

DOES ALGAL BIODIVERSITY AFFECT COMMUNITY BIOMASS IN NATURAL LAKE
ECOSYSTEMS?

by

Emily K. Zimmerman

A thesis submitted
in partial fulfillment of the requirements
for the degree of
Master of Science
(Natural Resources and Environment)
In the University of Michigan
April 2013

Thesis Committee:

Associate Professor Bradley J. Cardinale, Chair
Professor David Allan
Professor Donald R. Zak

ABSTRACT

Over the past several decades, a large body of research has examined how biodiversity loss influences the functioning of ecosystems, as well as the cascading impacts on the goods and services ecosystems provide to humanity. The relationship between biodiversity and various ecosystem-level functions quantified in experiments to date suggests that initial losses of biodiversity have relatively small impacts on properties like community biomass production; however, beyond some threshold, increasing losses lead to accelerating declines in function. Some have questioned whether a saturating relationship between diversity and community biomass production is an artifact of overly simplified experiments that manipulate diversity in homogeneous conditions over short time-scales in which niche differences may not be realized. Others have questioned whether even the modest effects of biodiversity observed in experiments would be discernible in natural systems where they could be over-ridden by the stronger influence of abiotic factors.

Here, I used a biogeographic dataset to assess how the taxonomic richness of aquatic primary producers relates to community biomass in unmanipulated lake ecosystems, and then compared these findings to prior experiments. I used Structural Equations Modeling (SEM) to quantify statistical relationships between algal richness and community biomass while accounting for covariance with environmental parameters measured in the USEPA's National Lakes Assessment (NLA), which sampled 1157 freshwater lakes across the U.S. These analyses converged on a single best-fit model ($X^2 = 0.31$, $P = 0.58$) wherein total community algal biomass was a function of three explanatory variables – nitrogen, phosphorus, and algal taxa richness. The quantitative magnitude of the algal diversity (x) - biomass (y) relationship in the NLA dataset suggests that experiments to date have, if anything, underestimated the relationship

between diversity and biomass production in more natural, unmanipulated systems. I discuss possible implications of this finding to future experimental manipulations and conservation strategies.

<u>TABLE OF CONTENTS</u>	<u>PAGE</u>
Abstract	iii
Table of Contents	iv
Introduction	1
Methods	4
Results	12
Discussion	15
Tables	21
Figures	27
Literature Cited	31
Supplemental Text	36

INTRODUCTION

Over the past two decades, accelerating biodiversity loss has led to formation a rapidly expanding field of biology that has become colloquially known as the field of Biodiversity and Ecosystem Functioning (hereafter, BEF) (Loreau, 2010). The goal of BEF research is to understand how biodiversity loss (genes, species, functional traits) impacts the suite of ecological processes that collectively control the 'functioning' of ecosystems and, in turn, the services that ecosystems provided to humans (Tilman et al., 1999; Naeem et al., 2002). As of 2009, researchers had completed nearly 600 manipulations of species richness for >500 types of organisms inhabiting 30 biomes to examine how diversity impacts community and ecosystem-level processes like resource use and biomass production (Cardinale et al., 2011). Meta-analyses of these experiments have shown that the efficiency of resource use, and production of community biomass, increase as a function of species richness (Balvanera et al., 2006; Cardinale et al., 2006; Schmid et al., 2009; Cardinale et al., 2011). However, these same analyses have revealed that most ecological processes saturate at relatively low levels of diversity, with half-saturation constants suggesting the majority of function is achieved by the first 2-4 species (Cardinale et al., 2006; Schmid et al., 2009; Cardinale et al., 2011).

The preponderance of positive, but quickly saturating curves relating biodiversity to ecosystem function has led to speculation about whether results from prior BEF experiments accurately portray the 'true' effects of biodiversity of the properties of natural ecosystems. Some researchers have argued that these quickly saturating curves might be artifacts of the overly simplified conditions of experimental studies (Duffy, 2009; Hillebrand & Matthiessen, 2009). While levels of 'realism' vary considerably among experiments, it is noteworthy that the median experiment to date has been performed at a spatial scale that is approximately equal to the size of

a 5-gallon bucket, and has run for < 1 generation of the focal organisms (Cardinale et al., 2009a; Cardinale et al., 2011). This has led some to hypothesize the true effects of biodiversity might be greater than that revealed by experiments (Fig. 1A) because increasing spatial and temporal scales should allow for more environmental variation to be realized and, in turn, more of species niche differences to be realized (Tylianakis et al., 2008; Duffy, 2009). Consistent with this hypothesis, the limited number of BEF experiments that have been run for many generations do, in fact, find that the impacts of biodiversity on community biomass tend to grow stronger through time (Cardinale et al., 2007, Stachowicz et al., 2008, Reich et al., 2012). Furthermore, some evidence suggests the effects of biodiversity on community biomass grow stronger as the spatial scale of experiments increase (Dimitrakopoulos & Schmid, 2004; Cardinale et al., 2011; Griffin & Cardinale, 2013).

While many have claimed that biodiversity effects should grow stronger in more realistic environments with greater spatial and temporal heterogeneity, there is an alternative hypothesis that is rarely cited, and which has received far less attention. Some have argued that, when compared to experiments, species diversity should become less important in controlling ecosystem properties in natural systems because other environmental variables (nutrients, temperature, light, herbivores, etc.) will have far stronger impacts on ecosystem functioning (Fridley, 2002; Wardle & Zackrisson, 2005; Grace et al., 2007). Proponents of this hypothesis argue that, while diversity may have significant impacts in more natural systems, those impacts will be trivial compared to those of other abiotic controls over ecosystem processes. In turn, diversity effects will be rendered undetectable in anything other than highly controlled, contrived experiments. There have been few direct tests of this hypothesis, and for those that do exist, conclusions are somewhat mixed. Grace et al. (2007), for example, found that the relationship

between plant diversity and community biomass production in unmanipulated grassland ecosystems is weak relative to other environmental variables. In contrast, Paquette & Messier (2011) found that the relationship between tree diversity and annual wood production in forests was positive and significant after statistically controlling for other environmental variables that influence tree growth.

Here, I present results from a study that was designed to quantify the relationship between species richness of primary producers and community biomass in natural lakes after accounting for environmental correlates, and then compare the magnitude of the richness-biomass relationship to that measured in past biodiversity experiments. Using a large biogeographic dataset of algal diversity and community biomass in freshwater lakes, I demonstrate that species richness is positively related to algal biomass, and can be quantified by the same power function that is common in experimental manipulations. The scaling coefficient of this power function relating algal richness and biomass is significantly positive, and remains so after statistically controlling for a suite of other environmental variables that are also known to influence the biomass of primary producers. But the scaling coefficient quantified for natural lake ecosystems in this study is significantly larger than what has been previously quantified in experimental manipulations of biodiversity for comparable organisms. This latter finding is most consistent with the hypothesis that experiments underestimate the relationship between biodiversity and ecosystem-level properties in natural ecosystems. If this conclusion holds broadly true as more surveys amass in natural systems, then one of the key challenges will be to determine what sources of 'realism' are missing from BEF experiments, and how the downward bias might be corrected to arrive at a more predictable diversity-function relationship.

METHODS

United States Environmental Protection Agency's National Lakes Assessment – Overview

The bulk of data used in this study came from a large observational dataset that was collected by the United States Environmental Protection Agency (hereafter, U.S. EPA) referred to as the National Lakes Assessment (hereafter, NLA). The original purpose of the NLA was to collect base-line measures of the condition of the U.S.'s lake ecosystems. To be included in the NLA, lakes had to have been a natural or man-made freshwater lake, pond, or reservoir occurring in the Continental United States, be greater than 10 acres (4 ha) in area, be at least one meter in depth, be accessible by field sampling crews, and have a minimum of a quarter acre of open water. The Great Lakes and the Great Salt Lake were not included in the survey. Lakes were chosen from the National Hydrography Dataset using two methods; 1,033 lakes were chosen randomly using a statistical survey design and 124 lakes were chosen as reference lakes, with the latter selected non-randomly by the U.S. EPA to represent the 'least-disturbed' lakes in the U.S. (Fig. 1B).

Collection of data for the NLA took place during the summer of 2007. Field sampling for each of the lakes was completed in one-day increments by 86 teams composed of two to four technicians trained and deployed by the U.S. EPA. At each lake, the field sampling crew followed standardized protocols (publicly available at http://water.epa.gov/type/lakes/lakessurvey_index.cfm) to collect samples at the deepest point of the lake and at ten different locations around the perimeter. Data collected on site, such as temperature and pH, were entered by field sampling crews into standardized data forms, whereas other samples were shipped to common laboratories for additional analyses. More than 680,000

measurements of the chemical, physical, biological, and recreational characteristics of the lakes were quantified.

The U.S. EPA employed a Quality Assurance Project Plan to ensure quality control at all levels of the study, from data collection by field sampling crews to standardized and central data management. All water chemistry samples were analyzed at the same laboratory under standard operating protocols administered by the U.S. EPA's Western Ecology Division. Zooplankton and diatom samples were sent to four different laboratories around the country, whereas phytoplankton samples were all processed at one laboratory. Laboratories were audited for adherence to the NLA standard operating protocols for benthic processing. All laboratories were subject to internal quality control on sorting and identification using the Integrated Taxonomic Information System and external quality control by independent taxonomists contracted to audit ten percent of each laboratory's samples. Analyzed data were organized and entered into a series of spreadsheets that are publically available at http://water.epa.gov/type/lakes/NLA_data.cfm.

National Lakes Assessment – Extraction and Summary of Relevant Variables

I extracted variables from the U.S. EPA NLA that quantified algal diversity, algal biomass, and a suite of environmental variables known to influence algal biomass. Algal richness was taken to be the taxonomic richness in samples taken from a lake (lowest possible taxonomic unit, which was usually genus), and biomass was estimated as the summed biovolume across all algal species in the sample ($\log [\mu\text{m}^3/\text{L}]$). Algal taxon richness and biomass were taken from the “Lake Phytoplankton Soft Algae Count Data” spreadsheet and the “Lake Phytoplankton Diatom Count Data” spreadsheet. I further selected ten environmental variables from the U.S. EPA NLA to include in analyses, such as measures of nutrients, zooplankton diversity and abundance, and water chemistry. The rationale for including each of these

variables in the analyses, the sources of data for each variable, and the mean, variation and range of each variable measured in NLA lakes are given in Table 1.

