COMMONNESS AND RARITY: EXPLAINING PATTERNS OF ABUNDANCE

AND OCCURRENCE IN DIAPTOMID COPEPODS, AND THEIR ROLE IN LAKE

FOOD WEBS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By Rachel Lynn Wilkins

August 2022

COMMONNESS AND RARITY: EXPLAINING PATTERNS OF ABUNDANCE AND OCCURRENCE IN DIAPTOMID COPEPODS, AND THEIR ROLE IN LAKE FOOD WEBS

Rachel Lynn Wilkins, Ph. D.

Cornell University 2022

Describing species distributions patterns and understanding the mechanisms responsible for producing those patterns is a fundamental aspect of ecology. The positive relationship between local abundance and regional occupancy of an assemblage of species is among the most ubiquitous patterns in nature, however the mechanisms that produce such a pattern are varied. We conducted two meta-analyses of the correlation-coefficients and slopes of abundance-occupancy relationships (AORs) to explain the mechanisms that govern species distributions in aquatic systems. We found AORs are stronger among habitats and organisms with elevated dispersal potential and wider niche breadths. We found consistent slopes of AORs among all aquatic systems, suggesting abundance and occupancy are similarly coupled. Exploring the differences between species can help us understand the reasons species fall along specific regions of AORs. To understand the drivers of the disparity in occurrence between two species, I studied the difference in environmental condition, diet, and population genetic structure, of two diaptomid copepod species, one common and one rare. Differences in occurrence between these two species is not due to a difference in diet, but may be due to different levels of dispersal, and local

adaptation to a wide range of environmental conditions. Variation in diet based on lake conditions suggests diaptomid copepods play a complex role in lake food webs. To understand how diaptomid copepods diets vary according to lake environmental conditions, we measured trophic position in five species of diaptomid copepod from 43 lakes. Copepods had a variable and predominantly predatory diet, that changed given the trophic state or size of the lake. For diaptomid copepods, responding to various environmental conditions through dispersal and diet provide insight into the mechanisms underlying distribution patterns, and the fundamental drivers of patterns of abundance and occupancy.

BIOGRAPHICAL SKETCH

Rachel Lynn Wilkins received her B.S. from the State University of New York College of Environmental Science and Forestry in 2010. She worked with Dr. Kimberly Schulz studying vernal pool ecology and the role of poly-unsaturated-fatty-acids in zooplankton diets. Rachel then managed the laboratories of Dr. Jonathan Shurin and Dr. Elsa Cleland at the University of California San Diego. With Dr. Shurin she studied applications of ecological diversity to enhance algae-biofuel production, and with Dr. Cleland she studied the effect of climate change on coastal sage scrub plant communities. Rachel began her dissertation research with Dr. Nelson Hairston, Jr. in the Department of Ecology & Evolutionary Biology at Cornell University in 2013.

DEDICATION

Dedicated to my family, especially my mom and my husband, for encouraging and

supporting me through it all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: THE REMARKABLE CONSISTENCY OF THE ECOLOGICAL

ABUNDANCE-OCCUPANCY RELATIONSHIP ACROSS AQUATIC SYSTEMS:

A META-ANALYSIS

Rachel L. Wilkins[1,2] & Nelson G. Hairston, Jr.[1]

1. Department of Ecology & Evolutionary Biology, Cornell University

2. eCornell, Office of External Education, Cornell University

Abstract

The positive relationship between local abundance and regional occurrence of an

assemblage of species is among the most ubiquitous patterns in nature. Species with

high local population abundance are generally common across a landscape, while

species with low local abundance tend to occur in fewer locations. Because of the

broad consistency of this abundance-occupancy relationship (AOR) a large number of

hypotheses have been put forward as explanations. Our objective here is to take

advantage of the wide variety of aquatic systems, representing a diversity of

environmental conditions, dominant taxa, and connectedness among locations, to

assess first the consistency of AORs and second to evaluate the viability of key

explanatory hypotheses. We performed two meta-analyses of AORs in published

studies of aquatic systems including marine open ocean, reef and benthos, intertidal

and estuary, and freshwater rivers, streams, lakes and ponds. We collected 247 AOR

correlation-coefficients from 108 published studies, and the linear slope terms (β-

coefficients) for 71 AORs from 58 published studies. There was a high level of

heterogeneity in correlation coefficients among AORs, most of it unexplained in our

statistical models, however, the patterns that we did find are most consistent the hypotheses that AORs are made stronger by elevated dispersal potential and wider niche breadths, and do not support the suggestion that they are statistical or sampling artifacts. At the same time, we found exceptionally high consistency and very little unexplained heterogeneity in the slopes of AORs among all aquatic systems, suggesting that abundance and occupancy are similarly coupled and underpinned by one or more common ecological processes.

Introduction

One of the most common definitions of ecology is that it is the study of the distribution and abundance of organisms in nature (Krebs 2009, Bowman et al. 2017). Universal patterns in distribution and abundance would imply that there are one or more fundamental underlying mechanisms that determine general patterns governing population sizes and patterns of distribution across a landscape. Examples include the species-area relationship (de Candolle 1855, Arrhenius 1921, Gleason 1922), Rapoport's rule (Rapoport 1982, Stevens 1989), and rank-abundance relationships (Fisher et al. 1943, Hairston 1959, Whittaker 1965). Among these broad patterns in nature is the nearly ubiquitous relationship between the average population size and level of spatial occurrence among assemblages of species, termed the abundance-occupancy relationship (Hanski 1982, Gaston *et al*. 2000, Blackburn *et al*. 2006). A positive interspecific relationship between local abundance and geographic occurrence was first pointed out by Charles Darwin (1859) and first formally described by John C. Willis (1922) who found that in an assemblage of species, those

with large populations tend also to occur in more locations across a specified area while species with small population sizes are typically found in fewer locations. This pattern is commonly referred to as an abundance-occupancy relationship (Box 1). A Positive abundance-occupancy relationship (hereafter, AOR) has been documented globally in taxonomic groups from bacteria to birds across a variety of habitats (Blackburn *et al*. 2006). The ubiquitous positive relationship between local abundance and regional occupancy has even been suggested as a general "rule in nature" (Hanski 1982), and a diversity of hypotheses have been proposed as explanations from mechanistic ecological processes to statistical artifact or publication bias. Although positive AORs are the general rule, the correlation between abundance and occupancy can be highly heterogeneous, varying by habitat, biogeographic region, and taxonomic group (Blackburn *et al*. 2006). It has been difficult to explain the residual variation in the relationship between abundance and occupancy, and Holt & Gaston (2003) suggested that AORs may very well be system specific. However, whereas significant AOR heterogeneity has been found among assemblages in terrestrial environments, and in assemblages of parasites, Blackburn *et al*. (2006) suggested that assemblages in aquatic systems not to be significantly heterogeneous, suggesting aquatic habitats may have a single common abundance-occupancy relationship. It may be that aquatic environments differ from terrestrial ones in their degree of isolation or connectivity, diversity of sizes, and, for streams and coral reefs, linearity. Birds and insects from terrestrial environments have in the past represented the majority of abundance-occupancy literature (Blackburn *et al*. 2006). Aquatic habitats such as lakes, streams and ocean gyres are fundamentally different from terrestrial

environments. They often operate as "islands" with wide uninhabitable expanses between them, preventing species dispersal among sites, although aquatic organisms from microbes to fish can vary greatly in potential to disperse.

Studies of AORs in aquatic systems have become common during the past two decades, providing a foundation for a meta-analysis to synthesize this literature and use emergent patterns to assess how ubiquitous AORs truly are, and to evaluate existing hypotheses for what produces them. The characteristics of aquatic systems such as variation in the presence or absence of discrete boundaries and accompanying diversity of degrees isolation, and variation in the dispersal potential of resident species, provide an opportunity to address some of the dominant hypotheses for why we so often observe AORs.

Least interesting is the possibility that AORs may simply be an artefact resulting from the way the data were collected or reported. If locally rare species are less likely to be detected than locally abundant species, artificially low occupancy levels may be reported for these species (McArdle *et al*. 1990). Alternatively, artefactual AORs may result when sampling occurs at the edge of a species' geographic range and if there is a gradient of abundance from the center to the edge where abundance and occupancy are both lower that at the center (Wright 1991; Hartley 1998). Artificially strong AORs could also result when regression analyses do not correct for phylogenetic non-independence so that the degrees of freedom are inflated when closely related species share abundance and distribution patterns (Harvey & Pagel 1991). Lastly, in a meta-analysis context, a reporting bias could lead to an incorrect conclusion that strong AORs are common if studies finding only a

weak relationship, or none at all, are not reported in publication.

Ecologically interesting hypotheses include attribution of AORs to variation in niche breadth, niche position, meta-population dynamics. While there are in fact a total of seven mechanistic ecological explanations for AORs (Borregaard & Rahbek 2010), though several are related, so the number depends on how mechanisms are counted. Here we explore three that are particularly germane to aquatic systems. According to the niche breadth hypothesis (Brown 1984), generalist species that have wide niches should have both higher population densities and more extensive distributions than specialist species with narrower niches which support fewer individuals and occur in fewer places. Alternatively, because the niche position of a species is determined by the prevalence across the landscape of the resources or habitat conditions it requires, when those widespread and abundant across the landscape, the species will itself be widespread and abundant regardless of niche breadth (Hanski *et al.* 1993; Venier & Fahrig 1996).

Third, spatial scale might affect the strength of the AOR: that is, different mechanisms can operate across different scales (Gaston & Lawton 1990; Borregaard & Rahbek 2010). Meta-population dynamics or density dependent habitat selection may be responsible for smaller scale abundance-occupancy relationships that encompass interacting populations. Brown (1984) folded his niche breadth explanation in with a meta-population perspective noting that if a species does well in some environmental conditions where there are enough resources to support a high density of individuals, then it should also be able to occur in other places as well. In a meta-population context, the dispersal ability of individual species could play an important

5

role in defining AORs since organisms that disperse more readily and further could have higher occupancy all else being equal (Hanski *et al*. 1993). Other mechanisms such density-independent factors related to the niche or population growth rate measures may operate at larger scales (Shepherd & Litvak 2004). For example, in a study of stream diatoms across the United States, species traits and population densities were most closely related to level of occupancy, while geographic range size was better predicted by niche breadth (Passy 2012). Weak or no AORs have been observed at larger scales for stream diatoms but become stronger as the scale of study goes from global distribution, to provincial, to regional occupancy (Riis & Sand-Jensen 2002). This has also been shown for stream bryophytes as well as aquatic angiosperms (Heino & Virtanen 2006, Soininen & Heino 2005).

We synthesize here the extensive and growing abundance-occupancy literature on aquatic systems and use this to evaluate proposed mechanisms underlying the abundance-occupancy relationship, including critiques of AORs as statistical, sampling or publication-bias artefacts, and mechanistic ecological hypotheses positing the nature of its slope and variation. We used a meta-analytical approach to study AORs from a systematic literature search spanning 120 years of publications, representing 247 abundance-occupancy relationships obtained from 108 studies. Our study explores the source of variation, or heterogeneity, observed in aquatic AORs, and the extent to which those relationships vary – or do not vary – among different biogeographic regions (Afrotropics, Nearctic, Neotropics and Palearctic), aquatic realms (marine, intertidal, freshwater), habitats within realm (e.g., freshwater streams, lakes; intertidal – rocky and estuary; marine benthos, coral reef, open ocean),

taxonomic groups (micro-organism, invertebrate, vertebrate) and sampling design. We take advantage of differences in habitat connectivity and organism dispersal potential, presumed differences in niche breadth among different kinds of taxa, and differences in niche similarity among habitats for different aquatic systems and their resident taxa to evaluate whether they support the key hypotheses, raised above, for the existence of AORs. We found positive abundance-occupancy relationships across all aquatic systems, with strikingly little difference in the magnitude of the slope among studies. At the same time, we noted marked heterogeneity in the strength of AORs, as measured by correlation coefficient, provides scope to evaluate different hypotheses. We find, for example the strongest AORs for taxa in intertidal systems and vertebrates (primarily fish) across all aquatic systems for which dispersal potential is high.

Methods

Literature search and data extraction

We searched five databases to retrieve as many titles in the aquatic sciences as possible with AOR data. On February 12-13, 2020 we performed searches in BIOSIS previews, Fish, Fisheries & Aquatic Biodiversity Worldwide (years 1933-2019), SCOPUS (years 1911-2020), Web of Science (years 1900-2020), and Wildlife & Ecology Studies Worldwide (years 1911-2020). Search terms were designed to capture studies that reported the relationship or correlation between abundance and occupancy of organisms in aquatic systems: (Relationship) AND (Abundance) AND (Occupancy) and (Aquatic characteristics) AND (Aquatic organisms) (see Table S1 for a complete list of search terms).

The search retrieved 31,141 total studies. We added seven additional studies that we knew of but that were not detected by our search terms. After removal of duplicates and studies that did not encompass aquatic systems, we screened the full text of 699 publications for eligibility, and excluded 591 that did not fit our criteria, leaving a total of 108 studies included in the meta-analysis (Figure 1). A study was included in the analysis if it met our inclusion criteria: 1) a natural biological community 2) free-living (non-parasitic) organisms that spend all or part of their life cycle partially or fully submerged in aquatic systems. To include as many aquatic studies as possible for this meta-analysis and to assess publication bias, we included studies that did not explicitly report an AOR but included raw data or summarized data on both abundance and occupancy from an assemblage of species from which we calculated an AOR. We excluded studies that measured occupancy as the extent of a species total geographic range size, as we felt this question tangential, albeit related, to AOR for this meta-analysis (Borregaard & Rahbek 2010). Studies that used fossil remains to quantify abundance or occupancy were not included due to the incomplete nature of fossil record. Similarly, studies that quantified abundance or occupancy using eDNA were excluded since methods are still being developed to measure abundance of individuals with precision (Ruppert *et al*. 2021).

In total, 247 abundance-occupancy relationships were extracted from 108 studies. Of these 108 studies, 45 (encompassing 165 separate abundance-occupancy relationships) explicitly examine abundance-occupancy relationships and presented a relationship between the two variables. We calculated the correlation coefficient ourselves for the 63 studies (82 abundance-occupancy relationships) that presented

8

both abundance and occupancy data but did not report an AOR. Where data were available, we extracted β-coefficients of linear AOR regression models, obtaining a total of 74 β-coefficients for our meta-analysis. Of these, 66 regressions were calculated by us using raw data, while five abundance-occupancy β-coefficients were reported by the author(s) of the studies. We are interested in both the correlation coefficients and slopes of AORs because correlation coefficients provide a measure of the reliability or scatter of data points around the regression, the slope indicates rate at which abundance changes as a function of occupancy. Interestingly correlation coefficients were more frequently reported in the literature than were slopes, which were mostly collected from papers that reported both abundance and occupancy, but did not calculate the relationship between the two.

For studies where a regression was not performed and original data were not provided, we extracted the raw data from figures or tables with DigitizeIt V 2.4.1 (Bormann 2020) and calculated correlation and β-coefficients. Transformation of abundance and occupancy varied across studies, however when we performed regressions on data extracted from figures or tables, we logit transformed proportion of sites occupied to normalize the data and meet statistical assumptions for linear regression analysis using $(\log(\frac{p}{1-p})$, where p is proportion of sites occupied) and $\log_{10}$ transformed mean abundance (Hanski and Gyllenberg 1997, Williamson and Gaston 1999). We transformed data from studies that plotted abundance-occupancy data but did not report regression statistics in the way the data were presented by the authors. For example, if the data were plotted on a log-log scale in a figure, but no correlation coefficient was calculated, we did a log-log transformation when

9

performing a linear regression to calculate the slope and correlation coefficient. When

necessary. We replaced logit proportion of sites occupied of 1.0 with 0.9999.

For each study we coded the realm as marine, intertidal or freshwater and habitat

within realms – the same categories used by Blackburn *et al*. (2006), taxonomic group

(micro-organisms, invertebrates and vertebrates, biogeographic region (Olson *et al*.

2001) (Table S2A), midpoint absolute latitude of the study area, the size of the study

area, number of sites surveyed, number of species or taxonomic units, the timespan

over which the study was conducted, and whether abundance or occupancy was

presented as the response (dependent) variable (Table S2B). Where authors did not

report the scale over which sites were surveyed, the area of study was measured by

drawing a polygon around the most outer study sites using ImageJ V. 1.52q (Schneider

et al. 2012) when a map, or GPS locations of survey sites were available.

Abundance-occupancy correlation- and beta-coefficients

For each AOR, we calculated the Pearson's product-moment correlation

coefficient (*r*) from the coefficient of determination: $r = \sqrt{R^2}$ when it was not directly

reported by the authors. We then transformed Pearson's correlation coefficients to

Fisher's Z to normalize their distribution:

$$Z = \frac{1}{2} ln \left(\frac{1+r}{1-r} \right) \tag{1}$$

To analyze the slope of AORs, we calculated β-coefficients and standard errors from

linear regression models, when raw data were available, using the lm function in R

version 4.0.4 (R Core Team 2019).

Thirty percent of studies reported more than one AOR, which are more similar

to one another than abundance-occupancy relationships from different studies because

they were collected from similar sets of organisms, the same habitats, by the same

research group, same sampling procedures, etc. This introduces dependency within the

dataset which violates the assumptions of independence of effect-sizes required for

traditional meta-analysis (Fernández-Castilla *et al.* 2020). Multi-level meta-analytical

techniques better account for this non-independence among effect-sizes, and prevent

artificial reduction of heterogeneity and thereby decreasing chances of type I errors

(i.e., false-positives) by accounting for the variation among abundance-occupancy

relationships within studies (Cheung 2014, Hedges 2009). To calculate the effect size

for the correlation coefficients of AORs we used a three-level meta-analysis modeling

structure with random-effects both at the individual-correlation and at the study level

(Cheung 2014, Konstantopoulos *et al.* 2011). The first level estimates sampling

variance of the effect sizes, in second level variance is calculated from the effect sizes

within each study, and the third level variance is that among studies (Cheung 2014,

Assink & Webbelink 2016). We assumed AORs in the same study share the same

covariance and AORs of different studies were independent from one another (Cheung

2014). The effect sizes of the random-effects model ($\hat{y}_{ij}$) were calculated according to

Chenug (2014), where *yi* is the *i*th effect size (i.e., AOR Fisher's Z) in study *j*:

$$\hat{y}_{ij} = \beta_{0+} u_{(2)\,ij+} u_{(3)j+} e_{ij.}.$$ (2)

β_0 is the average AOR effect size, $u_{(2)ij}$ is the within study heterogeneity, $u_{(3)j}$ is the

between study heterogeneity, and e_{ij} is the known sampling error in the *i*th AOR of the

*j*th study. Effect sizes and associated confidence intervals were fitted using the rma.mv

function in the R package metafor with restricted maximum-likelihood estimation

(Viechtbauer 2010).

A multilevel model structure was not necessary for a meta-analysis of the slope of AORs. There was no difference between mutlilevel and traditional two-level analysis in model fit according to likelihood-ratio tests when variance within and among studies was fixed at zero (Table S3). Therefore, we used a random-effects model to estimate the pooled β-coefficient (effect size) of each AOR and the Sidik-Jonkman estimator for the amount of heterogeneity using the rma function in the R package metafor (Viechtbauer 2010). The Knapp & Hartung (2003) method was used to adjust the standard errors of the estimated model coefficients and confidence intervals, therefore tests of individual coefficients and confidence intervals were based on the t-distribution with k-p degrees of freedom where k is the number of studies and p is the total number of model coefficients.

Heterogeneity and variance calculations

For variance quantification of AOR correlation coefficients, we calculated the proportion of total variation due to between-study and within-study heterogeneity using the I^2 statistic (Higgins & Thompson 2002, Koricheva et al 2013). Within study (level 2), I^2 is calculated as

$$I^2_{(2)} = \frac{\hat{\tau}^2_{(2)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{v}} , \qquad (3)$$

where $\hat{\tau}^2_{(2)}$ and $\hat{\tau}^2_{(3)}$ are the variances of the true effect sizes among Level 2 and Level 3, and $\tilde{v}$ is the within-study sampling variance of the ith study (Borenstein *et al*. 2009, Cheung 2014).

Between studies, the percent of variance due to level 3 between-study

heterogeneity ($I^2_{(3)}$) is

$$I^2_{(3)} = \frac{\hat{\tau}^2_{(3)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{v}} \ . \tag{4}$$

Variance distribution calculations were computed with the R package dematr (Harrer et al 2019a & 2019b). We determined whether the heterogeneity we calculated was statistically significant within and among studies through with log-likelihood ratio tests when variance either within or among studies is manually fixed at zero, and then compared to the same model which included each variance component (Konstantopoulos et al 2011). Variance estimates displayed a peak in restricted log-likelihood values, ensuring variance components at both levels were identifiable. Heterogeneity in the slope of AORs was assessed with Cochran's Q-test (Cochran 1954), τ^2, and I^2 statistics computed with the rma function in the R package metafor (Viechtbauer 2010).

Outlier and Influence Diagnostics

Outliers of AOR correlation coefficients were detected by examining Cook's distance and DFBETAS of the effect sizes of random-effects multi-level model when the ith correlation is excluded from the model fitting. A correlation was influential if the Cook's distance was greater than a threshold of $\frac{4}{n}$ (Cook & Weisberg 1982), and the DFBETAS value was greater than a size-adjusted value of $\pm\frac{2}{\sqrt{n}}$ (Belsley *et al*. 1980), where n is the number of AORs. Of 247 abundance-occupancy correlation coefficients initially included in our meta-analysis, we identified the 11 AORs as outliers (Figure S1, Table S4).

Statistical outliers of AOR slopes (β-coefficients) were identified using two methods. First, AORs were identified as outliers if their 95% confidence intervals lay outside of the 95% confidence interval of the pooled effect size. We used the find.outliers and gosh.diagnostics functions in the R package dmetar to identify these outliers (Harrer *et al*. 2019a, Olkin *et al*. 2012). Second, 10^4 data points from a graphical display of study heterogeneity plots were randomly sampled to identify studies that contribute to cluster imbalance using k-means, density reachability and connectivity clustering (DBSCAN) and Gaussian Mixture Model clustering algorithms (Harrer *et al*. 2019a). Three AORs were identified with this method, two of which were identified previously according to their 95% confidence intervals. Of the 74 abundance-occupancy β-coefficients included in our study, we excluded all three AORs identified (Table S5).

Publication bias

Publication bias can occur when researchers only publish results that are statistically significant or conform to a particular hypothesis, while omitting those that do not (the "file drawer problem"). This can be a concern in meta-analyses if the data sets identified are not an accurate representation of the phenomenon being studied. Because of the non-independence of each effect-size, it may not be possible to detect publication bias in three-level meta-analyses through examination of funnel plots (Light & Pillemer 1984) or statistical tests such an Egger's regression test (Egger *et al*. 1997). To overcome this issue, we collected data from both studies that explicitly measured AORs and studies that simply reported abundance and occurrence data but

were focused other topics and so did not explore AORs. By comparing correlations between the two groups, we can determine if studies that set out to measure AORs tended to have higher effect-sizes. We maintained the three-level random-effects model described above, and analyzed if a study's authors reported an AOR as a bivariate subgroup moderator with post-hoc comparisons using Holm p-value adjustments. Publication bias of β-coefficients were analyzed with Egger's regression test (Egger *et al*. 1997) to identify funnel plot asymmetry with the regtest function in the R package metafor (Viechtbauer 2010).

Subgroup Analyses and Meta-Regression

We explored differences among predictors by adding them as moderators to each meta-analysis. The moderators included were both categorical and continuous. Continuous variables (number of sites, area of study, duration of study) were scaled and centered around the mean before analysis. We excluded AORs with missing values or categories that represented a combination of disparate taxonomic categories (e.g., invertebrates combined with micro-organisms). At the same time, we combined some taxonomic categories in order to satisfy sample size requirements: invertebrates (insects with other invertebrates), micro-organisms (microalgae, protozoans, and bacteria), plants (vascular plants, non-vascular plants, and macroalgae), and vertebrates (fish, amphibians, reptiles, and reptiles & amphibians, mammals). Categories with less than $k = 10$ AORs were excluded from subgroup analyses to satisfy sample size requirements for robust parameter estimation (Higgins & Thompson 2004)

We did not include an intercept in the models when analyzing individual moderators, since there was not one true "reference" category from which to make comparisons, so the omnibus test for significance included all moderator coefficients. Test statistics were calculated using a t-distribution with k-p degrees of freedom, where k is the number of studies and p is the total number of model coefficients for individual coefficients and confidence intervals. We used an F-distribution with m and k-p degrees of freedom for the omnibus test statistic, where m is the number of moderators included in the model (Viechtbauer 2010). We compared model-coefficients among categorical predictors using post-hoc comparisons with contrast matrices and a Holm p-value adjustment for multiple comparisons using the R package multcomp (Wright 1992, Hothorn et al 2008).

In subgroup analyses, we performed tests for heterogeneity using the Q_E test for residual heterogeneity to determine if variability in the observed AORs, not accounted for by the moderators, is larger than expected based on sampling variability. We used the profile function from the R base stats package to verify parameter estimates (R Core Team 2020). All parameter estimates were verified with profile likelihood plots of the variance components to be sure there was a clear peak in the REML parameter estimates across levels. For multi-level models of AOR correlation coefficients, significance of within- and among-study variance was examined for all models with significant moderators using likelihood-ratio-tests described above.

Results

Abundance-occupancy relationships are overwhelmingly positive

In total, we analyzed 237 AOR correlation coefficients from 97 studies, and 71 AOR β-coefficients from 58 studies. All but one of the AORs were positive, with the result that the mean correlation coefficient (Fisher's Z) of all the AORs included in our study was positive (Table 1, Figure 2A). There was significant heterogeneity among AOR correlation coefficients, with the large amount of unexplained heterogeneity mostly distributed among studies (Table 1). Even though variance at the level of AORs nested within each study was lower than at the study level, there was still statistically significant variability at both levels according to a likelihood-ratio test when variance components were removed from the model, justifying the use of a mutli-level meta-analysis for these data (Table S6).

The mean slope (β-coefficient) of AORs was positive (Figure 2B), however unlike the correlation coefficients (*r*) of the AORs, there was very little unexplained heterogeneity (I^2=7.24%, Table 1). This suggests strongly that aquatic systems, broadly, have a single universal slope in AORs.

Vertebrates have stronger abundance-occupancy correlations, but similar slopes to other taxonomic groups

The large amount of unexplained heterogeneity in the pooled estimate of AOR correlation coefficients (in essence the "strength" of the relationships) suggests the presence of subgroups in the dataset. Among taxonomic groups, vertebrates had significantly higher correlation coefficients than invertebrates, while micro-organisms

were not statistically significantly different from either vertebrates or invertebrates

(Figure 3A, Table S7). Unexplained heterogeneity remained high even after the

addition of taxonomic group to multi-level random effects model (total $I^2 = 83.4\%$, Q_E

$_{226} = 1038.4$, $p < 0.0001$) and variance among studies remained much larger than

variance within studies ($\sigma^2 = 0.01$ within studies, and $\sigma^2 = 0.08$ among studies). There

was no difference in AOR slopes (β-coefficients) between invertebrates and

vertebrates (Figure 3B, Table S7), and residual heterogeneity remained low after

taxonomic group was added to the model ($I^2 = 5.0\%$, $\tau^2 = 0.009$, $Q_{E,64} = 18.9$). There

were insufficient data to compare AOR slopes for micro-organisms.

Intertidal systems have stronger abundance-occupancy correlations

The slope and correlation coefficients of AORs were stronger for intertidal

systems than marine or freshwater systems, which were not significantly different

from one another (Figure 3C,D, Table S8). In our meta-analysis of slopes,

heterogeneity remained low after realm was added to the model ($I^2 = 6.7\%$, $\tau^2 = 0.013$,

$Q_{E,68} = 20.3$). In contrast, even though intertidal realms had a significantly higher

pooled correlation coefficient, this did not reduce the large amount of heterogeneity

(total $I^2 = 83.8\%$, $Q_{E\,233} = 1375.3$, $p < 0.0001$), and variance among and within studies

remained significant (σ^2_{97} among studies $= 0.081$, σ^2_{235} within studies $= 0.009$).

Similar abundance-occupancy relationships among habitats within intertidal, marine, and freshwater systems

The habitats (within aquatic realms) included in the meta-analyses of correlation coefficients and slopes were not statistically significantly different from one another (Figure S2, Table S9). For slopes, heterogeneity remained low when habitat type was included ($I^2 = 6.7\%$, $\tau^2 = 0.014$, $Q_{E,44} = 12.1$, $p = 1.0$). For correlation coefficients, adding habitat as a moderator did not reduce the high amount of heterogeneity (total $I^2 = 83.9\%$, $Q_{E\,210} = 1239.6$, p <0.0001), and within-study and between-study variance remained statistically significant (σ^2_{96} among = 0.078, σ^2_{232} within studies = 0.008). This result suggests that habitat type is not responsible for producing the large amount of heterogeneity of correlation coefficients in the model.

Similar abundance-occupancy relationships globally

We found no consistent patterns of AORs geographically. The four biogeographic regions best represented in our meta-analysis (Afrotropics, Nearctic, Neotropics and Palearctic) all had positive average effect-sizes that were not statistically significantly different from one another (Table S10, Figure S2C). Biogeographic region did not account for the high variance and heterogeneity observed in the model estimates of correlation coefficients (total $I^2 = 83.3\%$, $Q_{E\,216} = 1013.8$, p <0.0001), nor the significant variance observed within and among studies (σ^2_{86} among studies = 0.08, σ^2_{219} within studies = 0.01). The Palearctic and Nearctic were the only two biogeographic regions included in the analysis of slopes since other

regions had low representation in the dataset, and there was no difference in mean slopes between them (Figure S2D, Table S10). Heterogeneity remained low when biogeographic region was added to the model of slopes ($I^2 = 6.7\%$, $Q_{E,44} = 12.06$, $p = 1.0$). Similarly, we found no relationship between AOR correlation coefficients and midpoint absolute latitude (distance from equator) of the study ($F_{1,228} = 0.748$, $p = 0.39$), or between AOR slope and absolute mid-latitude (distance from the equator: $F_{1,68} = 0.08$, $p = 0.78$).

No effect of study design

The design of the study had very little impact on the strength or the slope of abundance-occupancy relationships. The spatial scale over which AORs were measured ranged from very small (0.00016 km²) to studies encompassing sites across the globe. However, there was no significant relationship between the size of the study area and the correlation coefficient ($F_{1,195} = 1.25$, $p = 0.265$) or β-coefficient $F_{1,54} = 0.43$, $p = 0.512$). Similarly, there was no significant relationship between the number of sites sampled and the correlation-coefficient ($F_{1,187} = 0.512$, $p = 0.48$) or β-coefficient ($F_{1,66} = 2.99$, $p = 0.09$). The number of years over which data were collected ranged between one and 45 years, however there was no relationship between the duration of the study and the correlation coefficient ($F_{1,199} = 0.406$, $p = 0.524$) or β-coefficient ($F_{1,60} = 0.25$, $p = 0.61$).

The direction of causality defined by the authors varied among studies for which the authors calculated AORs; 39 studies placed measures of abundance as the response variable, while 121 studies placed occupancy as the response variable. There

was no statistically significant difference in correlation coefficients between either type of study ($F_{1,161} = 0.71, p = 0.170$). Statistical transformations of AORs varied, and we found a total of 10 possible combinations in the literature. The four most common were log-none ($k = 75$), followed by log-logit ($k = 34$), log-log ($k = 22$), and none-none ($k = 16$). When examined as moderators, none of the statistical transformations of AORs calculated by the study authors different significantly from one another (Table S11).

No evidence of publication bias

We did not detect publication bias among the 108 studies in the AOR correlation-coefficient meta-analysis. AORs that were explicitly measured did not have significantly different Fisher's-Z from studies that did not report an AOR but that provided abundance and occupancy data from which we could calculate the relationship (Figure S3, $F_{1,234} = 2.58, p = 0.109$). There was also no publication bias detected in the meta-analysis of slopes in a regression test for funnel plot asymmetry (t $= 0.55, p = 0.58$, Figure S4), and we would not have expected to find one since only $k = 5$ out of the $k = 66$ β-coefficients included in the meta-analysis were reported from studies explicitly measuring AORs.

Discussion

In two meta-analyses of 247 abundance-occupancy relationships from 108 published studies, we found the strength and slope of the relationship between abundance and occupancy to be overwhelmingly positive across all aquatic realms,

habitats, taxonomic groups, biogeographical locations (Figure 2A). This must mean that either the pattern is an artifact and so uninteresting, or a real ecological relationships that begs for an explanatory mechanism. Our result is consistent with a previous, though earlier and so less extensive meta-analyses (Blackburn *et al*. 2009) that also found that AORs are largely positive across aquatic environments. Furthermore, in our analysis freshwater and marine systems had stronger mean pooled correlation coefficients than previously documented (Blackburn *et al*. 2009). Because our data do not support any kind of statistical or sampling artifact, we conclude that the pattern is of fundamental ecological importance: species that occur widely across a region of any dimension, from local to global, are nearly always abundant where they occur, while species that are only found in relatively few locations tend to be of relatively low abundance.

Aquatic abundance-occupancy correlations are positive, but highly heterogeneous – There is not one "universal" AOR for aquatic systems, even though it is overwhelmingly positive; correlation coefficients are highly heterogeneous (Table 1). Furthermore, heterogeneity remained high when moderators taxonomic group, realm, habitat, or biogeographic region were added to the model, suggesting there are other sources of heterogeneity that we were unable to identify in our analysis. While in contrast to our results, Blackburn *et al*. (2009) did not find significant heterogeneity among aquatic AORs, we were able to include a much larger number of AORs in our analysis (247 AORs from 108 studies) than they did (57 AORs from 16 studies). This heterogeneity in correlation coefficients found our meta-analysis provides an

opportunity to consider some of the main hypotheses for the existence of AORs.

Not sampling bias – In addition to ecological explanations for positive AORs, the broad generality of the relationship has been suggested simply to be the result of statistical artefact. One suggestion has been that a positive AOR would be produced where locally rare species are less likely to be detected than locally abundant species, therefore producing artificially low occupancy levels for these taxa (Bock & Ricklefs 1983). Although this is a concern in studies with low sampling intensity, positive AORs have been observed in studies with near complete sampling (Borregaard & Rahbek 2010). If sampling artefacts were a concern in our analysis, we would have expected to find weaker AORs in studies with more exhaustive sampling (more sites or more species), and this was not the case. We did not find that the number of sites sampled ($F_{1,187} = 0.41, p = 0.48$), nor the area of the study ($F_{1,195} = 1.25, p = 0.27$) had a moderating effect on the strength of abundance-occupancy relationships in aquatic systems.

Another suggestion has been that if population sizes near the center of a species geographic range are higher and decline towards the range edges, then any survey done on the edge of the range might produce artificially low levels of local abundance or occurrence for that species (Bock & Ricklefs 1983). We note for our meta-analysis that large-scale studies ought to include a greater proportion of a species' geographic range than smaller-scale studies, and if this sampling bias were important, studies with larger ranges should have weaker AORs than those covering smaller ranges. In contrast, we found that range position does not affect the strength or

slope of the AOR in aquatic systems: those covering less than 1 km² had similar

strengths of AORs as studies at the global scale ($F_{1,195} = 1.25, p = 0.27$).

Finally, we found no evidence of publication bias since studies reporting

AORs did not have relationships at all different from those that only reported the

relevant data from which we calculated correlation coefficients and slopes (Figure S3).

Taken together these analyses strongly suggest that AORs generally are genuine

ecological relationships which must have one or more mechanistic explanations.

Ecological mechanisms considered

Meta-population dynamics – The correlation between abundance and occupancy may

be a result of the role that dispersal plays in extinction and colonization dynamics in a

spatially structured population. Because likelihood of dispersal varies markedly

among taxonomic groups independent of whether populations are structured as a meta-

population, it is possible to evaluate this hypothesis using our meta-analysis. First,

many species do not occur as meta-populations, so that hypothesis that this structure

plays a role in AORs may not be appropriate for studies at larger scales (Hanski 1994).

In our meta-analysis, we found no differences among taxonomic groups where we

would expect to find species that can easily disperse. For instance, there were no

differences in AORs in vertebrates from lakes and ponds where we would expect to

find greater barriers for dispersal compared with vertebrates in ocean environments

(Table S12).

At the same time we found, as did Blackburn *et al*. (2006), that species in the

intertidal realm had significantly stronger AOR correlations than those in freshwater

or marine systems (Figure 3C). Furthermore, we found that the mean pooled β-coefficient (slope) was higher in intertidal than either freshwater or marine systems (Figure 3D). This stronger AOR for intertidal systems may be due to the comparative lack of dispersal barriers as well as the life histories of resident organisms. In our meta-analysis of the strength of AORs, studies of intertidal habitats where dominated by invertebrates (Figure 4) and there were no statistically significant interactions of AOR correlation coefficients between taxonomic group and realm, therefore the stronger mean correlation coefficient is a characteristic of the system as a whole, rather than the type of organism that lives in each realm. Nevertheless, because invertebrates that make up the majority of AORs for studies of intertidal habitats, it might be that their pelagic larvae have higher dispersal than other organisms. Studies that have explicitly examined larval dispersal in these systems have found species with planktonic larval dispersal have higher AOR intercepts, but no differences in the slopes (Foggo *et al*. 2007). Other studies that explicitly examined intertidal invertebrates found no differences among the slopes of AORs among taxonomic groups within the invertebrate category (Foggo *et al*. 2003; Frost *et al*. 2004).

