

TROPHIC METACOMMUNITIES:  
LESSONS FROM STUDYING BROMELIADS  
AND  
TEACHING PROGRAMMING IN  
BIOSTATISTICS

*by*

LAURA MELISSA GUZMAN URIBE

B.Sc. (Hons), McGill University, 2012

M.Res. (Distinction), Imperial College London, 2013

A thesis submitted in partial fulfillment of  
the requirements for the degree of

DOCTOR OF PHILOSOPHY

*in*

The Faculty of Graduate and Postdoctoral Studies  
(Zoology)

The University of British Columbia  
(Vancouver)

APRIL 2019

© Laura Melissa Guzman Uribe, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a dissertation entitled:

Trophic metacommunities: Lessons from studying Bromeliads  
And Teaching programming in biostatistics

---

submitted Laura Melissa Guzman Uribe in partial fulfillment  
by \_\_\_\_\_ of the requirements for  
the degree Doctor of Philosophy  
of \_\_\_\_\_  
in Zoology  
\_\_\_\_\_

**Examining Committee:**

Diane Srivastava, Zoology

---

Supervisor

Mary O'Connor, Zoology

---

Supervisory committee member

Leticia Aviles, Zoology

---

University Examiner

Jeannette Whitton, Botany

---

University Examiner

**Additional Supervisory Committee Members:**

Sarah Otto, Zoology

---

Supervisory committee member

Loren Rieseberg, Botany

---

Supervisory committee member

Vinicius Farjalla, UFRJ

---

Co - Supervisor

## ABSTRACT

---

Trophic metacommunity ecology brings together the spatial thinking of metacommunity ecology and the complexity of food web ecology. While theoretical development in this field has been bountiful, empirical development has been slower. Using a diverse methodology, I bring together three different empirical approaches to understand trophic metacommunities as exemplified by bromeliads macro-invertebrates. First, I used Markov network analysis to study the effect of regional environmental gradients on community composition and trophic interactions. I found that a gradient in precipitation underlies the spatial turnover of some species and that the interactions of certain predators differed due to differences in bromeliad water volume. Second, I combined experimental feeding trials and a food web model to study the effect of body size diversity at the local scale on food web dynamics. I found that predator persistence was maximized when the minimum prey size in the community was intermediate, but as prey diversity increased the minimum body size could take a broader range of values due to Jensen's inequality. Third, I used population genetics to estimate dispersal kernels of a predator and a prey. I then used these empirical estimates of dispersal kernels and feeding rates to parameterize a trophic metacommunity model, to study the effect of differences in dispersal between a predator and a prey on persistence. From the empirical dispersal kernel estimates, I found that the prey dispersed up to 25 km whereas the predator dispersed up to 200 m. From the trophic metacommunity model, I found that differences in dispersal rates were sufficient to generate differences in occupancy of our modelled landscape, without requiring variation in the abiotic niche. None of this

work would have been possible without strong programming skills and a good understanding of statistics. In my final chapter, I studied the effect of using cognitive load theory to design R programming assignments for undergraduate biostatistics courses. I found that students that learned R through our assignments rated their programming ability higher and were more likely to put the usage of R as a skill in their CVs than control students.

## LAY SUMMARY

---

Food webs show which species eat each other. For my Ph.D., I investigated how food webs change when: (i) the environment changes, (ii) species have different body sizes, (iii) or species move different distances between generations. I studied a food web of aquatic invertebrates that live inside bromeliad plants. First, I found that a predatory crane fly interacted negatively with other species when the water volume in the plant was low. Second, I found that the damselfly predator needed mid-sized prey to survive. Third, I found that the crane fly moved up to 23 km(!), while the damselfly moved only 200m. Finally, I wanted to teach undergraduate students in biology how to program and do statistics at the same time. I designed homework assignments using educational theories about how to teach two concepts simultaneously. These were successful: students felt motivated and confident in their skills.

## PREFACE

---

CHAPTER 2 has been previously published as:

Guzman, L. M., Vanschoenwinkel, B., Farjalla, V. F., Poon, A., and Srivastava, D. S. (2018). A precipitation gradient drives change in macroinvertebrate composition and interactions within bromeliads. *PloS one*, 13(11), e0200179.

I conceived of the idea, alongside Vinicius Farjalla and Diane Srivastava. I collected and processed the data. Anita Poon assisted in processing the data. I analyzed the data and Bram Vanschoenwinkel and Diane Srivastava supervised the analysis. I wrote the manuscript with comments from Diane Srivastava, Anita Poon and Bram Vanschoenwinkel. Permit number 47164-1 was provided by Ministério do Meio Ambiente (MMA), Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) and Sistema de Autorização e Informação em Biodiversidade (SISBIO). This field study did not involve endangered or protected species.

CHAPTER 3 has been submitted to the Proceedings of the Royal Society B and it is under review.

I conceived the idea, ran the experiments, collected the data, analyzed the data, wrote the model and wrote the manuscript under the supervision of Diane Srivastava.

### CHAPTER 4

I conceived the idea alongside Vinicius Farjalla and Diane Srivastava. I collected the invertebrates and extracted the DNA. Extracted DNA of the

odonate predator was sent to the Cornell Institute for Genomic Diversity and extracted DNA of the tipulid prey was sent to University of Wisconsin-Madison Biotechnology centre to conduct GBS and sequencing. I performed the bioinformatics, wrote the model and the manuscript under the supervision of Diane Srivastava.

## CHAPTER 5

I conceived the idea. I wrote the homework assignments with the help of Ellen Nikelski and the supervision of Matthew Pennell and Diane Srivastava. I analyzed the data and wrote the manuscript with comments from Ellen Nikelski, Matthew Pennell and Diane Srivastava. This work was conducted with review and approval by the Behavioural ethics research board of the University of British Columbia, H16-02319. The project title was "Evaluation of R instruction in Zoology undergraduate statistics courses"

# CONTENTS

---

ABSTRACT . . . . .	iii
LAY SUMMARY . . . . .	v
PREFACE . . . . .	vi
CONTENTS . . . . .	viii
LIST OF TABLES . . . . .	xi
LIST OF FIGURES . . . . .	xii
ACKNOWLEDGEMENTS . . . . .	xiv
DEDICATION . . . . .	xvi
1 INTRODUCTION. . . . .	1
2 A PRECIPITATION GRADIENT UNDERLIES CHANGE IN MACROINVERTEBRATE COMPOSITION AND INTERACTIONS WITHIN BROMELIADS.. . . .	10
2.1 Chapter summary . . . . .	11
2.2 Introduction . . . . .	12
2.3 Methods . . . . .	16
2.4 Results . . . . .	22
2.5 Discussion . . . . .	27
2.6 Conclusion . . . . .	32
3 PREY BODY MASS AND DIVERSITY DETERMINE FOOD WEB PERSISTENCE. 33	
3.1 Chapter summary . . . . .	34
3.2 Introduction . . . . .	35
3.3 Methods . . . . .	38
3.4 Results . . . . .	47
3.5 Discussion . . . . .	52
4 EMPIRICALLY DETERMINED DIFFERENCES BETWEEN A PREDATOR AND ITS PREY ALLOW INCREASED PERSISTENCE IN A METACOMMUNITY 60	
4.1 Chapter summary . . . . .	61

4.2	Introduction . . . . .	62
4.3	Methods . . . . .	65
4.4	Results . . . . .	72
4.5	Discussion . . . . .	80
5	SUCCESSFUL INTEGRATION OF DATA SCIENCE IN UNDERGRADUATE BIOSTATISTICS COURSES USING COGNITIVE LOAD THEORY. . . . .	84
5.1	Chapter summary . . . . .	85
5.2	Introduction . . . . .	86
5.3	Methods . . . . .	89
5.4	Results . . . . .	97
5.5	Discussion . . . . .	105
5.6	Limitations . . . . .	108
5.7	Conclusion . . . . .	108
6	DISCUSSION . . . . .	109
	BIBLIOGRAPHY . . . . .	114
	APPENDICES . . . . .	137
A	SUPPLEMENTARY INFORMATION TO CHAPTER 1. . . . .	137
A.1	Abstract . . . . .	138
A.2	Introduction . . . . .	138
A.3	Past and present metacommunity theory . . . . .	141
A.4	Spatial use properties and their consequences for pairwise trophic interactions across scales . . . . .	148
A.5	Predicted effects of (co)variation in spatial use properties on trophic metacommunity dynamics . . . . .	160
A.6	Future directions: building and testing future metacommunity theory based on spatial use properties . . . . .	162
A.7	Conclusion . . . . .	170
B	SUPPLEMENTARY INFORMATION TO CHAPTER 2. . . . .	173
B.1	Partitioning beta diversity . . . . .	173
B.2	Environmental variation between sites . . . . .	174
B.3	Validation of Markov Network method . . . . .	179
B.4	Supporting results . . . . .	181
C	SUPPLEMENTARY INFORMATION TO CHAPTER 3. . . . .	193
C.1	Supporting results . . . . .	193
C.2	Asymmetric competition . . . . .	202

D	SUPPLEMENTARY INFORMATION TO CHAPTER 4. . . . .	207
D.1	Modifications to DNA extraction protocols . . . . .	207
D.2	Supporting results . . . . .	208
E	SUPPLEMENTARY INFORMATION TO CHAPTER 5. . . . .	213
E.1	Homework assignments . . . . .	213
E.2	Surveys . . . . .	220
E.3	Codebooks . . . . .	228
E.4	Supporting results . . . . .	237

## LIST OF TABLES

---

TABLE 3.1	Parameter values . . . . .	46
TABLE 3.2	Estimated parameters in the functional response . . . . .	48
TABLE 4.1	Parameter values used for simulations. . . . .	73
TABLE 5.1	Course structure . . . . .	91
TABLE 5.2	Self-perceived emotions of the students . . . . .	100
TABLE A.1	Competitive and trophic metacommunity response variables 144	
TABLE A.2	Synthesis of metacommunity ecology . . . . .	149
TABLE A.3	Spatial use properties and measurable traits . . . . .	164
TABLE B.1	Environmental gradients between sites . . . . .	175
TABLE B.2	Interaction strengths in known module . . . . .	180
TABLE B.3	Pairwise Tukey tests. . . . .	181
TABLE B.3	Pairwise Tukey tests. . . . .	182
TABLE B.3	Pairwise Tukey tests. . . . .	183
TABLE B.3	Pairwise Tukey tests. . . . .	184
TABLE B.4	Linear regression for relative interaction strengths . . . . .	184
TABLE B.4	Linear regression for relative interaction strengths . . . . .	185
TABLE B.5	Permutation test p-value . . . . .	191
TABLE C.1	Prey abundance used in feeding trials . . . . .	194
TABLE E.1	Perceived difficulty of the material . . . . .	238

## LIST OF FIGURES

---

FIGURE 1.1	<i>Neoregelia cruenta</i> at Jurubatiba national park . . . . .	5
FIGURE 1.2	Bromeliad patch at Jurubatiba national park . . . . .	6
FIGURE 1.3	Bromeliads form larger patches of habitat . . . . .	7
FIGURE 2.1	Map of sites along the coast of Brazil . . . . .	18
FIGURE 2.2	Community composition across sites . . . . .	24
FIGURE 2.3	Relative strength of species interactions . . . . .	25
FIGURE 2.4	Tipulid's interactions . . . . .	26
FIGURE 3.1	Type II functional response. . . . .	48
FIGURE 3.2	Time series of predator and prey abundance . . . . .	50
FIGURE 3.3	Predator persistence with different prey size . . . . .	51
FIGURE 3.4	Prey diversity has a unimodal relationship with predator persistence . . . . .	58
FIGURE 3.5	Food web persistence . . . . .	59
FIGURE 4.1	Map of sites along the coast of Brazil . . . . .	66
FIGURE 4.2	Odonate and tipulid's population structure . . . . .	76
FIGURE 4.3	Site of origin for odonate and tipulid individuals . . . . .	77
FIGURE 4.4	Estimated dispersal rate . . . . .	78
FIGURE 4.5	Persistence in trophic metacommunity model . . . . .	79
FIGURE 5.1	Students responses to the surveys. . . . .	99
FIGURE 5.2	Probability that students used R . . . . .	104
FIGURE A.1	Three forms of movement . . . . .	150
FIGURE A.2	Trophic metacommunity simulation outcomes . . . . .	152
FIGURE A.3	Spatial use properties in a three species module . . . . .	171
FIGURE B.1	Total precipitation using WorldClim data . . . . .	176
FIGURE B.2	Total precipitation using weather stations . . . . .	177
FIGURE B.3	Mean water volume increases with precipitation . . . . .	178
FIGURE B.4	Predators have more negative interactions and prey have more positive interactions . . . . .	186
FIGURE B.5	Relative interaction strength and water volume . . . . .	187
FIGURE B.6	Permutation of interaction terms . . . . .	188

FIGURE B.7	Permutation of negative slopes . . . . .	.189
FIGURE B.8	Permutation of positive slopes . . . . .	.190
FIGURE B.9	Presence vs. relative interaction strength . . . . .	.192
FIGURE C.1	Parameter functions vs. body size . . . . .	.195
FIGURE C.2	Parameter space explored by MCMC . . . . .	.196
FIGURE C.3	Time series of two prey species with high competition . . . . .	.197
FIGURE C.4	Time series of two prey species with low competition . . . . .	.198
FIGURE C.5	Jensen's inequality . . . . .	.199
FIGURE C.6	Predator persistence and prey diversity . . . . .	.200
FIGURE C.7	Predator persistence and prey body size . . . . .	.201
FIGURE C.8	Time series of two prey with assymetric competition . . . . .	.204
FIGURE C.9	Predator persistence with assymetric prey competition . . . . .	.205
FIGURE C.10	Prey diversity decreases predator persistence . . . . .	.206
FIGURE D.1	Identity by state similarity . . . . .	.209
FIGURE D.2	Cross validation error and loglikelihood of major clades . . . . .	.210
FIGURE D.3	Population segregation of potential sub-clades . . . . .	.211
FIGURE D.4	Cross validation error of different population structures . . . . .	.212

## ACKNOWLEDGEMENTS

---

Diane Srivastava has been a very patient mentor and supervisor. I am grateful for all her guidance and her endless effort to improve my writing.

I want to thank my committee members: Mary O'Connor, Sally Otto, Loren Rieseberg and Vinicius Farjalla. Your guidance and support have been indispensable.

In particular I want to thank certain members of the committee for exceptional help: Vinicius for Brazilian fieldwork logistics, Sally for help with model construction, Loren for providing lab space free. I also want to thank Bram Vanschoenwinkel for allowing me to come and do research in Brussels.

The working groups during my Ph.D. have been an integral part of my education. The trophic metacommunity working group was one of the best part of my Ph.D. Thank you for meeting every week and having awesome discussions about ecology. I want to thank: Patrick Thompson, Rachel Germain, Sam Strauss, Coreen Forbes, Dominique Gravel, Adam Ford, Mary O'Connor and Diane Srivastava. The sTURN working group helped me push the boundaries of what metacommunity theory means, and what we can do with it. In particular I want to thank: Jon Chase, Zsofi Horváth, Robert Ptacnik, Luc De Meester, Stéphanie Gascón, Maria Anton-Pardo, Pieter Lemmens, Alienor Jeliaskov, Duarte Viana and Bram Vanschoenwinkel. Finally, the bromeliad working group let me meet all the great people that love bromeliads, in particular Regis Cereghino, Ignacio Barberis and Kurt Trzcinski.

This whole Ph.D. couldn't have happened without key support of numerous people. I want to thank some of the people who made this possible: in the field by Juliana Leal and Nayara Gomez, in the lab Anita Poon, Sadie Garcia,

Kathleen Higgins and Winnie Cheung, and finally for the R assignments Ellen Nikelski.

A large part of my Ph.D. was trying to improve the way we teach R. I wanted to thank the people who showed me that research on teaching is possible, and fun: Miranda Meets, Lacey Samuels and Sunita Chowira.

I want to thank past, present and new members of the Srivastava lab: Sarah Amundrud, Natalie Westwood, Pierre Rogi, Nadia Paez, Keerthikrutha Seetharaman, Angie Nicolas, Alatheia Lethaw, and last but definitely not least Andrew McDonald.

I am thankful to everyone in the BRC (and adjacent buildings) that have made these past years a lot of fun. In particular I want to thank Lucho Camacho, Alejandra Echeverri, Santiago David, Manny Boehm.

I want to thank my family Sonia Uribe and Carlos Guzman for all your support, enthusiasm and love that have helped me get here. I also want to thank my new family, Brenna, Richard, Alec, Steve and Laura Pennell. Thank you for welcoming me to your family.

Matthew Pennell, thank you. Thank you for celebrating everything with me and for being a shoulder to cry on. Your love and support have made this experience be the best it could have been.

DEDICATION

*To my parents*

Sonia Uribe and Carlos Guzman

INTRODUCTION

---

The theories of island biogeography (MacArthur and Wilson, 1967), and metapopulation ecology (Hanski *et al.*, 1997) have long encouraged ecologists to think deeply about the role of space and dispersal on community dynamics. More recently, the theory of metacommunity ecology has highlighted that species interactions, in combination with space and dispersal, also have dynamical consequences for communities at local and regional scales (Leibold *et al.*, 2004). In a pioneering paper, which set the stage for the next 15 years of research in metacommunity ecology, Leibold *et al.* (2004) summarized the theoretical literature and found four major combinations of processes that enabled species that would otherwise competitively exclude each other, to coexist regionally (i.e., across multiple populations). These four combinations, which they originally term paradigms (and more recently renamed as "archetypes"; (Leibold and Chase, 2017)) were: i) species sorting, where intermediate levels of dispersal allow species to track their environmental optima; ii) mass effects, where high dispersal allows species to persist in sub-optimal environments; iii) patch dynamics, where dispersal-colonization trade-offs between species could allow regional persistence; and iv) neutral dynamics, where species are ecologically equivalent but stochastic dispersal and drift leads to co-existence. These four paradigms have been so successful as a rhetorical tool that they have become nearly synonymous with the term metacommunity ecology.

After a decade or so of empirical studies classifying metacommunities into these paradigms (typically following variance partitioning procedures outlined by Cottenie (2005)), researchers have increasingly recognized that the four metacommunity paradigms are neither mutually exclusive nor comprehensive, and that multiple mechanisms may promote coexistence simultaneously (Logue *et al.*, 2011; Thompson *et al.*, 2017) (to their credit, Leibold *et al.* (2004) recognized this in this original paper). Not only can multiple mechanisms promote coexistence, but these paradigms are just points along

a multi-dimensional continuum of community dynamics that results from variation in dispersal, the strength of the abiotic niche and the outcome of species interactions (Thompson *et al.* in prep).

Furthermore, while species interactions are widely recognized as key to understanding the dynamics of communities, both theoretical and empirical work, guided by the paradigms, has focused primarily on interactions between competitors. A number of researchers have recognized this limitation and tried to incorporate trophic interactions into metacommunity thinking (e.g. Holt, 2002; Gravel *et al.*, 2011); however, these efforts are quite scattered and disconnected. During my Ph.D. I co-led a collaborative network of researchers that aimed to create a coherent vision for a truly multi-trophic theory of metacommunities (see Appendix A of this thesis and Guzman *et al.* (2019)). This conceptual work, while not part of my dissertation, provides an overarching theme for the research presented here. In our synthesis paper, we argue that a theory of trophic metacommunities is necessary (i) when we need to predict food web properties, which are incompatible with a competitive framework, and (ii) when interacting species use space at different scales; for example, when a predator population interacts with multiple smaller-scale prey populations. In addition, we argue that trophic metacommunity dynamics arise when interacting trophic levels differ in the way they use space. These differences in space use arise from differences in 'spatial use properties', which we define as population level properties that reflect how species use space. These spatial use properties include three forms of movement — dispersal, migration, and foraging —, as well as the abiotic niche and spatial information processing. Spatial information processing refers to the ability of individuals to direct and control their movement based on biotic and abiotic conditions.

This trophic metacommunity framework is well suited for a system like the macroinvertebrate food webs that inhabit bromeliads. Bromeliads have inter-

locking leaves which form a tank that collects water, referred to as a phytotelm (Figure 1.1); the phytotelm provides a habitat for aquatic macroinvertebrates. Bromeliads are an ideal system to study metacommunity ecology since every bromeliad is a discrete patch of habitat (i.e., a sensu stricto metacommunity), and bromeliads are clustered in space, often forming patches of bromeliads (Figure 1.2 and 1.3). They create a hierarchical patchy spatial structure for these insect communities. In my research I focused on the communities living in a particular species of bromeliad — *Neoregelia cruenta*. *N. cruenta* is found along the sand dunes (called restingas) of coastal Brazil, from the north of Rio de Janeiro state to São Paulo state. Bromeliad macroinvertebrate communities are often comprised of aquatic larvae which include filter feeders, scrapers, shredders, collectors, and top predators. Both the spatial distribution of the plants and the diversity of invertebrates make the phytotelms of *N. cruenta* an ideal system to study trophic metacommunities.

As a result of decades of observational and experimental studies (many of these spearheaded by my advisor, Dr. Srivastava and other members of the Bromeliad Working Group), we know a lot about the ecology of these communities including that there is substantial variation in different species spatial use properties. For instance, the species within this food web have different tolerances to drought (Amundrud and Srivastava, 2015), which results in variation in species composition across the landscape. We also know that there is substantial variation in body size among the invertebrates (Céréghino *et al.*, 2018), which suggests that they might also differ in foraging scale or dispersal distance. However despite these results, our knowledge of the spatial use properties of the insects remains far from complete.

For my Ph.D., I decided to study trophic metacommunities in the bromeliad system using three different approaches. For my first chapter, I studied the effect of environmental gradients on species interactions. We often think of an

FIGURE 1.1: Collecting water from a bromeliad using a turkey baster. Some insects, especially culex mosquitoes that live in the water column are collected this way. This picture was taken at the Jurubatiba national park in Rio de Janeiro, Brazil.



FIGURE 1.2: A patch of bromeliads close together under a shrub. This picture was taken at the Jurubatiba national park in Rio de Janeiro, Brazil.



FIGURE 1.3: Under each shrub a patch of bromeliads aggregates forming larger patches of habitat. This picture was taken at the Jurubatiba national park in Rio de Janeiro, Brazil.



environmental gradient as a series of filters selecting which species can persist at different points on the gradient. However, environmental conditions can also change the way species interact. I evaluated the effect of a regional environmental gradient on community composition and the strength of species interactions. The species interactions were inferred from co-occurrences using Markov Networks.

For my second and third chapters, I parameterized theoretical models with data from bromeliad communities (consumption rates and dispersal, respectively). I was motivated to parameterize these models because, despite the remarkable theoretical advances that these models represent, many community ecology models are primarily heuristic and are often disconnected from natural systems. In particular, although the general predictions of some theoretical models have been tested using clever experimental designs, these models are rarely parameterized with empirical data. For my second chapter, I specifically looked at the persistence and coexistence mechanisms that shape food webs at the local scale, which will in turn affect trophic metacommunity dynamics at the regional scale. To do this, I studied the effect of trait diversity and body size at the local scale on food web dynamics. I did feeding trials with the top predator and multiple prey species of different size and I used this data to parameterize a food web model. For my third chapter I empirically estimated the dispersal kernel of a predator and a prey pair using population genetics, and I combined this information with the feeding rate of predators to create a trophic metacommunity model. I evaluated whether observed differences in dispersal were sufficient to generate differences in the spatial occupancy of landscapes between the predator and the prey.

None of this work could have been possible without the strong statistical and programming skills, that I developed over the course of my Ph.D., and which, more broadly, have been found to be important and necessary for

many quantitative areas in ecology. Using this knowledge, I wanted to empower future generations of biologists before they became graduate students by teaching them programming. I decided to approach this task with the same scholarly rigor that I applied to my ecological research. I studied whether we could teach R programming in biostatistics courses using cognitive load theory. The results of this educational research comprise my final chapter.

A PRECIPITATION GRADIENT UNDERLIES CHANGE IN  
MACROINVERTEBRATE COMPOSITION AND INTERACTIONS  
WITHIN BROMELIADS. <sup>1</sup>

---

---

<sup>1</sup>Previously published as Guzman, L. M., Vanschoenwinkel, B., Farjalla, V. F., Poon, A., Srivastava, D. S. (2018). A precipitation gradient drives change in macroinvertebrate composition and interactions within bromeliads. *PloS one*, 13(11), e0200179.

## 2.1 CHAPTER SUMMARY

Ecological communities change across spatial and environmental gradients due to (i) changes in species composition and (ii) changes in the frequency or strength of interactions. Here we use the communities of aquatic invertebrates inhabiting clusters of bromeliad phytotelmata along the Brazilian coast as a model system for examining variation in multi-trophic communities. We first document the variation in the species pools of sites across a geographical climate gradient. Using the same sites, we also explored the geographic variation in species interaction strength using a Markov network approach. We found that community composition differed along a gradient of spatially correlated water volume within bromeliads due to the spatial turnover of some species. From the Markov network analysis, we found that the interactions of certain predators differed due to differences in bromeliad water volume. Overall, this study illustrates how a multi-trophic community can change across an environmental gradient through changes in both species and their interactions.

## 2.2 INTRODUCTION

Ecological communities can change across spatial and environmental gradients in three main ways: the composition of species can change, the strength of interactions between species can change, or the presence of the interactions can change (Tylianakis and Morris, 2017; Poisot *et al.*, 2012). Species composition can vary across an environmental gradient if the environment filters particular traits (Leibold *et al.*, 2004; Kraft *et al.*, 2015) and across space if species differ in their dispersal abilities (Leibold *et al.*, 2004; Kraft *et al.*, 2015). Even when species are found together across a gradient, the presence or strength of interactions between these species can vary between sites on the gradient. For example, consumers may find prey more efficiently in structurally simple habitats, resulting in stronger interactions than in complex habitats (Laliberté and Tylianakis, 2010). Consumption rates can also be higher in warmer sites, due to temperature-dependence of metabolic rates (Rall *et al.*, 2012). Here we combine multiple analyses to show how environmental and spatial gradients affect both the composition and interactions of species in a multi-trophic ecological community.

Estimating changes in species interaction strengths between sites is notoriously difficult, much more so than estimating changes in community composition (e.g. Anderson, 2001; Borcard *et al.*, 2011). For example, pairwise competition experiments consider interactions between two species. These experiments would need to be performed at multiple environments to estimate changes in interaction strengths (Maser *et al.*, 2007). Even then, such experiments would ignore the influence of other species in the estimates of interaction strengths. In order to measure interactions in a community context (i.e. including indirect effects), researchers have experimentally removed one species from the system and assessed the impact on the whole community. However, this approach cannot reconstruct the strength of the interactions be-

tween all members of a community, only the interactions between the removed species and the rest of the community (Paine, 1966).

Inferring species interactions from observational data, as opposed to experimental manipulations, has the advantage of observing the end result of multiple direct and indirect interactions. For instance, combining observations of prey abundance and predator foraging rates can provide information on interaction strengths (Wootton, 1997). However, this method cannot estimate indirect interactions. Another popular method, checkerboard analyses, can determine if observational patterns in species co-occurrence differ from random assembly (Stone and Roberts, 1990; Gotelli, 2000); that is, checkerboard analyses attempt to estimate the effect of competition in shaping the distribution of species. However, such analyses do not explicitly test for differences in interaction strengths.

Markov networks are a promising method to get information about species interaction strengths from observational data while controlling for indirect interactions between species (Harris, 2016). A Markov network relates the probability of the occurrence of multiple species at a site to (i) one parameter for every species  $\alpha$  which determines how much the presence of a given species contributes to the probability of observing the presences and absences of all species in that site and (ii) one parameter for every pair of species  $\beta$  which determines how much the co-occurrence of a given pair of species contributes to the same probability. Given an observed vector of presences and absences, we can use maximum likelihood estimation to obtain the parameters  $\alpha$  and  $\beta$  (Harris, 2016). If species are less likely to occur together, their interaction strength (parameter  $\beta$ ) will be negative. And, conversely, if species are more likely to occur together, their interaction strength (parameter  $\beta$ ) will be positive. Some weaknesses of this approach are that Markov network analyses only calculate symmetrical interaction strengths, if a species is very

rare at the regional scale it may appear to be negatively interacting with many species, and this approach does not explicitly relate occurrences to environmental drivers. This method was developed for competitive communities that show a checkerboard distribution. A checkerboard distribution refers to an arrangement where two species are found to always occupy different patches. This distribution might be the outcome of some exclusion process (competition or predation) (Stone and Roberts, 1990). This reasoning suggests that we can infer interaction strengths in certain types of simple multi-trophic communities that also display checkerboard distributions (see also Harris, 2016).

Although a predator cannot persist in the absence of its prey in a closed system, open systems with a high colonization rate of the prey and a high predation rate can also display a checkerboard distribution between the predator and the prey. When the predator consumes its prey to extinction, we may find the predator on its own. If the prey has a high colonization rate, it can colonize patches where the predator is absent. These colonization - extinction dynamics can lead to a system with patch dynamics (e.g. Englund *et al.*, 2009). In other words, the spatial scale can affect the degree of co-occurrence observed between predators and prey; at small scales, effective predators should reduce or eliminate their prey (negative co-occurrence) while at larger scales predators and prey should positively co-occur (Freilich *et al.*, 2018).

Here we define interaction strength as a measure of the degree of co-occurrence between pairs of species, akin to measuring the correlation between the occurrence of two species (Berlow *et al.*, 2004). Note that this definition of interaction strength does not map to biomass or energy flux between trophic levels, but rather conforms to one of Berlow *et al.*'s (2004) definitions of in-

teraction strengths as a statistical pattern of co-occurrence at a given spatial scale.

Despite their potential, Markov methods have thus far not been used to reconstruct interactions in real food webs along environmental and spatial gradients. Good candidate ecosystems for such analyses are insular systems with simple food webs that occur over wide geographic areas. In such ecosystems, species interactions are contained within each replicate of the system and the environment can vary between systems. A classical food web model system that fits these criteria are the aquatic communities that live inside bromeliad plants in the Neotropics; these communities often occur as clusters that exchange species via dispersal. Bromeliad plants have leaves that interlock, forming a cavity where water accumulates. Inside these cavities, communities of aquatic invertebrates form a food web (Figure 2.1). In these communities, a suite of voracious predators can limit the abundance of prey species (Hammill *et al.*, 2015a), and prey colonization is rapid (Hammill *et al.*, 2015b). In addition, multiple studies have shown that environmental variation (such as water volume) can determine the presence of certain species and mediate the interactions between some predators and their prey (Amundrud and Srivastava, 2016, 2015). Thus bromeliad communities are a suitable model system to test how environment and space can affect species interactions and community composition (Petermann *et al.*, 2015; Farjalla *et al.*, 2012).

Making use of this model system, we explored three main questions. First, we tested whether environmental conditions varied between our sampling sites, located along a geographic gradient. Due to spatial variation in precipitation at the time of sampling, we would expect that sites will vary in the amount of water present in the plants. Second, we described change in community composition along this geographic gradient, and then partitioned beta diversity into either spatial turnover or nestedness of species assemblages

— specifically nestedness of community composition as a geographic pattern. We expect that beta diversity would be driven mostly by nestedness of species assemblages. Specifically, since the amount of water in the bromeliads determines habitat size, we expect that lower water volumes reduce diversity in the community, and that sites with lower water volumes would have a subset of the species of the sites with higher water volumes (Patterson and Atmar, 1986; Baselga, 2010). Third, we used Markov networks to quantify species interactions at each site. We explored whether differences between sites in the strength of species interactions could be explained by geographic variation in environmental conditions. We expect that species interactions would vary along this gradient, because water volumes determine the ability of some predators to persist.

## 2.3 METHODS

### 2.3.1 *Model system*

Tank bromeliads accumulate water inside their leaf axils, providing habitat for communities of aquatic macroinvertebrates (Kitching, 2000). Inside each bromeliad, these aquatic macroinvertebrates interact to form a food web comprised of detritivores, filter feeders, intermediate predators and top predators. Bromeliad macroinvertebrate communities are known to be particularly sensitive to changes in precipitation, since this can change the amount of habitat available for the invertebrates (Pires *et al.*, 2016). For example, drought in bromeliads is known to reduce growth rates of some invertebrate species (Amundrud and Srivastava, 2015). Therefore, we expect that changes in precipitation have the potential to substantially affect species interactions and community composition.

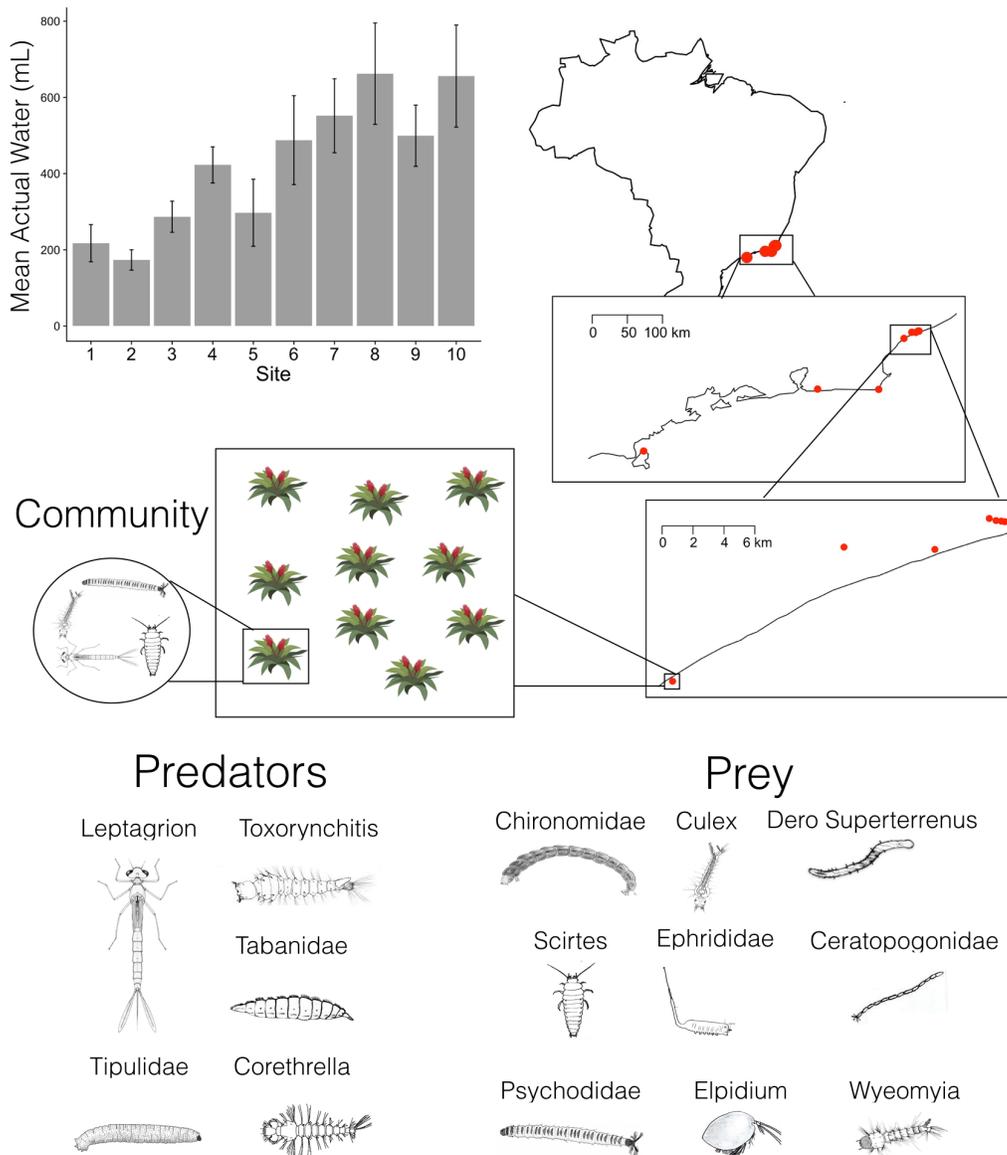
### 2.3.2 *Study Area*

The study area was located in the sand dunes of coastal Brazil (Figure 2.1), in the states of Rio de Janeiro and São Paulo. We sampled ten sites, seven of which were within the Jurubatiba National Park in Rio de Janeiro state, Brazil. The other three sites were located in the sand dunes of Arraial do Cabo (Rio de Janeiro), Marica (Rio de Janeiro), and Ilha Bela (São Paulo). This sampling design resulted in the sites closest to Jurutabita National Park receiving low precipitation, the sites close to Marica receiving intermediate precipitation, and the sites closest to Ilha Bela and Arraial do Cabo receiving high precipitation in the month immediately before sampling (February and March 2015, Appendix B, Figure B.1, B.2).

### 2.3.3 *Sampling*

We sampled all macroinvertebrate communities between March and May 2015. In each site, we dissected ten bromeliads (totalling 100 bromeliads across all sites) to collect all the invertebrates in each plant. The invertebrate samples were preserved in 99% ethanol. Macroinvertebrates were counted and identified to genus level whenever possible. Overall we identified between 11 and 16 genera for each site. For every bromeliad, we measured a suite of environmental variables to assess the amount and quality of habitat available to the invertebrates (Appendix B). For example, actual water volume (mL) was measured by emptying the plant and calculating how much water it contained and maximum water volume (mL) was calculated by emptying the plant and calculating how much water the plant could hold before it overflowed. Since bromeliad invertebrates prefer particular bromeliad sizes (Srivastava *et al.*, 2008), we chose the same broad range of bromeliad sizes for every site, to ensure that we obtained the spectrum of species present in the site.

FIGURE 2.1: Sites are located along the eastern coast of Brazil. A site is comprised of ten bromeliads found within 100 meters of each other. Ten sites were sampled, with a hierarchy of distances between bromeliads (nested boxes, right side of diagram). The mean water volume found in the bromeliads from each site is shown. Bars represent mean and standard error of the mean. Sites 1 to 10 are ordered from north to south. A community is the set of species found in one bromeliad. The bromeliad macroinvertebrate community is comprised of predators and prey.



#### 2.3.4 *Data analysis*

ENVIRONMENTAL VARIATION BETWEEN SITES — In order to test whether sites did indeed vary in environmental conditions, we performed an ANOVA for most of the environmental variables. For two variables measured on a percentage scale, oxygen saturation and canopy cover, this ANOVA procedure was inappropriate so we used an analogous generalized linear model specifying a binomial family error distribution (Appendix B: Table B.1).

COMPOSITIONAL VARIATION BETWEEN SITES — We tested for differences in community composition between sites using permutational multivariate analysis of variance using distance matrices (function `adonis` in R package `vegan`, hereafter referred to as `Adonis`). Multivariate tests of dispersion (function `betadisper` in R package `vegan`) were used to test for differences in community variation (beta diversity) between geographic sites. We summarized abundances according to genus so that these results would be comparable with the species interactions analyses. For the `Adonis` analysis, we tested if bromeliads from different sites and containing different water volumes differed in community composition. We used water volume in this analysis since, of all the environmental variables, it differed the most between sites (Appendix B, Table B.1). For the multivariate test of dispersion, we tested if bromeliads from different sites differed in their beta diversity, where within-site beta diversity was measured as the average dissimilarity of bromeliad invertebrate communities from the centroid in multivariate space (Anderson, 2001, 2006). To visualize the differences in community composition and dispersion, we used non-metric multidimensional scaling (NMDS) plots (Anderson, 2001) (Figure 2.2b). An NMDS plot shows both the differences between sites in their average community composition (position of centroids) as well as

differences between sites in beta diversity (the standardised residuals around the centroids).

To further understand our results, we partitioned beta diversity: by nestedness of assemblages and by spatial turnover of species. Nestedness of species assemblages occurs when some sites have a smaller subset of the species than other richer sites (Baselga, 2010). This pattern could result if lower water volumes in bromeliads exclude certain species without replacement. Spatial turnover of species occurs when some species are replaced by others (Baselga, 2010). This pattern could result if some species can persist in low water volumes and other sets of species can persist in high water volumes. To calculate the different portions of beta diversity we used Baselga’s method (Baselga, 2010), where Sørensen dissimilarity ( $\beta_{SOR}$ ) is partitioned into pure spatial turnover ( $\beta_{SIM}$ ) and nestedness ( $\beta_{NES}$ ) (See Appendix B for details in the equations used).  $\beta_{NES}$  is not an absolute measure of nestedness but instead a measure of the dissimilarity of communities due to the effect of nestedness patterns. To visualize nestedness and turnover we used a nestedness and degree fill plot (Figure 2.2c).

Overall, we tested differences in community composition between sites using Adonis, compared the differences in beta diversity between sites using Betadisper, and finally evaluated if the differences in community composition between sites can be attributed to species nestedness or turnover using partitioning of beta diversity. Adonis tests if the differences in community composition between sites are significant, and partitioning of beta diversity relates patterns in either nestedness or turnover to the compositional dissimilarity between sites.

**SPECIES INTERACTIONS** — To obtain species interactions strengths, we used Markov network analysis (Harris, 2016). This method does not make any assumptions about the topology of the food web, nor do we have to

define which species might interact with each other. The method calculates the conditional species interaction strength given the presence/absence data using maximum likelihood estimation. We summarized abundances according to genus, to reduce computational and taxonomic complexity. The trophic role of bromeliad aquatic invertebrates is highly conserved at the genus level (Poff *et al.*, 2006), so we likely have not averaged over different trophic interactions with this approximation. The abundance data of each genus were transformed into presence/absence data. We performed Markov Network analysis separately for each site (Harris, 2016). The output of this analysis is the relative interaction strength for every pair of species in the site. We used a logistic density function for the prior distribution of interaction strengths; after running the model the final distribution of interaction strengths tended to be normal with a mean close to zero. We performed two validations for the Markov Network analysis (Appendix B).

EFFECT OF ENVIRONMENT ON SPECIES INTERACTIONS — Once we were able to confirm that Markov Network analysis correctly distinguished between predators and prey in terms of the predominant sign of interactions, we could then examine if the environment explained differences between sites in the relative strength of either positive or negative interactions. For this analysis, we separated negative from positive interactions to assess how interaction strength (within a particular sign) changes with the environmental variables, based on linear regression. We also used quantile regression to assess how interaction strengths (positive and negative) change with the environmental variables. Quantile regressions are useful when there is unequal variation in the data and therefore there might be more than one slope describing the relationship between response variable and predictor. Quantile regression is also more robust to outliers than mean regressions (Cade and Noon, 2003). The linear regression and quantile regression p-values were adjusted using

the Holm correction for multiple comparisons. To confirm the robustness of our results, we performed a permutation analysis by shuffling community composition (Appendix B: Permutation results, Figure B.6, B.6, B.6, Table B.5).

For the species interaction analyses we used the *rosalia* package (Harris, 2015), all multivariate analyses were performed using the *vegan* package (Oksanen *et al.*, 2018), mixed effect models were performed using *lme4* (Bates *et al.*, 2015) and *car* (Fox and Weisberg, 2011), and all analyses were done using the R programming language (Team, 2018).

## 2.4 RESULTS

### 2.4.1 *Environmental variation between sites*

The only two environmental variables that significantly differed between sites were maximum and actual water volume in bromeliads (Appendix B: Table B.1), and of these two, the most pronounced gradient was observed in the actual water volume in the bromeliads ( $F_{10,90} = -3.854$ ,  $P = 0.0003$ , Figure 2.1). We therefore focus on actual water volume as the major environmental gradient for the remainder of the analyses, hereafter just water volume.

### 2.4.2 *Community variation along an environmental gradient*

Community composition differed between sites, depending on the water volume in the bromeliads ( $F_{9,90} = 4.649$ ,  $P = 0.001$ , Figure 2.2a, b). However, beta diversity, measured as multivariate dispersion in composition around site centroids, differed only marginally among sites ( $F_{9,90} = 1.966$ ,  $P = 0.052$ ). These site differences in beta diversity were mainly driven by the sites that were the furthest apart geographically and differed the most in the water contained in bromeliads (Appendix B: Pairwise multivariate analysis of variance and pairwise Tukey tests for community dispersion, Table B.3). The difference

in community composition between sites was mostly due to species turnover (70%) and not due to nestedness (30%, Figure 2.2c). Therefore, contrary to our initial predictions, species were not progressively lost along the gradient of water in the sites and species richness per site was relatively constant across the gradient (Figure 2.2a).

#### 2.4.3 *Effect of environment on species interactions*

As the majority of genera were found in most sites, we could ask how each genus differed across the large scale environmental gradient in terms of interaction type (i.e. sign) and strength (i.e. magnitude) with other community members. For every pair of genus we obtained an interaction strength in every site (Figure 2.3). These interactions, however, did not correlate between sites either using the Pearson's correlation between interactions or using a Spearman's rank correlation. Therefore species interactions are changing from site to site. Using site means of water as the environmental gradient, we found that the relative strength of positive and negative interactions remain constant between sites for most genera, but for *Tipulidae*, *Wyeomyia* and *Elpidium* the relative strength of negative interactions diminished with site water volume (Appendix B: Regression results, Figure B.5, Table B.4). That is, sites whose bromeliads contained less water on average tended to have stronger negative interactions between community members and either *Tipulidae* (linear regression:  $\beta = 1.179 \times 10^{-3}$ , P value = 0.017), *Wyeomyia* ( $\beta = 8.254 \times 10^{-4}$ , P value = 0.062) or *Elpidium bromeliarum* ( $\beta = 1.257 \times 10^{-3}$ , P value = 0.016, all P values have been adjusted for multiple comparisons). Arguably, quantile regression might be better suited to detecting changes in the distribution of interactions, in which case only the *Tipulidae* interactions remain related to the mean water volume, even after the results were adjusted for multiple comparisons (first quantile regression:  $\beta = 1.151 \times 10^{-3}$ , P value = 0.05, Figure 2.4a).

FIGURE 2.2: Community composition across sites. a) Richness is relatively constant between sites. b) The community composition of every bromeliad is compressed into two axes. Each site is represented by a polygon containing all bromeliads within that site. The polygons with higher overlap suggest that those sites have more similar community composition. The area of the polygons represents the differences in community composition within a site (beta diversity). While most of the polygons have a relatively similar area (resulting only in a marginal difference in community dispersion), we find that there are three main clusters of overlap between the sites (Sites 1, 2, 3, 4 and 7 overlap and then sites 5 and 8, and 10 and 9 overlap). The stress of the NMDS (non-metric multidimensional scaling) plot was 0.26, consistent with a good, but not great, representation of the communities in two dimensions. c). A genera (columns) by sites (rows) matrix with sites ordered based on mean actual water, where sites with low water volumes are at the top of the graph and sites with high water volumes are at the bottom of the graph. If differences in beta diversity occur mainly through nestedness, the sites at the top would have emptier communities (more white) than the sites at the bottom. Beta diversity is mostly due to species turnover and not nested loss of species along the gradient.

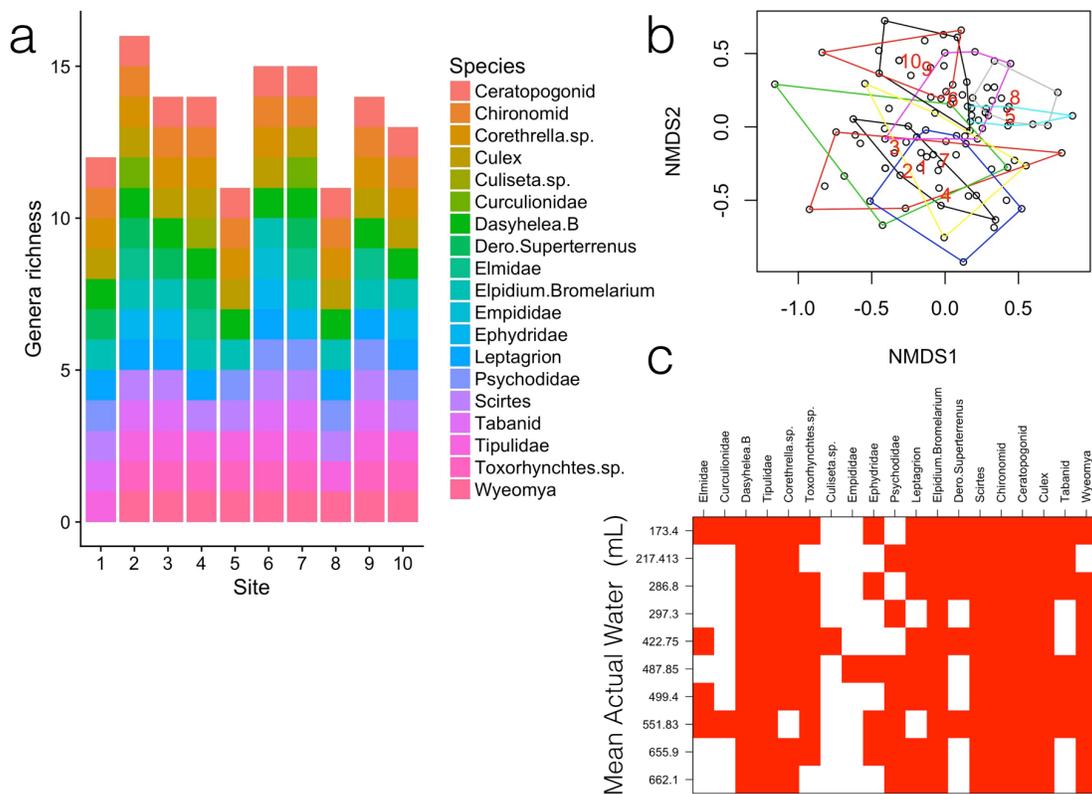


FIGURE 2.3: Relative strength of species interactions in every site. Species interactions are scaled to 1. Where rows or columns are empty, that particular species is not in that site. Blue indicates positive interactions and red indicate negative interactions. Positive interactions represent species than tend to co-occur, negative interactions represent species that do not tend to co-occur. The size of the points and intensity of the colour represent the interaction strength

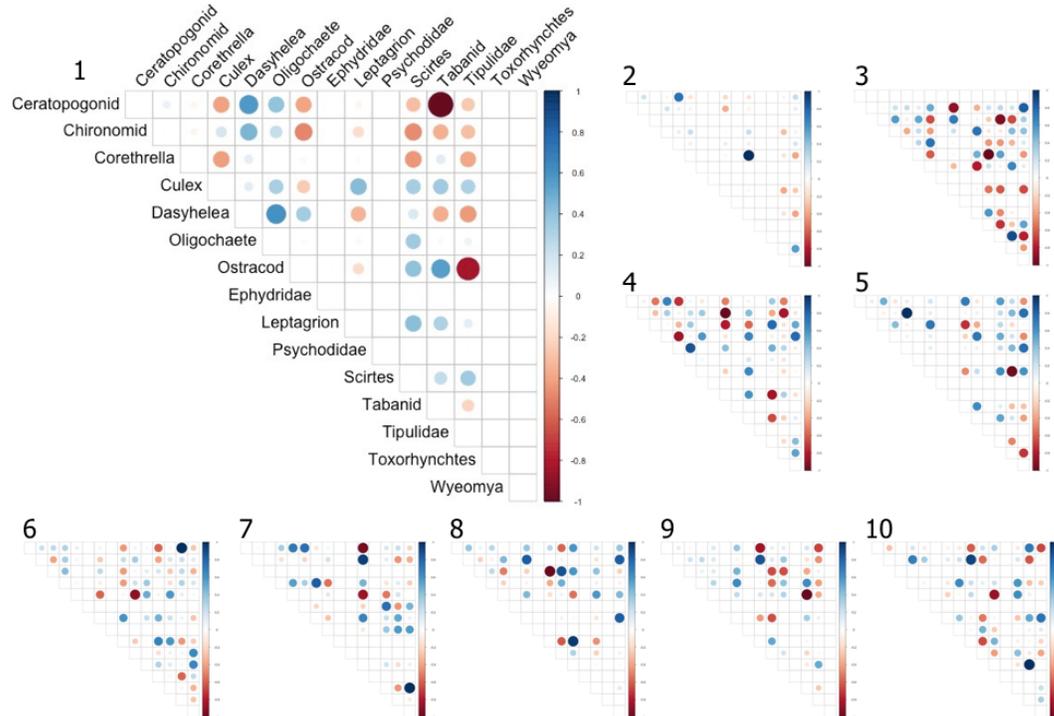
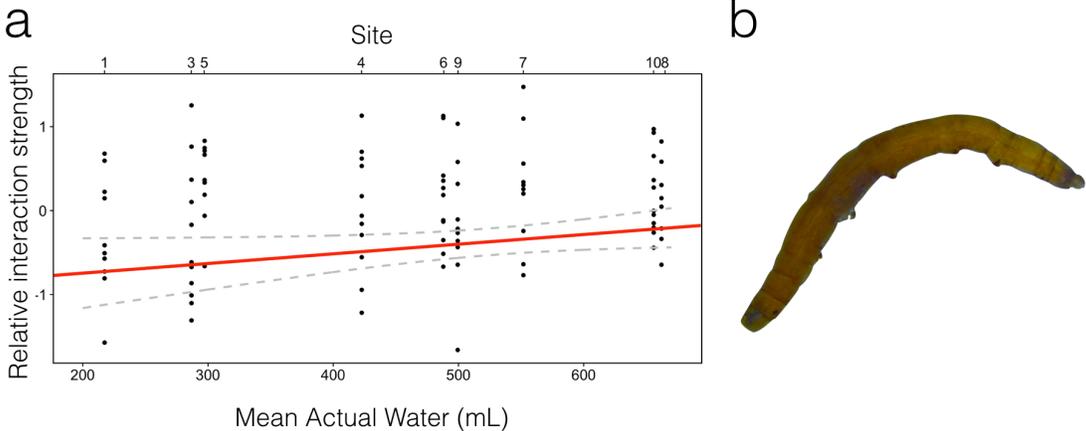


FIGURE 2.4: a) The tipulid has more negative interactions at low water volumes (Intercept = -0.98, se = 0.27, slope = 0.001, se = 0.0004). The red line represents the first quantile regression, because the major change in interaction strength occurred with the negative interactions. Tipulids are absent from site 2, so this site is absent from this regression. Dashed lines represent predicted confidence intervals. b). Image of a Tipulid larvae.