It is important to note that the NLA dataset only had measures of standing stock biomass of algae at a single time point, and that rates of primary production (neither gross nor net) were measured in this study. The spatial extent of this dataset is unrivaled (~1,200 lakes across the Continental U.S.), but the large extent prohibited taking measures of biomass at multiple time points, or measuring gas exchange rates by phytoplankton in sealed bottles to estimate productivity *per se*. So it should be kept in mind that my analyses focus on how species richness relates to community-level biomass of algae measured at a single point in time across a large biogeographic gradient of many lakes. While the distinction between standing biomass (a stock) vs. primary production (a flux) is important, as these do not necessarily respond to diversity in the same way (Weis et al., 2007; Power & Cardinale, 2009), it should also be noted that most studies of biodiversity-"productivity" relationships actually measure standing stock biomass at a single point in time, not primary production (Mittleback et al., 2001; Cardinale et al., 2011).

Biotic data available from the NLA (zooplankton abundance and richness, algal taxa richness, total algal community biovolume) required a good deal of post-processing to be made useful for these analyses. Zooplankton were labeled with their functional feeding group by the U.S. EPA. Only herbivorous zooplankton were included in the dataset. Zooplankton abundance was standardized across sites based on volume (abundance/mL), whereas zooplankton taxa richness was calculated by summing the number of taxa present in each lake. Algal taxa richness was calculated by summing the taxa richness of algae and the taxa richness of diatoms for each site from their respective spreadsheets. Total algal community biovolume for each site was calculated by summing individual taxa biovolume for each site. Any taxa for which biovolume

data was not provided by the U.S. EPA was excluded from the analyses. In cases for which biovolume data was not available for a given sample, but was available for the same taxa elsewhere in the dataset, that biovolume was assigned to the given sample.

Outliers ± 2 SD from the mean were flagged for additional consideration. I considered characteristics of sites with extreme outliers in attempt to determine whether these were ‘real’ data points; however, no outliers were removed from my analyses because I could not find adequate biological or methodological evidence to remove any data points. Distributions of all variables were initially screened for linearity, normality, and correlations. Variables were log-transformed when necessary to improve the normality of the distributions.

Variables from Other Sources

Light was not explicitly measured during the U.S. EPA NLA; yet, its role in photosynthesis and primary production is obvious. Therefore, it was necessary to gather data for light intensities from an alternative source. Data for the monthly average solar radiation incident on the surface of the earth was taken from the National Aeronautics and Space Administration Atmospheric Science Data Center by entering geographical coordinates and month of sampling for each of the 1,157 lakes in the U.S. EPA NLA. The geographical coordinates for each lake were provided in the U.S. EPA NLA. Monthly average solar radiation incident ($\text{kWh m}^{-2} \text{day}^{-1}$) is a direct measure of the light available on the surface of the water (Table 1).

Analyses – Structural Equations Modeling

In the first step of the analyses, I used structural equations modeling (SEM) to quantify the relationship between algal taxa richness and algal community biomass while statistically controlling for the potential influence of other environmental variables. The initial model included 12 explanatory variables (Table 1, Fig. 2). While the NLA measured many more

variables than this (e.g., land use surrounding sampling sites, maximum depth of sampled lake, perimeter of lake, lake trophic condition, etc.), many of the variables have only indirect effects on algal biomass, and represent proxies for variables that have a more direct influence on algal biomass. For example, land use does not have a direct effect on algae, but instead, leads to changes in nutrient loading, conductivity, and other metrics of water quality that are of direct consequence. I identified 10 explanatory variables from the NLA dataset that are known to have direct influence on algal richness or biomass (Table 1), and I specified five covariance paths in the initial model (Fig. 2) based on previously documented correlations in the literature. To confirm that the covariance paths predicted from the literature were actually covarying in the NLA dataset, I ran a correlation matrix in Rv 2.15.0 and verified that all correlations specified based on the literature had $P < 0.1$. These were included in the initial model (Fig. 2). For example, prior studies have shown that nitrogen and phosphorus loading are highly correlated in lakes, and that the two inorganic nutrients have synergistic effects on algal growth that lead to more biomass in tandem than individually (Harpole et al., 2011). In the NLA dataset, the correlation between TN and TP was, in fact, strong and statistically significant ($r = 0.81$, $P < 0.01$). Thus, I included a covariance term between TN and TP in the initial structural equations model.

Beginning with the initial structural equation model (Fig. 2), I used the Lavaan package in Rv 2.15.0 (Rosseel, 2012) to parameterize the relationship between algal taxa richness and algal community biomass while considering the specified covariance terms and the effects of the selected environmental variables. I removed variables that improved the overall goodness of fit of the structural equation model by increasing parsimony. Specifically, I iteratively removed pathways with the high P -values from the full model that improved fit (decrease in AIC,

RMSEA). I used the chi-square (χ^2) statistic to assess the overall significance of model fits to data ($P > 0.05$) and Akaike Information Criteria (AIC) and root mean square error of approximation (RMSEA) to compare models of differing complexity (Akaike, 1974; Grace, 2006). In addition, I calculated the Akaike weights to estimate the likelihood that each model was the best fit to the observed data, given the suite of models hypothesized (Johnson & Omland, 2004).

Analyses – Parameter Comparisons

After finding the most parsimonious, best-fit Structural Equation Model, I used the resulting parameter estimates from that model to compare the quantitative magnitude of the algal diversity (x) - biomass (y) relationship in unmanipulated lakes to the magnitude of the relationship that has been quantified in prior experimental manipulations of diversity. I made the comparison using the scaling coefficient b from the power function, $y=ax^b$ where y is algal community biomass, and x is algal species richness. The power function has previously been used to summarize this same diversity-function relationship in prior BEF experiments (Cardinale et al., 2011). The power function was the best-fit in only 14% of prior experiments, which compares to 53% for the better-fitting Michaelis-Menten function. However, the power function is considerably more flexible than the Michaelis-Menten, and can fit a wider variety of functional relationships. In addition to its greater flexibility, experimental data fit to power functions had R-squared values that were similar to the Michaelis-Menten function ($R^2_{\text{Power}}=0.71$, $R^2_{\text{MM}}=0.73$) (Cardinale et al., 2011). Thus, while the power function is not the single best fit to prior experimental data, its greater flexibility to fit a variety of datasets coupled with its comparable explanatory power makes it useful for my purposes (see the same argument used to justify analyses by Reich et al., 2012).

For BEF experiments, values for the scaling parameter b were taken from Cardinale et al. (2011), which assessed curve fits to 24 experiments. The average value of b in the selected experiments was 0.21 ± 0.53 (Cardinale et al., 2011). The value of b for natural lakes was taken directly from my best-fitting SEM. To calculate values of b in a manner comparable to how they are calculated for experiments, I first standardized biomass at each level of richness to that of the average monoculture ($a = 1$) to represent the proportional increase in biomass per increasing levels of richness. In the natural lakes dataset, three lakes had only one algal taxa present; these three samples were averaged and taken as the average monoculture. I compared mean b values from the literature to the b value extracted from the best-fitting SEM for quantitative comparison.

Variables and Relationships Not Considered

The single greatest challenge in quantifying diversity-function relationships in natural systems is statistically controlling for the myriad of potentially confounding environmental variables. The second greatest challenge is limiting one's search for confounding variables to a list that is manageably sized, and which only includes variables that have a direct causal influence on the focal relationship of interest (in this case, richness and biomass). I extracted variables from the NLA that have been shown to influence the biomass of aquatic primary producers via well-known biological mechanisms (Table 1). At the same time, I excluded from consideration certain relationships or variables for which I could not envision a direct causal pathway. One variable that is prominently missing from my analyses is lake area. While many studies have demonstrated a positive correlation between area and both biodiversity and biomass production (Conner & McCoy, 1979; McGuinness, 1984; Lomolino, 2001), it is generally understood that area *per se* is not the direct causal factor. Rather, area is associated with other

factors such as habitat heterogeneity, resource loading, or species population sizes that do directly influence biodiversity or ecosystem function (Williams, 1964; Harner & Harper, 1976). I recognize that other researchers might have a differing opinion on the utility of surrogates like area, so I have included supplemental analyses (see page 36, Supplemental Text) that shows lake area has no bearing on the conclusions that follow from my analyses of this dataset.

It is also worth noting that I specifically chose to examine how biodiversity influences biomass, rather than the converse relationship. I recognize that a large body of historical research has examined how the "productivity" of ecosystems influences species diversity (Connell & Orias, 1964; MacArthur, 1965). Even so, recent theoretical and empirical work has made it clear that resource supply rates - not actual biomass production - are what influence biodiversity by regulating mechanisms of competition and coexistence (Loreau et al., 2001, Schmid 2002; Cardinale & Gross, 2007, Cardinale et al., 2009b, Cardinale et al., 2009c). Yet, empiricists routinely quantify standing biomass (which is easy to measure) as a proxy for resource supply rates (which are hard to measure), and then they plot richness as a function of biomass as if this is a causal relationship. This model of causality between diversity and biomass production is fundamentally incorrect, and is inconsistent with all known ecological theory (see Cardinale & Gross, 2007). Nevertheless, in the supplemental text (see page 41, Supplemental Text) I present analyses in which I contemplate the historically considered direction of causality and model algal species richness as a function of algal biomass. I show that the model in which the causal arrow between algal taxa richness and biomass production is reversed (biomass production drives algal taxa richness), was a poor fit because the expected and observed covariance matrices were significantly different ($\chi^2 = 6.75, P < 0.01$). Thus, the historical perspective that productivity drives diversity can be rejected as a viable explanation of the data.