Differences in organism size provides another way to assess the potential role of dispersal in AORs, since size covaries with taxonomic group (micro-organisms < invertebrates < vertebrates), and is very likely related to dispersal ability, with smaller organisms potentially dispersing more readily (Tales *et al*. 2004; O'Sullivan *et al*. 2014). Conceptually, species with smaller sizes should have higher occupancy because they may face fewer dispersal barriers, and may have a relatively higher population growth rates than larger species (Blackburn & Gaston 2001). We found that

abundance and occupancy were more tightly correlated for vertebrates than for invertebrates or micro-organisms (Figure 3A), but no difference in AOR slope for vertebrates and invertebrates (there were too few studies in our analysis that included micro-organisms to be useful for this comparison) (Figure 3B). We note that the vertebrates in our meta-analysis consisted primarily of fish ($k = 95$; mammals $k = 1$; reptiles and amphibians $k = 2$) and the great majority of studies of fish in our dataset were from marine and freshwater environments with very few in intertidal settings (Figure 4) so that taxon and realm were not entirely independent. In any case, since the vertebrates in the studies we analyzed were quite generally larger than the invertebrates, and both were larger than the micro-organisms, the finding that vertebrates had the highest AOR correlation coefficients is opposite to what we would expect if small organisms have higher dispersal potential.

Interestingly, the relationships between body size and abundance, and body size and occupancy are not the same. Several studies have found, within taxonomic groups, a negative relationship between body size and abundance. We note, however, that the relationship is often relatively weak in fishes (Macpherson 1989; Pyron 1999; Tales et al. 2004), and, indeed, a monotonic relationship between body size and abundance may not be appropriate for fish at all, since intermediate body sizes tend to have higher abundances for freshwater fish across the United States and Europe (Gaston & Lawton 1990; Passy 2012). In contrast, for micro-organisms, body size may play a role in determining abundance. In a large study of stream diatoms across the United States, smaller species tended to have larger population sizes, at least at local scales (Passy 2012; Rocha et al. 2018). Comparisons of AORs for stream

26

diatoms with those for larger multicellular organisms found no differences, suggesting similar underlying mechanisms (Soininen & Heino 2005) and no effect of organism size. Similarly, we did not detect any differences in the correlation coefficient of AORs between micro-organisms and either invertebrates or vertebrates in our analysis (Figure 3A), nor was there any difference in AOR slope between invertebrates and vertebrates. Of the studies for which we have examined the role of body size in AORs, results are scale specific. The relationship between body size and occupancy is often only significant in studies that include many sites, or large areas (Pyron 1999; Tales *et al*. 2004).

Unlike fish, micro-organisms do conform to our expectations of the relationship between body size and occupancy. For example, studies of stream diatoms have demonstrated a negative relationship between cell size and range size in the context of AORs, with smaller species exhibiting larger geographic range sizes and higher levels of occupancy (Passy 2012; Rocha *et al*. 2018). The stronger AORs for vertebrates may be because fish do not conform to our expected relationships of body size vs abundance and occupancy. Within taxonomic groups, larger species may have higher abundance and occupancy, whereas among taxonomic groups this expected relationship breaks down. Given that fish tend not to follow a consistent pattern between body size and abundance and body size and occupancy, the difference in strength of AORs for fish may be attributable to this.

These considerations taken together, lead us to conclude that the meta-population and related dispersal hypotheses as explanations of AORs is only weakly supported in the results of our meta-analysis.

Niche dimensions – The niche breadth hypothesis first proposed by Brown (1984) suggests that abundant and wide-spread species should be generalists, found in a broad range of conditions or using a wide variety of resources, while rare species should be specialists using a narrower set of resources or occurring in a narrower range of conditions, or both. This explanation assumes that population density reflects the probability density distribution of environmental variables, that the environment is autocorrelated, and that ecologically similar species differ in only a small number of niche dimensions (Brown 1984).

Examination of AORs among biogeographic regions using our meta-analysis provides some insight into the validity of the niche breadth hypothesis. The tropics tend to have more specialists than generalists (Huey 1978; Laurance *et al*. 2011; Perez *et al*. 2016), which may manifest several ways, for example, tropical species tend to have narrow ranges in temperature tolerance (Janzen 1967; Ghalambor *et al*. 2006). Thus, if this were a dominant mechanism producing AORs, we would expect to find stronger relationships in the tropics than in the temperate zones, and we do not (correlation coefficients and β-coefficients were not statistically significantly different among biogeographic regions; Figure S2C, Table S10). It was also not the case that there were more species sampled in the tropics. There was no relationship between absolute latitude (distance from equator) and number of species in each relationship ($F_{1,288}= 0.75, p = 0.39$). Thus, at least at this broad biogeographic scale, we find no support for the niche breadth explanation for AORs.

Conclusion

Where does this leave us? First we return to the main finding of our meta-analysis, that there is a clear, exceptionally consistent relationship between the abundance of species and their spatial occurrence among aquatic organisms. This pattern, particularly consistent in AOR slopes, but more variable in correlation coefficients, has been noted by many before us (Balckburn *et al*. 2006, Gaston *et al*. 2000), but our study gathered data from a substantially larger number of studies, and still found that AORs hold across taxa, habitats, realms and biogeography. Further, the breadth of our study made it possible to reject hypotheses that AORs result from ecologically uninteresting sampling or reporting biases of one kind or another can be rejected using our results.

Given that a major goal of Ecology is understanding the distribution and abundance of species in nature (Krebs 2009, Bowman et al. 2017), the broad consistency of the AOR relationship among aquatic systems suggests – as many have speculated – a very general underlying ecological mechanism. We used the results of our meta-analysis to explore whether the heterogeneity of correlation coefficients among the taxa, habitats, realms and biogeography provides a foundation for determining whether one or more of the key ecological mechanistic hypotheses raised in the Introduction explained statistically any of this variation.

In contrast to our ability to reject sampling and statistical artefacts and biases as explanation for AORs, we found it considerably more difficult to discern patterns in our results strongly consistent with any of the three ecological mechanisms, reviewed in the Introduction, that have been hypothesized to produce AORs. While the single

comparison we evaluated that might have supported the niche breadth hypothesis failed, the meta-community and related dispersal hypotheses were consistent with some, but not all of the patterns in our data. It may be that while the meta-analysis we carried out was well suited to illuminating the broad consistency of the relationship between species abundance and spatial occurrence, it did not contain enough species- and environment-specific information to provide strong tests of ecological mechanisms.

References

Assink, M. & Wibbelink, C.J.M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *Quantitative Methods Psychology*, 12, 154–174.

Arrhenius, O. (1921). Species and area. *Journal of Ecology,* 9(1), 95-99.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics*. New York: Wiley.

Blackburn, T.M., Cassey, P. & Lockwood, J.L. (2009). The role of species traits in the establishment success of exotic birds. *Global Change Biology*, 15, 2852–2860.

Blackburn, T. M., & Gaston, K. J. (2001). Linking patterns in macroecology. *Journal of Animal Ecology*, *70*(2), 338–352.

Blackburn, T. M., Cassey, P., & Gaston, K. J. (2006). Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *Journal of Animal Ecology*, *75*(6), 1426–1439.

Bock, C. E., & Ricklefs, R. E. (1983). Range size and local abundance of some North American songbirds: A positive correlation. *The American Naturalist*, 122(2), 295–299.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.

Bormann, I. (2020). DigitizeIt. Retrieved from http://digitizeit.de

Borregaard, M. K., & Rahbek, C. (2010). Causality of the relationship between geographic distribution and species abundance. *Quarterly Review of Biology*, *85*(1), 3–25.

Bowman, WD, SD Hacker & ML Cain. 2017. Ecology. 4th ed. Sinauer Associates, Sunderland, MA

Brown, J. H. (1984). On the relationship between abundance and distribution of species. *American Naturalist*, *124*(2), 255–279.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229.

Cochran, W.G. (1954). The combination of estimates from different experiments. *Biometrics, 10*(1), 101-129

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. London: Chapman and Hall.

de Candolle, A. P. (1855). *Géographie Botanique Raisonnée*. 2 vols. Paris: Masson

Darwin, C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: John Murray

Egger, M., Smith, G.D., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315, 629.

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*(5), 2031-2052.

Fisher, R.A., A.S. Corbet & C.B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12: 42-58.

Foggo, A., Bilton, D. T., & Rundle, S. D. (2007). Do developmental mode and dispersal shape abundance-occupancy relationships in marine macroinvertebrates? *Journal of Animal Ecology*, *76*(4), 695–702.

Foggo, A., Frost, M. T., & Attrill, M. J. (2003). Abundance-occupancy patterns in British estuarine macroinvertebrates. *Marine Ecology-Progress Series*, *265*, 297–302.

Frost, M. T., Attrill, M. J., Rowden, A. A., & Foggo, A. (2004). Abundance-occupancy relationships in macrofauna on exposed sandy beaches: Patterns and mechanisms. *Ecography*, *27*(5), 643–649.

Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. (2000). Abundance–occupancy relationships. *Journal of Applied Ecology*, 37, 39–59.

Gaston, K. J., & Lawton, J. H. (1990). Effects of scale and habitat on the relationship between regional distribution and local abundance. *Oikos*, *58*(3), 329–335.

Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., & Lawton, J. H. (2000). Abundance-occupancy relationships. *Journal of Applied Ecology*, *37*, 39–59.

Ghalambor, C. K., Huey, R. B., Martin, P. R., Tewksbury, J. J., & Wang, G. (2006). Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, *46*(1), 5–17.

Gleason, H.A. (1922). On the relation between species and area. *Ecology*, *3*, 158-162

Hairston, N.G. (Sr.) 1959. Species abundance and community organization. *Ecology 40*:404-416.

Hanski, I. (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology*, *63*(1), 151–162.

Hanski, I., & Gyllenberg, M. (1997). Uniting two general patterns in the distribution of species. *Science*, *275*(5298), 397–400.

Hanski, I., Kouki, J., & Halkka, A. (1993). Three explanations of the positive relationship between distribution and abundance of species. Pp.108-116. In R. E. Ricklefs & D. Schluter (Eds.), *Species diversity in ecological communities*. Chicago: Species diversity in ecological communities: historical.

Hartley, S. (1998). A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same). *Journal of Animal Ecology*, *67*(6), 992–994.

Harrer, M., Cuijpers, P., Furukawa, T.A, & Ebert, D. D. (2019a). *Doing Meta-Analysis in R: A Hands-on Guide*. DOI: 10.5281/zenodo.2551803. Chapter 6.2.

Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. (2019b). dmetar: Companion R Package For The Guide "Doing Meta-Analysis in R". R package version 0.0.9000. http://dmetar.protectlab.org.

Harvey, P. H., & Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press.

Hedges, L. V. 2009. Statistical Considerations. Pp. 38-47. In: *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation New York.

Heino, J., & Virtanen, R. (2006). Relationships between distribution and abundance vary with spatial scale and ecological group in stream bryophytes. *Freshwater Biology*, *51*(10), 1879–1889.

Higgins, J. P., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, *23*(11), 1663–1682.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal* 50(3), 346–363

Holt, A. R., & Gaston, K. J. (2003). Interspecific abundance–occupancy relationships of British mammals and birds: is it possible to explain the residual variation? *Global Ecology and Biogeography*, *12*(1), 37–46.

Huey, R.B. (1978). Latitudinal pattern of between-altitude faunal similarity: mountains might be "higher" in the tropics. *The American Naturalist, 112*(983), 225-254.

Janzen, D. H. (1967). Why mountain passes are higher in the tropics. *The American Naturalist*, *101*, 233–249.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2, 61–76.

Krebs, C. J. (2009) *Ecology: The Experimental Analysis of Distribution and Abundance*. 6th ed. Pearson.

Laurance, W.F., Useche, D.C., Shoo, L.P., Herzog, S.K., Kessler, M., Escobar, F., *et al*. (2011).

Global warming, elevational ranges and the vulnerability of tropical biota. *Biological Conservation*, *144*, 548–557.

Light, R. J., Richard, J., Light, R., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Macpherson, E. (1989). Influence of geographical distribution, body size and diet on population density of benthic fishes off Namibia (South West Africa). *Marine Ecology-Progress Series*, *50*, 295–299.

McArdle, B. H., Gaston, K. J., & Lawton, J. H. (1990). Variation in the size of animal populations: Patterns, problems and artefacts. *Journal of Animal Ecology*, *59*(2), 439–454.

Olkin, I., Dahabreh, I. J. and Trikalinos, T.A. (2012). "GOSH–a graphical display of study heterogeneity." *Research Synthesis Methods* 3(3), 214–223.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., *et al*. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *Bioscience*, 51, 933–938.

O'Sullivan, D., Benton, T. G., & Cameron, T. C. (2014). Inter-patch movement in an experimental system: the effects of life history and the environment, *Oikos, 123*(5), 623-629.

Passy, S. I. (2012). A hierarchical theory of macroecology. *Ecology Letters*, *15*(9), 923–934.

Pyron, M. (1999). Relationships between geographical range size, body size, local abundance, and habitat breadth in North American suckers and sunfishes. *Journal of Biogeography*, *26*(3), 549–558.

Rapoport, E.H. 1982. *Areography: Geographical Strategies of Species*. New York: Pergamon Press.

Riis, T., & Sand-Jensen, K. (2002). Abundance-range size relationships in stream vegetation in Denmark. *Plant Ecology*, *161*, 175–183.

Rocha, M. P., Bini, L. M., Siqueira, T., Hjort, J., Grönroos, M., Lindholm, M., et al. (2018). Predicting occupancy and abundance by niche position, niche breadth and body size in stream organisms. *Oecologia*, *186*(1), 205–216.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2021). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675.

Shepherd, T. D., & Litvak, M. K. (2004). Density-dependent habitat selection and the ideal free distribution in marine fish spatial dynamics: considerations and cautions. *Fish and Fisheries*, *5*(2), 141–152.

Soininen, J., & Heino, J. (2005). Relationships between local population persistence, local abundance and regional occupancy of species: distribution patterns of diatoms in boreal streams. *Journal of Biogeography*, *32*(11), 1971–1978.

Stevens, G.C. 1989. The latitudinal gradients in geographic range: How so many species coexist in the tropics. *American Naturalist* 132, 240-256.

Tales, E., Keith, P., & Oberdorff, T. (2004). Density-range size relationships in French riverine fishes. *Oecologia*, *138*(3), 360–370.

Venier, L. A., & Fahrig, L. (1996). Habitat availability causes the species abundance-distribution relationship. *Oikos*, *76*(3), 564–570.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Willis, J. C. 1922. Age and area. The University Press, Cambridge, England.

Whittaker, R. H. (1965). Dominance and diversity in land plant communities: Numerical relations of species express the importance of competition in community function and evolution. *Science*, 147(3655), 250-260.

Wright, D. H. (1991). Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, *18*(4), 463–466.

Box 1. Abundance-occupancy relationships in nature

Abundance-occupancy relationships essentially describe the connection between local population size and the level of occurrence across a landscape. Among the patterns to describe how species are distributed in nature, the positive relationship between the average abundance where a species occurs and the number of places it occupies (occupancy) is almost always positive (Gaston *et al.* 2000).

For example, in a survey of zooplankton from 28 lakes in the Adirondack Mountains, NY, USA (Leach *et al.* 2018), the abundance and occurrence of each species varies systematically.

Species	Average abundance where species occurs (individuals L^{-1})	Proportion of lakes occupied (occupancy)
Mesocyclops edax	2.6	0.93
Daphnia ambigua	0.55	0.64
Daphnia parvula	0.55	0.07
Skistodiaptomus oregonensis	0.033	0.04
etc.	*etc.*	*etc.*
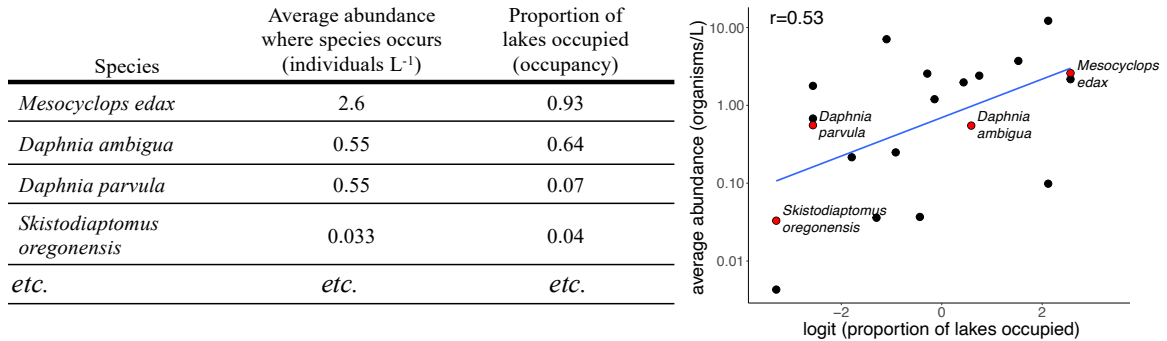
When abundance and occupancy of the whole assemblage of zooplankton found across this landscape are plotted on a log-logit scale, there is a positive linear relationship as a result of this difference among abundance and level of occurrence. In the graph above of 20 species of zooplankton, the red points highlight a few species of interest. The copepod *Mesocyclops edax* occurs in many more lakes and at higher levels of abundance than *Skistodiaptomus oregonensis* which is found in few lakes and at low levels of abundance. In essence, this positive relationship describes the continuum of two axes of commonness and rarity in a natural community (Rabinowitz 1981).

Hypotheses
There are as many as 13 hypotheses to explain the mechanisms producing a positive abundance-occupancy relationship (Borregaard & Rahbek 2010). In this meta-analysis, in addition to considering possible statistical or sampling artefacts, we focus on three ecological mechanisms that are hypothesized to produced positive abundance-occupancy relationships.
1. *Differences in niche breadth among species*: Positive abundance-occupancy relationships are a result from niche differences among generalist and specialist species. If we assume the environment is variable, and that variation increases with increasing scale, generalists that are abundant and widespread have niches that encompass a wide range of conditions, while specialists with a narrow niche are less abundant and occur infrequently (Brown 1984).
2. *Differences in niche position or resource abundance*: If the niche a species occupies is widespread across a region, then we would expect that species to be both abundant and occur in more locations than a species whose specific niche conditions occurs infrequently across the landscape, producing differences in abundance and occupancy among species (Hanski 1993, Venier 1996).
3. *Metapopulation dynamics*: Variation in dispersal (immigration) and extinction rates in a metacommunity may produce positive abundance-occupancy relationships if widespread species have a lower extinction rates and/or higher dispersal rates than less abundant and therefore less common species (Gyllenber and Hanski 1992).

TABLES

Table 1. Random-effects model results of meta-analysis of the correlation coefficient, expressed as Fisher's-Z, and the slope, expressed as the β-coefficient, of abundance-occupancy relationships in aquatic systems. Heterogeneity is expressed as Higgin's & Tompson's I^2 and Cochran's Q. * indicates significance level of p<0.0001 for model results and level of heterogeneity.

	Model estimate (95% CI)	I^2	Q	Among studies σ^2	Within study σ^2	Among studies I^2	Within study I^2
Fisher's-Z	0.900* (0.831-0.968)	84.2%	1388.1*	0.082	0.010	74.9%	9.3%
β-coefficient	0.318* (0.262-0.373)	7.24%	24.05	--	--	--	--

FIGURES

Figure 1. Study selection workflow after retrieval of articles
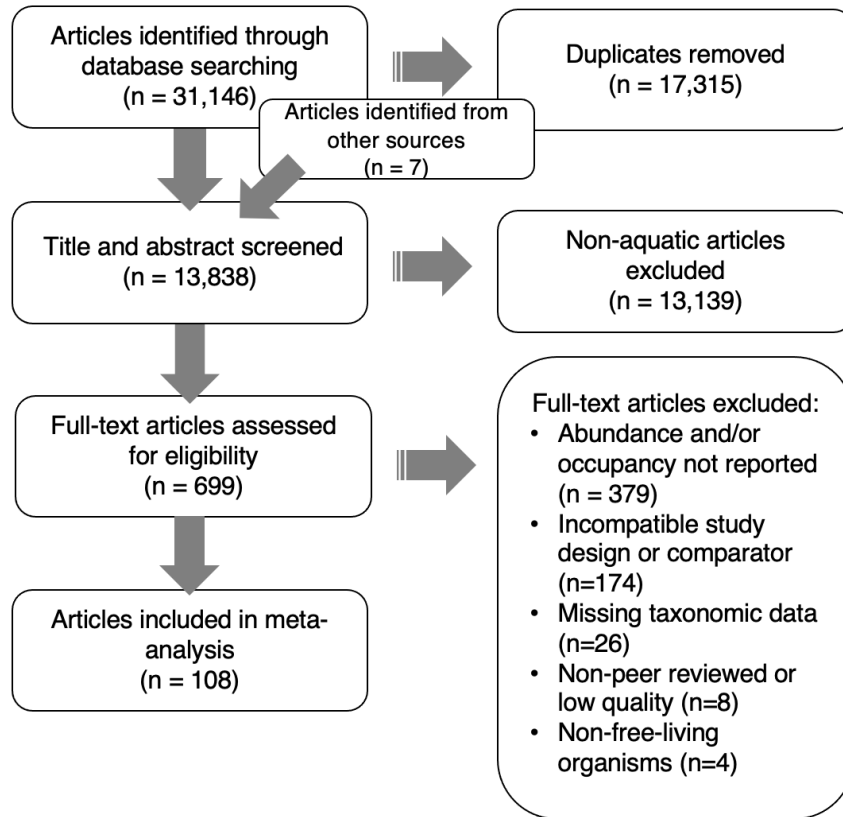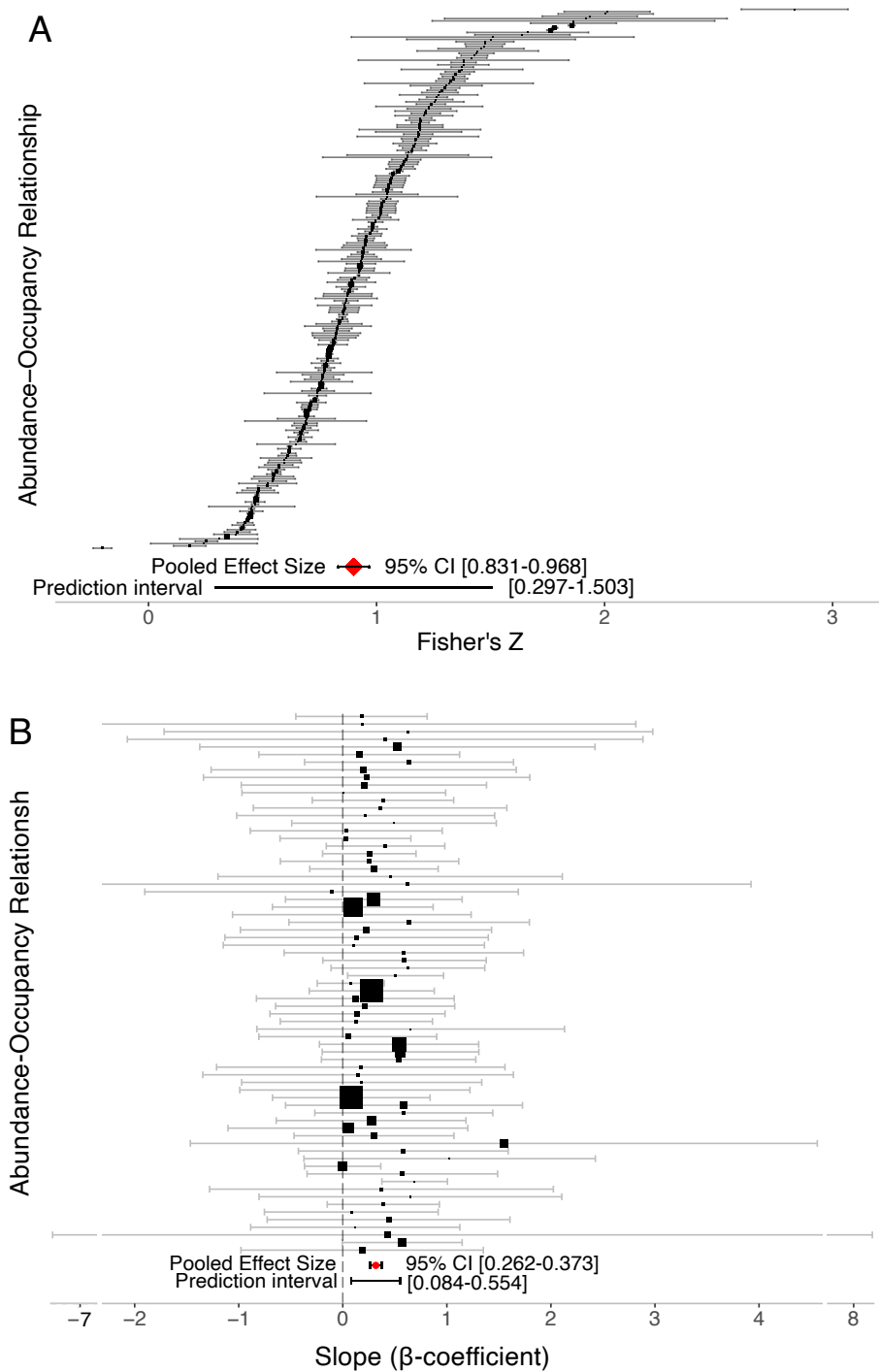
Figure 2. Random-effects model meta-analyses for of abundance-occupancy relationships in aquatic systems. A. The correlation coefficient (Fisher's Z): each point represents one AOR, the size of each point is proportional to the weight of the AOR in each model ordered by their effect-size. B. The slope (β-coefficient): each point represents one AOR, the size of each point is proportional to the weight of the AOR in each model ordered by their effect-size. Error bars are the 95% confidence intervals.
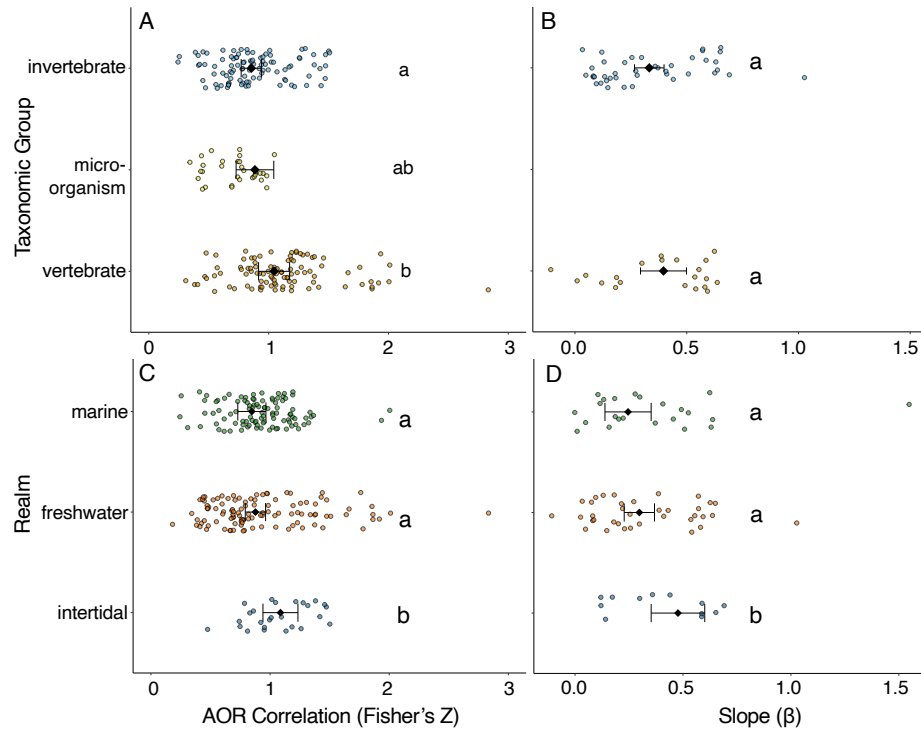
Figure 3. Subgroup model results for abundance-occupancy relationship (AOR) correlation coefficients and slopes among taxonomic groups (A,B) and aquatic realms (C,D). Each points is one abundance-occupancy relationship and black diamond points are the average pooled effect size from each subgroup from the random-effects multilevel model, in the case of correlation coefficients, and random-effects model in the case of AOR slopes. Error bars are 95% confidence intervals of model estimates.
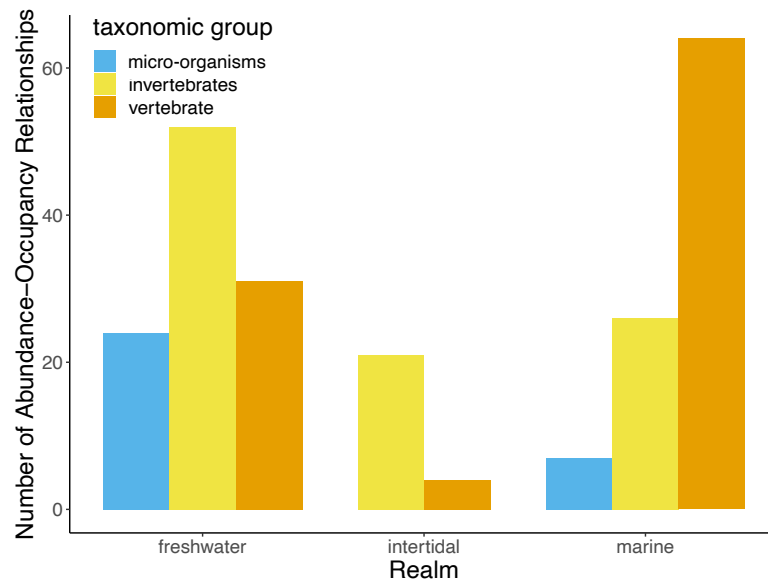
Figure 4. Distribution of the number of abundance-occupancy relationships for studies of invertebrates, micro-organisms, and vertebrates among aquatic realms included in our meta-analysis of abundance-occupancy correlation coefficients.

Table S1. Search terms included in systematic literature search for studies measuring abundance-occupancy relationships. Terms within each category were concatenated with OR, categories were concatenated with AND terms.

Relationship	Abundance	Occupancy
corre*	abundance	occup*
relationship*	density	occur*
regress*	"population size"	
	count	
Aquatic characteristics	Aquatic organisms	
aquatic	*bacteria	macroinvertebrate*
marine	bacteria*	insect*
freshwater	microb*	infauna*
"fresh water"	cell*	beetle*
lake*	protist*	fly
stream*	alga*	flies
river*	phytopl*	*fish
ocean*	plankton	fish*
intertidal	plant	vertebrate*
subtidal	macrophyte	mammal
estuar*	kelp	animal*
benth*	weed	organism*
pelagic	bryophyte*	meiofauna
littoral	moss	macrofauna
	angiosperm*	infauna
	zooplankton*	coral
	invertebrate*	reef

Table S2. A. Groupings of habitats within Realms, and subcategories within Taxonomic groups.

Realms	Habitats
Freshwater	streams, rivers, wetlands, ponds, lakes
Intertidal	rocky intertidal, estuaries
Marine	coral reef, benthic, open ocean

Taxonomic groups	Subcategories of organisms
Micro-organisms	attached algae, phytoplankton, protozoans, bacteria
Invertebrates	corals, sponges, crustaceans, insects
Vertebrates	fish, amphibians, reptiles, mammals

Table S2. B. Categories of studies included in the meta-analysis of abundance-occupancy relationships after outliers were removed; k is the number of abundance-occupancy relationships from each category.

Region	Taxonomic group & Habitat	k	Region	Taxonomic group Habitat	k
Afrotropic	Invertebrates		Neotropic	Invertebrates	
	marine benthos	3		river & stream	5
	river & stream	2		marine benthos	2
	Vertebrates			intertidal	2
	offshore ocean	6		lake & pond	2
				Microbe	
Antarctic	Invertebrates			offshore ocean	1
	marine benthos	2		Microbe & Invertebrates	
				intertidal	1
Australasia	Invertebrates			Vertebrates	
	lake & pond	1		intertidal	1
	Vertebrates			lake & pond	1
	coral reef	2		offshore ocean	10
	estuary	2		river & stream	1
Global	Microbe		Nearctic	Invertebrates	
	offshore ocean	4		lake & pond	6
				offshore ocean	6
Indomalaya	Invertebrates			river & stream	2
	marine benthos	1		freshwater wetland	1
	coral reef	1		Microbes	
	estuary	1		river & stream	3
	river & stream	1		freshwater wetland	2
	Vertebrates			lake & pond	1
	offshore ocean	1		Vertebrates	
				estuary	1
Nearctic	Invertebrates			lake & pond	1
	lake & pond	6		offshore ocean	49
	offshore ocean	6		river & stream	8
	river & stream	2		freshwater wetland	2
	freshwater wetland	1			
	Microbes				
	river & stream	3			
	freshwater wetland	2			
	lake & pond	1			
	Vertebrates				
	estuary	1			
	lake & pond	1			
	offshore ocean	49			
	river & stream	8			
	freshwater wetland	2			

Table S3. Results of likelihood ratio tests of three-level random-effects models of abundance-occupancy relationships. Meta-analysis of correlation coefficients (Fisher's Z) and slope (β) were run separately. Comparisons are the result of variance among and within studies removed compared to the full three-level random effects model including variance partitioned among all three levels. AICc is the corrected Akike Index Criterion, ΔAICc is the difference between AICc of the reduced model vs the full model. LRT is result of a likelihood-ratio-test between the two models. *** indicated $p<0.001$

Model with variance component removed	ΔAICc	AICc	LRT
Fisher's Z model excluding within-study variance	122.7	256.1	124.7***
Fisher's Z model excluding between-study variance	18.3	162.2	30.8***
β model excluding within-study variance	1.9	106.1	0.29
β model excluding between-study variance	2.2	105.8	0

Table S4. Studies excluded from the meta-analysis of abundance-occupancy correlation coefficients. Correlation coefficients calculated from a study's raw data indicated by *.

Study	Pearson r	Fisher's Z	Variance	Realm	Biogeographic region	Midpoint latitude	Taxonomic group	Habitat	Species	Sites	Scale (km2)	Study length (years)	Dependent variable	Independent variable
Choi et al. 2011	0.97	2.03	0.20	intertidal	Indomalayan	36N	invertebrates	intertidal	8	28	16.6	1	abundance	occupancy
Counsell et al. 2018	-0.03	-0.03	0.04	marine	Australasian	21N	vertebrates & invertebrates	reef	31	19	1693	1	abundance	occupancy
Dantas et al. 2010	0.99	2.51	0.20	intertidal	Neotropics	7S	vertebrates	intertidal	8	3	28.5	1	abundance	occupancy
Liu et al. 2015	0.41*	0.43	0.00	marine	Palearctic	37N	microbes	lake	10559	42	2563042	1	abundance	occupancy
Martins et al. 2014	0.93	1.67	0.01	marine	Palearctic	39N	invertebrates	marine benthos	140	145	59030	1	abundance	occupancy
Richards et al. 2013	0.98*	2.25	0.01	marine	Australasian	3N	invertebrates	reef	82	100	1458000	1	abundance	occupancy
Schmidl et al. 2008	0.95	1.79	0.14	freshwater	Palearctic	49N	invertebrates	pond	10	50	NA	1	abundance	occupancy
Simkhada et al. 2006	0.12	0.12	0.03	freshwater	Indomalayan	27N	microbes	pond	35	16	261	1	abundance	occupancy
Skvortsov et al. 1997	0.34	0.35	0.03	freshwater	Palearctic	67N	invertebrates	lake	43	19	NA	3	abundance	occupancy
Tongnunui et al. 2016	0.89	1.40	0.01	freshwater	Indomalayan	14N	vertebrates	river	79	96	NA	1	abundance	occupancy
Travers et al. 2012	0.11	0.11	0.03	marine	Australasian	20S	vertebrates	offshore ocean	33	7	89241	1	abundance	occupancy

Table S5. Studies excluded from the meta-analysis of abundance-occupancy slopes.