Even though difference between sites in interactions strengths could be due either to changes in the per capita interaction strengths between specific taxa, or changes in the pool of species available for interactions, we find that low volume sites do not progressively lose species (i.e. there is no nested loss of species, Figure 2.2c). We also checked whether if a species was missing due to low water volume (slope between presence and absence of a species vs water volume) was related to the interaction strength between the given species and the tipulid. We focused on the tipulid because the tipulid consistently showed a pattern of interaction strength related to water volume. (Appendix B: Relative interaction strength vs. presence, Figure B.9)

## 2.5 DISCUSSION

The main conclusions of this study were threefold. First, we found that, due to variation in precipitation, the water contained in the bromeliads was the main variable that consistently differed between sites. Second, we found that sites differed in both the average composition within bromeliads and the difference between bromeliads in composition (beta diversity). However, most of the effect of sites on beta diversity was due to the turnover of some species and not due to sequential loss of species being filtered by the environmental gradient (i.e. between-site turnover vs. between-site nestedness). Third, after validating that the Markov network approach could identify trophic levels, we found that interactions between tipulids and other species changed along a site gradient in actual water volume; sites with lower water volume had more intense negative interactions.

Our sites were located along a precipitation gradient with those in the south-west of our gradient receiving less rainfall than those towards the north-east (Appendix B: Figure B.1, B.2). This gradient was reflected in the amount

of actual water found in the bromeliads (Appendix B: Figure B.3), but not other aspects of the bromeliad environment (e.g. water chemistry). Low water volumes can affect bromeliad communities through a multitude of mechanisms: (i) low water volumes can select species whose traits allow them to be tolerant to drought (Dézerald *et al.*, 2015); (ii) low water volumes decrease habitat size thereby decreasing the size of the community (Dézerald *et al.*, 2014) (iii) low water volumes tend to lose higher trophic levels due to demographic stochasticity (Amundrud and Srivastava, 2015).

Overall, community composition differed along the geographic gradient in bromeliad water volume. However, this difference is not driven by the sequential loss of species along this gradient, but instead turnover in species identity. For example, the oligochaete *Dero superterrenus* was more common in the drier sites, and *Polypedilum* chironomids in the wetter sites. Such turnover may be related to the life history of organisms: oligochaetes reproduce within bromeliads, and so are resident year round, whereas larval chironomids require terrestrial adults to oviposit eggs, and adults may delay oviposition until most bromeliads in the site are water-filled. Previous studies have shown that the functional traits of bromeliad invertebrates determine their response to water levels within bromeliads: taxa able to survive low water conditions are characterized by small size and deposit or filter feeding whereas taxa able to rapidly colonize full bromeliads are characterized by drought-tolerant eggs and short generation times (Dézerald *et al.*, 2015). Since traits determine the response of species to altered environmental conditions, selection of species through their traits can alter not only the size of a community but also its structure (Tylianakis and Morris, 2017). More generally, if species differ in their optimal environment due to their life history and tolerance traits, we would expect that the arrangement of sites along an environmental gradient would cause a turnover in species composition due to species sorting mech-

anisms or when early successional species are gradually lost (Leibold and Chase, 2017; Brendonck *et al.*, 2015). Note that this species turnover occurs despite constant species richness between sites, regardless of water volume. However, within site, species richness increases with bromeliad water volume, as shown in previous studies of bromeliad invertebrates (Jabiol *et al.*, 2009).

The biggest advantage of the Markov network approach is that it considers indirect interactions such as intraguild predation, which commonly occurs in container habitat food webs (Edgerly *et al.*, 1999; Fincke, 1994). However, the low number of degrees of freedom in compositional data only allow us to estimate one interaction value per species pair (Harris, 2016). Even though Markov network analyses only calculate symmetrical interaction strengths, we argue that it is suitable for analysing trophic asymmetrical interactions when other information is available (Harris, 2016). For example, we can use information from the natural history of the system to allow us to interpret these interactions (Harris, 2016; Freilich *et al.*, 2018). We looked at the type of interactions that prey and predators participated in, knowing prior to the analysis which species were predators and which species were prey. We found that the top predator *Leptagrion andromache* dominates negative interactions, as expected from a generalist predator known to have high per capita impact on its prey, and therefore it is more likely to eat its prey at the local scale (Hammill *et al.*, 2015b). Furthermore, we found that predatory species were more likely to participate in net negative interactions and prey species were more likely to participate in net positive interactions, meaning that the Markov Network approach has different outcomes for different trophic levels. This, however does not mean that all predator-prey relationships are necessarily detected via negative interaction strengths. Another shortcoming of this method is that, if a species is very rare at the regional scale due to dispersal limitation or habitat filtering, it may appear to be negatively interacting with many species. This

occurs because a species that is rare in the region will not co-occur with the majority of the species. To reduce this problem, we only used species that were present in at least two bromeliads and the analysis was done at the site scale where most bromeliads experience the same climatic conditions.

Our Markov analyses indicated that, while overall the mean value of species interactions are similar in sign and strength along the water volume gradient, for three genera there is a consistent pattern of strengthening negative interactions in sites with lower water level. This pattern was particularly robust for tipulids. There are two possible mechanisms for this result. First, species that have only weak interactions with the three genera may become absent at sites with low water volumes, allowing stronger negative interactions to influence the mean. If this mechanism was operating, then the pattern should weaken with quantile regression. Indeed for two genera it does, but not for tipulids. The second mechanism is that many of the negative interactions intensify in strength as site water volumes diminish. This mechanism is consistent with the patterns seen in the tipulids. Tipulids may show stronger negative interactions at low water volumes because they switch from detritivores to generalist predators. This mechanism is supported by previous research, which found that tipulids in Costa Rican bromeliads supplement detritivory with opportunistic predation under drought (Amundrud and Srivastava, 2016). These researchers hypothesized that decreasing water volume in a bromeliad restricts the space for prey movement, and therefore the tipulids can become more effective predators. Other manipulative experiments confirm that bromeliad predators are more effective in smaller water volumes (Srivastava, 2006). Generally, from a biomechanical perspective, consumption rates of predators should depend on habitat dimensionality because it influences the cost of locomotion and the probability of prey escape (Pawar *et al.*, 2012).

Tipulids thus appear to be facultative predators, opportunistically switching from detritivory to predation. Facultative predators feed both on plant matter and animals at the same developmental stage; they represent a case of non-obligate omnivory (Albajes and Alomar, 2004). Facultative predation may constitute an adaptive strategy in habitats with high variability of food sources and allow species to withstand changing environments (Srivastava *et al.*, 2008; Albajes and Alomar, 2004). Bromeliad habitats are known to be highly variable, with water levels that fluctuate year around (Scarano, 2002). Therefore facultative predation may be a favourable strategy in these systems.

Our study adds to the evidence that trophic interactions may change with climate in bromeliad infauna. Over a much larger geographical gradient, Romero *et al.* (2016) found that cooler, less seasonal climatic conditions resulted in stronger top-down control from predators, based on biomass ratios of top predators to detritivores as a proxy for interaction strength. In their study, Romero *et al.* focused on odonate larvae as top predators. Here, we find that tipulid predation intensified in warmer, more seasonal sites. An intriguing topic for future study is whether seasonal droughts in Rio de Janeiro state, Brazil, shift predation from odonates to tipulids. Our study cannot pull apart other gradients that vary simultaneously with our sites such as latitude or time — since our sites were sampled at different times. To de-confound these effects, future studies can sample communities in the same site through time after a precipitation event, where bromeliads would start empty and eventually fill up. In addition, future studies can also perform experiments to establish the causal effect of water volume on interaction strength, not only for the tipulid, but also for the other predators in this system.

Our study reinforces the general point that ecological communities can change along an environmental gradient through three main mechanisms (i) through the turnover of species, (ii) through the change in species interactions,

or (iii) through the presence or absence of interactions (Tylianakis and Morris, 2017). Here we found the first two mechanisms are contributing to the changes in community composition along a gradient of water within bromeliads.

## 2.6 CONCLUSION

In this study we provided evidence for changes in community structure along an environmental gradient through two mechanisms. First we showed that community composition differed along a gradient of water volumes in macroinvertebrate communities due to the turnover of some species. Second we showed that species interactions also differed along this gradient. In our system, lower water levels likely changed the effectiveness of different predation strategies reflected in negative species interactions. Broader applications of the Markov approach to assess interaction strengths could assist studies that aim to explain differences in functional aspects of ecosystems that cannot be attributed to differences in species composition.

PREY BODY MASS AND DIVERSITY DETERMINE FOOD WEB  
PERSISTENCE.

---

### 3.1 CHAPTER SUMMARY

Predators and prey often differ in body mass. The ratio of predator to prey body mass influences the predator's functional response (how consumption varies with prey density), and therefore the strength and stability of the predator-prey interaction. The persistence of food chains is maximized when prey species are neither too big nor too small relative to their predator. Nonetheless, we do not know if (i) food web persistence requires that all predator-prey body mass ratios are intermediate; nor (ii) if this constraint depends on prey diversity. We experimentally quantified the functional response for a single predator consuming prey species of different body masses. We used the resultant allometric functional response to parameterize a food web model. We found that predator persistence was maximized when the minimum prey size in the community was intermediate. We also found that as prey diversity increased the range of values that this intermediate body mass prey could have become broader. This last result occurs because of Jensen's inequality: the average handling time for multiple prey of different sizes is higher than the handling time of the average sized prey. Our results demonstrate that prey diversity mediates how differences between predators and prey in body mass determine food web stability.

### 3.2 INTRODUCTION

Classical food web theory concludes that abstracted complex ecological networks are unstable (May, 1972). Yet most natural communities are comprised of complex, interconnected networks where species depend on other species for food and compete with other species for resources (Pascual *et al.*, 2006). To address this discrepancy, researchers have shown that natural food webs, rather than having random network structure as assumed in the classical approach, have more weak than strong interaction strengths (Gross *et al.*, 2009; McCann *et al.*, 1998; Neutel *et al.*, 2002), and that this non-random network structure actually promotes stability compared to May.

The strength of a trophic interaction that occurs via consumption (as opposed to non-consumptive effects), is determined by the flux of biomass transferred from prey to predator (McCann, 2011). A strong interaction between a single predator population and single prey population is unstable, leading to the extinction of the species pair (McCann, 2011; Berlow *et al.*, 2004; Holling, 1959). When we model food webs, reducing the flux of biomass transferred from the prey population to the predator population stabilizes the interaction. Similarly, when multiple species interact in a food web, one weak interaction among multiple species interactions can stabilize other interactions. For example, when two prey species share a predator population, the prey with the weaker interaction with the predator can reduce the efficiency of the predator in attacking the other prey, stabilizing an otherwise strong and destabilizing interaction (McCann *et al.*, 1998). Generally, food webs become stable when they contain more weak interactions. However, the increase in stability depends on the size of the food web: small food webs become stable, whereas larger food webs do not (Gross *et al.*, 2009).

The stability of food webs depends on the mechanisms that determine interaction strengths between species. One important determinant of interaction

strengths, at least amongst animals, is the relative body masses of predators and prey (Emmerson and Raffaelli, 2004; Kalinkat *et al.*, 2013). For example, an empirical test with crustacean predators showed that the predator-prey body mass ratio is correlated with the per capita interaction strength (Emmerson and Raffaelli, 2004). The per capita interaction strength between predators and prey depends on the predator's functional response (the relationship between predation rate and prey density). How the predator's functional response is modeled is in turn determined by parameters that depend on the body mass of both the predator and the prey (Kalinkat *et al.*, 2013). This functional response is composed of the attack rate (a measure of a predators hunting efficiency), the handling time (the amount of time needed to kill, ingest and digest an individual prey) and the efficiency rate (the fraction of prey biomass consumed and turned into predator biomass) (McCann, 2011; Berlow *et al.*, 2004; Holling, 1959; Jeschke *et al.*, 2002). This relationship between the ratio of predator and prey body masses and interaction strength has been integrated into population dynamic models, specifically by relating interaction strengths to the mean body masses of interacting species. Using these models, ecologists have found that species coexistence is restricted to a narrow range of prey body mass relative to that of their predator — specifically predator-prey body mass ratios between 10 and 100 for both invertebrate and vertebrate predator-prey pairs (Brose, 2010; Brose *et al.*, 2006; Otto *et al.*, 2007). Coexistence is restricted to these intermediate body mass ratios because at high body mass ratios, population dynamics are unstable and at low body mass ratios, the predators do not consume enough energy to persist in these models.

This body of literature has allowed ecologist to understand when pairwise predator and prey interactions are unstable, and how the presence of other species' populations may stabilize those interactions. Yet, we still do not know if predator and prey persistence requires that all the body mass ratios between

the predator and its prey species are intermediate in value. Jensen's inequality suggests that when relationships are not linear, such as the relationship between attack rate and handling time with body mass, then variation in body size within a prey population (or among prey species consumed by the predator) can alter the average interaction strength, thereby influencing the dynamics of the system (Bolnick *et al.*, 2011). Given Jensen's inequality, it is not clear if the narrow range of intermediate body mass ratios that stabilize food chains (one consumptive link between trophic levels) are also required to stabilize large food webs (multiple consumptive links between trophic levels). We took a three-pronged approach to answering these questions. First, we empirically determined how body mass ratios between a generalist predator and various prey species of different size affects the functional response parameters of a natural food web: the aquatic invertebrates that inhabit bromeliads. Although previous studies have examined this question by pooling data between systems, the same pattern may not apply within a system — and it is at this scale that allometric effects of body size on food web persistence is relevant (Kalinkat *et al.*, 2013). We then used this empirically-derived allometric functional response to simulate a food web with one predator and multiple prey, and confirmed that predator and prey persistence was maximized at intermediate prey body masses when the predator had only one prey. Finally, we tested (i) how differences between prey species in body mass affects the persistence of the entire food web or subcomponents and (ii) how the number of prey affects the distribution of prey body masses required for the entire food web or subcomponents.

### 3.3 METHODS

#### 3.3.1 *Study system*

Tank bromeliads accumulate water inside their leaf axils, providing habitat for communities of aquatic macroinvertebrates (Kitching, 2000). Inside each bromeliad, these aquatic macroinvertebrates interact to form a food web comprised of detritivores, filter feeders, intermediate predators and top predators. All the species we used in our experiment were identified to morphospecies level. We distinguished the morphospecies based multiple characteristics and not body mass.

#### 3.3.2 *Predator consumption rates*

We quantified the consumption rate for one damselfly larvae predator and many of its prey. The top predator *Leptagrion andromache* (Zygoptera: Odonata, dry mass = 3.31 mg, se = 2.45, n = 29) was fed several densities of each prey (Table c.1). The prey were chosen because they were the most abundant prey in bromeliads. All prey are aquatic insect larvae. We chose *Culex sp 1* (Culicidae: Diptera, density range = 1- 50, mean dry mass = 0.17 mg, se = 0.04, n = 25), *Culex sp 2* (Culicidae: Diptera, 1-20, 0.09 mg, 0.02, 14), *Forcipomyia* (Ceratopogonidae: Diptera, 1-60, 0.07 mg, 0.01, 5), *Dero superterrenus* (Naididae: Haplotaxida, 1-60, 0.12 mg, 0.01, 2), *Psychodidae* (Psychodidae: Diptera, 1-30, 0.22 mg, 0.11, 18), *Scirtes sp 1* (Scirtidae: Coleoptera, 1-30, 0.33 mg, 0.12, 13), *Scirtes sp 2* (Scirtidae: Coleoptera, 1-6, 0.43 mg, 0.27, 65), and *Trentepohlia* (Tipulidae: Diptera, 1-7, 0.29 mg, 0.19, 43). The predators were acclimatized to laboratory conditions for 72 h or more; the prey, were acclimatized to laboratory conditions for 6 h. Prior to each trial, the predator was starved for 24 h. Kairomones — signaling chemicals — emitted by predators have been shown to alter the behavior of the prey, altering functional responses

(Hammill *et al.*, 2015a), but because our focus was on body size we did not want interspecific differences in kairomone sensitivity between prey to affect our estimations of allometric functional responses. Therefore, for each trial, the predator was placed in a 250 mL dark cup with 200 mL of fresh mineral water and the prey were immediately introduced, preventing the water to be filled with kairomones before the prey was introduced. We recorded the number of prey individuals that were eaten during a two hour period. For the *Trentepohlia* and *Scirtes* trials, a leaf of approximately 25 cm<sup>2</sup> was placed in the experimental containers, as these species move along benthic surfaces rather than swim in the water column. We used five to eight densities per species increasing the total density, until feeding rates reached an asymptote (Table c.1). All species-by-density combinations were replicated three times for a total of 162 trials.

### 3.3.3 *Body Mass*

The average measured dry body mass of each prey species was obtained from previously established allometric equations (Median number of individuals = 21 and  $R^2 = 0.75-0.94$ , depending on the species) (Marino, 2015). We measured all prey used to the nearest mm and calculated the average dry body mass of that prey species in mg.

### 3.3.4 *Allometric functional response*

We fit a general functional response to the consumption data of each prey species:

$$N_e = \frac{aN^{q+1}}{1 + ahN^{q+1}} \quad (3.1)$$

where  $N_e$  is the number of prey consumed,  $N$  is the starting prey density,  $a$  is the attack rate of the predator,  $h$  is the handling time of the predator, that is, the time taken to search and consume the prey (Real, 1977). The parameter  $q$  determines the shape of the functional response, allowing a range of responses including type II ( $q = 0$ ) and type III ( $q = 1$ ). We initially fitted a full model with all three parameters ( $a$ ,  $h$  and  $q$ ). However  $q$  was not significantly different from zero, and we thus restricted  $q = 0$  throughout our analyses (type II functional response). Because the prey were not replaced during the experiment, we used numerical integration to calculate the true proportion of prey consumed (Hammill *et al.*, 2015a, 2010; Bolker, 2012).

We calculated a single allometric functional response that allows the attack rate and handling time to vary with the body mass of the prey and the predator. Since we used the same species of predator in all experiments, this allometric functional response need only consider variation in attack rate and handling time with respect to the mean prey body mass. We fit a simplified power relationship with fixed allometric-scaling exponents (eq. 3.2 and 3.3) following Yodzis and Innes (1992):

$$a = a_0 m_i^{-1} \quad (3.2)$$

$$h = h_0 m_i \quad (3.3)$$

as well as an allometric relationship (eq.3.4 and 3.5) based on Kalinkat *et al.* (2013):

$$a = a_0 m_i^{\alpha \frac{3.3}{m_i}} e^{\delta \frac{3.3}{m_i}} \quad (3.4)$$

$$h = h_0 m_i^{\phi} m_c^{\gamma} \quad (3.5)$$

The allometric relationship between attack rate and body mass based on Kalinkat *et al.* (2013) combines a power function with a Ricker function that allows the attack rate to have a humped relationship with body mass.

Because the body mass of our predator was constant, we used a simplified version of equation 3.5 (eq.3.6):

$$h = h_0 m_i^\varphi \quad (3.6)$$

where  $m_i$  is the body mass of the prey and  $m_c$  is the body mass of the predator. For equation 3.4  $m_c = 3.3$ .

### 3.3.5 Food web model

We built a stochastic food web model for a predator that consumes any number of prey with a type II functional response. If the model contained two or more prey, then the prey competed for resources. We chose a stochastic model to allow for competitive exclusion, as under a deterministic model, prey would continue to co-exist even if their body sizes were identical. The stochasticity was incorporated in the birth terms of the prey and predator equation. For both predators and prey this means that variance in lambda was obtained from a Poisson distribution. For the prey, lambda was calculated by multiplying the the intrinsic growth rate and the abundance of the prey in the previous time step. For the predator lambda was calculated from the consumption of prey, the predator's conversion efficiency and the abundance of predators in the previous time step.

The abundance of prey is given by:

$$R_i(t+1) = R_i(t) + \theta(m, t) \left( 1 - \frac{R_i(t) + \sum_{j=1}^R \alpha_{ij} R_j(t)}{K(m_i)} \right) - \frac{a(m_i) R_i(t) P(t)}{1 + \sum_{i=1}^R a(m_i) h(m_i) R_i(t) P(t)} \quad (3.7)$$

and

$$\theta(m, t) \sim \text{Poisson}(\lambda = r(m_i) R_i(t)) \quad (3.8)$$

In eqns. 3.7 and 3.8,  $R_i(t)$  is the abundance of prey  $i$  at time  $t$ . The growth rate of prey  $i$  is given by a stochastic Poisson process, where  $\lambda$  is the product of  $r(m_i)$ , the intrinsic growth rate of prey (e.q. 3.9), and  $R_i(t)$ , the abundance of that prey at time  $t$ . The prey experiences logistic growth in where its carrying capacity  $K(m_i)$  is determined by its body mass  $m_i$  (e.q. 3.10).

$$r = r_{max} m_i^o \quad (3.9)$$

$$K = k_k m_i^k \quad (3.10)$$

We explicitly added a term,  $\alpha_{ij}$ , signifying the intensity of competition between prey  $i$  and  $j$ , to our equation, as otherwise non-linear functional responses can transform apparent competition into apparent mutualism (Abrams *et al.*, 1998). The attack rate,  $a(m_i)$ , as well as the handling time,  $h(m_i)$ , are determined by the body mass  $m_i$  of prey  $i$  as given by eqns. 3.4 and 3.6.

The intensity of competition  $\alpha_{ij}$  was dependent on the amount of niche overlap, which was given by the body mass of the prey:

$$\alpha_{ij} = \alpha(m_i, m_j) = (1 + (m_i - m_j)^2 / 2\sigma_\alpha^2)^{-1} \quad (3.11)$$

where  $m_i$  is the body mass of species  $i$  and  $m_j$  is the body mass of species  $j$ .  $\sigma_\alpha$  represents the niche width. The competition coefficient is based on a decaying function between the body masses of species  $i$  and  $j$  (Johansson, 2008). This equation assumes that competition is symmetrical. Evidence suggests that larger organisms outcompete smaller ones, and that competition is often asymmetrical (Persson, 1985; Lawton and Hassell, 1981). We evaluate this scenario in Appendix c using equation 3.12:

$$\alpha_{ij} = \left( \frac{1}{1 + (m_i - m_j + \beta)/(2\sigma_\alpha^2)} \right) \left( 1 + \frac{\beta^2}{2\sigma_\alpha^2} \right) \quad (3.12)$$

where  $\beta$  determines the asymmetry of the competition coefficients (Van Den Elzen *et al.*, 2017). When  $\beta = 0$  competition is symmetric (Johansson, 2008), on the other hand when  $\beta > 0$  the competition is asymmetric, where larger individuals outcompete smaller individuals.

The predator will saturate not only from consuming prey  $i$  but also from other prey that it also consumes (McCann *et al.*, 2005). The abundance of the predator is therefore given by:

$$P(t+1) = P(t) + \vartheta(m, t) - CP(t) \quad (3.13)$$

with

$$\vartheta(m, t) \sim \text{Poisson} \left( \lambda = \frac{P(t)B \sum_{i=1}^R a(m_i)R_i(t)}{1 + \sum_{i=1}^R a(m_i)h(m_i)R_i(t)P(t)} \right) \quad (3.14)$$

where  $P(t)$  is the abundance of the predator at time  $t$ . The growth of the predator is given by a stochastic Poisson process, where  $\lambda$  is equal to the amount of prey (all species) consumed at time  $t$ . The caloric value of a captured prey individual is assumed to be constant across prey species, and is signified by  $B$ . The per capita mortality rate of the predator is  $C$ .

### 3.3.6 Simulations based on all possible combinations

We ran the simulation for prey body masses starting at 0.05 mg and increasing at 0.01mg intervals until reaching 1 mg.

For the one and two prey scenario, we ran every combination of prey body masses 10 times. The starting abundances were 10 individuals for each prey and 7 individuals for the predator. All simulations were run for 500 generations. As previously mentioned, the attack rate, handling time, intrinsic growth rate of the prey and the carrying capacity were all determined by the body mass of the prey (Table 3.1, Figure c.1).

We used a random Poisson process to simulate stochastic dynamics (e.g. 3.8 and 3.13), and one property of the Poisson distribution is that the variance is equal to the mean. Therefore, any increase in  $\lambda$  increases the possible range of values that the Poisson process can take, resulting in large spikes in population size as  $\lambda$  increases. To stabilize the model and reduce the amplitude of the spikes in population size, we decreased the maximum value of the attack rate and handling time but kept the shape of the relationship constant for all parameter combinations (Figure c.1).

### 3.3.7 Simulations based on sampled parameter space

Scenarios with three or more prey are computationally difficult, as the potential combinations of prey body masses increases exponentially with the number of prey. Fortunately, a recently proposed Monte Carlo strategy for sampling parameter space helps to make this tractable (Leigh and Bryant, 2015). In this sampling process, we begin by defining an binary outcome of interest as  $\mathcal{R}$  specific (for example, whether the predator persisted). For each outcome  $\mathcal{R}$  we followed the following five recursive steps:

S1: If at parameter  $\theta$ , then propose a move to  $\theta'$ . A step size drawn from a

random normal distribution centered at zero, truncated between a minimum and maximum value. The first step in the MCMC is to propose a new set of parameters  $\theta'$  given an initial value of  $\theta$ . For example, if the initial body mass of three species was a vector (0.1, 0.3, 0.4), we might draw a random value from a normal distribution (say 0.1) and add it to each body mass. This new vector (0.2, 0.4, 0.5) would be  $\theta'$ .

S2: Run the simulation using  $\theta'$  and assess  $\mathcal{R}$ . The second step of the MCMC is to run the simulation using the new parameters and to determine whether  $\mathcal{R}$  occurred (e.g. whether the predator persisted or not).

S3: If

$\mathcal{R}$

occurs, then accept  $\theta'$ , otherwise stay at  $\theta$ . In this third step, if  $\mathcal{R}$  occurred, we take  $\theta'$  as the new starting value, otherwise we go back to the initial value in step 1. This ensures that as the number of iterations increases, the MCMC converges towards the body masses ( $\theta$ ) where the predator is more likely to persist.

S4: Record  $\theta$ ,  $\theta'$  and  $\mathcal{R}$  as well as the final abundance of all species in the community at the end of the simulation

S5: Return to S1

This sampling process allowed us to sample through the space of  $\theta$  without having to run the simulation for every combination of prey body masses. In all cases, the only parameters that were sampled were the body masses of the prey. Therefore they could be sampled from the same distribution.

The outcomes  $\mathcal{R}$  that we required were that (i) all species survived, (ii) only prey survived and (iii) at least one prey and the predator survived.

We ran the sampling process for 10,000 steps for one and two prey and for 20,000 steps for three or more prey. For the first 10,000 steps, we used a step

size of 0.3, and then increased this to 0.5 for the next 10,000 steps. Parameter space was well sampled with the Monte Carlo sampling strategy (Figure c.2).

TABLE 3.1: The parameter values used for the simulations of the food web and their relationship with the body mass of the prey. Assigning  $k = -1$  implies that  $K * m_i$  is constant.

Parameter	Description	Body mass scaling	Estimates
a	Attack efficiency of the predator	$a = a_0 m_i^{\frac{3.3}{m_i}} e^{\frac{3.3}{m_i}}$	$a_0 = 0.0005,$ $\alpha = 1.30, \delta = 0.071$
h	Handling time of the predator	$h = h_0 m_i^\varphi$	$h_0 = 5, \varphi = 2.41$
K	Carrying capacity of the prey	$K = k_k m_i^k$	$k_0 = 6, k = -1$
r	Maximum population growth rate of the prey	$r = r_{max} m_i^0$	$r_{max} = 0.05,$ $\rho = -0.20$
B	Caloric value of a capture individual of prey species i (this is assumed to be constant across species)	Constant	0.2
C	Predator death rate	Constant	0.4
$\sigma$	Niche width	Constant	0.025
$\beta$	Degree of asymmetry in competition function	Constant	0 (0.5 in Appendix c)

To examine the impact of Jensen's inequality on food web dynamics, we compared parameters of the functional response calculated as the mean for all species versus based on the mean mass of the species. Specifically, for every run where all the prey species and the predator persisted, we calculated the attack rate and the handling time for each prey species and averaged these values over species to obtain  $\overline{a(m)}$  and  $\overline{h(m)}$ . Then we averaged the body mass of all prey species before calculating the attack rate and the handling time to

obtain  $a(\bar{m})$  and  $h(\bar{m})$ . Due to the non-linearity and the upward curvature of the attack rate and the handling time with body mass (Figure C.1), we would expect that  $\overline{a(m)} = a(\bar{m})$  and  $\overline{h(m)} = h(\bar{m})$  only when prey diversity is one. As diversity increases we expect that  $\overline{a(m)} > a(\bar{m})$  and  $\overline{h(m)} > h(\bar{m})$  (Figure C.5).

All analyses were performed using R (Team, 2018) and the *bbmle* package (Bolker and Team, 2017).

## 3.4 RESULTS

### 3.4.1 Allometric feeding rates

We compared two potential models that have often been used in the literature, describing the relationship between feeding rate and prey body mass: a power function model and an allometric model (Kalinkat *et al.*, 2013). In the power function model, the attack rate decreases proportionally to the negative power of body mass and the handling time increases linearly with body mass. In the allometric model, the attack rate decreases exponentially with body mass and the handling time increases in a power function (Figure C.1). We fit each model to all prey simultaneously, using the mean body mass of each species as a fixed effects (Figure 3.1, Table 3.2). Of these two models, the best fit model was the allometric model (Allometric model BIC = 621.74, df = 5, Power model BIC = 732.38, df = 3). We therefore use the allometric model for our simulations.

### 3.4.2 Simulation:

We found that when only one prey species was available for the predator and when this prey species was very small, this one predator-one prey food chain did not persist. When prey body masses were small, the attack rate

FIGURE 3.1: Fitted type II functional response for all prey separated between small, medium and large prey.

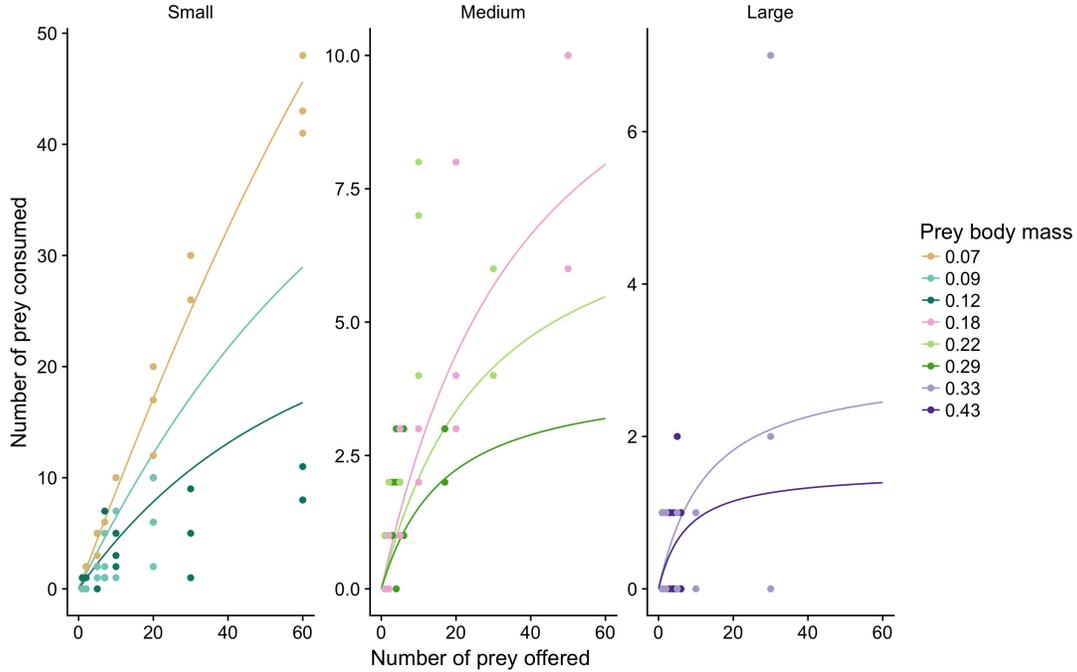


TABLE 3.2: Estimated parameter values for the relationship between attack rate, handling time and prey body mass.

	Parameter	Estimate	Std error	P value
$a = a_0 m_i^{\frac{3.3}{m_i}} e^{\frac{3.3}{m_i}}$	$a_0$	0.012	0.008	0.121
	$\alpha$	1.30	0.6113	0.033
	$\delta$	0.071	0.022	0.001
$h = h_0 m_i^\varphi$	$h_0$	19.837	0.002	$2 \times 10^{-16}$
	$\varphi$	2.418	0.083	$2 \times 10^{-16}$

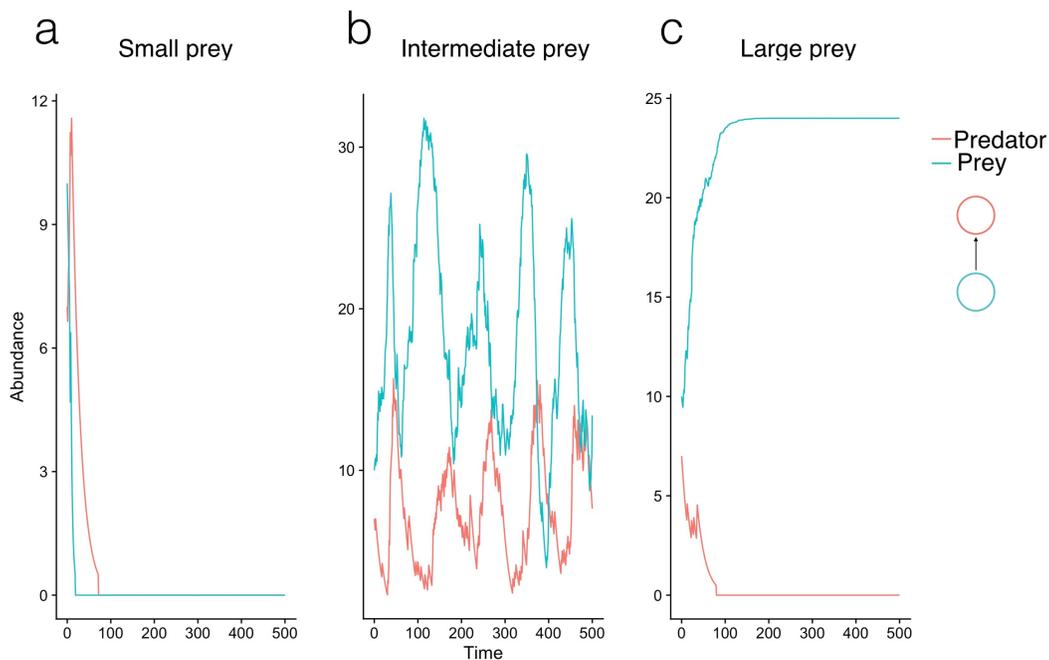
of predators on prey is so high that the predator consumed the prey quickly, driving the prey species extinct, which consequently led to the collapse of the predator population (Figure 3.2a). On the other hand, when the one prey was large, the handling time was so long that the predator could not meet

its energetic demands (Figure 3.2c). Only at intermediate body masses of the prey were both prey and predator able to persist through time.

Similarly, when two prey were present, the predator required at least one prey to have an intermediate body mass in order for the predator to persist (Figure 3.3a). Surprisingly, the body mass of the second prey could be almost any mass except for very small, including a prey body mass that was so large that, on its own, the prey could not sustain the predator. When the body mass of the second prey was small, only the predator and the larger of the two small prey persisted, as the smallest prey was driven to extinction due to the high attack rate of the predator (Figure 3.3c). By contrast, when both prey species were very small, the whole community went extinct, and when both prey species were very large, the prey persisted but the predator went extinct (Figure 3.3b). We only observed competitive exclusion in the presence of the predator. When the predator went extinct, competition between prey only generated differences in their abundances (Figure c.3, and c.4).

In food webs with more than two prey species, predator persistence shows a unimodal relationship with prey diversity, increasing with modest numbers of prey species but decreasing with larger numbers of prey species (Figure 3.4a). Within this overall effect of prey diversity, different effects of prey body mass can be discerned. At low prey diversity, predator persistence requires the prey communities have neither small or large species (Figure 3.4b, c). That is, the prey species are the 'optimal' intermediate mass. At high prey diversity, predator persistence increases when large prey are added to the community (Figure 3.4c). Regardless of prey diversity, predator persistence decreased as the proportion of the community with small prey increased (Figure 3.4b). By contrast, the proportion of prey species that strongly competed had only modest effects on predator persistence (Figure c.6). Although these results

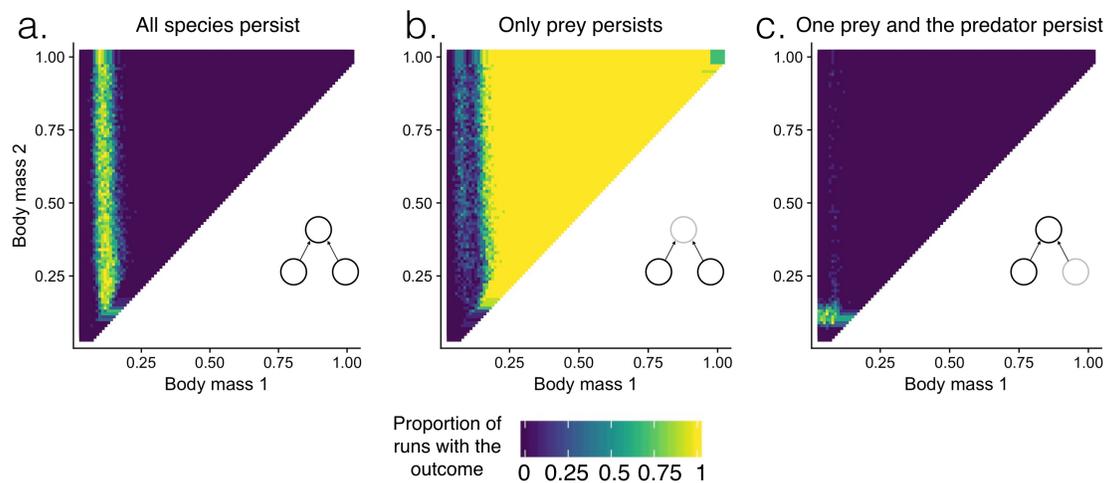
FIGURE 3.2: Time series for the abundance of the predator and one prey at three different prey body masses. Predator and prey persistence is maximized at intermediate prey body masses. a) When the prey is small (0.07), the attack rate of the predator is very high and the predator consumes all the prey, driving the prey and itself extinct. b) When the prey has intermediate body mass (0.12) both predator and prey can persist. c) When the prey has high body mass (0.25) the handling time is too long and the predator is limited by the amount of energy it can obtain.



pertain to the initial prey diversity, the patterns are unchanged even when we use the final prey diversity as the determinant (Figure c.7).

For a given diversity of prey, we found that predator persistence requires a minimum body mass for at least one prey species. As prey diversity increases, this minimum body mass first takes a wider range of values, but at high prey diversity again is constrained to a narrow range of values (Figure 3.5a). The broadening of the minimum body mass that allows all species to persist is due to the increasing importance of Jensen's inequality as we increase

FIGURE 3.3: Predator persistence is constrained only by the body mass of one prey, where one prey must have an intermediate body mass but the second prey can be of any mass. On the x axis is the body mass of prey 1, and on the y axis is the body mass of prey 2. Panels a-c show the proportion of the runs with a certain outcome, where yellow represents all the runs and purple none of the runs: a) both prey species and the predator persists; b) the predator goes extinct and only the two prey persist; c) one prey and the predator persist and one prey goes extinct (in every case the smaller prey). These scenarios are detailed in the food web modules: the grey circles represent species that go extinct and the black circles represent species that persist.



prey diversity. When we plotted  $\overline{a(m)}$  against  $a(\overline{m})$  and  $\overline{h(m)}$  against  $h(\overline{m})$  (Figure 3.5b and c), we find that for both parameters the effective attack rate or handling time initially increases with diversity and then decreases. All scenarios, when prey diversity is greater than one, fall above the 1:1 line. Therefore initially, as prey diversity increases, the attack rate and the handling time increase allowing then predator to persist at larger body masses (Figure 3.5a). As diversity continues to increase, the effective attack rate decreases more than the effective handling time, reducing the body masses of the prey where the predator can persist.

### 3.5 DISCUSSION

In this study, we tested: (i) how differences in body mass between prey species affect the stability of the entire food web or its subcomponents; and (ii) how the diversity of prey determines the distribution of prey body masses needed for food web persistence. To answer these questions, we first experimentally determined how body mass affected the parameters of functional responses within a single food web. We then used these parameters to simulate a food web consisting of a generalist predator and multiple prey; we kept the body mass of the predator constant and we varied the diversity and body mass distribution of the prey. We found that when the predator had one prey, predator persistence required the prey to be of intermediate body mass. As prey diversity increased, the range of prey body masses that allowed for predator persistence became first broader and then narrower. We also found that prey diversity had a unimodal relationship with predator persistence. At low prey diversity, increasing diversity *increased* predator persistence, whereas, at high prey diversity, increasing diversity *decreased* predator persistence. Furthermore, the effects of prey body mass on predator persistence reversed between low and high prey diversity conditions. We now consider each of these main results.

We found that the effect of prey body mass on the functional response of damselfly larvae in bromeliads largely conformed to the general form of the allometric functional response found by Kalinkat *et al.* (2013). This convergence in results suggests a generality to these allometric functional responses, despite differences in scale (our analysis was within a food web, Kalinkat *et al.*'s was between food webs) and hunting mode (our predator was sit-and-wait, whereas multiple hunting modes were considered by Kalinkat *et al.*). While this points to a qualitative generality in the effect of body mass on functional responses, we caution against any quantitative extension of the

parameters to other systems. Our estimates of the predator's functional response may underestimate the attack rate for other types of predators or other types of prey. The effectiveness of a predator depends not only on prey body mass, but also predator hunting mode, prey defensive behaviour or traits, the relative preference of the predator for prey, prey refuges and the identity of other species present in the community (Jonsson *et al.*, 2018). Predator hunting mode (such as sit-and-wait, sit-and-pursue or active predators) can affect the activity of the prey and therefore the prey mortality. Sit-and-wait predators have been found to cover small distances and kill fewer prey than either sit-and-pursue and active predators (Miller *et al.*, 2014). Therefore, our estimates of the predator's functional response may underestimate the attack rate for other types of hunting strategies. In natural systems, even when a predator is generalist, it may only rely on a handful of species due to differences in prey vulnerability and habitat complexity (Weber *et al.*, 2010). Furthermore, any variation in the biomass of prey species and conspecific density can result in preferences of predators for particular prey sizes (Costa-Pereira *et al.*, 2018). Finally, as we only allowed the predator to eat one type of prey at a time, our estimates of the functional response may not be representative of the predator's effectiveness in real bromeliads, which are complex habitats (Srivastava, 2006).

Our first main result from the model was that predator persistence increased when its sole prey had an intermediate body mass. As in other studies (Brose *et al.*, 2006; Otto *et al.*, 2007), this occurred because of the opposing effects of body mass on two parameters of the predator's functional response. If the prey had a very small body mass, then the predator had a high attack rate and consumed all prey individuals, driving the prey and itself extinct. If the prey had a large body mass, then the predator had a long handling time and it could not offset its mortality via reproduction and went extinct.

Even when the food web contained multiple prey, only the body mass of the smallest prey largely constrained predator persistence; the body mass of the other prey species could be either large, intermediate or small. This result is consistent with the case of asymmetrical competition between the prey, as long as the prey can coexist (Appendix c). These results extend previous studies of the effects of prey body mass on food web stability, which either considered only a single prey species (Otto *et al.*, 2007) or multiple prey species but with a narrow range of body masses (Brose *et al.*, 2006). These studies found that intermediate predator-prey body mass ratios maximize food web persistence, similar to our single prey results. Our novel result is that with a wide range of prey body masses, persistence depends on the smallest prey being of intermediate body size.

Even though the body mass of only one prey species constrained predator persistence, this optimal body mass may still change with increasing diversity. Indeed, we found that increasing prey diversity broadens and then narrows the range of body masses that promote predator persistence. Prey with body masses too high to allow predator persistence on their own, could, as part of a diverse prey community, now allow the predator to persist. The mechanism here relates to Jensen's inequality: since the attack rate and the handling time are non-linear with respect to prey body mass, variation around the mean body mass (i.e. when diversity increases) increases the effective attack rate and handling time. Therefore the predator experiences a much higher attack rate and handling time than expected from the mean body mass of all prey, allowing the predator to persist when otherwise it would not (Bolnick *et al.*, 2011). Intermediate prey diversity effectively relaxes the bottom-up energetic constraint imposed by having only one prey, allowing the attack rate and the handling time to be higher and resulting in an overall high total intake for the predator. Gibert and DeLong (2017) found a similar pattern, where the

predator's total intake is maximized at low diversity when prey is optimally matched in phenotype (e.g. body mass) to the predator and at higher diversity when the prey is poorly matched in phenotype to the predator.

Our results also are consistent with previous observations that weak interactions can stabilize a system with strong interactions (McCann *et al.*, 1998). In our study, the differential in body mass between predators and prey determined interaction strength: prey species with a small body mass had strong interactions with the predator, and prey species with a large body mass had weak interactions with the predator. When a predator had a strong interaction with one prey, leading to the extinction of both the predator and the prey, adding a larger prey species would often allow the predator to persist. Even though adding a large prey would stabilize many strong interactions between a predator a small prey, as shown by McCann *et al.* (1998), we found that some strong interactions between a predator and a very small prey could not be stabilized even in diverse communities (e.g. Figure 3.3).

So far, we have shown that predator persistence is constrained by the presence of at least one prey of intermediate body mass. This result helps explain why, at low prey diversity, adding more prey species can increase persistence of the predator: increasing diversity increased the probability that at least one prey has the optimal body mass. This sampling mechanism is consistent with models that show only intermediate body mass ratios between predators and prey can result in a positive relationship between food web stability and prey diversity (Brose *et al.*, 2006). However, our analysis not only allows us to confirm this result, but also to test the interaction between body mass ratios and diversity. We discovered that prey diversity had a unimodal relationship with predator persistence, which is mediated by the body mass distribution of the prey. The increasing portion of this relationship, at low prey diversity, reflects sampling of an optimal body mass prey. At high prey diversity, adding

prey species generally decreased predator persistence but this depends on the body mass distribution of added prey. Specifically, adding large prey species increased predator persistence whereas adding intermediate and small prey species decreased predator persistence. As before, the mechanism here is a weakening in interaction strength as prey body size increases relative to that of their predator, allowing for stabilization of the whole food web (McCann *et al.*, 1998). Consequently, we found that the food webs with the highest prey diversity at the end of the simulation were comprised of many large prey species. This result, however, is different when prey compete asymmetrically. In this scenario, higher diversity only reduces predator persistence. Adding small prey species has the same effect as in the symmetrical competition scenario, but adding large prey species now decreases predator persistence since large species drive intermediate species extinct due to competition (Appendix c). The implication of the symmetrical competition results is that we would expect the largest food webs in nature to have large-sized prey relative to their predators. Some food webs have shown a pattern where slow energy channels, which have weaker interactions, are more diverse than fast energy channels (Rooney and McCann, 2012). While some invoked explanations of this pattern suggest that is due to higher habitat complexity, another explanation can be that slow channels with weaker interactions are more stable. Our results would support the latter explanation.

Taken together, these results suggest that: (i) Only a few prey species, not all, constrained predator persistence; (ii) increased prey diversity can lead to increased predator stability if the additional prey contain at least one species of optimal body mass (low prey diversity) or enough large prey to weaken interactions with the predator (high diversity). That is, not all diversity is equal in terms of food web stability: the body mass of those gained or lost species has major consequences for food webs stability. This conclusion but-

tresses earlier findings that food web stability is more influenced by variation in the body mass ratios of predators and prey than by prey diversity alone (Brose *et al.*, 2006); (iii) The food webs we see in nature are only the observed surviving configurations. We expect that larger food webs will be skewed towards having larger prey species per predator when competition between prey is symmetrical.

FIGURE 3.4: Prey diversity has a unimodal relationship with predator persistence. a) Increasing diversity initially increases the proportion of runs where the predator and at least one prey survived. At higher diversity the trend reverses. b) As the proportion of small prey in the community increases, the proportion of runs where the predator persists decreases. The unimodal relationship is maintained regardless of the proportion of small prey in the community. c) At low diversity, a high proportion of large species in the community reduces the proportion of runs where the predator persists. At high diversity, the trend reverses and predator persistence is maximized when communities have a high proportion of large prey species.