RESULTS

The U.S. EPA NLA dataset was characterized by a large amount of variation in algal taxa diversity, primary production, and the different environmental parameters, many of which ranged by several orders of magnitude. A total of 1,006 taxa of phytoplankton were identified across the 1,157 lakes, which ranged from a minimum taxon richness of 22 to a maximum of 85. Taxa representing all major taxonomic groups of freshwater algae were found, including Bacillariophyceae, Chlorophyta, Chrysophyta, Cryptophyta, Cyanophyta, Euglenophyta, and Pyrrhophyta. The most commonly occurring taxonomic group in the dataset was Bacillariophyceae. Total community biovolume in each lake spanned 7 orders of magnitude (22 to 1,310,659,599 μm^3 biovolume L^{-1}). The range of the various environmental parameters is given in Table 1. The extreme levels of variation are ideal for purposes of this study, as modeling approaches like SEM become increasingly powerful and more reliably quantify relationships when variables span a large range.

Results of the SEM analysis indicate that the initial model (Fig. 2), which included all variables and specified covariance terms, was not the best-fit model. To find the best-fit model, I iteratively removed variables to improve the goodness-of-fit between the predicted and observed covariance matrices. This led to 15 additional models (Table 2), all of which represented alternative hypotheses to explain patterns of covariance in the dataset. In all models, including the initial model presented in figure 2, the pathway between algal taxa richness and total community biomass production was positive and significant. The best-fitting model to explain total algal community biomass was far more parsimonious than the original model, and included just three explanatory variables – total nitrogen, total phosphorus, and algal taxa richness (Fig. 3). The strongest predictor of total algal community biomass in this final model was total

nitrogen with a standardized partial regression coefficient of 0.26 ($P < 0.05$). The standardized partial regression coefficient from total phosphorus to total algal community biomass was 0.15 ($P < 0.05$), and the standardized partial regression coefficient from algal taxa richness to total algal community biomass was 0.23 ($P < 0.05$). Additional positive covariance terms existed between algal taxa richness and total nitrogen and between total nitrogen and total phosphorus. Only the latter of these was statistically significant ($P < 0.05$). The best-fit model with total nitrogen, total phosphorus, and algal taxa richness (Fig. 3) explained ~21% of the total variation in algal community biomass across all 1,157 lakes with a chi-square (χ^2) of 0.31 ($P = 0.58$) and a root mean square error of approximation (RMSEA) of 0.07 ($P = 0.88$). For the χ^2 statistic, a *P*-value greater than 0.05 indicates that the expected and observed covariance matrices are not significantly different; thus, the hypothesized model cannot be rejected as a viable explanation of the data.

I calculated the Akaike weights of all 16 SEM models, which use the likelihood values to compare how probable each model *i* is as an explanation for the observed data, given the suite of models hypothesized. The Akaike weight for the best-fit model (Fig. 3) was 0.74, indicating that the best-fit model in Figure 3 had a 74% chance of being the best fit model for the observed data, given the suite of models hypothesized (Table 2). The second most likely model, which only differed in that the covariance term between algal richness and total nitrogen was deleted, had a 26% chance of being the best fit model for the observed data (Table 2). All 14 remaining models had Akaike weights of 0.00, indicating that they were highly unlikely to be viable explanations of the observed dataset (Table 2).

One surprising result of the model comparison is that several models that included variables routinely thought to control algal biomass proved to be inferior explanations of the

observed covariance matrix, leading me to reject them as viable explanations of the data. For example, zooplankton taxa richness and zooplankton abundance - both of which can influence the magnitude of herbivory on algae, were not retained in my best-fitting models. This was surprising given the wealth of studies demonstrating the role of herbivory on biomass production in aquatic ecosystems (Cry & Pace, 1993). In addition to herbivory, light was also not retained in my best-fitting models, which was surprising given the obvious role of light in maintaining primary production.

After identifying the single best-fitting model, I used the parameter values from this model to compare the strength of the algal taxa richness (x) - biomass (y) relationship in the NLA dataset (Fig. 3) to the strength of relationships documented in prior experiments using the power function, $y=ax^b$. The unstandardized scaling coefficient b relating biovolume to taxa richness from the best-fitting SEM was 0.52 ± 2.41 (note that richness and biomass are both on a log scale in the SEM; thus the estimated slope is equal to b from the power function). Comparatively, the value of b that has been estimated from past experiments that explicitly manipulated the richness of freshwater algae and examined the impacts on biomass production was 0.21 ± 0.53 (Cardinale et al. 2011). Thus, the scaling coefficient relating algal biomass to diversity in natural lakes (this analysis) exceeds experimental estimates (past estimates) by a mean factor of 2.35, suggesting that experiments have potentially underestimated the diversity-productivity relationship in natural, unmanipulated systems (Fig. 4A). To understand the potential magnitude of underestimation, consider that experiments performed with a maximum of 32 species have estimated average polyculture produces a mean $2.07\times$ the biomass of the average monoculture (Fig. 4A). In contrast the average 32-taxa polyculture in the NLA dataset produces a mean $6.06\times$ as much biomass as the lowest diversity lakes (Fig. 4A). The predicted

experimental underestimation is relatively small at low levels of diversity loss, but grows disproportionately large as diversity loss increases (Fig. 4B). For example, when 10% of species are lost, the potential underestimate by experiments is ~3% (experimental estimates are relatively close to the relationship in real lakes) (Fig. 4B). However, when 50% of species are lost, the amount of function predicted to be lost differs by $\geq 16\%$ between prior experiments and real lake ecosystems (Fig. 4B). These results suggest caution may be warranted when attempting to extrapolate the results of biodiversity-ecosystem functioning experiments to predict the functional role of biodiversity, or the consequences of extinction, in more natural systems.

DISCUSSION

Over the past two decades, accelerating rates of biodiversity loss, and the growing concern about how these losses impact ecosystem functions, have led to an increase in the number of experiments focused on understanding how biodiversity impacts ecosystem-level processes like community biomass production (Loreau, 2010). These experiments have provided strong evidence that community biomass increases as a function of biodiversity; yet, many of the experiments have been completed at relatively small spatial and temporal scales (Cardinale et al., 2011). While many have called on researchers to increase the spatial and temporal scales used in BEF experiments in order to provide a more ‘realistic’ comparison to natural ecosystems, the practical limitations (funding, personnel, experimental control, etc.) to performing large-scale, long-term manipulative experiments make it unlikely that we will routinely mimic realistic scenarios of extinction any time soon. As an alternative, some have suggested that we complement the expanding number of BEF experiments with analyses of large, observational datasets that are representative of unmanipulated ecosystems (Loreau et al., 2001, Cardinale et

al., 2011); thus allowing us to compare the results of mechanistic, highly replicated BEF relationships performed in relatively simplistic environments to the BEF relationship that occurs in more natural ecosystems.

Several studies have now explicitly used large observational datasets to characterize the relationship between biodiversity and various ecosystem-level properties; but the results of these studies have proven to be mixed. Paquette & Messier (2011) used a large observational dataset from the Quebec Forest Survey to examine the relationship between tree biodiversity and wood production. They found a significant, positive relationship between tree functional richness and forest productivity that was strong even after holding several climatic variables constant. Mora et al. (2011) used a global survey of 1,906 reefs to evaluate the relationship between reef fish diversity and standing biomass of reef fishes. Using structural equation modeling to account for the potential influence of environmental, physiographic, and anthropogenic variables, Mora et al. (2011) suggested the relationship between fish biodiversity and standing biomass was stronger than that documented in prior experiments. In contrast, Grace et al. (2007) investigated the relationship between plant biodiversity and community biomass, while statistically controlling for environmental variables in 12 natural grassland systems spanning 4 countries. Using structural equations modeling, they found no significant relationship between plant diversity and community biomass in natural grasslands. Instead, Grace et al. (2007) suggested that producer biomass was more consistently and strongly controlled by abiotic variables.

Findings from this study are most consistent with the proposition that biodiversity's impact on community biomass in natural, unmanipulated systems is stronger than revealed by prior experiments (Fig. 1A). Indeed, the scaling coefficient relating biomass to diversity in natural lakes exceeds experimental estimates, suggesting that experiments underestimate the

relationship between diversity and community biomass in nature (Fig. 4). For those who have hypothesized diversity has a stronger impact on community biomass in natural ecosystems than in controlled experiments, the most common explanation for this proposed difference is that natural, unmanipulated ecosystems have more spatial and temporal heterogeneity, and that the additional heterogeneity allows more species niche differences to be expressed in unmanipulated systems than in relatively homogenous experiments. This is an assumption that has yet to be directly tested or verified in empirical studies of natural systems, and we have no ability to determine if this is indeed the underlying cause of a stronger diversity-biomass relationship in the NLA dataset. However, it is worth noting that select experimental studies do suggest that diversity effects grow increasingly strong at larger spatial and temporal scales as heterogeneity is manifest. For example, Stachowicz et al. (2008) used a dataset with a temporal scale of three years to compare the relationship between macroalgal diversity and community biomass accrual in marine intertidal zones, and found that the positive relationship between diversity and productivity became stronger over time likely because of facilitation and differential use of the heterogeneous environment. Similarly, Reich et al. (2012) showed that biodiversity effects on community production increased through time in two long term studies (≤ 13 years) at the Cedar Creek Natural History area, and that this trend occurred because of increasing complementarity of species over time.

A number of variables, such as light and zooplankton abundance, which would have been expected to be important drivers of algal biomass production based on studies by others (Hill & Knight, 1988; Feminella, 1995; Cry & Pace, 1993), were not included in the best-fit model. I can only speculate on reasons why variables were not significant. One possible reason for the lack of an effect of light could be the relatively coarse measure that was available for analyses.

Light obtained from NASA represented the total solar radiation incident on a horizontal surface at Earth's surface for a given month, with that value representing the average of a 22-year collection period. Thus, light intensities did correspond to actual condition on, or immediately preceding, the day of sampling for the NLA project, and that may explain why there was no signal on this particular date.