Study	β coefficient	Standard error	Realm	Biogeographic region	Midpoint latitude	Taxonomic group	Habitat	Species	Sites	Scale (km²)	Study length (years)
Dolan *et al.* 2007	10.56	17.95	marine	Neotropics	25S	microbes	Offshore ocean	149	21	719,238	1
Fisher *et al.* 2004	3.85	0.678	marine	Nearctic	44N	vertebrates	Offshore ocean	24	166	96,783	32
Hwang *et al.* 2010	0.72	0.064	intertidal	Indomalayan	25N	invertebrates	estuary	28	24	62.2	3

Table S6. Results of likelihood ratio tests of three-level random-effects models of abundance-occupancy relationships. Meta-analysis of correlation coefficients (Fisher's Z) and slope (β) were run separately. Comparisons are the result of variance among and within studies removed compared to the full three-level random effects model including variance partitioned among all three levels. AICc is the corrected Akaike Index Criterion, ΔAICc is the difference between AICc of the reduced model vs the full model. LRT is result of a likelihood-ratio-test between the two models. *** indicated p<0.001

Model with variance component removed	ΔAICc	AICc	LRT
Fisher's Z model excluding within-study variance	122.7	256.1	124.7***
Fisher's Z model excluding between-study variance	18.3	162.2	30.8***
β model excluding within-study variance	1.9	106.1	0.29
β model excluding between-study variance	2.2	105.8	0

Table S7. Model Results with 95% confidence intervals and post-hoc comparisons of taxonomic groups in a three-level random effects model of correlation coefficients and random-effects model of abundance-occupancy beta-coefficients. Significance value * indicates p<0.05, **** p<0.0001

Taxonomic Group	Fisher's-Z	β
Invertebrate	0.86 [0.77 – 0.94] ****	0.33 [0.26 - 0.40]****
Microbe	0.89 [0.73 – 1.04] ****	
Vertebrate	1.05 [0.92 – 1.17] ****	0.40 [0.29 - 0.50]****
Comparison	Fisher's-Z Difference	β Difference
Microbe vs Invertebrate	0.03	
Vertebrate vs Invertebrate	0.19*	0.06
Vertebrate vs Microbe	0.16	

Table S8. Model Results with 95% confidence intervals and post-hoc comparisons of realms in a three-level random effects model of correlation coefficients and random-effects model of abundance-occupancy beta-coefficients. Significance value * indicates $p<0.05$, ** $p<0.01$, **** $p<0.0001$

Realm	k	Fisher's-Z	k	β
Freshwater	111	0.88 [0.79 – 0.96]****	36	0.30 [0.23 – 0.37]****
Intertidal	26	1.09 [0.94 – 1.23]****	12	0.48 [0.35 – 0.60]****
Marine	99	0.85 [0.73 – 0.97]****	23	0.25 [0.14 – 0.35]****

Comparison	Fisher's-Z Difference	β Difference
Intertidal vs Freshwater	0.21*	0.18*
Marine vs Freshwater	-0.03	-0.05
Marine vs Intertidal	-0.24**	-0.23*

Table S9. Meta-analysis results and post-hoc comparisons of habitats in a three-level random effects model of correlation coefficients and random-effects model of abundance-occupancy β-coefficients. Significance values ***$p<0.001$ and **** $p<0.0001$

Habitat	k	Fisher's-Z [95%CI]	k	β [95% CI]
Lake & Pond	34	0.98 [0.84 – 1.11] ****	12	0.23 [0.10 – 0.36] ***
Open Ocean	78	0.90 [0.75 – 1.06] ****	15	0.29 [0.12 – 0.46] ***
River & Stream	70	0.88 [0.77 – 0.98] ****	20	0.33 [0.25 – 0.42] ****
Marine Benthos	23	0.87 [0.66 – 1.07] ****		
Intertidal	10	0.79 [0.77 – 0.98] ****		

Table S10. Model Results with [95% confidence intervals] and post-hoc comparisons of biogeographic regions in a three-level random effects model of correlation coefficients and random-effects model of abundance-occupancy beta-coefficients. Significance values **** $p<0.0001$

Biogeographic Region	k	Fisher's-Z [95%CI]	k	β [95%CI]
Afrotropics	11	0.77 [0.43 – 1.12] ****		
Nearctic	82	1.00 [0.86 – 1.14] ****	18	0.36 [0.26 – 0.47] ****
Neotropics	26	0.99 [0.82 – 1.17] ****		
Palearctic	101	0.82 [0.72 – 0.92] ****	31	0.28 [0.21 – 0.35] ****

Table S11. Model results with [95% confidence intervals] and post-hoc comparisons abundance-occupancy relationship statistical transformation conducted by the study author(s) as a moderator of Fisher's-Z. Significance values ****p<0.0001

Transformation	k	Fisher's-Z [95%CI]
log-log	22	0.95 [0.76-1.13]****
log-logit	34	0.88 [0.71-1.05]****
log-none	75	0.76 [0.59-0.94]****
none-none	16	0.89 [0.70-1.08]****

Comparison	Fisher's-Z Difference
log-logit vs log-log	-0.07
log-none vs log-log	-0.18
none-none vs log-log	-0.06
log-none vs log-logit	-0.12
none-none vs log-logit	0.01
none-none vs log-none	0.13

Table S12. Model results with [95% confidence intervals] and post-hoc comparisons abundance-occupancy relationship of vertebrates from marine, freshwater, or intertidal environments as a moderator of Fisher's-Z. Significance values ****p<0.0001

Realm	k	Fisher's-Z [95%CI]
freshwater	31	0.94 [0.72 -1.15]****
intertidal	4	1.33 [0.86 -1.80]****
marine	64	1.11 [0.87 -1.36]****

Comparison	Fisher's-Z Difference
intertidal vs freshwater	0.39
marine vs freshwater	0.17
marine vs intertidal	-0.22
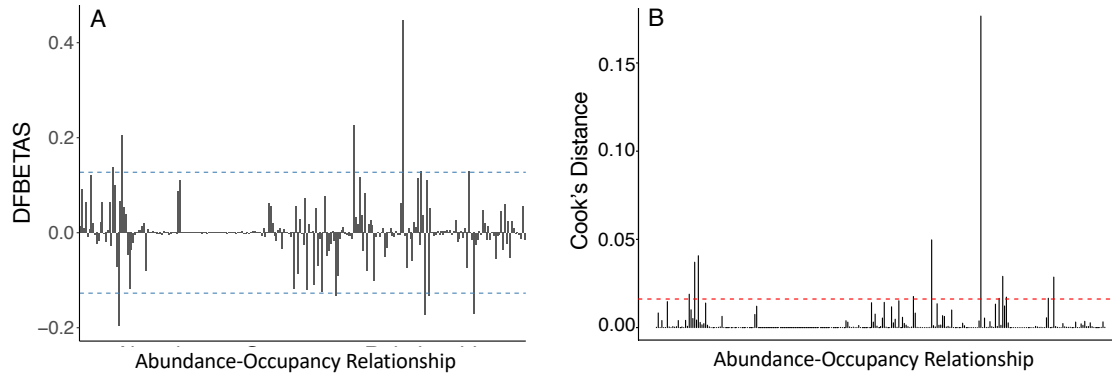
Figure S1. Identification of outliers and highly influential abundance-occupancy correlation coefficients (Fisher's Z). Each bar represents one abundance-occupancy correlation coefficient. A) abundance-occupancy relationships above or below dashed line identified at $\pm\frac{2}{\sqrt{n}}$ where n = number of abundance-occupancy relationships were identified as outliers. B) abundance-occupancy relationships above red dashed line at $\frac{4}{n}$ were identified as highly influential studies.
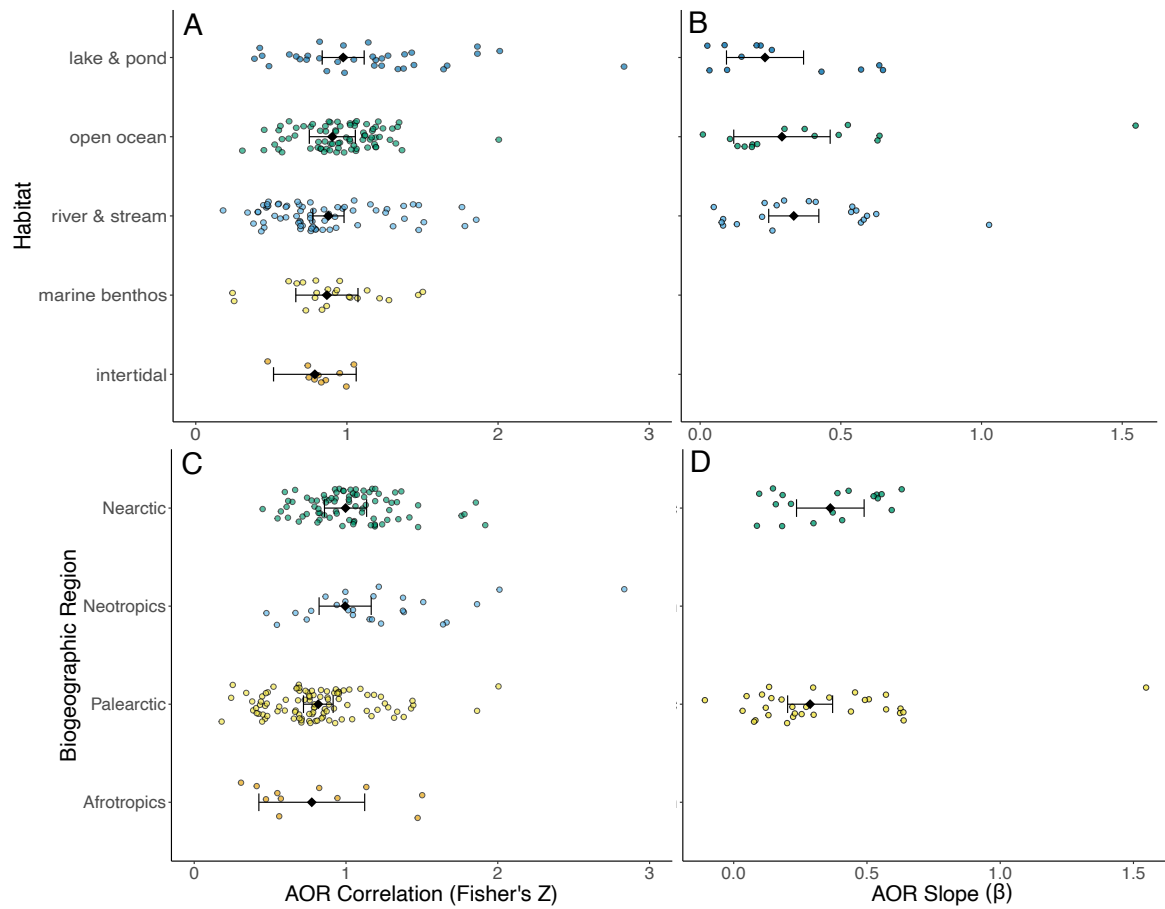
Figure S2. Subgroup model results for abundance-occupancy correlation coefficients and slopes among taxonomic groups and aquatic realms. Each point is one abundance-occupancy relationship. Diamond points are the average pooled effect size and error bars are 95% confidence intervals of model estimates. Points jittered on y-axes for clarity.
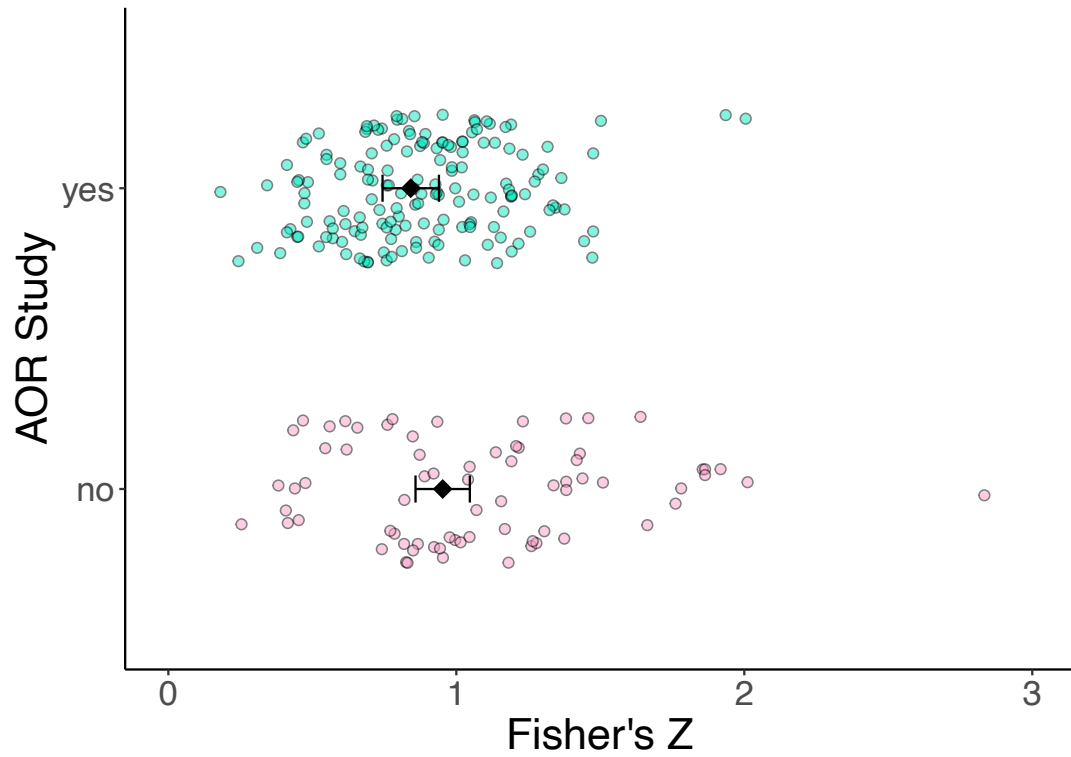
Figure S3. Fisher's Z-scores of abundance-occupancy relationships from n=163 effect sizes calculated by the study author(s) and n=73 effect sizes not calculated by the study author(s). Diamond is the pooled effect size from three-level random effects model for each group, error bars are 95% confidence intervals. Points jittered on y-axis for clarity.
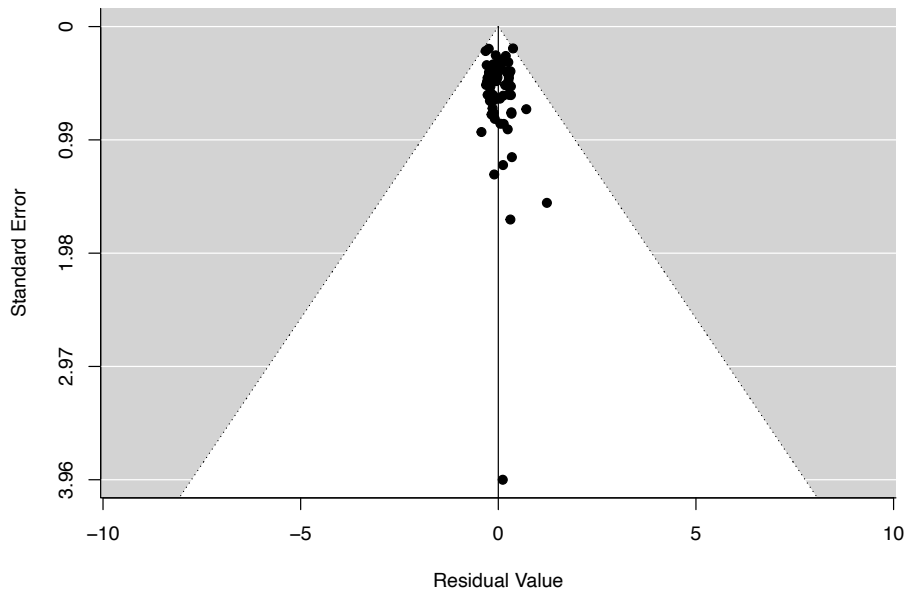
Figure S4. Funnel plot for evaluation of potential publication bias in beta-coefficients of abundance-occupancy relationships. Each point is one abundance-occupancy β-coefficient. Symmetry of points around the residual value 0 indicates no publication bias.

CHAPTER 2: ABUNDANT EVERYWHERE OR ABUNDANT SOMEWHERE: THE TROPHIC, GENETIC, AND HABITAT DIFFERENCES OF A COMMON AND A RARE COPEPOD

Rachel L. Wilkins[1,2]

1. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York

2. eCornell, Office of External Education, Cornell University, Ithaca NY

Abstract

Commonness is rare in nature. Very few species have large geographic ranges, occur in a variety of environmental conditions, and have large population sizes. To examine the underlying mechanisms that make a common species so distinct from a rare one, I measured the habitat specificity, trophic position, and population genetic structure of two closely related species of freshwater copepods with different levels of occurrence, *Leptodiaptomus minutus* (Lilljeborg 1889) and *Aglaodiaptomus leptopus* (Forbes 1882). Although both species occur over a wide geographic range in lakes across the Northern United State and Southern Canada, *A. leptopus* occurs much less frequently than *L. minutus*. To understand the drivers of the disparity in occurrence between these two species, I sampled nine lakes in the Adirondack Region, NY, USA, that fall within their geographic ranges. *L. minutus* occurred in a range of lake conditions, from shallow stained lakes with high productivity, to deep clear lakes with low productivity, whereas *A. leptopus* was only found in small lakes with low levels of hypolimnetic oxygen. There was no difference in trophic position between the two species, however I found evidence

that their trophic position varied according to lake productivity, having a more omnivorous or predatory diet in lakes with low concentrations of algal biomass. Populations of *L. minutus* had more genetic admixture, and less genetic differentiation than *A. leptopus* populations. I detected candidate loci under selection related to lake productivity using genotype-environment analyses, suggesting there is a genetic basis for adaptation to a variety of lake environments contributing to the success of *L. minutus* across a wide geographic area and diverse lake environments. Differences in level of occurrence between these two species is therefore not due to a difference in diet, but may be due to different levels of dispersal, and local adaptation to a wide range of conditions.

Introduction

In any given habitat, typically only a few species account for the majority of individuals and a high proportion of the total biomass in a community (Gaston and Fuller 2008). Ecosystem processes are often determined by the traits of those common taxa that contribute the most to biomass (Grime 1998, Geider et al. 2001, Winfree et al. 2015). Species often considered "common" have a combination of large local population size, wide habitat generality, or occurrence in a range of different habitat conditions, and have large geographic ranges (Rabinowitz 1981). According to Rabinowitz's framework, there are seven different types of rareness defined by eight possible combinations of population distribution characteristics: These are large and dominant somewhere vs. small and non-dominant, wide vs. narrow habitat specificity, and large vs. small geographic range size. Each combination of these categories is

considered rare, except for truly common species which have large geographic range, wide habitat specificity and large population size where they occur. While many species fit into one of the seven types of rareness Rabinowitz describes, very few fit the common criteria. Commonness is in fact a rare phenomenon in nature (Orme et al. 2006, Gaston 2011, Godet et al. 2015).

There may be several traits that distinguish rare species from common species, and understanding the difference between rare and common species requires an understanding of drivers of both abundance and occurrence (also termed site occupancy). In this study, I focus on understanding the causes of difference in in one kind of commonness – the level of occurrence (the fraction of sites where they are found) between a rare and common species that have similar levels of local abundance where they do occur.

Dispersal ability is often a trait exhibited by common species. For example, in aquatic insects, wing length is correlated with occupancy, suggesting one key characteristic of common species is longer wing length, thus increased dispersal ability (Malmqvist 2000). In studies from microorganisms to fish, species that are a smaller size (another proxy measure for dispersal ability) tend to occur in more locations (Wilkinson 2001, Tales et al. 2004). Dispersal becomes particularly important when favorable environmental conditions are not continuous on the landscape. Isolated habitat patches can be surrounded by inhospitable environments where species are not able to colonize, such as islands surrounded by ocean for terrestrial organisms and lakes surrounded by land for aquatic organisms. Distribution across these inhospitable environments can be achieved though active dispersal so that

the difference between rare and commonly occurring species may be due to dispersal ability.

Many zooplankton living in lakes and ponds readily disperse passively between habitats (Bilton et al. 2001). For organisms like zooplankton that depend upon transport by birds or wind, the production of resistant diapause eggs may facilitate dispersal (Cohen and Shurin 2003, Simonis and Ellis 2014, Hessen et al. 2019), although dispersal events over-land via wind are typically infrequent and of short distance (Cohen and Shurin 2003, Moreno et al. 2016, Sirianni 2017). If a species has the ability to disperse across inhospitable areas, successful establishment may nevertheless be prevented due to poor habitat match, Alee effects, hybrid incompatibility, and reproductive interference among congeneric species (Thum 2007) so that the new locations it reaches may not be suitable enough for a population to maintain a positive population growth rate. As a result, dispersal ability is one trait of closely related taxa that could distinguish common species from rare ones. Populations can even be maintained in locations outside of optimal conditions via high dispersal from source populations (Hanski 1994, Pulliam 2000).

While dispersal ability may play a role in an organism's chance at arriving at a new potentially habitable location, large populations enhance the chance of a dispersal event, and as the number of locations where a species occurs across a landscape increases, the chances of a dispersal event also increases (Hanski 1994, Gaston et al. 1997). This may be related to the size of the organisms, for example within micro-organisms, smaller taxa tend to have larger population sizes and therefore may be less vulnerable to local extinction (Wilkinson 2001, Finlay and Fenchel 2004). The

ubiquitous relationship between abundance and occupancy in nature (Chapter 1, Blackburn et al. 1997) suggests that population abundance plays a role in the level of occurrence across the landscape. Locally abundant species are more likely to have high occupancy if they have a lower extinction rate, a higher colonization rate than a rare species, or both (Gotelli 1991). In the absence of other factors, abundance is often the most reliable to explain occurrence (Faulks et al. 2015).

Common and rare species may also be different in their diet, or occur across different ranges of abiotic conditions, broadly defined as having different niches. Common and rare species may differ in their niche breadth, or the range of diet or abiotic conditions in which they occur. Common species are expected to differ along just a few of their niche axis, such that common species have a wider niche breadth than rare ones: they may eat a wider range of diet items, or occur in a wider range of environmental conditions (Brown 1984). A wider diet range may be explored by examining trophic position. Populations of common species may exhibit a more diverse, omnivorous, diet than populations that specialize in either grazing (or filter feeding) on primary producers, or by being exclusively predators. Since omnivores consume resources from several different trophic levels, I would expect common species to have trophic positions indicating that they are consuming both primary producers and primary or secondary consumers. Differences in niche breadth may explain differences in the abundance of species (Hurlbert and White 2007, Faulks et al. 2015), but it only more rarely explains differences in occurrence (e.g., Gregory and Gaston 2000, Frost et al. 2004). Occurrence is more often found to be correlated with niche position than niche breadth: that is, how often the species' set environmental

requirements are met across a landscape, regardless of the actual range of conditions where it is found. This has been shown to be true across a variety of organisms such as trees (Díaz et al. 2020), fishes (Rocha et al. 2018), invertebrates (Heino and Grönroos 2014), and algae (Soininen and Heino 2005, Vilmi et al. 2019).

Abiotic constraints, dispersal, niche requirements, and biotic interactions all influence species distributions (Wiens 2011). However, no single factor operates independently, and ultimately a species' distribution and its local abundance are determined by a combination of intrinsic species-specific, environmental, and historical factors that all interact to influence its distribution and abundance (Brown et al. 1996, Gregory and Gaston 2000). It was my goal in this study to discover which traits distinguish a representative common species from a closely related representative rare one by comparing habitat specificity, trophic position, and population genetic structure for two planktonic freshwater copepods in the family Diaptomidae. *Leptodiaptomus minutus* (Lilljeborg 1889) and *Aglaodiaptomus leptopus* (Forbes 1882) both occur across the Adirondack Mountains, New York State, USA. This region has a complex geology comprised of a diverse mosaic of bedrock due to glacial influence, and contains over 3,000 lakes, of a wide range of sizes and environmental conditions (Driscoll et al. 1991). Species within the genera *Leptodiaptomus* and *Aglaodiaptomus* are both within the North American clade of Diaptomidae, with *Aglaodiaptomus* closer to the ancestral node in the diaptomid phylogeny (Thum 2004). Among the thousands of Adirondack lakes, *L. minutus* is one of the most frequently occurring species of crustacean zooplankton, with the largest population sizes, while in contrast, *A. leptopus* only occurs in a few lakes, but has

large population sizes where it does occur (Figure S1) (Siegfried et al. 1989, Leach et al. 2018). According to Rabinowitz's topology of rarity, *L. minutus* is defined as common, while *A. leptopus* meets the criteria for a "predictable" rare species (large geographic range, narrow habitat specificity, large local population size). This difference in distribution is representative of the geographic occurrence of the two species more broadly. Based on a literature survey of 459 publications from years 1882-2001, the geographic ranges of *A. leptopus* and *L. minutus* are both relatively large occurring throughout the northeastern United States and Canada (Andrew Robertson, *pers. comm.*), with the range of *A. leptopus* even extending down into the southeastern United States, and in a comprehensive survey of 499 Wisconsin lakes where *L. minutus* was found in 112 and *A. leptopus* in only six (Torke 2001). According to Rabinowtiz's definition of rarity, I sought to examine what drives this difference in occurrence, for two species that have similar geographic range sizes and large populations sizes.

As the name suggests, *L. minutus* is one of the smallest members of the North American Diaptomidae (ca. 0.78-1.04 mm total length, 28.9 μg wet weight) (Leach et al. 2018), a characteristic that may contribute to its ability to persist in a variety of lake environments, whereas *A. leptopus* is larger (ca 1.5-2.4 mm total length, 70.5 μg wet weight) and therefore more vulnerable to visual predators such as zooplanktivorous fish. Shallow, fishless lakes tend to have higher abundances of *A. leptopus* than those with fish (Torke 2001), although the presence of fish does not always result in the exclusion of *A. leptopus*: salmonids in particular do not exclude this species (Donald et al. 2001). Nor is *A. leptopus* present in all fishless lakes in regions where it is

known to occur, due at least in part to presence of other visually hunding predators such as salamander larvae (Keller and Conlon 1994, Malkin et al. 2006). After fish had been removed from lakes where they were previously introduced, recolonization of *A. leptopus* populations is very slow (Donald et al. 2001), especially if its diapausing egg bank had been depleted during the period that zooplanktivorous fish were present (Parker et al. 1996). Non-visually feeding predatory fly larvae, *Chaoborus* spp. are known to prey on both *L. minutus* and *A. leptopus*, but *A. leptopus* is preyed upon at a lower rate (Blais and Maly 1993), presumably due to gape limitation in the larvae.

In other systems, predation has been implicated in limiting the geographic range and habitat preference of diaptomid copepods (Elmore 1983, Torke 2001). Previous work has also suggested that species with narrow habitat preferences are better at avoiding predation than those found in a broad range of habitats, although this may also come at the cost - in this case lower resource use efficiency (McPeek 1996). Copepods have specific predator avoidance adaptations such as diel vertical migration (Gliwicz 1986), which at the same time comes at a cost since individuals experience lower temperatures and lower food concentrations near the lake bottom, resulting in slower growth rate and lower reproductive output (Loose and Dawidowicz 1994). These general features could contribute to the observed distribution differences between *A. leptopus* and *L. minutus* since the former is more vulnerable to fish predation (Arnott and Vanni 1993, Keller and Conlon 1994, Donald et al. 2001, Torke 2001). While exclusion by fish is a possibility, there is no evidence that interspecific competition from other crustacean zooplankton excludes *A. leptopus* (Olenick 1983).

To gain a fuller understanding of the difference in occurrence patterns between *L. minutus* and *A. leptopus,* it is important to assess the environmental and genetic underpinnings of the two distinct patterns of occurrence. Exclusion by zooplanktivors doesn't fully explain the patchy occurrence of *A. leptopus*, and the additional causes of rareness in this species, namely limited dispersal ability, and narrow niche (biotic and abiotic in terms of food sources) could contribute to, or be the cause of, their patchy distribution. Additionally, difference in susceptibility in fish predation doesn't fully explain how *L. minutus* is able to exist in so many lakes, and at such high population densities.

My study consists of nine Adirondacks lakes (Figure 1) all of which contain populations of *L. minutus,* but only three of which contain populations of *A. leptopus,* a pattern representative of Adirondack lakes more generally (Sutherland 1989, Leach et al. 2018). I estimated the range of environmental conditions where each species occurred by measuring lake characteristics. Using stable isotopes, I measured each population's trophic position to examine the differences in diet between the common and rare species. To determine the degree to which each population is isolated from others, and how that pattern differs between the two species, I assessed genetic differences among populating using next generation reduced representation sequencing to generate single nucleotide polymorphisms (SNPs).

I show that *L. minutus* occurs in a wider range of lake conditions than *A. leptopus,* but that there is no difference in trophic position between the two species, both being omnivorous and adjusting their trophic position based on the quality of the seston in each lake. Populations of *L. minutus* exhibited more genetic admixture than

A. leptopus while also exhibiting a signature of genetic isolation-by-distance. Additionally, genomic analyses reveal SNPs in *L. minutus* populations associated with environmental gradients of algal biomass, seston quality, and pH among the lakes. I evaluate the extent to which these differences help explain the difference in occurrence between these two species, one common and the other rare across the Adirondack region.

Methods

Lake Sampling

I chose nine lakes for study across the Adirondack State Park, New York, USA (Figure 1) based on known populations of *L. minutus and A. leptopus* (Leach et al. 2018), to represent a range of lake conditions found throughout the Adirondacks. The mean ± 1 standard deviation in lake depth and surface area in our study was 9.7 m ± 7.2 m and 55.6 ha ± 74.3 ha., respectively. This range is similar to the larger Adirondack Lake Survey where mean lake depth was 11.3 m ± 6.8 m, and surface area was 60.6 ha ± 104.3 ha. Helldiver, Hoel, Limekiln, and Round Lakes were sampled three times (in late May, early July, and late August 2014). Falls, Indian, Squaw, West, and Wolf Lakes were sampled two times (in early July and late August 2014). At the deepest point in each lake, I measured the thermal profile and pH with an *in situ* YSI multiprobe sonde at 0.5 m intervals between the surface and 20 m or to 0.5 m above the sediment in lakes less than 20 m deep. Water samples were collected from the epilimnion and hypolimnion using a high-volume peristaltic pump to capture the entire depth range of each layer. Water samples were stored on ice in 3 L low density

polyethylene containers until being processed in the laboratory. Prior to laboratory processing and analysis, all water samples were filtered with a 75 μm mesh sieve to remove large zooplankton. Three replicates from each layer of each lake were transferred to acid washed high density polyethylene (HDPE) bottles and frozen until total nitrogen (TN) and particulate phosphorus (PP) was measured in the laboratory (I also analyzed soluble reactive phosphorus, but all values were below the limit of detection (0.025 μM). I did not measure total phosphorus, however since soluble reactive phosphorus was below the limit of detection, PP likely accounts for a large portion of total phosphorus in the lakes. Water samples for dissolved organic carbon (DOC) were stored in amber glass bottles at 4 °C until analysis. Both TN and DOC samples were collected only during July and August sampling events.

Seston and Water Chemistry Analysis

Seston from the epilimnion and hypolimnion were filtered onto a pre-combusted (500 °C for 4 hours) 25 mm Whatman GF/F glass fiber filters with effective pore size 0.7 μm, and stored at -20 °C until analysis. I measured seston phosphorus (PP) using a colorimetric assay with modifications to the color reagent using smaller volumes according to (Stainton et al. 1977). Absorbance was measured using a spectrophotometer at 880 nm. Chlorophyll-*a* was extracted with 90% ethanol and measured fluorometrically according to (Wetzel and Likens 1991). Filters containing seston for measuring seston carbon and seston nitrogen were dried at 60 °C for 48 hours and rolled into aluminum tins. The carbon and nitrogen content were measured using a Thermo Finnigan Flash 1112 elemental analyzer.

Dissolved organic carbon (DOC) was analyzed from the filtrate of lake water passed through a pre-combusted (500 °C for 4 hours) Whatman GF/F filter using a Shimadzu TOC auto-analyzer according to the manufacturer's instructions. Total nitrogen was measured on unfiltered water samples using an Astoria Pacific auto-analyzer using a modification of the Astoria Pacific standard methodology for cadmium reduction in freshwater samples (A173), adjusted to increase accuracy at concentrations below 1 μM (Hayn et al. 2014).

I calculated the average seston phosphorus, seston nitrogen, total nitrogen, dissolved organic carbon, and seston carbon for the whole water column in each lake by calculating the weighted mean based on the depth of the epilimnion and hypolimnion in each lake, since water samples for each layer were taken separately. I then calculated the average for each water chemistry parameter across the two or three summer sampling events.

Copepod Field Sampling

The nine populations of *L. minutus*, and three populations of *A. leptopus* in were sampled during July 2014 (Figure 1). Vertical plankton tows were taken at the deepest point of each lake with a 0.3 m diameter 75 μm mesh plankton net. Live zooplankton samples were stored on ice, and then rinsed and incubated in 1.2 μm filtered (Whatman GF/C glass fiber filter) lake water at room temperature for at least 30 min to permit gut clearance prior to preservation (Dam and Peterson 1988). Copepods were identified to species according to Reid and Williamson (2010).

Trophic position

To estimate trophic position of each copepod population, I measured their $\delta^{15}N$ isotope signatures, which become enriched with each trophic transfer within in a food chain (Deniro and Epstein 1981, Post 2002). Live zooplankton were individually picked, rinsed with distilled water, and frozen in scintillation vials until stable isotope analysis. Each vial contained between 30-100 individuals, depending on the body size of the species, to obtain enough biomass for analysis. Samples were lyophilized, ground, weighed, packed in tin, and stored in a desiccator at room temperature until analysis. $\delta^{15}N$ stable isotopes were measured on a Thermo Scientific Flash 2000 EA-IRMS at Eawag (Swiss Federal Institute for Aquatic Science and Technology).

Trophic position for each copepod population was calculated with a single trophic base according to Post (2002) using $\delta^{15}N$ isotope signatures. For each lake, I designated one of two indiscriminate filter feeding zooplankton *Daphnia* spp. or *Bosmina* spp. (depending upon which was present) as an integrated herbivore signature at trophic position 2. Trophic enrichment (Δn) of $\delta^{15}N$ in each lake was calculated as the difference between the filter feeding grazer and a co-occurring known obligate predator, either phantom midge larvae *Chaoborus* spp*.,* or the predatory copepod *Epischura* spp*.* For lakes without an obligate predator, I used the mean trophic enrichment factor $\Delta n = 1.727$ ‰ calculated from 6 nearby Adirondack lakes (Kjeldgaard et al. 2021, Supplement Table S1). Although 3.4‰ is often used as a trophic enrichment factor, trophic enrichment can be highly variable among lakes (Post 2002), and even variable depending on the food items consumed (Blanke et al. 2017). The calculation of trophic enrichment is made more precise by using a lake-

specific trophic enrichment factor calculated from co-occurring non-selective filter

feeding herbivore and an obligate zooplankton predator. I making the assumption with

this method that zooplankton grazers and predators are feeding along the same food-

chain, and the adult zooplankton predators are consuming the zooplankton grazers,

which has been observed for both *Epischura* (Wong 1981, Chow-Fraser and Wong

1986) and *Chaoborus* (Riessen et al. 1988). I did not mathematically correct the $\delta^{15}N$

signatures for lipids because $\delta^{15}N$ values are not systematically affected by lipids

accumulated in tissues (Logan et al. 2008).

I analyzed which environmental variates significantly correlate with trophic

position using multiple linear regression and AIC model selection. Seston N, Seston

C, Seston P, C:P, C:N, chlorophyll-*a*, DOC, and pH were all scaled in the linear

model. I used stepAIC with forward and backward model selection in the MASS

library (Venables and Ripley 2002) in R version 4.1.0 (Team 2021) to evaluate model

fit.

DNA Extraction and Sequencing

Entire zooplankton samples were concentrated with a 75 μm sieve, preserved

in 99% ethanol and stored at room temperature in 250 mL HDPE collection bottles

until DNA extraction. Individual adults were picked from zooplankton samples using

a sterile stainless-steel filiform needle and rinsed with 99% ethanol to remove any

algae or debris. Eggs were separated from females when present and discarded. In

total, 12 males and 13 females were isolated from each population for genetic analysis.

Copepods were stored individually in 1.5 mL microcentrifuge tubes at 0 °C for up to

24 hours until DNA extraction.

DNA was extracted from each individual with a Qiagen QIAamp DNA Micro Kit according to the manufacture's protocol for isolation of genomic DNA from less than 10 mg of tissue with carrier RNA to increase yield. Briefly, samples were lysed overnight in at water bath at 56 ℃, purified with QIAamp MiniElute columns, washed, and eluted with 40 μL of Buffer AE after a five-minute incubation period at room temperature. DNA was stored at -20 ℃ until digestion (1-23 days).

For each copepod, a double-digest restriction site associated (ddRadSeq) library was prepared according to (Peterson et al. 2012). Undiluted DNA extracts were digested with two restriction enzymes, SbfI-HF (New England Biolabs R3642), and MspI (New England Biolabs R0106). I used a two-tiered indexing scheme that allows for demultiplexing individuals after Illumina sequencing. Each population of both copepod species was split into two libraries, with 12-13 individuals per population in each library. Libraries were pooled and sequenced at the Cornell University Institute of Biotechnology on an Illumina single-end 75 bp NextSeq 500 using Illumina TruSeq indices.

Genotyping

I genotyped all copepods using the Stacks2 pipeline V.2.53 (Catchen et al. 2013). Samples from each sequencing run were first demultiplexed with process_radtags with -r to "rescue" RAD-tags and barcodes. Poor quality reads were removed with -c option to remove any reads with an uncalled base and -q to discard low quality reads. After demultiplexing I removed eight *L. minutus* individuals and

two *A. leptopus* individuals with a high number of reads with no rad-tags, a low number of retained reads, or both.

Before running the *de novo* pipeline on all samples, I optimized the number of raw reads required to form an allele (m), the maximum number of mismatches allowed between putative alleles (M), and the number of mismatches allowed between sample loci when building the catalog (n) using a subset of six individuals per population for *A. leptopus* and two individuals per population for *L. minutus* with the highest read coverage using the procedure outlined in Paris et al. (2017). For *L. minutus*, final stacks parameters were set to m = 4, M = 6, n = 7, and for *A. leptopus* final stacks parameters were set to m = 3, M = 6, n = 6.