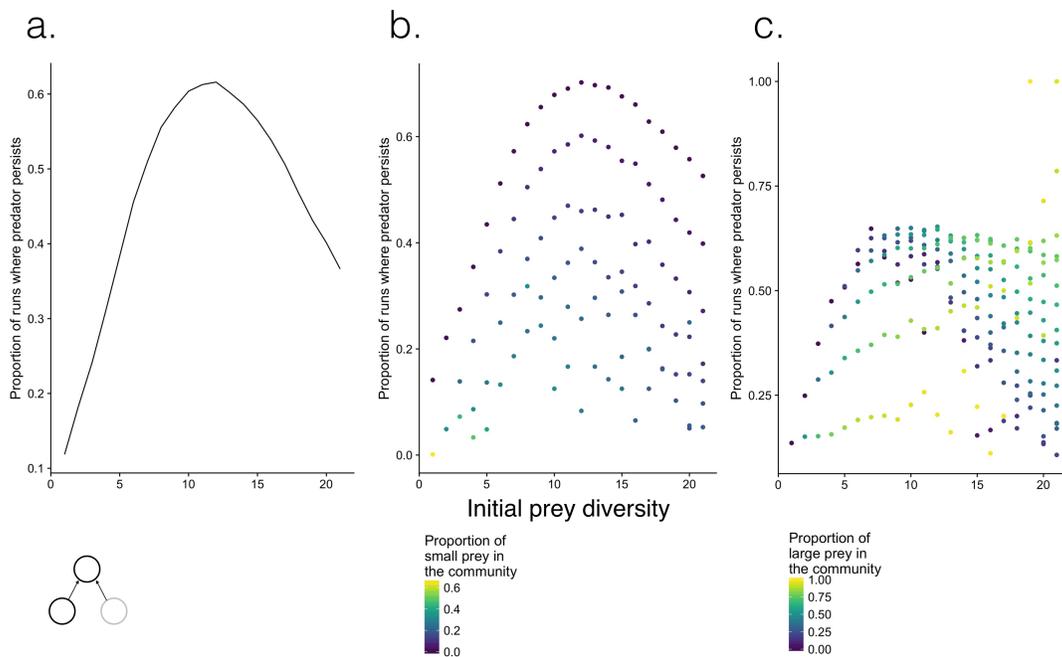
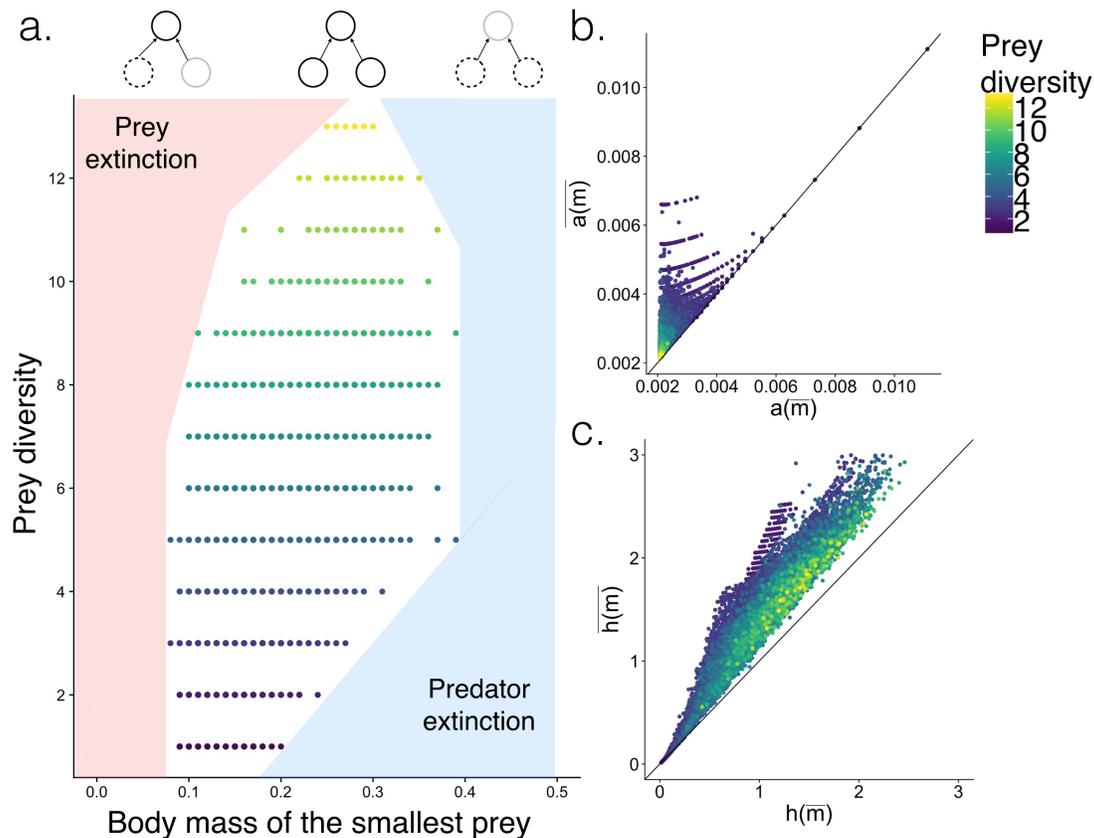


FIGURE 3.5: Food web persistence depends on the number of prey and the body mass of the smallest prey species. Circles of different colours represent food webs that successfully persisted. a) Some combinations of prey number and smallest body mass did not lead to food web persistence, due either to extinction of at least one prey (shaded red area) or predator extinction (shaded blue area). The food web modules illustrate these scenarios of food web persistence and extinction; the dashed circles represent species that may, but do not necessarily go extinct, the grey circles represent species that go extinct and the black circles represent species that persist. b) Jensen's inequality is demonstrated by the mean attack rate,  $\overline{a(m)}$ , being higher than the attack rate on a species with mean mass  $a(\overline{m})$ , but this effect is strong at low diversity and weak at high diversity given the deviation from the black 1:1 line. As diversity increases, both  $\overline{a(m)}$  and  $a(\overline{m})$  decrease. c) Jensen's inequality has a greater impact on handling time, as the mean handling time,  $\overline{h(m)}$ , is higher than the handling time on a species with mean mass,  $h(\overline{m})$ . The deviation from the 1:1 line is particularly pronounced at low prey diversity, and approaching the 1:1 line at high diversity.



EMPIRICALLY DETERMINED DIFFERENCES BETWEEN A  
PREDATOR AND ITS PREY ALLOW INCREASED PERSISTENCE IN  
A METACOMMUNITY

---

#### 4.1 CHAPTER SUMMARY

Trophic metacommunity theory has pushed the boundaries of our understanding of spatial dynamics in food webs, yet empirical tests of theory have lagged behind. A key reason for this mismatch is that most empirical studies of metacommunities do not parameterize models. To bridge this gap, we estimated the dispersal rate and kernel of a predator-prey pair using genotyping-by-sequencing (GBS) and a Bayesian approach. We found that the prey dispersed up to 25 km while the predator dispersed up to 200 m. We also measured the interaction strength between these two species experimentally (from a previous study). Using our estimates of dispersal and feeding rates, we parameterized a metacommunity model and found that differences in dispersal rates were sufficient to generate differences in occupancy of our modelled landscape, without requiring variation in the abiotic niche. More surprisingly, the observed asymmetry in dispersal was more likely to generate differences in occupancy than the reverse asymmetry (higher predator than prey dispersal) generated.

## 4.2 INTRODUCTION

Food webs are inherently spatial, and species within a food web often live in different parts of a landscape. The movement of organisms via dispersal is critical to determining food web structure at small and large scales (Amarasekare, 2008). A food web metacommunity then, is a set of connected food webs which inhabit patches linked by dispersal (Leibold *et al.*, 2004). Trophic metacommunity theory has provided valuable insight into the stability and persistence of food webs (Holt and Hoopes, 2005). For example, trophic metacommunity theory has provided us with tools to understand the persistence of otherwise unstable predator-prey dynamics in a landscape. Specifically, the movement of individuals can create negative density dependence that stabilizes multi-species communities, allowing recovery when populations reach low abundances (Holt, 1984, 1985; Huxel and McCann, 1998; Briggs and Hoopes, 2004). Trophic metacommunity theory has also allowed us to understand the emergence of complexity and higher trophic levels in food webs (Pillai *et al.*, 2011). However, until recently, most metacommunity studies have relied on the assumption that all species within a food web disperse equally. In natural communities, this assumption is unlikely to be met (Guzman *et al.*, 2019). When species at different trophic levels differ in dispersal, classic patterns in metacommunity ecology, such as the humped relationship between diversity and dispersal rate, may be altered (Haegeman and Loreau, 2014). More broadly, differences in dispersal ability — and the directionality of the difference — between any two interacting species are expected to have diverging consequences on metacommunity dynamics, (Guzman *et al.*, 2019). For example, when prey populations disperse less than predators, the prey density will vary in the landscape. On the other hand, when the predator disperses less than the prey, the prey is able to persist in patches free of predators (McCauley *et al.*, 1993; Pedersen and Guichard, 2016).

While both theoretical and empirical research on metacommunity ecology has pushed the field in remarkable directions, research that integrates both at once has lagged behind. Specifically, the parameters often used in metacommunity models are rarely quantified empirically. The majority of empirical metacommunity studies have instead compared patterns observed in real communities with potential patterns originating from metacommunity models. For example, variation partitioning analysis (Logue *et al.*, 2011) estimates the amount of community variation explained by environmental or spatial components, and then these estimates are interpreted as signatures of a particular metacommunity dynamic (Cottenie, 2005). This approach has been extended to study differences between generalist and specialist species (Pandit *et al.*, 2009), species of different body sizes and dispersal modes (Bie *et al.*, 2012), and even among different taxonomic groups along an isolation gradient (Driscoll and Lindenmayer, 2009). In parallel, other empirical studies have also considered the effects of dispersal mode (Jones *et al.*, 2015) or constraints of spatial drivers (Grainger *et al.*, 2017) on metacommunity distribution and diversity, yet these studies were not designed to estimate the parameters that are used in metacommunity theory (e.g., dispersal rates). Studies that compare observed and predicted patterns, without parametrizing models, may fall into the trap of assuming that a particular process underlies the pattern when in fact the pattern could be generated in multiple ways (Gilbert and Bennett, 2010). The main reasons we find this mismatch between theoretical and empirical trophic metacommunities is: (i) the lack of clear and concise predictions of trophic metacommunity theory (Guzman *et al.*, 2019) and, (ii) the challenge of quantifying dispersal kernels - especially as we expect each species to have a unique dispersal kernel (Borthagaray *et al.*, 2015).

To integrate trophic metacommunity theory with empirical estimates, we used a model system for metacommunity research (Srivastava *et al.*, 2004),

which consists of multiple interacting macro-invertebrates that inhabit the aquatic habitat created by the leaves of bromeliads. From this system, we estimated (i) the dispersal rate and kernel and (ii) the feeding rates of a predator-prey pair (the top predator *Leptagrion andromache* (Zygoptera: Odonata) and one of its prey *Trentepohlia sp.* (Tipulidae: Diptera)). To estimate the dispersal kernel, we sampled individuals from multiple populations at different spatial distances and then used population genetics to estimate dispersal between these populations. Specifically, we used a genotyping-by-sequencing approach (Elshire *et al.*, 2011) with a Bayesian method that uses individual genotypes to estimate rates of dispersal between populations (Wilson and Rannala, 2003). Genotyping by sequencing (GBS) does not rely on any previous knowledge of a species genome (for example to develop primers) and does not require a reference genome. Instead, GBS uses restriction enzymes to digest the genome at multiple sites and PCR to introduce adapters used for Illumina sequencing. The result is short sequences that can be aligned to identify individual genotypes. Once the genotypes have been identified, Bayesian analysis can extract information about recent dispersal from disequilibrium at individual genotypes of migrants.

Using these estimates of dispersal, we asked: (i) whether the predator and the prey were dispersal limited by the same geographical barriers at large spatial scales; (ii) whether the dispersal rate and kernel were the same for the predator and the prey; (iii) whether a simple metacommunity model could predict differences in space use based on the estimated differences in dispersal parameters; and (iv) whether the observed dispersal kernels led to greater persistence or occupancy of either the predator or the prey.

## 4.3 METHODS

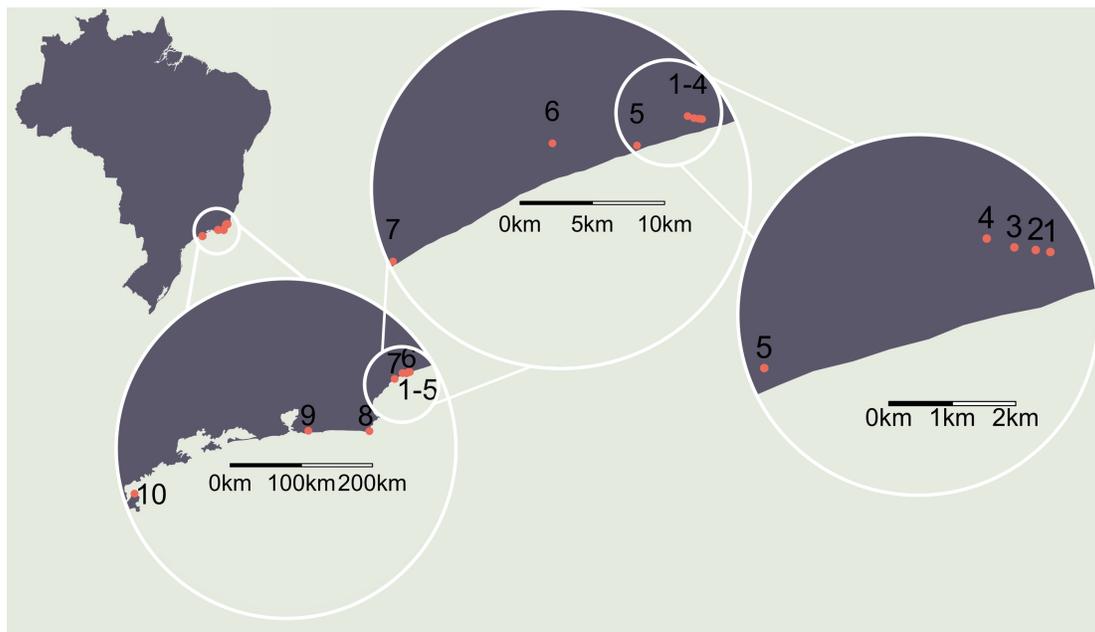
### 4.3.1 *Study system*

Tank bromeliads accumulate water and detritus inside their leaf axils, providing habitat for communities of aquatic macroinvertebrates (Kitching, 2000). Inside each bromeliad, these aquatic macroinvertebrates interact to form a food web comprised of detritivores, filter feeders, intermediate predators and top predators. These communities have been identified as useful study systems for metacommunity ecology (Srivastava *et al.*, 2004). Each bromeliad is perceived as a patch of habitat and overall bromeliads form a naturally patchy landscape for the food webs of macroinvertebrates. Previous studies on bromeliad metacommunities have shown that ten isolated bromeliads are insufficient habitat to sustain the metacommunity through time (LeCraw *et al.*, 2014).

### 4.3.2 *Sampling*

The study area was located in the sand dunes of coastal Brazil (Figure 1), in the states of Rio de Janeiro and São Paulo. We sampled ten sites, seven of which were within the Jurubatiba National Park in Rio de Janeiro state, Brazil. The other three sites were located in the sand dunes of Arraial do Cabo (Rio de Janeiro), Marica (Rio de Janeiro), and Ilha Bela (Sao Paulo). Our sites were selected so that the distance between adjacent sites increased logarithmically (Figure 4.1). Sampling was done between February and May of 2015. In each of our sites we collected 10-20 individual larvae of two species in order to have enough replication of genetic variation within each site and make meaningful comparisons between sites; the top predator *Leptagrion andromache* (Zygoptera: Odonata) and one of its prey *Trentepohlia sp.* (Tipulidae: Diptera).

FIGURE 4.1: Sites were located in the sand dunes of coastal Brazil, in the states of Rio de Janeiro and São Paulo. We sampled the sites such that sites were increasingly further apart. Sites one to four were within 1 km, while sites one to ten were 430 km apart.



#### 4.3.3 DNA extractions

We extracted DNA from the odonate predator larvae using Qiagen DNA blood and tissue kit (Qiagen, Inc., Valencia, California, USA) and concentrated the DNA using Agencourt AMPure XP beads. We extracted the tipulid prey larvae using Qiagen DNA blood and tissue kit and the OmniPrep™ kit from G-Biosciences (St. Louis, MO) for the smaller individuals (modified protocols are in Appendix D).

#### 4.3.4 Genotyping - by - sequencing (GBS)

Extracted DNA of the odonate predator was sent to the Cornell Institute for Genomic Diversity (Ithaca, NY, USA) and extracted DNA of the tipulid prey was sent to University of Wisconsin-Madison Biotechnology centre since the Cornell Institute for Genomic Diversity lost its patent throughout our study. In both cases, the extracted DNA was processed with GBS. GBS (Elshire *et al.*, 2011) is a simple technique for constructing reduced representation libraries for the Illumina sequencing platform. In GBS, DNA from each individual is digested using a restriction enzyme (in this case PstI (CTGCAG)). The fragmented DNA is then ligated to an adaptor that matches long PCR primers. During PCR, these long primers add a length of sequence to the fragments, that bind to the Illumina flow cell.

#### 4.3.5 SNP calling

Because neither of our species has a reference genome, raw sequences were converted into individual genotypes using a UNEAK pipeline (Lu *et al.*, 2013). The UNEAK pipeline clusters identical reads into tags. All unique tags are merged and their counts stored. Then pairwise alignments of tags are performed and tags with 1-bp mismatch are considered candidate SNPs. Of

those candidate SNPs, only reciprocal tags, which involve only two tags with 1-bp mismatch are more likely to be true SNPs than those which belong to complicated networks of tags. Rare tags are considered to be sequencing errors, so tags that have relative counts of less than 0.03 are removed. After identifying the SNPs, counts of each tag (or allele) are output for each locus and each individual. Within the pipeline we excluded tags that were present less than 15 times, and set the minimum minor allele frequency to 5% and the minimum proportion of individuals covered by at least one tag to 10% (Stansell *et al.*, 2018; Lu *et al.*, 2013; Michalski *et al.*, 2017).

After the UNEAK pipeline output the SNPs for each individual, we filtered the SNPs so that the maximum number of times a tag was present was 2000, heterozygosity at each locus was less than 0.5, and loci had at most 40% missing data. For this filtering we used the Tassel 5.0 software (Bradbury *et al.*, 2007). The resulting data consisted of 4161 SNPs for the odonate predator and 1675 SNPs for the tipulid prey.

#### 4.3.6 *Historical ancestry*

We estimated the genetic distance between each individual pair using the 1 - identity-by-state metric (1 - IBS), which is the probability that alleles drawn at random from two individuals pairs at the same locus are the same (Bradbury *et al.*, 2007). After finding that some individuals were very similar and other individuals were very dissimilar (Figure D.1), we ran an admixture analysis to identify major groups of individuals (Alexander *et al.*, 2009). We ran the admixture analysis five times with a different random seed and performed a cross validation procedure to find the best number of clusters (from one to five). We chose the number of clusters that had the lowest cross-validation value, yielding two clusters for both the odonate predator and the tipulid prey.

We separated the individuals for each of these clusters and then re-ran the admixture analysis for each cluster for both species.

#### 4.3.7 *Recent dispersal*

To estimate recent rates of dispersal (within the last few generations) we used the software Bayesass (Wilson and Rannala, 2003). Bayesass is a Bayesian method that uses individual genotypes to estimate rates of recent dispersal using transient disequilibrium observed in individuals that migrated. This method does not assume that genotypes are at Hardy-Weinberg equilibrium. We ran Bayesass for 107 iterations, discarding the first 106 iterations as burn-in, and sampled every 1000 iterations. We adjusted the mixing parameters for the migration rates to 0.3, allele frequencies to 0.6 and inbreeding coefficients to 0.1 for the odonate and 0.2 for the tipulid such that the acceptance rate of these parameters was within the recommended range by the software. We repeated this process four times for each species, assessed the convergence of the MCMC and pooled the outcomes of the four chains.

#### 4.3.8 *Dispersal kernel*

We inferred a dispersal kernel for both the odonate predator and the tipulid prey by regressing the estimated dispersal rates between each population (from Bayesass) with the distance between each population. We filtered the distance to only include sites one to seven, because only two individuals from site eight were present in the major clade of the odonate and the tipulid. We fit: (i) a GLM with a Gamma family and a log link; (ii) an exponential decay function; and (iii) model the distance between patches as a continuous function of their physical distance,  $x$ , but treated sites within a patch as spatially identical ( $x=0$ ). We also used a hurdle function, which allowed for a

probability,  $D_0$ , that an individual would stay in the same site, coupled with an exponential distribution among those individuals that leave a site:

$$D(x) = \begin{cases} \text{if } x = 0 : & D_0 \\ \text{if } x > 0 : & D_1 e^{-bx} \end{cases} \quad (4.1)$$

#### 4.3.9 Feeding trials

We fit a type II functional response (eq. 4.2) to a consumption experiment of the odonate predator with the tipulid prey. This experiment was previously reported in Guzman Srivastava (In review) and the data is publicly available (Guzman and Srivastava, 2018). We fit a type II functional response because the parameters that determined a type III functional response were not significantly different from zero.

$$N_e = \frac{aN}{1 + ahN} \quad (4.2)$$

$N_e$  is the number of prey consumed,  $N$  is the starting prey density,  $a$  is the attack rate of the predator,  $h$  is the handling time of the predator, that is, the time taken to search and consume the prey (Real, 1977).

#### 4.3.10 Trophic metacommunity model

Using the dispersal and consumption parameters, we built a simple metacommunity model, based on a Rosenzweig-MacArthur (Rosenzweig and MacArthur,

1963) model with a type II functional response in a 2-dimensional grid with 25 patches. The abundance of the prey  $N$  at time  $t + 1$  in patch  $j$  was given by:

$$N_{t+1,j} = N_{t,j} + rN_{t,j}\left(1 - \frac{N_{t,j}}{K}\right) - \frac{aN_{t,j}P_{t,j}}{1 + ahN_{t,j}} + \sum_{i \neq j} D_N(x)N_{t,i} - (1 - D_{N0})N_{t,j} \quad (4.3)$$

where  $N_{t,j}$  is the abundance of the prey at time  $t$  in patch  $j$ ,  $r$  is the intrinsic growth rate of the prey,  $K$  is the carrying capacity of the prey,  $a$  is the attack rate of the predator and  $h$  is the handling time of the predator.  $D_N(x)$  is the dispersal rate of the prey as a function of the distance  $x$  between the patches  $j$  and  $i$  and  $N_{t,i}$  is the abundance of the prey at time  $t$  in patch  $i$ .  $D_{N0}$  is the proportion of individuals of the prey that stay in patch  $j$ . The abundance of the predator at time  $t + 1$  in patch  $j$  was given by:

$$P_{t+1,j} = P_{t,j} + r_p P_{t,j} \left(1 - \frac{P_{t,j}}{K_p}\right) + \frac{P_{t,j} B a N_{t,j}}{1 + a h N_{t,j}} - C P_{t,j} + \sum_{i \neq j} D_P(x) P_{t,i} - (1 - D_{P0}) P_{t,j} \quad (4.4)$$

where  $P_{t,j}$  is the abundance of the predator at time  $t$  in patch  $j$ . Since the odonate in this system is top predator that could persist through the consumption other prey species, we included a growth term  $r_p$  and carrying capacity term  $K_p$  which allow the predator to persist without the tipulid prey. The caloric value of a captured prey individual is signified by  $B$ . The per capita mortality rate of the predator is  $C$ .  $D_P(x)$  is the dispersal rate of the predator as a function of the distance  $x$  between the patches  $j$  and  $i$  and  $P_{t,i}$  is the abundance of the prey at time  $t$  in patch  $i$ .  $D_{P0}$  is the proportion of individuals of the predator that stay in patch  $j$ .  $D_P(x)$  and  $D_N(x)$  were determined using equation 4.1.  $D_N(x)$  and  $D_{N0}$  are parameterized based on the tipulid's dispersal kernel,  $D_P(x)$  and  $D_{P0}$  are parameterized based on the

odonate's dispersal kernel. Because the odonate has a generation time of 9 months while the tipulid larvae has a generation time of 2 months, and the dispersal rates were estimated per generation, we divided the dispersal rate by the generation time to obtain a dispersal rate per day. The attack rate and the handling time were parameterized based on the consumption data (Table 4.1).

We ran the simulation model over several combination of parameters of  $B$ ,  $C$ ,  $r$ ,  $K$ ,  $r_p$  and  $K_p$  and found combinations of parameters that yielded persistence -non-zero abundance after 2000 time steps- for both the predator and the prey. We evaluated the occupancy of the predator and the prey, where they differed in occupancy as long as the prey was present in at least one patch without the predator. We also ran the simulation (across the same combination of parameter space) by interchanging the dispersal parameters of the predator and the prey.

The dispersal kernel was fitted using the package `nlsTools` (Baty *et al.*, 2015) and all the simulations were run with the programming language R (Team, 2018). This research was enabled in part by support provided by WestGrid ([www.westgrid.ca](http://www.westgrid.ca)) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)).

## 4.4 RESULTS

### 4.4.1 *Historical ancestry*

Each species was clearly split into two distinct clusters (Figure D.1) both in the identity-by-state distance and in an admixture analysis. The admixture analysis showed that the cross validation error was minimized between two and five clusters depending on the run. However, the log likelihoods of these cases differed by less than four. Therefore, we chose the simplest model for both species (Figure D.2, D.3): two clusters for the odonate predator and two

TABLE 4.1: The parameter values used for the simulations of the metacommunity model.

Parameter	Description	Empirical estimates	Theoretical values
$a$	Attack efficiency of the predator	0.170	
$h$	Handling time of the predator	1.170	
$D_{P0}$	Proportion of the predator population staying in patch j	0.850	
$D_{P1}$	Initial dispersal rate of the predator	0.020	
$b_P$	Predator's exponential decay constant	0.003	
$D_{N0}$	Proportion of the prey population staying in patch j	0.807	
$D_{N1}$	Initial dispersal rate of the prey	0.035	
$b_N$	Prey's exponential decay constant	0.030	
$B$	Caloric value of a capture individual		0.45, 0.48, 0.51, 0.54, 0.57, 0.60, 0.63, 0.66, 0.69, 0.72, 0.75
$C$	Predator death rate		0.10, 0.13, 0.16, 0.19, 0.22, 0.25, 0.28, 0.31, 0.34
$r$	Growth rate of the prey		0.5, 0.6, 0.7, 0.8, 0.9, 1.0
$K$	Carrying capacity of the prey		9, 12, 15, 18, 21, 24, 27, 30
$r_P$	Growth rate of the predator		0.01, 0.05, 0.10, 0.50
$K_P$	Carrying capacity of the predator		1, 5, 10

clusters for the tipulid prey. For the odonate predator, we found that two individuals from site 8 are very similar to sites 1 to 4, and we found two individuals from site 4 that are very similar to site 8. This result may be due to swapped labels. Since we used the entire individual for this analysis, we cannot confirm the provenance of these individuals, and we will assume from now they were indeed errors. We ran the analysis with the original labels but did not interpret dispersal from these individuals.

After separating the individuals of each species into the two separate clusters, we re-ran the admixture analysis separately for each cluster for each species (Figure D.4). Hereafter, we refer to the major clade as the cluster that had the majority of the individuals (odonate had 67 and the tipulid 64 individuals). For the odonate, we found that the major clade (Figure 4.2a) included individuals from sites 1 to 7 and two individuals from site 8. The individuals from the major clade form two sub-clusters, where some individuals draw some admixture. The odonate's minor clade (Figure 4.2c), included two individuals from site 4, and all other individuals from site 8, 9 and 10. The individuals of the minor clade, form three sub-clusters which mirror major geographical barriers, as site 10 is in a small island and the city of Rio de Janeiro separates sites 9 and 10. These are the sites that are furthest apart.

For the tipulid, we found that the major clade (Figure 4.2b) has no sub-clusters, and therefore all individuals at this scale are inter-breeding. The major clade includes individuals from site 1 to 8. The minor clade (Figure 4.2d) includes three sub-clusters, two of which mirror the same geographical barriers as the odonate (from sites 9 and 10). The last sub-cluster includes individuals from sites 4 to 6 and site 8. This sub-cluster co-occurs in the same sites as the major cluster, yet they do not interbreed. Due to the prevalence of cryptic species in insects (Bickford *et al.*, 2007) and this evidence of reproduc-

tive isolation, we decided to treat these major and minor clusters as separate taxa for both the odonate and the tipulid and ran subsequent analysis only on the taxa with most individuals.

#### 4.4.2 *Recent dispersal*

Bayesass provides the most likely source population for each individual, the generation at which they dispersed and the dispersal rate between each pair of populations. We plotted the source and sampling site for each tipulid and odonate individual from the taxa with most individuals (Figure 4.3). We find that the majority of odonate individuals were assigned to the source site from which they were sampled (Figure 4.3a). We interpret this result as indicating that individuals move little between one generation and the next. The only individual that was assigned a different source site, had dispersed two generations ago (Figure 4.3a) from site three to two, moving approximately 350 m. On the other hand, many tipulid individuals moved from their assigned source site and the site where they were sampled (Figure 4.3b). Some of these individuals moved within one generation or two generations ago. The longest distance that a tipulid individual moved was from site 4 to 7, approximately 23 km.

We next considered how the dispersal rate between each pair of sites related to the geographic distance between each pair of sites (Figure 4.4). We only build the dispersal kernels up to 25 km since we found that the longest any individual moved was 23 km and we were confident that the individuals at this scale belonged to the same taxa. We used AIC to distinguish between three different functions for this relationship. For the odonate predator (Figure 4.4a), the hurdle function that included the exponential decay ( $AIC = -283.235$ ,  $df = 4$ ) was a better function than the exponential decay alone ( $AIC = -234.655$ ,  $df = 3$ ) and the generalized linear model with

FIGURE 4.2: The odonate major clade (a) has two sub-clusters while the tipulid's major clade (b) does not. Both the odonate (c) and the tipulid (d) minor clades have three sub-clusters. Every vertical bar represents an individual, and the colours represent the ancestral population. Individuals who have multiple colours are said to be drawing 'admixture', that is, individual who likely resulted from the interbreeding of multiple populations. The different colours delineate the ancestral populations for those individuals.

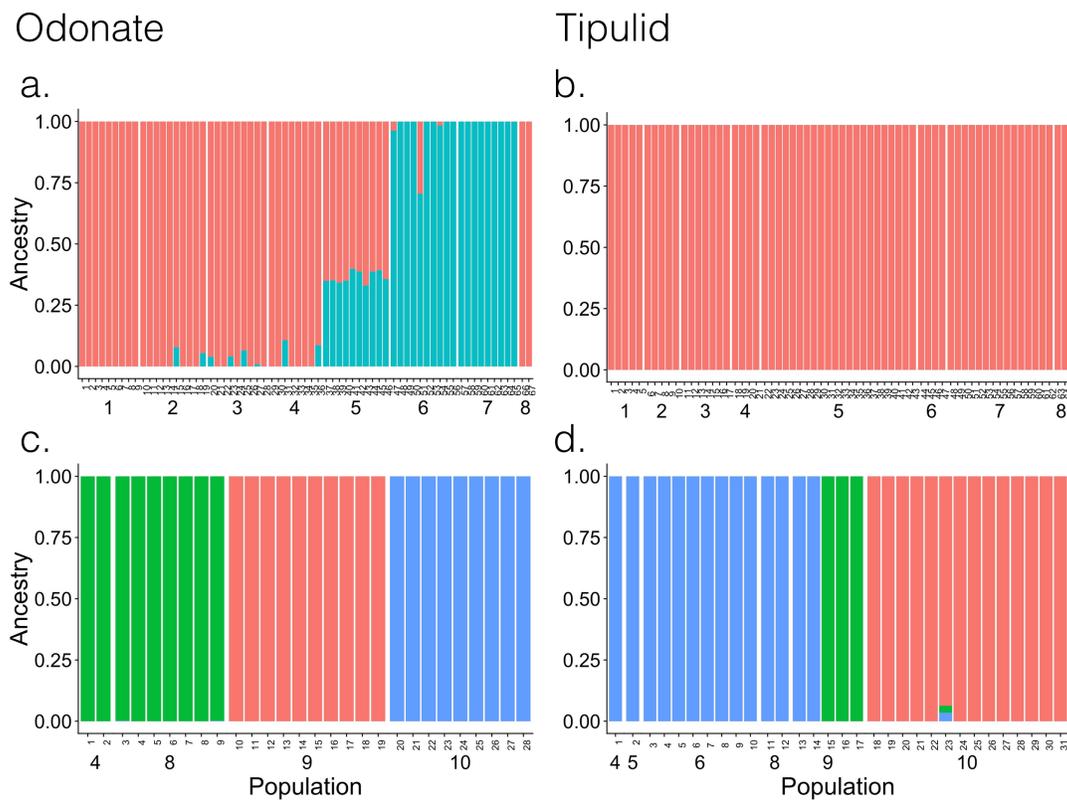
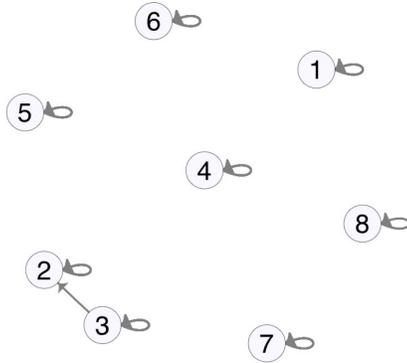


FIGURE 4.3: The ancestry of individual odonates (a) and tipulids (b) is reconstructed from their genetic composition. Dispersal events from one site to the next are represented by arrows, each circle represents a site. Loops represent individuals whose source site was the same as their collection site. The ancestry of the odonate individuals was commonly assigned to the site where they were collected (loops in a) and only one odonate individual was assigned as dispersing within two generations. On the other hand, the ancestry of tipulid individuals was commonly assigned to other sources site (b) including dispersal event within one generation and two generations.

a. Odonate



b. Tipulid

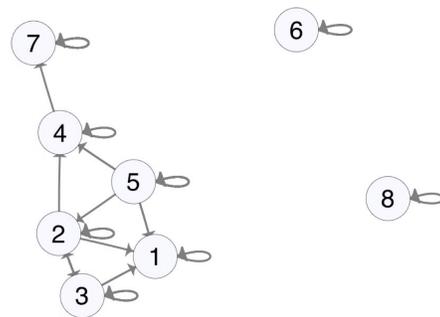
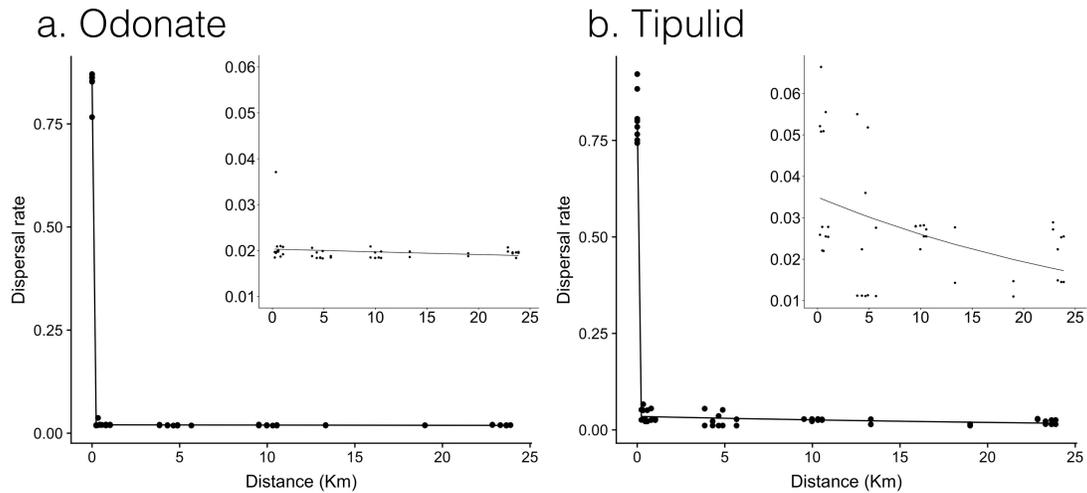


FIGURE 4.4: Estimated dispersal rate between sites at different distances for the odonate predator (a) and the tipulid prey (b). Insets highlight the dispersal rate when the distance between sites is greater than zero.

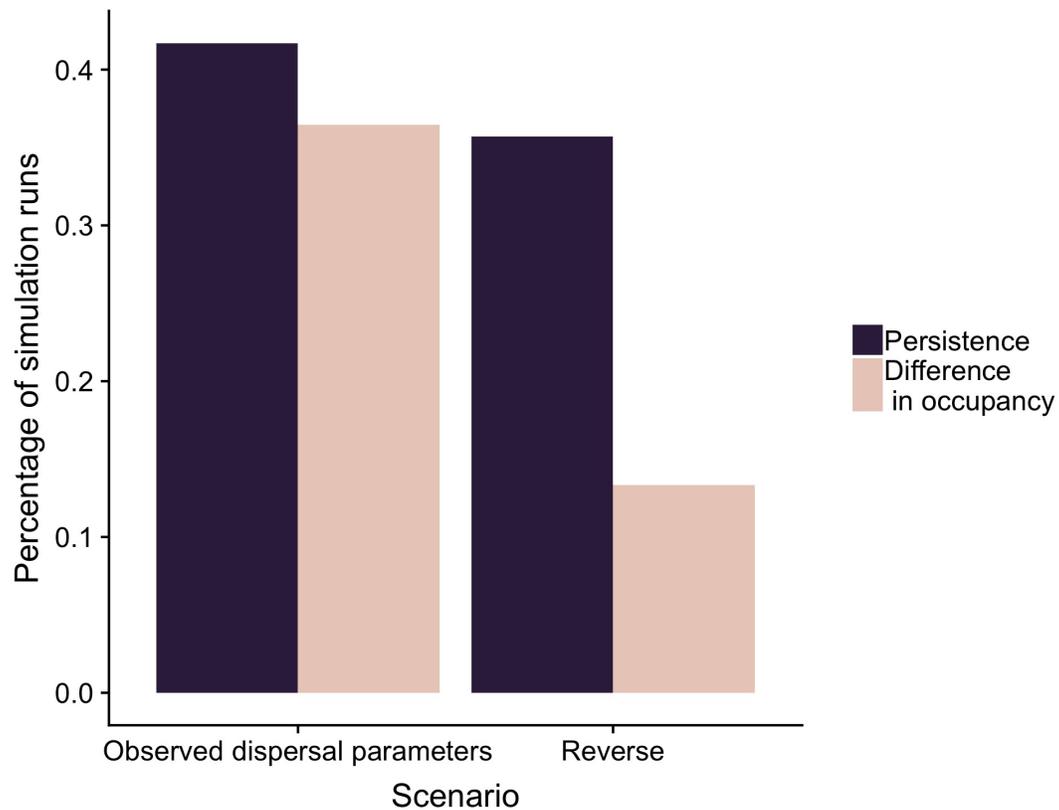


gamma distribution ( $AIC = -203.989, df = 3$ ). Similarly, for the tipulid prey (Figure 4.4b) the hurdle function that included the exponential decay ( $AIC = -213.281, df = 4$ ) was better than the exponential decay alone ( $AIC = -187.039, df = 3$ ) and the generalized linear model with gamma distribution ( $AIC = -179.799, df = 3$ ). Overall we found that the tipulid's dispersal rate is higher than the odonate's dispersal rate at all distances.

#### 4.4.3 Trophic metacommunity model

We used the best fit models for the dispersal kernels of odonates and tipulids to parameterize a simple trophic metacommunity model. We tested whether the observed differences between predator and prey in dispersal rate would be predicted to result in differences in occupancy. We also tested whether the reverse pattern (i.e. the predator having a higher dispersal rate than the prey) would also produce the same difference in occupancy. Of the parameter

FIGURE 4.5: Persistence of the predator and prey at the end of the simulation under different parameter combinations. The percentage of simulation runs where both predator and prey persists varies depending on the dispersal scenario. Similarly, the percentage of simulations runs where the predator and prey differ in their occupancy depends on the dispersal scenario.



combinations tested, the observed dispersal rates resulted in more scenarios where both the predator and the prey persisted than the reverse scenario (Figure 4.5), and more cases where the predator and the prey differed in occupancy -the prey was found without the predator in at least one patch- than the reverse scenario (Figure 4.5).

## 4.5 DISCUSSION

We found first, that at large spatial scales (100, 160 and 430 km), both the predator and the prey are similarly structured — in each taxa, individuals from sites 8, 9 and 10 separated into different clusters. However, at intermediate spatial spaces (0-23 km for the odonate and 0-100 km for the tipulid) the predator and the prey had a different structure — the prey consists of one interbreeding population, while the predator consists of two spatially structured clusters. Second, we found parallel prey species or sub-species that were not interbreeding while occupying the same sites. Third, we estimated higher dispersal rates for the tipulid prey than the odonate predators. This result is consistent with the previous finding at the same intermediate spatial scale: the prey had one interbreeding cluster, while the predator had two clusters, with some individuals drawing admixture. In terms of the dispersal kernel, we found that the rate of dispersal of the tipulid prey was higher at longer distances. Overall we found that while the odonate predator barely dispersed, the tipulid prey was able to disperse up to 25 km. Finally, we found that when we parameterized the simple metacommunity model with the estimated dispersal kernels, the predator and the prey were likely to differ in their use of space. This was less frequently the case when the predator and the prey had the reverse dispersal kernels.

Dispersal variation within a metacommunity, while often deemed necessary for trophic metacommunity theory, has proven hard to quantify (Borthagaray *et al.*, 2015). For some systems, like streams, dispersal of multiple species has been easier to measure, due to the constrained nature of the system where individuals can be physically trapped by mesh (Elliott, 2003). Similarly, wind dispersed organisms from temporary rock pools have also been readily measured, where individuals can be physically trapped using sticky surfaces, and show higher rates of dispersal than previously thought

(Vanschoenwinkel *et al.*, 2008). Yet for many other systems, similar direct measures of dispersal are not easily obtained, especially at a community level. In such systems, population genetics has been identified as useful tool for estimating dispersal (Robledo-Arnuncio and García, 2007; Ouborg *et al.*, 1999; Broquet and Petit, 2009). Many studies have used different estimates of population differentiation to quantify dispersal for single species. In a community context, measures of population differentiation have most commonly been used in marine systems (Kinlan and Gaines, 2003; Weersing and Toonen, 2009). However, these studies often ignore how dispersal combines with species interactions to affect metacommunity dynamics. Here we show not only the dispersal estimate of two interacting species using population genetics, but we also show the consequences of the observed pattern of differential dispersal for metacommunity dynamics using simulations.

Both the odonate predator and the tipulid prey seemed to be comprised of two separate taxa. The identity-by-state results showed a bimodal distribution of distances between individuals, meaning that some individuals seem to be closely related while others are not. Similarly, the admixture analysis initially separated all individuals in two separate clusters, and by re-running the analysis on each cluster separately we were able to find population structure within each cluster. The strongest evidence that supports the existence of two separate taxa, at least for the tipulid prey, is that individuals that occupied the same sites still segregated into different clusters that did not interbreed. This would be the case of sympatric cryptic species (Bickford *et al.*, 2007). In the case of the odonate predator, we do not have the evidence to differentiate between cryptic species or simply isolation by distance since the two separate clades were allopatric. In our study we identified the insects at the larval stage, which for tipulidae hinders the proper identification of species that are often done by rearing the larvae to adults (Gelhaus, 2009).

Based on scaling relationships between body mass and dispersal (Jenkins *et al.*, 2007), we expected that the odonate predator, being larger, would disperse further than the tipulid larvae. However, we found the opposite result. One possibility is that, even though odonates have the capacity to fly further, some species are philopatric and therefore might oviposit in the same area where they emerged as adults (Conrad *et al.*, 1999; McCauley, 2007). While we do not have evidence of whether the odonate predator used in this system is philopatric, our population structure analyses are consistent with this hypothesis. Another possibility is that the tipulid is not just an active disperser, but it may have a mixed strategy which is facilitated by wind or birds. For example, tipulid larvae have been found to be viable after being ingested by birds (Frisch *et al.*, 2007) which may help explain the inferred dispersal over 20 km. Overall, we find that common scaling relationships — such as that between dispersal distance and body mass (Jenkins *et al.*, 2007) — may apply generally to fauna but not necessarily to any particular species pair. However, our results only encompass two interacting species in a food web. The question of whether species within a food web follow the same scaling relationships for dispersal as entire taxonomic groups remains an understudied question in ecology (Guzman *et al.*, 2019).

The question of whether predators or prey disperse more or less than one another is most important when we are interested in the dynamic consequences of dispersal in a metacommunity. Trophic metacommunity models have found that differential dispersal may stabilize predator-prey dynamics. Many metacommunity models incorporate fixed differences between patches in their conditions, which causes the dynamics of the patches to be asynchronous and therefore more stable (Briggs and Hoopes, 2004). For example, models of source-sink dynamics assume patches differ in quality (Holt, 1984). However, many of these models incorporate the differences in dispersal as ex-

treme asymmetries (Holt, 1984) or by incorporating differences in the dispersal rate and not the dispersal kernel (Rohani *et al.*, 1996). We expect that if species have differences in dispersal parameters, we would observe differences in the way they use space. Indeed, we found that the observed differences in the dispersal rate and kernel, in combination with the way these two species interact, were sufficient to generate differences in space use between the predator and the prey without invoking any differences in the abiotic niche. Surprisingly, we found in our model that the observed directionality in dispersal asymmetry, with the prey having a higher dispersal rate than the predator, was more likely to generate differences in space use than the reverse. We urge ecologists to use population genetic methods to explicitly parameterize dispersal kernels of interacting species, and combine these with studies of interaction strengths between species. Only with empirical estimates of the relevant parameters can we hope to apply metacommunity models to understanding community dynamics over space.

SUCCESSFUL INTEGRATION OF DATA SCIENCE IN  
UNDERGRADUATE BIostatISTICS COURSES USING COGNITIVE  
LOAD THEORY.

---

## 5.1 CHAPTER SUMMARY

Biostatistics courses are integral to many undergraduate biology programs. Such courses have often been taught using point-and-click software, but these programs are now seldom used by researchers or professional biologists. Instead, professionals typically use programming languages, such as R, which are better suited to analysing large and complex datasets. These coding skills are valued not only in common "tech" fields, but also in non-traditionally-tech fields, such as healthcare. As a result, many biology programs are not providing their students the skills they will need in the future. However, teaching statistics and programming simultaneously has the potential to overload the students and hinder their learning. We sought to mitigate this potential overload by using cognitive load theory to develop assignments for two biostatistics courses (introductory and advanced). We evaluated the effectiveness of these assignments by comparing cohorts that were taught R using these assignments to those that were taught R just through example scripts or were instructed on a point-and-click software program. We surveyed all cohorts after the courses, and analyzed statistical and programming ability through students' lab reports or final exams. Students that learned R through our assignments rated their programming ability higher and were more likely to put the usage of R as a skill in their CVs than control students. We also found that the treatment students were more motivated, less frustrated and less stressed when using R. There was no cost of learning R for their understanding of statistical concepts. These results suggest that we can use cognitive load theory to teach challenging material and better prepare students for modern careers in biology.

## 5.2 INTRODUCTION

Today, more than ever, biology graduates need to be equipped with statistical and programming skills. At the interface between statistics and programming is data science. Biology graduates require these data science skills in the job market or in graduate programs. Data science has been one of the fastest growing careers in North America. The employment site Glassdoor rated data science as the best job in the USA in 2018 (Glassdoor, 2018) and being a data scientist is regarded as the "sexiest job of the 21st century" (Davenport and Patil, 2012). In environmental and conservation sectors, employers (including government, non-profit, and private) list technical and statistical skills as important requirements they look for in their potential hires (Blickley *et al.*, 2012). These technical skills are not only important for data science or professional biologist jobs, but also employers in 'non-tech' jobs are demanding them, such as employers in marketing, engineering, finance, manufacturing, design and even healthcare (Dishman, 2016). Biology graduates interested in research also need strong statistical and programming skills. With biology firmly in the era of big data, biologists from multiple disciplines are starting to grapple with handling, processing and analyzing large data (Marx, 2013). Not only is data becoming larger, but analyses are becoming more sophisticated. In our own discipline of ecology, sophisticated and computationally intensive statistical techniques, such as mixed models and Bayesian statistics, are replacing more traditional frequentist-based tests such as ANOVAs or t-tests (Barraquand *et al.*, 2014; Touchon and McCoy, 2016). The increased need for data science solutions for biological data has resulted in a growing demand for software that can do these analyses. Specifically, the programming language R is in much higher demand and more commonly used both in commercial applications and in academic research than point-and-click statistical software such as JMP, SAS, and SPSS (Touchon and McCoy, 2016; Muenchen, 2017).

Biology education, at undergraduate and graduate levels, rarely provides students with the statistical and programming skills that they need for their future careers. One proposed solution to this problem, at the graduate student level, is to provide students with accelerated learning programs at the beginning of their graduate programs (Stefan *et al.*, 2015; Vale *et al.*, 2012). However, a recent study by Feldon *et al.* (2017) found that short format training courses, such as "bootcamps", do not provide students with the desired skills. One explanation for this result is that students learn quantitative skills best when taught incrementally over a long time frame rather than intensively (Rohrer, 2015). It seems then, that a better place to introduce programming and statistical skills is at the undergraduate level (Michener and Jones, 2012). Teaching data science skills to biology undergraduates will provide them with the skills they need, not only for graduate school but also for a demanding job market.

Given that biology undergraduates require simultaneous training in statistics and programming, the question is how this can be most effectively achieved. Teaching either statistics or programming is challenging. Both statistics and programming are courses in which students report high levels of anxiety, with debilitating effects on academic performance (Onwuegbuzie and Wilson, 2003; Wilson and Shrock, 2001). For example, the main predictors of student success in introductory programming courses is feeling comfortable while working on computer assignments and being able to ask questions (Wilson and Shrock, 2001; Simon *et al.*, 2006). Statistics and programming courses not only induce high anxiety in students, they also are perceived to be hard courses to learn. Programming for example, requires that students use both deep (understanding application of concepts) and surface (e.g. memorization of syntax) learning at the same time, and therefore students have trouble learning when instruction is primarily through lectures (Bellaby *et al.*, 2003) or when

they do not have adequate support on assignments (Bellaby *et al.*, 2003; Wilson and Shrock, 2001; Jenkins, 2002). The simultaneous instruction of statistics and programming will only increase the cognitive load on students. One strategy for this problem is to use cognitive load theory to design hands-on assignments (Wilson, 2018). Cognitive load theory deals with how cognitive resources are distributed during learning and problem solving (Sweller *et al.*, 1990).

Cognitive load theory suggests that learners have a limit in their working memory. There are three components of cognitive load: (i) Intrinsic load is the inherent difficulty of the instructional material. It is related to the number of elements that learners need to consider simultaneously to learn a particular procedure and the prior knowledge of the learner (Sweller and Chandler, 1994). (ii) Extraneous load is determined by the manner in which the instructional materials are presented. Since students have limited cognitive resources, using cognitive resources to process the extraneous load reduces the available resources for the intrinsic load (Sweller, 1993). Finally (iii) the germane load is the processing and creation of mental models. The germane load can be modified by instructors through the materials presented (Paas *et al.*, 2004). By recognizing these three aspects of cognitive load, instructors can design the scope and nature of their teaching so as to minimize the intrinsic and extrinsic loads while emphasizing the germane load.

We used cognitive load theory to design regular homework assignments to teach R programming in two biostatistics courses. In particular, we used three pedagogical methods based on cognitive load theory to design our assignments: the worked-example effect, where studying worked examples results in better performance of the students (Renkl, 2005); the completion effect, where we required students to complete partially solved problems (Paas and Van Merriënboer, 1994); and the split-attention effect, where an integrated

teaching of multiple concepts can improve learning compared to presenting the concepts separately but concurrently in a "split" format (Ayres and Sweller, 2005). We compared student cohorts that applied R using assignments based on cognitive load theory with cohorts that either applied R strictly through reference to example scripts or applied a point-and-click software. We investigated whether (i) the students effectively learned to use R, (ii) the introduction of R programming hindered the learning of statistics, (iii) the students felt that they learned a useful skill, (iv) the students felt frustrated or overwhelmed with the assignments, and (v) the students were motivated because they felt they were learning something useful and challenging.