The lack of a signal of zooplankton is perhaps more perplexing given the well known influence of grazers on algal biomass and turnover rates in freshwater ecosystems (Cry & Pace, 1993). I offer two possible explanations of this lack of a signal. First, the NLA only had measures of zooplankton abundance available, but did not have estimates of biomass. Biomass would be a more appropriate measure of the potential influence of herbivores on algal biomass. Second, it is plausible that zooplankton influence rates of algal production and turnover, but the herbivory does not translate into an influence on standing stock biomass. For example, if an algal community is dominated by a large, inedible algal species, zooplankton may not strongly influence standing stock biomass. The turnover of rarer, edible algal species may be influenced by herbivory, but standing stock biomass may remain relatively unaffected due to the proportion of large, inedible algal species dominating the community. Unfortunately, measures available in the NLA dataset do not allow us to examine these possibilities; as such, I suggest interpreting the non-significant impacts of zooplankton (and light) cautiously.

As with all biogeographic studies, my work has several obvious limitations and boundaries. I address just such limitations here. First, while this dataset has unrivaled spatial resolution (variables from ~1,200 lakes across the Continental U.S. measured in a consistent way), all lakes were sampled at just a single time point. Because the data is only a snapshot of in time, I cannot test the temporal dynamics of the relationship between algal taxa richness and

community biomass, and cannot rule out the possibility that the data expressed a particularly strong relationship between algal richness and biomass at the particular time of sampling.

Second, structural equation modeling (SEM) is a means to test hypotheses about multivariate causal relationships; in the case of my study, the causal relationship between richness and community biomass while accounting for the influence of other environmental variables that affect biomass production. But conclusions from these SEMs are only as reliable as the hypothesized models. If the suite of hypothesized models does not contain the 'true' set of causal relationships, then my analyses would identify a 'best-fit' model - but one that is fundamentally incorrect. I made every attempt to evaluate a variety of hypothetical models showing how variables influence algal biomass; nevertheless, future work should continue to pit new, alternative hypotheses against the best fit obtained in Figure 3.

In conclusion, my study finds that in unmanipulated, heterogeneous ecosystems, biodiversity is more strongly related to important ecosystem processes like primary production than predicted from two decades of prior BEF experiments. There have been many hypotheses proposed for why the relationship between biodiversity and ecosystem functioning may be stronger in natural environments than prior experiments, including that greater spatial and temporal heterogeneity in natural environments allows more species niche differences to be expressed. However, as I mentioned above, the use of a large observational dataset such as the NLA does not allow me to differentiate between mechanisms that may be causing this stronger relationship, and it is important to emphasize that I cannot verify any assumptions as the underlying cause of the differences seen in Figure 4. Therefore, in order to more accurately inform conservation strategies in unmanipulated ecosystems, BEF experiments may need to increase their focus to determining the specific mechanisms that may be driving the stronger

relationship between biodiversity and ecosystem functioning that have been observed in a diverse number of natural systems.

TABLES

Table 1: Description of environmental variables taken from the USEPA NLA and NASA. The asterisk (*) in the units column indicates that the variable was log transformed before use in the analysis.

Variable	Units	Rationale (Reference)	Source	Mean \pm SD, (Min, Max)
Total Nitrogen	($\mu\text{g N/L}$) [*]	Important limiting nutrient to primary production in freshwater ecosystems (Elsner et al., 1990)	NLA Lake Water Quality Data	1162 \pm 2154 (10, 26100)
Total Phosphorus	($\mu\text{g P/L}$) [*]	Important limiting nutrient to primary production in freshwater ecosystems (Elsner et al., 1990)	NLA Lake Water Quality Data	107 \pm 270 (1, 4679)
Silica	($\mu\text{g SiO}_2/\text{L}$) [*]	Critical to the growth and ecology of diatoms, and may be a third limiting nutrient in systems lacking nitrogen and phosphorus (Martin-Jezequel et al., 2000; Evans et al., 2011)	NLA Lake Water Quality Data	8755 \pm 10630 (25, 91907)
Growing Days	Latitude (Decimal Degrees)	Latitude intended as a proxy for degree days, which are frequently used as a measure of growing period, or in the case of algae, ice-off period (Weyhenmeyer et al. 2013)	NLA Information for Lakes that were Sampled	41 \pm 5 (27, 49)
Dissolved organic carbon	($\mu\text{g DOC/L}$) [*]	Negatively impacts primary production, possibly through shading, of freshwater plankton communities (Carpenter et al., 1998)	NLA Lake Water Quality Data	8857 \pm 16846 (340, 290570)
Zooplankton taxa richness	Lowest taxonomic unit possible	Significant influence of consumer diversity on rate of herbivory of algal biomass (Naeem & Li, 1998)	NLA Lake Zooplankton Count Data	7 \pm 3 (0, 18)
Zooplankton abundance	(density/mL) [*]	Aquatic herbivores remove as much as 79% of algal primary production (Cyr & Pace, 1993)	NLA Lake Zooplankton Count Data	70 \pm 203 (0, 2602)
Water temperature	Degrees Celsius	Influences photosynthesis (Davinson, 1991) as well as lake turnover and stratification (Kalf, 2012)	NLA Lake Profile Data	21 \pm 5 (7, 34)
Conductivity	($\mu\text{S/cm at } 25\text{ }^\circ\text{C}$) [*]	Frequently noted environmental variable in freshwater algal studies; ions in solution may affect algal growth rates and uptake of phosphorus (Tilman et al.	NLA Lake Water Quality Data	662 \pm 2455 (4, 50590)

1982)

Water pH	pH units	Frequently noted environmental variable in freshwater algal studies; hypothesized to play a role in nutrient availability and uptake of algal species (Tilman et al. 1982)	NLA Lake Water Quality Data	8 ± 1 (4, 10)
Light	kWh/m ² /day	Important environmental variable in primary production; measured as monthly average solar radiation incident (kWh/m ² /day), which is a direct measure of the light available on the surface of the water.	NASA	5 ± 1 (3, 8)

Table 2: Summary of 18 SEMs run to determine influences of algal diversity and environmental variables on total algal community biovolume. Models arranged using model identification number. Best-fit model is the first listed. Models were selected based on significant X^2 values ($P > 0.05$) and lowest AIC value. The difference between the model i and the best-fit model (Δ_i) was calculated for each model. Using these values, the likelihood of a model, m_i , given the data, y , ($L[m_i | y]$) was calculated as $L[m_i | y] = \exp(-1/2 \Delta_i)$. The Akaike weight (w_i) was calculated using the ($L[m_i | y]$) calculations. Akaike weight provides a relative weight of evidence for each model, and can be interpreted as the probability of model i being the best-fit model of the candidate models, given the data. Akaike weight was calculated by normalizing model likelihood values ($L[m_i | y]$) across all models.

Model (m_i)	Model Description	df	X^2	P	AIC	Δ_i	$L(m_i y)$	w_i
15	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L, algal taxa richness})$ covariance: $\mu\text{g nitrogen/L -- } \mu\text{g phosphorus/L, } \mu\text{g nitrogen/L -- algal taxa richness}$	1.00	0.31	0.58	5952.45	0.00	1.00	0.74
16	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L, algal taxa richness})$ covariance: $\mu\text{g nitrogen/L -- } \mu\text{g phosphorus/L}$	2.00	4.35	0.11	5954.49	2.04	0.36	0.26
11	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L, algal taxa richness, growing period})$ covariance: $\mu\text{g nitrogen/L -- } \mu\text{g phosphorus/L, } \mu\text{g nitrogen/L -- algal taxa richness}$	4.00	42.98	0.00	12932.43	6979.99	0.00	0.00
12	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L, algal taxa richness, temperature})$ covariance: $\mu\text{g nitrogen/L -- } \mu\text{g phosphorus/L, } \mu\text{g nitrogen/L -- algal taxa richness}$	4.00	222.80	0.00	12953.72	7001.27	0.00	0.00
10	community biomass of algae = $f(\mu\text{g nitrogen/L, temperature, algal taxa richness, growing period})$ covariance: $\text{growing period -- temperature, } \mu\text{g nitrogen/L -- algal taxa richness}$	4.00	225.50	0.00	18501.55	12549.11	0.00	0.00

Model (m_i)	Model Description	df	X^2	P	AIC	Δ_i	L (m_i y)	w_i
14	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ growing period, temperature) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, growing period -- temperature	4.00	283.44	0.00	18845.74	12893.30	0.00	0.00
13	community biomass of algae = $f(\mu\text{g phosphorus/L, temperature,}$ algal taxa richness, growing period) covariance: growing period -- temperature	5.00	252.80	0.00	19369.95	13417.50	0.00	0.00
9	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ growing period, temperature, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature	7.00	297.91	0.00	19582.13	13629.69	0.00	0.00
8	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ light, growing period, temperature, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature	12.00	367.21	0.00	22349.97	16397.53	0.00	0.00
7	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ light, growing period, temperature, zooplankton abundance/mL, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature, zooplankton abundance/ mL -- algal taxa richness	17.00	481.13	0.00	24854.30	18901.86	0.00	0.00