Each of the programs under the core Stacks *de novo* pipeline (*ustacks, cstacks, sstacks, tsv2bam, gstacks, populations*) were run manually. After examining the per-locus coverage of all samples, individual samples that had high coverage and were representative of the genetic diversity in the dataset were selected to be included in the reference catalog. I included 15 individuals per population for *A. leptopus* and 10 individuals per population for *L. minutus*. Before running the sstacks program to match putative loci against the constructed catalog, samples with coverage less than 10× were removed. In total, the Stacks *de novo* pipeline genotyped 192,442 loci with mean effective per-sample coverage of 45.4× (st. dev. 25.1×) for *L. minutus,* and 271,227 loci with mean effective per-sample coverage of 71.3× (st. dev. 22.9×) for *A. leptopus*.

SNP Filtering

SNPs were called with the Stacks *population* program. For both species, I restricted loci to those occurring in at least seven of the nine *L. minutus* populations and two of the three *A. leptopus* populations. Loci were processed when they occurred in a minimum of 75% and 80% of individuals in each population for *L. minutus* and *A. leptopus,* respectively. To process a nucleotide site at a locus, minor allele frequencies were set to 0.05, and maximum observed heterozygosity of 0.70. I restricted data analysis to only the first SNP per locus. There was a higher frequency of SNPs at the beginning of the read sequences of *L. minutus* individuals, potentially due to sequencing errors. To resolve this, I filtered SNPs within the first 10 positions that were outside 1.5 times the interquartile range of SNP frequencies across the radtags (Polato et al. 2017). Additionally, individuals with more than 50% missing data were excluded from the dataset.

Population structure

To compare genetic distance among lakes, pairwise SNP-level measures of F_{ST} were calculated with the *populations* module of STACKS for every combination of sampling locations for both species (Catchen et al. 2013). I examined population genetic structure with discriminant analysis of principal components (DAPC) with the R library adegenet (Jombart and Ahmed 2011). I chose the number of clusters based on the lowest BIC. To avoid over fitting the model, I optimized the number of principal components included in the analysis through DAPC-cross validation procedures (Jombart and Ahmed 2011). Final DAPC analyses were conducted with 20

principal components for *L. minutus* and one principal component for *A. leptopus*.

Additional population structure was assessed using ADMIXTURE. I evaluated cluster

support to find the optimal number of clusters (K) with the greatest support through

ADMIXTURE's cross validation procedure (Alexander et al. 2009). I tested K = 1

through K = n + 2, where n is the number of populations for each species.

Genetic isolation and trophic level isolation by distance

      I examined the correlation of both the genetic distance and trophic level

distance among populations with geographic distance among lakes using a Mantel test

(Mantel 1967). I used mantel.rtest with 9999 permutations in the R package ade4

(Dray and Dufour 2007) to measure genetic isolation by geographic distance.

Genotype-environment association

      Redundancy analysis was used to determine how loci covary with

environmental conditions in each lake and identify potential loci under selection using

the R package psych (Revelle 2021) and vegan (Oksanen et al. 2020) according to

recommended procedures by (Forester et al. 2018). The matrix of alleles from *L.*

minutus contained 21% missing data, therefore missing genetic data were imputed

using the most common genotype at each SNP across all individuals. Environmental

variables that were correlated r > |0.7| were excluded from the analysis. Food

abundance (chlorophyll-*a*) and quality (seston C:N, and seston C:P) as well as water

quality indicators (hypolimnetic dissolved oxygen and pH) were included in the final

RDA analysis. I verified that variance inflation factors for each environmental

predictor were below five to ensure there was no multi-collinearity among variables. Significance of the full model and individual constrained axes were evaluated with a permutation test with $n = 999$ permutations in the R package vegan (Oksanen et al. 2020). Candidate loci under potential selection were identified as those with loadings greater than ± three standard deviations of the distribution of SNP loadings along the first five RDA axes (Forester et al. 2018). SNPs were evaluated based on the predictor with which they were most strongly correlated, and their position in ordination space. Due to the strong genetic structure in *A. leptopus* populations, and the small number of populations I surveyed (it is after all a rare species)*,* I had to exclude it from genetic and trophic isolation by distance analyses and genotype-environment association analyses.

Results

I chose nine study lakes that all contained populations of the common copepod *L. minutus*. The rare species, *A. leptopus,* in contrast, occurred in only three. Where *A. leptopus* was found, *L. minutus* also always occurred. This distribution is representative of the region: for example, in a survey of 28 Adirondack lakes, *L. minutus* occurred in 25 lakes and *A. leptopus* occurred in seven (Leach et al. 2018). Despite the difference in occurrence between the two species, there was little difference in the environmental characteristics of the lakes in which they were found. There was no statistically significant difference in the mean levels of chlorophyll-*a,* DOC, seston P, seston N, seston C:P and C:N, total nitrogen, pH, surface area, or depth between lakes where *A. leptopus* was present and where it was absent. The only

environmental characteristics that was distinctly different was hypolimnetic dissolved oxygen which was significantly lower for the three lakes where *A. leptopus* occurred compared with the six where it was absent (Student's t-Test, $p = 0.024$, Figure 2).

Among the nine lakes sampled, I detected a suite of correlated environmental variables indicative of a potential gradient in productivity. Positive correlations between chlorophyll-*a*, dissolved organic carbon, and seston phosphorus indicate some lakes are highly stained and have more standing algal biomass than other lakes that are more oligotrophic (Figure 3, Table 1). These Adirondack lakes also varied in watercolor, from very stained lakes with high dissolved organic carbon (DOC), to lakes with low levels of DOC and high water transparency (Table 1). Seston quality, as assessed by the ratio of nutrients, was negatively correlated with total N: lakes with higher Total N tended to have seston of lower quality (higher C:N, $r = 0.68$), but Total N was unrelated to seston P (PP) (Figure 3). The N:P of the lakes ranged from 4.3 to 15.6, suggesting that they are phosphorus limited (Downing and McCauley 1992).

In addition to the gradient in standing phytoplankton biomass (chlorophyll-*a*) and DOC among lakes, seston quality, expressed as C:N and C:P were closely associated, and opposite to pH in ordination space, despite the lack of statistically significant correlation (Figure 3, Figure 4). Acidic lakes tended to have lower quality seston, while lakes with higher pH had higher quality seston (lower seston C:N and C:P). The quality of the seston was more closely associated with the pH of the lake, and less closely associated with productivity. Compared with the range of nutrient ratios of different algal resources, the C:N and C:P ratios I measured in these lakes was still high compared with the range of nutrient ratios found in freshwater

environments generally (Sterner and Elser 2002). Vectors for chlorophyll-*a* and hypolimnetic dissolved oxygen are opposite to one another along at least one RDA axis, illustrating the pattern among lakes that lakes with low levels of standing algal biomass typically have more oxygenated hypolimnetic zones, and lakes with higher standing algal biomass had hypolimnetic zones with lower levels of oxygen.

Trophic position

Both *L. minutus* and *A. leptopus* had $\delta^{15}N$ signatures indicating an omnivorous or predatory diet, rather than being purely herbivorous (Figure 5). The mean trophic position of *L. minutus* populations was 2.61 ± 0.64 (s.d), while the mean trophic position of *A. leptopus* populations was 3.20 ± 0.82 (s.d.). Although the mean trophic position value for the larger copepod species *A. leptopus* was higher than for *L. minutus,* the difference was not statistically significant (t = -0.77, d.f. = 1, *p* = 0.47). *L. minutus* trophic position ranged from a value very similar to co-occurring obligate grazers (i.e., trophic position two) to trophic position three, indicating a predatory diet. While I do not have an explanation for the unexpectedly low tropic position near one for *L. minutus* in Helldiver Lake, I chose not to exclude it from our data. Trophic positions below 2 indicate *Daphnia* had higher $\delta^{15}N$ signatures than the co-occurring *L. minutus*. It is possible that *L. minutus* is feeding within a different food chain than *Daphnia,* causing the calculation of the base of the food chain to be invalid if *L. minutus* and *Daphnia* are feeding on different diets such as components of the microbial chain starting with methanogens versus phytoplankton since Helldiver Lakes has a markedly anoxic hypolimnion. The final multiple regression linear model

of trophic positions of both *L. minutus* and *A. leptopus* and the eight environmental variables in our data set, supported by the lowest AIC, included seston N, seston P, C:N, DOC, and chlorophyll-*a* (adjusted R^2 = 0.683, *p* = 0.045, Table 2). Average pH, seston C, and C:P were excluded from the final model (Table 3). Hypo DO was not included in model selection due to multi-collinearity.

Despite the range of trophic positions exhibited by *L. minutus* populations, there was no spatial pattern in trophic position among lakes. Geographic distance was not statistically significantly correlated with difference in trophic position among populations (Mantel test, *r* = -0.26*, p* = 0.94). Copepods from lakes that were located closer together geographically did not have more similar trophic positions to one another.

Population structure

Observed heterozygosity was lower than Hardy-Weinberg expectation for populations of both species, indicating that they exhibit some degree of population genetic structure (Figure S2). As expected, *L. minutus* had less structure and more admixture among populations than *A. leptopus.* According to cross-validation error, K = 7 followed closely by K = 5 population clusters were best supported for the nine populations of *L. minutus* (Figure 6A, Figure S3). At K = 5, the populations from Limekiln and Round lakes had shared ancestry, as did populations from West, Falls, Wolf, and Indian lakes. At K = 7, two populations in proximity, Falls and Wolf had shared ancestry, while populations from West and Indian had shared ancestry although they are not nearby geographically (Figure 1). Among the three populations *A.*

leptopus, there was the greatest support for K = 3 population clusters with virtually no admixture among them (Figure 7A, Figure S4).

According to discriminate analysis of principal components (DAPC), many, but not all lakes that are close together geographically also clustered close together genetically, supporting the ADMIXTURE results. Populations from Indian, Falls, and Wolf clustered close together, and had shared ancestry (Figure 6B). Similarly, Limekiln, Round, and Hoel lakes were distinct, and were the lakes that were the furthest from one another geographically. Squaw was genetically distinct according to ADMIXTURE, and was the furthest from other lakes in multivariate space according to DAPC (Figure 6B), although Hoel was the greatest distance geographically from other lakes. Among *L. minutus* populations, genetic distance increased significantly with geographic distance (mantel test, n=9999 replicates, r = 0.33, p = 0.017).

Based on DAPC cross-validation procedures, a single principal component axis was sufficient to describe the genetic variation among *A. leptopus* populations (Figure 7B). Individuals from Squaw and Falls were clustered closer together along the first discriminate axis than individuals from West, likely due to the larger geographic distance between West, and Squaw and Falls (Figure 1), although there were only four admixed individuals among lakes (Figure 7A). Pair-wise F_{ST} values were much lower among populations of *L. minutus* than *A. leptopus*, again indicating greater genetic differentiation among *A. leptopus* populations than *L. minutus* populations (Table 4).

Genotype-environment associations in <u>L. minutus</u> populations

I was unable to conduct a genotype-environment association study for *A. leptopus* due to a lack of statistical power since there were only three populations (again, it is rare). As a result, I confine my analysis to *L. minutus.* I found a significant relationship between SNPs of *L. minutus* and five of the environmental predictors: hypolimnetic dissolved oxygen, pH, seston C:N, seston C:P, and chlorophyll-*a*. When included in the RDA analysis of genotype-environment associations, these five were statistically significant ($p < 0.001$) and explained about 10% of the genetic variation among *L. minutus* populations (adjusted $R^2 = 0.100$). Separately, each of the five constrained RDA axes were statistically significant ($p < 0.001$, Table S2).

According to the redundancy analysis, *L. minutus* individuals from several lakes were easily separated in ordination space along environmental vectors. Individuals from Helldiver Lake, the lake with the highest standing algal biomass and among the highest nutrient levels and hypolimnetic dissolved oxygen (Table 1), were easily distinguishable from individuals from Hoel Lake, one of the more oligotrophic lakes, along RDA2 (Figure 4A). Individuals from lakes that were less acidic, namely Round and Limekiln, were also distinct genetically and were located toward the higher end of the vector representing the range of pH along RDA1, opposite of acidic lakes such as Indian, Falls, and Wolf (Figure 4). Individuals from Squaw lake were also easily distinguishable from other individuals along RDA2 and RDA3 (Figure 4A), consistent with their distinct genetic clustering that I observed along linear discriminant (LD) axis 2 in the DAPC analysis (Figure 6B). Individuals from Indian, Falls, and Wolf lakes were clustered together and indistinguishable from one another

along all three RD axes. Individuals from these three lakes tended to be located along vectors of high C:N and high C:P, low pH, and low hypolimnetic dissolved oxygen (Figure 4).

In total, I detected 51 SNPs potentially under selection that were associated with lake standing biomass, seston quality, and acidity. Many SNPs were most strongly correlated with environmental predictors associated with lake trophic state such as chlorophyll-*a* (12 SNPs) and hypolimnetic dissolved oxygen (5 SNPs). Sixteen SNPs were most strongly correlated with pH, while 18 SNPs were associated with food quality (nutrient ratios) of the seston (12 SNPs correlated with C:N, and 6 SNPs correlated with C:P). Along RDA1, I detected multilocus sets of SNP genotypes associated with seston quality (seston C:N, and seston C:P), while on RDA2, SNPs were associated with chlorophyll-*a* (phytoplankton abundance). On RDA3 I detected multilocus sets of SNP genotypes most strongly associated with lake acidity (Figure 8).

Discussion

Throughout the northern United Sates, including the Adirondack region of New York State, *L. minutus* is a common copepod, found in high abundance in a large proportion of lakes. In contrast, populations of the related copepod *A. leptopus* are rare, found in relatively lakes in the region, although average population abundance tends to be high where they do occur (Leach et al. 2018). I contrasted the trophic position, population genetic structure, and lake environmental conditions where each species is found to understand the factors that contribute to the differences between

these two species, one common and one rare. I tested whether a broad generalist diet, tolerance to a wide range of lake conditions, and a high level of population connectance contribute to the observed distribution differences.

Evidence of higher geneflow among populations of the common copepod L. minutus than the rare copepod A. leptopus – I found signatures of higher gene flow among lake populations of *L. minutus* than I did for *A. leptopus*, with higher levels of admixture (Figure 6A) and lower relative pairwise F_{ST} values (Table 4), suggesting *L. minutus* faces fewer barriers to dispersal. Still, there was clear genetic structure among *L. minutus* populations, with significant isolation by distance. Therefore, despite the population connectedness I detected, this species is nevertheless dispersal limited. In contrast, there was virtually no admixture among the three populations of *A. leptopus*, and higher pairwise F_{ST} values among populations compared with *L. minutus* (Figure 7A, Table 4). This greater genetic isolation of *A. leptopus* populations indicates that they face greater barriers to dispersal, or are unable to inhabit many lakes across the region, and so is presumably tied to being rare, in the sense of occurring in relatively few lakes. The lower population structure observed in *L. minutus* populations is likely due to dispersal events among populations that do not occur as frequently among populations of *A. leptopus*. Therefore, the lower genetic structure observed in *L. minutus* is what I would expect to find in an abundant species found in many lakes, and so likely is a consequence of of being common.

One very likely method of dispersal for calanoid copepods is transport of diapausing eggs among lakes most likely attached to feathers or feet water birds, or

moved along with fish by anglers. Both species produce diapausing eggs (Watras 1980, Hairston and Brunt 1994, Torke 2001), but because *L. minutus* occurs in more lakes, I would expect there to be more opportunities for its eggs to be dispersed across the landscape. Based on the lack of admixture among populations of *A. leptopus*, it is clear that even if diapausing eggs are dispersed to nearby lakes, populations rarely become established. For example, *A. leptopus* occurs in both Falls and Squaw Lakes, which are approximately five km from one another, however in my population samples only a single individual from Falls Lake had a portion of shared genetic ancestry with the genetic cluster assigned to Squaw Lake, and two individuals from Squaw Lake had shared ancestry with those from Falls Lake (Figure 7A). Despite the potential for dispersal through diapausing eggs, establishment of individuals into a lake that already has an established population may be inhibited by a strong priority effect of the founding genotypes. In essence, rapid population growth and local adaptation when a water body is colonized may exclude new individuals from invading a resident population because of monopolization of resources and a large bank of diapausing eggs further strengthening the predominance of one genotype in a population, and preventing other genotypes from other populations from establishing (Boileau et al. 1992, Meester et al. 2002). Based on the very low admixture among populations, and the ability of *A. leptopus'* to produce diapausing eggs that could potentially be easily dispersed, the genetic isolation among populations may be due to priority effects of founding populations.

Among *L. minutus* populations, individuals from Helldiver and Hoel Lakes were assigned to separate genetic clusters and had relatively little admixture compared

with other populations (Figure 6A). Helldiver and Hoel Lakes are approximately 80 km apart (Figure 1), and are separated in environmental ordination space, especially along RDA2, which encapsulates a productivity gradient (Figure 4A). In this case, two lakes are separated geographically, have different levels of productivity, and the populations of *L. minutus* are also distinct genetically (Figure 6B). In contrast, individuals from Falls, Indian, and Wolf Lakes are clustered closely together in ordination space and are assigned the same ancestry (Figure 6A). These lakes are located close together geographically, therefore individuals are either able to disperse easily among the three lakes, or were colonized from a common source population. The similarity in environmental conditions of these lakes may also make it possible that genetic lineages that share adaptations to local environmental conditions are able to persist and become dominant.

Geographic proximity, however, is not the sole force shaping the genetic clustering of *L. minutus* populations. Squaw Lake and Indian Lake are only about 2 km apart, but individuals do not cluster together in ordination space and are assigned different ancestry groups (Figure 6A). Differences in the environmental characteristics, specifically productivity, between Squaw Lake and Indian Lake may be responsible for the genetic isolation observed. Squaw Lake has higher standing phytoplankton biomass and nutrient levels than Indian lake (Table 1). Lake productivity level has been shown to exclude calanoid copepods from lake environments, especially when food availability is low and competition is high (Elmore 1983). Among the *L. minutus* populations in this study, one third of the 51 candidate SNPs under selection were correlated with measures of lake productivity,

therefore productivity may also be shaping the structure of *L. minutus* populations if the genetic differences observed prevent individuals in one lake type from colonizing or establishing in another.

Trophic position flexibility in response to available food sources – The trophic signatures of *L. minutus* and *A. leptopus* are statistically indistinguishable (Figure 5). On average, *A. leptopus* and *L. minutus* both had δ^{15}N values above the co-occurring obligate herbivores *Daphnia* spp. or *Bosmina* spp., indicating an omnivorous to predatory diet for the copepods. Calanoid copepods are known for their ability to select of food items preferentially, which is made possible by their sophisticated chemo- and mechanosensory capabilities (Friedman and Strickler 1975, Buskey 1984, Steinke et al. 2006, Kiørbe 2008, Almeda et al. 2018). It seems likely that the trophic positions of the copepods were elevated by the selective consumption of planktonic heterotrophic ciliates and flagellates (Christoffersen et al. 1990, Faithfull and Goetze 2019). Such consumption can be substantial, ultimately influencing the composition of the microbial community in lakes (Burns and Schallenberg 1996, 2001). Additionally, diaptomid copepods can preferentially consume herbivorous rotifers (Williamson and Butler 1986, 1987, Arnott and Vanni 1993), which would also elevate their trophic position. Preferential consumption of heterotrophic ciliates and flagellates, as well as rotifers, essentially inserts a trophic level between phytoplankton and copepods, thereby making them more similar to primary predators such as the *Chaoborus* and *Epischura* that I used to establish the δ^{15}N signature for the third trophic level in my study lakes.

The concentration and ratio of nutrients (C:N:P) in a lake (i.e., its trophic state), influences the community composition and nutritional quality of phytoplankton and microbes found there (Hutchinson 1967, Watson et al. 1997, Sterner and Elser 2002, Moody and Wilkinson 2019). Diet generalists can survive and reproduce while consuming a wide variety of food resources, so a common species such as *L. minutus*, that occurs in lakes across the spectrum of lake trophic states, may consume the wide variety of resources that co-vary with lake productivity. In this study, model selection for a description of trophic position of both *L. minutus* and *A. leptopus* included seston N, seston P, C:N, DOC, and chlorophyll-*a* in the linear model with the most statistical support (Table 2). Model coefficients for seston N, C:N, and chlorophyll-*a* were all positive, indicating the copepods had a more predatory diet in lakes dominated by lower-quality phytoplankton. Seston phosphorus had a negative coefficient, suggesting copepods were more herbivorous in lakes with seston rich in phosphorus, and more predatory in lakes with low-P seston where they likely consume rotifers and heterotrophic ciliates and flagellates. Phosphorus is an essential element in the diet of calanoid copepods and all zooplankton (Sterner and Elser 2002), deficiencies can prevent successful development from the naupliar to the juvenile copepodid stage (Villar-Argaiz and Sterner 2002). Therefore, adjusting diet to meet minimum nutrient requirements could be the reason the final model describing copepod trophic position included seston P, indicating that copepods adjust their feeding behavior to meet minimum nutrient requirements. Given that *A. leptopus* is included in the model of trophic position, diet flexibility is a trait shared by both species and is likely not the reason for the difference in rarity versus commonness (i.e., occurrence) between

A. leptopus and *L. minutus.*

Adaptation of L. minutus to a variety of lake conditions – By affecting diet composition and behavior, lake trophic state has apparently driven natural selection acting on the copepods. In my analysis of genotype-environment associations in *L. minutus*, multi-locus genotypes were statistically related to correlates of lake productivity (Figure 8). One-third of candidate SNPs were most highly correlated with chlorophyll-*a* and hypolimnetic dissolved oxygen. Lakes with a large amount of standing phytoplankton biomass (high concentrations of chlorophyll-*a*) tended to have low hypolimnetic dissolved oxygen, likely due to nutrient release from decomposition of dead phytoplankton (Figure 4), and copepods from these lakes tended to feed at lower trophic levels, since trophic position was negatively correlated with DOC. Due to the nature of redundancy analysis, I could not examine all environmental correlates simultaneously. However, the SNPs that are correlated with measures of lake productivity (chlorophyll-*a* and dissolved oxygen) are likely also correlated with other environmental variables associated with chlorophyll-*a*, especially seston phosphorus. Therefore, candidate SNPs may be under selection based on the type of lake (low-productivity-clear vs high-productivity-stained) in which a copepod population resides as well as copepod trophic position via the type of diet they consume (rotifers, heterotrophic ciliates and flagellates versus photosynthetic algae) with a resulting genetic basis for copepod trophic position (Moosmann et al. 2021).

   Along with a gradient in productivity and water clarity, I also observed a gradient in seston quality and pH among the study lakes. Seston C:N and C:P were

opposite to pH along the first RDA axis, which accounts for almost thirty percent of the environmental variation (Figure 4). Lakes with high pH tended to have higher quality seston (lower C:N and C:P), and acidic lakes tended to have lower quality seston (higher C:N and C:P). Two-thirds of the *L. minutus* SNPs detected as potentially under selection were most highly correlated with pH, seston C:N, and seston C:P (Figure 8), suggesting that these variables may exert selection pressure on the copepods. Further evidence for local adaptation by diaptomid copepods to different environments is found in a study by Hausch et al. (2013) showing that population origin (lake) better explained morphological variation in the shapes of the prosome, urosome, and antennae than did species identity.

Relative vulnerability of A. leptopus *and* L. minutus *to predation* – In the end, it may be that the reason for the commonness of *L. minutus* is a more interesting question than the reason for the rarity of *A. leptopus*, simply because the latter, larger species is more vulnerable to predation by visually foraging fish. While the three lakes in our survey where *A. leptopus* was found were not completely fishless, the community of fishes in West, Squaw, and Falls Lakes consisted of Brook Trout, which are not strictly zooplanktivores (Werner 2004) and bottom-feeding Brown Bullhead (Table S3). The lakes where I found *A. leptopus* had significantly lower levels of hypolimnetic dissolved oxygen (Figure 2), suggesting greater oxygen demand leading to hypoxia and likely fish-kills under the ice during the winter, a finding consistent with the distribution of *A. leptopus* in lakes in Wisconsin (Torke 2001). In contrast to *A. leptopus,* it is clear that *L. minutus* can persist in lakes with zooplanktivorous fish,

including the smelt and sunfish present in our study lakes (Table S3). The small body size and fast swimming acceleration (Drenner and McComas 1980) of *L. minutus* may allow it either to evade detection and capture by fish and other visual predators, or make them less valuable as food when larger slow swimming taxa such as *Daphnia* spp. are present.

Lastly, more *L. minutus* individuals in the Adirondack region as a whole may provide a greater chance for genetic variation to be present to facilitate local adaptation, and for those adapted genotypes to make it into any given lake through dispersal. Even if there is little to no difference in dispersal ability between *L. minutus* and *A. leptopus,* there are simply many more populations from which dispersal can occur.

Conclusion

The common occurrence of *L. minutus* across a range of lake types in the Adirondack region of New York, USA, and across much of North America, may be a result of both its dispersal capabilities, and its ability to adapt to local lake environments. High genetic similarity of populations of *L. minutus* among lakes, compared to with the rare copepod *A. leptopus,* suggests that its success may partially be due to its ability to disperse to new lake environments, and to succeed in establishing a population by adjusting its diet based on the food resources available. I detected candidate SNPs under selection related to lake productivity, and thus potentially also diet, suggesting there is a genetic basis for adaptation to a variety of lake environments contributing to the success of *L. minutus* across a wide geographic

area and a diverse set of lake environments.

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Almeda, R., H. S. Gréve, and T. Kiørboe. 2018. Prey perception mechanism determines maximum clearance rates of planktonic copepods. Limnology and Oceanography 63:2695–2707.

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

Bilton, D. T., J. R. Freeland, and B. Okamura. 2001. Dispersal in freshwater invertebrates. Annual Review of Ecology and Systematics 32:159–181.

Blackburn, T. M., Kevin. J. Gaston, and R. D. Gregory. 1997. Abundance-range size relationships in British birds: is unexplained variation a product of life history? Ecography 20:466–474.

Blais, J. M., and E. J. Maly. 1993. Differential predation by *Chaoborus americanus* on males and females of two species of *Diaptomus*. Canadian Journal of Fisheries and Aquatic Science 50:410–415.

Blanke, C. M., Y. Chikaraishi, Y. Takizawa, S. A. Steffan, P. S. Dharampal, and M. J. V. Zanden. 2017. Comparing compound-specific and bulk stable nitrogen isotope trophic discrimination factors across multiple freshwater fish species and diets. Canadian Journal of Fisheries and Aquatic Sciences 74:1291–1297.

Boileau, M. G., P. D. N. Hebert, and S. S. Schwartz. 1992. Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. Journal of Evolutionary Biology 5:25–39.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. The American Naturalist 124:255–279.

Brown, J. H., G. C. Stevens, and D. M. Kaufman. 1996. The geographic range: Size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics 27:597–623.

Burns, C. W., and M. Schallenberg. 1996. Relative impacts of copepods, cladocerans and nutrients on the microbial food web of a mesotrophic lake. Journal of Plankton Research 18:683–714.

Burns, C. W., and M. Schallenberg. 2001. Calanoid copepods versus cladocerans: Consumer effects on protozoa in lakes of different trophic status. Limnology and Oceanography 46:1558–1565.

Buskey, E. J. 1984. Swimming pattern as an indicator of the roles of copepod sensory systems in the recognition of food. Marine Biology 79:165–175.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124–3140.

Chow-Fraser, P., and C. K. Wong. 1986. Dietary change during development in the freshwater calanoid copepod *Epischura lacustris* Forbes. Canadian Journal of Fisheries and Aquatic Sciences 43:938–944.

Christoffersen, K., B. Riemann, L. R. Hansen, A. Klysner, and H. B. Sørensen. 1990. Qualitative importance of the microbial loop and plankton community structure in a eutrophic lake during a bloom of cyanobacteria. Microbial Ecology 20:253–272.

Cohen, G. M., and J. B. Shurin. 2003. Scale-dependence and mechanisms of dispersal in freshwater zooplankton. Oikos 103:603–617.

Dam, H. G., and W. T. Peterson. 1988. The effect of temperature on the gut clearance rate constant of planktonic copepods. Journal of Experimental Marine Biology and Ecology 123:1–14.

De Meester, L. D., A. Gómez, B. Okamura, and K. Schwenk. 2002. The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. Acta Oecologica 23:121–135.

Deniro, M. J., and S. Epstein. 1981. Influence of diet on the distribution of nitrogen isotopes in animals. Geochimica et Cosmochimica Acta 45:341–351.

Díaz, D. M. V., C. Blundo, L. Cayola, A. F. Fuentes, L. R. Malizia, and J. A. Myers. 2020. Untangling the importance of niche breadth and niche position as drivers of tree species abundance and occupancy across biogeographic regions. Global Ecology and Biogeography 29:1542–1553.

Donald, D. B., R. D. Vinebrooke, R. S. Anderson, J. Syrgiannis, and M. D. Graham. 2001. Recovery of zooplankton assemblages in mountain lakes from the effects of introduced sport fish. Canadian Journal of Fisheries and Aquatic Sciences 58:1822–1830.

Downing, J. A., and E. McCauley. 1992. The nitrogen:phosphrous relationship in lakes. Limnology & Oceanography 37:936–945.

Dray, S., and A. Dufour. 2007. The ade4 Package: Implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

Driscoll, C. T., R. M. Newton, C. P. Gubala, J. P. Baker, and S. W. Christensen. 1991. Adirondack Mountains. Pages 133–202 *in* D. F. Charles, editor. Acidic Deposition and Aquatic Ecosystems. Springer, New York, NY.

Elmore, J. L. 1983. Factors influencing *Diaptomus* distributions: An experimental study in subtropical Florida. Limnology and Oceanography 28:522–532.

Faithfull, C., and E. Goetze. 2019. Copepod nauplii use phosphorus from bacteria, creating a short circuit in the microbial loop. Ecology Letters 22:1462–1471.

Faulks, L., R. Svanbäck, H. Ragnarsson-Stabo, P. Eklöv, and Ö. Östman. 2015. Intraspecific niche variation drives abundance-occupancy relationships in freshwater fish communities. The American Naturalist 186:272–283.

Finlay, B. J., and T. Fenchel. 2004. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 155:237–244.

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology 27:2215–2233.

Friedman, M. M., and J. R. Strickler. 1975. Chemoreceptors and feeding in calanoid copepods (Arthropoda: Crustacea). Proceedings of the National Academy of Sciences 72:4185–4188.

Frost, M. T., M. J. Attrill, A. A. Rowden, and A. Foggo. 2004. Abundance-occupancy relationships in macrofauna on exposed sandy beaches: patterns and mechanisms. Ecography 27:643–649.

Gaston, K. J. 2011. Common ecology. BioScience 61:354–362.

Gaston, K. J., T. M. Blackburn, and R. D. Gregory. 1997. Interspecific abundance-range size relationships: range position and phylogeny. Ecography 20:390–399.

Gaston, K. J., and R. A. Fuller. 2008. Commonness, population depletion and conservation biology. Trends in Ecology & Evolution 23:14–19.

Geider, R. J., E. H. Delucia, P. G. Falkowski, A. C. Finzi, J. P. Grime, J. Grace, T. M. Kana, J. L. Roche, S. P. Long, B. A. Osborne, T. Platt, I. C. Prentice, J. A. Raven, W. H. Schlesinger, V. Smetacek, V. Stuart, S. Sathyendranath, R. B. Thomas, T. C. Vogelmann, P. Williams, and F. I. Woodward. 2001. Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. Global Change Biology 7:849–882.

Gliwicz, M. Z. 1986. Predation and the evolution of vertical migration in zooplankton. Nature 320:746–748.

Godet, L., P. Gaüzere, F. Jiguet, and V. Devictor. 2015. Dissociating several forms of commonness in birds sheds new light on biotic homogenization. Global Ecology and Biogeography 24:416–426.

Gotelli, N. J. 1991. Metapopulation models: The rescue effect, the propagule rain, and the core-satellite hypothesis. The American Naturalist 138:768–776.

Gregory, R. D., and K. J. Gaston. 2000. Explanations of commonness and rarity in British breeding birds: separating resource use and resource availability. Oikos 88:515–526.

Grime, J. P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. Journal of Ecology 86:902–910.

Hanski, I. 1994. A practical model of metapopulation dynamics. The Journal of Animal Ecology 63:151.

Hausch, S., J. B. Shurin, and B. Matthews. 2013. Variation in body shape across species and populations in a radiation of diaptomid copepods. PLoS ONE 8:e68272.

Hayn, M., R. Howarth, R. Marino, N. Ganju, P. Berg, K. H. Foreman, A. E. Giblin, and K. McGlathery. 2014. Exchange of nitrogen and phosphorus between a shallow lagoon and coastal waters. Estuaries and Coasts 37:63–73.

Heino, J., and M. Grönroos. 2014. Untangling the relationships among regional occupancy, species traits, and niche characteristics in stream invertebrates. Ecology and Evolution 4:1931–1942.

Hessen, D. O., T. C. Jensen, and B. Walseng. 2019. Zooplankton diversity and dispersal by birds: Insights from different geographical scales. Frontiers in Ecology and Evolution 7:74.

Hurlbert, A. H., and E. P. White. 2007. Ecological correlates of geographical range occupancy in North American birds. Global Ecology and Biogeography 16:764–773.

Hutchinson, G. E. 1967. A Treatise on Limnology. II. Introduction to Lake Biology and the Limnoplankton. Wiley, New York.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 24:1403–1405.

Hairston, N. G., Jr. and R. A. Van Brunt. 1994. Diapause dynamics of two diaptomid copepod species in a large lake. Hydrobiologia 292:209–218.

Keller, W., and M. Conlon. 1994. Crustacean zooplankton communities and lake morphometry in Precambrian Shield lakes. Canadian Journal of Fisheries and Aquatic Science 51:2424–2434.

Kiørbe, T. 2008. A Mechanistic Approach to Plankton Ecology. Princeton University Press, Princeton, New Jersey.

Kjeldgaard, M. K., J. A. Hewlett, and M. D. Eubanks. 2021. Widespread variation in stable isotope trophic position estimates: patterns, causes, and potential consequences. Ecological Monographs 91.

Leach, T. H., L. A. Winslow, F. W. Acker, J. A. Bloomfield, C. W. Boylen, P. A. Bukaveckas, D. F. Charles, R. A. Daniels, C. T. Driscoll, L. W. Eichler, J. L. Farrell, C. S. Funk, C. A. Goodrich, T. M. Michelena, S. A. Nierzwicki-Bauer, K. M. Roy, W. H. Shaw, J. W. Sutherland, M. W. Swinton, D. A. Winkler, and K. C. Rose. 2018. Long-term dataset on aquatic responses to concurrent climate change and recovery from acidification. Scientific Data 5:180059.

Logan, J. M., T. D. Jardine, T. J. Miller, S. E. Bunn, R. A. Cunjak, and M. E. Lutcavage. 2008. Lipid corrections in carbon and nitrogen stable isotope analyses: comparison of chemical extraction and modelling methods. Journal of Animal Ecology 77:838–846.

Loose, C. J., and P. Dawidowicz. 1994. Trade-offs in diel vertical migration by aooplankton: The costs of predator avoidance. Ecology 75:2255–2263.

Malkin, S. Y., O. E. Johannsson, and W. D. Taylor. 2006. Small-bodied zooplankton communities yet strong top-down effects on phytoplankton in the absence of fish. Archiv für Hydrobiologie 165:313–338.

Malmqvist, B. 2000. How does wing length relate to distribution patterns of stoneflies (Plecoptera) and mayflies (Ephemeroptera)? Biological Conservation 93:271–276.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220.

McPeek, M. A. 1996. Trade-offs, food web structure, and the coexistence of habitat specialists and generalists. The American Naturalist 148:S124–S138.

Moody, E. K., and G. M. Wilkinson. 2019. Functional shifts in lake zooplankton communities with hypereutrophication. Freshwater Biology 64:608–616.

Moosmann, M., M. Cuenca-Cambronero, S. D. Lisle, R. Greenway, C. M. Hudson, M. D. Lürig, and B. Matthews. 2021. On the evolution of trophic position. Ecology Letters 24:2549–2562.

Moreno, E., C. Pérez-Martínez, and J. M. Conde-Porcuna. 2016. Dispersal of zooplankton dormant propagules by wind and rain in two aquatic systems. Limnetica 35:323–336.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, H. H. H. Stevens, E. Szoecs, and H. Wagner. 2020. vegan: Community Ecology Package. R package version 2.5-7.

Olenick, R. J. 1983. The effects of interspecific competition on *Diaptomus leptopus* in a montane lake: an experimental field study. Canadian Journal of Zoology 61:2273–2287.

Orme, C. D. L., R. G. Davies, V. A. Olson, G. H. Thomas, T.-S. Ding, P. C. Rasmussen, R. S. Ridgely, A. J. Stattersfield, P. M. Bennett, I. P. F. Owens, T. M. Blackburn, and K. J. Gaston. 2006. Global patterns of geographic range size in birds. PLoS Biology 4:e208.

Paris, J. R., J. R. Stevens, and J. M. Catchen. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8:1360–1373.

Parker, B. R., F. M. Wilhelm, and D. W. Schindler. 1996. Recovery of *Hesperodiaptomus arcticus* populations from diapausing eggs following elimination by stocked salmonids. Canadian Journal of Zoology 74:1292–1297.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Polato, N. R., M. M. Gray, B. A. Gill, C. G. Becker, K. L. Casner, A. S. Flecker, B. C. Kondratieff, A. C. Encalada, N. L. Poff, W. C. Funk, and K. R. Zamudio. 2017. Genetic diversity and gene flow decline with elevation in montane mayflies. Heredity 119:107–116.

Post, D. M. 2002. Using stable isotopes to estimate trophic position: Models, methods, and assumptions. Ecology 83:703–718.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters 3:349–361.