## 5.3 METHODS

### 5.3.1 *Target courses*

We implemented this experiment at the University of British Columbia (Canada), in an introductory biostatistics course, Fundamentals of Biostatistics (hereafter Biostatistics), and an advanced ecological statistics course, Ecological Methodology (hereafter Eco-Methods). Biostatistics introduces the concepts of hypothesis testing, probability, experimental design, and statistical tests such as student's t-test, linear regression and ANOVA. Biostatistics includes three 50 minute lectures and one 2-hour optional computer laboratory per week. Eco-Methods introduces the concepts of experimental design, statistical power and sample size, mark and recapture methods, metrics of community diversity and composition, as well as statistical tests such as ANOVA, multiple regression, ordination and clustering. Eco-Methods includes includes two 60 minute lectures and one 3-hour field and/or computer laboratory per week.

For each course, we had a control and a treatment term (Table 5.1). All courses included homework assignments for a relatively small reward in terms

of overall marks. The main difference between the treatment and control terms was the teaching of R using cognitive load theory in the homework assignments (Box 5.1, Appendix E). The assignments taught and tested the ability to apply the statistical concepts in R. In Biostatistics we aggregated the previous homework assignments and introduced cognitive load theory for conceptual questions taken from the textbook. In this course we included two R questions in the midterm and the final exams. The control terms were different for each course. In Biostatistics the students in control term learnt how to use the point-and-click software JMP. In Eco-Methods the students in the control term learnt how to use R using example scripts. In all courses and terms, the in-class sessions consisted of Socratic lecturing.

Although the instructor differed between the control (2016) and treatment (2017) terms for Eco-Methods, both instructors taught from the same lecture slides. We note that in 2015 instructor D.S.S taught R from the same example scripts as M.K.T. in 2016, and that their teaching evaluations were comparable between these two years, suggesting that there was not a strong effect of instructor identity.

### 5.3.2 *Homework assignments*

We designed ten homework assignments for Biostatistics and seven homework assignments for Eco-Methods. In each of these assignments, we applied cognitive load theory to (i) reduce the extraneous load of students by taking advantage of the split-attention effect, the worked-example and completed problem effect, (ii) reduce the intrinsic load of the material by managing the element interactivity and (iii) increased germane load by scaffolding the material with self-explanation questions (See examples in Box 5.1 and Appendix E - All the materials have been submitted to CourseSource).

TABLE 5.1: Course structure in control vs. treatment terms. The treatment groups for both courses completed assignments designed using the ideas of cognitive load theory (CLT) as homework

	Biostatistics		Eco-Methods	
	Control	Treatment	Control	Treatment
Year	2016	2018	2016	2017
n	155	116	26	30
Instructor	M.W.P.	M.W.P.	M.K.T.	D.S.S.
TAs	5	5	2	2
Grade breakdown	Assignments (3): 10% Homework (10): 10% Midterm exam: 30% Final exam: 50%	Homework assignments (10): 20% Midterm exam: 30% Final exam: 50%	Homework assignments (5): 25% Formal lab reports (two at 15% each): 30% Research proposal, group project: 10% Group project presentations: 5% Group project written report: 25% Participation: 5%	Homework assignments (7): 28% Formal lab reports (three at 11% each): 33% Research proposal, group project: 11% Group project presentation: 5% Group project written report: 21% Participation: 2%
Labs	Labs used JMP	Labs used R	Labs used R/Microsoft Excel <sup>TM</sup>	Labs used R
Homework assignments	Homework were conceptual problems from the textbook	Homework assignments used R and CLT	R scripts that they had to run on their own time and do conceptual statistics homework	Homework assignments used R and CLT

---

**Box 5.1 - Homework assignment examples**

Selected examples from the assignments showing how we used cognitive load theory to introduce R programming concepts in the statistics exercises.

---

(i) Reducing the extraneous load

---

*Split attention effect:*

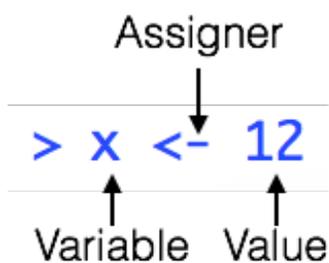
Code is often presented as multiple sources of information. We incorporated the code and the explanations as a single source to reduce the split attention effect.

*Worked example effect:*

We presented worked examples of simple and complex problems, both involving how to write code to run statistical tests. All worked examples were partitioned into different parts.

*Completed problem effect:*

After presenting worked examples, we presented partially completed problems were the scaffolding was introduced in the steps to solve a question and the code needed to run a statistical test.



**Question:** Calculate the mean of a vector of all the integers from 1 to 50.

**First,** we must create the vector.

```
vector <- 1:50
```

**Third,** construct your box plot using ggplot. Fill in the blanks in the following code to do

so:

```
> ggplot(data =
```

**Second**, we must calculate the mean.

```

> mean(vector)
[1] 25.5

```

**Finally**, we now have our answers calculated by R. The mean of a vector from 1 to 50 is 25.5.

---

### (ii) Reducing the intrinsic load

---

We reduced the element interactivity of the material by:

1. Presenting only one way to do a task. In R, every task can be done by multiple functions. While understanding these different function is useful for more advanced programming, beginners can be overwhelmed by learning multiple functions simultaneously.
2. Presenting only the functions that were needed for a given statistical test.

---

### (iii) Increasing germane load

---

In both worked examples and in partially completed problems, we asked the students to reflect on a part of the question to engage in germane load activities such as self-explaining.

*Self-explanation questions:*

**For you to think:** Why did you use ":" instead of "c" to create a vector in "vector <- 1:50"?

---

### 5.3.3 Evaluating the success of the homework assignments

**SURVEYS** — We evaluated students' perceptions using a survey at the end of the course (Appendix E). Student participation in the survey was requested by

L.M.G., who was not a course instructor or TA, during lecture time. We did not offer any incentive to complete the survey but it also presented no cost to the students, it was anonymous and it was conducted with the instructors absent from the room. Both surveys consisted of 29 closed-response questions and 3 open-response questions. The survey included questions on the frequency of using R before and after the course took place, as well as attitudes about the difficulty of the course and their emotional response to the data science material during the course.

ASSESSMENTS OF LAB REPORTS AND FINAL EXAMS — To address the question of whether the students learned R effectively, we used analyses particular to each course. For Eco-Methods we evaluated graphs produced by students for their final report. In these graphs we assessed whether student were able to customize graphs relative to the example graph provided in the assignments. For example, we recorded whether the students changed the colour, type of lines, font, background of their figures, etc. We compared the customization of graphs between the control and the treatment groups. For Biostatistics, we asked the students to upload and submit graphs as part of their assignments. The students were required to do nine graphs as the course progressed. We compared the proportion of students creating graphs by hand, in Microsoft Excel<sup>TM</sup>, or in R, each week, from week 1 to week 10. We only did this temporal comparison in the treatment group, as we were interested primarily in the progression to R from other methods.

We were also interested in determining whether the introduction of R programming would hinder the learning of statistics. Biostatistics was the only course with a final exam. We ensured that this exam had one conceptual question in common between the control and the treatment group. We then compared the scores for this question.

#### 5.3.4 *Data analysis*

**SURVEYS: LIKERT-TYPE QUESTIONS** — Likert-type responses were converted into numerical values from 1 to 5. For these types of questions, we used a generalized linear model to test for differences in the response between control and treatment terms. We used a Poisson family with a logarithmic link function since these data were comprised of integer responses from 1 to 5.

**SURVEYS: CONCEPTUAL DIFFICULTY QUESTIONS** — We tested whether students perceived the difficulty of the conceptual material differently from that of the application of the material, and whether this difference depended on the topic taught (e.g. ANOVA v.s. linear regression) and the course treatment (control v.s. treatment). Here our expectation was that as the course progressed, the perceived difficulty of the conceptual material would increase while the perceived difficulty of the application would decline in the treatment group but not the control. To analyze these predictions, we used a generalized linear model with a Poisson family and a logarithmic link function.

**SURVEYS: EMOTIONAL RESPONSE QUESTIONS** — We asked the students to assess their emotions towards both the conceptual parts of the course and the use of R or JMP. We transformed all positive feelings into values of 1 and all negative feelings into values of 0. For these types of questions, we used a generalized linear model to test for differences in the response due to the treatment (control vs. treatment) or due to the use of R v.s. JMP (for Biostatistics). We used a Binomial family with a logit link function.

To investigate which particular feelings contributed most to this difference, we evaluated, for each feeling separately, the difference between the treatments using a Chi-squared contingency test. We corrected the p-values for multiple

comparisons using the false discovery rate method (Benjamini and Hochberg, 1995). We excluded from this analysis all feelings that had less than ten responses.

**SURVEYS: OPEN-RESPONSE QUESTIONS** — We developed codes for each of these questions using the method described in Guest *et al.* (2012). Two observers, E.N. and L.M.G., generated and reviewed the codes, the themes and the codebook (Appendix E).

**ASSESSMENTS OF LAB REPORTS, MIDTERM AND FINAL EXAMS** —  
Eco-Methods Lab reports:

To analyze the degree of customization of the graphs in the Eco-Methods lab reports we summed the total number of customizations per person and then used a generalized linear mixed effects model to test for differences between the courses. We used the number of customized elements per person as the response and the treatment as the fixed effect. Since the lab reports were done in groups of four students, the reports generated were not independent, therefore we used the group id as the random effect. We used a Poisson family with a logarithmic link function.

Graphs in the Biostatistics assignments:

We used multinomial logistic regression to test for the odds of using R vs other methods as the term progresses.

Biostatistics Final exam marks:

We tested the differences in scores between the control and treatment groups

in Biostatistics using paired questions in the final exam. Here we used a generalized linear model using a Poisson family and logarithmic link function where the response was the question score on this question and the explanatory variable was the course.

All analyses were done using the R programming language (Team, 2018). Mixed effect models were performed using *lme4* R package (Bates *et al.*, 2015) and *lmerTest* (Kuznetsova *et al.*, 2017), analysis of variance was done using *car* (Fox and Weisberg, 2011), and multinomial regressions were done using *nnet* R package Venables and Ripley (2002).

#### 5.3.5 *Human Subjects Oversight*

This work was conducted with review and approval by the Behavioural ethics research board of the University of British Columbia, H16-02319.

## 5.4 RESULTS

### 5.4.1 *Surveys*

LIKERT-TYPE QUESTIONS — Students in the control and the treatment cohorts, for both courses, rated similarly their initial programming skills and the frequency of using R (Figure 5.1: "Before this course started").

For Biostatistics (Introduction to Biostatistics), we found that during the course the students in the treatment group self-rated their frequency of using R higher in lab assignments and outside of class than the control group rated their frequency of using JMP (Figure 5.1: "During this course"). This difference occurred even though the students in the control cohort started the course feeling more comfortable using statistical software than the experimental cohort. By the end of the course, the students who used the R assignments

designed with cognitive load theory self-rated a higher proficiency in R, a higher willingness to put R as a skill in their CV and a higher chance that they would continue using R in their future graduate and undergraduate studies.

In Eco-Methods, we could compare the students' comfort in using R when it was taught traditionally (control cohort) or using cognitive load theory (experimental cohort). Here there were no treatment effects before or during the course, but by the end of the course students in the experimental cohort felt more proficient in R and were more likely to include it as a skill on their CV (Figure 5.1: "Having completed this course"). Both cohorts were likely to use R in future projects.

**CONCEPTUAL DIFFICULTY QUESTIONS** — For both courses, we found no effect of teaching treatment on the perceived difficulty of the course, including the conceptual v.s. applied parts, or the different materials taught (Appendix E - Table E.1).

**EMOTIONAL RESPONSE QUESTIONS** — We found that both the Biostatistics and the Eco-Methods students in treatment cohorts had more positive feelings than the students who were taught JMP traditionally (Biostatistics) or were taught R traditionally (Eco-Methods) ( $X^2_{1,517} = 25.68, P \ll 0.001$  and  $X^2_{1,163} = 11.57, P \ll 0.001$  respectively). Specifically, the students in the Biostatistics treatment cohort felt more excited, happy, motivated, proud and less bored, and the students in the Eco-Methods experimental cohort felt less frustrated (Table 5.2).

**OPEN-RESPONSE QUESTIONS** — We focus on two questions (out of three questions asked) for each course about the way the statistical software (either R or JMP) was taught. The third question, "did you have any other comments about the course", was too broad and resulted in many comments not relevant

FIGURE 5.1: Students responses to the survey questions before, during and after the course, in relation to teaching treatment (control vs. cognitive load theory treatment) and course identity. Student responses are ranked on a Likert scale (all but first question) or referring to year of undergraduate (first question). Points and bars represent means and standard errors respectively. Control groups are coloured red and treatment groups are coloured blue. Significance is noted with stars.

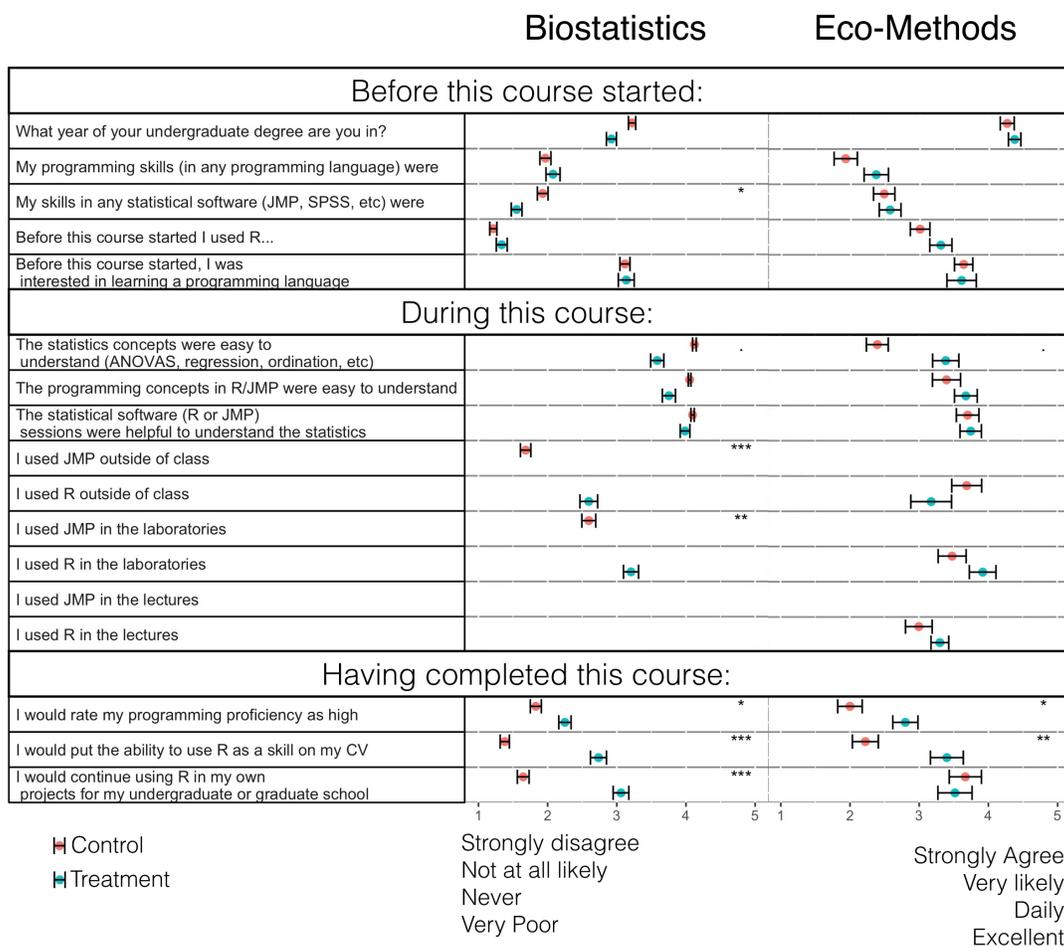


TABLE 5.2: Treatment students in Biostatistics significantly felt less bored, and more excited, happy, motivated and proud than control students. Treatment students in Eco-Methods felt less frustrated than control students.  $X^2$  and P-value of the  $X^2$  test for each emotion.

Emotion	Biostatistics		Eco-Methods	
	$X^2$	P	$X^2$	P
angry	0.02	0.98		
annoyed	0.005	0.98	0.73	0.59
anxious	0.55	0.62	1.45	0.45
bored	6.52	0.03 **		
excited	17.94	< 0.001 ***	0.20	0.66
frustrated	0.51	0.62	10.89	0.01 **
happy	8.92	0.009 ***		
motivated	30.04	< 0.001 ***	3.73	0.16
overwhelmed	1.05	0.50	1.33	0.45
proud	10.66	0.005 ***	0.26	0.66
scared	0.001	0.98		
stressed	3.52	0.13	5.24	0.10
supported	1.53	0.40	0.29	0.66

to the assignments. For each question and each course we emphasize the top 5 themes.

### Biostatistics

Question: If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

The most prevalent themes we found in this question were (i) "Theme A: Course should use other software", which had 47 responses in total (Control = 47 out of 157 students, Treatment = 0 out of 117 students). In Biostatistics, the control cohort learnt JMP in the labs whereas the treatment cohort learnt R. Overall, 30% of the students wanted to use another software, the students mentioned both R and Excel in their answers. (ii) "Theme N: The R assignments need improvement". Theme N had 29 responses (0, 29). 25% of the students responded in this theme. Specifically, seven students mentioned that the assignments needed more clarity, or better instructions. Four students mentioned that the assignments were too easy, while six mentioned it was too challenging and four mentioned that they were too long. Two students mentioned that the assignments were disconnected from the lectures or the labs and finally four students did not like the layout in the online learning management system (Canvas) since they had to scroll between instructions and questions. (iii) "Theme B: The students want more activities to help or force them to learn the software" which had 23 responses (8, 15). (iv) "Theme H: The course should provide an incentive to come to the labs to learn the statistical software". Theme H had 19 responses (13, 6). And (v) "Theme C: The course should provide more support learning the statistical software" which had 16 responses (5, 11). Some of the suggestions that the students

provided were: more help troubleshooting, provide demos or explain better new commands, functions and concepts, give more walkthroughs and provide more information on online resources.

Question: If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

The most prevalent themes we found in this question were (i) "Theme K: Keep some part of the canvas R assignments" which had 49 responses (42% of the students) (Control = 0 out of 157 students, Treatment = 49 out of 117 students). In particular, 18 students suggested to keep the walkthroughs, 12 students the detailed instructions, 5 students the step-by-step questions, 3 students the fill-in-the blank questions, the expected codes and graphs. 4 students mentioned the assignments were informative and not overwhelming. (ii) "Theme A: Keep the lab manual" which had 39 responses (26, 13). (iii) "Theme C: Keep the labs" which had 15 responses (12, 3). (iv) "Theme N: Liked how R was taught/ no changes" which had 14 responses, all in the treatment group. (v) "Theme G: They did not like how JMP was taught" which had 9 students responses, all in the control group.

#### Eco-Methods

Question: If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

The most prevalent themes we found in this question were (i) "Theme B: Students want more R instructions on functions and packages" which had 28 responses (Control = 17 out of 27 students, Treatment = 11 out of 30 students).

(ii) "Theme E: More synchrony between the assignments, the lectures and labs" which had 13 responses (6, 7). (iii) "Theme C: More support outside of class" which had 11 responses (10, 1). (iv) "Theme A: the student wants to learn more R" which had 9 responses (6, 3). Finally, (v) 8 students responded with "Theme G: the teaching of R was not good enough", which had 8 responses (7, 1).

Question: If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

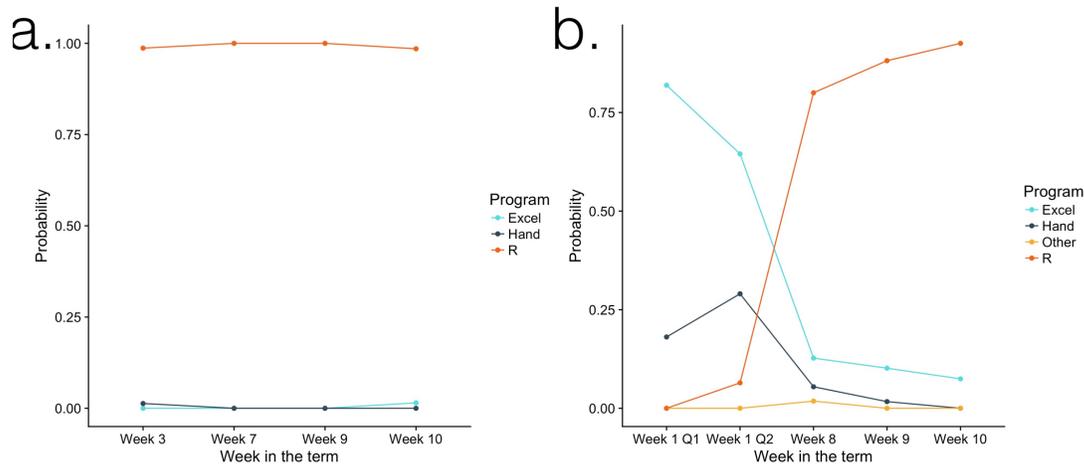
The most prevalent themes we found in this question were (i) "Theme C: R was taught well", which had 21 responses (Control = 4 out of 27 students, Treatment = 17 out of 30 students). (ii) "Theme E: the students liked having an R Workshop" which had 19 responses (15, 4). Here we found that the students liked having the first in-lab session devoted to learning to start using R, which occurred for both the treatment and the control groups. (iii) "Theme I: Liked the R/stats assignments" which had 16 responses (2, 14). (iv) "Theme F: The students are grateful to have learned R" which had 5 responses (2, 3). Finally, "Theme A: Learning packages/analyses/functions was useful" which had 4 responses (2, 2).

#### 5.4.2 *Assessments of lab reports and final exams*

LAB REPORTS — Students in the control and treatment cohorts of Eco-Methods did not differ in the number of customized elements on their graphs ( $X^2_{1,58} = 1.67, P = 0.19$ , project group as random effect).

GRAPHS IN THE ASSIGNMENTS — The students in the treatment group of Biostatistics had to do two types of graphs: graphs from data provided in the

FIGURE 5.2: a) The probability that Biostatistics students in the treatment cohort made their graphs using R from previous assignment examples is high from the beginning of the course. b) The probability that Biostatistics students in the treatment cohort made their graphs using R from the textbook increases as the term progresses and replaces use of Microsoft Excel™ ("Excel") or hand drawing ("Hand") or other software.



textbook, where they had to input and graph the data without an example, and graphs from data provided in the labs, where the data was already formatted and easy to input and the graph was based on an example. We found that when the students had an example, they were able to produce a graph of their data using R from the beginning of the term (Figure 5.2a). However, without an example, we found that at the beginning of the term they used Microsoft Excel™ or drew the graph by hand, but by the end of the term, the majority of the students were able to input and graph their data using R (Figure 5.2b).

FINAL EXAM MARKS — The treatment and control cohorts of Biostatistics did not differ in their scores on the same question in the final exam ( $X^2_{1,244} = 0.64, P = 0.42$ ).

## 5.5 DISCUSSION

Overall we found that students not only learnt to use R, but they felt that they mastered a useful skill, they had positive feelings when using the assignments and liked the assignments. Examining the final reports (Eco-Methods) and the progression of the use of R for the graphs through the term (Biostatistics), we found that the students learnt how to input, plot and analyze data. We also found that the introduction of R programming did not hinder the learning of statistics (at least in Biostatistics), since the results from the final exam were not significantly different between the control and treatment.

Overall we found that students appreciated learning R, regardless of the format in which it was taught. For example, a student from the Biostatistics control group (which used the JMP program) wrote: *"I wish I learned R because it seems more relevant to my degree and I wish it was part of homework and assignments [C97]."* As well, those students who were taught R generally felt it was valuable; one student wrote that they were *"glad [they] learned R, as [they]'ve heard it's very useful in biology especially."* [E52 - Biostatistics] and another thought the course could be improved by adding even more R into the class as this was *"probably the most useful part of this course moving forward"* and that they *"would have liked more assignments that required more problem solving"* [E113 - Biostatistics].

Self-determination theory states that there are multiple sources of extrinsic motivation. When a student identifies the value or utility of a task, the extrinsic goal is self-endorsed and thus adopted. Identifying the utility of task is a form of extrinsic motivation that has been associated with greater engagement, performance, higher quality learning, among other outcomes (Ryan and Deci, 2000). Students who learnt R using our assignments based on cognitive load theory felt that they learnt a useful skill and they would rate their programming proficiency as high after the treatment level. Similarly,

they would put the ability to use R as a skill in their CV. Regarding the student's affect, we found that the students reported feeling more motivated when learning R than when learning JMP. Additionally, we found that the students felt more positive when using the treatment assignments to learn R than when either learning JMP or using only scripts to learn R. Specifically, the students who used the R assignments in Biostatistics felt more excited, happy, motivated, proud and less bored than the students who used JMP. In Eco-Methods, the students who used the CTL-based R assignments felt less frustrated than the control students who used the R scripts. Part of this positive response may be due to the students liking some elements of cognitive load theory that we introduced in the assignments. For examples, when we asked the students what they would keep about the way the software was taught they wrote that they liked how the assignments *"walked you through the questions almost step-by-step" [E3]*, how *"everything was broken down and explained to a very basic level [as] it made it very enjoyable to learn for someone who really struggles with computer programming" [E19]* and how they *"made sure your code was right and gave hints too if you were on the right track" [D14]*. Consistent with the principles behind cognitive load theory, we also found that the design of the assignments influenced whether students perceived that they were able to be successful. For examples, one student wrote: *"I liked the fill-in-the-blanks especially the question with the expected graphs because I could test it out and it gave me some sense of support" [E74]*; and another *"really liked how the instructions walked us through the process so it was less overwhelming" [D7]*.

Previous studies have found that when teaching novice students, boredom and frustration were negatively correlated with learning, while transitioning between confusion and engagement were positively correlated with learning (Bosch and D'Mello, 2015). While we do not measure affect throughout the term, we found that using assignments designed with cognitive load theory

reduces the frustration compared to plain R scripts, and reduces boredom compared to a point and click software JMP.

Cognitive load theory has been used successfully in a variety of courses. For example, Mason *et al.* (2016) used cognitive load theory to re-design a course in database systems. They found that the failing rate of mid to lower performing students was reduced by 34% after the redesign on identical final exams. Student satisfaction also increased and feedback was very positive (Mason *et al.*, 2016). Similarly, on an Advanced Web Applications course for graduate students, cognitive load theory was used to develop an online programming tool, and they found that students performed best when they were able to view examples of code during the learning of new material (Heo and Chow, 2005). Similarly, when cognitive load theory was applied to teaching math to middle school students, researchers found that student performance was improved by signaling important information, improving the aesthetic of item organization and removing extraneous content (Gillmor *et al.*, 2015). Previous studies on teaching programming to novice learners have also found that using cognitive load theory led to better learning as well as an increase in self-efficacy and reductions in the perception of difficulty (Mason and Cooper, 2013).

When we designed the assignments, we included multiple types of scaffolding, including procedural scaffolding (helps the learners use appropriate resources as well as tools), and metacognitive scaffolding (helps the learners to reflect about what they are learning). Metacognitive scaffolding and self-questioning have been shown to support student learning of programming (Nurulain *et al.*, 2017). In addition, affect can have metacognitive effects, such as feelings of difficulty (Efklides, 2017). For example, a negative mood can increase the self-reported difficulty in math problem solving (Efklides and Petkaki, 2005).

## 5.6 LIMITATIONS

The perceptions expressed by the students may not be generalizable to a larger population. The students who were surveyed were those present on the last day of class, which may reflect a more motivated subset of the class. Furthermore, many students in the University of British Columbia Biology program who take these classes are interested in medical school or graduate school, and this motivation may not extend to students situated in other environments or those enrolled in other programs. This study was unable to control for the possibility of temporal differences in either course or that instructors differed in teaching ability (Eco-Methods). We view these explanations less likely since similar effects were seen in both courses.

## 5.7 CONCLUSION

This is the first evidence, to our knowledge, that using cognitive load theory increased learning success for the introduction of data sciences practices and the integration of programming and statistics, based on two courses in the biology undergraduate program. Each course teaches different concepts in bio-statistics, but we found congruent results in terms of affect and performance of the students. The findings presented here suggest that data science is of interest to students, and cognitive load theory can be useful in introducing programming not only in statistics but also in other courses. Even though we designed these assignments with biology students (and novice programmers) in mind, other disciplines will face the same data heavy method demands and challenges of having to teach quantitative skills to novice undergraduate students. We think that these methods can be applied to other disciplines with discipline-specific examples.

DISCUSSION

---

More generally, through the projects that I did during my Ph.D. I found that empirical approaches that infer processes from observational data are limited. While both competitive and trophic metacommunity theory have advanced incredibly quickly, and have improved our understanding of the role of space in community dynamics, our empirical approaches remain behind. Experiments have been critical at testing whether the predictions made by theoretical models *could* happen in nature (e.g. Forbes and Chase, 2002; Staddon *et al.*, 2010; Cadotte, 2006b; Altermatt *et al.*, 2011). Yet we are still far from confirming these predictions using observational surveys. Often, we use statistical tools to analyze observational data, that do not infer the processes we are interested in. For instance, variation partitioning is a general tool that allows us to examine the degree to which variation in environment or in space can explain variation in community composition. However, this tool is only loosely tied to the processes we are interesting in inferring, i.e. dispersal, abiotic niche and species interactions. In fact, while we attribute processes to the different variance components, multiple processes can produce the same patterns (Gilbert and Bennett 2010). Without explicitly testing and comparing these processes, we will be hopeless in trying to infer process from pattern. Some of the statistical tools that metacommunity ecologists can use, to link the processes we are interested in to the patterns we observed, have already been developed and adopted in other fields. For example, population genetics and phylogenetics often use Bayesian statistics (Falush *et al.*, 2007), approximate bayesian computation (Aeschbacher *et al.*, 2012) and machine learning (Sukumaran *et al.*, 2016). These tools allow them to infer population structure from allelic data or whether dispersal is constrained by traits in radiations of birds.

New studies in ecology seem to be incorporating specific statistical models (Martin *et al.*, 2018; Pontarp *et al.*, 2019), and it is time we start applying them

to metacommunity ecology. A potential worry for ecologists, is that some of these methods do not encapsulate all possible processes that could be driving certain patterns. However, by using specific models that tie the patterns we measure to the processes we want to infer, we show explicitly which processes we are taking into account and which ones we are not. Using these methods that tie directly our observations to the inferences we actually are interested in, will push the field further than using general methods that do not answer the questions we are asking.

The development of new inference techniques for metacommunity ecology is a huge mountain to climb, that will take decades of research to develop. In the meantime, we can use better approximations to directly link patterns to the processes we are interested in inferring. For example, while not for my thesis, I am currently using machine learning methods identify which metrics can allow us to distinguish metacommunity processes from time series data. This concept has been applied previously to differentiate the different metacommunity paradigms (Münkemüller *et al.*, 2011), but since the metacommunity paradigms are only a subset of all combinations of the possible metacommunity processes (Thompson *et al.*, in prep), we need an update on the metrics that will allow us to differentiate the three metacommunity processes — the abiotic niche, dispersal and the strength of local competition — rather than the paradigms. This will enable ecologists to have a richer understanding of the processes that maintain biodiversity.

After completing this dissertation, I found some aspects which could be expanded to allow a more thorough understanding of the bromeliad system. Given that species interactions vary with environmental variables, I think feeding trials should be done in multiple environments, especially those identified as important for the organisms. While metabolic theory focuses mostly on the variation of feeding rates due to body size and temperature, bromeliad

macroinvertebrates might be more sensitive to changes in water volume than to changes in temperature. The relationship between water volume and feeding rate has only been quantified for one predator in the bromeliad food web (Amundrud accepted paper), and could profitably be expanded to many more. Since Jensen's inequality is important whenever the relationship between two variables is not linear, future experiments should repeat the feeding experiment while incorporating intra and not only interspecific variation in body size.

Parameterizing theoretical models in metacommunity ecology allow us to get closer to understand the effect of the different processes in nature. While I parameterized a simple model with the dispersal kernel of two species and their feeding interaction, there are still plenty of parameters that I had to ignore. For example, growth rate, carrying capacity and mortality of the predator are all parameters which I did not estimate. Next steps should also include, not just which environmental conditions promote growth of the different species, but also, how do abiotic gradients change the functional response or dispersal propensity.

Throughout my dissertation, I realized how important computational approaches are for biology. Every single one of my chapters relied on intense computation, often through the access of clusters. I often helped other graduate students with their statistical and computational needs. This led me to realize that many biology programs are not preparing their students for the demands of the job industry or academia. While not all biology undergraduates need strong computational skills, those going into ecology and evolution are likely to need them. All of my chapters inspired me help undergraduate students in biology studying biostatistics to learn R programming. One of the largest difficulties in educational research is replication at the scale of multiple courses. Collaborative networks can help by allowing researchers to

come together and replicate the same study in multiple universities. Graduate students may be more motivated to learn this material already, or may have higher expertise and therefore might not benefit as much. I think it would also be interesting to see if cognitive load theory can help graduate students learning programming.

## BIBLIOGRAPHY

---

- Abrams P.A., Holt R.D., and Roth J.D. 1998. Apparent competition or apparent mutualism? shared predation when populations cycle. *Ecology* 79:201–212.
- Aeschbacher S., Beaumont M.A., and Futschik A. 2012. A novel approach for choosing summary statistics in approximate bayesian computation. *Genetics* 192:1027–1047.
- Albajes R. and Alomar O. 2004. Facultative predators. *In Encyclopedia of Entomology*, pages 818–823, Springer.
- Alerstam T., Hedenstrom A., and Akesson S. 2003. Long-distance migration: evolution and determinants. *Oikos* 103:247–260.
- Alexander D.H., Novembre J., and Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.
- Altermatt F., Bieger A., Carrara F., Rinaldo A., and Holyoak M. 2011. Effects of connectivity and recurrent local disturbances on community structure and population density in experimental metacommunities. *PLoS One* 6:e19525.
- Amarasekare P. 2008. Spatial dynamics of foodwebs. *Annu. Rev. Ecol. Evol. Syst.* 39:479–500.
- Amundrud S.L. and Srivastava D.S. 2015. Drought sensitivity predicts habitat size sensitivity in an aquatic ecosystem. *Ecology* 96:1957–1965.
- Amundrud S.L. and Srivastava D.S. 2016. Trophic interactions determine the effects of drought on an aquatic ecosystem. *Ecology* 97:1475–1483.
- Anderson M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26:32–46.
- Anderson M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62:245–253.
- Anderson R.A. and Karasov W.H. 1981. Contrasts in energy intake and expenditure in sit-and-wait and widely foraging lizards. *Oecol.* 49:67–72.
- Ayres P. and Sweller J. 2005. The Split-Attention principle in multimedia learning. *In The Cambridge Handbook of Multimedia Learning*, pages 135–146, Cambridge University Press, New York.

- Baird Day L., Crews D., and Wilczynski W. 1999. Relative medial and dorsal cortex volume in relation to foraging ecology in congeneric lizards. *Brain Behav. Evol.* 54:314–322.
- Baiser B., Buckley H.L., Gotelli N.J., and Ellison A.M. 2012. Predicting food-web structure with metacommunity models. *Oikos* 122:492–506.
- Banasek-Richter C., Bersier L.F., Cattin M.F., Baltensperger R., Gabriel J.P., Merz Y., Ulanowicz R.E., Tavares A.F., Williams D.D., de Ruiter P.C., Winemiller K.O., and Naisbit R.E. 2009. Complexity in quantitative food webs. *Ecology* 90:1470–1477.
- Barraquand F., Ezard T.H.G., Jørgensen P.S., Zimmerman N., Chamberlain S., Salguero-Gómez R., Curran T.J., and Poisot T. 2014. Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ* 2:e285.
- Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* 19:134–143.
- Bates D., Mächler M., Bolker B., and Walker S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48.
- Baty F., Ritz C., Charles S., Brutsche M., Flandrois J.P., and Delignette-Muller M.L. 2015. A toolbox for nonlinear regression in r: the package nlstools. *J. Stat. Softw.* 66:1–21.
- Beger M., Grantham H.S., Pressey R.L., Wilson K.A., Peterson E.L., Dorfman D., Mumby P.J., Lourival R., Brumbaugh D.R., and Possingham H.P. 2010. Conservation planning for connectivity across marine, freshwater, and terrestrial realms. *Biol. Conserv.* 143:565–575.
- Bellaby G., McDonald C., and Patterson A. 2003. Why lecture? *In* 4th Annual LTSN-ICS Conference.
- Benjamini Y. and Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57:289–300.
- Berlow E.L., Neutel A.M., Cohen J.E., De Ruiter P.C., Ebenman B., Emmerson M., Fox J.W., Jansen V.A., Iwan Jones J., Kokkoris G.D., *et al.* 2004. Interaction strengths in food webs: issues and opportunities. *J. Anim. Ecol.* 73:585–598.
- Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.K., and Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22:148–155.

- Bie T., Meester L., Brendonck L., Martens K., Goddeeris B., Ercken D., Hampel H., Denys L., Vanhecke L., Gucht K., Wichelen J., Vyverman W., and Declerck S.A.J. 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.* 15:740–747.
- Blickley J.L., Deiner K., Garbach K., Lacher I., Meek M.H., Porensky L.M., Wilkerson M.L., Winford E.M., and Schwartz M.W. 2012. Graduate student's guide to necessary skills for nonacademic conservation careers. *Conserv. Biol.* 27:24–34.
- Bolker B. 2012. Rogers random predator equation: extensions and estimation by numerical integration.
- Bolker B. and Team R.D.C. 2017. *bbmle: tools for general maximum likelihood estimation*. R package version 1.0.20.
- Bolnick D.I., Amarasekare P., Araújo M.S., Bürger R., Levine J.M., Novak M., Rudolf V.H., Schreiber S.J., Urban M.C., and Vasseur D.A. 2011. Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* 26:183–192.
- Borcard D., Gillet F., and Legendre P. 2011. *Numerical ecology with R*. 2 ed., Springer, New York Dordrecht London Heidelberg.
- Börger L., Dalziel B.D., and Fryxell J.M. 2008. Are there general mechanisms of animal home range behaviour? a review and prospects for future research. *Ecol. Lett.* 11:637–650.
- Borthagaray A.I., Berazategui M., and Arim M. 2015. Disentangling the effects of local and regional processes on biodiversity patterns through taxon-contingent metacommunity network analysis. *Oikos* 124:1383–1390.
- Bosch N. and D'Mello S. 2015. The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education* 27:181–206.
- Bowman J., Jaeger J.A.G., and Fahrig L. 2002. Dispersal distance of mammals is proportional to home range size. *Ecology* 83:2049.
- Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., and Buckler E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635.
- Brendonck L., Jocqué M., Tuytens K., Timms B.V., and Vanschoenwinkel B. 2015. Hydrological stability drives both local and regional diversity patterns in rock pool metacommunities. *Oikos* 124:741–749.
- Briggs C.J. and Hoopes M.F. 2004. Stabilizing effects in spatial parasitoid-host and predator-prey models: a review. *Theor. Popul. Biol.* 65:299–315.

- Broquet T. and Petit E.J. 2009. Molecular estimation of dispersal for ecology and population genetics. *Annu. Rev. Ecol. Evol. Syst.* 40:193–216.
- Brose U. 2010. Body-mass constraints on foraging behaviour determine population and food-web dynamics. *Funct. Ecol.* 24:28–34.
- Brose U., Williams R.J., and Martinez N.D. 2006. Allometric scaling enhances stability in complex food webs. *Ecol. Lett.* 9:1228–1236.
- Cade B.S. and Noon B.R. 2003. A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* 1:412–420.
- Cadotte M. 2006a. Dispersal and species diversity: a meta-analysis. *Am. Nat.* 167:913.
- Cadotte M.W. 2006b. Metacommunity influences on community richness at multiple spatial scales: a microcosm experiment. *Ecology* 87:1008–1016.
- Cain M.L., Milligan B.G., and Strand A.E. 2000. Long-distance seed dispersal in plant populations. *Am. J. Bot.* 87:1217–1227.
- Calcagno V., Mouquet N., Jarne P., and David P. 2006. Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecol. Lett.* 9:897–907.
- Cazelles K., Araújo M.B., Mouquet N., and Gravel D. 2016. A theory for species co-occurrence in interaction networks. *Theor. Ecol.* 9:39–48.
- Céréghino R., Pillar V.D., Srivastava D.S., de Omena P.M., MacDonald A.A.M., Barberis I.M., Corbara B., Guzman L.M., Leroy C., Ospina Bautista F., *et al.* 2018. Constraints on the functional trait space of aquatic invertebrates in bromeliads. *Funct. Ecol.* 32:2435–2447.
- Chapman J.W., Drake V.A., and Reynolds D.R. 2011. Recent insights from radar studies of insect flight. *Annu. Rev. Entomol.* 56:337–356.
- Chase J.M. and Leibold M.A. 2003. *Ecological niches: linking classical and contemporary approaches.* University of Chicago Press.
- Chittka L., Skorupski P., and Raine N.E. 2009. Speed-accuracy tradeoffs in animal decision making. *Trends Ecol. Evol.* 24:400–407.
- Conrad K.F., Willson K.H., Harvey I.F., Thomas C.J., and Sherratt T.N. 1999. Dispersal characteristics of seven odonate species in an agricultural landscape. *Ecography* 22:524–531.
- Costa-Pereira R., Araújo M.S., Olivier R.d.S., Souza F.L., and Rudolf V.H. 2018. Prey limitation drives variation in allometric scaling of predator-prey interactions. *Am. Nat.* 192:E139–E149.
- Cottenie K. 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.* 8:1175–1182.

- Davenport T.H. and Patil D.J. 2012. Data scientist: the sexiest job of the 21st century. *Harv. Bus. Rev.* 90:70–6, 128.
- de Roos A.M., McCauley E., and Wilson W.G. 1998. Pattern formation and the spatial scale of interaction between predators and their prey. *Theor. Popul. Biol.* 53:108–130.
- De Ryck D.J.R., Robert E.M.R., Schmitz N., Van der Stocken T., Di Nitto D., Dahdouh-Guebas F., and Koedam N. 2012. Size does matter, but not only size: two alternative dispersal strategies for viviparous mangrove propagules. *Aquat. Bot.* 103:66–73.
- DeLong J.P., Hanley T.C., and Vasseur D.A. 2014. Competition and the density dependence of metabolic rates. *J. Anim. Ecol.* 83:51–58.
- Dézerald O., Céréghino R., Corbara B., Dejean A., and Leroy C. 2015. Functional trait responses of aquatic macroinvertebrates to simulated drought in a neotropical bromeliad ecosystem. *Freshwater Biol.* 60:1917–1929.
- Dézerald O., Talaga S., Leroy C., Carrias J.F., Corbara B., Dejean A., and Céréghino R. 2014. Environmental determinants of macroinvertebrate diversity in small water bodies: insights from tank-bromeliads. *Hydrobiologia* 723:77–86.
- Díaz S., Kattge J., Cornelissen J.H.C., Wright I.J., Lavorel S., Dray S., Reu B., Kleyer M., Wirth C., Prentice I.C., Garnier E., Bönisch G., Westoby M., Poorter H., Reich P.B., Moles A.T., Dickie J., Gillison A.N., Zanne A.E., Chave J., Wright S.J., Sheremet'ev S.N., Jactel H., Baraloto C., Cerabolini B., Pierce S., Shipley B., Kirkup D., Casanoves F., Joswig J.S., Günther A., Falczuk V., Rüger N., Mahecha M.D., and Gorné L.D. 2016. The global spectrum of plant form and function. *Nature* 529:167–171.
- Dishman L. 2016. Why coding is still the most important job skill of the future. <https://www.fastcompany.com/3060883/why-coding-is-the-job-skill-of-the-future-for-everyone>, accessed: 2019-1-21.
- Dobson A. 2009. Food-web structure and ecosystem services: insights from the serengeti. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:1665–1682.
- Driscoll D.A. and Lindenmayer D.B. 2009. Empirical tests of metacommunity theory using an isolation gradient. *Ecol. Monogr.* 79:485–501.
- Dunne J.A., Williams R.J., and Martinez N.D. 2002. Food-web structure and network theory: the role of connectance and size. *Proc. Natl. Acad. Sci. U. S. A.* 99:12917–12922.

- Edgerly J., Willey M., and Livdahl T. 1999. Intraguild predation among larval treehole mosquitoes, *aedes albopictus*, *ae. aegypti*, and *ae. triseriatus* (diptera: Culicidae), in laboratory microcosms. *J. Med. Entol.* 36:394–399.
- Edmunds N.B., Laberge F., and McCann K.S. 2016. A role for brain size and cognition in food webs. *Ecol. Lett.* 19:948–955.
- Efklides A. 2017. Affect, epistemic emotions, metacognition, and self-regulated learning. *Teach. Coll. Rec.* 119:130305.
- Efklides A. and Petkaki C. 2005. Effects of mood on students' metacognitive experiences. *Learning and Instruction* 15:415–431.
- Ehrlén J. and Morris W.F. 2015. Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* 18:303–314.
- Elliott J.M. 2003. A comparative study of the dispersal of 10 species of stream invertebrates. *Freshw. Biol.* 48:1652–1668.
- Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., and Mitchell S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379.
- Elton C.S. 1966. Wytham hill. *In* *The Pattern of Animal Communities*, pages 45–61.
- Emmerson M.C. and Raffaelli D. 2004. Predator–prey body size, interaction strength and the stability of a real food web. *J. Anim. Ecol.* 73:399–409.
- Engelbrecht B.M.J., Comita L.S., Condit R., Kursar T.A., Tyree M.T., Turner B.L., and Hubbell S.P. 2007. Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* 447:80–82.
- Englert D.L., Manson M.D., and Jayaraman A. 2009. Flow-based microfluidic device for quantifying bacterial chemotaxis in stable, competing gradients. *Appl. Environ. Microbiol.* 75:4557–4564.
- Englund G., Johansson F., Olofsson P., Salonsaari J., and Öhman J. 2009. Predation leads to assembly rules in fragmented fish communities. *Ecol. Lett.* 12:663–671.
- Fagan W.F., Lewis M.A., Auger-Méthé M., Avgar T., Benhamou S., Breed G., LaDage L., Schlägel U.E., Tang W.W., Papastamatiou Y.P., Forester J., and Mueller T. 2013. Spatial memory and animal movement. *Ecol. Lett.* 16:1316–1329.
- Falush D., Stephens M., and Pritchard J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7:574–578.

- Farjalla V.F., Srivastava D.S., Marino N.A., Azevedo F.D., Dib V., Lopes P.M., Rosado A.S., Bozelli R.L., and Esteves F.A. 2012. Ecological determinism increases with organism size. *Ecology* 93:1752–1759.
- Feldon D.F., Jeong S., Peugh J., Roksa J., Maahs-Fladung C., Shenoy A., and Oliva M. 2017. Null effects of boot camps and short-format training for PhD students in life sciences. *Proc. Natl. Acad. Sci. U. S. A.* 114:9854–9858.
- Fincke O.M. 1994. Population regulation of a tropical damselfly in the larval stage by food limitation, cannibalism, intraguild predation and habitat drying. *Oecol.* 100:118–127.
- Forbes A.E. and Chase J.M. 2002. The role of habitat connectivity and landscape geometry in experimental zooplankton metacommunities. *Oikos* 96:433–440.
- Fox J. and Weisberg S. 2011. *An R companion to applied regression*. Second ed., Sage, Thousand Oaks CA.
- Freilich M.A., Wieters E., Broitman B.R., Marquet P.A., and Navarrete S.A. 2018. Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 99:690–699.
- Frisch D., Green A.J., and Figuerola J. 2007. High dispersal capacity of a broad spectrum of aquatic invertebrates via waterbirds. *Aquat. Sci.* 69:568–574.
- Fronhofer E.A., Klecka J., Melián C.J., and Altermatt F. 2015. Condition-dependent movement and dispersal in experimental metacommunities. *Ecol. Lett.* 18:954–963.
- Galiana N., Lurgi M., Claramunt-López B., Fortin M.J., Leroux S., Cazelles K., Gravel D., and Montoya J.M. 2018. The spatial scaling of species interaction networks. *Nat Ecol Evol* 2:782–790.
- Gamfeldt L., Hillebrand H., and Jonsson P.R. 2005. Species richness changes across two trophic levels simultaneously affect prey and consumer biomass. *Ecol. Lett.* 8:696–703.
- Gelhaus J.K. 2009. Tipulidae (crane flies, tipúlidos). *In* *Manual of Central American Diptera* (B.V. Brown, ed.), pages 193–236, NRC Research Press.
- Gibert J.P. and DeLong J.P. 2017. Phenotypic variation explains food web structural patterns. *Proc. Natl. Acad. Sci. U. S. A.* page 201703864.
- Gilbert B. and Bennett J.R. 2010. Partitioning variation in ecological communities: do the numbers add up? *J. Appl. Ecol.* 47:1071–1082.
- Gillmor S., Poggio J., and Embretson S. 2015. Effects of reducing the cognitive load of mathematics test items on student performance. *Numeracy* 8.

- Giroux M.A., Berteaux D., Lecomte N., Gauthier G., Szor G., and Bêty J. 2012. Benefiting from a migratory prey: spatio-temporal patterns in allochthonous subsidization of an arctic predator. *J. Anim. Ecol.* 81:533–542.
- Glassdoor. 2018. 50 best jobs in america. [https://www.glassdoor.com/List/Best-Jobs-in-America-LST\\_KQ0,20.htm](https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm), accessed: 2018-11-28.
- González A.L., Dézerald O., Marquet P.A., Romero G.Q., and Srivastava D.S. 2017. The multidimensional stoichiometric niche. *Front Ecol Evol* 5:110.
- Gotelli N.J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.
- Gounand I., Harvey E., Little C.J., and Altermatt F. 2017. Meta-ecosystems 2.0: rooting the theory into the field. *Trends Ecol. Evol.* .
- Gounand I., Mouquet N., Canard E., Guichard F., Hauzy C., and Gravel D. 2014. The paradox of enrichment in metaecosystems. *Am. Nat.* 184:752–763.
- Grainger T.N., Germain R.M., Jones N.T., and Gilbert B. 2017. Predators modify biogeographic constraints on species distributions in an insect metacommunity. *Ecology* 98:851–860.
- Grainger T.N. and Gilbert B. 2016. Dispersal and diversity in experimental metacommunities: linking theory and practice. *Oikos* 125:1213–1223.
- Gravel D., Albouy C., and Thuiller W. 2016a. The meaning of functional trait composition of food webs for ecosystem functioning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371.
- Gravel D., Canard E., Guichard F., and Mouquet N. 2011. Persistence increases with diversity and connectance in trophic metacommunities. *PLoS One* 6:e19374.
- Gravel D., Guichard F., Loreau M., and Mouquet N. 2010a. Source and sink dynamics in meta-ecosystems. *Ecology* 91:2172–2184.
- Gravel D., Massol F., and Leibold M.A. 2016b. Stability and complexity in model meta-ecosystems. *Nat. Commun.* 7:12457.
- Gravel D., Mouquet N., Loreau M., and Guichard F. 2010b. Patch dynamics, persistence, and species coexistence in metaecosystems. *Am. Nat.* 176:289–302.
- Greenleaf S.S., Williams N.M., Winfree R., and Kremen C. 2007. Bee foraging ranges and their relationship to body size. *Oecol.* 153:589–596.
- Grigaltchik V.S., Ward A.J.W., and Seebacher F. 2012. Thermal acclimation of interactions: differential responses to temperature change alter predator-prey relationship. *Proc. Biol. Sci.* 279:4058–4064.