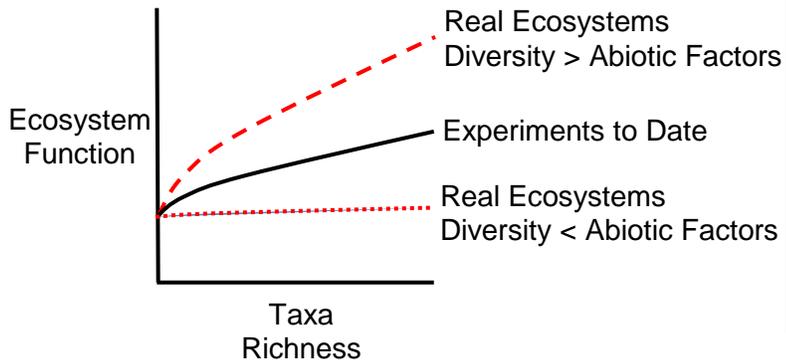
Model (m_i)	Model Description	df	X^2	P	AIC	Δ_i	L (m_i y)	w_i
6	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ light, growing period, temperature, zooplankton abundance/mL, zooplankton taxa richness, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature, zooplankton abundance/ mL -- algal taxa richness	24.00	732.33	0.00	30634.57	24682.12	0.00	0.00
5	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L,}$ conductivity, light, growing period, temperature, zooplankton abundance/ mL, zooplankton taxa richness, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature, zooplankton abundance/ mL -- algal taxa richness	32.00	1420.71	0.00	32612.12	26659.67	0.00	0.00
4	community biomass of algae = $f(\mu\text{g nitrogen/L, } \mu\text{g phosphorus/L, pH,}$ light, growing period, temperature, zooplankton abundance/mL, zooplankton taxa richness, algal taxa richness) covariance: $\mu\text{g nitrogen/L -- } \mu\text{g}$ phosphorus/L, $\mu\text{g nitrogen/L -- algal}$ taxa richness, growing period -- temperature, zooplankton abundance/mL -- algal taxa richness	32.00	1172.98	0.00	33235.18	27282.73	0.00	0.00

Model (m_i)	Model Description	df	X^2	P	AIC	Δ_i	L (m_i y)	w_i
3	community biomass of algae = f(μg nitrogen/L, μg phosphorus/L, pH, conductivity, light, growing period, temperature, zooplankton abundance/mL, zooplankton taxa richness, algal taxa richness) covariance: μg nitrogen/L -- μg phosphorus/L, μg nitrogen/L -- algal taxa richness, conductivity -- pH, growing period -- temperature, zooplankton abundance/mL -- algal taxa richness	40.00	1466.48	0.00	34307.38	28354.93	0.00	0.00
2	community biomass of algae = f(μg nitrogen/L, μg phosphorus/L, μg silica/L, pH, conductivity, light, growing period, temperature, zooplankton abundance/mL, zooplankton taxa richness, algal taxa richness) covariance: μg nitrogen/L -- μg phosphorus/L, μg nitrogen/L -- algal taxa richness, conductivity -- pH, growing period -- temperature, zooplankton abundance/mL -- algal taxa richness	50.00	1702.08	0.00	36297.63	30345.18	0.00	0.00
1	community biomass of algae = f(μg nitrogen/L, μg phosphorus/L, μg silica/L, μg DOC/L, pH, conductivity, light, growing period, temperature, zooplankton abundance/mL, zooplankton taxa richness, algal taxa richness) covariance: μg nitrogen/L -- μg phosphorus/L, μg nitrogen/L -- algal taxa richness, conductivity -- pH, growing period -- temperature, zooplankton abundance/mL -- algal taxa richness	61.00	2929.46	0.00	37244.75	31292.30	0.00	0.00

FIGURES

Figure 1: **A)** Positive, but quickly saturating curves of experiments to date (continuous black line), and possible outcomes of BEF in real ecosystems (dashed red lines). Upper dashed red line represents hypothesis 1, wherein the diversity is greater in nature than in prior experiments; lower dashed red line represents hypothesis 2, wherein the diversity signal is lower in nature than in prior experiments. **B)** Map of the Continental United States with lakes sampled by the USEPA noted as red circles; darker circles indicate multiple lakes sampled in a close proximity.

A)



B)



Figure 2: Full structural equation model including all variables and specified covariance terms. Solid lines with arrow head represent causal pathways. Dashed lines with dual arrow heads represent covariance terms. The model's chi-square was 2952.79, and the AIC was 36,874.602 ($P = 0.000$). Variance explained by the model was 0.209. Analyses were completed in Lavaan package in R. Growing season is measured by latitude (decimal degrees); though growing season is usually measured by degree days, latitude is highly correlated with degree days and could be log transformed to achieve normality. Asterisks indicate significant pathways.

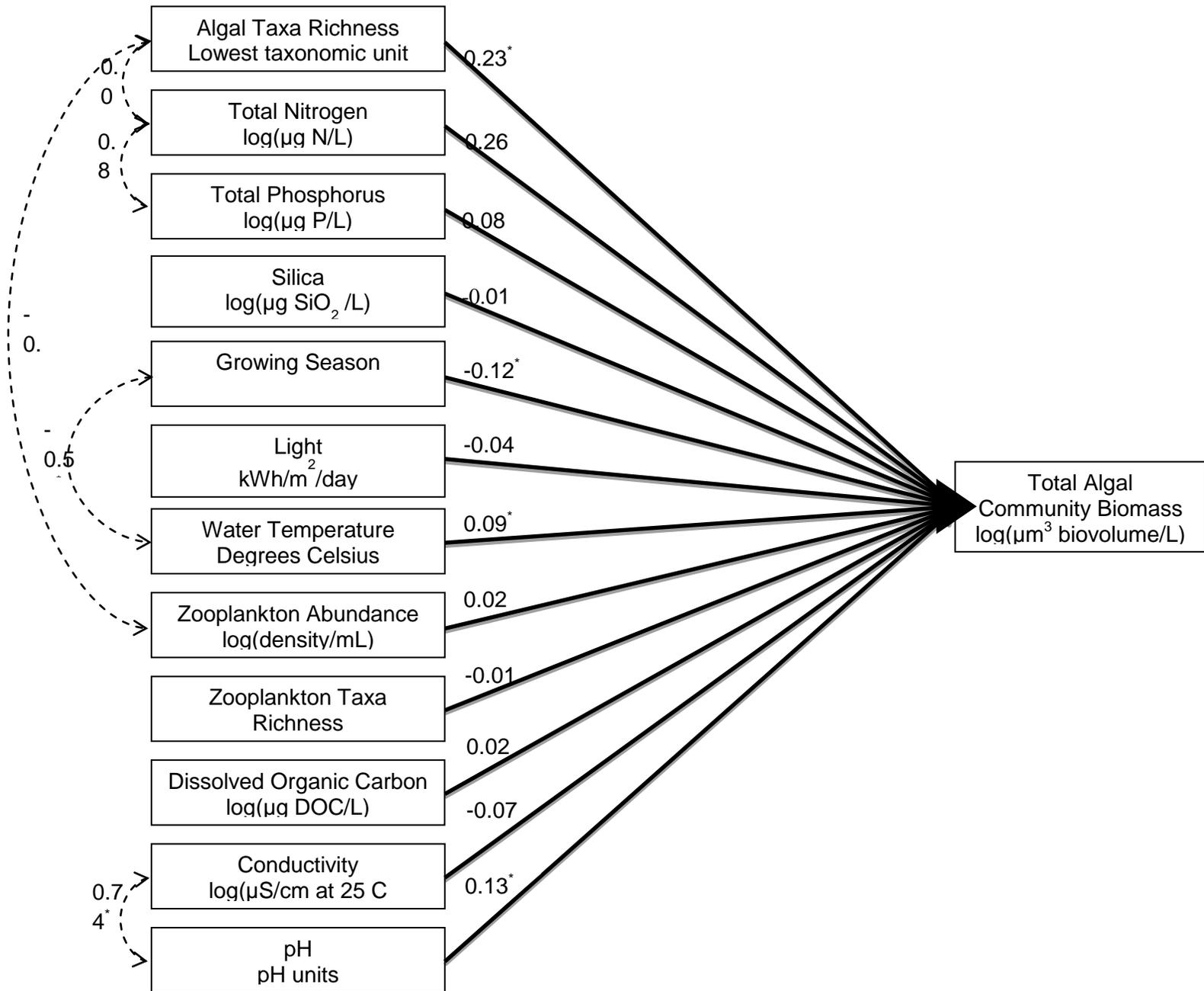


Figure 3: Best-fit structural equation model. Solid lines with arrow head represent causal pathways. Dashed lines with dual arrow heads represent covariance terms. Standardized partial regression coefficients are noted first. Unstandardized partial regression coefficients are noted second in parentheses. Significance is indicated by asterisks. The model's chi-square was 0.307, and the AIC was 5952.447 ($P = 0.580$). Variance explained by the model was 0.209. Analyses were completed in Lavaan package in R.

