that the type and number of best fitting statistical models depended on spatial scale. However, neither the power law, nor its truncated version, was ever a plausible model fit regardless of spatial extent, but most NDD models fell either in the single-scale or broad-scale categories. Finally, I observed taxonomic patterns in the model fits, transitioning from primarily broad-scale to primarily single scale with increasing spatial extent (diatoms) and vice versa for fish. This work, being the first to comprehensively demonstrate explicit spatial and taxonomic constraints on co-occurrence network NDDs across hundreds of networks, broadens the understanding of the origins of network variability.

Biography

William R. Budnick was born on October 10, 1989 and grew up in Seymour, CT. He graduated from Auburn University with a B.S. in Fisheries and Applied Aquacultures in May 2012. He then attended Louisiana State University for his Master's degree, where he studied geographic distributions of crayfish in Louisiana under his advisor, Michael Kaller. He received his Master's degree in May 2015. He then attended University of Texas at Arlington for his Ph.D., where he studied and worked under the guidance of his doctoral advisor Sophia Passy to learn how to apply complex ecological theory towards the study of aquatic biodiversity and conservation. As of this publication, he returned to the fisheries realm with a Post-doc position at Michigan State University under Brian Roth to pursue his passions in aquatic conservation, fisheries management, applied theoretical ecology, and, of course, astacology.