- Grilli J., Barabás G., and Allesina S. 2015. Metapopulation persistence in random fragmented landscapes. *PLoS Comput. Biol.* 11:e1004251.
- Gross T., Rudolf L., Levin S.A., and Dieckmann U. 2009. Generalized models reveal stabilizing factors in food webs. *Science* 325:747–750.
- Guest G., MacQueen K., and Namey E. 2012. *Applied Thematic Analysis*.
- Guzman L.M., Germain R.M., Forbes C., Straus S., O'Connor M.I., Gravel D., Srivastava D.S., and Thompson P.L. 2019. Towards a multi-trophic extension of metacommunity ecology. *Ecol. Lett.* 22:19–33.
- Guzman L.M. and Srivastava D.S. 2018. Feeding trials of leptagrion andromache and prey species. Title of the publication associated with this dataset: Zenodo.
- Haegeman B. and Loreau M. 2014. General relationships between consumer dispersal, resource dispersal and metacommunity diversity. *Ecol. Lett.* 17:175–184.
- Hammill E., Atwood T.B., Corvalan P., and Srivastava D.S. 2015a. Behavioural responses to predation may explain shifts in community structure. *Freshwater Biol.* 60:125–135.
- Hammill E., Atwood T.B., and Srivastava D.S. 2015b. Predation threat alters composition and functioning of bromeliad ecosystems. *Ecosystems* 18:857–866.
- Hammill E., Petchey O.L., and Anholt B.R. 2010. Predator functional response changed by induced defenses in prey. *Am. Nat.* 176:723–731.
- Hanski I. 1991. Single-species metapopulation dynamics: concepts, models and observations. *Biol. J. Linn. Soc. Lond.* 42:17–38.
- Hanski I. 1999. Habitat connectivity, habitat continuity, and metapopulations in dynamic landscapes. *Oikos* 87:209.
- Hanski I., Erälahti C., Kankare M., Ovaskainen O., and Sirén H. 2004. Variation in migration propensity among individuals maintained by landscape structure. *Ecol. Lett.* 7:958–966.
- Hanski I., Gilpin M.E., and McCauley D.E. 1997. *Metapopulation biology*, vol. 454. Elsevier.
- Harris D.J. 2015. *rosalia*: Exact inference for small binary Markov networks. R package version 0.1.0.
- Harris D.J. 2016. Inferring species interactions from co-occurrence data with markov networks. *Ecology* 97:3308–3314.

- Harris S., Cresswell W.J., Forde P.G., Trehwella W.J., Woollard T., and Wray S. 1990. Home-range analysis using radio-tracking data—a review of problems and techniques particularly as applied to the study of mammals. *Mamm. Rev.* 20:97–123.
- Haskell J.P., Ritchie M.E., and Olff H. 2002. Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature* 418:527–530.
- Hauzy C., Gauduchon M., Hulot F.D., and Loreau M. 2010. Density-dependent dispersal and relative dispersal affect the stability of predator–prey metacommunities. *J. Theor. Biol.* 266:458–469.
- Hein A.M., Hou C., and Gilgooly J.F. 2012. Energetic and biomechanical constraints on animal migration distance. *Ecol. Lett.* 15:104–110.
- Hein A.M. and McKinley S.A. 2012. Sensing and decision-making in random search. *Proc. Natl. Acad. Sci. U. S. A.* 109:12070–12074.
- Heo M. and Chow A. 2005. The impact of computer augmented online learning and assessment tool. *Journal of Educational Technology & Society* 8:113–125.
- Higginson A.D. and Ruxton G.D. 2015. Foraging mode switching: the importance of prey distribution and foraging currency. *Anim. Behav.* 105:121–137.
- Hirt M.R., Lauermann T., Brose U., Lucas P J, and Dell A.I. 2017. The little things that run: a general scaling of invertebrate exploratory speed with body mass. *Ecology* 98:2751–2757.
- Hobson K.A. 1999. Tracing origins and migration of wildlife using stable isotopes: a review. *Oecol.* 120:314–326.
- Holdo R.M., Holt R.D., and Fryxell J.M. 2009. Opposing rainfall and plant nutritional gradients best explain the wildebeest migration in the serengeti. *Am. Nat.* 173:431–445.
- Holling C.S. 1959. The components of predation as revealed by a study of small-mammal predation of the european pine sawfly. *The Canadian Entomologist* 91:293–320.
- Holt R.D. 1984. Spatial heterogeneity, indirect interactions, and the coexistence of prey species. *Am. Nat.* 124:377–406.
- Holt R.D. 1985. Population dynamics in two-patch environments: some anomalous consequences of an optimal habitat distribution. *Theor. Popul. Biol.* 28:181–208.

- Holt R.D. 1993. Ecology at the mesoscale: the influence of regional processes on local communities. *Species Diversity in Ecological Communities* pages 77–88.
- Holt R.D. 1996. Food webs in space: an island biogeographic perspective. *In Food Webs*, pages 313–323.
- Holt R.D. 2002. Food webs in space: on the interplay of dynamic instability and spatial processes. *Ecol. Res.* 17:261–273.
- Holt R.D. 2007. Ijee soapbox: cultural amnesia in the ecological sciences. *Isr. J. Ecol. Evol.* 53:121–128.
- Holt R.D. 2009. Bringing the hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc. Natl. Acad. Sci. U. S. A.* 106 Suppl 2:19659–19665.
- Holt R.D. and Hoopes M.F. 2005. Food web dynamics in a metacommunity context. *Metacommunities: Spatial dynamics and Ecological Communities* .
- Holt R.D., Lawton J.H., Polis G.A., and Martinez N.D. 1999. Trophic rank and the species-area relationship. *Ecology* 80:1495.
- Holyoak M., Leibold M.A., and Holt R.D. 2005. *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press.
- Hubbell S.P. 2001. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press.
- Huey R.B. and Kingsolver J.G. 1989. Evolution of thermal sensitivity of ectotherm performance. *Trends Ecol. Evol.* 4:131–135.
- Huffaker C.B. 1958. Experimental studies on predation: dispersion factors and predator-prey oscillations. *Hilgardia* 27:343–383.
- Huxel G.R. and McCann K. 1998. Food web stability: the influence of trophic flows across habitats. *Am. Nat.* 152:460–469.
- Jabiol J., Corbara B., Dejean A., and Céréghino R. 2009. Structure of aquatic insect communities in tank-bromeliads in a east-amazonian rainforest in french guiana. *For. Ecol. Manag.* 257:351–360.
- Jacquet C., Mouillot D., Kulbicki M., and Gravel D. 2017. Extensions of island biogeography theory predict the scaling of functional trait composition with habitat area and isolation. *Ecol. Lett.* 20:135–146.
- Jansen V.A. 2001. The dynamics of two diffusively coupled predator-prey populations. *Theor. Popul. Biol.* 59:119–131.

- Jenkins D.G., Brescacin C.R., Duxbury C.V., Elliott J.A., Evans J.A., Grablow K.R., Hillegass M., Lyon B.N., Metzger G.A., Olandese M.L., Pepe D., Silvers G.A., Suresch H.N., Thompson T.N., Trexler C.M., Williams G.E., Williams N.C., and Williams S.E. 2007. Does size matter for dispersal distance? *Glob. Ecol. Biogeogr.* 16:415–425.
- Jenkins T. 2002. On the difficulty of learning to program. *In* 3rd Annual LTSN-ICS Conference.
- Jeschke J.M., Kopp M., and Tollrian R. 2002. Predator functional responses: discriminating between handling and digesting prey. *Ecol. Monogr.* 72:95–112.
- Johansson F. 1991. Foraging modes in an assemblage of odonate larvae? effects of prey and interference. *Hydrobiologia* 209:79–87.
- Johansson J. 2008. Evolutionary responses to environmental changes: how does competition affect adaptation? *Evolution* 62:421–435.
- Jones N.T., Germain R.M., Grainger T.N., Hall A.M., Baldwin L., and Gilbert B. 2015. Dispersal mode mediates the effect of patch size and patch connectivity on metacommunity diversity. *J. Ecol.* 103:935–944.
- Jonsson T., Kaartinen R., Jonsson M., and Bommarco R. 2018. Predictive power of food web models based on body size decreases with trophic complexity. *Ecol. Lett.* 21:702–712.
- Kalinkat G., Jochum M., Brose U., and Dell A.I. 2015. Body size and the behavioral ecology of insects: linking individuals to ecological communities. *Current Opinion in Insect Science* 9:24–30.
- Kalinkat G., Schneider F.D., Digel C., Guill C., Rall B.C., and Brose U. 2013. Body masses, functional responses and predator–prey stability. *Ecol. Lett.* 16:1126–1134.
- Kelt D.A. and Van Vuren D. 1999. Energetic constraints and the relationship between body size and home range area in mammals. *Ecology* 80:337.
- Kiltie R.A. 2000. Scaling of visual acuity with body size in mammals and birds. *Funct. Ecol.* 14:226–234.
- Kinlan B.P. and Gaines S.D. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84:2007–2020.
- Kitching R.L. 2000. *Food webs and container habitats: the natural history and ecology of phytotelmata.* Cambridge University Press.
- Kondoh M. 2003. Foraging adaptation and the relationship between food-web complexity and stability. *Science* 299:1388–1391.

- Kraft N.J., Adler P.B., Godoy O., James E.C., Fuller S., and Levine J.M. 2015. Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29:592–599.
- Krebs J.R., Sherry D.F., Healy S.D., Perry V.H., and Vaccarino A.L. 1989. Hippocampal specialization of food-storing birds. *Proc. Natl. Acad. Sci. U. S. A.* 86:1388–1392.
- Kuznetsova A., Brockhoff P.B., and Christensen R.H.B. 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82:1–26.
- Laca E.A., Sokolow S., Galli J.R., and Cangiano C.A. 2010. Allometry and spatial scales of foraging in mammalian herbivores. *Ecol. Lett.* 13:311–320.
- Laliberté E. and Tylianakis J.M. 2010. Deforestation homogenizes tropical parasitoid–host networks. *Ecology* 91:1740–1747.
- Law R. and Morton R.D. 1993. Alternative permanent states of ecological communities. *Ecology* 74:1347–1361.
- Lawton J. and Hassell M. 1981. Asymmetrical competition in insects. *Nature* 289:793.
- LeCraw R.M., Srivastava D.S., and Romero G.Q. 2014. Metacommunity size influences aquatic community composition in a natural mesocosm landscape. *Oikos* 123:903–911.
- Leibold M.A. 1998. Similarity and local co-existence of species in regional biotas. *Evol. Ecol.* 12:95–110.
- Leibold M.A. and Chase J.M. 2017. *Metacommunity ecology*, vol. 59. Princeton University Press.
- Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F., Holt R.D., Shurin J.B., Law R., Tilman D., Loreau M., and Gonzalez A. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7:601–613.
- Leibold M.A. and Loeuille N. 2015. Species sorting and patch dynamics in harlequin metacommunities affect the relative importance of environment and space. *Ecology* 96:3227–3233.
- Leigh J.W. and Bryant D. 2015. Monte carlo strategies for selecting parameter values in simulation experiments. *Syst. Biol.* 64:741–751.
- Levin S.A., Muller-Landau H.C., Nathan R., and Chave J. 2003. The ecology and evolution of seed dispersal: a theoretical perspective. *Annu. Rev. Ecol. Evol. Syst.* 34:575–604.
- Levin S.A. and Paine R.T. 1974. Disturbance, patch formation, and community structure. *Proc. Natl. Acad. Sci. U. S. A.* 71:2744–2747.

- Levins R. and Culver D. 1971. Regional coexistence of species and competition between rare species. *Proc. Natl. Acad. Sci. U. S. A.* 68:1246–1248.
- Logue J.B., Mouquet N., Peter H., Hillebrand H., and Metacommunity Working Group. 2011. Empirical approaches to metacommunities: a review and comparison with theory. *Trends Ecol. Evol.* 26:482–491.
- Loreau M., Mouquet N., and Gonzalez A. 2003. Biodiversity as spatial insurance in heterogeneous landscapes. *Proc. Natl. Acad. Sci. U. S. A.* 100:12765–12770.
- Lu F., Lipka A.E., Glaubitz J., Elshire R., Cherney J.H., Casler M.D., Buckler E.S., and Costich D.E. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215.
- MacArthur R.H. and Wilson E.O. 1967. *The theory of island biogeography*, vol. 1. Princeton university press.
- Magnuson J.J., Crowder L.B., and Patricia A. Medvick. 1979. Temperature as an ecological resource. *Am. Zool.* 19:331–343.
- Marino N.A. 2015. Allometric equations for some aquatic macroinvertebrates found in tank bromeliads at parque nacional da restinga de jurubatiba.
- Martin B.T., Munch S.B., and Hein A.M. 2018. Reverse-engineering ecological theory from data. *Proceedings of the Royal Society B: Biological Sciences* 285:20180422.
- Marx V. 2013. The big challenges of big data. *Nature* 498:255–260.
- Maser G.L., Guichard F., and McCann K.S. 2007. Weak trophic interactions and the balance of enriched metacommunities. *J. Theor. Biol.* 247:337–345.
- Mason R. and Cooper G. 2013. Mindstorms robots and the application of cognitive load theory in introductory programming. *Computer Science Education* 23:296–314.
- Mason R., Seton C., and Cooper G. 2016. Applying cognitive load theory to the redesign of a conventional database systems course. *Computer Science Education* 26:68–87.
- May R.M. 1972. Will a large complex system be stable? *Nature* 238:413.
- McCann K., Hastings A., and Huxel G.R. 1998. Weak trophic interactions and the balance of nature. *Nature* 395:794.
- McCann K.S. 2011. *Food webs (MPB-50)*, vol. 57. Princeton University Press.
- McCann K.S., Rasmussen J.B., and Umbanhowar J. 2005. The dynamics of spatially coupled food webs. *Ecol. Lett.* 8:513–523.

- McCauley E., Wilson W.G., and de Roos A.M. 1993. Dynamics of age-structured and spatially structured predator-prey interactions: individual-based models and population-level formulations. *Am. Nat.* 142:412–442.
- McCauley S.J. 2007. The role of local and regional processes in structuring larval dragonfly distributions across habitat gradients. *Oikos* 116:121–133.
- Merrill R.M., Gutiérrez D., Lewis O.T., Gutiérrez J., Díez S.B., and Wilson R.J. 2008. Combined effects of climate and biotic interactions on the elevational range of a phytophagous insect. *J. Anim. Ecol.* 77:145–155.
- Michalski S.G., Malyshev A.V., and Kreyling J. 2017. Trait variation in response to varying winter temperatures, diversity patterns and signatures of selection along the latitudinal distribution of the widespread grassland plant *Arrhenatherum elatius*. *Ecol. Evol.* 7:3268–3280.
- Michener W.K. and Jones M.B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol. Evol.* 27:85–93.
- Miller J.R., Ament J.M., and Schmitz O.J. 2014. Fear on the move: predator hunting mode predicts variation in prey mortality and plasticity in prey spatial response. *J. Anim. Ecol.* 83:214–222.
- Millspaugh J.J. 2001. A manual for wildlife radio tagging. *Auk* 118:812.
- Mitchell M.S. and Powell R.A. 2004. A mechanistic home range model for optimal use of spatially distributed resources. *Ecol. Modell.* 177:209–232.
- Mitchinson B., Martin C.J., Grant R.A., and Prescott T.J. 2007. Feedback control in active sensing: rat exploratory whisking is modulated by environmental contact. *Proc. Biol. Sci.* 274:1035–1041.
- Molnár P.K., Kutz S.J., Hoar B.M., and Dobson A.P. 2013. Metabolic approaches to understanding climate change impacts on seasonal host-macroparasite dynamics. *Ecol. Lett.* 16:9–21.
- Mouquet N. and Loreau M. 2003. Community patterns in source-sink metacommunities. *Am. Nat.* 162:544–557.
- Muenchen R.A. 2017. The popularity of data science software. <http://r4stats.com/articles/popularity/>, accessed: 2018-11-28.
- Münkemüller T., de Bello F., Meynard C.N., Gravel D., Lavergne S., Mouillot D., Mouquet N., and Thuiller W. 2011. From diversity indices to community assembly processes: a test with simulated data. *Ecography* 35:468–480.
- Nathan R., Getz W.M., Revilla E., Holyoak M., Kadmon R., Saltz D., and Smouse P.E. 2008a. A movement ecology paradigm for unifying organismal movement research. *Proc. Natl. Acad. Sci. U. S. A.* 105:19052–19059.

- Nathan R., Schurr F.M., Spiegel O., Steinitz O., Trakhtenbrot A., and Tsoar A. 2008b. Mechanisms of long-distance seed dispersal. *Trends Ecol. Evol.* 23:638–647.
- Nee S., Hassell M.P., and May R.M. 1997. Two-species metapopulation models. *In* *Metapopulation Biology* (I. Hanski and M.E. Gilpin, eds.), pages 123–147, Academic Press, San Diego.
- Neutel A.M., Heesterbeek J.A., and de Ruiter P.C. 2002. Stability in real food webs: weak links in long loops. *Science* 296:1120–1123.
- Norberg J., Urban M.C., Vellend M., Klausmeier C.A., and Loeuille N. 2012. Eco-evolutionary responses of biodiversity to climate change. *Nat. Clim. Chang.* 2:747–751.
- Nurulain S., Rum M., and Ismail M.A. 2017. Metacognitive support accelerates computer assisted learning for novice programmers. *Journal of Educational Technology & Society* 20:170–181.
- Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., and Wagner H. 2018. vegan: community ecology package. R package version 2.5-3.
- Onwuegbuzie A.J. and Wilson V.A. 2003. Statistics anxiety: Nature, etiology, antecedents, effects, and treatments—a comprehensive review of the literature. *Teaching in Higher Education* 8:195–209.
- Orians G.H. and Wittenberger J.F. 1991. Spatial and temporal scales in habitat selection. *Am. Nat.* 137:S29–S49.
- Otto S.B., Rall B.C., and Brose U. 2007. Allometric degree distributions facilitate food-web stability. *Nature* 450:1226.
- Ouborg N.J., Piquot Y., and Van Groenendael J.M. 1999. Population genetics, molecular markers and the study of dispersal in plants. *J. Ecol.* 87:551–568.
- Ovaskainen O., Tikhonov G., Norberg A., Guillaume Blanchet F., Duan L., Dunson D., Roslin T., and Abrego N. 2017. How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecol. Lett.* 20:561–576.
- Paas F.G.W.C., Renkl A., and Sweller J. 2004. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science* 32:1–8.
- Paas F.G.W.C. and Van Merriënboer J.J.G. 1994. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *J. Educ. Psychol.* 86:122–133.

- Paine R.T. 1966. Food web complexity and species diversity. *Am. Nat.* 100:65–75.
- Palumbi S.R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. *Ecol. Appl.* 13:146–158.
- Pandit S.N., Kolasa J., and Cottenie K. 2009. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* 90:2253–2262.
- Parmesan C., Gaines S., Gonzalez L., Kaufman D.M., Kingsolver J., Townsend Peterson A., and Sagarin R. 2005. Empirical perspectives on species borders: From traditional biogeography to global change. *Oikos* 108:58–75.
- Pascual M., Dunne J.A., *et al.* 2006. *Ecological networks: linking structure to dynamics in food webs.* Oxford University Press.
- Patterson B.D. and Atmar W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biol. J. Linnean. Soc.* 28:65–82.
- Pawar S., Dell A.I., and Savage V.M. 2012. Dimensionality of consumer search space drives trophic interaction strengths. *Nature* 486:485–489.
- Pedersen E.J. and Guichard F. 2016. At what scales does aggregated dispersal lead to coexistence? *Oikos* 125:1677–1687.
- Persson L. 1985. Asymmetrical competition: are larger animals competitively superior? *Am. Nat.* 126:261–266.
- Petermann J.S., Farjalla V.F., Jocque M., Kratina P., MacDonald A.A.M., Marino N.A., de Omena P.M., Piccoli G.C., Richardson B.A., Richardson M.J., *et al.* 2015. Dominant predators mediate the impact of habitat size on trophic structure in bromeliad invertebrate communities. *Ecology* 96:428–439.
- Pillai P., Gonzalez A., and Loreau M. 2011. Metacommunity theory explains the emergence of food web complexity. *Proc. Natl. Acad. Sci. U. S. A.* 108:19293–19298.
- Pillai P., Loreau M., and Gonzalez A. 2010. A patch-dynamic framework for food web metacommunities. *Theor. Ecol.* 3:223–237.
- Pires A.P., Marino N.A., Srivastava D.S., and Farjalla V.F. 2016. Predicted rainfall changes disrupt trophic interactions in a tropical aquatic ecosystem. *Ecology* 97:2750–2759.
- Poff N.L., Olden J.D., Vieira N.K., Finn D.S., Simmons M.P., and Kondratieff B.C. 2006. Functional trait niches of north american lotic insects: traits-based ecological applications in light of phylogenetic relationships. *J. N. Am. Benthol. Soc.* 25:730–755.

- Poisot T., Canard E., Mouillot D., Mouquet N., and Gravel D. 2012. The dissimilarity of species interaction networks. *Ecol. Lett.* 15:1353–1361.
- Poisot T., Gueveneux-Julien C., Fortin M.J., Gravel D., and Legendre P. 2016. Hosts, parasites, and their interactions respond to different climatic variables. *Glob. Ecol. Biogeogr.* .
- Polis G.A., Anderson W.B., and Holt R.D. 1997. Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. *Annu. Rev. Ecol. Syst.* 28:289–316.
- Polis G.A., Holt R.D., Menge B.A., and Winemiller K.O. 1996. Time, space, and life history: influences on food webs. *In Food Webs*, pages 435–460.
- Pontarp M., Brännström Å., and Petchey O.L. 2019. Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and approximate bayesian computation (ABC). *Methods Ecol. Evol.* .
- Pyke G.H. 1984. Optimal foraging theory: a critical review. *Annu. Rev. Ecol. Syst.* 15:523–575.
- Rahel F.J. and Nutzman J.W. 1994. Foraging in a lethal environment: fish predation in hypoxic waters of a stratified lake. *Ecology* 75:1246–1253.
- Rall B.C., Brose U., Hartvig M., Kalinkat G., Schwarzmüller F., Vucic-Pestic O., and Petchey O.L. 2012. Universal temperature and body-mass scaling of feeding rates. *Phil. Trans. R. Soc. B* 367:2923–2934.
- Real L.A. 1977. The kinetics of functional response. *Am. Nat.* 111:289–300.
- Renkl A. 2005. The worked examples principle in multimedia learning. *In The Cambridge Handbook of Multimedia Learning*, pages 391–412, Cambridge University Press, New York.
- Resetarits Jr W.J. 2005. Habitat selection behaviour links local and regional scales in aquatic systems. *Ecol. Lett.* 8:480–486.
- Resetarits Jr W.J. and Silberbush A. 2015. Local contagion and regional compression: habitat selection drives spatially explicit, multiscale dynamics of colonisation in experimental metacommunities. *Ecol. Lett.* .
- Robledo-Arnuncio J.J. and García C. 2007. Estimation of the seed dispersal kernel from exact identification of source plants. *Mol. Ecol.* 16:5098–5109.
- Rohani P., May R.M., and Hassell M.P. 1996. Metapopulation and equilibrium stability: the effects of spatial structure. *J. Theor. Biol.* 181:97–109.
- Rohrer D. 2015. Student instruction should be distributed over long time periods. *Educ. Psychol. Rev.* 27:635–643.

- Romero G.Q., Piccoli G.C., de Omena P.M., and Gonçalves-Souza T. 2016. Food web structure shaped by habitat size and climate across a latitudinal gradient. *Ecology* 97:2705–2715.
- Ronce O. and Clobert J. 2012. Dispersal syndromes. *Dispersal ecology and evolution* .
- Rooney N. and McCann K.S. 2012. Integrating food web diversity, structure and stability. *Trends Ecol. Evol.* 27:40–46.
- Rooney N., McCann K.S., and Moore J.C. 2008. A landscape theory for food web architecture. *Ecol. Lett.* 11:867–881.
- Rosenzweig M.L. and MacArthur R.H. 1963. Graphical representation and stability conditions of predator-prey interactions. *Am. Nat.* 97:209–223.
- Ryan R.M. and Deci E.L. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.* 25:54–67.
- Ryberg W.A. and Chase J.M. 2007. Predator-dependent species-area relationships. *Am. Nat.* 170:636–642.
- Scarano F.R. 2002. Structure, function and floristic relationships of plant communities in stressful habitats marginal to the brazilian atlantic rainforest. *Ann. Bot.* 90:517–524.
- Schimper A.F.W., Balfour I.B., Fisher W.R., and Groom P. 1903. *Plant-geography upon a physiological basis*. Oxford: Clarendon Press.
- Sexton J.P., McIntyre P.J., Angert A.L., and Rice K.J. 2009. Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.* 40:415–436.
- Shanks A.L., Grantham B.A., and Carr M.H. 2003. Propagule dispersal distance and the size and spacing of marine reserves. *Ecol. Appl.* 13:S159–S169.
- Shmida A. and Wilson M.V. 1985. Biological determinants of species diversity. *J. Biogeogr.* 12:1–20.
- Simberloff D. 2009. The role of propagule pressure in biological invasions. *Annu. Rev. Ecol. Evol. Syst.* 40:81–102.
- Simon, Fincher S., Robins A., Baker B., Box I., Cutts Q., Raadt M.d., Haden P., Hamer J., Hamilton M., Lister R., Petre M., Sutton K., Tolhurst D., and Tutty J. 2006. Predictors of success in a first programming course. *In Proceedings of the 8th Australasian Conference on Computing Education*, vol. 52, pages 189–196.
- Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236:787–792.

- Sol D., Bacher S., Reader S.M., and Lefebvre L. 2008. Brain size predicts the success of mammal species introduced into novel environments. *Am. Nat.* 172 Suppl 1:S63–71.
- Speakman J.R. 2005. Body size, energy metabolism and lifespan. *J. Exp. Biol.* 208:1717–1730.
- Srivastava D.S. 2006. Habitat structure, trophic structure and ecosystem function: interactive effects in a bromeliad–insect community. *Oecol.* 149:493–504.
- Srivastava D.S., Kolasa J., Bengtsson J., Gonzalez A., Lawler S.P., Miller T.E., Munguia P., Romanuk T., Schneider D.C., and Trzcinski M.K. 2004. Are natural microcosms useful model systems for ecology? *Trends Ecol. Evol.* 19:379–384.
- Srivastava D.S., Trzcinski M., Richardson B., and Gilbert B. 2008. Why are predators more sensitive to habitat size than their prey? insights from bromeliad insect food webs. *Am. Nat.* 172:761–771.
- Staddon P., Lindo Z., Crittenden P.D., Gilbert F., and Gonzalez A. 2010. Connectivity, non-random extinction and ecosystem function in experimental metacommunities. *Ecol. Lett.* 13:543–552.
- Stansell Z., Hyma K., Fresnedo-Ramírez J., Sun Q., Mitchell S., Björkman T., and Hua J. 2018. Genotyping-by-sequencing of brassica oleracea vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Horticulture Research* 5.
- Stefan M.I., Gutlerner J.L., Born R.T., and Springer M. 2015. The quantitative methods boot camp: teaching quantitative thinking and computing skills to graduate students in the life sciences. *PLoS Comput. Biol.* 11:e1004208.
- Sterner R.W. and Elser J.J. 2002. *Ecological stoichiometry: the biology of elements from molecules to the biosphere.* Princeton University Press.
- Stevens V.M., Whitmee S., Galliard L., Clobert J., and others. 2014. A comparative analysis of dispersal syndromes in terrestrial and semi-terrestrial animals. *Ecology* .
- Stier A.C., Hein A.M., Parravicini V., and Kulbicki M. 2014. Larval dispersal drives trophic structure across pacific coral reefs. *Nat. Commun.* 5:5575.
- Stone G., Florez-Gonzalez L., and Katona S. 1990. Whale migration record. *Nature* 346:705–705.
- Stone L. and Roberts A. 1990. The checkerboard score and species distributions. *Oecol.* 85:74–79.

- Sukumaran J., Economo E.P., and Lacey Knowles L. 2016. Machine learning biogeographic processes from biotic patterns: A new Trait-Dependent dispersal and diversification model with model choice by Simulation-Trained discriminant analysis. *Syst. Biol.* 65:525–545.
- Sweller J. 1993. Some cognitive processes and their consequences for the organisation and presentation of information. *Aust. J. Psychol.* 45:1–8.
- Sweller J. and Chandler P. 1994. Why some material is difficult to learn. *Cogn. Instr.* 12:185–233.
- Sweller J., Chandler P., Tierney P., and Cooper M. 1990. Cognitive load as a factor in the structuring of technical material. *J. Exp. Psychol. Gen.* 119:176–192.
- Team R.C. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Thompson P.L. and Gonzalez A. 2017. Dispersal governs the reorganization of ecological networks under environmental change. *Nat Ecol Evol* 1:162.
- Thompson P.L., Rayfield B., and Gonzalez A. 2017. Loss of habitat and connectivity erodes species diversity, ecosystem functioning, and stability in metacommunity networks. *Ecography* 40:98–108.
- Tilman D. 1982. Resource competition and community structure. *Monogr. Popul. Biol.* 17:1–296.
- Touchon J.C. and McCoy M.W. 2016. The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* 7:e01394.
- Trebilco R., Baum J.K., Salomon A.K., and Dulvy N.K. 2013. Ecosystem ecology: size-based constraints on the pyramids of life. *Trends Ecol. Evol.* 28:423–431.
- Treml E.A., Roberts J.J., Chao Y., Halpin P.N., Possingham H.P., and Riginos C. 2012. Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integr. Comp. Biol.* 52:525–537.
- Tunney T.D., McCann K.S., Lester N.P., and Shuter B.J. 2012. Food web expansion and contraction in response to changing environmental conditions. *Nat. Commun.* 3:1105.
- Tylianakis J.M. and Morris R.J. 2017. Ecological networks across environmental gradients. *Annu. Rev. Ecol. Evol. Syst.* 48.
- Vale R.D., DeRisi J., Phillips R., Mullins R.D., Waterman C., and Mitchison T.J. 2012. Graduate education. interdisciplinary graduate training in teaching labs. *Science* 338:1542–1543.

- Van de Meutter F., De Meester L., and Stoks R. 2007. Metacommunity structure of pond macroinvertebrates: effects of dispersal mode and generation time. *Ecology* 88:1687–1695.
- Van Den Elzen C.L., Kleynhans E.J., and Otto S.P. 2017. Asymmetric competition impacts evolutionary rescue in a changing environment. *Proc. R. Soc. B* 284:20170374.
- Van der Graaf A.J. 2006. Geese on a green wave: flexible migrants in a changing world.
- Vanschoenwinkel B., Gielen S., Seaman M., and Brendonck L. 2008. Any way the wind blows - frequent wind dispersal drives species sorting in ephemeral aquatic communities. *Oikos* 117:125–134.
- Venables W.N. and Ripley B.D. 2002. *Modern Applied Statistics with S*. Springer, New York.
- Vickers N.J. 2000. Mechanisms of animal navigation in odor plumes. *Biol. Bull.* 198:203–212.
- Vincze O. 2016. Light enough to travel or wise enough to stay? brain size evolution and migratory behavior in birds. *Evolution* 70:2123–2133.
- Ward C.L., McCann K.S., and Rooney N. 2015. HSS revisited: multi-channel processes mediate trophic control across a productivity gradient. *Ecol. Lett.*
- Weber M.J., Dettmers J.M., Wahl D.H., and Czesny S.J. 2010. Effects of predator–prey interactions and benthic habitat complexity on selectivity of a foraging generalist. *T. Am. Fish. Soc.* 139:1004–1013.
- Webster M.S., Marra P.P., Haig S.M., Bensch S., and Holmes R.T. 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol. Evol.* 17:76–83.
- Weersing K. and Toonen R.J. 2009. Population genetics, larval dispersal, and connectivity in marine systems. *Mar. Ecol. Prog. Ser.* 393:1–12.
- Wilson B.C. and Shrock S. 2001. Contributing to success in an introductory computer science course. *In* Proceedings of the thirty-second SIGCSE technical symposium on Computer Science Education - SIGCSE '01.
- Wilson G. 2018. *Teaching Tech Together*. Lulu.com.
- Wilson G.A. and Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191.
- Winder M., Reuter J.E., and Schladow S.G. 2008. Lake warming favours small-sized planktonic diatom species. *Proceedings of the Royal Society B: Biological Sciences* 276:427–435.

- Winemiller K.O., Pianka E.R., Vitt L.J., and Joern A. 2001. Food web laws or niche theory? six independent empirical tests. *Am. Nat.* 158:193–199.
- Woodward G., Ebenman B., Emmerson M., Montoya J.M., Olesen J.M., Valido A., and Warren P.H. 2005. Body size in ecological networks. *Trends Ecol. Evol.* 20:402–409.
- Wootton J.T. 1997. Estimates and tests of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds. *Ecol. Monographs* 67:45–64.
- Yodzis P. and Innes S. 1992. Body size and consumer-resource dynamics. *Am. Nat.* 139:1151–1175.

SUPPLEMENTARY INFORMATION TO CHAPTER 1 - TOWARDS A  
MULTI-TROPHIC EXTENSION OF METACOMMUNITY ECOLOGY.<sup>2</sup>

---

---

<sup>2</sup>Previously published as Guzman L.M., Germain R.M., Forbes C., Straus S., O'Connor M.I., Gravel D., Srivastava D.S., and Thompson P.L. 2019. Towards a multi-trophic extension of metacommunity ecology. *Ecology letters* 22

## A.1 ABSTRACT

Metacommunity theory provides an understanding of how spatial processes determine the structure and function of communities at local and regional scales. Although metacommunity theory has considered trophic dynamics in the past, it has been performed idiosyncratically with a wide selection of possible dynamics. Trophic metacommunity theory needs a synthesis of few influential axis to simplify future predictions and tests. We propose an extension of metacommunity ecology that addresses these shortcomings by incorporating variability among trophic levels in "spatial use properties". We define "spatial use properties" as a set of traits (dispersal, migration, foraging, and spatial information processing) that set the spatial and temporal scales of organismal movement, and thus scales of interspecific interactions. Progress towards a synthetic predictive framework can be made by (i) documenting patterns of spatial use properties in natural food webs, and (ii) using theory and experiments to test how trophic structure in spatial use properties affects metacommunity dynamics.

## A.2 INTRODUCTION

Metacommunity theory formalizes the role that dispersal plays in determining the diversity, stability, and function of ecological communities at local and regional scales (Leibold *et al.*, 2004; Holyoak *et al.*, 2005). This rich body of theory has allowed ecologists to understand that ecological dynamics observed at the scale of local habitat patches are, in part, determined by dynamics in other habitat patches via the exchange of dispersing individuals. To date, metacommunity ecology has been most successful at providing a theoretical predictive framework for competitive metacommunities (Holyoak *et al.*, 2005). However, we still lack a cohesive framework for trophic metacommunities (Leibold and

Chase, 2017). The need for a general theory of trophic metacommunity arises (i) when we need to predict food web properties, which are incompatible with a competitive framework, and (ii) when interacting species use space at different scales; for example, when a predator population interacts with multiple smaller-scale prey populations.

A growing effort has been dedicated to exploring the consequences of trophic interactions in metacommunities (Holt, 2002; Pillai *et al.*, 2011; Gravel *et al.*, 2011; Haegeman and Loreau, 2014; Beger *et al.*, 2010; Trembl *et al.*, 2012). Despite recent empirical and theoretical advances that have laid a solid foundation for a synthetic theory of trophic metacommunities, our understanding remains fragmented due to the diversity of response variables (e.g., diversity, stability, network structure, energy flow) and representations of spatial constraints (eg., perception of scale, types of movement). Traditional metacommunity theory focuses on only one type of movement: dispersal. Dispersal is often related to reproduction (e.g., seed, larvae, and gamete dispersal, or dispersal in search of mates), and therefore relates to only one particular component of life-history. A recent review of metacommunity ecology suggested that future development of this theory must allow species to vary in their abilities to experience the spatial environment (Leibold and Chase, 2017). As a consequence, we suggest that it is time to rebuild trophic metacommunity theory, using spatial processes as pillars of a more cohesive theory for metacommunity dynamics. We focus on five characteristics of "spatial use properties" that we suggest should be at the center of a coherent and broad theory of trophic metacommunities. We define "spatial use properties" as population-level properties that reflect how species use space, and include three forms of movement relevant to trophic metacommunities — dispersal, migration, and foraging (Gounand *et al.*, 2017). We emphasize that species vary in their responses to the environment and to each other, that movement

is not just about dispersal, but an array of processes that each have their own consequences for population dynamics and, we highlight that differences in the way species use space — a dynamic critical to metacommunity dynamics — is due to differences in these spatial use properties.

We propose a framework as a first step to bridge the rapidly advancing fields of spatially-structured food web ecology, movement ecology and metacommunity ecology. This framework: (i) builds on competitive metacommunity theory to make it applicable to trophic dynamics, and (ii) explicitly considers a set of 5 spatial use properties relevant to the spatial and temporal dimensions of trophic interactions. We emphasize the distinction between the three forms of movement — dispersal, migration and, foraging — because they occur at different stages of an organism's life-cycle, they couple different habitat types (e.g., different nearby habitat patches vs. summer and winter habitats), they are initiated by different environmental cues, and they generally occur over different spatial and temporal scales. Therefore, we expect each form of movement to differ in their consequences for metacommunity dynamics. Future progress in trophic metacommunity ecology can be made by documenting the distribution and variation of these five spatial use properties within and among food webs to generate empirical and theoretical predictions for how patterns in spatial use properties within a food web can affect metacommunity dynamics the diversity and structure of food webs at local and regional scales. We also outline empirical and theoretical avenues to test our predicted consequences of spatial use properties in trophic metacommunities.

### A.3 PAST AND PRESENT METACOMMUNITY THEORY

#### A.3.1 *Recent advances and challenges in trophic metacommunity research*

While the theory for competitive metacommunities offers clear predictions, trophic metacommunity theory is remarkable in the diversity of topics explored despite lacking an overarching organizational framework. The first models were inspired by Huffaker's 1958 famous experiment exploring the population dynamics of herbivorous and predatory mites in an experimental metacommunity. Seminal metapopulation models by Holt and Hoopes (2005), Hanski (1999), and others investigated how spatial predator-prey dynamics can contribute to regional coexistence. For example, predators may stabilize prey populations that would otherwise overexploit their resources in the absence of predators (Holt, 2002). The spatial nature of food webs has also been considered previously (Holt, 2007). A greater geographic range of higher trophic level populations was noted by Elton (1966) and its implications for the spatial scale of communities by Holt (1996) and Polis *et al.* (1996). Species (and resources) moving on different scales was recognized to result in spatial subsidies between otherwise seemingly discrete food webs (Polis *et al.*, 1997). Despite these early advances, the effects of spatial processes on food webs dynamics has not been explored in a metacommunity context though they are becoming increasingly apparent (Ward *et al.*, 2015). Metacommunity configurations can determine whether dispersal stabilizes or destabilizes predator-prey dynamics (Jansen, 2001; McCann *et al.*, 2005; Amarasekare, 2008; Gravel *et al.*, 2016b) and this understanding has pushed food web models towards a more general patch dynamics approach of predator and prey assembly. In an effort to map different metacommunity paradigms to food webs, Baiser *et al.* (2012) found that pitcher plant inquiline community structure is best explained by the species-sorting archetype (because of co-variation in response to the environment) and patch dynamics (because of a predominance of local inter-

actions). Other studies have used the source-sink framework to investigate the maintenance of food web structure, not only directly through the dispersal of individuals to poor quality patches, but also indirectly via the spatial exchange of nutrients and energy (Gravel *et al.*, 2010b,a). Such exchanges were further shown to buffer spatial variation in patch productivity, potentially stabilizing trophic metacommunities subject to the paradox of enrichment (Gounand *et al.*, 2014).

Emerging models of trophic metacommunities have demonstrated how trophic interactions can help to understand basic ecological patterns and processes, such as species-area relationships, the co-distribution of predators and prey, range limits, and the restructuring of food webs in response to global change. For example, Holt *et al.* (1999), followed by Ryberg and Chase (2007), proposed that predator species richness should accumulate faster with increasing area than prey species richness. Similarly, Stier *et al.* (2014) showed that predator species richness is less sensitive to isolation than prey species richness. This difference between trophic levels has significant consequences for the interaction network-area relationship (Galiana *et al.*, 2018). This phenomenon results from a sequential assembly of food webs, starting with generalist species at the trophic base of the food web, followed by higher trophic levels and more specialized species (Pillai *et al.*, 2011; Gravel *et al.*, 2011). The co-distribution of predators and prey in trophic metacommunities appears to be key to understanding spatial variation in local network structure (Cazelles *et al.*, 2016). In addition to species turnover among patches, interaction networks also vary in space due to spatial turnover in the realization of potential interactions (Poisot *et al.*, 2012), with cold and hot spots of network beta-diversity (Poisot *et al.*, 2016; Stier *et al.*, 2014).

Although existing metacommunity theory provides a guiding predictive framework for how spatial processes affect the dynamics and structure of

species belonging to the same trophic level (Table A.1A; Calcagno *et al.*, 2006; Mouquet and Loreau, 2003), those predictions are not applicable to the unique response variables that arise when trophic levels interact. When trophic levels interact, the local and regional food webs that are formed can be characterized by network properties, such as connectance (Dunne *et al.*, 2002), diversity at each trophic level (Gamfeldt *et al.*, 2005), and spatial turnover in pairwise interactions in a network (Poisot *et al.*, 2012). Local communities that contain identical numbers of species might differ in their ratio of predators to prey, or in the average number of prey species that predators consume (i.e., linkage density Winemiller *et al.*, 2001; Banasek-Richter *et al.*, 2009). Additionally, because trophic levels are linked through consumption, the flow of energy and matter through local food webs might differ through space (Table A.1B). The greater array of metacommunity properties that characterize multi-trophic systems may reveal spatial processes that are missed by the traditional suite of metacommunity response variables (Pillai *et al.*, 2010) despite being essential to food web stability (Rooney and McCann, 2012; Dunne *et al.*, 2002).

### A.3.2 *Reformulating the assumptions of competitive metacommunity theory*

Leibold *et al.*'s 2004 proposal of four metacommunity paradigms has guided empirical research for much of the past decade (Table A.2A) though, subsequent research demonstrates that communities rarely conform to any single archetype (Cottenie, 2005; Leibold and Loeuille, 2015). Rather, the distribution of organisms across habitat patches can reflect a combination of mechanisms, such as species sorting into some patches and mass effects into others, even within a single species (Thompson *et al.*, 2017). Others have suggested that metacommunity dynamics do not fit into discrete paradigms and instead are better represented as a continuum (Holyoak *et al.*, 2005; Cottenie, 2005; Logue *et al.*, 2011; Thompson *et al.*, 2017). We argue that this continuum perspective

TABLE A.1: Comparison of response variables of competitive metacommunity vs. trophic metacommunity theory

<b>Response class</b>	<b>A. Competitive meta-community ecology</b>	<b>B. Trophic metacom-munity ecology</b>
Structure	Coexistence Diversity Species distribution	Coexistence Diversity Species distribution Species co-distribution Complexity/connectance Trophic length
Dynamics	Trophic modules Stability Synchrony Species turnover	Stability Synchrony Species turnover Interaction turnover Trophic regulation (top-down vs. bottom up)
Energy	Energy flow Productivity	Energy flow Productivity Trophic biomass pyramid

is more critical when we are interested in studying the dynamics of trophic metacommunities, since different trophic levels are more likely to differ in the way species use space than single trophic levels. Therefore, adopting this continuum perspective is necessary to extend metacommunity theory to encompass trophic interactions.

Competitive metacommunity theory assumes that the suitability of habitat patches is determined only by the abiotic environment, and competition can allow species to exclude one another (Leibold *et al.*, 2004). In a trophic metacommunity perspective, patch suitability also depends on the interactions between species, because predators can only persist in patches that have sufficient prey (Gravel *et al.*, 2011). Effectively, the presence of prey increases predator persistence (i.e., a form of "niche construction"), whereas the presence of predators decreases their persistence (i.e., "niche destruction") (Holt, 2009). Because the population dynamics of species linked by trophic interactions are interdependent, patch suitability is dynamic through space and time, even in the absence of abiotic heterogeneity. In this context, distinguishing among patch dynamics and species sorting archetypes becomes difficult (Table A.2A) because species sort into habitat patches based on the presence of predators and prey.

It is clear that these systems where scales of movement and dispersal differ among interacting species violate the assumption inherent in most metacommunity theory: species interacting and coexisting within the metacommunity experience the environment at the same spatial and temporal scales. This assumption is reflected in three ways in competitive metacommunity models: (i) by forcing species to share a common dispersal rate (the proportion of the population that disperses to another population in each generation), (ii) by considering only dispersal and not other forms of movement among populations, such as migration or foraging, and (iii) by assuming species share the

spatial resolution at which they perceive the environment and their ability to act on this information. Variation in dispersal rates has generally been considered in competitive contexts where competition-colonization trade-offs promote coexistence (Cadotte, 2006a). In empirical studies, however, bulk dispersal is the most commonly used method for altering dispersal rates, which prevents detection of interspecific differences in dispersal abilities (Grainger and Gilbert, 2016). Grainger and Gilbert (2016) argue that the heterogeneity that many experimenters choose to remove is necessary to detect metacommunity processes, leading to an inability to robustly test a growing body of theory. Variation in dispersal rates has also been applied to simple predator-prey systems where coexistence is promoted by a higher colonization rate in the prey species (Holt and Hoopes, 2005). In particular, studies of host-parasite interactions revealed that differences in dispersal rate and/or scale could have huge impacts on metacommunity dynamics because parasitoid infection was found to be dependant on host dispersal rate (Holt and Hoopes, 2005) and differences in host vs. parasitoid dispersal rate was found to destabilize dynamics (Rohani *et al.*, 1996). In a two-parasitoid model, the less mobile species was able to persist only in small pockets of high host density, resulting in a competition-colonization trade-off for the competing parasitoids (Nee *et al.*, 1997). However, beyond two-species systems, differences in dispersal between species of different trophic levels are only recently being considered (Haegeman and Loreau, 2014; Pedersen and Guichard, 2016; Thompson and Gonzalez, 2017; Jacquet *et al.*, 2017). Differences in dispersal rates between trophic levels are expected to be much greater than differences within trophic levels because, for example, species at different trophic levels tend to differ in body size and life history (Haskell *et al.*, 2002; McCann *et al.*, 2005). This has consequences for the structure of local and regional trophic networks (Woodward *et al.*, 2005). With larger body size also comes longer lifespans

and greater energetic requirements (Speakman, 2005), and thus the need for other forms of movement, such as foraging and migration, to track daily and seasonal variation in resource supply, respectively. It is for these reasons that we will explore the consequences of differences between interacting species not only in dispersal but also in foraging and migration, and how species-specific differences in these "spatial use properties" affect the structure of food webs.

Metacommunity theory has not yet explicitly integrated the effects of movement governed by perception of the environment on spatial biodiversity processes, even though perception and behaviour are central to the interactions between species (Table A.2B). Existing metacommunity models implicitly assume that demographic consequences of behaviour are captured in local population dynamics. Metacommunity models based on patch dynamics and species sorting assume that the probability that an organism exists in a habitat patch (often equated to a population) is based on its colonization and extinction probabilities (Levins and Culver, 1971; Law and Morton, 1993). In reality, this probability is not fixed but varies with patch quality and the experience of the dispersing organism through prey seeking, predator avoidance, avoidance of competition and selection of suitable habitat, all of which occur at the level of the organism but which have consequences for stability of the entire food web (Kondoh, 2003). For example, predators might leave habitat patches when their prey reach low abundances, buffering prey populations from extinction (Holt, 1984), or prey might avoid dispersing to habitat patches that contain predators, bolstering the prey's regional fitness and allowing predators and prey to coexist regionally (Resetarits, 2005). Similarly, individuals may choose to leave patches with high densities of intra- or interspecific competitors allowing more stable, regional coexistence (Fronhofer *et al.*, 2015). Movement and behaviour of individuals that link patches can affect the population dynamics

and persistence of other species. The latter is traditionally the domain of metacommunity concepts, but a food web perspective highlights that individual decisions about movement in space can couple these population dynamics (McCann *et al.*, 2005).

#### A.4 SPATIAL USE PROPERTIES AND THEIR CONSEQUENCES FOR PAIRWISE TROPHIC INTERACTIONS ACROSS SCALES

Spatial use properties must be considered beyond abiotic niches and dispersal in order to expand metacommunity theory. We propose to incorporate additional ones related to temporal and spatial scales of migration, foraging, and spatial information processing, all of which with very different implications for population dynamics. We propose to differentiate these forms of movement since they happen at different times in an organism's life, they couple different habitats in space and in time, they occur at different temporal frequencies and each may have varied consequences across scales of observation. In this section, we draw on theory from movement ecology and food web ecology to consider explicitly how to integrate the consequences of species' differences in spatial use properties (Figures A.4 and A.4.1), and provide examples where spatial use properties vary with trophic level in natural systems. We also consider how spatial use properties may be estimated in terms of measurable organismal traits.