Rabinowitz, D. 1981. Seven forms of rarity. Pages 205–217 *in* H. Synge, editor. The Biological Aspects of Rare Plant Conservation. John Wiley & Sons, Chichester, UK.

Reid, J. W., and C. E. Williamson. 2010. Copepoda. Pages 829–899 *in* J. H. Thorp and A. P. Covish, editors. Ecology and Classification of North American Invertebrates. 3rd edition. Elsevier, Academic Press, Oxford, UK.

Revelle, W. 2021. psych: procedures for psychological, psychometric, and personality research. Version = 2.1.6. Northwestern University, Evanston, Illinoise, USA.

Riessen, H. P., J. W. Sommerville, C. Chiappari, and D. Gustafson. 1988. *Chaoborus* predation, prey vulnerability, and their effect in zooplankton communities. Canadian Journal of Fisheries and Aquatic Sciences 45:1912–1920.

Rocha, M. P., L. M. Bini, T. Siqueira, J. Hjort, M. Grönroos, M. Lindholm, S.-M. Karjalainen, and J. Heino. 2018. Predicting occupancy and abundance by niche position, niche breadth and body size in stream organisms. Oecologia 186:205–216.

Siegfried, C. A., J. A. Bloomfield, and J. W. Sutherland. 1989. Planktonic rotifer community structure in Adirondack, New York, U.S.A. lakes in relation to acidity, trophic status and related water quality characteristics. Hydrobiologia 175:33–48.

Simonis, J. L., and J. C. Ellis. 2014. Bathing birds bias β-diversity: Frequent dispersal by gulls homogenizes fauna in a rock-pool metacommunity. Ecology 95:1545–1555.

Sirianni, K. M. 2017. Differential wind dispersal of cladoceran ephippia in a rock pool metacommunity. Aquatic Ecology 51:203–218.

Soininen, J., and J. Heino. 2005. Relationships between local population persistence, local abundance and regional occupancy of species: distribution patterns of diatoms in boreal streams. Journal of Biogeography 32:1971–1978.

Stainton, M. P., M. J. Capel, and F. A. J. Armstrong. 1977. The chemical analysis of fresh water. Fisheries and Environment Canada.

Steinke, M., J. Stefels, and E. Stamhuis. 2006. Dimethyl sulfide triggers search behavior in copepods. Limnology and Oceanography 51:1925–1930.

Sterner, R. W., and J. J. Elser. 2002. Ecological stoichiometry: the biology of elements from molecules to the biosphere. Princeton University Press, Princeton, New Jersey.

Sutherland, J. 1989. Field surveys of the biota and selected water chemistry parameters in 50 Adirondack mountain lakes. New York State Department of Environmental Conservation, Albany, New York, USA.

Tales, E., P. Keith, and T. Oberdorff. 2004. Density-range size relationships in French riverine fishes. Oecologia 138:360–370.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Thum, R. A. 2004. Using 18S rDNA to resolve diaptomid copepod (Copepoda: Calanoida: Diaptomidae) phylogeny: an example with the North American genera. Hydrobiologia 519:135–141.

Thum, R. A. 2007. Reproductive interference, priority effects and the maintenance of parapatry in Skistodiaptomus copepods. Oikos 116:759–768.

Torke, B. 2001. The distribution of calanoid copepods in the plankton of Wisconsin Lakes. Hydrobiologia 453–454:351–365.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth. Springer, New York.

Villar-Argaiz, M., and R. W. Sterner. 2002. Life history bottlenecks in *Diaptomus clavipes* induced by phosphorus-limited algae. Limnology and Oceanography 47:1229–1233.

Vilmi, A., K. T. Tolonen, S. M. Karjalainen, and J. Heino. 2019. Niche position drives interspecific variation in occupancy and abundance in a highly-connected lake system. Ecological Indicators 99:159–166.

Watras, C. J. 1980. Subitaneous and resting eggs of copepods: Relative rates of clutch production by *Diaptomus leptopus*. Canadian Journal of Fisheries and Aquatic Sciences 37:1579–1581.

Watson, S. B., E. McCauley, and J. A. Downing. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnology and Oceanography 42:487–495.

Werner, R. G. 2004. Freshwater Fishes of the Northeastern United States. Syracuse University Press, Syracuse, New York.

Wetzel, R. G., and G. E. Likens. 1991. Limnological Analyses:139–165. Springer Science + Business Media, New York.

Wiens, J. J. 2011. The niche, biogeography and species interactions. Philosophical Transactions of the Royal Society B: Biological Sciences 366:2336–2350.

Wilkinson, D. M. 2001. What is the upper size limit for cosmopolitan distribution in free-living microorganisms? Journal of Biogeography 28:285–291.

Williamson, C. E., and N. M. Butler. 1986. Predation on rotifers by the suspension-feeding calanoid copepod *Diaptomus pallidus.* Limnology and Oceanography 31:393–402.

Williamson, C. E., and N. M. Butler. 1987. Temperature, food and mate limitation of copepod reproductive rates: separating the effects of multiple hypotheses. Journal of Plankton Research 9:821–836.

Winfree, R., J. W. Fox, N. M. Williams, J. R. Reilly, and D. P. Cariveau. 2015. Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. Ecology Letters 18:626–635.

Wong, C. K. 1981. Predatory feeding behavior of *Epischura lacustris* (Copepoda, Calanoida) and prey defense. Canadian Journal of Fisheries and Aquatic Sciences 38.

Table 1. Summary of environmental characteristics of study lakes. All lakes contain *Leptodiatpomus minutus*; lake names with * denote locations where *Aglaodiatpomus leptopus* also occurs. All values except surface area and maximum depth (Z-max) are averages of two or three sampling periods (±1 standard deviation). C:N is molar carbon to nitrogen ratio, C:P is molar carbon to phosphorus ratio, Seston C, N, and P are nutrient contents of seston greater than 0.45µm size, TN is total nitrogen, DOC is dissolved organic carbon, Hypo DO is hypolimnetic dissolved oxygen.

Lake	Z-max (m)	Surface area (ha)	Secchi depth (m)	Chlorophyll-*a* (µg L⁻¹)	pH	Seston C (mg L⁻¹)	Seston N (mg L⁻¹)	Seston C:N	Seston C:P	Seston P (µg L⁻¹)	TN (mg L⁻¹)	DOC (mg L⁻¹)	Hypo DO (mg L⁻¹)
Falls*	9.25	37.5	2.58 (0.11)	0.46 (0.035)	5.02 (0.41)	1.65 (0.60)	0.10 (0.034)	17.57 (0.11)	95.48 (2.14)	10.51 (2.18)	0.45 (0.001)	5.14 (0.06)	0.95 (0.41)
Helldiver	3.63	6.5	0.54 (0.12)	6.88 (0.275)	5.33 (0.35)	1.38 (0.01)	0.15 (0.004)	10.90 (0.49)	33.74 (4.24)	28.12 (2.73)	0.43 (0.016)	14.50 (1.55)	1.08 (1.47)
Hoel	28.03	181.7	4.85 (0.52)	1.37 (0.042)	6.58 (0.02)	0.32 (0.08)	0.03 (0.007)	11.38 (0.61)	38.72 (8.00)	5.54 (1.76)	0.20 (0.014)	3.44 (0.34)	8.54 (1.18)
Indian	10.15	33.2	1.89 (0.41)	2.55 (0.054)	5.43 (0.16)	0.55 (0.05)	0.04 (0.007)	16.09 (1.70)	77.39 (16.2)	5.33 (2.28)	0.41 (0.021)	5.73 (0.02)	4.98 (1.29)
Limekiln	10.57	186.9	4.11 (1.96)	1.10 (0.034)	6.53 (0.50)	0.35 (0.04)	0.04 (0.005)	10.79 (0.21)	99.31 (76.4)	3.73 (2.00)	0.29 (0.064)	3.33 (0.30)	7.30 (4.32)
Round	6.60	3.6	3.80 (0.72)	1.35 (0.088)	5.66 (0.19)	0.89 (0.35)	0.10 (0.045)	10.51 (1.60)	36.33 (6.12)	23.38 (18.7)	0.22 (0.005)	3.33 (0.09)	5.17 (4.48)
Squaw*	7.05	36.4	3.46 (0.87)	1.83 (0.072)	6.85 (NA)	0.65 (0.28)	0.06 (0.029)	12.74 (0.61)	86.16 (32.7)	6.68 (3.71)	0.26 (0.011)	3.82 (0.16)	2.94 (0.79)
West*	4.20	10.4	1.38 (0.32)	1.40 (0.078)	5.83 (0.37)	1.08 (0.21)	0.09 (0.014)	13.67 (0.53)	92.79 (12.0)	9.75 (1.23)	0.41 (0.032)	6.82 (0.27)	0.44 (0.28)
Wolf	8.10	4.4	1.4 (0.00)	1.34 (0.026)	4.41 (0.33)	1.22 (0.63)	0.09 (0.032)	15.15 (2.39)	157.17 (50.3)	5.75 (3.28)	0.45 (0.029)	8.30 (0.20)	4.12 (2.62)

Table 2. Results of multiple linear regression after forward and backward AIC model selection of Leptodiaptomus minutus and Aglaodiaptomus leptopus trophic position and five environmental variables from nine Adirondack lakes. Full model adjusted R^2 = 0.683, p = 0.0454. Values in bold-face are statistically significant.

Predictor	Estimate	Standard Error	p Value
Intercept	**2.766**	**0.120**	**<0.0001**
Seston N	0.876	0.517	0.151
Seston P	-0.893	0.510	0.140
C:N	0.215	0.170	0.260
Dissolved Organic Carbon	**-1.603**	**0.489**	**0.022**
Chlorophyll-*a*	1.305	0.518	0.053

Table 3. Results of stepwise model selection by AIC of lake environmental predictors for copepods trophic position among nine Adirondack lakes. Each line is one linear model with the associated AIC value.

Model Predictors								AIC
Seston N	Seston P	Seston C:N	DOC	Chl-a	Seston C	Seston C:P	pH	-13.420
Seston N	Seston P	Seston C:N	DOC	Chl-a	Seston C	Seston C:P		-13.420
Seston N	Seston P	Seston C:N	DOC	Chl-a	Seston C:P			-15.418
Seston N	Seston P	Seston C:N	DOC	Chl-a				-17.010

Table 4. Pairwise F_{ST} estimates of two copepod species sampled from lakes across the Adirondack region, New York, USA. *Leptodiaptomus minutus* F_{ST} values are calculated using 880 SNP loci from 193 copepods among nine populations; *Aglaodiaptomus leptopus* F_{ST} values in bold face type are calculated using 4,598 SNP loci from 72 copepods among three population. Bold-face values are *Aglaodiaptomus leptopus*.

Lake Name	Helldiver	Hoel	Indian	Limekiln	Round	Squaw	West	Wolf
Falls	0.058	0.065	0.041	0.054	0.062	(0.055; **0.202**)	(0.047; **0.251**)	0.028
Helldiver		0.066	0.069	0.060	0.063	0.064	0.063	0.067
Hoel			0.079	0.050	0.059	0.063	0.069	0.069
Indian				0.060	0.073	0.060	0.056	0.056
Limekiln					0.053	0.053	0.059	0.061
Round						0.064	0.069	0.069
Squaw							(0.065; **0.238**)	0.064
West								0.059
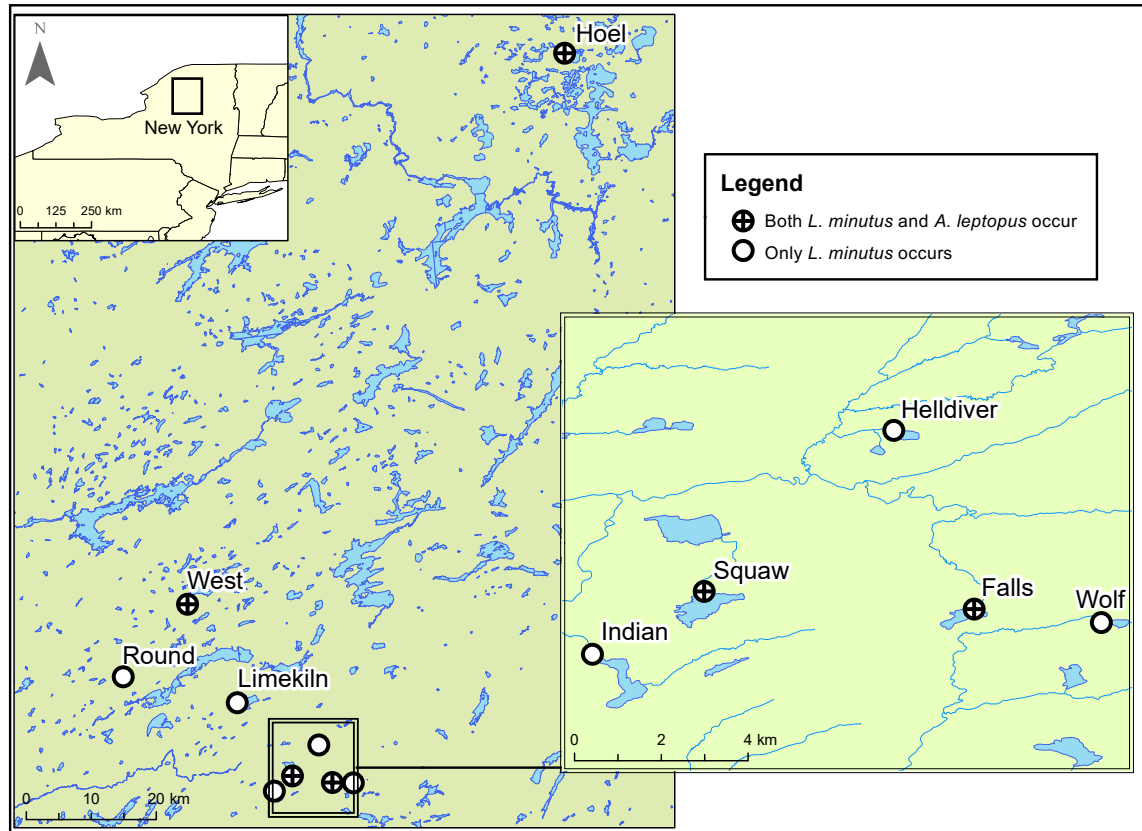
Figure 1. Lakes where the copepods *Aglaodiaptomus leptopus* and *Leptodiaptomus minutus* were sampled across the Adirondack region in New York, USA. *Leptodiaptomus minutus* occurred in all sampling locations, *Aglaodiaptomus leptopus* occurred in West, Squaw and Falls Lakes.
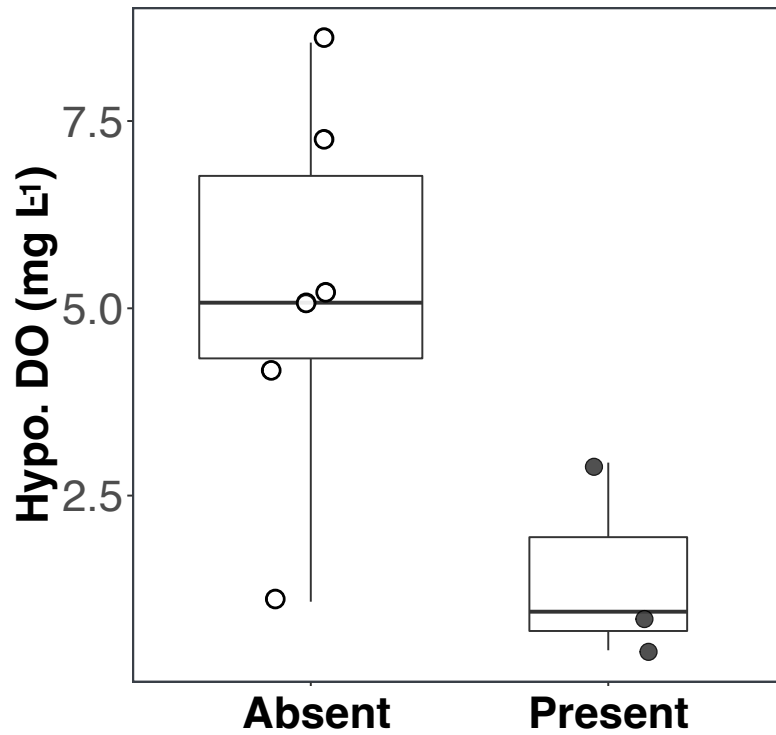
Figure 2. Comparison of the mean level of hypolimnetic dissolved oxygen (Hypo. DO) among nine lakes sampled across the Adirondack region, NY, USA where *Aglaodiaptomus leptopus* is present and absent. Hypo. DO is significantly lower in lakes where *A. leptopus* occurs versus lakes where it is absent $p = 0.024$.
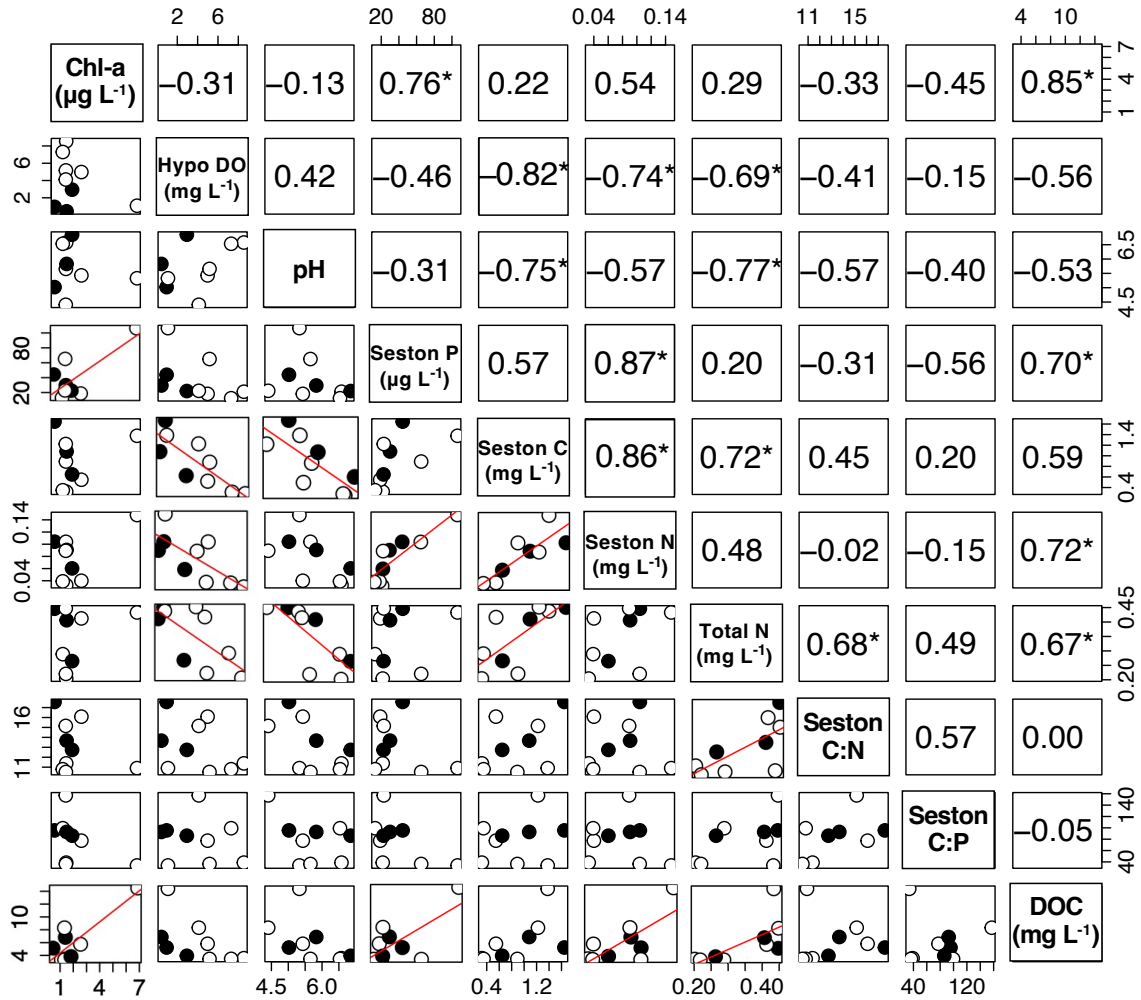
Figure 3. Pairwise comparisons of ten environmental variables from nine Adirondack lakes where *Aglaodiaptomus leptopus and Leptodiaptomus minutus* occur. Filled circles are lakes where both copepods occur, and red regression lines are included where correlations are statistically significant. Upper panels are the Pearson correlation coefficients among variables, with asterisks indicating $p < 0.05$ statistical significance. Correlations without asterisks were not statistically significant. Chl-a is chlorophyll-*a,* Hypo. DO is hypolimnetic dissolved oxygen, Seston C:N is seston molar carbon to nitrogen ratio, DOC is dissolved organic carbon, Seston C:P is seston molar carbon to phosphorus ratio.
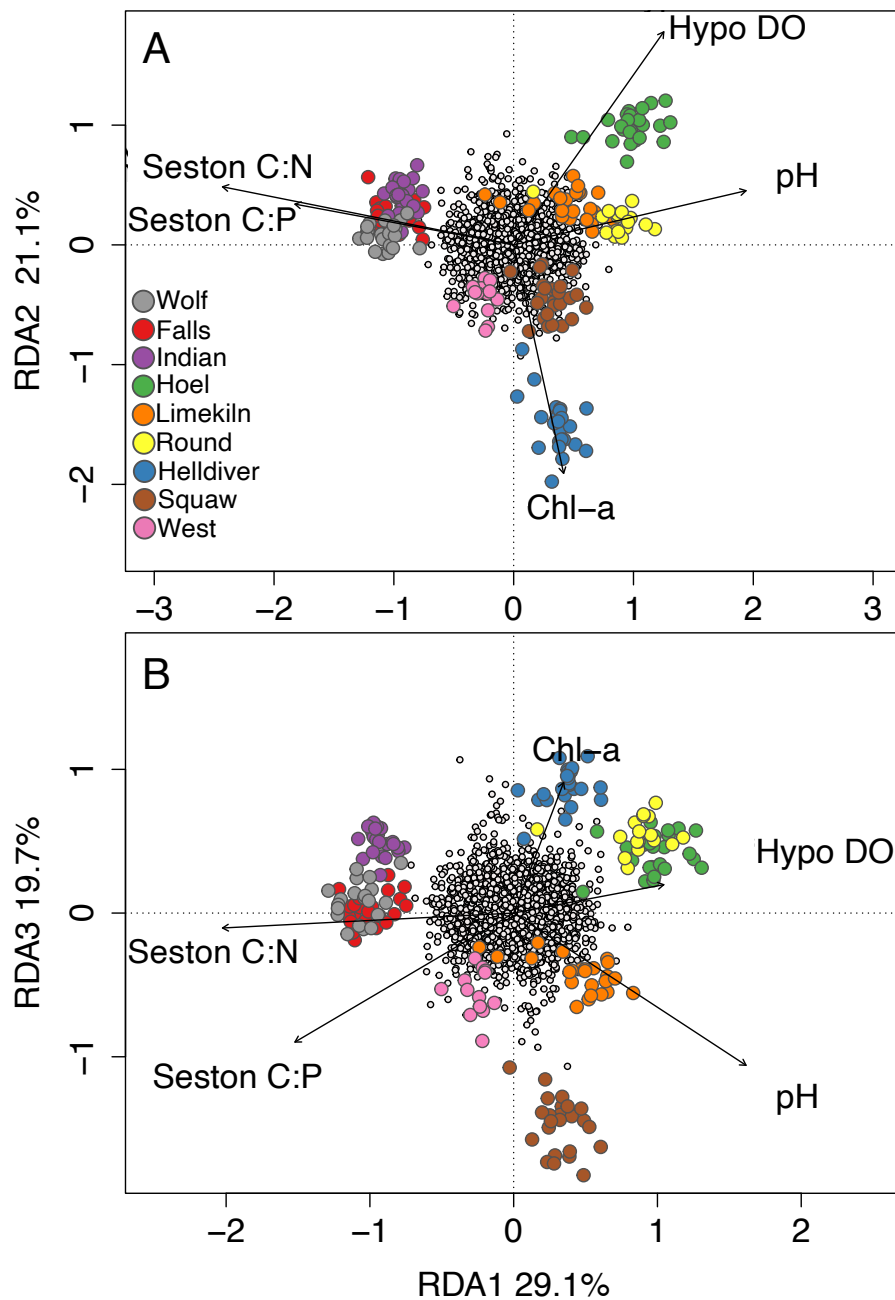
Figure 4. Redundancy analysis of five environmental characteristics of Nine lakes where *Leptodiaptomus minutus* is found in the Adirondack region New York, USA. A) RDA axes 1 and 2, B) RDA axes 3 and 4. Large colored points are individual copepods with colors coded by lake, small white points are single nucleotide polymorphisms (SNPs) from all individuals. Vectors are environmental characteristics. Chl-a is chlorophyll-a, Hypo DO is hypolimnetic dissolved oxygen, Seston C:N and Seston C:P are molar ratios of carbon to nitrogen and carbon to phosphorus, respectively.
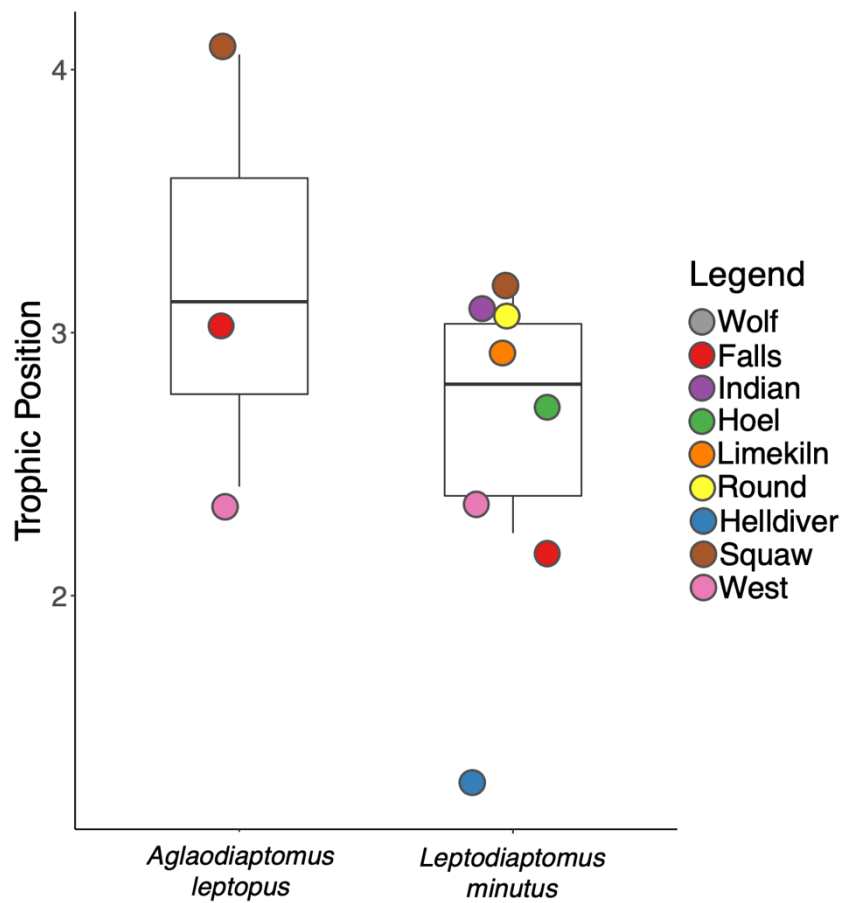
Figure 5. Trophic position of two copepod species based on $\delta^{15}N$ stable isotope signatures. Each data point is a population. Trophic position 2 is an herbivorous grazer *Daphnia spp.* or *Bosmina spp.* and Trophic Position 3 is a known co-occurring predator *Epischura spp.* or *Chaoborus spp.*
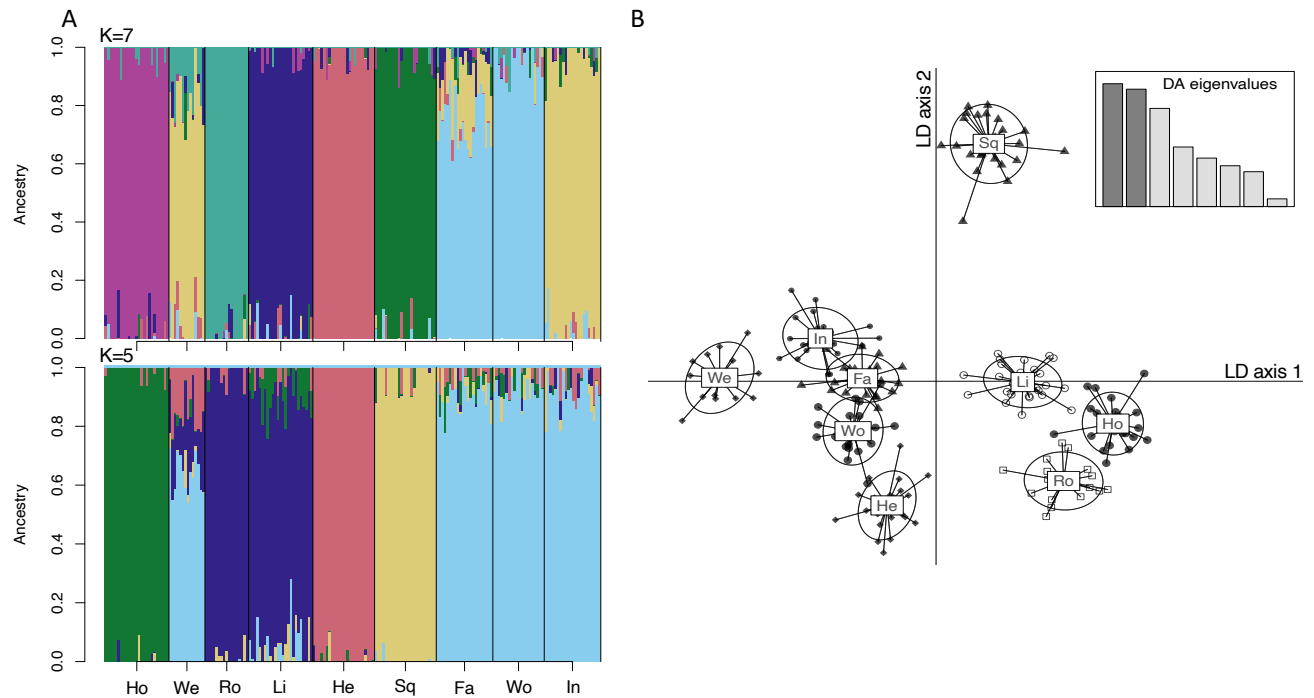
Figure 6. Results of individual ancestry estimation and genetic clustering of 193
Leptodiaptomus minutus individuals from populations in nine lakes across the
Adirondack region, New York, USA, ordered from North to South, based on 880 SNP
loci. A) Individual member assignment of 193 individuals where K=7 and K=5
genetic clusters identified with maximum likelihood estimation. B) Scatterplot of
genetic structure based on the first two axes of discriminate analysis of principal
component space. Individuals are points around an ellipse (populations) labeled with
an abbreviation for each lake: Fa is Falls, He is Helldiver, Ho is Hoel, In is Indian, Li
is Limekiln, Ro is Round, Sq is Squaw, We is West, Wo is Wolf.
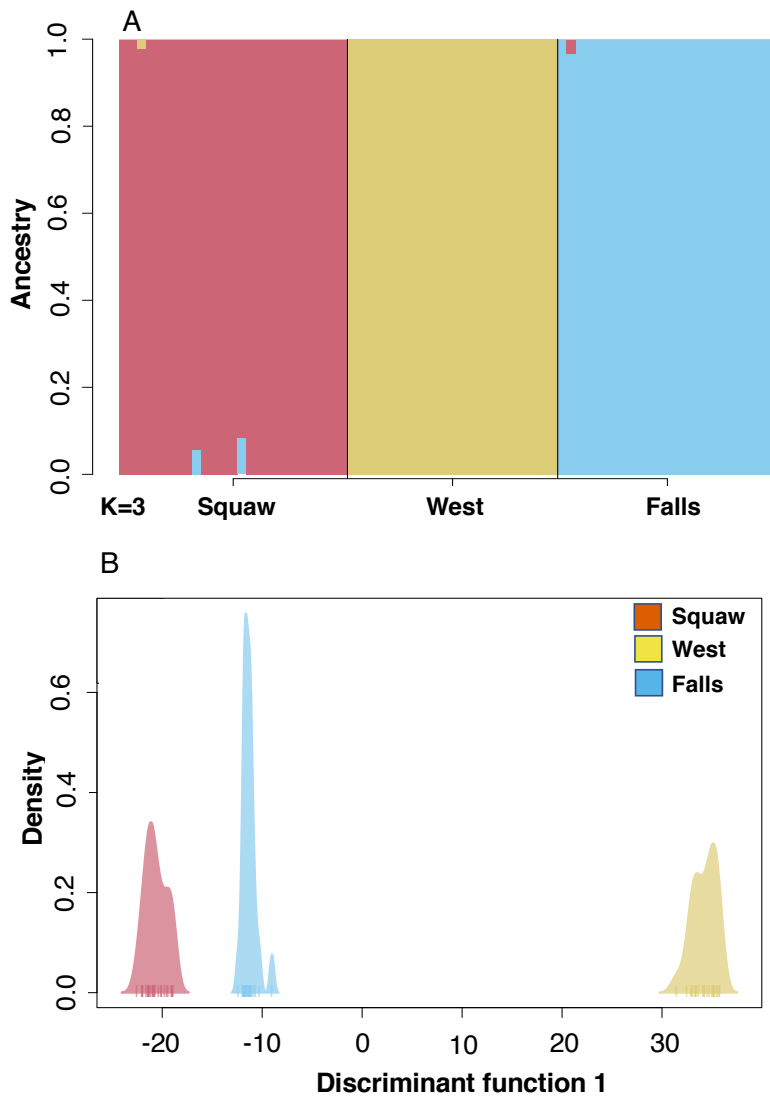
Figure 7. Results of individual ancestry estimation and genetic clustering of 72 *Algaodiaptomus leptopus* individuals populations in three lakes across the Adirondack region, New York, USA based on 4,598 SNP loci. A) Individual member assignments of 72 sampled animals where K = 3 genetic clusters identified with maximum likelihood estimation. B) Densities of individuals in each genetic cluster (lake) on discriminate function 1 from the discriminate analysis of principal components.
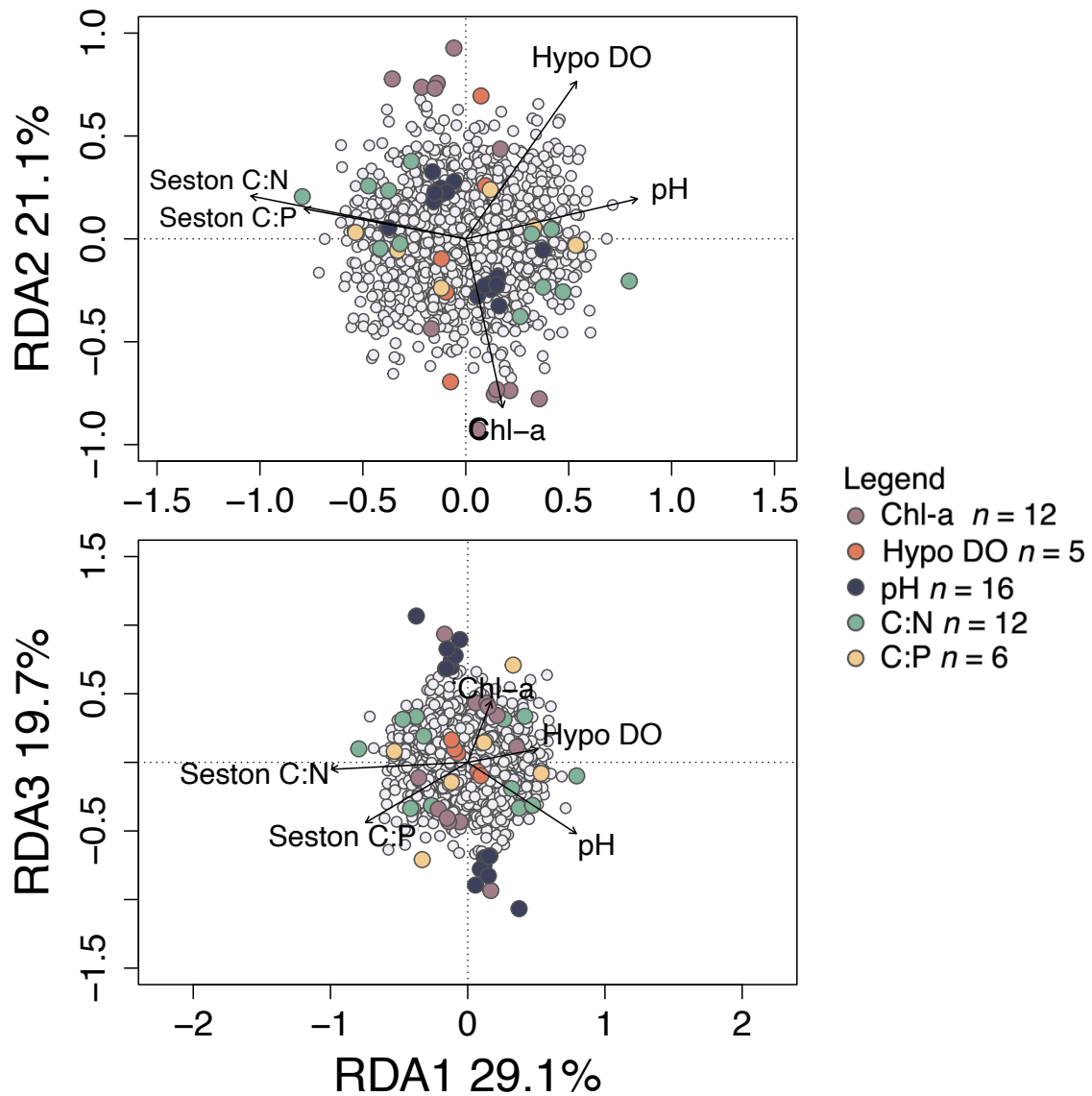
Figure 8. Results of redundancy analysis (RDA) of 880 SNP loci from 193
Leptodiaptomus minutus individuals among nine lakes, magnified to show
SNP points along RDA axes 1 and 2 (A), and RDA axes 2 and 3 (B) with
percent variation explained. Grey points are neutral SNPs, while colored
points are coded according to their mostly highly correlated environmental
variable. Vectors represent environmental characteristics among nine lakes
sampled; Chl-a is chlorophyll-a, Hypo DO is hypolimnetic dissolved oxygen,
Seston C:N and Seston C:P are molar ratios of carbon to nitrogen and carbon
to phosphorus, respectively.