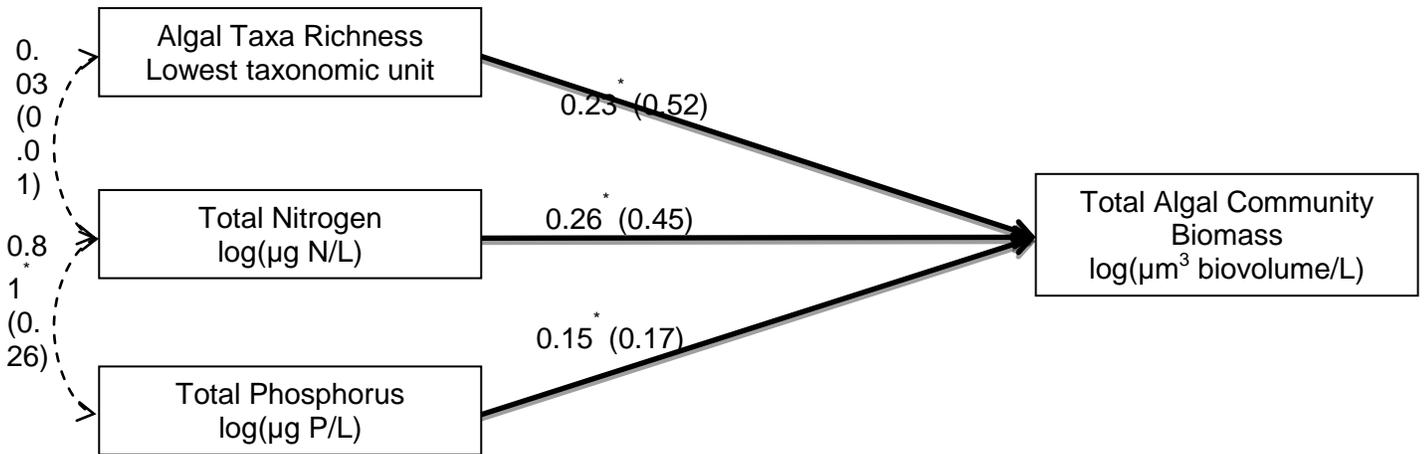
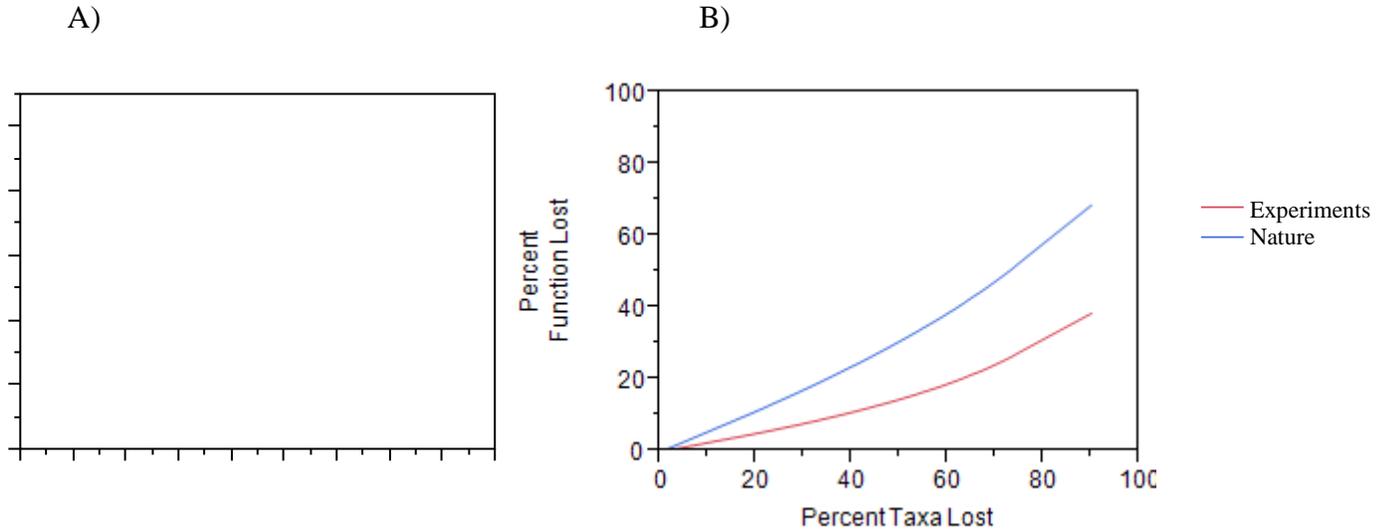


Figure 4: **A)** Comparison of the scaling coefficient relating biomass to diversity in natural lakes and experiments. Patterns calculated using power function $y=ax^b$, where y = standardized biomass $[B(p)/B(m)]$, $a=1$, x =taxa richness, and b =scaling coefficient. The scaling coefficient for experiments is 0.22 and was taken from Cardinale et al. (2011). The scaling coefficient for lakes is 0.52 and was taken directly from our best-fitting SEM. Taxa richness range is intentional; in experiments the maximum level of diversity is usually 32 species. In the dataset, the maximum number level of diversity is 85 species. **B)** Comparison of the magnitude of underestimation given scaling coefficient and known taxa richness maxima, using the power equation described above.



LITERATURE CITED

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**: 716-723.
- Balvanera, P., A.B. Pfisterer, N. Buchmann, J.S. He, T. Nakashizuka, D. Raffaelli, and B. Schmid. (2006). Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology Letters* **9**: 1146-1156.
- Cardinale, B.J., D.S. Srivastava, J.E. Duffy, J.P. Wright, A.L. Downing, M. Sankaran, and C. Jouseau. 2006. Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**: 989-992.
- Cardinale, B.J., J.P. Wright, M.W. Cadotte, I.T. Carroll, A. Hector, D.S. Srivastava, M. Loreau, and J.J. Weis. 2007. Impacts of plant diversity on biomass production increase through time due to complementary resource use: A meta-analysis. *Proceedings of the National Academy of Sciences*. **104**: 18123-18128.
- Cardinale, B.J. and K. Gross. 2007. Does species richness drive community production or vice versa? Reconciling historical and contemporary paradigms in competitive communities. *The American Naturalist* **170**: 207-220.
- Cardinale, B.J., E. Duffy, D. Srivastava, M. Loreau, M. Thomas, and M. Emmerson. 2009a. Towards a food web perspective on biodiversity and ecosystem functioning. Pages 105-120 *in* S. Naeem, D.E. Bunker, A. Hector, M. Loreau, and C. Perrings, editors. *Biodiversity, Ecosystem Functioning, and Human Wellbeing*. Oxford University Press, Oxford, UK.
- Cardinale, B.J., H. Hillebrand, W.S. Harpole, K. Gross, and R. Ptacnik. 2009b. Separating the influence of resource 'availability' from resources 'imbalance' on productivity-diversity relationships. *Ecology Letters* **12**: 475-487.
- Cardinale et al. 2009c, Does species diversity drive productivity or visa-versa? A test of the multivariate productivity diversity hypothesis in streams.
- Cardinale, B.J., K.L. Matulich, D.U. Hooper, J.E. Byrnes, E. Duffy, L. Gamfeldt, P. Balvanera, M.I. O'Connor, and A. Gonzalez. 2011. The functional role of producer diversity in ecosystems. *American Journal of Botany* **98**: 572-592.
- Carpenter, S.R., J.J. Cole, J.F. Kitchell, and M.L. Pace. 1998. Impact of dissolved organic carbon, phosphorus, and grazing phytoplankton biomass and production in experimental lakes. *Limnology and Oceanography* **43**: 73-80.
- Chase, J.M., and M.A. Leibold. 2002. Spatial scale dictates the productivity-biodiversity relationship. *Nature*. **416**: 427-430.
- Connell, J.H. & E. Orias. 1964. The ecological regulation of species diversity. *The American Naturalist* **98**: 399-414.

- Conner, E.F., and E.D. McCoy. 1979. The statistics and biology of the species-area relationship. *The American Naturalist* **113**: 791-833.
- Cyr, H. and M.L. Pace. 1993. Magnitude and patterns of herbivory in aquatic and terrestrial ecosystems. *Nature* **361**: 148-150.
- Dimitrakopoulos, P.G., and B. Schmid. 2004. Biodiversity effects increase linearly with biotope space. *Ecology Letters* **7**: 574-583.
- Dodson, S.I., S.E. Arnott, and K.L. Cottingham. 2000. The relationship in lake communities between primary productivity and species richness. *Ecology* **81**: 2662-2679.
- Duffy, E.J. 2009. Why biodiversity is important to the functioning of real-world ecosystems. *Frontiers in Ecology and the Environment* **7**: 437-444.
- Elsner, J.J., E.R. Marzolf, and C.R. Goldman. 1990. Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: A review and critique of experimental enrichments. *Canadian Journal of Fisheries and Aquatic Science* **47**: 1468-1477.
- Evans, M.A., G. Fahnenstiel, and D. Scavia. 2011. Incidental oligotrophication of North American Great Lakes. *Environmental Science and Technology* **45**: 3297-3303.
- Feminella, J.W., and C.P. Hawkins. 1995. Interactions between stream herbivores and periphyton: A quantitative analysis of past experiments. *Journal of the North American Benthological Society* **14**: 465-509.
- Fridley, J.D. 2002. Resource availability dominates and alters the relationship between species diversity and ecosystem productivity in experimental plant communities. *Oecologia* **132**: 271-277.
- Grace, J.B. 2006. *Structural Equation Modeling and Natural Systems*. Cambridge University Press, Cambridge, United Kingdom.
- Grace, J.B., T.M. Anderson, M.D. Smith, E. Seabloom, S.J. Andelman, G. Meche, E. Weiher, L.K. Allain, H. Jutila, M. Sankaran, J. Knops, M. Ritchie, and M.R. Willig. 2007. Does species diversity limit productivity in natural grassland communities. *Ecology Letters* **10**: 680-689.
- Griffin, J.N., and B.J. Cardinale. 2013. Predator richness and prey suppression: meta-analysis reveals importance of scale and phylogenetic diversity. *Journal of Animal Ecology*, in review.
- Harner, F.R., and K.T. Harper. 1976. The role of area, heterogeneity, and favorability in plant species diversity of Pinyon-Juniper ecosystems. *Ecology* **57**: 1254-1263.
- Harpole, W.S., J.T. Ngai, E.E. Cleland, E.W. Seabloom, E.T. Borer, M.E.S. Bracken, J.J. Elser, D.S. Gruner, H. Hillebrand, J.B. Shurin, and J.E. Smith. 2011. Nutrient co-limitation of primary producer communities. *Ecology Letters* **14**: 852-862.