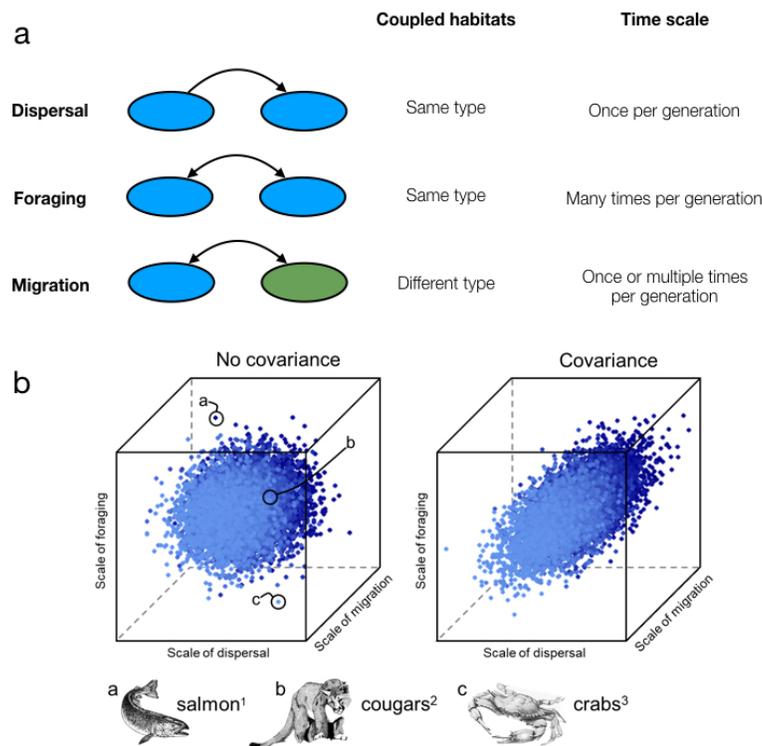
##### A.4.1 *Abiotic niches*

Species' abiotic niches, and their overlap, play a major role in determining the spatial distribution of species in metacommunities. Interactions are only possible between species that overlap in their abiotic niches, except for species that can transiently forage in or disperse through environments that are other-

TABLE A.2: A synthesis of ideas in metacommunity ecology (formalized in Leibold *et al.* (2004)) and food web ecology

Process	Application to competitive communities	Extension to trophic metacommunities	Incompatibility with trophic metacommunities	Spatial use properties relevant to coexistence
<b>A. Metacommunity ecology</b>				
Patch dynamics (Levins and Culver, 1971; Levin and Paine, 1974)	Competition-colonization tradeoffs allow the regional coexistence of species that differ in competitive ability	Prey must disperse more than their predators; predator distributions must be a nested subset of their prey	None	Dispersal; niches (incl. biotic environment)
Species sorting (Tilman, 1982; Leibold, 1998; Chase and Leibold, 2003)	Species differ in which patches are suitable, with suitability defined by abiotic conditions and competitive interactions	The presence of predators and prey (i.e., trophic interactions) also affect patch suitability	Patches must contain prey to be suitable to a predator, thus predators and prey can never completely sort into different patches	Dispersal; Niches (incl. biotic environment)
Mass effects (Shmida and Wilson, 1985)	High dispersal erodes the effects of species sorting such that abundance does not fully reflect patch suitability	The presence of predators and prey also affects patch suitability; predators and prey maintained in neighbouring patches can impact each other	None	Dispersal; Niches (incl. biotic environment)
Neutral interactions (Hubbell, 2001)	Species are competitively equivalent, consuming the same resources	Not extendable	Neutral interactions are not possible between species that do not consume the same resources	Dispersal
<b>B. Food web ecology*</b>				
Process	Application to food web ecology	Application to food web ecology	Extension to trophic metacommunities	Incompatibility with trophic metacommunities
Spatial coupling (McCann <i>et al.</i> , 2005)	Predators forage at larger spatial scales than their prey, linking local food webs together	Promotes food web stability when predators are generalists	Predators forage at larger scales than prey	Foraging scale
Behavioural adaptive foraging (Kondoh, 2003)	Dynamic shifts in foraging strategies to optimize prey capture	Currently not incorporated	Current formulations of metacommunity theory do not allow for changes in interspecific interactions and the response to space due to behaviour	Foraging scale and spatial information processing

FIGURE A.1: (a) Schematic representation of the three forms of movement highlighting the differences of the three forms of movement based on habitat and timescale. The differences between the types of movement for spatial scale are dependent on the organism. (b) A hypothetical distribution of species spatial use properties, where each axis dimension is one of the three movement types, and each point is a species' characteristic movement distance. Dark blue points correspond to those with high levels of migration, and light blue to those with low levels of migration. Data are simulated with multivariate normal distributions, and the empty regions represent ecological or evolutionary constraints. We present two scenarios with no covariance between movement types or covariance between movement types. When there is no covariance between movement types, knowing the scale of one type of movement can not allow for predictions about the others. We highlight three examples of organisms that vary in their scales of movement among the three movement types (1 = Tallman Healey 1994, 2 = Dickson Beier 2006), 3 = Pineda *et al.*, 2007). Three spatial use properties are shown for visualisation, but all five are possible.



wise lethal (Holt, 1993; Mouquet and Loreau, 2003; Rahel and Nutzman, 1994). We expect abiotic niches to correlate between trophic levels when predators are specialists and need to track their prey. Generalist predators are not constrained to their prey distribution and therefore their abiotic niches may not correlate with their prey's abiotic requirements. Incomplete overlap of abiotic niches may allow for spatial refugia from predation or competition, or constrain species ranges for species that depend on other species for persistence (e.g., specialist predators). In addition, we expect that changes in environmental conditions will determine not only a species' spatial distribution, but also food web responses such as food chain length and the shape of biomass pyramids (Tunney *et al.*, 2012).

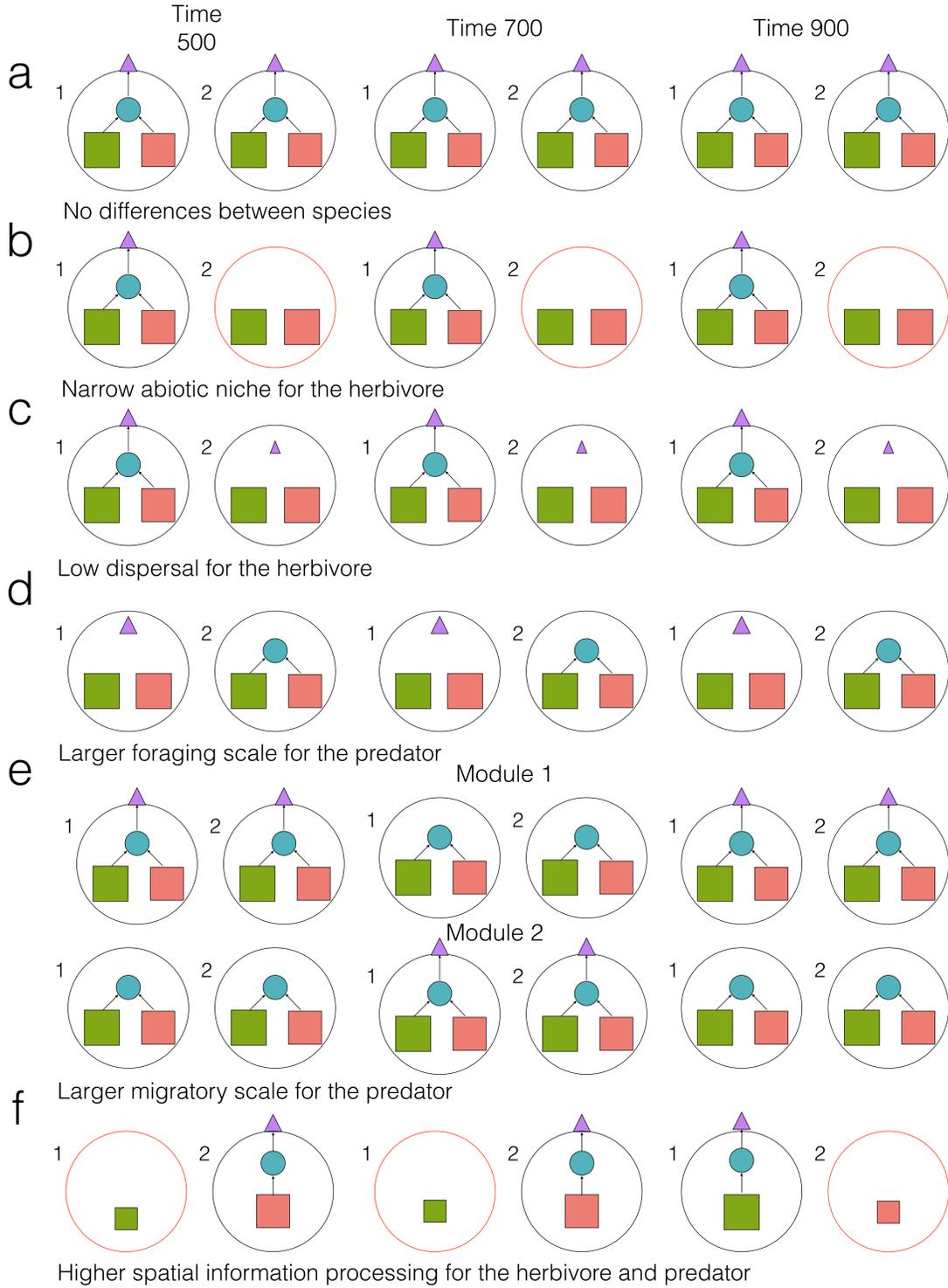
We expect that the consequences of partial overlap in abiotic niches should depend on which trophic level has the narrower niche. If species belonging to lower trophic levels have wider niches, then prey populations might benefit from spatial refugia in environments outside their predators' niche envelope. For example, mosquitofish are more tolerant to warm temperatures compared to their bass predators, which allows them to escape predation by residing in warm habitats (Grigaltchik *et al.*, 2012). Conversely, the abiotic niches of species at higher trophic levels may be limited if their prey have lower abiotic tolerances (e.g. Figure A.4.1b). For example, a butterfly with the physiological tolerance to handle high elevations, can be limited to low elevations by the low elevation range of its host plant (Merrill *et al.*, 2008).

#### A.4.2 *Three forms of movement*

Dispersal, migration, and foraging differ in their frequency and timing within an organism's life cycle; more importantly they have different consequences for the dynamics of trophic metacommunities (Figure A.4a). We first define each movement process, and then unpack their unique consequences

FIGURE A.2: Schematic representation of simulation outcomes for food web structure in two patches and three time points. The food web consists of two primary producers (squares), one herbivore (circle) and one predator (triangle). We show five scenarios of species differences in spatial use properties. (a) The species in this food web do not vary in spatial use properties and therefore the food web does not vary through space or time. (b) The environment varies between patch 1 and patch 2 and because the herbivore has a narrower abiotic niche, it cannot persist in the red patches. The predator consequently also goes extinct in these patches. (c) In this food web, the herbivore has a much lower dispersal than any other species in the food web. Over time, the herbivore is able to reach distant patches. (d) In this food web, the herbivore has a larger foraging scale than any other species in the food web. The herbivore alternates foraging between patches, frequently enough to allow the predator to persist. (e) The top predator migrates in and out of the region, (here presented as a module, each module has two patches) with little effect on the persistence of other species (although abundances may change). (f) In this food web, plants have lower spatial information processing and they are unable to track changing abiotic conditions in patches. The herbivore and the predator track the abundance of the plants.

**Theoretical predictions**



for spatial food web structure, particularly when they differ among trophic levels. The first form of movement, dispersal, occurs once in a lifetime via the movement of individuals to new habitat patches. When an individual disperses, it permanently leaves a patch and enters another patch of the same type of habitat. As such, dispersal can allow a species to colonize a patch that was previously unoccupied or can contribute individuals to existing populations, affecting population size and stability. The second form of movement, migration, is the tracking of seasonally available resources or mates by individuals, and typically occurs annually or once in a lifetime. Unlike dispersal, individuals or their progeny complete migration by returning to their original habitat type; migration does not act to link populations together like dispersal but tends to move single populations through many food webs. Migration is a more predictable event, with proximate cues depending on body condition and climatic and phenological processes. The last form of movement, foraging, is the frequent exploration of space by organisms as they search for resources. Species differ in the spatial scales over which foraging takes place, from very localized for some (e.g., plants through roots) to highly mobile and integrating resources across many habitat patches in a region (Dobson, 2009). Foraging behaviour is highly variable, with decisions depending on the availability and preference of different prey types as well as predator risk (Figure A.4a). We provide the following example to clarify the differences between the types of movement: Dragonflies forage when they are larvae within ponds. Although they are sit and wait predators, they can still forage within ponds to find higher abundances of prey (Johansson, 1991). Then they undergo metamorphosis, an ontogenetic habitat shift. We categorize this ontogenetic habitat shift as the first part of a migration, as aquatic habitats and terrestrial habitats are different types of habitats, and terrestrial adults will eventually return to ponds to oviposit, completing the cycle between

habitats. Finally, adult dragonflies may stay in their natal pond, or disperse to a different pond (McCauley, 2007). Here the movement as adults between ponds is a dispersal event since they are moving between habitats of the same type. While each of these three forms of movement may occur over distinct spatial and temporal scales, each form of movement will likely have different consequences for food web structure and stability across multiple scales of space and time. For example, at the local scale of a marine rocky reef system, sea urchins may be extirpated by foraging sea otters, however, urchins may recover locally due to the dispersal of urchin gametes from distant populations or alternatively, by the migration of transient orcas that eat sea otters. Here, multiple types of movement, each primarily occurring over different scales of space and time and by organisms at different trophic levels, interact to produce dynamics at a single scale that could not be fully understood without consideration of each simultaneously. We consider dispersal, migration and foraging to be different types of movement (i) because they are not necessarily correlated, and therefore, it is difficult to infer anything about the scale or dynamic consequences of these processes from knowledge of another of these processes; (ii) they have dramatically different effects on the trophic dynamics of a metacommunity since they occur at different stages in an organism's life cycle and because they couple different habitats. These three spatial use properties have recently been highlighted as key to understanding spatial flows of energy in meta-ecosystems (Gounand *et al.*, 2017), but their effects on metacommunity dynamics and coexistence are not well understood.

CONSEQUENCES OF DISPERSAL — Differences in dispersal rates among trophic levels may stabilize population dynamics and lead to more complex food webs than would exist in the absence of dispersal (Hauzy *et al.*, 2010). Spatial asymmetries in dispersal among interacting consumer and resource populations can produce distinct spatial distribution of resources. For ex-

ample, when resource populations have limited dispersal and the consumer has global dispersal, the resource density becomes highly variable in space (e.g. Figure A.4.1c). On the other hand, when the consumer has limited dispersal and the prey disperses regionally, the prey is able to persist in subsets of patches that do not contain their predators (de Roos *et al.*, 1998; McCauley *et al.*, 1993; Pedersen and Guichard, 2016). It is expected that rates of dispersal often vary systematically with trophic level, e.g., larval dispersal is greater in predator versus prey species in Pacific reefs (Stier *et al.*, 2014). More generally, we expect that specialist predators require a higher dispersal rate than generalist predators because they need that a particular prey species to be present before it can colonize new habitats (Holyoak *et al.*, 2005).

CONSEQUENCES OF MIGRATION — Species' migration determines the movement of species among habitat type patches, for reproduction and resource consumption. Migration links patches of different habitat type, where species composition is different. In contrast, when individuals disperse and forage, they typically move between habitats of the same type, with similar species composition. In the case of foraging, the movement of individuals between these compositionally-similar habitats may be driven by variation in resource abundance. Classic examples of migration include whales migrating towards the poles in the search of food resources during the summer and migrating towards the tropics during their breeding season in the winter months (Stone *et al.*, 1990), wildebeest following the flush of grass growth across the Serengeti (Holdo *et al.*, 2009), and waterfowl migrating across latitudes to follow the growing season of plants (Van der Graaf, 2006). In addition, some species switch habitats at some point in their life cycle if they require sequential hosts (i.e., parasites) or resources to complete development (Molnár *et al.*, 2013). For example, many insects transition between aquatic and terrestrial life histories, or between belowground and aboveground dwellers as they develop from

larvae to adults. We aggregate migration and habitat switches as both act as spatial subsidies for the receiving food web and link different habitats.

Migration can influence the structure and dynamics of local food webs, involving non-migrating species, by providing a temporal influx of energy, nutrients, and temporary competitors, natural enemies or facilitators. For example, migration may allow the maintenance of populations in low productivity ecosystems such as the Arctic, where large populations of migratory birds disrupt the trophic interaction between terrestrial carnivores and small rodents (Giroux *et al.*, 2012). These spatial subsidies can occur at different trophic levels, for example, a prey species may migrate into a community and provide resources to predators, which can release local prey from risk. Alternatively, predators may migrate which can depress prey populations and have either stabilizing or destabilizing effects (Polis *et al.*, 1997).

CONSEQUENCES OF FORAGING — Foraging movements are within-population movements of one species that can affect the dynamics of other species. Food webs across habitats may be coupled when predators and prey differ in the spatial scales at which they forage (e.g. Figure A.4.1d) (Polis *et al.*, 1997; McCann *et al.*, 2005). For example, if predators forage at broader spatial scales (meaning, over greater areas) than their prey, prey populations in one habitat patch can increase the abundance of predators in an adjacent habitat patch. Because organisms can forage in habitat patches that are outside their abiotic niches (Rahel and Nutzman, 1994), a predicted outcome of linking local food webs via foraging activities is the realization of a greater range of trophic interactions than a given habitat patch would otherwise support. As a consequence of spatial coupling among habitat patches, local dynamics may be decoupled, leading to otherwise unstable food web structures (McCann *et al.*, 2005).

The consequences of foraging on metacommunity structure depends on the mismatch between species at different trophic levels in the use of space; some predators forage over smaller spatial areas than their prey, whereas others forage over larger areas than their prey. We must first consider two general constraints to understand the causes of mismatches in foraging extent among trophic levels: i) consumptive interactions are energetically inefficient, with only 10% energy transfer from food consumed into the bodies of individuals of the consumer population (Trebilco *et al.*, 2013), ii) foraging is also costly because of energetic demands and lethal risks of movement (Anderson and Karasov, 1981; Pyke, 1984). The spatial scale of foraging should therefore reflect the minimum area needed to meet energetic and nutritional requirements given the spatial distribution of prey (DeLong *et al.*, 2014; Laca *et al.*, 2010). Foraging can be highly localized for predators with locally replenishing prey (e.g., web-building spiders with a sit-and-wait strategy), or integrate over much larger spatial scales for predators with scarce, depleted and patchily distributed prey (e.g., predatory birds that must actively seek prey). Mismatches in the spatial scale of foraging occur among trophic levels when their constituent species differ in spatial scales at which energetic/nutritional requirements are met (Higginson and Ruxton, 2015). Such differences in foraging scale between trophic levels will create spatiotemporal dynamics in food webs, and therefore are a critical part of understanding trophic metacommunities.

#### A.4.3 *Spatial information processing*

Traditional metacommunity ecology assumes that dispersal is passive. This assumption becomes problematic when studying food webs, especially for higher trophic levels where movement involves cognitive and information processing systems that allow organisms to actively determine when and where

to move (e.g. Figure A.4.1f). Movement therefore often requires the capacity to receive, store, and process spatially explicit information about the environment; we refer to this capacity as spatial information processing. Spatial information processing can affect any of the three forms of movement - dispersal, migration and foraging- and encompasses 'habitat selectivity', or the degree to which individuals control their movement based on local conditions. However, spatial information processing requires organisms not only to sense their local environment (requiring ability to perceive environment), but also the environment of adjacent patches (requiring spatial memory to integrate perceptions). Organisms must then use this information to aid their navigation and decide where to go (Nathan *et al.*, 2008a). Spatial information processing can have large consequences for the distribution of species in space. For example, colonization rates can depend not only on the perceived quality of one patch, but also that of surrounding suitable patches, leading to spatial contagion (Resetarits and Silberbush, 2015). Mathematical models of animal movements suggest that perception of environmental stimuli affects movement decisions (Hein and McKinley, 2012), and that increased spatial memory optimizes time spent foraging in suitable patches (Fagan *et al.*, 2013).

The ability to process spatial information likely differs between trophic levels. In general, we expect selectivity to increase with trophic level, with plants and microbes being the least selective and top predators being the most selective (but there are also counter examples). Organisms at higher trophic levels tend to have greater cognitive function and brain size, both of which correlate with greater habitat selectivity (Rooney *et al.*, 2008). In particular, actively foraging consumers require more spatial memory to efficiently exploit their environment (Edmunds *et al.*, 2016), and so have larger hippocampal complexes and putative hippocampal homologues both across and within taxa (Krebs *et al.*, 1989; Baird Day *et al.*, 1999). Similarly, animals with larger brains,

for example mammals, have a greater degree of behavioural flexibility and are better able to successfully colonize new environments (Sol *et al.*, 2008). However, increased brain size also comes with increased energetic demands (Fagan *et al.*, 2013) and thus the need for increased foraging.

The scale at which organisms perceive their environment reflects the scale at which they use that environment. Some organisms, such as seabirds, forage across multiple habitat types to meet their nutritional requirements (Orians and Wittenberger, 1991) and so must be able to perceive the patchiness of the landscape and select for certain patches. Habitat specialists may perceive a higher degree of habitat heterogeneity than generalists, resulting in their restriction to small amounts of suitable habitat surrounded by perceived barriers (Holyoak *et al.*, 2005). Larger species have longer viewing distances and therefore a wider scale of perception (Kiltie, 2000), allowing them to move farther and survive for longer in novel environments (Sol *et al.*, 2008). However, faster moving animals also have less accurate perception, potentially explaining changes in visual acuity with trophic level (Chittka *et al.*, 2009). Indeed, the ability to navigate through sensory perception and memory has likely co-evolved with movement capacity, and together these factors influence how and where an individual may move (Fagan *et al.*, 2013; Nathan *et al.*, 2008a).

#### A.5 PREDICTED EFFECTS OF (CO)VARIATION IN SPATIAL USE PROPERTIES ON TROPHIC METACOMMUNITY DYNAMICS

Predicting the consequences of different distributions of spatial use properties will require deeper theoretical investigation than is possible here, but we nonetheless propose a few general patterns as a starting point. As an illustration, we examine the dynamics of a simple food web of four species

(two plants consumed by one herbivore, which itself is preyed upon by one predator) in two patches using the model presented in appendix S2. We parameterized this model with simple scenarios where at least one species in the food web varies in their spatial use properties from the rest of the food web (Figure A.4.1). The same food web will occur in all patches at all timepoints if species do not differ in their spatial use properties, the environment is homogenous and dispersal between patches is null (Figure A.4.1a). Changes in diversity and food web composition through space or time arise with variability in spatial use properties (Figure A.4.1). For example, the herbivore cannot persist in the red patch of Figure A.4.1b, only on the black patch when it has a narrower abiotic niche and species do not disperse between patches. Since the herbivore is permanently absent from the red patch, the predator is also absent because of starvation. Similarly, the herbivore will be absent from the second patch if it has very low dispersal (Figure A.4.1c). In this case however, the predator is only present via dispersal, where the second patch becomes a sink population for the predator, given that there is no prey present (Figure A.4.1c). The predator can persist if it has a larger foraging scale because of the consumption of herbivores on both patches. Even if there are no herbivores on the patch that contains predators, the predator will be able to persist by coupling the two patches and foraging on the second patch (Figure A.4.1d). In the case where the predators migrate in and out of a metacommunity module, they will affect the abundance of the herbivores only when they are present in that module (Figure A.4.1e). Finally, when the plants have lower spatial information processing, they will be slower at tracking changes in the abiotic conditions of patches. This inertia could cascade to other trophic levels if upper trophic levels track their resources more closely than they do their environment (Figure A.4.1f). Overall, we expect increased network diversity, complexity, and stability when trophically-linked species are dissimilar in

their spatial use properties. This should coincide with greater difference in the spatio-temporal dynamics of each species. These results, based on a relatively simple model, show how differences in spatial use properties across trophic levels can impact the dynamics, diversity, and food web structure of trophic metacommunities. Further work is now needed to fully integrate spatial use properties into trophic metacommunity models and theory.

#### A.6 FUTURE DIRECTIONS: BUILDING AND TESTING FUTURE METACOMMUNITY THEORY BASED ON SPATIAL USE PROPERTIES

We have argued that incorporating spatial use properties will provide a deeper understanding of trophic metacommunities; our challenge is now to use this perspective to develop, test, and refine a body of trophic metacommunity theory. To accomplish this goal, efforts are now needed to (i) document these five spatial use properties within food webs, (ii) use meta-analytical approaches to investigate patterns of spatial use properties across scales of space time and organization, within and among food webs, (iii) develop new theory for how the relative scales of spatial use properties across trophic groups affects metacommunity dynamics and their outcomes, and (iv) test whether empirical biodiversity patterns in trophic metacommunities can be explained by the scales of ecological processes related to spatial use properties.

(I) DOCUMENTING SPATIAL USE PROPERTIES WITHIN FOOD WEBS — Before new theory about trophic metacommunity dynamics (goals iii-v) can be tested, we require quantitative measures of spatial use properties (using traits) within food webs. This is a challenge, because spatial use properties themselves are rarely quantified directly in empirical studies. We propose a set of measurable traits that can be used as proxies of spatial use properties (Table

A.3), to quantitatively compare differences in the spatial scales and extents movement among interacting species. A single measurable trait may not be suitable to estimate differences in spatial use properties across all trophic levels ranging from microbes to top predators. Experiments coupled with observations from multiple techniques may be required to estimate spatial use properties for whole food webs. For example, bacterial movement can be studied using microfluidic devices (Englert *et al.*, 2009), insect movement with harmonic radar (Chapman *et al.*, 2011), and mammal movement with radio tags (Millsaugh, 2001).

(II) USING META-ANALYTICAL APPROACHES TO INVESTIGATE PATTERNS OF SPATIAL USE PROPERTIES WITHIN AND AMONG FOOD WEBS — Documenting the scales and mechanisms associated with spatial use properties will provide the empirical evidence needed to answer the question of whether these properties vary systematically within and across food webs, using meta-analytical approaches. It will also allow us to test whether spatial use properties are constrained by physiological, morphological or evolutionary trade-offs. In other words, can we use knowledge of one spatial use property within a food web to infer the structure of another spatial use property in that food web? (Figure A.4b).

The synthesis of metacommunity and spatial food web concepts we have reviewed here implies that within a food web, organisms vary in their spatial use properties and that this variation affects metacommunity dynamics. Species at different trophic levels have very different energetic needs and life history strategies (Trebilco *et al.*, 2013). Furthermore, both trophic level and spatial use properties such as dispersal, migration and foraging scale with body size (Kalinkat *et al.*, 2015; McCann *et al.*, 2005; Hein *et al.*, 2012). However, these scaling relationships have been generated by aggregating species across many food webs and therefore little is known about how spatial use properties

TABLE A.3: Spatial use properties and how they correspond both to measurable traits and parameters in the modelling framework described in the text.

<b>Spatial use properties</b>	<b>Measurable organismal traits</b>	<b>Model parameters</b>
Abiotic niches	Temperature tolerance (Magnuson <i>et al.</i> , 1979; Huey and Kingsolver, 1989) Drought tolerance (Engelbrecht <i>et al.</i> , 2007; Schimper <i>et al.</i> , 1903) Range limits (Ehrlén and Morris, 2015; Parmesan <i>et al.</i> , 2005; Sexton <i>et al.</i> , 2009) Stoichiometric niche (González <i>et al.</i> , 2017; Sterner and Elser, 2002)	Species-specific environmental optima and environmental breadth
Dispersal scale	Maximum dispersal distance (Nathan <i>et al.</i> , 2008b; Levin <i>et al.</i> , 2003; Cain <i>et al.</i> , 2000) Dispersal rate (Hanski, 1991) Number of propagules (Shanks <i>et al.</i> , 2003; Simberloff, 2009) Gene flow (Palumbi, 2003; Slatkin, 1987) Mode of locomotion (Stevens <i>et al.</i> , 2014; Ronce and Clobert, 2012)	Dispersal rate and distance
Migration scale	Migration propensity (Hanski <i>et al.</i> , 2004; Alerstam <i>et al.</i> , 2003) Migration distance (Webster <i>et al.</i> , 2002) Stable isotopic ratios (Hobson, 1999)	Migration rate and distance
Foraging scale	Home range size (Börger <i>et al.</i> , 2008; Mitchell and Powell, 2004) Radio collars for daily movement (Harris <i>et al.</i> , 1990)	The number of patches that each species uses to forage
Spatial information processing	Relative brain size (Fagan <i>et al.</i> , 2013) 2d vs. 3d perception (Pawar <i>et al.</i> , 2012) Sensing appendages (Vickers, 2000; Mitchinson <i>et al.</i> , 2007) Active vs. passive dispersal (Cottenie, 2005; Van de Meutter <i>et al.</i> , 2007)	Changes to movement due to environmental variation

are structured within individual food webs. A related question that could be answered through meta-analysis of is whether trophic structure in spatial use properties varies systematically across different ecosystem types. We might expect such systematic differences between ecosystem types to arise because of differences in the evolutionary histories of their constituent species (e.g., aquatic vs. terrestrial), regional environmental structure (e.g., patchiness), or bioclimatic differences on larger geographic scales (e.g., temperate vs. tropical). Identifying additional scaling relationships, and their causes, will not only allow application of this framework with efficient use of data for parameterizing models, but will also help understand how macro-ecological and physiological constraints may influence spatial processes in metacommunities.

Understanding covariances in spatial use properties across species can reveal biological or evolutionary constraints and tradeoffs of these properties, as well as variation in those constraints among ecosystems (Díaz *et al.*, 2016). For example, relative brain size (a proxy for spatial information processing) is smaller in migratory than in non-migratory birds due to an energetic trade-off between neural tissue volume and migratory flight (Vincze, 2016). Similarly, traits relevant to particular types of movement (foraging, migration, dispersal) might be positively correlated (Bowman *et al.*, 2002) as each has been shown to increase with body size (De Ryck *et al.*, 2012; Kelt and Van Vuren, 1999; Greenleaf *et al.*, 2007; Alerstam *et al.*, 2003; Hirt *et al.*, 2017); for this reason, we refrain from assigning body size as a proxy for any specific spatial use property (Table A.3). A high degree of covariation among spatial use properties might simplify predictions in some ecosystems. Non-random covariation among traits will constrain the range of local food web structures that are possible for theoretical studies (Gravel *et al.*, 2016a).

(III) DEVELOPING NEW THEORY FOR HOW THE RELATIVE SCALES OF SPATIAL USE PROPERTIES ACROSS TROPHIC GROUPS AFFECTS META-

COMMUNITY DYNAMICS AND THEIR OUTCOMES — We can use theoretical models to explore the consequences of different trophic structures in spatial use properties for the stability and network structure of food webs. By constraining this exploration based on documented patterns of spatial use properties (goals i and ii) will allow us to focus on and contrast the predicted outcomes of patterns that are found in specific food webs or ecosystem types. We have outlined how this could be done using a modelling approach that incorporates the five spatial use properties (Figure A.4.1). This constrained exploration will allow us to ask what are the commonalities and differences in how the five spatial use properties affect food web stability? and how does trophic metacommunity structure and persistence respond to environmental change and habitat loss? Given the importance of spatial use properties for the dynamics and stability of trophic metacommunities as we suggest here, we hypothesize then that diversity associated with trophic status in traits related to spatial use properties might be a particularly important dimension of diversity for spatially structured food webs (McCann *et al.*, 2005). For example, McCann *et al.* (2005) showed that when predators forage at larger spatial scales than prey, they can stabilize food webs. Similarly, differences in dispersal between predators and prey can result in stability of the interactions (Pedersen and Guichard, 2016). We suggest that it should be addressed with theory that is guided by observational patterns of spatial use properties (i.e. goals i and ii) and then tested using experiments. If, for example, fish in ponds are observed to forage at larger scales but disperse at smaller scales than invertebrate prey, we can develop models that provide theoretical predictions for how these movement differences affect the spatial distribution of the two trophic levels. We can test these predictions by experimentally manipulating fish foraging and dispersal via movement restriction (i.e., size-specific mesh) and assisted dispersal, respectively.

Movement is a key process that determines how communities respond to environmental change and habitat loss (Loreau *et al.*, 2003; Norberg *et al.*, 2012; Thompson *et al.*, 2017; Grilli *et al.*, 2015). Despite the fact that we know that trophic level is a key predictor of how species will respond to such changes, we have limited theory that links this response to movement within a food web context (Thompson and Gonzalez, 2017). Theoretical models offer the opportunity for developing expectations of how different patterns of spatial use properties affect the response of food webs to different forms of environmental change or habitat loss. This theory is needed for informing and interpreting experiments since the presence of predators has often interfered with our ability for experiments to match theoretical predictions (Grainger and Gilbert, 2016).

(IV) TESTING WHETHER EMPIRICAL BIODIVERSITY PATTERNS IN TROPHIC METACOMMUNITIES CAN BE EXPLAINED BY THE SCALES OF ECOLOGICAL PROCESSES RELATED TO SPATIAL USE PROPERTIES — Inferring the spatial processes that govern the diversity and functioning of communities is a major goal in metacommunity ecology (Leibold and Chase, 2017). Yet, methods for linking patterns of abundance to different metacommunity paradigms (Cottenie, 2005; Ovaskainen *et al.*, 2017) do not have a systematic way of incorporating trophic interactions, nor variation in movement between trophic levels. We demonstrate how our framework can be used to link patterns of abundance to spatial use properties with a food web module from a metacommunity of bromeliad-dwelling invertebrates (Box 1). This example shows how observational data may be coupled with structural equation models to untangle how space affects food web structure. Additional efforts to formalize these links in other systems with existing data would be one way to rapidly advance our empirical understanding of trophic metacommunities.

The questions and avenues of research we highlight are underexplored and promise rich research opportunity. The feasibility of answering these questions will undoubtedly vary among food webs, particularly those for which spatial use properties are difficult to quantify with reasonable certainty. Trophic metacommunities are complex and models will need to deal with the rich natural history that underlies species interactions and movement (such as omnivory, territoriality, ontogenetic niche shifts, non-consumptive effects and cross-ecosystem subsidies). Experiments and observational studies will guide theoretical studies to manage that complexity. Further development of trophic metacommunity theory requires a feedback between empirical observation, theory, and experiments. We believe that this approach offers exciting possibilities and has the potential to guide the development and testing of the next generation of trophic metacommunity theory.

---

**Box 1: Trophic metacommunities in bromeliad-dwelling insect**

---

We use a food web module from water-filled bromeliads in Costa Rica to consider how shifts in the relative abundances of species along a habitat size gradient can be understood in terms of species differences in their spatial use properties (colonization rates, abiotic niches) and susceptibility to predators. *Culex spp.* and *Wyeomyia spp.* mosquitoes are potential competitors and are both preyed upon by *Mecistogaster modesta* damselflies. All three taxa show strong patterns with bromeliad size, with the abundance of *Culex* and *Mecistogaster* increasing with bromeliad size, and *Wyeomyia* decreasing with bromeliad size (Figure A.6a). Bromeliad size affects species in three ways.

**(1) Numerical effects on colonization.** If colonization probability is related to available habitat, as often assumed in competitive metacommunity models, we would expect larger bromeliads to be colonized more frequently than small bromeliads, such that species with small regional populations and thus few colonists (*Mecistogaster*) occur entirely in the large bromeliads whereas species with larger regional populations (*Culex*, *Wyeomyia*) occupy mainly large but also some medium-sized bromeliads. Although such numerical effects explain the distribution of *Culex*, *Mecistogaster* still occurs in larger bromeliads than expected and *Wyeomyia* in smaller bromeliads than expected.

**(2) Abiotic niche differences.** Small bromeliads are at risk of drying out while insects are still aquatic larvae, and this risk is particularly acute for *Mecistogaster*, whose larvae require ca. 9 months to develop. *Culex* and *Wyeomyia* larvae require ca. 3 weeks to develop and have less exposure to drought risk (Figure A.6b). After correcting species abundance for numerical effects on colonization probability, residual *Mecistogaster* abundance is positively related to bromeliad size in a structural equation model — presumably reflecting its greater likelihood of drought exposure at some point during the larval stage. *Wyeomyia* residual abundance is negatively related to bromeliad size, potentially because drought-resistant eggs in this genus (unlike *Culex*) enable it to preferentially colonize small bromeliads.

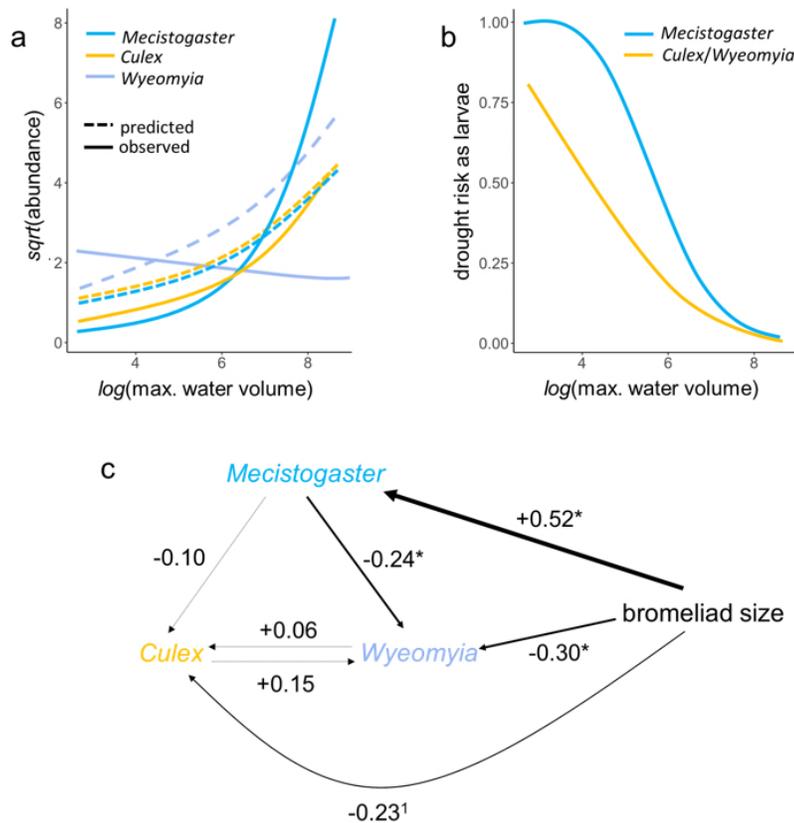
**(3) Trophic interactions.** Finally, bromeliad size may affect species indirectly via predation or competitive interactions. In our structural equation model, *Wyeomyia* occurs in smaller than expected bromeliads because it is negatively affected by its predator, *Mecistogaster*, which in turn occurs disproportionately in large bromeliads. By contrast, *Culex* abundance is unaffected by effects of bromeliad size mediated by *Mecistogaster* (Figure A.6c). This is consistent with the documented ability of *Culex* — but not *Wyeomyia* — to chemically detect *Mecistogaster* and avoid predation through a change in foraging behaviour (Hammill *et al.*, 2015a). This example shows the power of combining statistical analyses with documented differences between species in their response to abiotic stress and predation to understand the distribution of the food web module between habitat patches. Deeper understanding of this module could be achieved by studying how other spatial use properties differ between these species, such as spatial information processing by ovipositing adults, or dispersal (pupation) cues for mosquito larvae.

---

## A.7 CONCLUSION

We argue that metacommunity theory must incorporate trophic interactions to encompass the full range of dynamics that occur in real-world communities. We began by outlining challenges to extending metacommunity ecology beyond competitive systems, with suggestions for how to overcome those challenges by reformulating some basic assumptions. We then proposed that progress towards a trophic metacommunity framework could be achieved by accounting for a wider array of spatial use properties than the traditional metacommunity framework allows. These spatial use properties are (i) abiotic niches, (ii) spatiotemporal scales of dispersal, (iii) scales of migration,

FIGURE A.3: (a) The observed abundance of three genera of insects changes across a gradient in the maximum volume of bromeliads. A null model that assumes no per capita differences in colonisation probability, but purely numerical differences driven by differences in regional abundance ('predicted abundance'), can account for some of the positive effects of bromeliad size on *Culex* and *Mecistogaster*. (b) The *Mecistogaster* damselfly larvae experience greater risk of drought during their 9 month larval phase than the *Culex* and *Wyeomyia* mosquito larvae with a 3 week larval duration. (c) After correcting for the numerical effects of bromeliad size on colonisation rates, residual abundance of species may be related to bromeliad size either directly, for example by drought risk associated with small bromeliads, or indirectly, through effects on competitors and predators. Bromeliad size was corrected for numerical effects on colonisation. Path coefficients are standardised effect sizes from structural equation models described in full in the Supplementary Material; path widths are proportional to absolute standardised effect sizes. Significance of path coefficients: \* $P < 0.05$ ;  $1P < 0.10$ .



(iv) scales of foraging, and (v) spatial information processing. We end by reiterating priority questions to be answered towards a robust trophic metacommunity theory. Answering these questions would allow metacommunity ecology to fulfil its promise as a truly synthetic theory of food web ecology.

## SUPPLEMENTARY INFORMATION TO CHAPTER 2

## B.1 PARTITIONING BETA DIVERSITY

We used the procedure of Baselga (2010), where Sørensen beta diversity for multiple-sites can be expressed as:

$$\beta_{SOR} = \frac{\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_j S_i - S_T \right] + \left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]} \quad (\text{B.1})$$

Where  $S_i$  is the richness in each site,  $b_{ij}$  is the number of species in site  $i$  not in site  $j$  and  $b_{ji}$  is the number of species in site  $j$  not in site  $i$  and  $S_T$  is total richness across all sites. Sørensen dissimilarity ( $\beta_{SOR}$ ) accounts for both species turnover and nestedness. Beta diversity accounting only for pure spatial turnover is ( $\beta_{SIM}$ ):

$$\beta_{SIM} = \frac{\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_j S_i - S_T \right] + \left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right]} \quad (\text{B.2})$$

Therefore we can use Sørensen dissimilarity ( $\beta_{SOR}$ ) and spatial turnover ( $\beta_{SIM}$ ) to calculate the total nestedness of species assemblages ( $\beta_{NES}$ ):

$$\beta_{NES} = \beta_{SOR} - \beta_{SIM} \quad (\text{B.3})$$

Therefore,

$$\beta_{NES} = \frac{\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_j S_i - S_T \right] + \left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]} - \frac{\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_j S_i - S_T \right] + \left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right]} \quad (\text{B.4})$$

## B.2 ENVIRONMENTAL VARIATION BETWEEN SITES

For every bromeliad, we measured a suite of environmental variables to assess the amount and quality of habitat available to the invertebrates: the height (cm) and diameter (cm, measured as the maximum distance between leaf tips) of the plant, maximum water volume (mL, calculated by emptying the plant and calculating how much water the plant could hold before it overflowed), actual water volume (mL), longest leaf length (cm), longest leaf width (cm), number of leaves, canopy cover (% of shaded pixels in photos taken looking directly up from the bromeliad), total detritus (g dry mass), pH, oxygen concentration (% saturation), salinity (ppt), temperature (°C), and turbidity (NTU). Water chemistry and temperature variables were measured using a portable multiparameter waterproof meter in the field as soon as the water was collected from the plant.

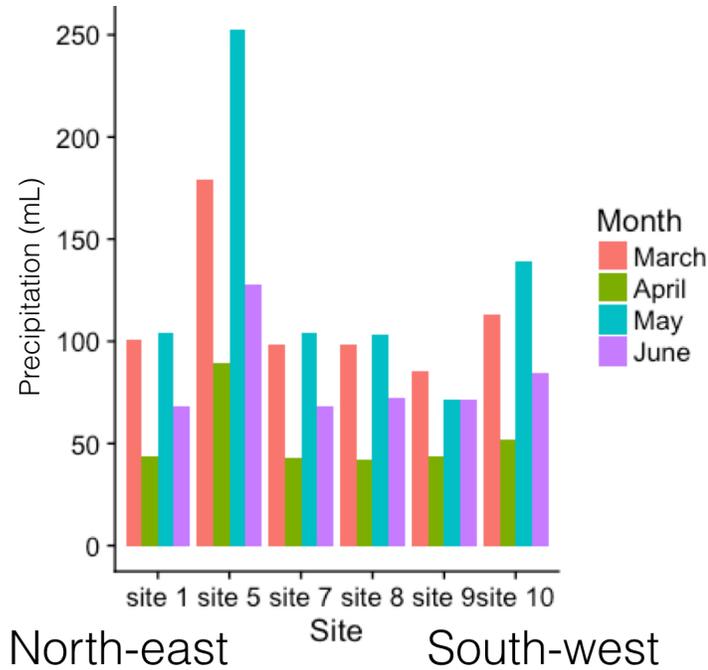
To test for differences between sites in environmental variables, we used a linear model and for oxygen saturation and canopy cover we used a generalized linear model with binomial family between the environmental variable and the site number.

We obtained precipitation information from World Clim for our sites. The finest resolution from World Clim is 30 seconds, or 1 Km<sup>2</sup>. Sites 1-4 of our

TABLE B.1: F value, Chisq value when using a binomial family, and the P value for comparing the environmental variables between sites. All treatment degrees of freedom are 9 and residual degrees of freedom are 90.

Environmental variable	F/Chisq	P value
Diameter	1.501	0.159
Height	0.987	0.455
Oxygen Saturation	3.914 (Chisq)	0.916
Salinity	0.860	0.563
Canopy cover	9.084 (Chisq)	0.429
Chlorophyll	0.441	0.908
Total detritus	1.417	0.192
Turbidity	1.764	0.086
Maximum Volume	2.232	0.0267**
Actual Volume	3.854	0.0003**

FIGURE B.1: Total precipitation (mL) for the closest 1  $Km^2$  of every site using WorldClim data. Sites not shown are within 1  $Km^2$ .



sampling are all within 1  $Km^2$ , therefore we cannot obtain independent precipitation data from World Clim for each of these sites [1].

We also obtained precipitation data from the year we collected from the weather stations located close to the metacommunities where we sampled [2].

We then related the cumulative precipitation in the closest weather station to each of our sites to the actual volume of water present in the bromeliads. For each site we added the precipitation from February 1st to the date the site was sampled

FIGURE B.2: Total precipitation (mL) for the closest weather stations to the transect. This data represents the months of sampling in 2015.

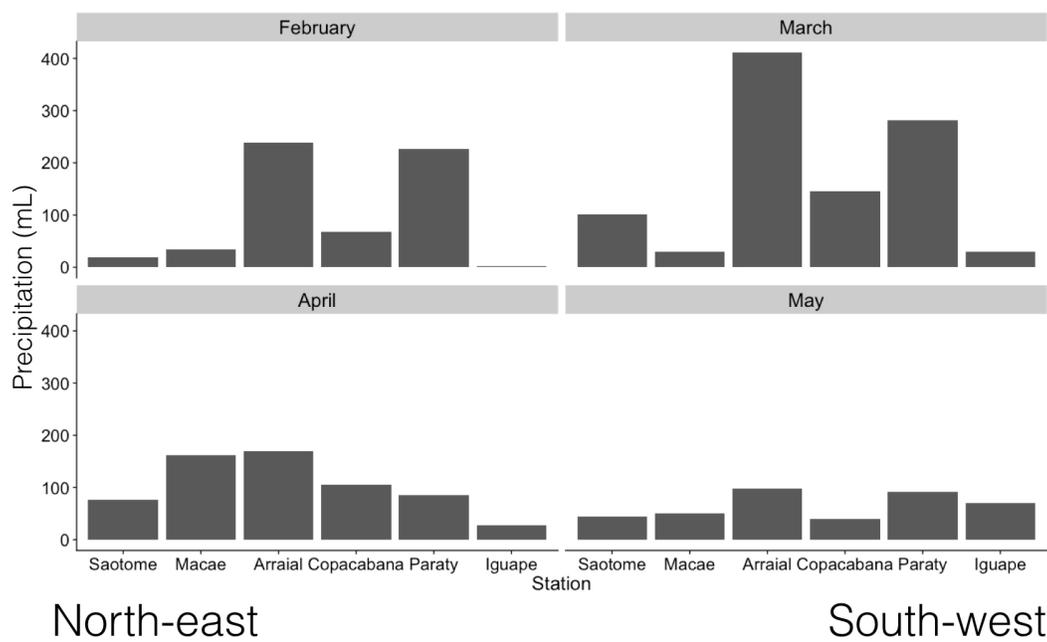
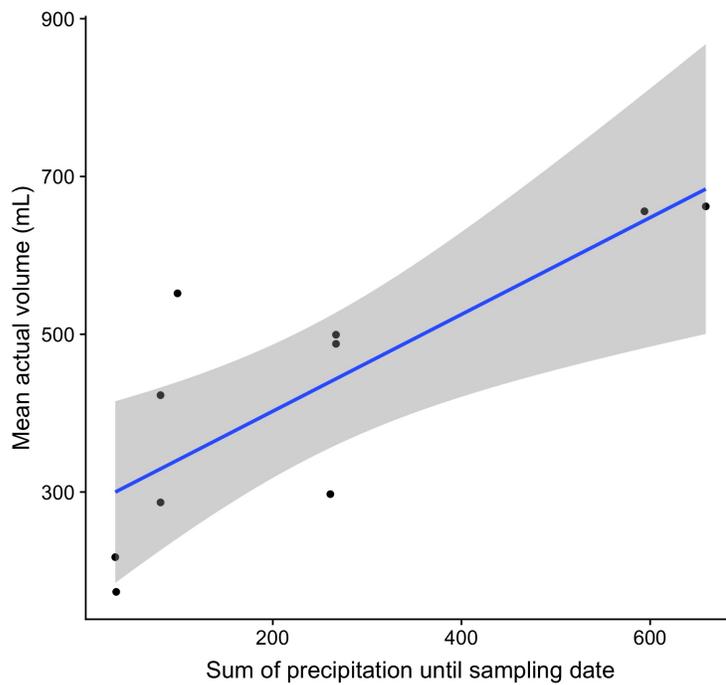


FIGURE B.3: The mean actual water volume in the bromeliads increases with the cumulative precipitation until sampling date (Intercept = 279.65, se = 53.92, slope = 0.613, se = 0.167). For each site we added the precipitation from February 1st to the date the site was sampled. We used the data from the closest weather station to the site.



### B.3 VALIDATION OF MARKOV NETWORK METHOD

We validated the method by confirming it gave the same results as known interaction strengths and could predict trophic interaction strengths in simple bromeliad food webs. We took two different approaches to this confirmation. First, we ran the Markov Network analysis on a three species module from Costa Rica where we compared the outcome of the Markov Network analysis on the distribution of these species in Costa Rica vs. interaction strengths that had been established based on experiments (Hammill *et al.*, 2015b). Second, because we have prior knowledge on the trophic ranks of every genera in the Brazilian dataset, we could test whether the Markov Network method could correctly assign the trophic positions of genera.

#### B.3.1 *Markov Network analysis on a known species module:*

We were able to get similar interaction strengths as those expected based on direct experimentation. Using a three species module prohibits us from accounting for the indirect interactions that these three species may have with the rest of the community. However, from the experimental evidence, these three species have been shown to have strong effects on each other and we had experimental data only on this module.

We used a known species module to test the outcome of the Markov Network analyses where species interactions have been studied in detail using experiments (Hammill *et al.*, 2015a). In this species module, both the prey *Culex* and the predator *Mecistogaster* increase in abundance with bromeliad size. However, wherever the predator is present, *Wyeomyia*'s numbers are greatly reduced. *Culex* is tolerant to predation due to behavioural responses and therefore is commonly found with the predator. Using Markov Network analysis, *Culex* and *Mecistogaster* have a positive interaction strength. On the

TABLE B.2: Interaction strengths of the three species module studied in Hammill *et al.* (2015a)

	<b>Mecistogaster</b>	<b>Culex</b>	<b>Wyeomyia</b>
<b>Mecistogaster</b>	0.000	1.013	-2.268
<b>Culex</b>	1.013	0.000	0.936
<b>Wyeomyia</b>	-2.268	0.936	0.000

other hand, *Wyeomyia* and *Mecistogaster* have a negative interaction strength (Table B.2).

### B.3.2 Markov Network analysis for assigning trophic rank:

For this analysis, we started by assessing the types and strengths of interactions for every combination of genera. Positive interactions mean that species are more likely to co-occur than expected by chance. Negative interactions mean that species are less likely to co-occur than expected by chance. We calculated the number of positive and negative interactions for every species, regardless of the strength across sites. We classified every species as a predator or a prey, to test if known differences in trophic positions have a different preponderance of positive (expected for prey) or negative (expected for predators) interactions. If such a pattern is found it must reflect realized trophic structure in the community since information on trophic position was not included in the estimation of interaction terms. The species or genera known to be predatory are: the damselfly *Leptagrion andromache* nymphs, elephant mosquito larvae *Toxorhynchites*, *Corethrella* midge larvae, horsefly larvae *Tabanidae* and crane fly larvae *Tipulidae* (Figure 2.1). For this analysis, we explained the total number of interactions for each genera as a function of the sign of the interaction (either positive or negative) and the trophic position. Since individual genera are present in different interactions, we included genera identity as a random effect in a generalized mixed effect model. The independent units of

replication are the sites. We used a Poisson error distribution, appropriate for left-skewed count data. The interaction term of this model, between trophic position and the sign of the between-genera interaction, tests if predators and prey differ in the sign of their biological interactions.

The top predator *Leptagrion andromache* dominated negative interactions (Figure B.4a), which is expected since it preys on most species in the community. In general, prey species had more positive interactions and predator species had more negative interactions compared to random expectations ( $\beta = 0.564$ ,  $z$  value = 4.456,  $P$  value =  $8.3 \times 10^{-6}$ , Figure B.4). This result was robust to the matrix permutations of presences ( $P$  value = 0, Permutation results, Figure B.6 - B.8, Table B.5). We therefore conclude that the Markov Network method assess trophic position from observational data.