SUPPLEMENT

TABLES

Table S1. Sources for mean trophic enrichment factor (Δn) for Adirondack lakes used to calculate the trophic position of *Leptodiaptomus minutus* and *Aglaodiaptomus leptopus*. δ^{15}N stable isotope values and Δn calculations are from one sampling period in early July 2014. Cholorphyll-*a*, pH, and Secchi depth are average values that were measured over two to three sampling events. Hoel, G, Moss, and North Lakes were sampled late May, early July, and late August. Squaw and West were sampled early July and late August. Values in parenthesis are ±1 standard deviation.

Lake	Maximum depth (m)	surface area (ha)	Chlorophyll-*a* (µg/L)	pH	Secchi depth (m)	grazer	grazer δ^{15}N (‰)	predator	predator δ^{15}N (‰)	Δn
Hoel	28.03	181.7	1.37 (0.042)	6.58 (0.02)	4.85 (0.52)	*Daphnia spp.*	0.843	*Epischura spp.*	3.13	2.29
G	9	39.9	1.19 (0.58)	6.02 (0.28)	3.33 (0.96)	*Daphnia spp.*	3.555	*Chaoborus spp.*	4.14	0.585
Moss	13.5	45.7	0.928 (0.17)	6.52 (0.39)	3.62 (0.49)	*Daphnia spp.*	2.48	*Epischura spp.*	5.11	2.63
North	15	176.8	1.35 (0.02)	5.29 (0.20)	1.8 (0.25)	*Daphnia spp.*	3.62	*Chaoborus spp.*	4.26	0.67
Squaw	7.05	36.4	1.83 (0.072)	6.85 (NA)	3.46 (0.87)	*Daphnia spp.*	1.83	*Epischura spp.*	3.56	1.73
West	4.2	10.4	1.40 (0.078)	5.83 (0.37)	1.38 (0.32)	*Daphnia spp.*	2.19	*Chaoborus spp.*	4.65	2.46

Table S2. Results of Redundancy Analysis (RDA) of environmental predictors (hypolimnetic dissolved oxygen, pH, seston C:N, seston C:P, and chlorophyll-*a*) in nine lakes where *Leptodiaptomus minutus* occurs across the Adirondack Region, New York, USA. All constrained axes are statistically significant *p* < 0.001.

Constrained axis	Eigenvalue	Proportion Variance Explained	Cumulative Proportion Variance Explained
RDA1	63.33	0.291	0.291
RDA2	45.84	0.211	0.501
RDA3	42.87	0.197	0.698
RDA4	38.42	0.176	0.875
RDA5	27.31	0.125	1.000

Table S3. Records of fish occurrence in each lake. * Indicates lakes where *A. leptopus* occurs, *L minutus* occurs in all lakes.

Lake	Year(s) sampled	Fish
Hoel[1]	1984,2002, 2004	Atlantic Salmon, B.Bullhead, Cisco, Fallfish, Golden Shiner, Lake Trout, Lake Whitefish, Largemouth Bass, Pumpkinseed, Rainbow Smelt, Round Whitefish, Smallmouth Bass, White Sucker, Yellow Perch
*West[1]	1996, 2006	Brook Trout, Brown Bullhead
Helldiver[1]	1998	Brown Bullhead, Brown Trout, Golden Shiner
Round[1]	2005	Central mudminnow, Pumpkinseed
Limekiln[1]	1997, 2004	Brown Bullhead, Brown Trout, Golden Shiner, Pumpkinseed, Rock Bass, Splake, White Sucker, Yellow Perch
*Squaw[1]	2000, 2001, 2003, 2006, 2008, 2010, 2012	Brook Trout, Creek Chub, Eastern Blacknose Dace, Rock Bass, Summer Sucker, Tessellated Darter, White Sucker
Indian[2,3]	1984, 2011	Brook Trout
*Falls[3]	1984	Brook trout
Wolf[3]	1984	None

1. Carlson, D. M., Daniels, R. A., & Wright, J. J. (2016). Atlas of Inland Fishes of New York. The New York State Education Department and Department of Environmental Conservation.
2. Baldigo, B. P., Roy, K. M., & Driscoll, C. T. (2016). Response of fish assemblages to declining acidic deposition in Adirondack Mountain lakes, 1984-2012. Atmospheric Environment, 146, 223–235.
3. Adirondack Lake Survey Corporation http://www.adirondacklakessurvey.org/historic.php
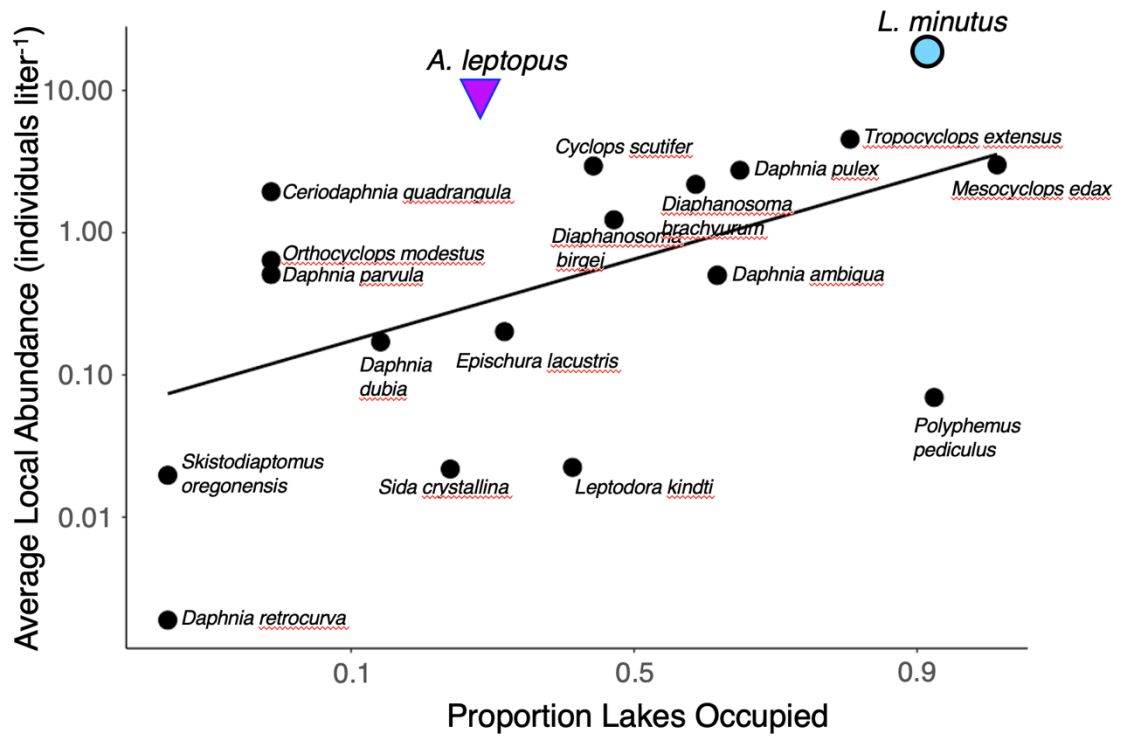
Figure S1. Relationship between zooplankton average local abundance and the proportion of lakes occupied from a long-term dataset of 28 Adirondack lakes (Leach *et. al.* 2018). There is a significant interspecific relationship between abundance and occupancy $\beta = 0.29$, $R^2 = 0.21$, $p = 0.017$.
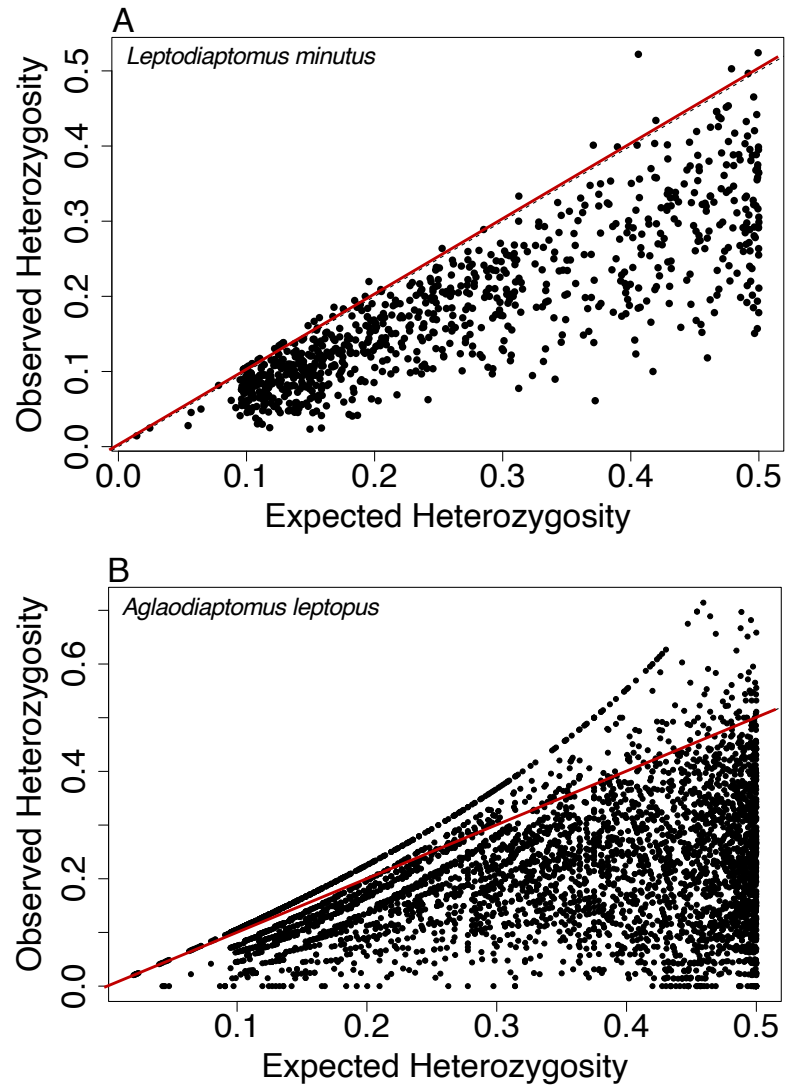
Figure S2. Expected vs observed heterozygosity of single nucleotide polymorphisms (SNPs) from two copepod species. Red line is the 1:1 relationship between observed and expected heterozygosity based on Hardy-Weinberg equilibrium. A) 880 SNP loci from 193 *Leptodiaptomus minutus* individuals sampled from nine lakes. B) 4,598 SNP loci from 72 *Aglaodiaptomus leptopus* individuals among three lakes.
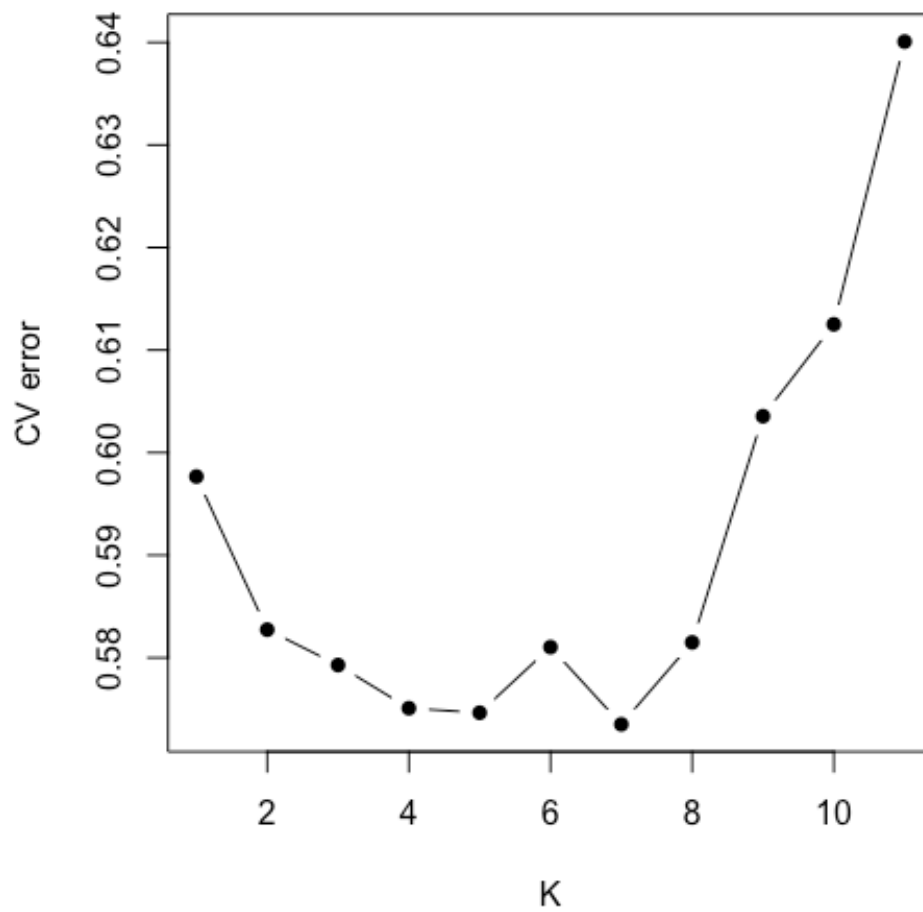
Figure S3. ADMIXTURE cross validation error from K=1 through K=11 populations for *L. minutus*. K=5 and K=7 had the lowest error and is used to analyze shared genetic ancestry among populations.
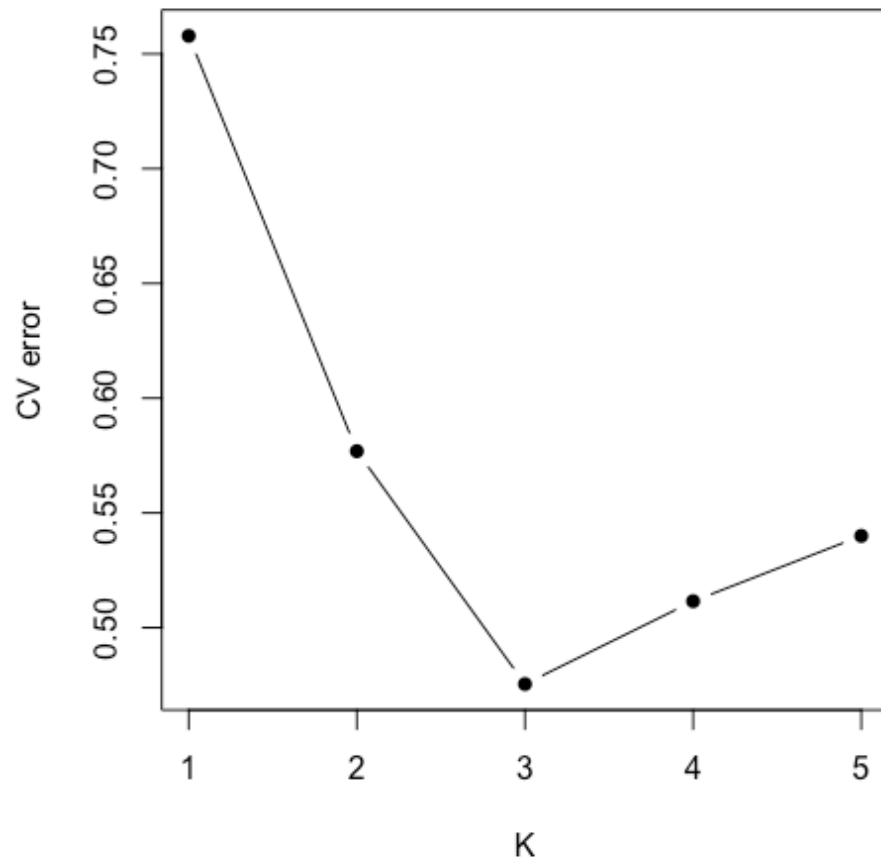
Figure S4. ADMIXTURE cross validation error from K=1 through K=5 populations for *A. leptopus*. K=3 had the lowest error and is used to analyze shared genetic ancestry among populations.

CHAPTER 3: EXPLAINING CONTEXT-DEPENDENT TROPHIC POSITION VARIATION IN DIAPTOMID COPEPODS

Rachel L. Wilkins[1,2], Blake Matthews[3], Asit Mazumder[4]

1. Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, New York, USA

2. eCornell, Office of External Education, Cornell University, Ithaca, New York, USA

3. Department of Fish Ecology and Evolution, EAWAG, Kastanienbaum, Switzerland

4. Water and Aquatic Sciences Research Program, Department of Biology, University of Victoria, Victoria, British Columbia, Canada

Abstract

Food chain length in lakes varies non-systematically with lake area, lake depth, and productivity. Researchers have found a positive relationship between lake size and food chain length, but not in all contexts, and others have found a positive relationship between productivity and food chain length, although, again, not in all lakes. Additionally, food chains are lengthened both via the addition of a top predator, and the elevation in trophic position of an existing top predator in a lake. Diaptomid copepods are an abundant, intermediate link in food chains, that themselves exhibit wide variation in trophic position, with the potential to affect the trophic position of predators higher in a food chain. We calculated the trophic position of five species of diaptomid copepods from 43 lakes of varying size and productivity across British Columbia, Canada. We used $\delta^{15}N$ stable isotope data to determine how trophic position of an intermediate consumer varies with body size, lake size, and lake

productivity, and found context-dependent relationships, modulated by body size, lake size, and lake water chemistry. There were no strong broadly consistent patterns in the data, although detailed analysis revealed some more subtle relationships. Trophic position was negatively correlated with body size only in mesotrophic lakes. Trophic position was positively correlated with lake surface area in lakes larger than 100 ha, but not in smaller lakes. Our results demonstrate that mechanisms of food chain lengthening and contraction along gradients of lake size and productivity may be caused as much by changes in the trophic position of intermediate consumers as by addition or subtraction of top predator species.

Introduction

In lakes, lengthening of food chains can occur when there are adequate resources for the addition of a top predator, such as a piscivorous fish species, or when a niche develops at an intermediate trophic link ultimately elevating the trophic position of a top predator (Post et al. 2000). Lake size and productivity are determinants of total food chain length and thus the trophic position of top predators (Jenkins et al. 1992, Vander Zanden et al. 1999, Post et al. 2000, Doi et al. 2009, Takimoto and Post 2013, Eloranta et al. 2015, Ward and McCann 2017), although whether food chain length is correlated with lake size, productivity, or both differs among studies. Some recent studies of food chain length have found that the significance of lake size and productivity are context-dependent, becoming more or less important as a function of size and trophic state, with different mechanisms lengthening or shortening the food chain (Tunney et al. 2012, Ward and McCann

2017). While the addition of a top predator to a food web lengthens a food chain, factors affecting the trophic position of intermediate consumers and how we measure it are important to consider when examining lake food webs as a whole. Lake characteristics and organisms' body size determine the diet of intermediate consumers, thus affecting their trophic position in a food chain, and their trophic connections with other organisms.

Food chain length may be a function of lake productivity as organisms are limited by energy availability; for example, (Persson et al. 1992) found that top predators tend to be absent from low productivity lakes. Longer food chains are often found in more productive lakes (Jenkins et al. 1992, Vander Zanden et al. 1999, Doi et al. 2009), although this is not always the case, since Vander Zanden et al. (1999), Post et al. (2000), and Post (2002) showed that lake size indirectly affects trophic position presumably through altered habitat availability, greater richness of resources, or gape limitation of intermediate consumers (Hairston and Hairston 1993). The effect of lake size and productivity on trophic position may, indeed, be context, or mechanism dependent. Lake size and productivity similarly impact food-chain length through different mechanisms (Takimoto and Post 2013). However, these relationships are not always monotonic, and some studies have found a positive relationship between food chain length and lake size, but not productivity, and others the opposite (Ward and McCann 2017).

Omnivory among consumers is common within aquatic systems (Havens et al. 1996), and the trophic position of the particular species in a food web can vary markedly due to the degree of omnivory. Among populations of consumers, the level

of omnivory can be determined by the relative availability of plant and animal prey (Hellmann et al. 2013, Chubaty et al. 2014, Tewfik et al. 2016), but is also often context-dependent. Within existing food web structure, trophic position may vary depending on feeding behaviors modified by the presence of predators and the intensity of competition (Zandonà et al. 2017). Understanding how omnivory changes along environmental gradients such as ecosystem size and productivity, can help understand how changes in the diets of intermediate consumers affects food chain length.

Body size has a well-established positive relationship with trophic position, spanning from zooplankton to cetaceans (Romanuk et al. 2011, Romero-Romero et al. 2016, Potapov et al. 2019). Fishes are largely gape limited consumers (e.g. Hambright et al. 1991), and because mouth gape is generally related to overall body size, the upper limit of the size of prey consumed is directly related to predator size. This is indirectly responsible for the positive relationship between body size and trophic position among fishes (Keppeler and Winemiller 2020). Additionally, there is often a positive relationship between body size and trophic position among populations of single fish species (Arim et al. 2010, Olson et al. 2020). This wide intraspecific variation in size means that the consumer community as a whole can access a wider variety of prey and have higher variation in diet than if size variation were minimal (Snowberg et al. 2015, Gibert and DeLong 2017). As a result, the variation in body size not only plays a role in determining trophic position, but also in the variation of trophic position and the degree of omnivory within and among species.

For lake fish assemblages there are well known relationships between tropic

position and fish body size, lake size and lake productivity. However, it is far less well known if this pattern is at all general for another major group of intermediate consumers in the pelagic zone of lakes: diaptomid copepods, which can be abundant and an important part of the food chain from phytoplankton to fish (Lazzaro 1987, Gliwicz 2003).

Despite the evidence of a positive relationship between trophic position and body size, planktonic food webs may not fit well into this framework. Although fishes are constrained in their prey size by gape limitation, zooplankton have different feeding modes, and their resource acquisition may not be similarly constrained. Specifically, planktonic members of the ubiquitous crustacean order Copepoda capture particles with their feeding appendages rather than with their mouths (Kohl & Strickler 1981, Kiørboe et al. 2009) and so are both capable of diet selectivity and may even preferentially feed on other consumers. Members of the freshwater Family Diaptomidae are known to consume both algae and rotifers (Williamson and Butler 1986, DeMott 1988, 1995) affecting their trophic position and energetic connections within lake food webs (Ventura and Catalan 2008). Additionally, other freshwater herbivorous zooplankton, such as *Daphnia* and *Holopedium*, are often larger than co-occurring predatory plankton such as many cyclopoid copepods, such that within plankton communities there may be no relationship between body size and trophic position as there is in fish. Calanoid copepods rely on sophisticated chemo- and mechanoreception to detect prey (Friedman and Strickler 1975, Buskey 1984, DeMott 1988, Steinke et al. 2006, Kiørbe 2008, Almeda et al. 2018), which allows them to select the food items that increase survival and reproduction (Williamson and Butler

1986, 1987). In addition to phytoplankton (Grosbois et al. 2017), diets of calanoid copepods may include prey ranging from rotifers (Williamson and Butler 1986, Arnott and Vanni 1993) to heteotrophic flagellates and ciliates (Burns and Schallenberg 1996), to nano-plankton (Hambright et al. 2007).

Our goal here was to understand how copepod body size and lake physical, chemical, and biological characteristics affect the trophic position among five species of diaptomid copepods distributed across 43 lakes of varying size and productivity in British Columbia, Canada. If copepod omnivory, trophic position, or both vary as a function of lake environment, this might then influence the trophic position of consumers further up the food chain, and in this way relate to the observed dependence of the trophic position of top predatory fish on lake size and productivity. We used $\delta^{15}N$ stable isotope signatures to calculate the trophic positions of the five most commonly occurring diaptomid species in these lakes to determine how trophic position varies within and among species as a function of adult body size as well as lake size, trophic state, and chemical environment. All five species had omnivorous to predatory trophic positions, which is higher than the herbivory typically expected of most diaptomid copepods. Despite the known positive relationship between body size and trophic position among pelagic lake fishes, we found no such relationship either within or among diaptomid copepod species. Rather, copepod trophic position was only related to body size in specific lake conditions.

Methods

Lake Sampling – Forty-three lakes were sampled for zooplankton throughout British Columbia, Canada (Figure 1), over the period 2001-2006 during several different field campaigns. Lakes in the interior of British Columbia (BC) were sampled as part of regular water quality monitoring performed by the BC Ministry of Water Land and Parks. These samples were shipped frozen to the University of Victoria (UVic), whereas samples from Vancouver Island were sampled by several members of the Environmental Management of Drinking Water Laboratory at UVic. Samples collected in 2001-2003 were stored overnight in filtered (Whatman GF/F or deionized) water and then sorted live and dried at 60 °C. Overnight storage of live zooplankton did not affect the δ^{15}N signature compared with samples that were immediately frozen (Matthews and Mazumder 2003). Samples collected after 2005 were stored frozen, and both picked and analyzed within several months. Freezing has no known effects on isotope values (Matthews and Mazumder 2006), so we pooled all the data for the current analysis. In total, 237 samples each consisting of a single date from a specific lake were analyzed, of which 123 have been published in previous papers (Matthews and Mazumder 2003, 2005, 2007). Sixty-seven samples from 34 lakes have not been included in previous work.

Environmental data were compiled from routine monitoring of lake water quality by the Government of British Columbia, Ministry of Water, Land, and Air Protection (WLAP), and from the University of Victoria Environmental Management of Drinking Water Laboratory (EMDWL), both of which follow standard nutrient analysis methods (APHA 1998). For example, the WLAP samples for total

phosphorus were determined by ascorbic acid digestion and total nitrogen by a persulfate digestion, and for the EMDWL samples, total nitrogen (TN) and total phosphorus (TP) were analyzed on a Lachat automated ion analyzer (Zellweger Analytics, QuickChem 8000) (see also Davies et al. 2004). Lake water total organic carbon (TOC) was measured via a wet oxidation TOC Analyzer (e.g. Shimadzu TOC). Lake water pH was measured using environmental probes (e.g. VWR PhD probe, model 2000). Median values of all the available data were used to characterize among-lake variation in environmental conditions.

Lakes were assigned a trophic state by TP concentration according to (Carlson 1977). Lakes with greater than 25 μgL^{-1} TP were categorized as eutrophic, lakes with $\leq$ 10 μgL^{-1} TP categorized as oligotrophic, and lakes with 10.1-24.9 μgL^{-1} TP as mesotrophic. In total, 23 of the sampled lakes were as oligotrophic, 17 were mesotrophic, and three were eutrophic. For the 42 lakes for which we have surface area data, those $\leq$ 100 hectares surface area were categorized as small ($n=20$), and those > 100 hectares were categorized as large ($n=22$) (Cael and Seekell 2016). There was no relationship between TP and lake size (Figure 2).

Zooplankton were collected with a Wisconsin net (30 - 50 cm diameter, 64 μm mesh) from the entire water column, or to a maximum depth of 30 m. The dominant diaptomid copepod, non-discriminate filter feeding herbivores (*Daphnia, Holopedium, Ceriodaphnia,* or *Bosmina*), and obligate predators (*Chaoborus*, *Epischura*) were picked from the sample. For this study a total of five diaptomid species were abundant enough to obtain a stable isotope sample, consisting of one species per lake: *Aglaodiaptomus denticornis, Hesperodiaptomus franciscanus, Leptodiaptomus*

ashlandi, Leptodiaptomus tyrelli, and *Skistodiaptomus oregonensis.* Zooplankton were either sorted live within 24 hours of collection (samples were stored at 4°C overnight, e.g. see Matthews and Mazumder (2003), or frozen after collection and individuals were hand-picked from thawed samples. The number of individuals per sample depended on the species and size (as reported in Matthews and Mazumder 2007) sufficient to obtain ca. 0.5 mg - 1 mg of dried zooplankton tissue (dried at 60°C for 24 h), which was stored in a desiccator until analysis. The majority of the stable isotope samples were run on a Finnegan Delta Plus Advantage stable isotope ratio mass spectrometer (IRMS) at the University of Victoria, with some samples from previous work being run on an Isochrom IRMS at the University of Waterloo (Matthews and Mazumder 2003, 2005). In all cases, the precision for $\delta^{15}N$ was <0.2 ‰ across all individual runs.

Trophic Position – Trophic position for each copepod population was calculated using a single trophic base as in Post (2002) using $\delta^{15}N$ isotope signatures. The indiscriminate filter feeding herbivorous zooplankton in the genera *Daphnia, Holopedium, Ceriodaphnia,* or *Bosmina* were used as integrated representations of trophic position 2. To calculate the most accurate trophic enrichment factor, where possible, we used the average $\delta^{15}N$ of more than one species of herbivore in each lake to represent the mean signature for trophic position 2 in that lake (Kjeldgaard et al. 2021). We calculated trophic position using the $\delta^{15}N$ signatures of both the grazer herbivore base and the diaptomid copepod or other zooplankton secondary consumer as:

$$\text{trophic position} = \lambda + \frac{\delta^{15}N \; secondary \; consumer - \delta^{15}N \; grazer}{\Delta_n},$$

where λ is the trophic position of the grazer, in this case 2, and Δ_n is the average

enrichment in δ^{15}N for each trophic level, 3.4 ‰ (Minagawa and Wada 1984, Post

2002). Trophic position was calculated for the diaptomid copepod species and any co-

occurring zooplankton predators at each sampling location and date.

Body Size – Average copepod body size was measured as the length of the prosome on

30-100 individuals per sample. Those same individuals were used in δ^{15}N samples.

Body size was not measured from one lake containing *H. franciscanus,* and one lake

containing *L. tyrelli*. Additionally, body size was not measured for all sampling

events, such that body size was measured for *n* = 15 samples for *A. denticornis, n* = 35

samples for *H. franciscanus, n* = 5 samples for *L. ashlandi, n* = 21 samples for *L.

tyrelli,* and *n* = 10 samples for *S. oregonensis.* Lengths were converted to mass (µg dry

weight per individual) using an established length-weight regression for diaptomid

copepods:

$$dry \; weight = e^{\ln(\alpha) + \beta \ln (L)},$$

where the α = 1.953, β = 2.40, and *L* is length of the copepod in mm (Watkins et al.

2011).

Data analysis – Before further analysis, for each species we removed copepods with

outlier trophic positions greater than four times the mean Cook's distance. In total,

nine outlier samples were removed from a total of 237 observations (Table S1). We

used Bayesian multiple regression and mixed-effects models created in the Stan

computational framework (http://mc-stan.org/) accessed with the R package brms

(Bürkner 2017, 2021) to evaluate differences in body size, trophic position, and the

relationships between trophic position and body size and lake environmental variables.

For all models, fit was first assessed visually with posterior predictive checks of the

distribution of the response variable with $n = 50$ draws with the R bayesplot package

(Gabry and Mahr 2022), second, by examining convergence of all four chains after

doubling the number of post-warmup draws for each model (Depaoli and Schoot

2017), and third, ensuring $\hat{R}$ values for all model parameters were less than 1.000. All

models were sampled in four chains with 5,000 warm-up draws and 20,000 total post-

warmup draws. To avoid bias in posterior estimates, unless specified otherwise, we

adjusted the NUTS sampler adaptation parameter adapt_delta to 0.99 and a maximum

tree depth to 15. All data analysis was conducted with R Version 4.1.0 (R Core Team

2021).

Interspecific Difference in Trophic Position – Interspecific differences in trophic

position were estimated using Bayesian mixed-effects models. The model was

structured without an intercept, with copepod species as a fixed effect, and lake

identity as a random effect. Many diaptomid copepods are omnivores (Hairston 1979,

Williamson and Butler 1986, Arnott and Vanni 1993), however because this varies

among studies, we used weakly informed priors for the model to detect differences in

trophic position among species. Model priors of trophic position for all species were

normally distribution with $\mu = 2.5$, $\sigma = 1$. The remaining model parameters (standard

deviation and intercept of the lake random effect, and residual standard deviation) were a student-t distribution with $\upsilon = 3$, $\mu = 0$, and $\sigma = 2.5$ (Table S2). We used a skew normal with a log link error distribution for the model. Differences in trophic position among species were assessed with pair-wise contrasts of posterior medians and distributions with the R package emmeans (Lenth 2022).

Trophic position and Body Size – Interspecific differences in adult body size among species were estimated using Bayesian linear models with skew normal error distribution. Body size priors were obtained for each species from those reported by (Reid and Williamson 2010). Dry weights for all five species from published size estimates ranged between 5.18 μg and 32.90 μg per individual. We used species-specific informed priors with μ set to the published median estimated body size range for each species, and $\sigma = 12$ (Table S2). Differences in body size among species were assessed with pair-wise contrasts of posterior medians and distributions with the R package emmeans (Lenth 2022). We estimated the difference in body size among populations in oligotrophic and mesotrophic lakes with a Bayesian linear mixed effects model with lake trophic state as a fixed effect and copepod species as a random effect. Model priors for the average copepod mass from each lake trophic state were the average and standard deviation of the range of size estimates from published size estimates for each species (Reid and Williamson 2010) (Table S2).

To determine the interspecific relationship between body size and trophic position, we used a Bayesian linear mixed effects model with copepod species as a fixed effect and lake as a random effect. The prior for the model intercept was the

dummy variable equivalent to the trophic position of *A. denticornis,* therefore prior for model intercept was normally distributed with μ = 2.5 and σ = 1. Weakly informed priors for all β coefficients (body size and copepod species identity) were normally distributed with μ = 0 and σ = 1 (Table S2).

We also examined the intraspecific relationship between body size and trophic position by splitting the data by species identity. The intraspecific model had body size as a fixed effect and lake as a random effect. Lake was not included as a random effect for *S. oregonensis* and *L. ashlandi* models since there was only a single sample per lake. Generic weakly informed priors were used; β copepod body size estimated as normally distributed with μ = 0, σ = 1 (Table S2).

In addition to investigating the relationship between body size and trophic position within species, we also examined the relationship for lakes of different trophic states and size (surface area) by running the model for oligotrophic lakes, mesotrophic lakes, small lakes, and large lakes separately. All models had body size and species identity as a fixed effect and lake as a random effect. We did not include eutrophic lakes in this analysis because there were only three in the data set. Small and large lake models were fit with a Gaussian error distribution, while lake trophic state models were fit with a skew normal and Gaussians error distribution for oligotrophic and mesotrophic lakes, respectively. To avoid bias in posterior estimates we adjusted the NUTS sampler adaptation parameter adapt_delta to 0.999. The prior for the model intercept was the dummy variable equivalent to the trophic position of *A. denticornis,* therefore a weakly informed prior for model intercept was normally distributed with μ=2.5 and σ=1. Priors for all β coefficients (body size and copepod species identity)

were generic weakly informed and normally distributed with μ=0 and σ=1 (Table S2).

Lake size and lake water chemistry – To analyze the relationship between copepod trophic position and the lake physical-chemical environment, we first calculated the average diaptomid trophic position in each lake over all *n* sampling events, which was between one and 32 samples (Table 1). Then, we $\log_{10}+1$ transformed total phosphorus, total organic carbon, and area to correct for skewed distributions in the raw data. To maximize the number of lakes included in the dataset, each incomplete variable was imputed using the mice package in R, which uses multivariate imputation by chained equations using Gibbs sampling (Buuren and Groothuis-Oudshoorn 2011). One missing value each was imputed for total organic carbon and lake area. One hundred imputed data sets were created and were passed onto models using the brm_multiple command in the brms package, which runs the same model on all 100 data sets (Bürkner 2017). The relationship between trophic position and lake environment was fitted with Bayesian multiple regression with $\log_{10}+1$ total phosphorus, $\log_{10}+1$ total organic carbon, pH, $\log_{10}+1$ lake area, and copepods species as predictors in the model. As with previous models, the prior for the model intercept was the dummy variable equivalent to the trophic position of *A. denticornis* (chosen based on alphabetical order of the five species in the model) therefore prior for model intercept was normally distributed with $\mu = 2.5$ and $\sigma = 1$. Priors for all β coefficients (copepod species, total phosphorus, total organic carbon, pH, area) were normally distributed with $\mu = 0$ and $\sigma = 1$ (Table S2).