- Hill, W.R., and A.W. Knight. 1988. Nutrient and light limitations of algae in two northern California streams. *Journal of Phycology* **24**: 125-132.
- Hillebrand, H., and B. Matthiessen. 2009. Biodiversity in a complex world: consolidation and progress in functional biodiversity research. *Ecology Letters* **12**: 1405-1419.
- Johnson, J.B. and K.S. Omland. 2004. Model selection in ecology and evolution. *TRENDS in Ecology and Evolution* **19**: 101-106.
- Kalff, J. 2002. *Limnology*. Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.
- Lomolino, M.V. 2001. Ecology's most general, yet protean pattern: the species-area relationship. *Journal of biogeography* **27**: 17-26.
- Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, M.A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D.A. Wardle. 2001. Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**: 804-808.
- Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, M.A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D.A. Wardle. 2001. Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**: 804-808.
- Loreau, M. 2010. Linking biodiversity and ecosystems: towards a unifying ecological theory. *Philosophical Transactions of the Royal B-Biological Sciences* **365**: 49-60.
- MacArthur, R.H. 1965. Patterns of species diversity. *Biological Reviews of the Cambridge Philosophical Society* **40**: 510-533.
- Martin-Jezequel, V., M. Hildebrand, and M.A. Brzezinski. 2000. Silicon metabolism in diatoms: implications for growth. *Journal of Phycology* **36**: 821-840.
- MCGuinness, K.A. 1984. Equations and explanations in the study of species-area curves. *Biological Reviews* **59**: 423-440.
- Mittelbach, G.G., C.F. Steiner, S.M. Scheiner, K.L. Gross, H.L. Reynolds, R.B. Waide, M.R. Willig, S.I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? *Ecology* **82**: 2381-2396.
- Mora, C., O. Aburto-Oropeza, A. A. Bocos, P. M. Ayotte, S. Banks, A. G. Bauman, M. Beger, S. Bessudo, D. J. Booth, E. Brokovich, A. Brooks, P. Chabanet, J. E. Cinner, J. Cortes, J. J. Cruz-Motta, A. C. Magana, E. E. DeMartini, G. J. Edgar, D. A. Feary, S. C. A. Ferse, A. M. Friedlander, K. J. Gaston, C. Gough, N. A. J. Graham, A. Green, H. Guzman, M. Hardt, M. Kulbicki, Y. Letourneur, A. L. Perez, M. Loreau, Y. Loya, C. Martinez, I. Mascarenas-Osorio, T. Morove, M. O. Nadon, Y. Nakamura, G. Paredes, N. V. C. Polunin, M. S. Pratchett, H. R. Bonilla, F. Rivera, E. Sala, S. A. Sandin, G. Soler, R. Stuart-Smith, E. Tessier, D. P. Tittensor,

- M. Tupper, P. Usseglio, L. Vigliola, L. Wantiez, I. Williams, S. K. Wilson, and F. A. Zapata. 2011. Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *Plos Biology* **9**.
- Naeem, S. and S. Li. 1998. Consumer species richness and autotrophic biomass. *Ecology* **79**: 2603-2615.
- Naeem, S. 2002. Ecosystem consequences of biodiversity loss: The evolution of a paradigm. *Ecology* **83**: 1537-1552.
- Naeem, S., M. Loreau, and P. Inchausti. 2002. Biodiversity and ecosystem functioning: the emergence of a synthetic ecological framework. Pages 3-17 *in* M. Loreau, S. Naeem, and P. Inchausti, editors. *Biodiversity and Ecosystem Functioning Synthesis and Perspectives*, Oxford University Press, Oxford, UK.
- Paquette, A., and C. Messier. 2011. The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecology and Biogeography* **20**: 170-180.
- Power, L.D., and B.J. Cardinale. 2009. Species richness enhances both algal biomass and rates of oxygen production in aquatic microcosms. *Oikos* **118**: 1703-1711.
- Reich, P.B., D. Tilman, F. Isbell, K. Mueller, S.E. Hobbie, D.F.B. Flynn, and N. Eisenhauer. 2012. Impacts of biodiversity loss escalate through time as redundancy fades. *Science* **336**: 589-592.
- Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* **48**: 1-36.
- Schmid, B. 2002. The species richness-productivity controversy. *Trends in Ecology and Evolution* **17**: 113-114.
- Schmid, B., P. Balvanera, B.J. Cardinale, J. Godbold, A.B. Pfisterer, D. Raffaelli, M. Solan, and D.S. Srivastava. 2009. Consequences of species loss for ecosystem functioning: meta-analyses of data from biodiversity experiments. Pages 14-29 *in* S. Naeem, D.E. Bunker, A. Hector, M. Loreau, and C. Perrings, editors. *Biodiversity, Ecosystem Functioning, and Human Wellbeing*. Oxford University Press, Oxford, UK.
- Stachowicz, J.J., M. Graham, M.E.S. Bracken, and A.I. Szoboszlai. 2008. Diversity enhances cover and stability of seaweed assemblages: The role of heterogeneity and time. *Ecology* **89**: 3008-3019.
- Tilman, D., S.S. Kilham, and P. Kilham. 1982. Phytoplankton community ecology: The role of limiting nutrients. *Annual Review of Ecology and Systematics* **13**: 349-372.
- Tilman, D. 1999. The ecological consequences of changes in biodiversity: A search for general principles. *Ecology* **80**: 1455-1474.

Tylianakas, J.M., T.A. Rand, A. Kahmen, A. Klein, N. Buchmann, J. Perner, and T. Tschardtke. 2009. Resource Heterogeneity Moderates the Biodiversity-Function Relationship in Real World Ecosystems. *PLoS Biology* **6**: 947-956.

Wardle, D.A., and O. Zackrisson. 2005. Effects of species and functional group loss on island ecosystem properties. *Nature* **435**: 806-810.

Weis, J.J., Cardinale, B.J., Forshay, K.J., and A.R. Ives. 2007. Effects of species diversity on community biomass production change over the course of succession. *Ecology* **88**: 929-939.

Weyhenmeyer, G.A., H. Peter, and E. Willén. 2013. Shifts in phytoplankton species richness and biomass along a latitudinal gradient – consequences for relationships between biodiversity and ecosystem functioning. *Freshwater Biology* **58**: 612-623.

Williams, C.B. 1964. Patterns in the balance of nature and related problems in quantitative ecology. Academic Press, New York, New York, United States of America.

Wright, D.H. 1983. Species-energy theory – an extension of species area theory. *Oikos* **41**: 496-506.

SUPPLEMENTAL TEXT

Relationships to Lake Area

I chose to limit my selection of environmental variables to those that have been shown to influence biomass production via a direct, known biological mechanism. As such, I did not include area in my analyses, since area is only thought to influence algal richness and biomass indirectly via other factors that covary with area. Nonetheless, I recognize that others may feel that area is an important explanatory variable that influences the diversity productivity relationship (Dodson et al., 2000, Chase & Leibold, 2002). I therefore explored the possibility that the productivity--diversity relationship was area-dependent using linear regression. Figure 6 shows there was no significant relationship between lake surface area and total algal community biomass (Fig.6).

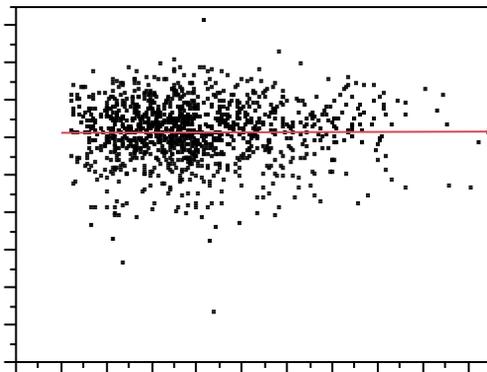


Figure 6: Simple linear regression demonstrating the influence of area (hectares) on biovolume ($\mu\text{m}^3/\text{L}$) ($p=0.81$); both variables were log-transformed to achieve normality.

Furthermore, I found no significant relationship between lake surface area and algal species richness (Fig. 7).

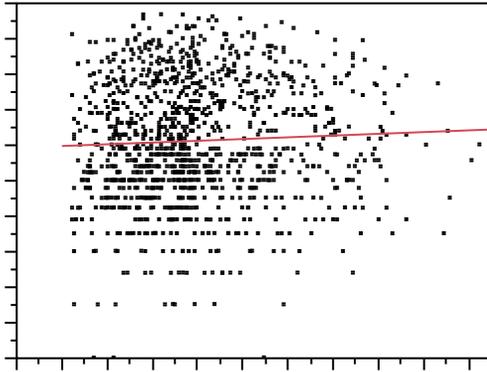


Figure 7: Simple linear regression demonstrating the influence of area (hectares) on algal taxa richness ($p=0.11$); both variables were log-transformed to achieve normality.

Given the lack of relationship between area and primary production and area and species richness, coupled with the lack of biological meaning of the influence of area on primary production, I found no evidence to suggest area should be included in the models.

Nutrients

Multiple theories have postulated how resource supply relates to diversity and productivity in ecosystems (Cardinale et al., 2009). The species energy theory (SET) focuses on how the summation of all available resources, such as nitrogen and phosphorus, may limit taxa population sizes (Wright, 1983). As more energy is available in a local community, population size of the abundant and rare local taxa increases, and subsequently, there is a reduction in the probability of stochastic extinction (Wright, 1983). SET has been widely used to explain the monotonically increasing and the increasing portion of the unimodal relationships between diversity and productivity (Cardinale et al., 2009). In our dataset, we chose to account for nitrogen and phosphorus based on their total amounts available in accordance with the SET.

In addition to SET, resource ratio theory (RRT) can be used to understand the relationship between taxa diversity and productivity. Conversely to SET, RRT is commonly used to explain the concave-down relationship between diversity and productivity (Cardinale et al., 2009). RRT is based on the principle that taxa have intrinsic trade-offs in the capture and utilization of required resources (nitrogen vs. phosphorus); changes in the ratio of required resources drive competition and coexistence within a community, and will ultimately influence diversity. Therefore, I felt it necessary to account for the possible influence of nutrient ratios, specifically of N:P.

Given SET and RRT, I felt it pertinent to examine the relationships between nitrogen, phosphorus, taxa richness, and productivity more carefully. I ran simple linear regressions to assess concavity, and found that total nitrogen, total phosphorus, and N:P ratio had a significant linear relationship with productivity (Fig. 8).