## B.4 SUPPORTING RESULTS

### B.4.1 *Pairwise multivariate analysis of variance and pairwise Tukey tests for community dispersion*

TABLE B.3: F value,  $R^2$ , p value and adjusted p value for pairwise multivariate analysis of variance (adonis tests) and the difference between the dispersion of communities, lower and upper boundaries of the differences and adjusted p values for pairwise tukey-tests.

Pairs of sites	F	Pairwise multivariate analysis of variance			Pairwise Tukey-test for community dispersion			
		R <sup>2</sup>	p.value	p.adjust	Diff	Lower	Upper	p.adjust
1 vs 2	1.634	0.083	0.205	1.000	0.043	-0.367	0.454	1.000
1 vs 3	1.907	0.096	0.135	1.000	0.151	-0.260	0.562	0.972
1 vs 4	1.892	0.095	0.158	1.000	0.178	-0.232	0.589	0.922

TABLE B.3: F value,  $R^2$ , p value and adjusted p value for pairwise multivariate analysis of variance (adonis tests) and the difference between the dispersion of communities, lower and upper boundaries of the differences and adjusted p values for pairwise tukey-tests.

Pairs of sites	F	Pairwise multivariate analysis of variance			Pairwise Tukey-test for community dispersion			
		R <sup>2</sup>	p.value	p.adjust	Diff	Lower	Upper	p.adjust
1 vs 5	12.887	0.417	0.001	0.045*	-0.159	-0.569	0.252	0.961
1 vs 6	9.186	0.338	0.001	0.045*	0.195	-0.216	0.605	0.873
1 vs 7	4.272	0.192	0.010	0.450	0.103	-0.308	0.513	0.998
1 vs 8	15.549	0.463	0.001	0.045*	-0.163	-0.573	0.248	0.955
1 vs 9	6.718	0.272	0.001	0.045*	0.062	-0.348	0.473	1.000
1 vs 10	11.157	0.383	0.001	0.045*	0.036	-0.374	0.447	1.000
2 vs 3	0.743	0.040	0.614	1.000	0.108	-0.303	0.518	0.997
2 vs 4	2.355	0.116	0.052	1.000	0.135	-0.276	0.545	0.987
2 vs 5	13.273	0.424	0.001	0.045*	-0.202	-0.613	0.208	0.846
2 vs 6	7.746	0.301	0.001	0.045*	0.151	-0.259	0.562	0.971
2 vs 7	3.849	0.176	0.007	0.315	0.059	-0.351	0.470	1.000
2 vs 8	15.564	0.464	0.001	0.045*	-0.206	-0.617	0.205	0.831
2 vs 9	7.690	0.299	0.001	0.045*	0.019	-0.392	0.429	1.000
2 vs 10	7.988	0.307	0.002	0.090	-0.007	-0.418	0.403	1.000
3 vs 4	1.717	0.087	0.165	1.000	0.027	-0.383	0.438	1.000
3 vs 5	8.485	0.320	0.001	0.045*	-0.310	-0.720	0.101	0.311
3 vs 6	4.413	0.197	0.003	0.135	0.044	-0.367	0.454	1.000
3 vs 7	4.304	0.193	0.002	0.090	-0.048	-0.459	0.362	1.000
3 vs 8	13.604	0.430	0.001	0.045*	-0.314	-0.724	0.097	0.294
3 vs 9	4.701	0.207	0.001	0.045*	-0.089	-0.500	0.322	0.999
3 vs 10	6.041	0.251	0.001	0.045*	-0.115	-0.525	0.296	0.996

TABLE B.3: F value,  $R^2$ , p value and adjusted p value for pairwise multivariate analysis of variance (adonis tests) and the difference between the dispersion of communities, lower and upper boundaries of the differences and adjusted p values for pairwise tukey-tests.

Pairs of sites	F	Pairwise multivariate analysis of variance			Pairwise Tukey-test for community dispersion			
		R <sup>2</sup>	p.value	p.adjust	Diff	Lower	Upper	p.adjust
4 vs 5	5.571	0.236	0.001	0.045*	-0.337	-0.747	0.074	0.205
4 vs 6	6.196	0.256	0.001	0.045*	0.016	-0.394	0.427	1.000
4 vs 7	3.214	0.152	0.010	0.450	-0.075	-0.486	0.335	1.000
4 vs 8	8.524	0.321	0.001	0.045*	-0.341	-0.751	0.070	0.193
4 vs 9	7.137	0.284	0.001	0.045*	-0.116	-0.527	0.295	0.996
4 vs 10	10.432	0.367	0.001	0.045*	-0.142	-0.553	0.269	0.981
5 vs 6	2.857	0.137	0.021	0.945	0.353	-0.057	0.764	0.155
5 vs 7	6.691	0.271	0.001	0.045*	0.261	-0.149	0.672	0.556
5 vs 8	4.122	0.186	0.007	0.315	-0.004	-0.414	0.407	1.000
5 vs 9	6.496	0.265	0.001	0.045*	0.221	-0.190	0.631	0.767
5 vs 10	12.187	0.404	0.001	0.045*	0.195	-0.216	0.605	0.872
6 vs 7	5.470	0.233	0.001	0.045*	-0.092	-0.503	0.319	0.999
6 vs 8	6.412	0.263	0.001	0.045*	-0.357	-0.768	0.053	0.145
6 vs 9	3.823	0.175	0.004	0.180	-0.132	-0.543	0.278	0.988
6 vs 10	7.344	0.290	0.001	0.045*	-0.158	-0.569	0.252	0.961
7 vs 8	10.299	0.364	0.001	0.045*	-0.265	-0.676	0.145	0.535
7 vs 9	7.227	0.286	0.001	0.045*	-0.041	-0.451	0.370	1.000
7 vs 10	10.481	0.368	0.001	0.045*	-0.066	-0.477	0.344	1.000
8 vs 9	8.375	0.318	0.001	0.045*	0.225	-0.186	0.635	0.748
8 vs 10	16.800	0.483	0.001	0.045*	0.199	-0.212	0.609	0.858
9 vs 10	2.282	0.113	0.073	1.000	-0.026	-0.437	0.385	1.000

TABLE B.3: F value,  $R^2$ , p value and adjusted p value for pairwise multivariate analysis of variance (adonis tests) and the difference between the dispersion of communities, lower and upper boundaries of the differences and adjusted p values for pairwise tukey-tests.

Pairs of sites	F	Pairwise multivariate analysis of variance			Pairwise Tukey-test for community dispersion		
		R <sup>2</sup>	p.value	p.adjust	Diff	Lower	Upper

#### B.4.2 Regression

TABLE B.4: Slope and the p value (corrected and uncorrected for multiple tests) for the linear regression between positive or negative relative interaction strengths and mean bromeliad volume.

Species	Negative slope	P value	Corrected p value	Positive slope	P value	Corrected p value
Tipulidae	1.179e-03	0.0012**	0.0178**	7.12e-05	0.838	1.00
Corethrella.sp.	3.58e-04	0.435	1.00	4.25e-04	0.291	1.00
Leptagrion	2.44e-04	0.488	1.00	1.116e-03	0.020**	0.312
Elpidium	1.251e-03	0.001**	0.0163**	1.82e-04	0.609	1.00
Dero	9.69e-05	0.885	1.00	-5.53e-04	0.418	1.00
Chironomid	3.42e-04	0.337	1.00	3.73e-04	0.209	1.00
Scirtes	2.97e-05	0.938	1.00	-3.47e-04	0.34	1.00
Ceratopogonid	6.69e-05	0.872	1.00	-3.43e-04	0.34	1.00
Culex	2.96e-04	0.295	1.00	3.27e-05	0.891	1.00
Tabanid	1.19e-03	0.256	1.00	8.08e-04	0.277	1.00
Toxorhynchtes	4.3e-04	0.216	1.00	5.49e-04	0.300	1.00
Wyeomya	8.24e-04	0.004**	0.06**	-1.08e-04	0.794	1.00
Psychodidae	-7.78e-04	0.175	1.00	-4.67e-04	0.428	1.00

TABLE B.4: Slope and the p value (corrected and uncorrected for multiple tests) for the linear regression between positive or negative relative interaction strengths and mean bromeliad volume.

Species	Negative slope	P value	Corrected p value	Positive slope	P value	Corrected p value
Ephydridae	3.00e-03	0.057	0.690	1.66e-03	0.38	1.00
Dasyhelea.B	-2.91e-04	0.513	1.00	-4.25e-04	0.180	1.00

#### B.4.3 *Permutation results*

Since we only obtain one set of interaction strength values for each site, and we wanted to ensure that the results we obtained were not due to a random combination of presence and absences, we permuted the presence-absence matrix for each site, and then re-calculated the interaction strengths. We permuted the presence of the species by keeping the number of species in each bromeliad constant across each permutation. For each of our sites, we created 10,000 reshuffled sites. After calculating the interaction strength for each site, we recalculated 1) the effect of the type of sign and trophic position on the number of interactions and 2) the effect of water volume on the positive or negative interaction strength of the species.

THE EFFECT OF THE TYPE OF SIGN AND TROPHIC POSITION ON THE NUMBER OF INTERACTIONS: — To assess the effect of type of sign of interaction and trophic position on the number of interactions we ran a generalized linear model. For this model, we used the number of interactions each species had for the full data set as a function of the the sign of the species interaction (either positive or negative) and the trophic position using the species identity as a random effect; the units of replication are the sites. The interaction term of this model, between sign of species interaction and trophic position, tests if

FIGURE B.4: Predators have more negative interactions and prey have more positive interactions. a) Negative interactions are dominated by Leptagrion andromache (top predator). b) Positive interactions are dominated by Culex (dipteran prey). Darker colours indicate known predatory species prior to the analysis and lighter colours indicate prey species. c) Predators have a higher number of negative interactions while prey have a higher number of positive interactions. Blue indicates positive interactions and red indicate negative interactions. Bars represent mean and standard error of the mean. Positive interactions represent species that tend to co-occur, negative interactions represent species that do not tend to co-occur.

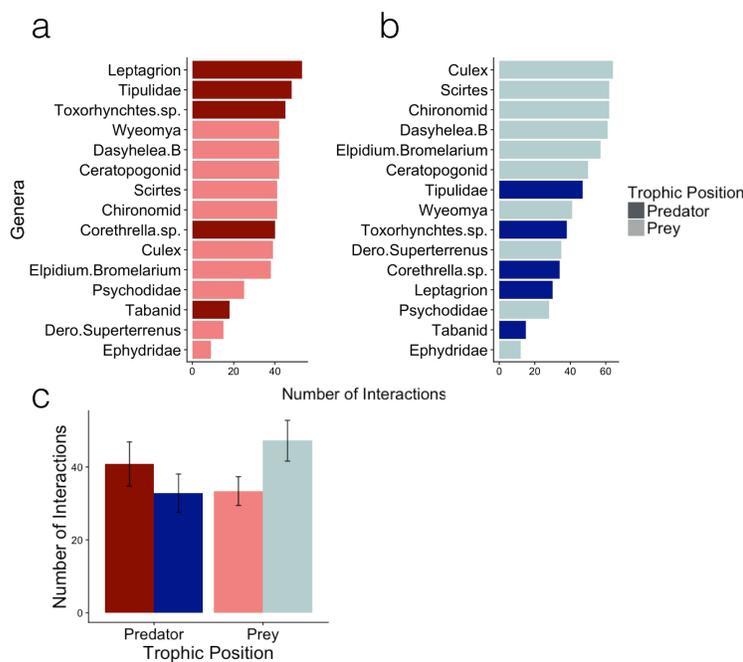


FIGURE B.5: Relative interaction strength against mean bromeliad water volume. Every panel represents a different focal species. Blue regression lines represent the positive interactions and the red regression lines represent the negative interactions.

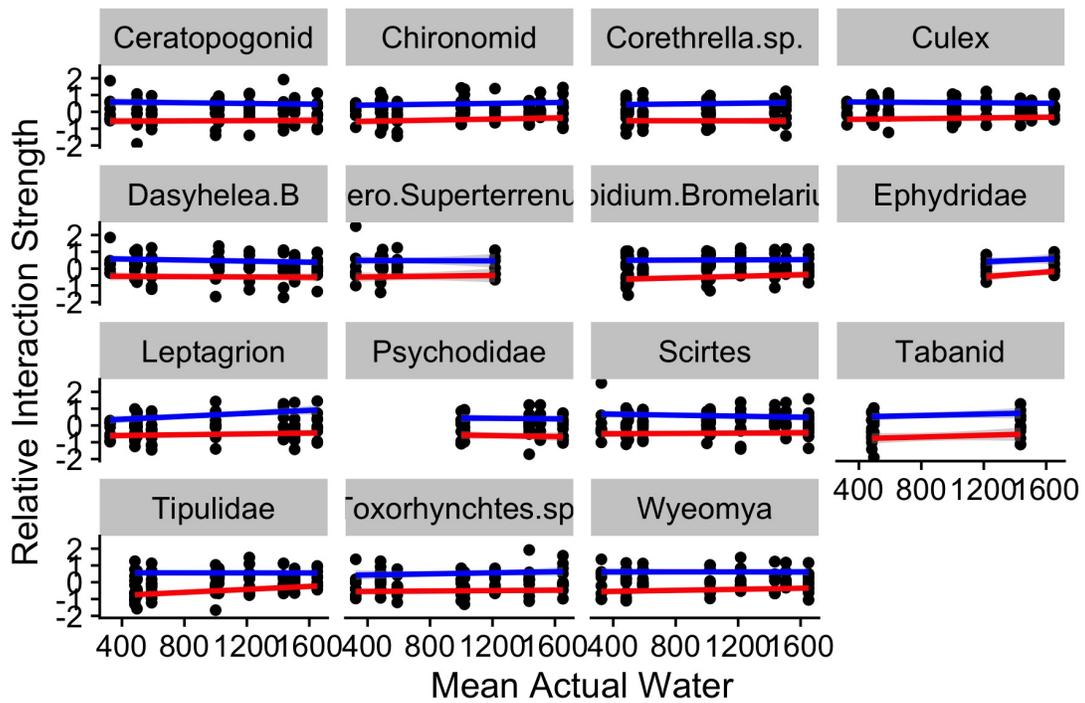
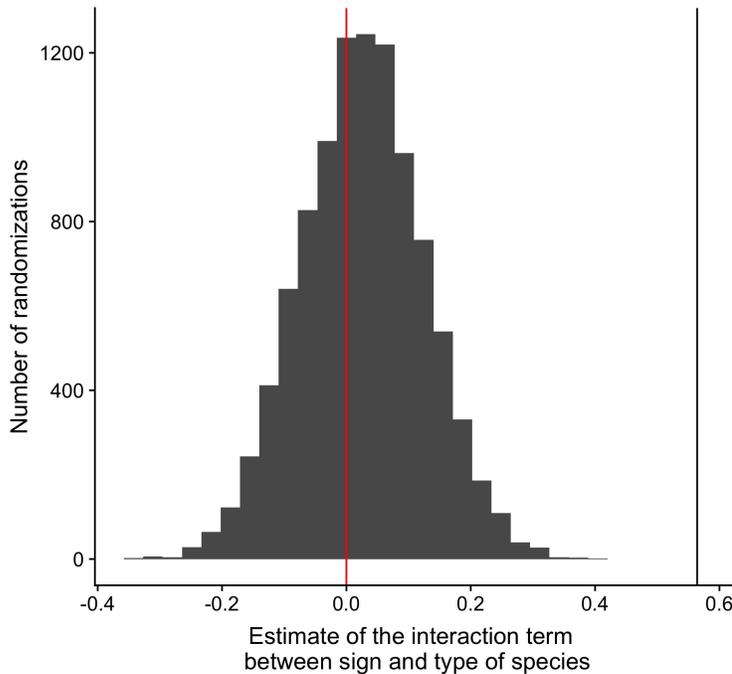


FIGURE B.6: Distribution of interaction terms after the presence of species was permuted 10,000 times. Red line represents 0 and the black line represents the actual interaction term found in our original data.

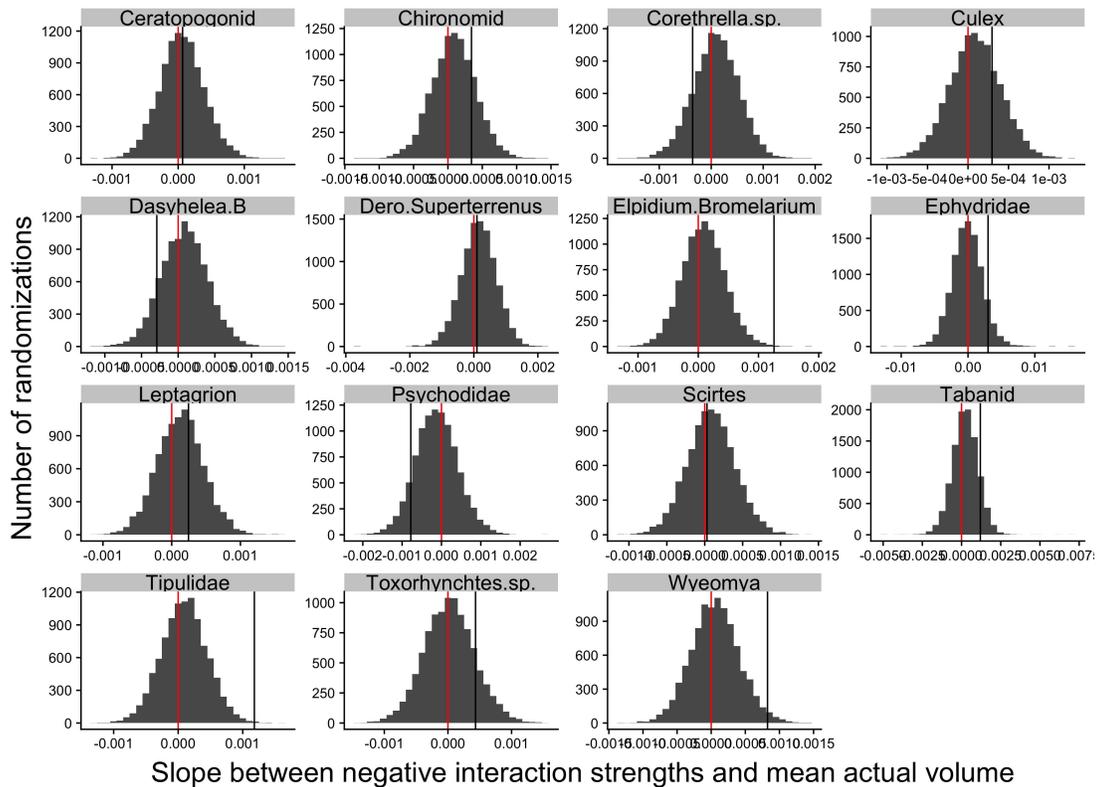


predators and prey have different distributions of the sign of their interactions. Here we show the distribution of the interaction term of the model from all the permutations vs the interaction term we obtained from our data.

Here we see that the interaction term of the model we obtained from our data does not overlap the distribution. We are calculating the P value for this permutation test by the proportion of values from the distribution that are greater than the value we obtained from the original data. Here the p-value is 0.

**THE EFFECT OF WATER VOLUME ON THE POSITIVE OR NEGATIVE INTERACTION STRENGTH OF THE SPECIES:** — To assess the effect of water volume on the positive or negative interaction strength of the species we used a linear regression. Here we compare the slope of that regression for each

FIGURE B.7: Distribution of negative slopes after the presence of species was permuted 10,000 times. Red line represents 0 and the black line represents the actual slope found in our original data.



species vs the distribution of slopes obtained from the permutation of the presence matrix.

We are calculating the P value for this permutation test by the proportion of values from the distribution that are greater than the value we obtained from the original data.

FIGURE B.8: Distribution of positive slopes after the presence of species was permuted 10,000 times. Red line represents 0 and the black line represents the actual slope found in our original data.

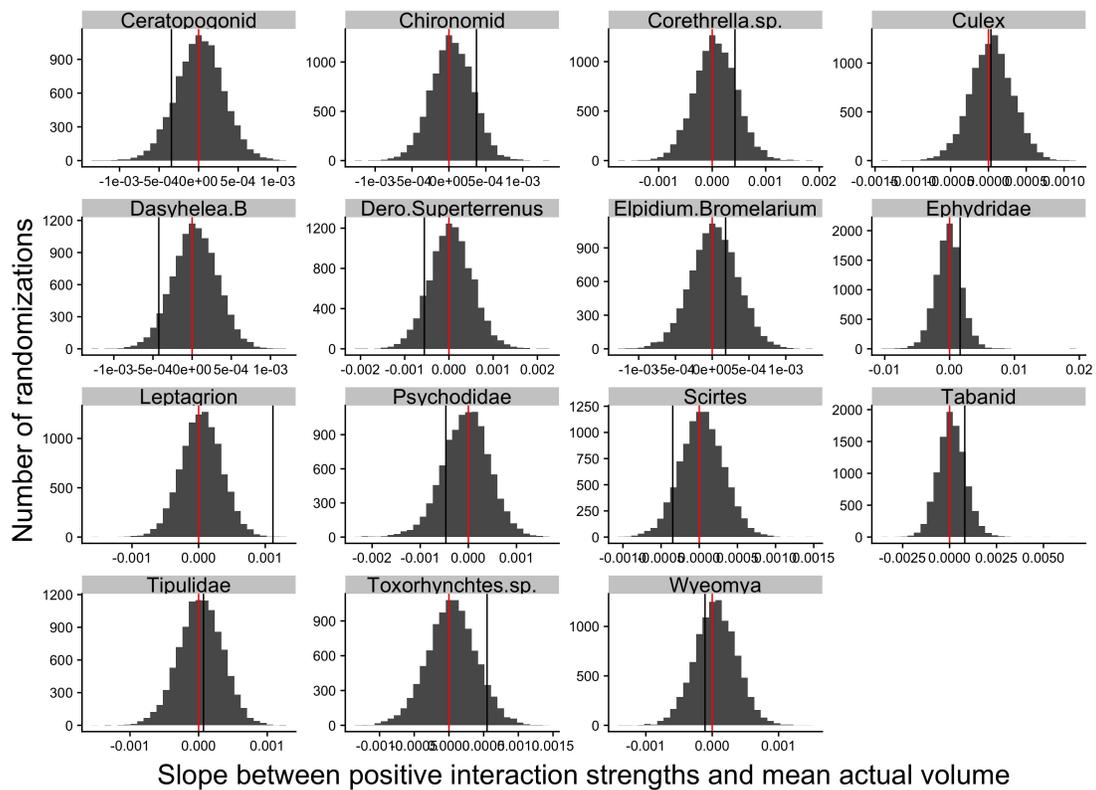


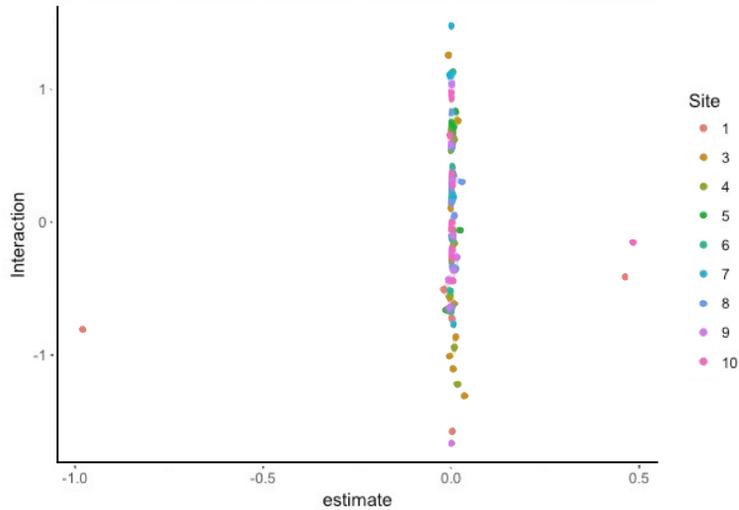
TABLE B.5: P value for the permutation test for linear regression between positive or negative relative interaction strengths and mean bromeliad volume.

Species	P value for positive interaction values	P value for negative interaction values
Tipulidae	0.448	0.001 **
Corethrella.sp.	0.175	0.845
Leptagrion	0.0005 **	0.360
Elpidium bromelarium	0.320	0.001 **
Dero superterrenus	0.897	0.545
Scirtes	0.911	0.569
Chironomid	0.118	0.209
Ceratopogonid	0.886	0.481
Culex	0.513	0.252
Tabanid	0.165	0.115
Toxorhynchtes.sp.	0.077	0.149
Wyeomya	0.698	0.019 **
Psychodidae	0.812	0.892
Ephydridae	0.169	0.080
Dasyhelea.B	0.947	0.881

#### B.4.4 *Relative interaction strength vs presence*

The negative interaction strength between two species can be due to the absence of one caused by the local community environment (for example, low water volume filters out certain species). Therefore, we wanted to test if

FIGURE B.9: The estimate of a logistic regression between actual water volume in the bromeliad and the presence of the species versus the relative interaction strength with the tipulid at each site.



the slope between the presence of any given species and water volume in a bromeliad was related to their interaction strength to other species. The expectation here was that if the interaction strength at the site scale is driven by the filtering of the local environment, then they would be more likely to engage in negative interactions if the slope between presence and water volume is positive. Here we plotted the slope of the relationship between the presence of each species and water volume in the bromeliad, against the interaction strength value that they have against the tipulid on that site. We find that there is no relationship between this slope and the interaction strength with the Tipulid.

## SUPPLEMENTARY INFORMATION TO CHAPTER 3

---

### C.1 SUPPORTING RESULTS

TABLE C.1: The prey species used for the feeding trials with the predator *Leptagrion*, their average body mass and the densities we used for the feeding trials

Prey	Mean Body Mass (mg)	Densities
Culex sp 1	0.1771	1, 2, 5, 7, 10, 20, 30, 50
Culex sp 2	0.0906	1, 2, 5, 7, 10, 20
Forcypomia	0.0737	1, 2, 5, 7, 10, 20, 30, 60
Oligochaete	0.1242	1, 2, 5, 7, 10, 20, 30, 60
Psychodidae	0.2198	1, 2, 5, 10, 30
Scirtes sp 1	0.3286	1, 2, 5, 10, 30
Scirtes sp 2	0.4344	1, 2, 3, 4, 5, 6
Tipulid	0.2935	1, 2, 3, 4, 5, 6, 10, 17

FIGURE C.1: The attack rate, carrying capacity and growth rate decrease with prey body size, the handling time increases with prey body mass. The attack rates and the handling times of the model had to be reduced in magnitude (but not in shape) to reduce the amplitude of the dynamics (Parameters in tables 1 and 2)

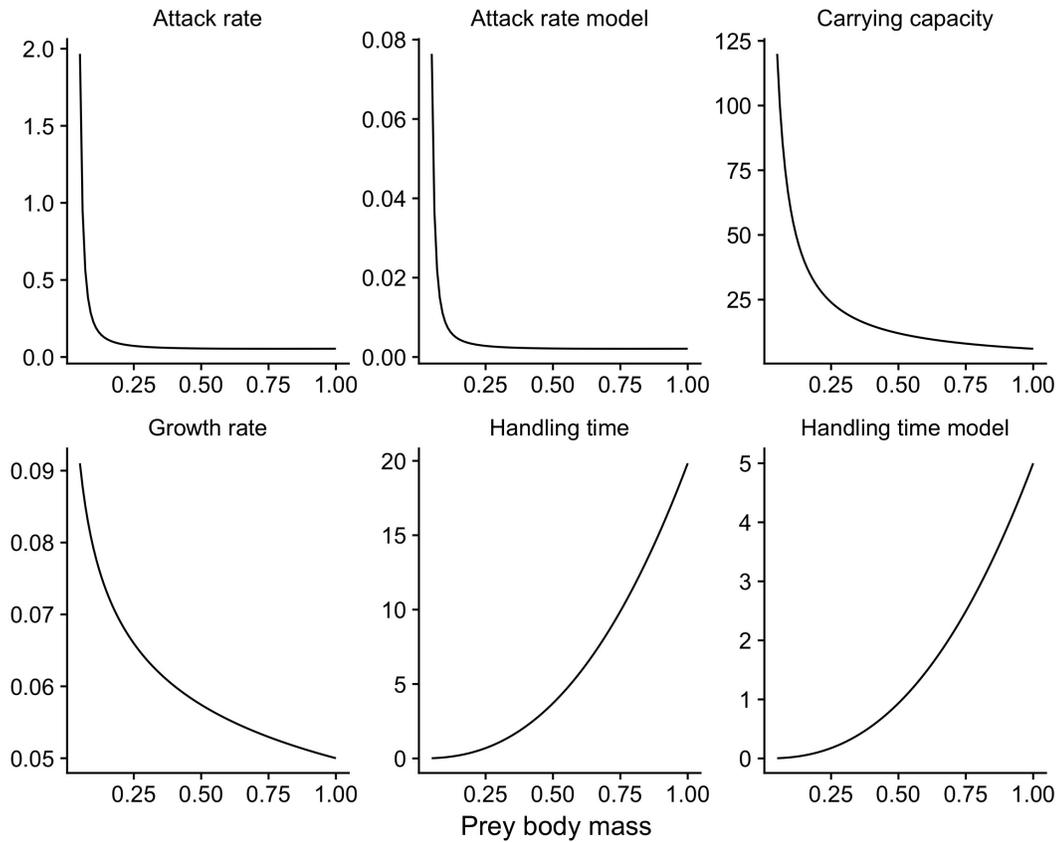


FIGURE C.2: Parameter space is well covered in the Monte Carlo sampling. As diversity increases, most values for body mass of the prey 1 and 2 are represented. All other prey show a similar coverage. Each panel shows the diversity being considered

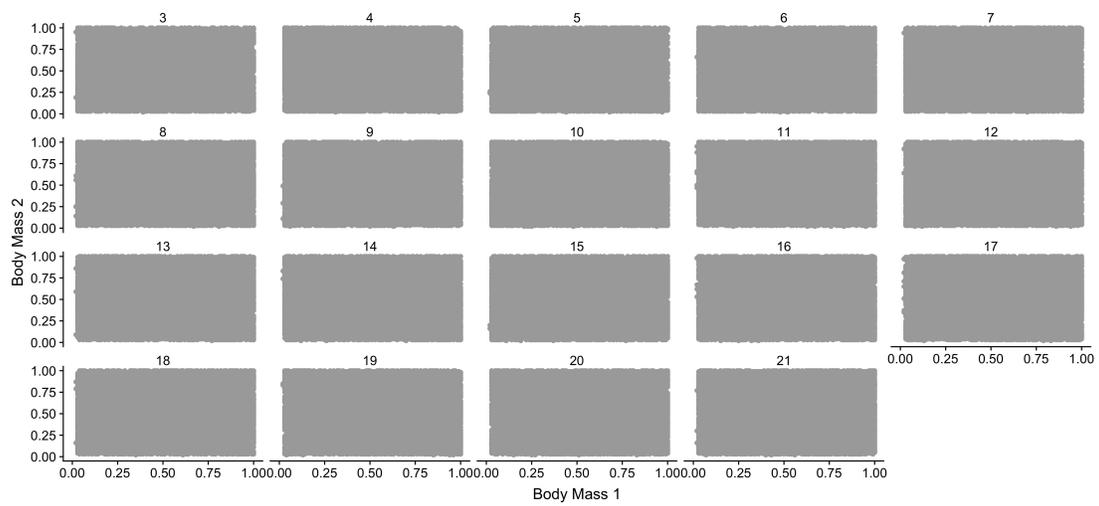


FIGURE C.3: Time series of two prey species and the predator. If the body mass of both prey is high the predator will go extinct. Body masses of each prey species are presented above each graph. All a-d show prey species that have high competitive interaction strengths since they are close in body mass. a) and b) show two prey species with intermediate body mass, c) and d) show two prey species with large prey mass

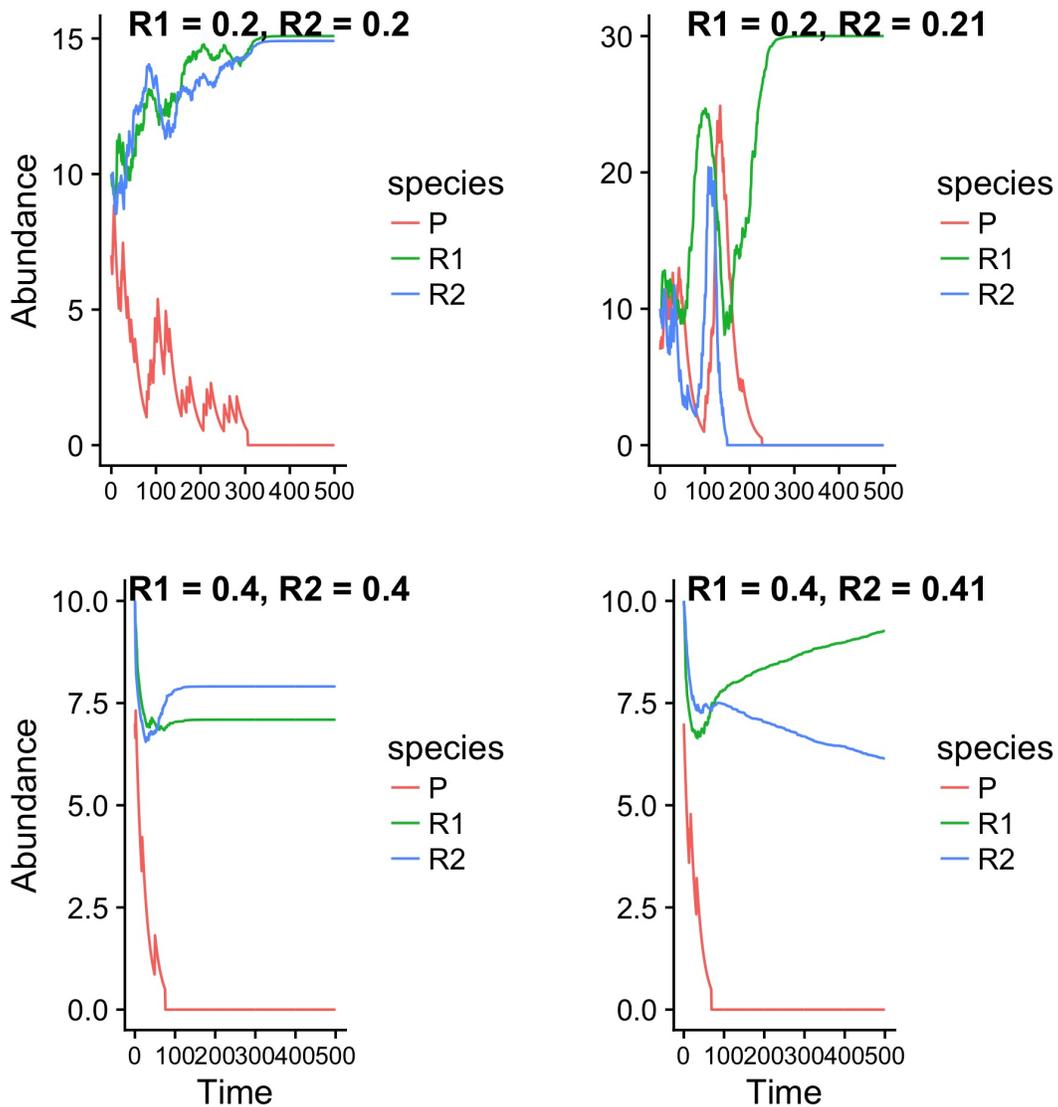


FIGURE C.4: Time series of two prey species and the predator. The body mass of each prey species is presented above each graph. Panels a- d show prey species that have low competitive interaction strengths since they differ in body mass. a) The predator and the largest prey persist, the prey with the smaller body mass goes extinct due to predation and competition. b) and c) Both prey species and the predator persist through time since the smallest prey has an intermediate body mass. d) The predator goes extinct when both prey are large

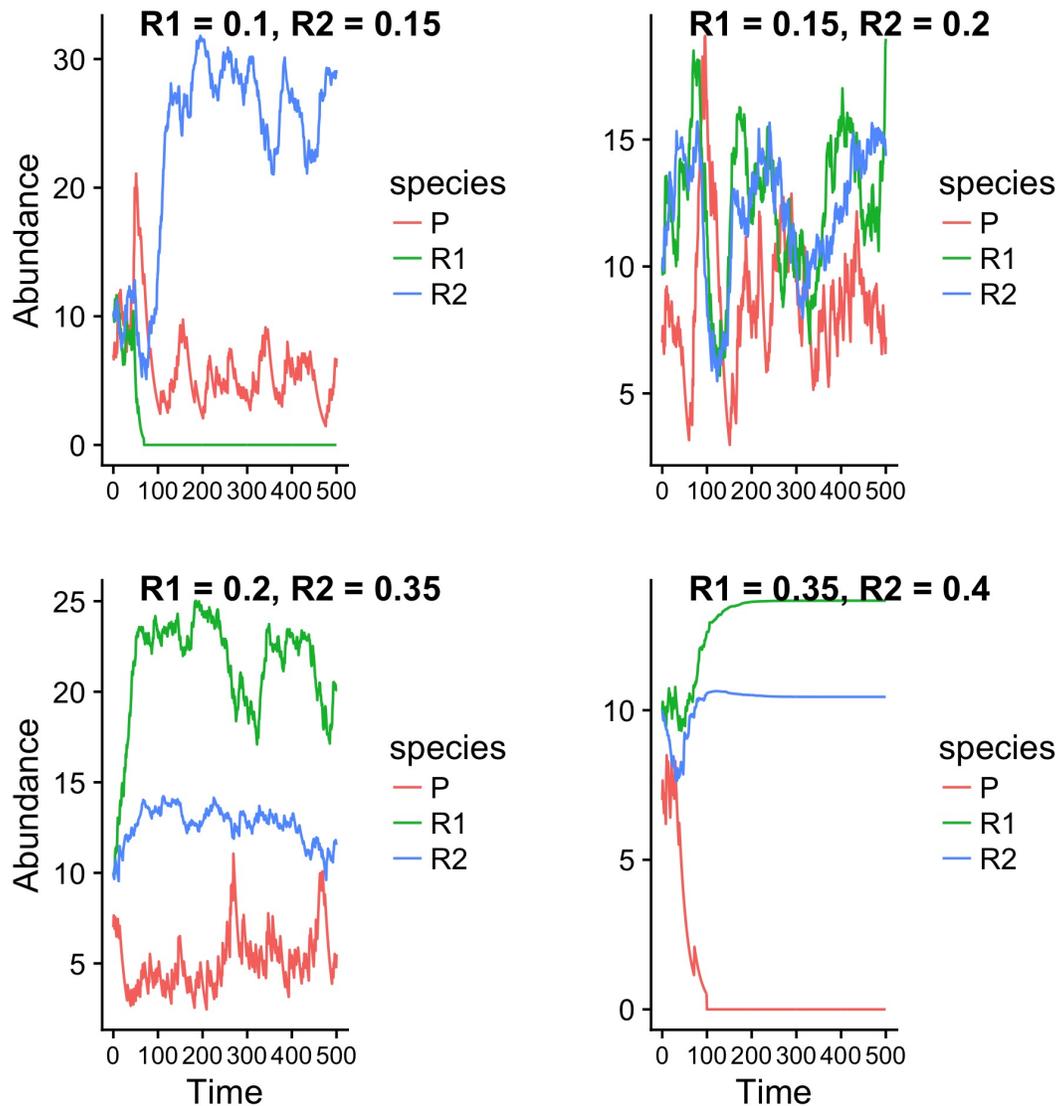


FIGURE C.5: Jensen's inequality alters the effective attack rate and handling time as prey diversity increases. a) As diversity increases, the average body mass of the prey increases causing an increase in the handling time  $h(\bar{m})$  but a decrease in the attack rate  $a(\bar{m})$ . b) Due to Jensen's inequality, increasing diversity increases the combination of effective attack rates and effective handling times (compare to panel a). Given the upward curvature of the attack rate c) and the handling time d), we expect that the effective attack rate and the effective handling time will be higher than the attack rate and the handling time of the mean body mass

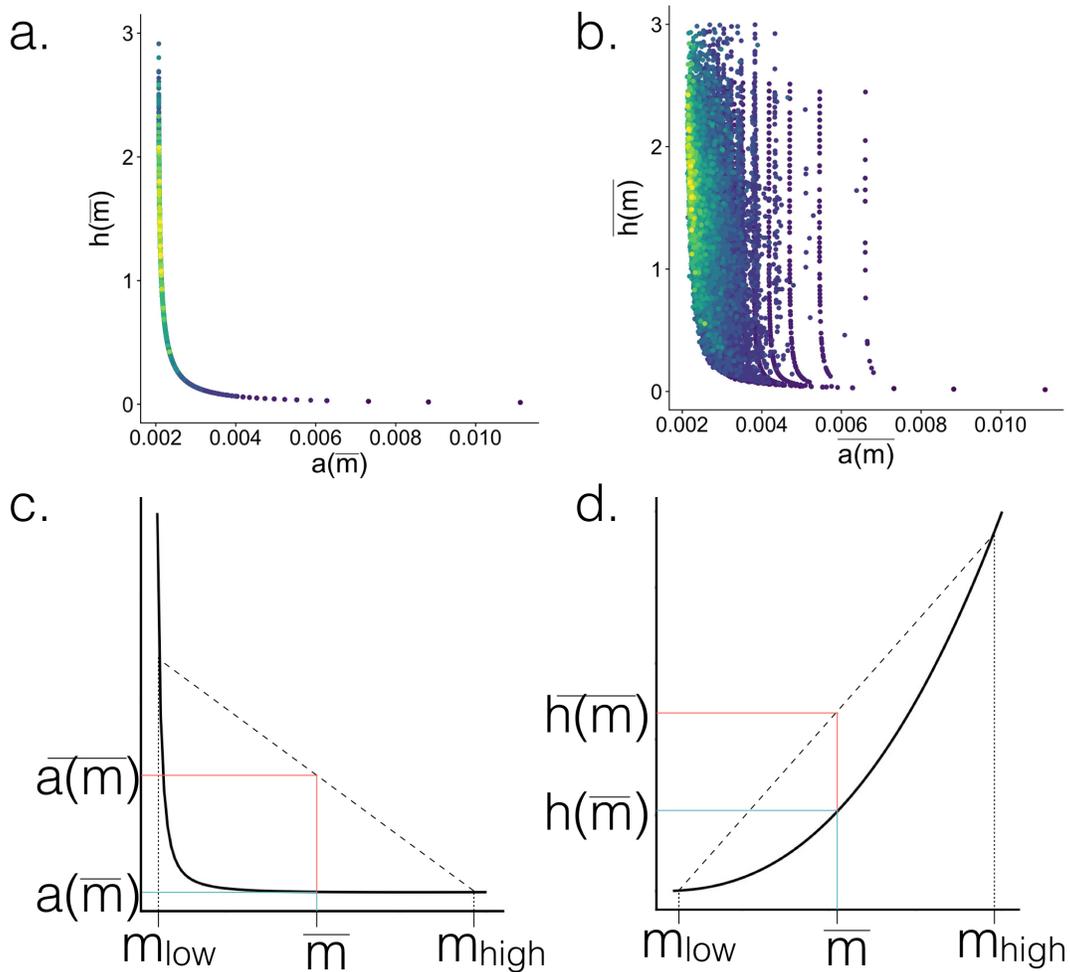


FIGURE C.6: The probability that the predator persists has a unimodal relationship with prey diversity. Since prey that are close to each other in body mass have higher competition coefficients, we calculated the proportion of species in a run that had the same body mass. As the proportion of prey species with the same body mass increases, predator persistence decreases. But this decrease is relatively small compared to the effects of having small or large prey (Figure 3.4)

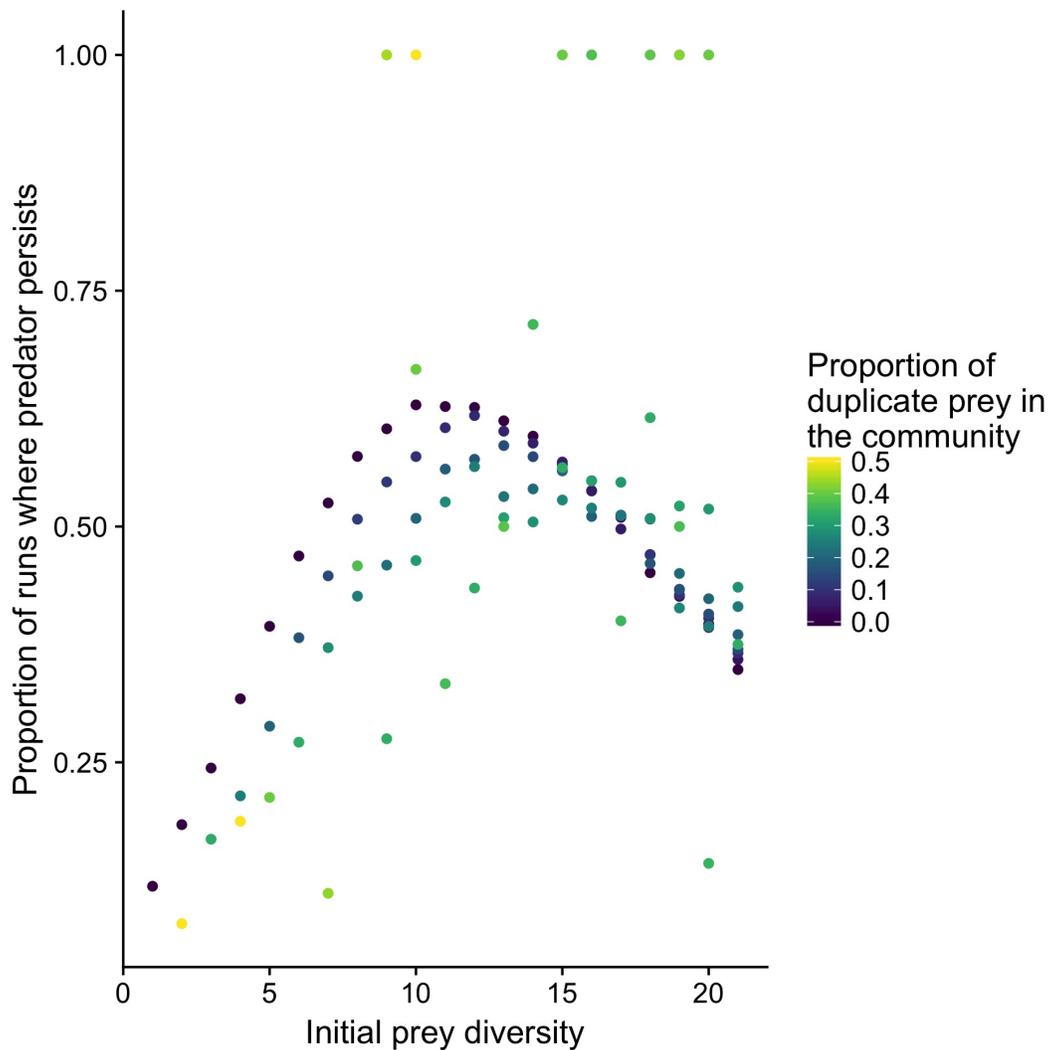
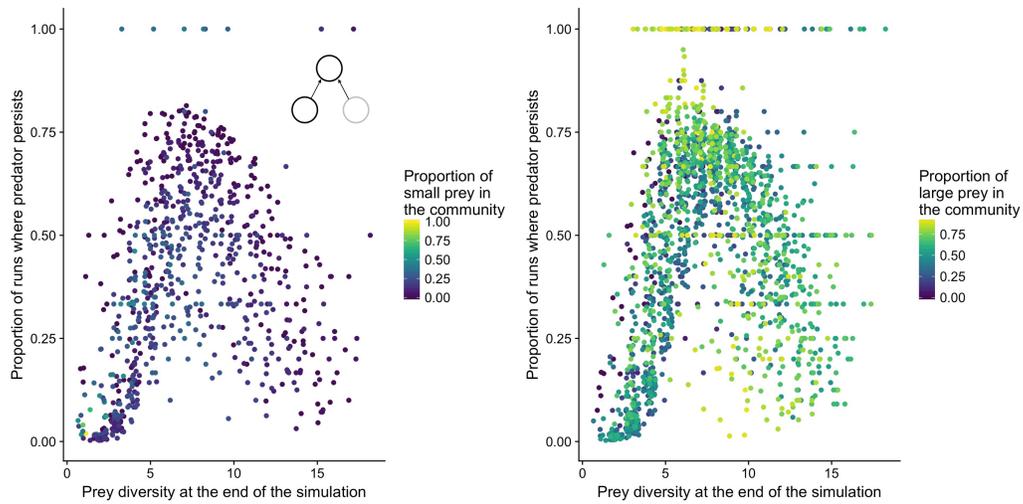


FIGURE C.7: a) As the proportion of small prey in the community increases, the proportion of runs where the predator persists decreases slightly. b) For any given predator persistence, higher final prey diversity occurs when the run had a higher proportion of large prey species



## C.2 ASYMMETRIC COMPETITION

The model we presented assumes that competition is symmetrical with regards to the trait. That is, there is no benefit of being large or small in the outcome of competition. Arguably, this may not be the case. Evidence suggests that the outcome of competition is dependent on the body size of the individuals, where larger individuals outcompete smaller individuals Persson (1985); Lawton and Hassell (1981). In certain cases smaller individuals can also outcompete larger individuals Winder *et al.* (2008). Here, however, we will only analyze the case where larger individuals outcompete smaller individuals.

Asymmetric competition reduces the window of persistence of the whole community (Figure c.8, c.9). Persistence only occurs when both prey are intermediate in body mass and the competition is not vastly asymmetric, that is, their body masses are close (Figure c.8c). If both prey are small, but one prey is very small, the latter will go extinct due to predator overconsumption (Figure c.8a). Similarly, if the both prey are large, where the predator cannot persist, and the larger prey will outcompete the smaller prey. Due to the asymmetry in competition, the smaller prey will extinct much faster than in the case of symmetric competition (Figure c.8d vs Figure c.4d).

Predator persistence occurs less frequently when prey species compete asymmetrically (Figure c.10a) and it decreases with diversity. Similarly, as in the case of symmetric competition, a higher proportion of smaller prey species decreases the predator's persistence. Contrary to the case of symmetric competition, a higher proportion of large prey species now decreases the predator's persistence. This occurs due to an increase in competitive exclusion at high diversity and larger prey outcompeting the intermediate prey necessary for predator persistence. In this scenario, large prey species do not interact weakly

in the system (so they do not stabilize the system), but instead destabilize them due to a larger competitive advantage.