To examine the relationship between copepod trophic position and lake physical-

chemical characteristics within lake classes, we also ran the separate Bayesian multiple regression models for oligotrophic and mesotrophic lakes, as well as large and small lakes. We ran the same Bayesian multiple regression model for each copepod species separately to determine the intraspecific relationship between trophic position and lake physical-chemical characteristics, with lake area, total phosphorus, pH and total organic carbon as fixed effects, and lake as a random effect. *L. tyrelli* (*n* = 6 samples) and *L. ashlandi* (n = 7 samples) were excluded because their low sample sizes prevented model convergence.

Results

Intraspecific differences in trophic position and body size – The mean trophic position among all five copepod species across 43 lakes in British Columbia, Canada was 3.04 $\pm$ 0.37 (sd). These were comparable to trophic positions of the co-occurring obligate predatory zooplankton in the genera *Chaoborus, Epischura,* and *Leptodora* (Figure 3). There were three levels of mean trophic position among the five copepod species. *H. franciscanus* and *L. ashlandi* had a higher posterior mean trophic positions than both *L. tyrelli* and *S. oregonensis*, while *A. denticornis* had an intermediate trophic position and was not significantly different from any other species. *S. oregonensis* and *L. tyrelli* had the lowest posterior trophic position estimates (Figure 4A). *L. ashlandi* was almost a half of a trophic position higher than *S. oregonensis* (Table S3). Copepod dry weight ranged between 4.51 and 32.07 μg individual^{-1}. *H. franciscanus* was similar in size to *L. tyrelli,* and *S. oregonensis* was similar in size to both *L. ashlandi* and *L. tyrelli* (Figure 4B, Table S3).

Trophic position vs body size – Body size was not a strong predictor of trophic position in this study (Figure 4). There is no statistically significant relationship between body size and trophic position within the entire five-species data set (β -0.008, 95% credible interval (CI) -0.029−0.012), nor among oligotrophic lakes, large lakes, or small lakes (Table 2). Additionally, there is no statistically significant difference in diaptomid body size between mesotrophic and oligotrophic lakes (posterior size difference estimate -0.10 µg, 95% CI -1.74−1.55). There, is however a statistically significant negative relationship between trophic position and body size (i.e., smaller copepods had a higher trophic position) among mesotrophic lakes in the model that included all copepod species (β -0.037, 95% CI -0.07−-0.001, Table 2). Not all five copepod species occur in mesotrophic lakes; *A. denticornis* occurs only in mesotrophic lakes, *L. ashlandi* and *L. tyrelli* occur only in oligotrophic lakes, and *H. franciscanus* and *S. oregonensis* occur in both mesotrophic and oligotrophic lakes.

Lake Size and Water Chemistry – The environmental variables from the study 43 lakes varied in systematic ways. There was a weak positive relationship between lake area and pH (Figure 5A, $p = 0.035$, $R^2 = 0.12$), while there is no statistically significant relationship between lake area and total phosphorus ($p = 0.28$, $R^2 = 0.03$). Additionally, there is no statistically significant difference in lake size between oligotrophic and mesotrophic lakes (T-test, $p = 0.67$, $F_{1,23} = 0.18$, Figure S1). Smaller lakes have a significantly slightly lower mean pH of 7.37 compared with 7.62 in large lakes (T-test, $p = 0.028$, $F_{1,40} = 5.22$, Figure 5C). Additionally, larger lakes had a slightly wider range of pH values than smaller lakes, 7.0-8.35 versus 6.9-7.8 respectively.

Taking all lakes as a group, lake area, total phosphorus, total organic carbon, and pH are not statistically significant predictors of copepod trophic position (Figure 6, Table S4). When the data are split according to lake size and lake trophic state, however, both pH and lake area become significant predictors of trophic position. Within oligotrophic lakes, pH is significantly positively associated with trophic position (β pH = 0.51, 95% CI 0.11 − 0.89 , Figure 7A), but not for mesotrophic lakes (β pH = 0.15 95% CI -0.39 − 0.68, Figure 7B). Among small lakes, pH is significantly positively associated with trophic position (β pH = 0.82, 95% CI 0.21 − 1.40, Figure 7C). In large lakes, only lake area is positively associated with trophic position (β area = 0.20, 95% CI 0.07 − 0.33, Figure 7D). Within species, neither water chemistry nor lake size is a significant predictor of trophic position (Table S5).

Discussion

The diaptomid copepods in our study had both surprisingly high mean trophic positions and surprisingly variable trophic positions both within and among species. These observations motivate questions of which ecological drivers affect trophic position in these intermediate consumers, which link primary producers to top predators, and what role they play in affecting food chain length. We expected the trophic positions of the copepods in this study to be between the reference herbivore and a comparable obligate predator based on evidence that they can be omnivorous and feed on phytoplankton and small zooplankton. However, trophic positions were markedly higher than previously suspected, approximately trophic position 3 and in many cases even higher (Figure 3). This is an unusual finding, even if we would

132

expect copepods to have a higher trophic position than *Daphnia* because of their diet selectivity, trophic positions of copepods from BC lakes were closer to what we would expect of a piscivorous fish, rather than omnivorous zooplankton.

Lake trophic state (a categorization of lake productivity), may modulate consumer trophic position based on the amount of recycled production and shifts in the ratio of autotrophic to heterotrophic production, causing $\delta^{15}N$ to become enriched as it cycles. In marine systems, zooplankton had elevated $\delta^{15}N$ levels during periods of low production when the microbial community is dominated by heterotrophic organisms, compared with periods of high production when the microbial community consisted mainly of autotrophic organisms (*i.e.*, phytoplankton). As a result, the $\delta^{15}N$ signature of the zooplankton becomes elevated as their diets shift from phytoplankton to heterotrophic microorganisms (Romero-Romero et al. 2016). If there was greater recycled production and a shift from autotrophic to heterotrophic production enriching $\delta^{15}N$ in oligotrophic lakes, we might expect to see a difference in mean trophic position between oligotrophic and mesotrophic lakes, which we do not (posterior trophic position difference estimate 0.10, 95% CI -0.15 − 0.36).

While $\delta^{15}N$ may not be enriched in lakes with different trophic states, heterotrophic production nevertheless likely plays a role in affecting the trophic position of diaptomid copepods in these lakes. Copepod trophic position was related to ecosystem size in large lakes, but not in small lakes where it was related to pH (Figure 7). As lake size increases, the ratio of the areas of the pelagic to littoral zones increases exponentially. This may contribute to the relationship between copepod trophic position and lake area; in lakes with a larger relative pelagic zone, copepods

may appear to feed at higher trophic levels because autochthonous production is constantly recycled through the microbial loop (Figure 8). Repeated cycles of microbial degradation followed by consumption by copepods (and other grazers) can enrich $\delta^{15}N$, potentially elevating the actual trophic position of next-level consumers as result of this cycling (Romero-Romero et al. 2019). With each connection in the microbial loop, with bacteria using dissolved and detrital particulate organic, ciliates and heterotrophic flagellates consuming bacteria, and ciliates also consuming heterotrophic flagellates, the $\delta^{15}N$ becomes enriched. The connections within the microbial loop effectively elevate the trophic position of consumers, independent of the food chain from primary producers (algae and cyanobacteria) to copepods and fish. If copepods consume more heterotrophic flagellates and ciliates from the microbial loop than co-occurring non-selective grazers, such as *Daphnia*, then we would expect $\delta^{15}N$ to be markedly elevated. The resulting effect would be to elevate the $\delta^{15}N$-determined trophic position of piscivorous fish, and this would be particularly noticeable in large lakes where copepods are relatively more abundant than cladoceran grazers (e.g., *Daphnia*). This could then explain the observed positive pattern between piscivore trophic position and lake size (Vander Zanden et al. 1999, Post et al. 2000).

Our method of calculating trophic position, by comparing the $\delta^{15}N$ stable-isotope signature of a co-occurring herbivorous grazer, assumes that the grazer and the copepod in each lake are feeding within the same food-chain. However, if copepods are not well connected to higher trophic levels in a lake's foodweb, then using *Daphnia* as a reference herbivore may not give us a clear understanding of where copepods are between the microbial loop and the food chain that connects primary

producers to piscivorous fish (Figure 8). Additionally, because of their short generation time, the $\delta^{15}N$ signature of *Daphnia* is less stable over time than the $\delta^{15}N$ signatures of co-occurring obligate predatory zooplankton such as *Chaoborus* and *Epischura,* as well as co-occurring copepods (Matthews and Mazumder 2007). If *Daphnia*, which can consume particles up to 70 µm in diameter (Hambright et al. 2007) are also to some extent consuming heterotrophic flagellates and ciliates, they would not be the pure herbivores that our feeding models assume, which leaves open the question of the actual trophic position of *Daphnia* in these lakes. To get are more accurate sense of trophic position based on $\delta^{15}N$, a more reliable baseline would be needed, one that separates phytoplankton from other heterotrophic microorganisms and is time-integrated. For example, Post (2002) used filter-feeding freshwater mussels for time-integration of the $\delta^{15}N$ signal.

In lakes where the predatory copepod *Epischura* coexisted with a diaptomid copepod, the $\delta^{15}N$ signatures of the diaptomids were nearly always higher than the predator, although the trophic position was similar (Figure S2). It is possible either that the diaptomids had a $\delta^{15}N$ signature because are feeding on different food chains with an enriched $\delta^{15}N$ signatures, such as the microbial loop or *Epischura* may be less predatory than previously thought (Figure S2). While some herbivorous calanoid copepods are a link to higher trophic levels, in marine systems several co-occurring calanoid copepod species are a part of secondary or tertiary food webs that do not link to higher trophic positions (Schukat et al. 2014).

Among the 43 British Columbia lakes we studied, trophic positions of five diaptomid copepods were not universally related to body size, lake area, or

135

productivity, but were instead context-dependent. Despite the relatively wide intraspecific variation in trophic positions exhibited by each of the five copepod species (Figure 3), it was not explained by the variation in body size within species, nor by the lake environment.

Trophic position and body size – In lakes, the relationship between $\delta^{15}N$ trophic position and body size among fishes is largely positive, but for the diaptomid copepods studied here, size only significantly explained the variation in trophic position in mesotrophic lakes where the relationship was actually negative (Table 2). In addition, variation in the physical-chemical environments of the lakes did not provide insight into copepod trophic position since it was unrelated to nutrient concentration, or pH at the time of collection, or the lake size (Figure 7B). Even among fishes, the positive relationship between body size and trophic position tends only to apply for piscivores and breaks down when non-piscivorous species are included (Ou et al. 2017, Keppeler et al. 2020).

For fishes globally, trophic position scales more closely with jaw length than overall body size (Kopf et al. 2021). The physical limitations that govern the relationship between trophic position and body size among fishes does not apply to organisms whose feeding habits are not associated with gape-size. Diaptomid copepods are simply not gape limited, therefore the mechanisms that govern the body-size-trophic-position relationships do not apply. They capture prey with a combination of behaviors including using their feeding appendages to set up a feeding current from which they selectively choose prey (Koehl and Strickler 1981), or quickly move to

ambush prey detected with sensory structures on their first antennae (Koehl and

Strickler 1981, Kiørbe 2008). There is a half a trophic position difference between the

largest and smallest copepods in mesotrophic lakes, which would indicate a transition

of a carnivore biased toward secondary carnivory and as body size increases

transitioning to carnivore biased toward herbivory (Figure 9). Although we did not

consider the sex of the copepods in this study, and simply measured average size of 30

− 100 adults in each sample, diaptomid copepods have substantial sexual body size

dimorphism. Males are between 64% and 97% of the mass of females (based on

published sizes for the two sexes: Table S6), with accompanying differences in

clearance rate (feeding imposed per capita prey mortality) and prey ingestion rate

related to size-specific differences in gut-capacity (Gréve et al. 2017). Additionally,

behavioral differences between male and female copepods can affect foraging if males

spend less time feeding because they are searching for mates (Gréve et al. 2017).

Among the five species in our data set, *A. denticornis* has the greatest sexual size

dimorphism, and only occurs in the mesotrophic lakes. Compensatory feeding (i.e.,

ingesting a greater amount of when food is of lower-quality) has been shown for

copepods, including diaptomids (Chen and Folt 1993, Calliari and Tiselius 2005,

Burian et al. 2018). An omnivory can supplement diet when other sources of nutrition

are deficient, such as consuming primary producers when prey quality is low in the

case of predators (Ugine et al. 2019). Therefore, if males have lower clearance and

ingestion rates because of their smaller size, and display compensatory feeding

behaviors to meet nutritional requirements, we might expect that they would feed at a

higher trophic level by selecting food sources with greater nutritional density, such as

rotifers, compared with phytoplankton. It might be that a difference in sex ratios among lakes could account for the differences in both body size and trophic position within mesotrophic lakes, although we did not determine sex ratios in this study. Furthermore, Hairston et al. (1983) found for *Onychodiaptomus sanguineus* that adult sex ratio became progressively more male biased as seasonal fish predation pressure increased (because females are more visible than males), making it possible that a copepod population's trophic position could change through the season, and so indirectly alter the $\delta^{15}N$ trophic position of top predators.

Trophic position and pH – There was a positive association between trophic position and pH in both oligotrophic lakes and small lakes (Figure 7 A,C). Additionally, pH was slightly higher in large lakes than small lakes (Figure 5). As pH changes among lakes, we might expect to see changes in phytoplankton and zooplankton communities. In lakes with a wider range of pH, phytoplankton community structure can vary with pH, potentially impacting the quality of the food resources available since food quality in terms of nutrient composition and edibility varies among phytoplankton species (Havens and De Costa 1986, Findlay and Kasian 1996). We did not measure the phytoplankton community in this study, however we did observe that the co-occurring herbivorous zooplankton community changes as pH changes among lakes. *Ceriodaphnia* spp. and *Diaphanosoma* spp. occurred at higher pH values, *Holopedium* spp. was found at lower pH, while *Daphnia* spp. occurred throughout the entire pH range (Figure 9). *Daphnia* always occurred where *Holopedium* was present. *Holopedium* occurred in lakes with lower pH, where copepod trophic position tended

to be slightly lower. This suggests that across the range of pH levels, there is a difference in co-occurring herbivore community as well. A larger abundance of cladoceran grazers (*Daphnia* and *Holopedium*) would increase resource competition for copepods. In oligotrophic lakes, where production is likely already low, high herbivore densities may graze down the phytoplankton below levels needed to support large populations of the copepods' animal prey such as rotifers, or reduce rotifer populations through mechanical interference (Gilbert 1988), thus lowering copepod trophic position in lakes with low pH.

Our results indicate that the trophic structure of zooplankton classically considered phytoplankton grazers is complex, containing omnivores and predators as well as herbivores. By placing all zooplankton into a single "grazer" trophic level, trophic connections will be overlooked that lengthen the food chain (Boyce et al. 2015), thus elevating the trophic position of predators higher in the food web. Given the context-dependent relationships between copepod trophic position and lake conditions, it is clear that the mechanisms operating to lengthen or shorten food chain length go beyond those operating at the level of the top predator such as access to littoral resources, or at intermediate trophic levels, such as the addition of an extra species at an intermediate trophic position. The variation in trophic position we observed among intermediate consumers indicates that changes in food chain length, measured at the level of a top predator, may not always be a result of additional levels or additional trophic links, but may be due to the dynamic, changing trophic position of individual intermediate consumer species. While we did not find exactly the same relationships between trophic position and physical and chemical characteristics of a

lake that others have found for fish, it is clear that the factors determining trophic positions of copepods are complex and significant. Since the trophic position of the intermediate consumer itself varies with body size, lake size, and lake water chemistry, studies of food chain length would benefit from examining the variation in trophic position of intermediate consumers that serve as food resources for high trophic levels to explain the lengthening or contraction of food chains.

References

Almeda, R., H. S. Gréve, and T. Kiørboe. 2018. Prey perception mechanism determines maximum clearance rates of planktonic copepods. Limnology and Oceanography 63:2695–2707.

APHA. 1998. Standard Methods for the Examination of Water and Wastewater. 20th ed. American Public Health Association, Washington, DC.

Arim, M., S. R. Abades, G. Laufer, M. Loureiro, and P. A. Marquet. 2010. Food web structure and body size: trophic position and resource acquisition. Oikos 119:147–153.

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

Boyce, D. G., K. T. Frank, and W. C. Leggett. 2015. From mice to elephants: overturning the 'one size fits all' paradigm in marine plankton food chains. Ecology Letters 18:504–515.

Burian, A., J. Grosse, M. Winder, and H. T. S. Boschker. 2018. Nutrient deficiencies and the restriction of compensatory mechanisms in copepods. Functional Ecology 32:636–647.

Bürkner, P.-C. 2017. brms: An R Package for Bayesian multilevel models using Stan. Journal of Statistical Software 80:1–28.

Bürkner, P.-C. 2021. Bayesian item response modeling in R with brms and Stan. Journal of Statistical Software 100:1–54.

Burns, C. W., and M. Schallenberg. 1996. Relative impacts of copepods, cladocerans, and nutrients on the microbial food web of a mesotrophic lake. Journal of Plankton Research 18:683–714.

Buskey, E. J. 1984. Swimming pattern as an indicator of the roles of copepod sensory systems in the recognition of food. Marine Biology 79:165–175.

Buuren, S. van, and K. Groothuis-Oudshoorn. 2011. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 43:1–67.

Cael, B. B., and D. A. Seekell. 2016. The size-distribution of Earth's lakes. Scientific Reports 6:29633.

Calliari, D., and P. Tiselius. 2005. Feeding and reproduction in a small calanoid copepod: *Acartia clausi* can compensate quality with quantity. Marine Ecology Progress Series 298:241–250.

Carlson, R. E. 1977. A trophic state index for lakes. Limnology and Oceanography 22:361–369.

Chen, C. Y., and C. L. Folt. 1993. Measures of food quality as demographic predictors in freshwater copepods. Journal of Plankton Research 15:1247–1261.

Chubaty, A. M., B. O. Ma, R. W. Stein, D. R. Gillespie, L. M. Henry, C. Phelan, E. Palsson, F. W. Simon, and B. D. Roitberg. 2014. On the evolution of omnivory in a community context. Ecology and Evolution 4:251–265.

Davies, J.M., W. H. Nowlin, and A. Mazumder. 2004. Temporal changes in nitrogen and phosphorus codeficiency of plankton in lakes of coastal and interior British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 61:1538–1551.

DeMott, W. R. 1988. Discrimination between algae and artificial particles by freshwater and marine copepods. Limnology and Oceanography 33:397–408.

DeMott, W. R. 1995. Optimal foraging by a suspension-feeding copepod: Responses to short-term and seasonal variation in food resources. Oecologia 103:230–240.

Depaoli, S., and R. van de Schoot. 2017. Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. Psychological Methods 22:240–261.

Doi, H., K. Chang, T. Ando, I. Ninomiya, H. Imai, and S. Nakano. 2009. Resource availability and ecosystem size predict food-chain length in pond ecosystems. Oikos 118:138–144.

Eloranta, A. P., K. K. Kahilainen, P.-A. Amundsen, R. Knudsen, C. Harrod, and R. I. Jones. 2015. Lake size and fish diversity determine resource use and trophic position of a top predator in high-latitude lakes. Ecology and Evolution 5:1664–1675.

Findlay, D. L., and S. E. Kasian. 1996. The effect of incremental pH recovery on the Lake 223 phytoplankton community. Canadian Journal of Fisheries and Aquatic Sciences 53:856–864.

Friedman, M. M., and J. R. Strickler. 1975. Chemoreceptors and feeding in calanoid copepods (Arthropoda: Crustacea). Proceedings of the National Academy of Sciences USA 72:4185–4188.

Gibert, J. P., and J. P. DeLong. 2017. Phenotypic variation explains food web structural patterns. Proceedings of the National Academy of Sciences USA 114:11187–11192.

Gilbert, J. J. 1988. Suppression of rotifer populations by *Daphnia*: A review of the evidence, the mechanisms, and the effects on zooplankton community structure. Limnology and Oceanography 33:1286–1303.

Gliwicz, M. Z. 2003. Between Hazards of Starvation and Risk of Predation: Ecology of Offshore Animals. International Ecology Institute, Oldendorf/Luhe, Germany.

Gréve, H. van S., R. Almeda, M. Lindegren, and T. Kiørboe. 2017. Gender-specific feeding rates in planktonic copepods with different feeding behavior. Journal of Plankton Research 39:631–644.

Grosbois, G., P. A. del Giorgio, and M. Rautio. 2017. Zooplankton allochthony is spatially heterogeneous in a boreal lake. Freshwater Biology 62:474–490.

Hairston, N. G. 1979. The adaptive significance of color polymorphism in two species of *Diaptomus* (Copepoda). Limnology and Oceanography 24:15–37.

Hairston, N. G., Jr., and N. G. Hairston, Sr. 1993. Cause–effect relationships in energy flow trophic structure and interspecific interactions. The American Naturalist 142:379–411.

Hairston, N. G., Jr., W. E. Walton, and K. T. Li. 1983. The causes and consequences of sex-specific mortality in a freshwater copepod. Limnology and Oceanography 28:935-947.

Hambright, K. D., R. W. Drenner, S. R. McComas, and Jr. Hairston. 1991. Gape-limited piscivores, planktivore size refuges, and the trophic cascade hypothesis. Archiv für Hydrobiologie 121:389–404.

Hambright, K. D., N. G. Hairston, W. R. Schaffner, and R. W. Howarth. 2007. Grazer control of nitrogen fixation: Synergisms in the feeding ecology of two freshwater crustaceans. Fundamental and Applied Limnology 170:89–101.

Havens, K. E., and J. De Costa,. 1986. A comparison of phytoplankton responses to nutrient additions in acidic and circumneutral pH lakewater. Hydrobiologia 137:211–222.

Hellmann, C., B. Wissel, and C. Winkelmann. 2013. Omnivores as seasonally important predators in a stream food web. Freshwater Science 32:548–562.

Gabry, J., and T. Mahr. 2022. Bayesplot: Plotting for Bayesian models. R package version 1.9.0.

Jenkins, B., R. L. Kitching, and S. L. Pimm. 1992. Productivity, disturbance and food web structure at a local spatial scale in experimental container habitats. Oikos 65:249.

Keppeler, F. W., C. G. Montaña, and K. O. Winemiller. 2020. The relationship between trophic level and body size in fishes depends on functional traits. Ecological Monographs 90: e01415.

Keppeler, F. W., and K. O. Winemiller. 2020. Incorporating indirect pathways in body size–trophic position relationships. Oecologia 194:177–191.

Kiørbe, T. 2008. A Mechanistic Approach to Plankton Ecology. Princeton University Press, Princeton, New Jersey.

Kjeldgaard, M. K., J. A. Hewlett, and M. D. Eubanks. 2021. Widespread variation in stable isotope trophic position estimates: patterns, causes, and potential consequences. Ecological Monographs 91: e01451.

Koehl, M. A. R., and J. R. Strickler. 1981. Copepod feeding currents: Food capture at low Reynolds number. Limnology and Oceanography 26:1062–1073.

Kopf, R. K., J. D. L. Yen, D. G. Nimmo, S. Brosse, and S. Villéger. 2021. Global patterns and predictors of trophic position, body size and jaw size in fishes. Global Ecology and Biogeography 30:414–428.

Lazzaro, X. 1987. A review of planktivorous fishes: Their evolution, feeding behaviors, selectivities and impacts. Hydrobiologia 146:97–167.

Lenth, R. V. 2022. emmeans: Estimated Marginal Means, aka least-squares means. R package version 1.7.3.

Matthews, B., and A. Mazumder. 2003. Compositional and interlake variability of zooplankton affect baseline stable isotope signatures. Limnology and Oceanography 48:1977–1987.

Matthews, B., and A. Mazumder. 2005. Temporal variation in body composition (C:N) helps explain seasonal patterns of zooplankton $\delta 13C$. Freshwater Biology 50:502–515.

Matthews, B., and A. Mazumder. 2006. Habitat specialization and the exploitation of allochthonous carbon by zooplankton. Ecology 87:2800–2812.

Matthews, B., and A. Mazumder. 2007. Distinguishing trophic variation from seasonal and size-based isotopic ($\delta 15N$) variation of zooplankton. Canadian Journal of Fisheries and Aquatic Sciences 64:74–83.

Minagawa, M., and E. Wada. 1984. Stepwise enrichment of 15N along food chains: Further evidence and the relation between δ15N and animal age. Geochimica et Cosmochimica Acta 48:1135–1140.

Olson, A., A. Frid, J. dos Santos, and F. Juanes. 2020. Trophic position scales positively with body size within but not among four species of rocky reef predators. Marine Ecology Progress Series 640:189–200.

Ou, C., C. G. Montaña, and K. O. Winemiller. 2017. Body size–trophic position relationships among fishes of the lower Mekong basin. Royal Society Open Science 4:160645.

Persson, L., S. Diehl, L. Johansson, G. Andersson, and S. F. Hamrin. 1992. Trophic interactions in temperate lake ecosystems: A Test of Food Chain Theory. The American Naturalist 140:59–84.

Post DM (2002) The long and short of food-chain length. Trends in Ecology and Evolution 17:269–277

Post, D. M., M. L. Pace, and N. G. Hairston. 2000. Ecosystem size determines food-chain length in lakes. Nature 405:1047–1049.

Potapov, A. M., U. Brose, S. Scheu, and A. V. Tiunov. 2019. Trophic position of consumers and size structure of food webs across aquatic and terrestrial ecosystems. The American Naturalist 194:823–839.

Reid, J. W., and C. E. Williamson. 2010. Copepoda. Pages 829–899 *in* J. H. Thorp and A. P. Covich, editors. Ecology and Classification of North American Invertebrates. 3rd edition. Elsevier, Academic Press, Oxford, UK.

Romanuk, T. N., A. Hayward, and J. A. Hutchings. 2011. Trophic level scales positively with body size in fishes. Global Ecology and Biogeography 20:231–240.

Romero-Romero, S., R. González-Gil, C. Cáceres, and J. L. Acuña. 2019. Seasonal and vertical dynamics in the trophic structure of a temperate zooplankton assemblage. Limnology and Oceanography 64:1939–1948.

Romero-Romero, S., A. Molina-Ramírez, J. Höfer, and J. L. Acuña. 2016. Body size-based trophic structure of a deep marine ecosystem. Ecology 97:171–181.

Schukat, A., H. Auel, L. Teuber, N. Lahajnar, and W. Hagen. 2014. Complex trophic interactions of calanoid copepods in the Benguela upwelling system. Journal of Sea Research 85:186–196.

Snowberg, L. K., K. M. Hendrix, and D. I. Bolnick. 2015. Covarying variances: more morphologically variable populations also exhibit more diet variation. Oecologia 178:89–101.

Steinke, M., J. Stefels, and E. Stamhuis. 2006. Dimethyl sulfide triggers search behavior in copepods. Limnology and Oceanography 51:1925–1930.

Takimoto, G., and D. M. Post. 2013. Environmental determinants of food-chain length: a meta-analysis. Ecological Research 28:675–681.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Tewfik, A., S. S. Bell, K. S. McCann, and K. Morrow. 2016. Predator diet and trophic position modified with altered habitat morphology. PLoS ONE 11:e0147759.

Tunney, T. D., K. S. McCann, N. P. Lester, and B. J. Shuter. 2012. Food web expansion and contraction in response to changing environmental conditions. Nature Communications 3:1105.

Ugine, T. A., S. B. Krasnoff, R. J. Grebenok, S. T. Behmer, and J. E. Losey. 2019. Prey nutrient content creates omnivores out of predators. Ecology Letters 22:275–283.

Ventura, M., and J. Catalan. 2008. Incorporating life histories and diet quality in stable isotope interpretations of crustacean zooplankton. Freshwater Biology 53:1453–1469.

Ward, C. L., and K. S. McCann. 2017. A mechanistic theory for aquatic food chain length. Nature Communications 8:2028.

Watkins, J., L. Rudstam, and K. Holeck. 2011. Length-weight regressions for zooplankton biomass calculations- A review and a suggestion for standard equations. Cornell University eCommons, Ithaca, NY.

Williamson, C. E., and N. M. Butler. 1986. Predation on rotifers by the suspension-feeding calanoid copepod *Diaptomus pallidus*. Limnology and Oceanography 31:393–402.

Williamson, C. E., and N. M. Butler. 1987. Temperature, food and mate limitation of copepod reproductive rates: separating the effects of multiple hypotheses. Journal of Plankton Research 9:821–836.

Vander Zanden, M. J., B. J. Shuter, N. Lester, and J. B. Rasmussen. 1999. Patterns of food chain length in lakes: A stable isotope study. The American Naturalist 154:406–416.

Zandonà, E., C. M. Dalton, R. W. El-Sabaawi, J. L. Howard, M. C. Marshall, S. S. Kilham, D. N. Reznick, J. Travis, T. J. Kohler, A. S. Flecker, S. A. Thomas, and C. M. Pringle. 2017. Population variation in the trophic niche of the Trinidadian guppy from different predation regimes. Scientific Reports 7:5770.

TABLES

Table 1. Physical and chemical characteristics of lakes sampled. Z_{max} is maximum depth, Total P is total phosphorus, Total N is total nitrogen, TOC is total organic carbon, and n is the number of times a lake was sampled. * Lakes from which zooplankton stable isotope data has been previously published in Matthews & Mazumder 2003 and/or Matthews & Mazumder 2007.

Lake Name	Z_{max} (m)	Area (ha)	Total P (μgL^{-1})	Total N (mgL^{-1})	pH	TOC (mgL^{-1})	n
Burns	40	1180.4	23	590	7.8	14.6	1
*Butchart	32	70.9	4	110.9	7.1	2.2	2
Buttle	120.7	3530	4	70	7.3	1.3	1
*Council Lake	17	16.37	3.6	115.5	7.5	2.5	14
Cowichan	152	6204.3	3	83	7.4	1	1
*Cusheon	9.5	31	18.6	416.3	7.6	5.3	1
*Deception Reservoir	8	75.8	10.3	237	7.2	3.8	1
Decker	16	1122.5	20	550	7.7	14.8	1
Durrance	16.6	8.37	5	200	7.8	2.6	6
*Elk	16.8	246.8	29.7	457.6	8.2	5.8	8
Florence	5.5	11.9	15	395	7.6	3.2	3
Fork	10	4.03	11.5	310	7.1	4.5	4
Glen	14	16.9	18	535	7.4	5.3	3
*Goldstream	27	74.9	4.6	132.4	7.1	2.7	2
Kathlyn	9.4	170	17	310	7.4	6	1
Langford	17	61.2	21	530	7.8	3.2	7
Long	14	33.6	7	420	7.7	2.7	2
*Lubbe Reservoir	23.8	54	4.6	125.2	7.1	2.5	1
Mabel	200.6	5990	3	140	7.6	1.65	1
Mara	45.7	1940	5.5	120	7.4	2.2	1
McIvor	45.5	123	2	70	7.2	1.2	1
Middle Quinsam	14.6	70.8	3	125	7.4	2	2
Mitchell	7.9	2.95	13.3	262.2	7.4	3.88	1
NoName	NA	NA	2	155	7	1.1	2
Okanagan	242	34442.8	7	210	8.2	4	1
Old Wolf	13	23.62	5	220	6.9	3.1	1
Paul	73	390	8	380	8.4	NA	1
Prospect	14	59.8	12	320	7.5	4.8	6
Quamichan	8	313.64	25	460	7	4.6	1