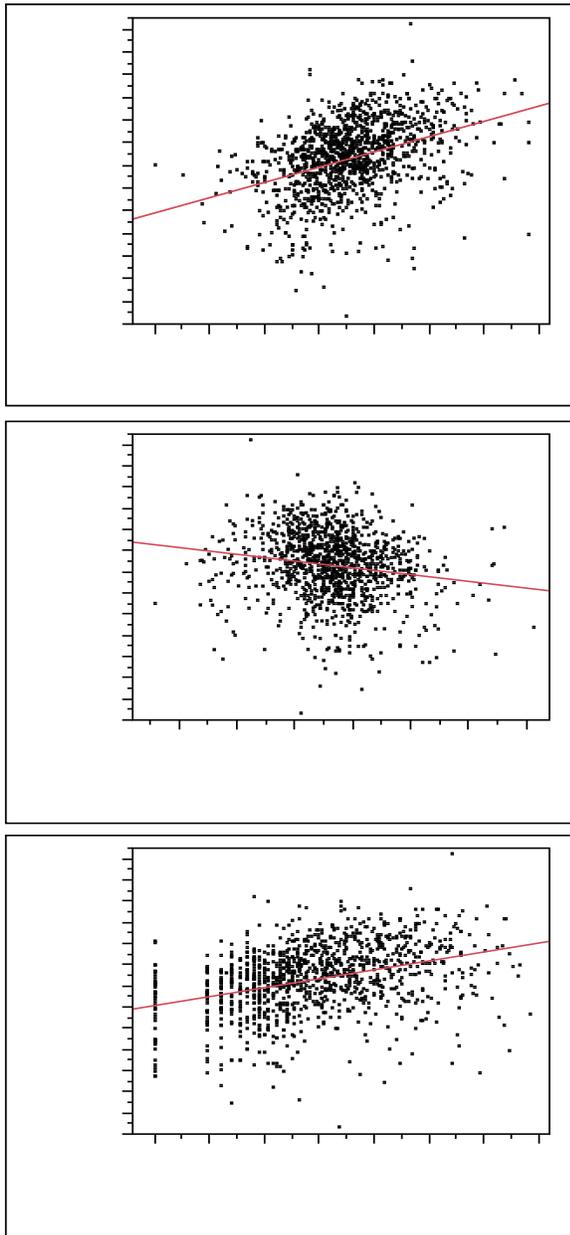


Figure 8: Simple linear regressions demonstrating the influence of total nitrogen ($p < 0.001$), total phosphorus ($p < 0.001$), and N:P ratio on total algal community biovolume ($p < 0.001$). All variables were log-transformed to achieve normality.

I also added a N:P ratio into our best-fitting SEM to assess the role of RRT in this study (Fig. 9). The partial regression coefficient describing the relationship between the N:P ratio and total algal community biomass was not significant ($P > 0.05$). The SEM model was not significant ($X^2 = 32987.673$, $P = 0.000$).

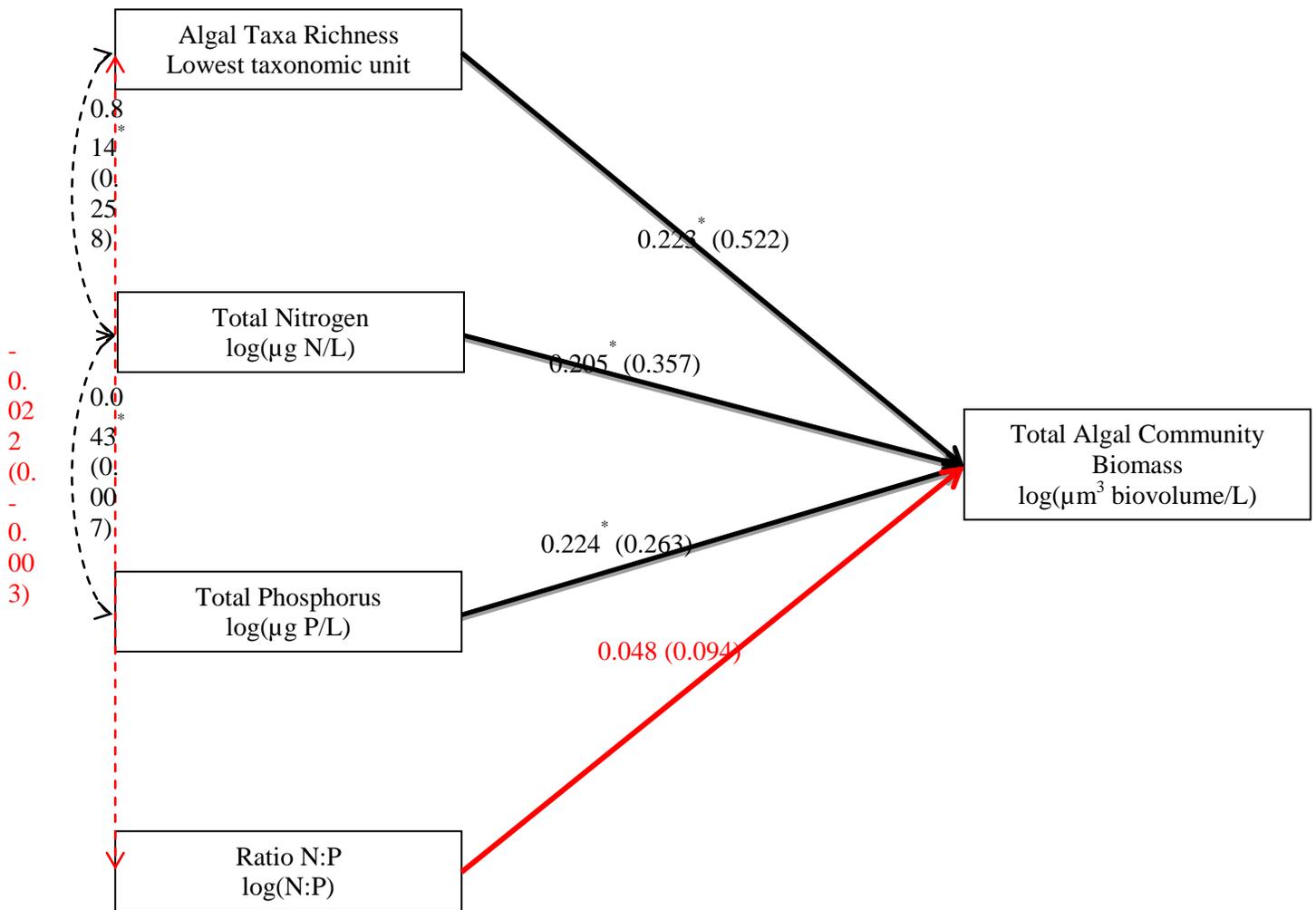


Figure 9: Best-fitting structural equation model with modification for N:P ratio. Solid lines with arrow head represent causal pathways. Dashed lines with dual arrow heads represent covariance terms. Red indicates paths that were added to include the N:P ratio. Standardized partial regression coefficients are noted first. Unstandardized partial regression coefficients are noted second in parentheses. Significance is indicated by asterisks. The model's chi-square was 32987.673 ($P = 0.000$). Analyses were completed in Lavaan package in R.

Direction of causal relationship between diversity and community biomass

The question of how taxa diversity relates to the productivity of ecosystems has been rigorously pursued in ecology since the time of Darwin. Historically, differences in biodiversity have been viewed as the consequence of changes in productivity, wherein biodiversity is driven primarily from differences in resource availability (Mittelbach et al., 2001). But, over the past two decades, there has been a fundamental shift in the approach to the relationship between diversity and productivity, and numerous studies have focused on how differences in biodiversity may be driving, as opposed to a consequence of, changes in productivity (Cardinale et al. 2011). Important steps in understanding how diversity relates to productivity have been hypothesized and mathematically modeled (see Loreau et al. 2001; Schmid, 2002; Cardinale & Gross, 2007). Mathematical models developed by Cardinale & Gross demonstrated that resource availability can drive differences in biodiversity, which can drive resource use and efficiency and biomass production (Cardinale & Gross, 2007). As a result of the volume of experimental, theoretical, and mathematical evidence demonstrating differences in biodiversity cause, rather than a consequence of, changes in productivity, I chose to model productivity as the dependent variable and diversity as an independent variable.

As an exercise, I reversed the causal relationship between diversity and productivity in our best-fitting SEM, while retaining all other relationships (Fig.10). The resulting SEM is not significant ($X^2 = 6.752, P < 0.01$).

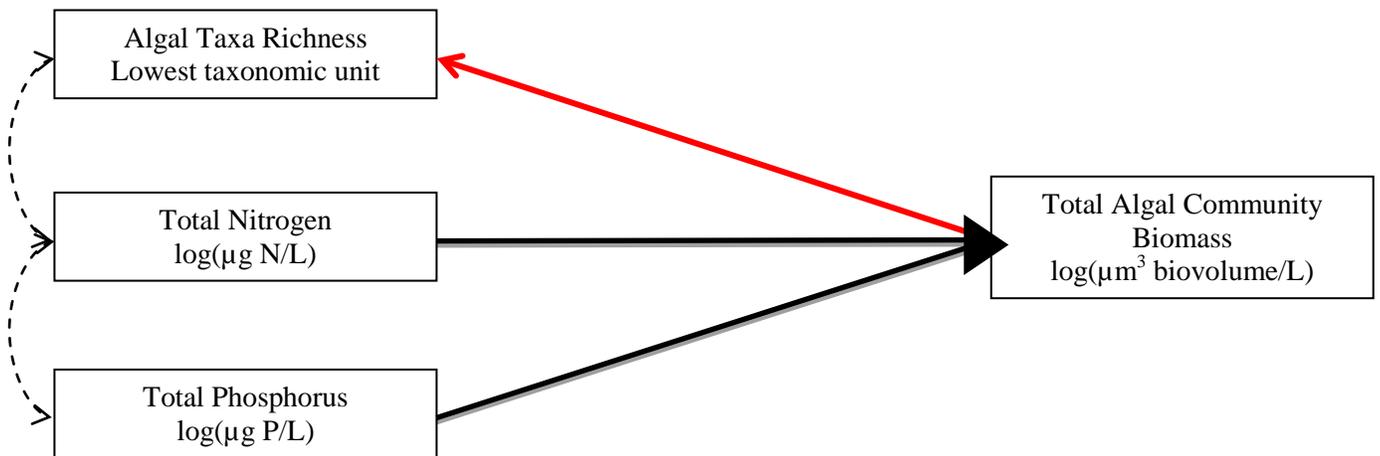


Figure 10: SEM identical to our best-fitting SEM, with one modification: the causal relationship between algal taxa richness and total algal community biomass has been reversed. Solid lines with arrow head represent causal pathways. Dashed lines with dual arrow heads represent covariance terms. The model's chi-square was 6.752 ($P = 0.009$). Analyses were completed in Lavaan package in R.