Table 1. Continued

| Lake Name	Z_{max} (m)	Area (ha)	Total P (μgL^{-1})	Total N (mgL^{-1})	pH	TOC (mgL^{-1})	*n*
Quennell	6.9	119.8	75.5	1110	7.4	10.8	1
Ross	8.2	33.2	16	480	7.1	8	3
Round	20.4	182.1	39	1110	8.1	11.7	2
Seeley	2.7	19.8	21	305	7.3	8.95	3
*Seymour	8	80.1	17.5	500	7.6	11.5	2
*Shawnigan	50	537	4.4	195.2	7	3.1	31
Shuswap	161.5	30960	4	120	7.7	2	1
Skaha	57	1960	8	220	8.2	4	1
SMary's	16.7	182.3	13.6	320.9	7.4	4	1
*Sooke Lake Reservoir	67	428	3.3	91.2	7.5	2.2	32
Sugar	82.6	2080	3	135	7.5	2	1
Teanook	7.6	3.3	7	270	7.3	4.3	3
Tyhee	43.3	308.5	20	690	8.3	9.4	2
Upper Quinsam	48	505.8	5	70	7.1	2	1
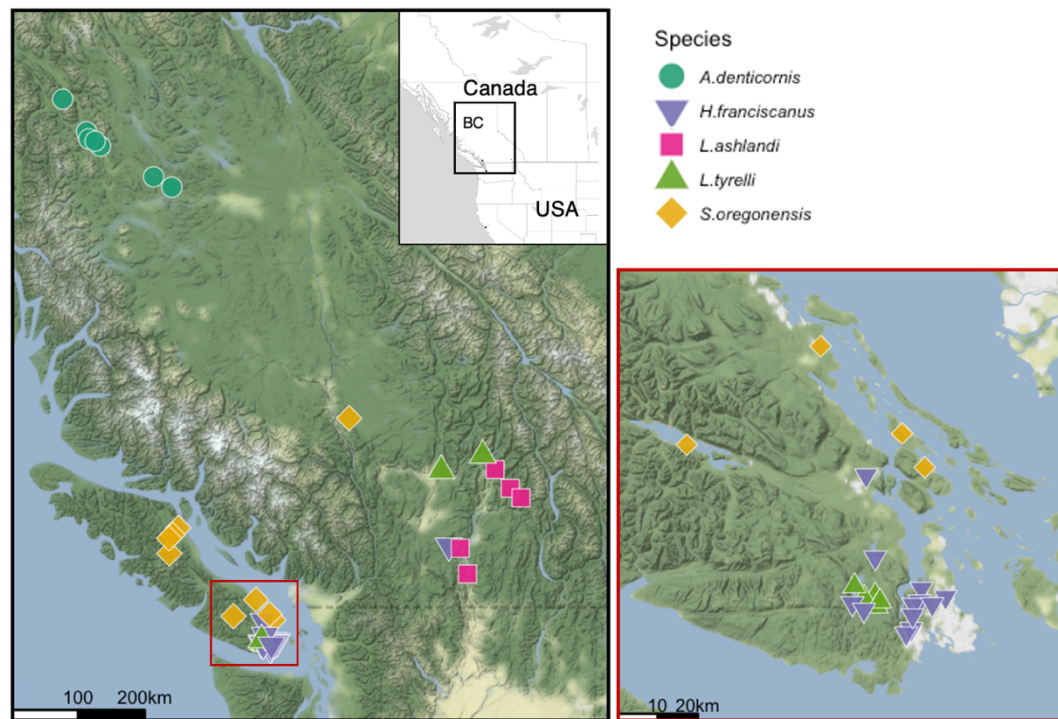
Figure 1. Lake sampling locations of five species of diaptomid copepods across British Columbia, Canada. Point shape and color is the most abundant copepod found in each lake, and used in this study. Inset: Lake sampling locations on Vancouver Island and around the city of Victoria, British Columbia.
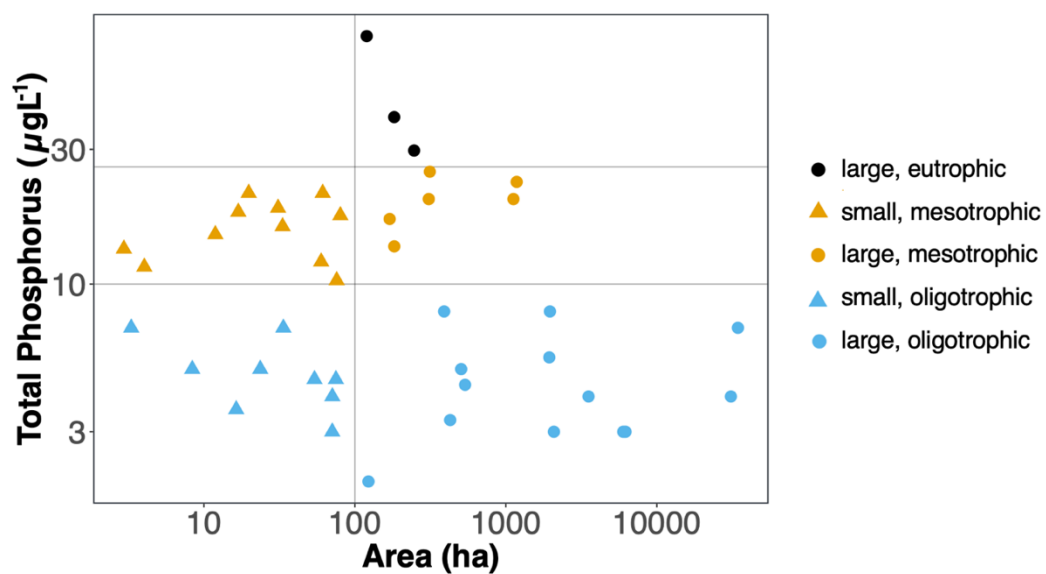
Figure 2. Distribution of total phosphorus and lake surface area of 43 lakes across British Colombia, Canada. Vertical line divides large lakes >100ha from small lakes. Horizontal lines divide lakes into eutrophic >25 μgL^{-1} total phosphorus, mesotrophic $10.1- 24.9$ μgL^{-1} total phosphorus, and oligotrophic <10 μgL^{-1} total phosphorus.
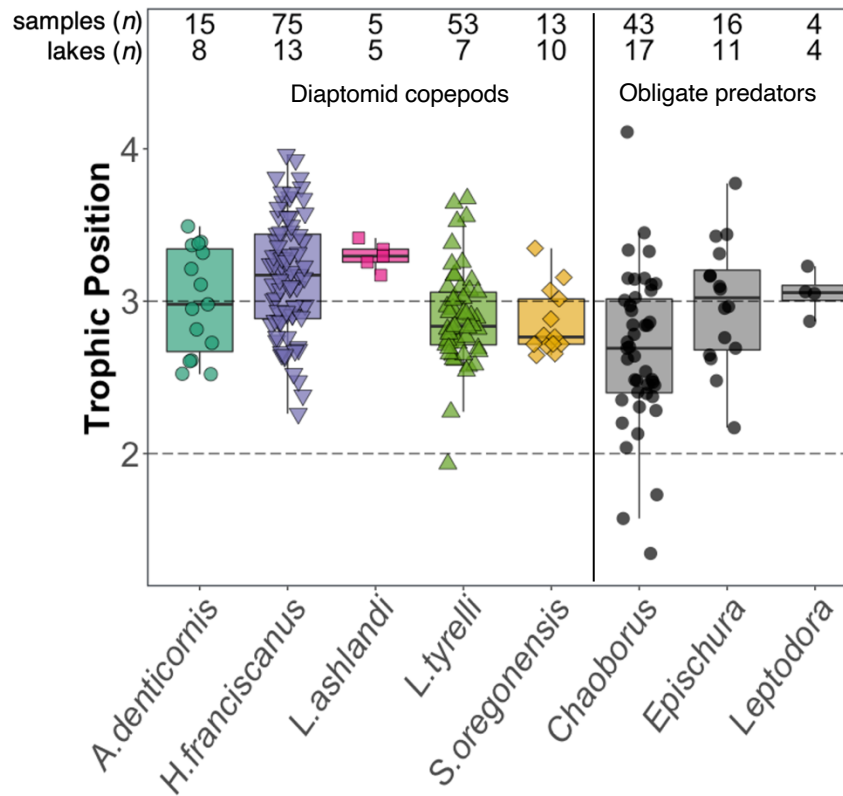
Figure 3. Trophic positions of five diaptomid copepod species and co-occurring obligate predatory zooplankton *Chaoborus*, *Epischura,* and *Leptodora*. Each point represents one sampling date; lakes were sampled unevenly, between one and 32 times over the course of the study. Trophic position 2 represents the expected trophic position of an herbivore, while trophic position 3 represents the expected trophic position of a predator.
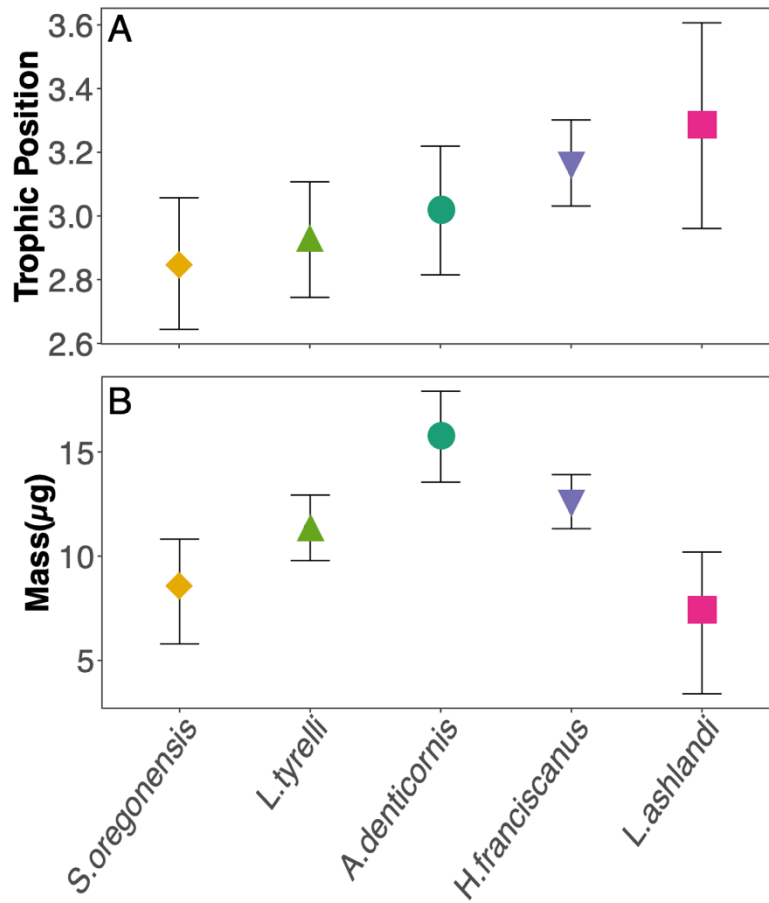
Figure 4. Mean posterior model estimates and 95% credible intervals of (A) trophic position and (B) body mass per individual of five diaptomid copepod species. Note the obvious lack of a relationship overall between trophic position and body mass (See Table 2). (A) *H. franciscanus* has a significantly higher trophic position than *S. oregonensis* or *L. tyrelli,* and *L. ashlandi* has a significantly higher trophic position than *S. oregonensis.* All contrasts were based on one sided non-linear hypothesis testing of posterior probabilities (see text). (B) All pairwise contrasts are based on non-linear hypothesis testing of posterior probabilities (see text). *S. oregonensis* is not significantly different in mass from *L. tyrelli* or *L. ashlandi. H. franciscanus* was not significantly different from *L. tyrelli* based on least squared means contrasts.
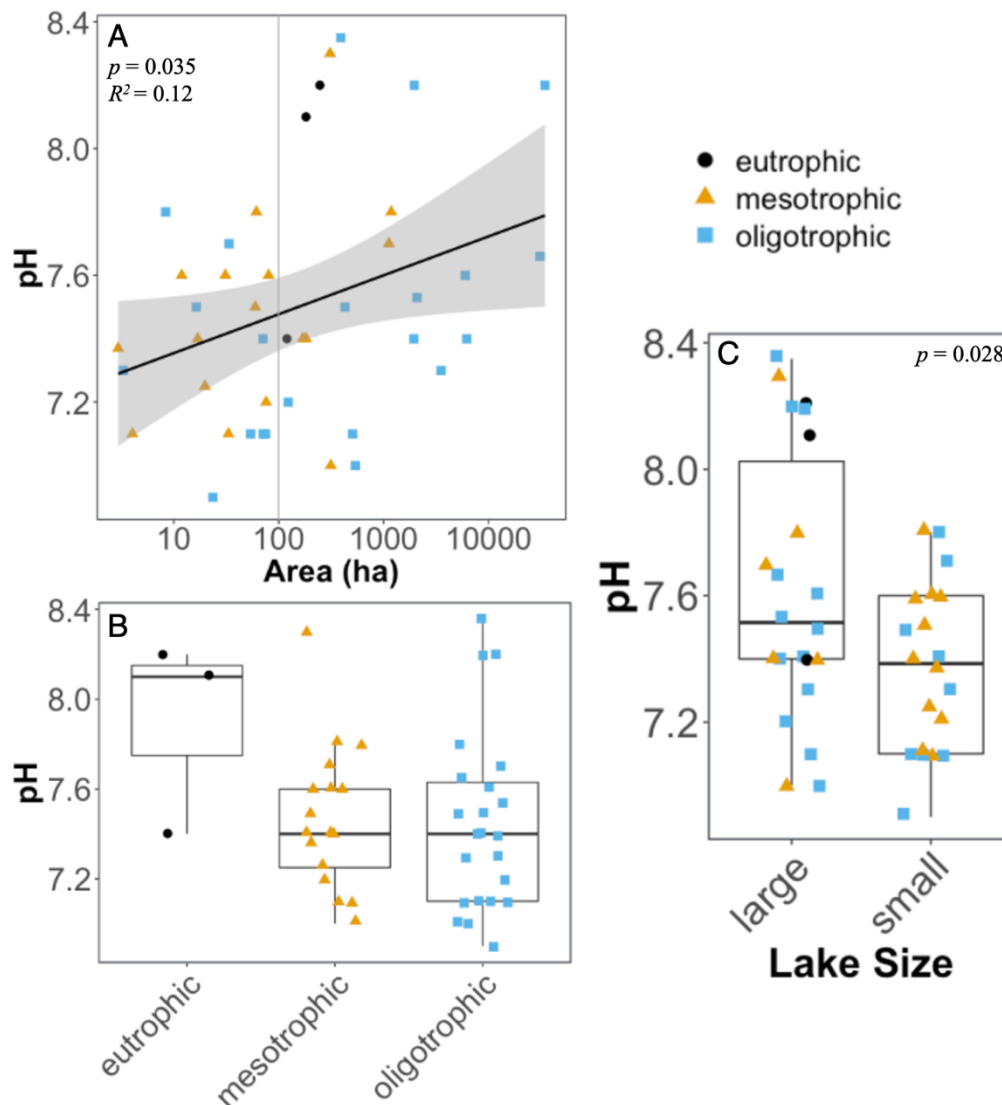
Figure 5. pH variation among 43 lakes across British Columbia, Canada. A) Positive linear relationship with 95% confidence intervals between lake area and pH. Vertical line at 100 ha divides small lakes from large lakes. B) Range of pH in eutrophic, mesotrophic, and oligotrophic lakes. Both mesotrophic and oligotrophic lakes had the same median pH, 7.4. C) Range of pH in large and small lakes
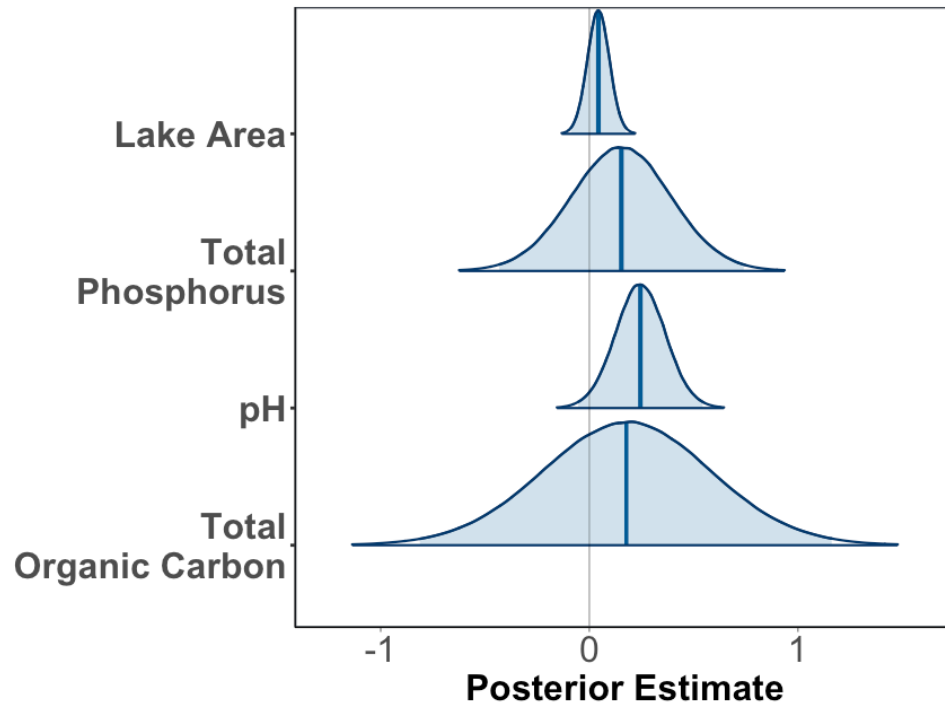
Figure 6. Posterior distributions correlation of lake physio-chemical variables on copepod mean trophic position from 43 lakes in British Colombia Canada. Solid vertical lines line are mean posterior estimates, shaded regions are the 50% uncertainty areas of each distribution. Posterior distributions model estimates for five copepod species, and model uncertainty not shown, see Table S4
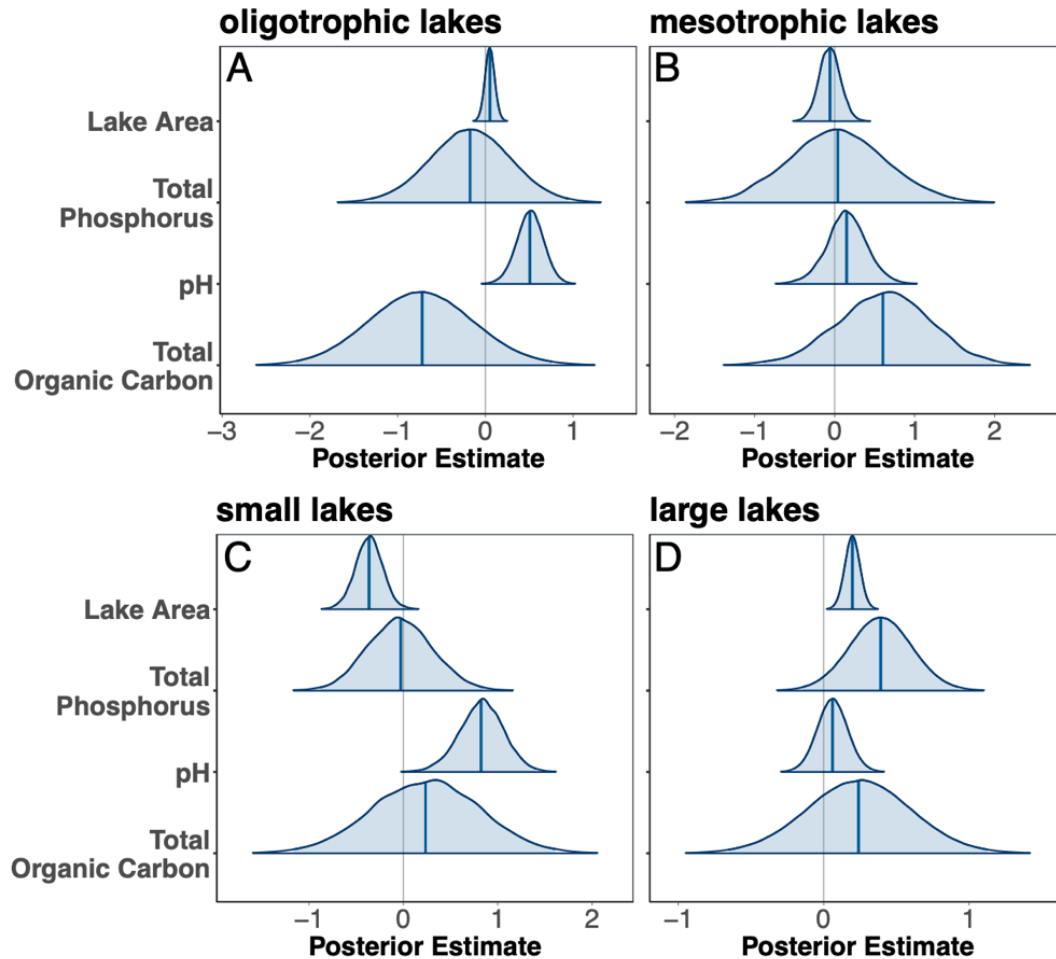
Figure 7. Posterior distributions of the effect of lake environment on trophic position among five species of diaptomid copepod. Vertical line within each distribution is the mean estimate, width of the distribution is the 95% credible interval. A) Posterior distributions of oligotrophic lakes only ($n = 23$), with total phosphorus less than or equal to 10 μgL^{-1}. pH was positively associated with trophic position B) Posterior distributions of mesotrophic lakes only ($n = 17$) with total phosphorus between 10.1-24.9 μgL^{-1}. C) Posterior distributions from 20 small lakes less than 100ha. pH was positively associated with trophic position. D) Posterior distributions from 22 large lakes greater than 100ha. Lake area was positively associated with trophic position.
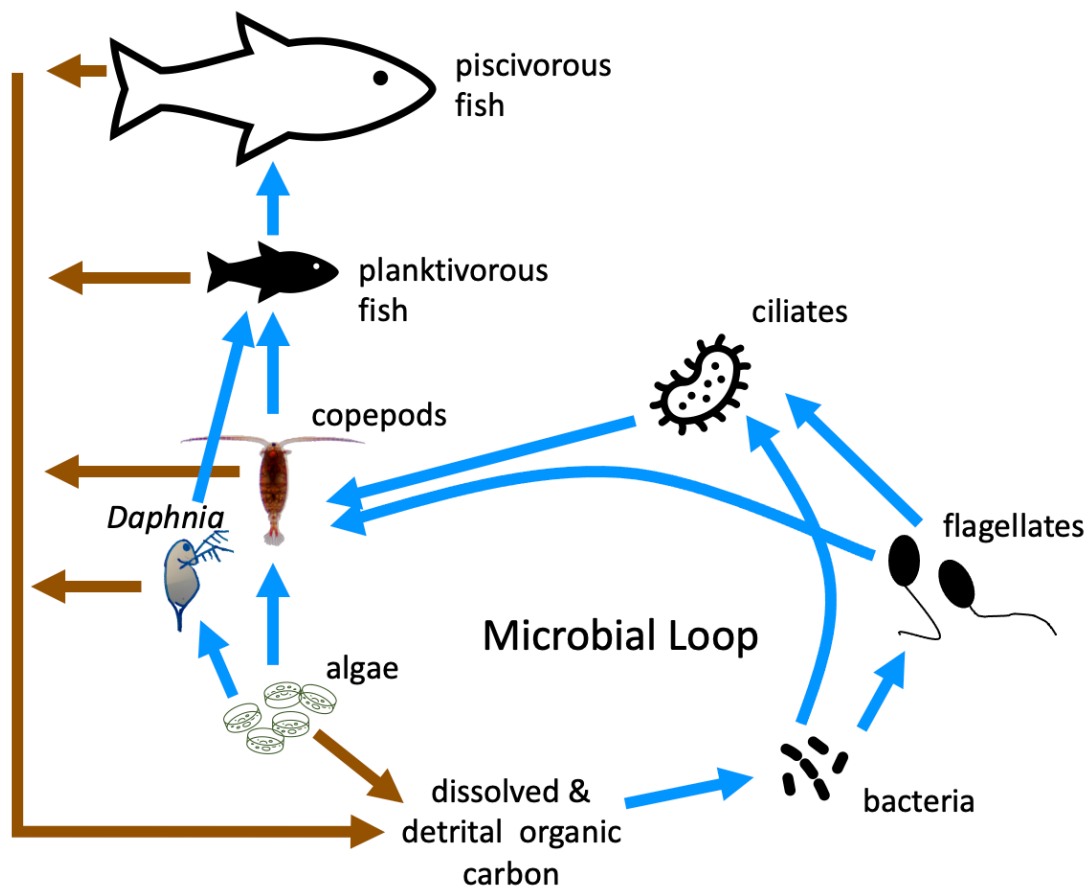
Figure 8. Diagram of a lake food chain, consisting of bacteria, heterotrophic flagellates, ciliates, and copepods that make up the microbial loop, as well as upper levels of a food chain consisting of fish. Blue arrows represent trophic pathways, while brown arrows represent waste products. Copepods connect the microbial loop to the linear food chain. As $\delta^{15}N$ becomes enriched with each trophic transfer, $\delta^{15}N$ signatures of copepods are elevated due to cycling of $\delta^{15}N$ in the microbial loop.
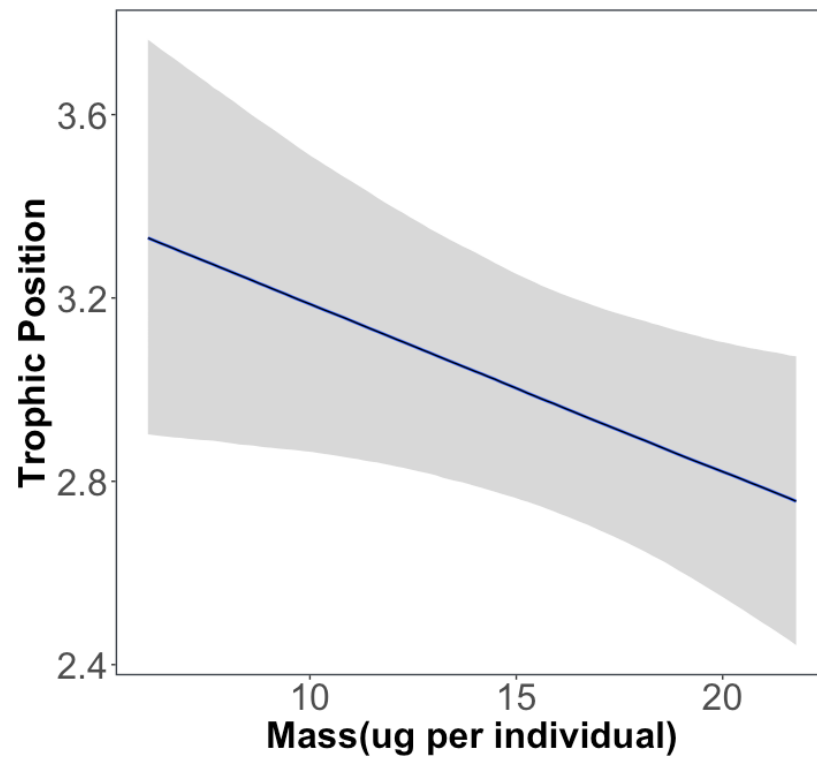
Figure 9. Posterior relationship between diaptomid trophic position and body mass in mesotrophic lakes. Shaded area is the 95% credible interval.
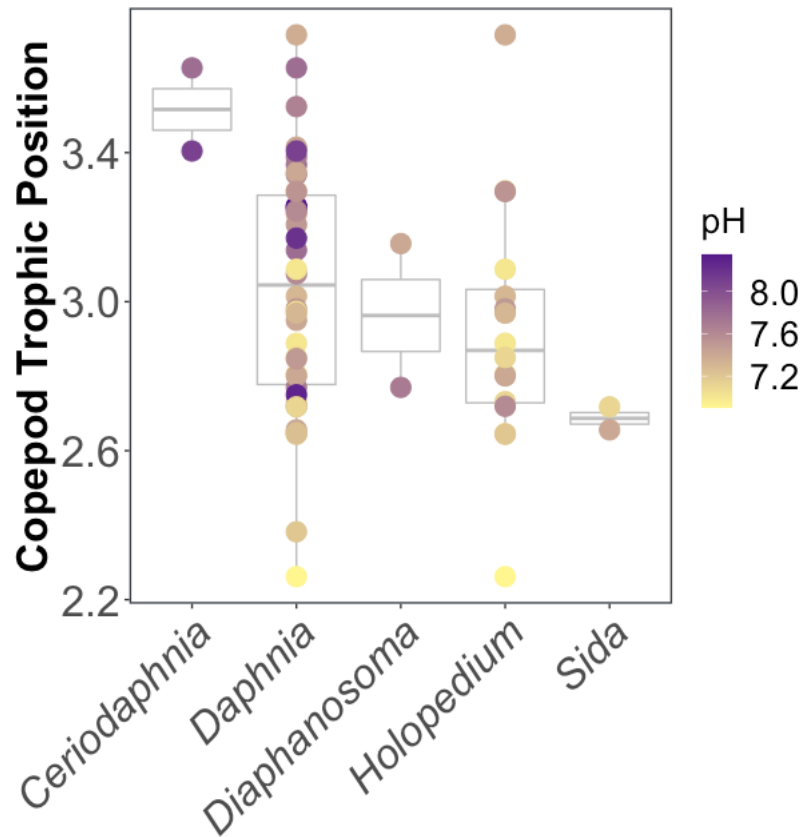
Figure 10. Five grazers used to set the $\delta^{15}N$ base for calculating copepod trophic position. Each point is the trophic position of the co-occurring copepod from each sampling date. Color of each point is the pH of the lake. Copepods co-occurring with *Daphnia* occur across a range of trophic positions and pH. *Holopedium* occurs in lakes with lower pH than *Daphnia.*

SUPPLEMENT

TABLES

Table S1. Outlier trophic position values removed from the data analysis because they are four times greater or less than the mean Cook's distance.

Species	Lake	$\delta^{15}N$	Trophic Position
L. tyrelli	Council	3.11	1.8
H. franciscanus	Fork	7.44	1.9
H. franciscanus	Langford	13.3	4.4
L. ashlandi	Kalamalka	9.95	2.7
L. ashlandi	Osoyoos	10.82	2.7
A. denticornis	Round	12.43	3.7
A. denticornis	Round	14.13	3.7
A. denticornis	Round	5.79	1.9
S. oregonensis	Long	11.65	3.9

Table S2 Prior distribution parameters for Bayesian models.

Model	Model parameter	Prior
Difference in trophic position among species: trophic position ~ 0 + copepod species + (1|lake)	α	normal μ=0, σ=4
	β *A. denticornis*	normal μ=2.5, σ=1
	β *H. franciscanus*	normal μ=2.5, σ=1
	β *L. ashlandi*	normal μ=2.5, σ=1
	β *L. tyrelli*	normal μ=2.5, σ=1
	β *S. oregonensis*	normal μ=2.5, σ=1
	st. dev. lake	student t υ=3, μ=0, σ=2.5
	st. dev. lake intercept	student t υ=3, μ=0, σ=2.5
	σ	student t υ=3, μ=0, σ=2.5
Difference in body size among species: copepod body size ~ 0 + copepod species	α	normal μ=0, σ=4
	β *A. denticornis*	normal μ=25.2, σ=12
	β *H. franciscanus*	normal μ=20.2, σ=12
	β *L. ashlandi*	normal μ=9.7, σ=12
	β *L. tyrelli*	normal μ=18.7, σ=12
	β *S. oregonensis*	normal μ=10.9, σ=12
	σ	student t υ=3, μ=0, σ=4.9
Difference in body size among lake trophic states copepod body size ~ lake trophic state + (1|copepod species)	intercept	normal μ=18.29, σ=12
	β oligotrophic	normal μ=0, σ=1
	sd copepod species	student t υ=3, μ=0, σ=4.6
	sd intercept copepod species	student t υ=3, μ=0, σ=4.6
	σ	student t υ=3, μ=0, σ=4.6
Relationship between trophic position and body size: trophic position ~ copepod body size + copepod species + (1|lake)	β copepod body size	normal μ=0, σ=1
	β *H. franciscanus*	normal μ=0, σ=1
	β *L. ashlandi*	normal μ=0, σ=1
	β *L. tyrelli*	normal μ=0, σ=1
	β *S. oregonensis*	normal μ=0, σ=1
	intercept	normal μ=2.5, σ=1
	st. dev. lake	student t υ=3, μ=0, σ=2.5
	st. dev. lake intercept	student t υ=3, μ=0, σ=2.5
	σ	student t υ=3, μ=0, σ=2.5
Intraspecific relationship between trophic position and body size: trophic position ~ copepod body size + (1|lake)	α	normal μ=0, σ=1
	β copepod body size	normal μ=0, σ=1
	st. dev. lake	student t υ=3, μ=0, σ=2.5
	st. dev. lake intercept	student t υ=3, μ=0, σ=2.5
	σ	student t υ=3, μ=0, σ=2.5

Table S2. Continued

Model	Model parameter	Prior
Difference in trophic position among lake trophic states or lake size: trophic position ~ lake trophic state + copepod species + (1| lake)	intercept	normal μ=2.5, σ=1
	β mesotrophic lake	normal μ=0, σ=1
	β oligotrophic lake	normal μ=0, σ=1
	β *H. franciscanus*	normal μ=0, σ=1
	β *L. ashlandi*	normal μ=0, σ=1
	β *L. tyrelli*	normal μ=0, σ=1
	β *S. oregonensis*	normal μ=0, σ=1
	st. dev. lake	student t υ=3, μ=0, σ=2.5
	st. dev. intercept lake	student t υ=3, μ=0, σ=2.5
	σ	student t υ=3, μ=0, σ=2.5
Relationship between trophic position and lake environment: trophic position ~ Total P + TOC + pH + Area + copepod species	β log Area	normal μ=0, σ=1
	β log TOC	normal μ=0, σ=1
	β log TP	normal μ=0, σ=1
	β pH	normal μ=0, σ=1
	β *H. franciscanus*	normal μ=0, σ=1
	β *L. ashlandi*	normal μ=0, σ=1
	β *L. tyrelli*	normal μ=0, σ=1
	β *S. oregonensis*	normal μ=0, σ=1
	intercept	normal μ=2.5, σ=1
	σ	student t υ=3, μ=0, σ=2.5
Intraspecific relationship between trophic position and lake environment: trophic position ~ Total P + TOC + pH + Area	β log Area	normal μ=0, σ=1
	β log TOC	normal μ=0, σ=1
	β log TP	normal μ=0, σ=1
	β pH	normal μ=0, σ=1
	intercept	student t υ=3, μ=3.1, σ=2.5
	σ	student t υ=3, μ=0, σ=2.5

Table S3. Posterior contrasts between pair-wise median trophic levels among five copepod species. Lower and Upper are the lower and upper bounds of the 95% credible intervals. Estimates with * are contrasts in which the median trophic positions between the two species were statistically significantly different.

Contrast	Trophic Position Difference			Mass Difference		
	Estimate	Lower	Upper	Estimate	Lower	Upper
A. denticornis - H. franciscanus	-0.15	-0.39	0.09	3.21*	0.64	5.93
A. denticornis - L. ashlandi	-0.26	-0.64	0.12	8.45*	4.51	12.89
A. denticornis - L. tyrelli	0.10	-0.17	0.37	4.48*	1.90	6.89
A. denticornis - S. oregonensis	0.16	-0.12	0.45	7.29*	3.70	10.92
H. franciscanus - L. ashlandi	-0.11	-0.46	0.23	5.18*	1.93	8.96
H. franciscanus - L. tyrelli	0.25*	0.02	0.46	1.30	-0.83	3.23
H. franciscanus - S. oregonensis	0.32*	0.07	0.56	4.02*	1.55	6.87
L. ashlandi - L. tyrelli	0.36	-0.02	0.73	-3.92*	-7.99	-0.46
L. ashlandi - S. oregonensis	0.43*	0.05	0.81	-1.15	-5.48	2.95
L. tyrelli - S. oregonensis	0.07	-0.20	0.35	2.74	-0.07	5.98

Table S4. Model posterior estimates of the relationship between trophic position and lake physico-chemical conditions.

Model Parameter	Posterior Estimate	95% Credible Interval
β log total phosphorus	0.153	$-0.43 - 0.74$
β log total organic carbon	0.178	$-0.81 - 1.16$
β pH	0.245	$-0.05 - 0.54$
β log area	0.044	$-0.09 - 0.18$
Intercept	3.04	$-1.39 - 2.83$
β *H. franciscanus*	0.204	$-0.18 - 0.58$
β *L. ashlandi*	0.288	$-0.19 - 0.76$
β *L. tyrelli*	0.108	$-0.32 - 0.53$
β *S. oregonensis*	0.170	$-0.41 - 0.44$
σ	0.037	$0.24 - 0.39$

Table S5. Posterior model estimates of the intraspecific relationship between trophic position and physico-chemical environment.

A. denticornis

Model Parameter	Posterior Estimate	95% Credible Interval
β log total phosphorus	0.51	-1.09 - 1.99
β log total organic carbon	0.53	-1.08 - 2.01
β pH	-0.15	-0.88 - 0.68
β log area	0.28	-0.24 - 0.78
Intercept	2.32	-2.82 - 7.38

H. franciscanus

Model Parameter	Posterior Estimate	95% Credible Interval
β log total phosphorus	0.11	-0.94 - 1.16
β log total organic carbon	0.07	-1.63 - 1.77
β pH	0.44	-0.28 - 1.13
β log area	-0.17	-0.53 - 0.19
Intercept	-0.07	-5.05 - 5.17

S. oregonensis

Model Parameter	Posterior Estimate	95% Credible Interval
β log total phosphorus	0.35	1.26 - 1.00
β log total organic carbon	0.19	1.86 - 1.01
β pH	-0.28	0.55 - 1.01
β log area	0.06	0.30 - 1.02
Intercept	4.37	10.09 - 1.02

Table S6. Published body size estimates and size differences between male and female copepods. Size estimates are body length reported in Reid and Williamson 2010 converted to body mass per individual based on length-weight regression from Watkins *et al.* 2011. Lake trophic states are the trophic states where each species occurs in this study.

| Species	Median size (range) (μg individual^{-1})		Size difference (μg individual^{-1}) and male to female size ratio	Lake trophic states
	Female	Male		
A. denticornis	29.05 (25.19 − 32.90)	18.66	10.39 0.64	mesotrophic
H. franciscanus	24.07 (10.92 − 37.21)	23.04 (8.86 − 37.21)	1.03 0.96	mesotrophic, oligotrophic
L. tyrelli	21.91 (10.92 − 32.90)	20.88 (8.86 − 32.90)	1.03 0.95	oligotrophic
L. ashlandi	10.80 (5.78 − 15.81)	10.50 (5.19 − 15.81)	0.30 0.97	oligotrophic
S. oregonensis	12.15 (8.48 − 15.81)	11.18 (6.55 − 15.81)	0.97 0.92	mesotrophic, oligotrophic
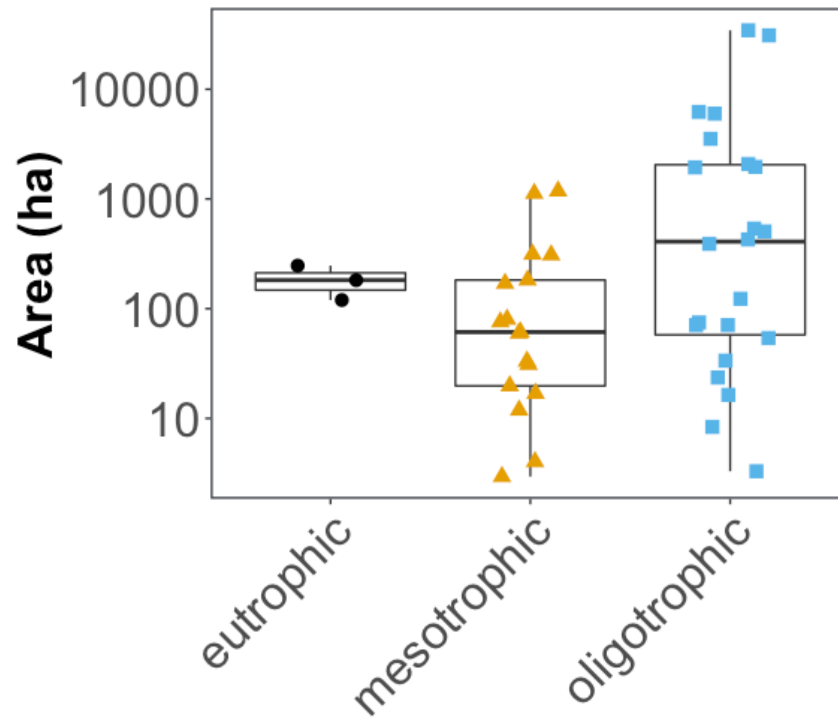
Figure S1. Distribution of lake sizes among three lake productivity levels. Despite the difference in median and spread, there is no statistically significant difference in area between mesotrophic and oligotrophic lakes (T-test, p = 0.12, $F_{1,40}$ = 2.57).
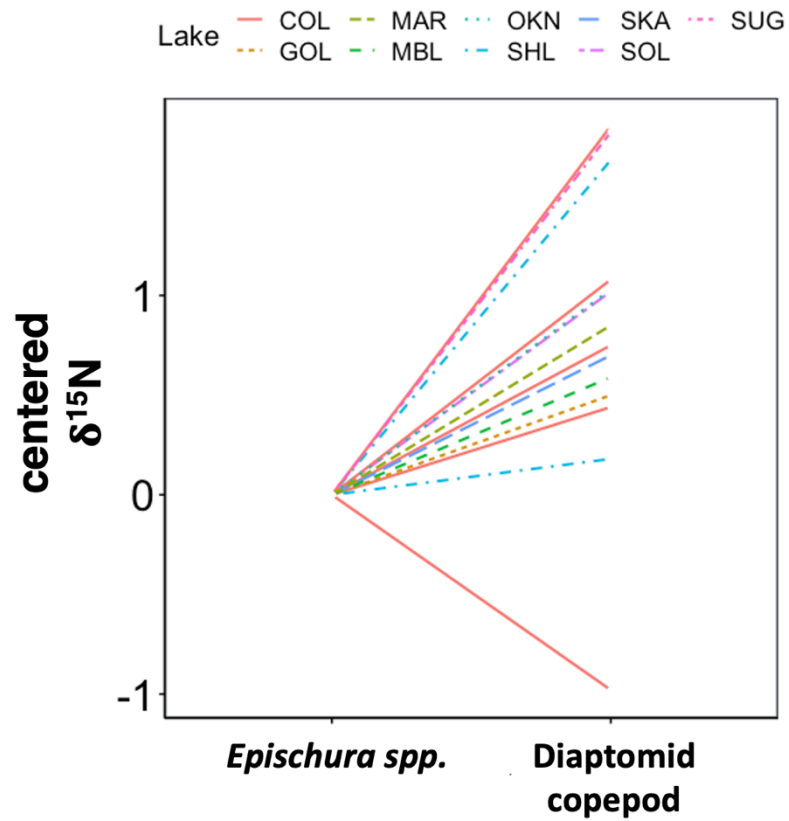
Figure S2. Comparison of $\delta^{15}N$ stable-isotope signatures of the predatory copepod *Epischura* spp. with the co-occurring diaptomid copepod. Each line represents one sampling point. $\delta^{15}N$ of *Epischura* spp. was set at 0, the $\delta^{15}N$ of the co-occurring diaptomid copepod is the relative $\delta^{15}N$ value compared to *Epischura* spp.