FIGURE C.8: Time series of two prey species and the predator. The body mass of each prey species is presented above each graph. Panels a- d show prey species that have low competitive interaction strengths since they differ in body mass. a) The predator and the largest prey persist, but the smaller prey goes extinct. b) and c) Both prey species and the predator persist through time since the smallest prey has an intermediate body mass. d) The predator goes extinct when both prey are large

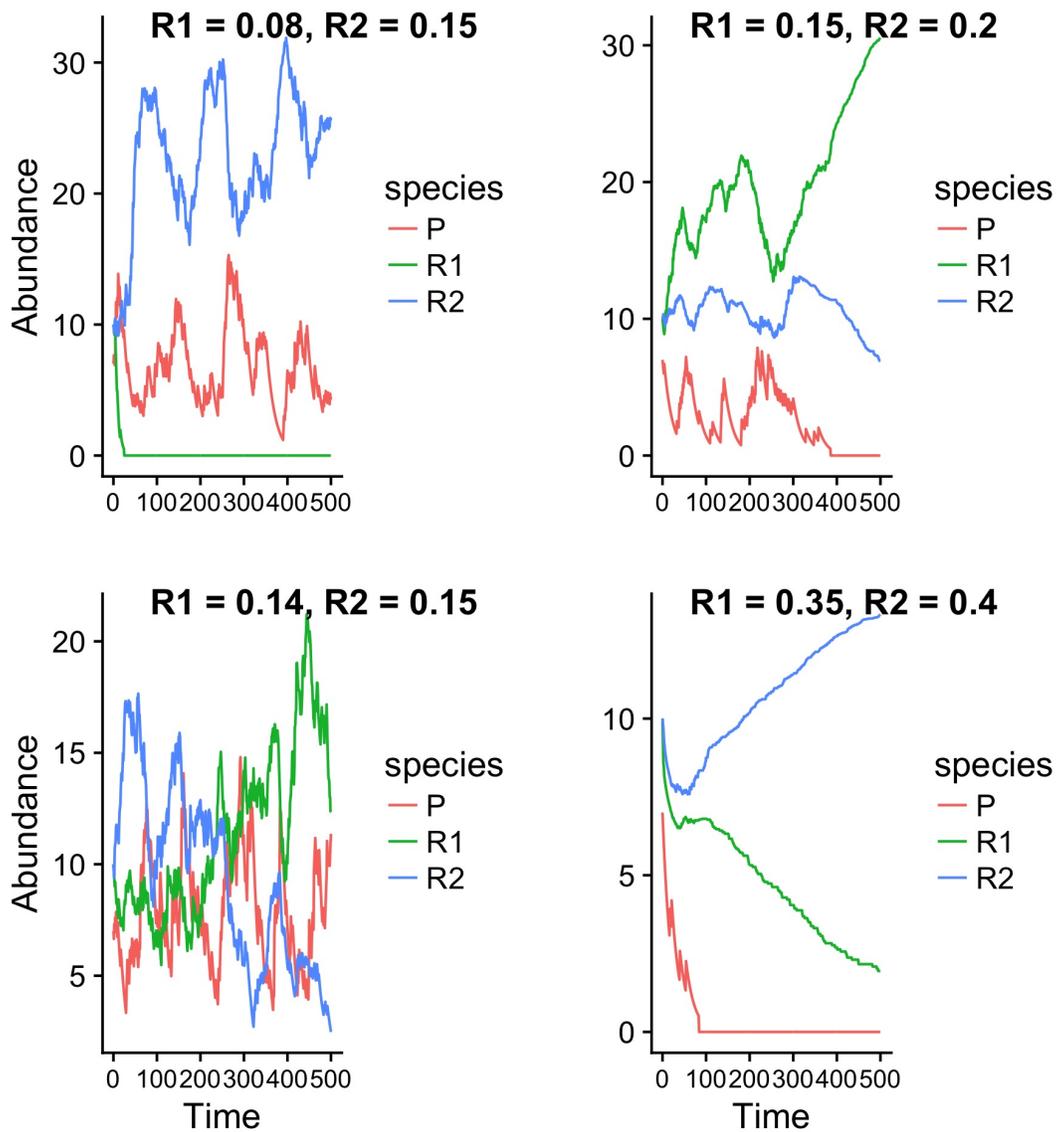


FIGURE C.9: Predator persistence is constrained to two intermediate similarly sized prey or one intermediate and one large prey. Prey competitively exclude each other unless they are similarly sized (1:1 edge). On the x axis is the body mass of prey 1, and on the y axis is the body mass of prey 2. Panels a-c show the proportion of the runs with a certain outcome, where yellow represents all the runs and purple none of the runs: a) both prey species and the predator persists; b) the predator goes extinct and only the two prey persist; c) one prey and the predator persist and one prey goes extinct (in every case the smaller prey). These scenarios are detailed in the food web modules: the grey circles represent species that go extinct and the black circles represent species that persist

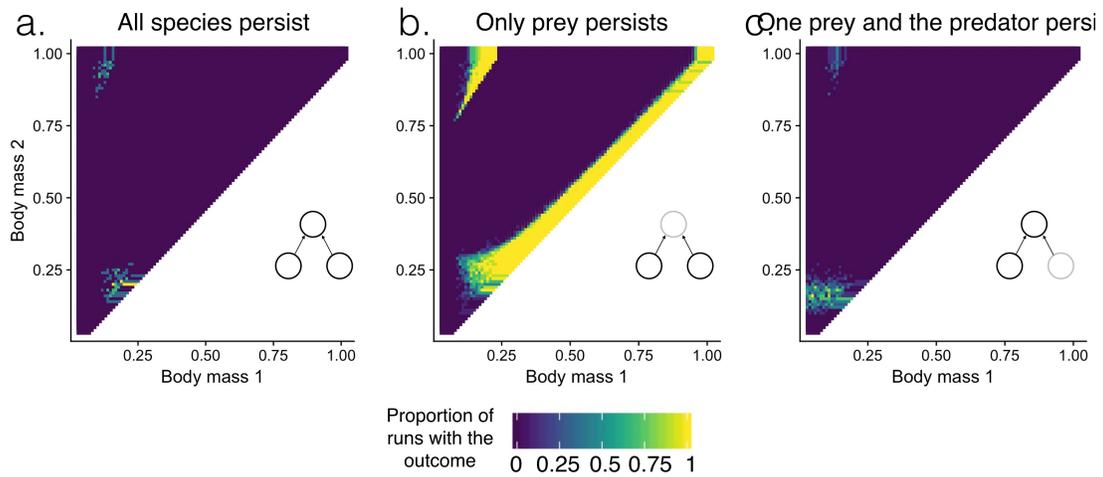
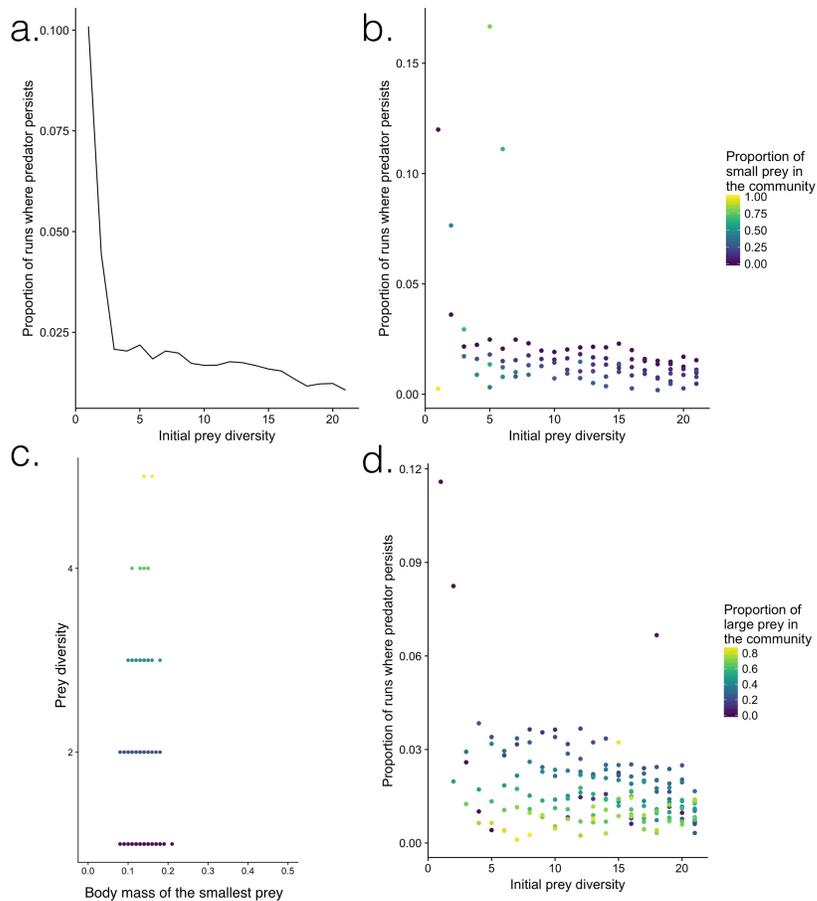


FIGURE C.10: Prey diversity decreases predator persistence. a) Increasing diversity decreases the proportion of runs where the predator and at least one prey survived. b) As the proportion of small prey in the community increases, the proportion of runs where the predator persists decreases. c) The range of body masses that allow the predator to persist becomes narrower with prey diversity. d) As the proportion of large prey in the community increases, the proportion of runs where the predator persists decreases



SUPPLEMENTARY INFORMATION TO CHAPTER 4

---

## D.1 MODIFICATIONS TO DNA EXTRACTION PROTOCOLS

D.1.1 *Modifications of the Qiagen protocol*

Based on the purification of Total DNA from Animal Tissues (Spin-Column Protocol):

1. Let ethanol evaporate and macerate samples before adding ATL buffer.
2. Add 30  $\mu\text{L}$  of proteinase K instead of 20  $\mu\text{L}$ .
3. Incubate for seven days instead of 6-8 hours. Vortex, flick and spin samples daily.
4. Add 250  $\mu\text{L}$  of buffer AW<sub>1</sub> and AW<sub>2</sub> instead of 500  $\mu\text{L}$ .
5. Incubate for 15 minutes in buffer AE instead of 1 minute.
6. Repeat AE incubation.

D.1.2 *Modifications of the Ommiprep protocol*

Based on the Ethanol or Formalin Fixed Tissue Protocol:

1. Let the insect dry and then re-hydrated in water for 1 hour.
2. Freeze in liquid nitrogen and lyse with beads.

3. Add 6  $\mu\text{L}$  of Proteinase K instead of 5  $\mu\text{L}$
4. Incubate the sample at 55-60°C overnight instead of 60 minutes.
5. Always centrifuge at 22°C.
6. Added 4  $\mu\text{L}$  instead of 2  $\mu\text{L}$  Mussel Glycogen.
7. Incubate for 1 hour after adding the Mussel Glycogen.
8. Centrifuge at max speed for 20 minutes instead of 1.
9. Add 35  $\mu\text{L}$  TE Buffer to the pellet, instead of 50  $\mu\text{L}$ . Incubate in the fridge at 4°C for several days.

## D.2 SUPPORTING RESULTS

FIGURE D.1: Individuals of both the odonate predator and the tipulid prey form two groups of individuals by the identity by state similarity.

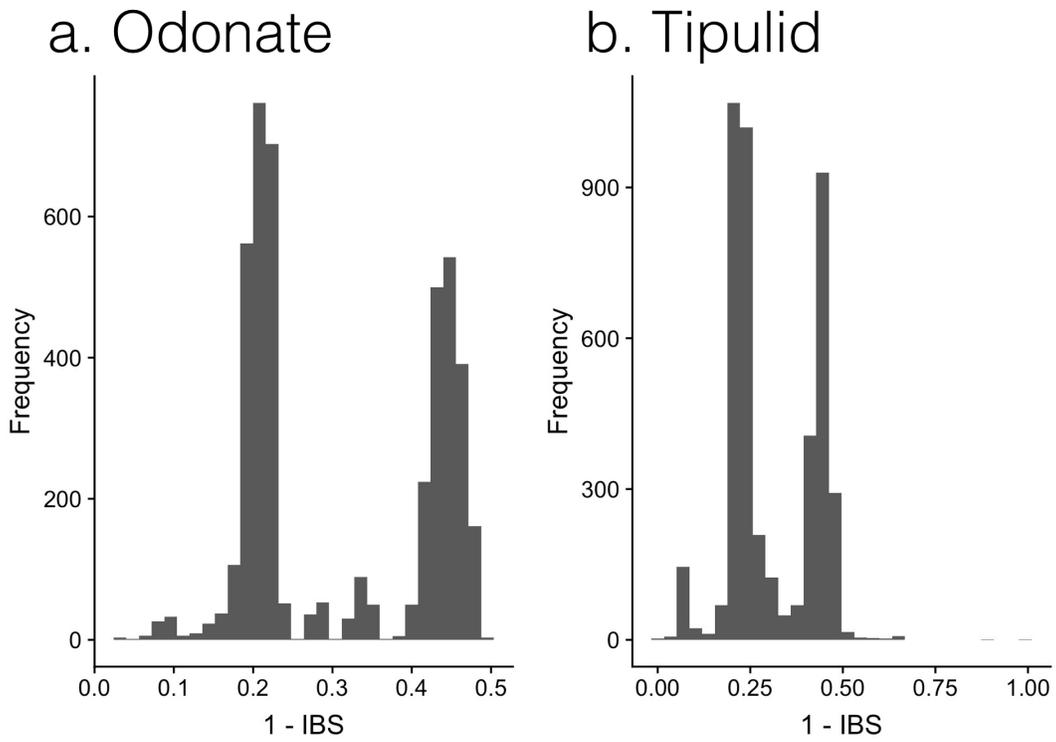


FIGURE D.2: Both the odonate predator (a, c) and the tipulid prey (b, d) have two major clusters of individuals as shown by the cross validation error and loglikelihood results from Admixture. Different colours represent different model runs with a different random seed.

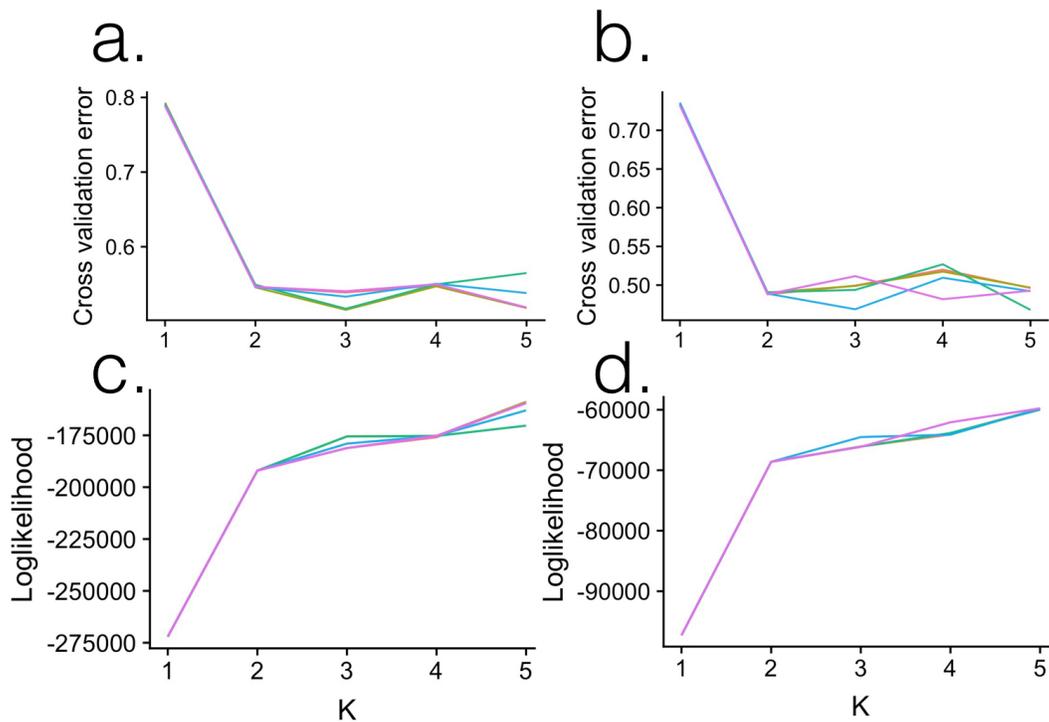


FIGURE D.3: The odonate predator (a) and the tipulid prey (b) both have two major clusters with little admixture. We separated these clusters for further analysis.

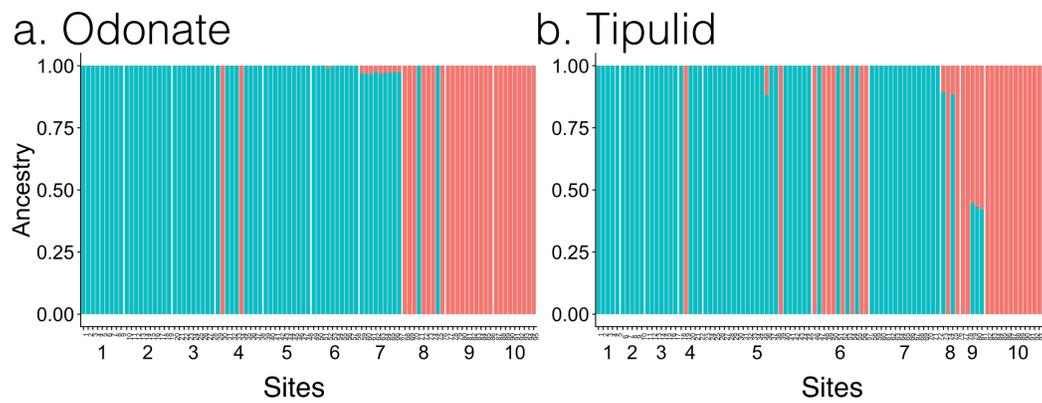
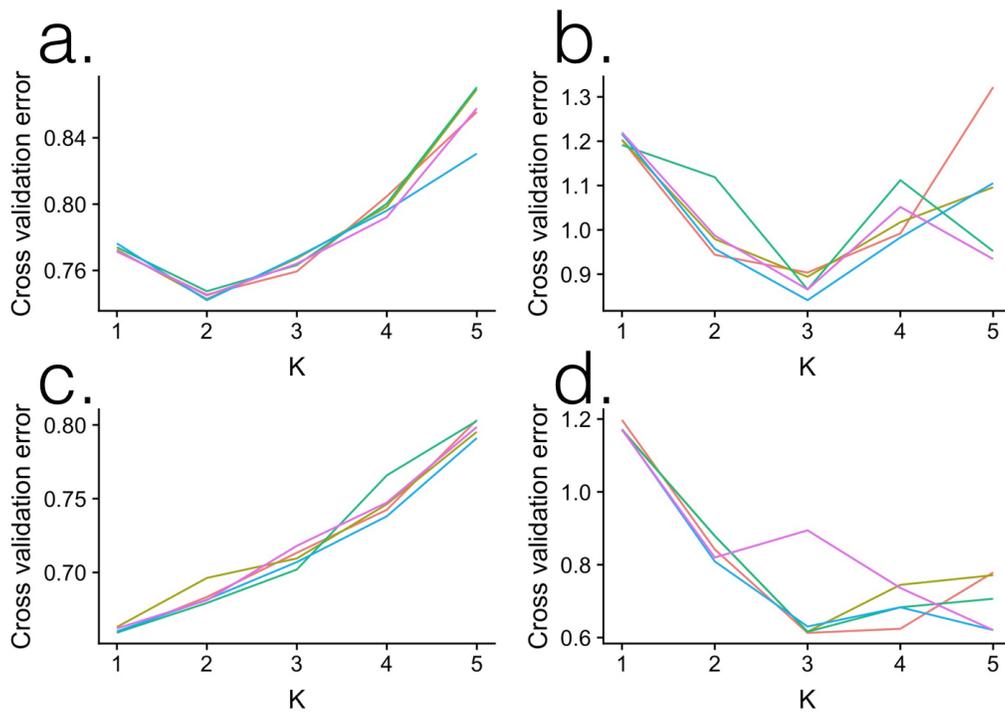


FIGURE D.4: The major cluster of the odonate predator (a) is best represented by two subclusters, while the minor cluster of the odonate predator (b) is best represented by three subclusters. The major cluster of the tipulid (c) is best represented by one clusters, while the minor cluster of the tipulid prey (d) is best represented by three subclusters.



## SUPPLEMENTARY INFORMATION TO CHAPTER 5

---

### E.1 HOMEWORK ASSIGNMENTS

The homework assignments were designed to (i) reduce the extraneous load, (ii) reduce the intrinsic load and (iii) increase the germane load.

#### E.1.1 (i) Reducing the extraneous load

**SPLIT ATTENTION EFFECT** — Code is often presented separate from its explanation, for example:

```
binom.test(c(463, 850), 0.5)
```

Where 0.5 is the probability of success, 463 is the number of successes, 850 is the number of failures and `binom.test` is the function to run a binomial test.

We minimized the split attention effect for the students learning new code by providing code with the english explanations of each part:

```

      Function    Successes    Probability
      ↓          ↓            ↓
> binom.test(c(463, 850), 0.5)
      ↑
    Failures
  
```

And by providing explanations of the code output:

Variable  
↓

```
> grouped_titanic
Source: local data frame [1,313 x 7]
Groups: Sex, Survive [4]
```

	Passenger.Class	Name	Age
	<fctr>	<fctr>	<dbl>
1	1st	Allen, Miss Elisabeth Walton	29.0000
2	1st	Allison, Miss Helen Loraine	2.0000

WORKED EXAMPLE EFFECT — For conceptual questions (in Biostatistics) and for programming questions, we included a worked example in every assignment. For each worked example, we also highlighted the different parts that must be completed to do a question correctly.

Conceptual question:

**From Whitlock and Schluter:**

Ch 3: 15 The data in the accompanying table are from an ecological study of the entire rainforest community at El Verde in Puerto Rico (Waide and Reagan 1996). Diet Breadth is the number of types of food eaten by an animal species. The number of animal species having each diet breadth is shown in the second column. The total number of species listed is  $n = 127$ .

- a) Calculate the median number of prey types consumed by animal species in the community.

**First**, order your list of numbers from smallest to largest.

Luckily, this step has already been completed for us in the table. The lowest diet breadth is 1 and the greatest diet breadth is  $>20$ . Keep in mind that this is a condensed list as multiples of each diet breadth are counted up in the frequency columns. If we wrote out this list with full, it would begin with twenty-one "1"s followed by eight "2"s and so on.

**Second**, determine how you will obtain or calculate your median number.

Because this list contains an odd number of observations, our median will be equal to  $Y_{([n+1]/2)}$  meaning that the median will be equal to  $Y_{64}$ .

**Third**, find the 64th number in your list. This will be your median.

Add together the multiples of each diet breadth starting at "1" until you reach 64.

$$21 + 8 = 29$$

$$21 + 8 + 9 = 38$$

$$21 + 8 + 9 + 10 = 48$$

$$21 + 8 + 9 + 10 + 8 = 56$$

$$21 + 8 + 9 + 10 + 8 + 3 = 59$$

$$21 + 8 + 9 + 10 + 8 + 3 + 4 = 63$$

$$21 + 8 + 9 + 10 + 8 + 3 + 4 + 8 = 71$$

From these calculators we know that numbers 64 - 71 in our list are "8"s. Therefore our median is 8 prey species.

Programming question:

**Q1** Using the "titanic" dataset mentioned above, create three histograms: one for "Passenger Class", one for "Age" and one for "Survive".

**First**, we need to read the "titanic" data frame into R. This is done using the following line of code:

```
> titanic <- read.csv(file = 'titanic.csv')
```

**\*\*Note**, after uploading your csv, you should see the item, "titanic", under the "Data" heading in the "Environment" tab of the upper right panel of your R console. If you would like to view the data frame in a spreadsheet format, simply click on this item or employ the View() function in the following line of code:

```
> View(titanic)
```

**Second**, now that we have read our data into R, we can use "ggplot2" to create our histograms. Start by using the library() function to load this package onto your console.

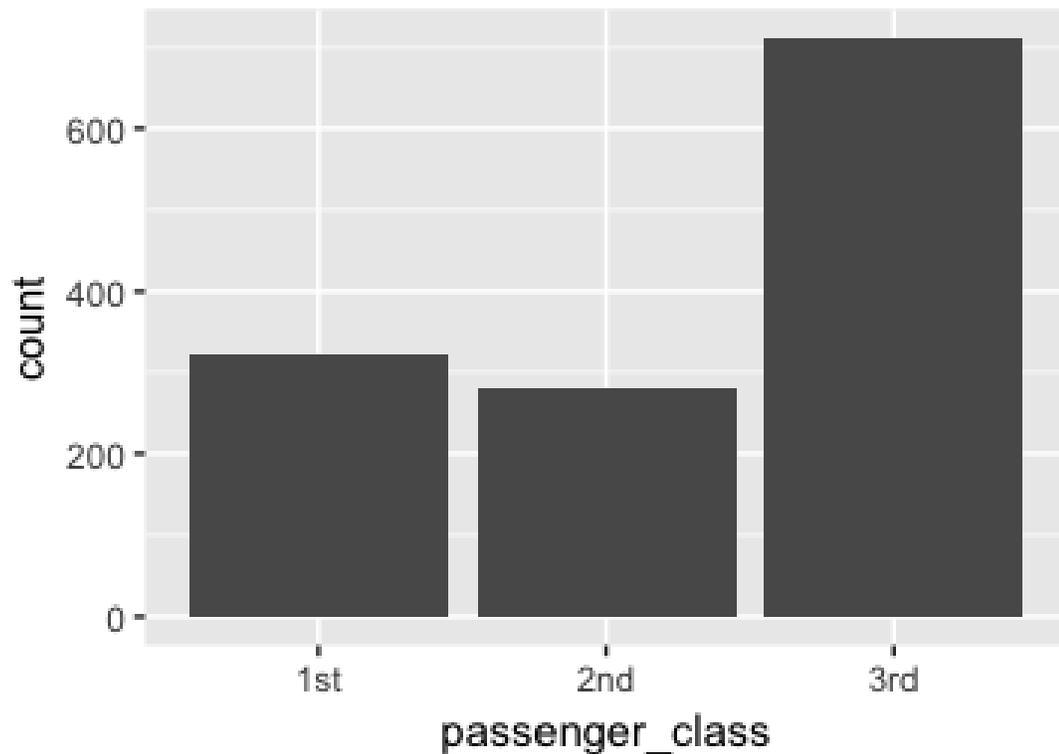
```
> library(ggplot2)
```

**Third**, construct your first histogram using the "Passenger Class" variable. This can be done using a single line of code. Please return to the previous sections if you do not remember what the different parts of the following code represent.

```
> ggplot(data = titanic, aes(x = Passenger.Class))+  
  geom_histogram(stat= 'count')
```

**For you to think:** What kind of data is "Passenger Class"? Why do we need to include the 'count' statistic in this line of the code?

**Fourth**, view your graph. It should look like this:



PARTIALLY COMPLETED PROBLEM — Similar to the worked example problems, we included partially completed questions in conceptual (Biostatistics) and programming questions. We scaffolded both the different parts to a question and the steps needed for each part of the question. As the assignment progressed, both types of scaffolding was reduced.

Conceptual question:

From Whitlock and Schluter:

Ch 3: 21 Researchers have created every possible "knockout" line in yeast. Each line has exactly one gene deleted and all the other genes present (Steinmetz et al.

2002). The growth rate—how fast the number of cells increase per hour—of each of these yeast lines has also been measured, expressed as a multiple of the growth rate of the wild type that has all the genes present. In other words, a growth rate greater than 1 means that a given knockout line grows faster than the wild type, whereas a growth rate less than 1 means it grows more slowly. Below is the growth rate of a random sample of knockout lines:

0.86, 1.02, 1.02, 1.01, 1.02, 1, 0.99, 1.01, 0.91, 0.83, 1.01

a) What is the mean growth rate of this sample of yeast lines?

**First**, review the formula for calculating the mean of a list of numbers.

$$mean = \frac{\sum_{i=1}^n Y_i}{n}$$

**Finally**, use this formula to calculate the mean

b) What is the median growth rate of this sample?

**First**, order your list of numbers from smallest to largest

**Second**, count how many observations are in your list.

**Finally**, based on the results from the previous step, determine what your median is.

Programing question:

**Question:** Construct a boxplot plotting the "growthexpt2" column against the "fertexpt2" column in the "fertilized\_block" data set.

**First**, make sure that you have ggplot2 installed by using the library() function.

**Second**, decide which column is your independent and dependent variable.

Independent:\_\_\_\_\_ Dependent: \_\_\_\_\_

**For you to think:** Can you explain how you made this distinction?

**Third**, construct your box plot using ggplot. Fill in the blanks in the following code to do so:

```
> ggplot(data =____, aes(x = _____, y = _____)) + _____()
```

### E.1.2 (ii) Reducing the intrinsic load

We reduced the element interactivity of the material by presenting (i) only one way to do one job and (ii) presenting only the functions that were needed for that assignment. In R, there are many synonymous ways to do the same task. For example, if we wanted to select one column out of a table these are the possible ways of doing it:

```
my_table$column1
my_table[,1]
select(my_table, column1)
```

We decided to choose only one option and consistently use it throughout the course. Additionally, in every assignment the students learned only a handful of new functions, but throughout the course they gained practice in a diverse assortment of functions. For example, they learned to input data into R, graph data, manipulate data and do multiple statistical tests.

### E.1.3 (iii) Increasing the germane load

Both in worked examples and in partially completed problems we asked the students to reflect on a part of the question to engage in germane load activities such as self-explaining.

Self-explanation questions:

**For you to think:** What kind of variable is "fertexpt2" and what kind of variable is "growthexpt2"?

## E.2 SURVEYS

### E.2.1 *Biostatistics*



## Students attitudes to programming in R for BIOL 300

### Page 1

What year of your undergraduate degree are you in?

- 1
- 2
- 3
- 4
- 5+

Before this course started...

	Excellent	Above Average	Average	Below Average	Very Poor
My programming skills (in any programming language) were	<input type="radio"/>				
My skills in any statistical software (JMP, SPSS, etc) were	<input type="radio"/>				

Before this course started I used R...

- Daily
- Weekly
- Monthly
- Yearly
- Never

Before this course started, I was interested in learning a programming language

- Strongly Agree
- Agree
- Neutral / Undecided
- Disagree
- Strongly Disagree

During this course I felt that the ...

	Strongly Agree	Agree	Undecided / Neutral	Disagree	Strongly Disagree	NA
statistics concepts were easy to understand (Normal distribution, t-test, ANOVA, regression)	<input type="radio"/>					
programming concepts in R were easy to understand (Data input, visualization, writing code)	<input type="radio"/>					
statistical software JMP was easy to understand	<input type="radio"/>					
statistical software (R or JMP) sessions were helpful to understand the statistics	<input type="radio"/>					

During this course, the level of difficulty of was...

The concept refers to the material you saw in the lecture and the application refers to doing the analysis in the software (either R or JMP)

	Too high	High	Right	Low	Too low
Probability (Concept)	<input type="radio"/>				
Probability (Application)	<input type="radio"/>				
Contingency analysis (Concept)	<input type="radio"/>				
Contingency analysis (Application)	<input type="radio"/>				
t-test (Concept)	<input type="radio"/>				
t-test (Application)	<input type="radio"/>				
ANOVA (Concept)	<input type="radio"/>				
ANOVA (Application)	<input type="radio"/>				

I felt ...

Mark all that apply

	Happy	Excited	Motivated	Supported	Overwhelmed	Anxious	Bored	Frustrated	Stressed	Angry	Annoyed	Proud	Scared	NA
While working on the conceptual parts of the statistics assignments	<input type="checkbox"/>													
While using the statistical software R	<input type="checkbox"/>													
While using the statistical software JMP	<input type="checkbox"/>													

I used ...

	Daily	Weekly	Monthly	Once in the term	Never
R outside of class	<input type="radio"/>				
JMP outside of class	<input type="radio"/>				
R in the laboratories	<input type="radio"/>				
JMP in the laboratories	<input type="radio"/>				

	Daily	Weekly	Monthly	Once in the term	Never
R in the lectures	<input type="radio"/>				
JMP in the lectures	<input type="radio"/>				

Having completed this course I would ...

	Extremely likely	Very likely	Moderately likely	Slightly likely	Not at all likely
rate my programming proficiency as high	<input type="radio"/>				
put the ability to use R as a skill on my CV	<input type="radio"/>				
continue using R in my own projects for my undergraduate or graduate school	<input type="radio"/>				

## Long answer questions

Do you have any other comments about the course?

Type here

If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

Type here

If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

Type here

E.2.2 *Eco-Methods*



## Students attitudes to programming in R for BIOL 404

### Page 1

What year of your undergraduate degree are you in?

- 1
- 2
- 3
- 4
- 5+

Who was your instructor

- Diane Srivastava
- Kurt Trzcinski

Before this course started...

	Excellent	Above Average	Average	Below Average	Very Poor
My programming skills (in any programming language) were	<input type="radio"/>				
My skills in any statistical software (JMP, SPSS, etc) were	<input type="radio"/>				

Before this course started I used R...

- Daily
- Weekly
- Monthly
- Yearly
- Never

Before this course started, I was interested in learning a programming language

- Strongly Agree
- Agree
- Neutral / Undecided
- Disagree
- Strongly Disagree

During this course I felt that the ...

	Strongly Agree	Agree	Undecided / Neutral	Disagree	Strongly Disagree	NA
statistics concepts were easy to understand (ANOVAS, regression, ordination, etc)	<input type="radio"/>					
programming concepts in R were easy to understand (Data input, visualization, writing code)	<input type="radio"/>					
statistical software JMP was easy to understand	<input type="radio"/>					
statistical software (R or JMP) sessions were helpful to understand the statistics	<input type="radio"/>					

During this course, the level of difficulty of was...

The concept refers to the material you saw in the lecture and the application refers to doing the analysis in the software (either R or JMP)

	Too high	High	Right	Low	Too low
Two way ANOVA (Concept)	<input type="radio"/>				
Two way ANOVA (Application)	<input type="radio"/>				
Randomized block ANOVA (Concept)	<input type="radio"/>				
Randomized block ANOVA (Application)	<input type="radio"/>				
Multiple regression (Concept)	<input type="radio"/>				
Multiple regression (Application)	<input type="radio"/>				
Ordination (Concept)	<input type="radio"/>				
Ordination (Application)	<input type="radio"/>				

I felt ...

Mark all that apply

	Happy	Excited	Motivated	Supported	Overwhelmed	Anxious	Bored	Frustrated	Stressed	Angry	Annoyed	Proud	Scared	NA
While working on the conceptual parts of the statistics assignments	<input type="checkbox"/>													
While using the statistical software R	<input type="checkbox"/>													
While using the statistical software JMP	<input type="checkbox"/>													

I used ...

	Daily	Weekly	Monthly	Once in the term	Never
R outside of class	<input type="radio"/>				
JMP outside of class	<input type="radio"/>				
R in the laboratories	<input type="radio"/>				
JMP in the laboratories	<input type="radio"/>				

	Daily	Weekly	Monthly	Once in the term	Never
R in the lectures	<input type="radio"/>				
JMP in the lectures	<input type="radio"/>				

Having completed this course I would ...

	Extremely likely	Very likely	Moderately likely	Slightly likely	Not at all likely
rate my programming proficiency as high	<input type="radio"/>				
put the ability to use R as a skill on my CV	<input type="radio"/>				
continue using R in my own projects for my undergraduate or graduate school	<input type="radio"/>				

## Long answer questions

If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

Type here

If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

Type here

Do you have any other comments about the course?

Type here

## E.3 CODEBOOKS

### E.3.1 *Biostatistics*

Biol 300 - Question: If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

Theme	Total responses	Child nodes	Keywords	Example answer	Control responses	Treatment responses
K - Keep some part the canvas R assignments	49	K1 - Step by step questions	Questions, step-by-step, assignments	"I like how they walked you through the questions almost step-by-step" [E3]	0	5
		K2 - Practice calculation/ walkthroughs	practice calculation, walkthrough, example, helpful	"I appreciated the walkthroughs- I learn by example so that was great." [E113]	0	18
		K3 - Informative and not overwhelming	informative, not overwhelming	"I love how informative the assignments are; not overwhelming to complete them." [E11]	0	4
		K4 - Detailed instructions/ intro	Instructions, intro, tutorial, guide	"The guide before each assignment. Very helpful." [E102]	0	12
		K5 - expected code/graph/values	code, expected graphs/values,	"Showing the code and the expected graphs/values to make sure you did it correctly in R." [E22]	0	3
		K6 - Fill in the blank	Fill in the blank	"I liked the fill-in-the-blanks especially the question with the expected graphs because I could test it out and it gave me some sense of support." [E74]	0	3
		K7 - Pictures	Pictures	"pictures were a great help." [E92]	0	1
		K8 - Assignments in general	questions, assignments, homework	"Assignments were helpful." [E95]	0	3
A - Keep the lab manual	39	A1 - It was clear and easy to understand	Lab manual, clear, understand, easy to follow	"Lab manual guides were easy to follow and understand." [C120]	9	2
		A2 - Detailed instructions and lots of examples	Examples, helpful, detailed, instructions	"Detailed instructions with lots of examples." [C110]	5	2
		A3 - Step-by-step	Step by step, lab manual	"I appreciate the step-by-step process outlined in the lab manual." [C38]	2	2
		A4 - Images (screen shots)	Screenshots, images, lab manual	"I think the images on the lab manuals for JMP instructions was very useful/helpful." [C5]	2	0
		A5 - Overview pages with formulas	summaries, formula, commands	"I liked the summaries of R commands." [E68]	0	4
		A6 - Liked the lab manual	lab manual, helpful, Keep	"The lab manual is a very useful way for new users to get started." [C28]	8	3
C - Keep the labs - students like being in the labs	15	C1 - How JMP was broken down in labs	JMP, labs	"I liked how JMP was broken down in the labs so that it became easy to use." [C149]	2	0
		C2 - Can work on JMP on computers in lab	computer, lab	"the computer lab space" [C143]	1	0
		C3 - No homework, purely practice	labs, marks	"I liked that the labs were straightforward and not for marks, it made it a very stress-free way to learn." [C49]	1	0
		C4 - Length was good	hours	"Two hours is great." [C18]	2	0
		C5 - Intro before labs	concepts, first, problems	"It was really useful to have the going over the concepts first then applying them to problems." [E57]	0	1
		C6 - Liked the labs	tutorial, lab, keep, teaching	"I liked that the labs were straightforward and not for marks, it made it a very stress-free way to learn." [C49]	6	2
N - Liked how R was taught	14	N1 - no changes	useful, keep, everything	"Everything! Taught well" [E73]	0	6
		N2 - Everything was broken down into small bite sized chunks	broken down, clear	"I liked that everything was broken down and explained to a very basic level; it made it very enjoyable to learn for someone who really struggles with computer programming." [E19]	0	2
		N3 - Examples	examples, code	"I really liked the examples. I would keep those." [E23]	0	2
		N4 - Application skills in questions	application, questions	"Good application skills put into questions." [E51]	0	1
		N5 - Going over concepts and then attempting problems	concepts, problems, example, question	"It was really useful to have the going over the concepts first then applying them to problems." [E57]	0	2
		N6 - well written and straightforward	straightforward, well written	"It was super straightforward and well written, despite being taught for the first time." [E71]	0	1
G - They did not like how JMP was taught	9	G1 - not useful	didn't like	"Nothing. Maybe we could learn a program that is still current in the field." [C6]	2	0
		G2 - Labs unhelpful and disorganized	didn't find useful	"I never went to labs as I didn't find them useful." [C97]	3	0
		G3 - Teach R instead or as well	teach, R	"JMP just seemed a bit dated. It would be nice if we were able to use R instead." [C61]	4	0
E - No changes to how JMP was taught	7	E1 - No changes	JMP, good	"JMP taught well, teaching of software and practicing using JMP was good. The lab manual was a little out of data w/ links but otherwise really good." [C34]	6	0
		E2 - Easy and straightforward	easy, straight-forward	"The easiness and straightforwardness." [C70]	1	0
		M1 - Nothing	Nothing	"Nothing at all." [E16]	0	1

M - Didn't like how R was taught	7	M2 - Wants more practical marked experience	practical, marks	"More practical [R] experience. Make it marked. Have a competition." [E10]	0	1
		M3 - Labs should be worth marks	grades, labs, marks	"Labs aren't worth marks, for correctness but we should get participation grades for attending." [E12]	0	1
		M4 - Wants reference command page	Reference, commands page	"Keep some key points for easier reference? Like the commands page in lab." [E20]	0	1
		M5 - More lessons on basics	basics, fundamentals	"more lessons on the basics" [E55]	0	2
		M6 - More class interaction	class, interaction	"More interaction in class." [E72]	0	1
		B - Ability to work on software and problems independently (ie from home)	6	B1 - Work from home	at home, online, independently	"I liked that it was mostly independently self-taught with the aid of the manual and TA." [C32]
B2 - Labs not mandatory	labs, optional, not mandatory	"I liked that labs were optional because if something conflicted you weren't stressed about it." [E15]		1	1	
D - Keep statistical software	6	D1 - Software makes calculations easier	calculate, homework	"Being able to use [software] for homework answers." C53	2	0
		D2 - Good for plotting graphs	plot, graph	"The way to quickly generate and analyze graphs." [C20]	2	0
		D3 - Put it on exams	exams	"If R was taught, then use R all the way, even for final exams." [E47]	0	2
F - TA/instructor assistance with stat software	5	F1 - TA walkthrough	TA, help, time, instructor, support	"Providing time to spend with an instructor who will walk you through it." [C52]	4	0
		F2 - Wish TAs would teach from lab manual before students attempt	teach, first, lab book	"We read the lab book and did it on our own; maybe teach it first (I can read it at home)" [C20]	1	0
H - JMP Activities/assignments	5	H1 - JMP Activities/assignments	Assignments, questions, keep, activities	"Keep the problems and questions." [C135]	5	0
O - Learning how to graph data with R	5	O1 - Learning how to graph data with R	ggplot, graphing	"I loved producing graphs- it felt like I had to process the information and explore." [E54]	0	5
L - Didn't like R assignments	4	L1 - Make instructions more concise	"instructions, concise" [E8]	Make the instructions more concise.	0	1
		L2 - Separate instruction doc	instructions, separate,	"Have the instructions written on a doc" [E62]	0	1
		L3 - Too easy	too easy	"I just think it was a bit too easy when [R] was in the week assignments" [E39]	0	1
		L4 - Link to previous assignment explanations	link, previous	"In [questions] link to previous lessons." [E55]	0	1
I - Use JMP in lecture	2	I1 - Use JMP in lecture	lecture, use, class, JMP	"Use it more during lecture to showcase its advantages/disadvantages." [C58]	2	0
J - Never used stat software	1	J1 - Never used stat software	Never, used	"Never used either" [C54]	1	0

Biol 300 Q2: Question: If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

Theme	Total responses	Child nodes	Keywords	Example answer	Control responses	Treatment responses
A - Course should use other software	47	A1 - R	R, teach, instead	"Learn more about R" [C95]	31	0
		A2 - Excel	Excel, teach, instead	"Excel should be used instead" [C134]	11	0
		A3 - other	other, teach, option, newer, not JMP	"Teach excel or a newer program" [C59]	5	0
N - The R assignments need improvement	29	N1 - Assignments need more clarity, better instructions	Confusing, understand, difficult	"Have hints or more thorough walk-throughs- sometimes I would do exactly what the instructions said and nothing outputted." [E24]	0	7
		N2 - R questions too easy	Unnecessary, assignments, easy, challenging	"I think the assignments shouldn't have the R walkthrough. It made them too easy and encouraged me not to go to lab." [E39]	0	4
		N3 - Disconnect between assignments and lecture	Compared, assignment, class	"Conceptually some of the assignments seemed off compared to what we saw in class" [E5]	0	1
		N4 - Didn't like help questions	Help question	"I feel like instead of having to debunk the code through the "Help" tab, it is more efficient if it's laid out because it should be an application question rather than trying to spend time trying to find it." [E7]	0	1
		N5 - Scrolling	Separate, document, scrolling	"Have a separate document instead of a walkthrough in the assignment- too much scrolling." [E22]	0	5
		N6 - Graphing questions were challenging	graphs, help, explain, graphing	"Explain how to deal with graphing questions more." [E49]	0	3
		N7 - assignments were too long	too long, instructions, assignments, quicker	"The questions were long and difficult, and sometimes way too many numbers needed to be typed in to do a test." [E87]	0	4
		N8 - Disconnect between assignments and lab	Assignment, lab, reflect	"Assignment R questions were not what was done in labs should reflect lab material" [E40]	0	1
		N9 - Too challenging	Difficult, easier, questions	"One change would be to make the R parts easier/quicker to do for a beginner (mostly)" [E92]	0	3
B - The students want more activities to help or force them to learn the software	23	B1 - Make learning it more interesting	Interesting more	"Make it more interesting" [C78]	1	0
		B2 - Show application of software	Application, program, use, implement	"How to use R in more applicable ways" [E79]	1	2
		B3 - Present exam style R questions	R, exam, tested, questions	"More exam style R questions on assignments" [E112]	0	3
		B4 - big project	project	"create a project to demonstrate our [R] skills" [E54]	0	1
		B5 - More practice software	More, practice, program, software, homework, R	"Make homework or assignments that test knowledge of the programs" [C52]	6	9
H - The course should provide an incentive to come to the labs to learn the statistical software	19	H1 - Make labs mandatory	labs, mandatory	"some of the labs should be made mandatory" [C74]	7	4
		H2 - Labs worth marks	labs, marks, marking scheme, quiz, attendance	"Maybe make labs mandatory and with some form of marking scheme or small quiz?" [C18]	3	1
		H3 - Force people to work together/in groups	group, partners, work, learning	"I would like to have worked in the [lab] portion in a group or in parnters" [C116]	1	1
		H4 - Provide incentive for labs	incentive, labs, forces, mandatory	"Just make there an incentive to go to labs and to learn [JMP]" [C129]	2	0
C - The course should provide more support learning the statistical software	16	C1 - Make learning more organized and structured	organized, sturctured, even spread	"I wish it could be an even spread between R material and lecture material" [E73]	1	1
		C2 - Help troubleshooting	troubleshooting, tips	"some troubleshooting tips for us" [E92]	0	1
		C3 - Demo/ better explanation of new commands/functions/ concepts	Demo, concepts	"Demo new R commands" [E44]	0	4
		C4 - Provide reading	provide, readings	"encourage further exploration by providing reading" [E54]	0	1
		C5 - More walkthroughs	explanation, functions, work	"better explanation of how functions work" [E86]	0	2
		C6 - Provide information on online resources	online, education, sites	"Maybe do R-online assignments [with] online education sites such as edx." [E104]	0	1
		C7 - More support	teach, TA, help, program, R	"Have the TA teach us how to use the program, rather than have us self-taught through the lab manual" [C147]	4	1
		I1 - Improve in general	labs, waste, improve	"Improve tutorial/labs" [E10]	3	1
		I2 - Integrate labs into lecture	labs, integrated, lecture	"[Labs] were not integrated into lecture" [C123]	1	0

		I3 - Too long	labs, shorter	"Making [labs] shorter and more helpful[.] We never finished a lab!" [C98]	1	0
		I4 - Review at beginning of lab	review, before, labs	"Provide a 15 minute review session before students begin working on the labs." [C12]	0	2
		I5 - Provide answers for lab questions	answer, key, questions, manual, assignments	"It would be useful to have an answer key for the lab assignments to be able to see if I did the questions correctly." [E85]	0	6
		I6 - More R activities	more, activites, R	"The labs could have more activities that enhance R" [E56]	0	1
I - Improve labs	16	I7 - More comprehensive	comprehensive, tutorials	"Have slightly more comprehensive tutorials [near] the end." [E95]	0	1
P - Liked how R was taught	13	P1 - Liked how R was taught	Understand, taught, good, clear, R	"No changes! It was easy to understand and easy to complete. Thanks for making it doable." [E58]	0	13
D - Integrate statistical software into lecture	11	D1 - Integrate statistical software into lecture	Integrate, lecture, software, class, R, JMP, demo, teach	"Teach R in lectures as well as lab[.] Maybe cover concepts simultaneously in written and software based formats" [C113]	10	1
G - JMP was a waste of time/take out	10	G1 - JMP was a waste of time/take out	waste of time, JMP, didn't learn, Don't teach	"I would get rid of JMP and use the more useful and universal R." [C51]	10	0
F - Liked how JMP was taught/keep it	9	F1 - Liked how JMP was taught/keep it	Straightforward, teach, JMP, fine	"JMP is excellent as it is" [C106]	9	0
		O1 - Instructions were very clear	Clear, help, instructions	"All the instructions on how to use R for the assignments was very clear." [E13]	0	6
O - Liked R assignments	7	O2 - Visualization of code	Visualization	"The visualization part was really good." [E101]	0	1
L - Wants online video tutorials	5	L1 - Wants online video tutorials	online, tutorial, video	"I find video tutorials of anything is more useful/effective than reading a bunch of instructions." [E8]	2	3
Q - Provide list of commands	5	Q1 - Provide list of commands	list, command, cheat sheet	"Provide an abbreviated list of all commands used in the course." [E27]	0	5
M - Ensure TAs understand statistical software	4	M1 - Ensure TAs understand statistical software	better training, TA, understanding, program	"Ensure that the TAs actually have a thorough understanding of the program before they teach it to us." [C6]	2	2
J - Make stastical software mandatory	3	J1 - Make stastical software mandatory	software mandatory	"For statistical software to be mandatory" [C110]	3	0
K - Liked lab manual	3	K1- Liked lab manual	lab manual, helpful, guide	"Following a guide and doing things along with the guide helped familiarize myself with the program" [C19]	3	0
E - Doesn't want statistical software taught	1	E1 - Doesn't want statistical software taught	Remove	"Remove [R] entirely" [E16]	0	2
R - Liked labs	1	R1 - Liked labs	Liked, labs, useful	"I really liked the labs. I felt that the TAs were really helpful in explaining R and how to use it." [E41]	0	1

E.3.2 *Eco-methods*

Biol 404 - Question: If you could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

Theme	Total responses	Child nodes	Keywords	Example answer	Control responses	Treatment responses
C - R was taught well	21	C1 - R was taught well	R, taught	"Overall, I liked the way R was taught and that we were provided with a lot of reasons behind the code we were typing so we could use them going forward." [D19]	0	1
		C2 - Liked examples and clarification in R	code, examples, clarification, hints	"Like how assignments made sure your code was right and gave hints to if you were on the right track" [D14]	2	1
		C3 - Liked having code to go along with lectures	lectures, code, work-alongs	"Including code to follow along with lectures was helpful for learning new concepts." [B14]	2	0
		C4 - Explanations for each question	Explanations, question, assignments	"Including code to follow along with lectures was helpful for learning new concepts." [D1]	0	1
		C5 - Explanations for each element of the code	explanation, code, R, element, part	"Present R code, explaining what each element does" [D2]	0	7
		C6 - Had lots of opportunities for practice	practice	"The quizzes provided lots of opportunity for practice" [D4]	0	1
		C7 - Fill in the blank questions	fill in the blank	"The fill in the blank questions were helpful in figuring stuff out" [D16]	0	3
		C8 - Liked instructions in the quiz	instructions, quiz	"The instructions on the quizzes were very helpful" [D11]	0	2
		C9 - Liked hints	hints, quizz	"Like how assignments made sure your code was right and gave hints to if you were on the right track" [D14]	0	1
E - Liked having an R workshop	19	E1 - Liked having an R workshop	R, workshops	"Keep the R-script tutorials. Very helpful." [B19]	6	0
		E2 - Script was step by step/ having a script	code, step by step, example	"Walking through programming techniques step by step was valuable" [B8]	4	0
		E3 - Workshop during lab hours	Workshop, lab	"First introductory tutorial allowed for smooth entry into R" [D6]	5	2
		E4 - Step by step instructions	instructions, guided, walk through	"I really liked how the instructions walked us through the process so it was less overwhelming" [D7]	0	2
I - Like the stats assignments	16	I1 - Like the stats assignments	assignments, understanding, learn, explain, quiz, practice, R	"Liked the way the R assignments guided us through the steps and showed us how to do the parts necessary" [D21]	2	14
F - The students liked learning something useful	5	F1 - They learned something they wanted to learn	skill, wanted, learn	"The course forced me to learn a skill I wanted to learn, but I didn't know where to start." [B8]	1	0
		F2 - They learned something that has practical usage	practical, realistic, going forward	"Overall, I liked the way R was taught and that we were provided with a lot of reasons behind the code we were typing so we could use them going forward." [D19]	1	2
		F3 - They learned to simplify code	simpler, code	"Learning about the different functions and how they can be used to make code simpler" [D10]	0	1
A - Learning packages/analyses/ functions were useful	4	A1 - Variety of things they found useful	functions, code	"Learning about the different functions and how they can be used to make code simpler" [D10]	1	2
		A2 - Anova	ANOVA, R	"I think the part of the course on ANOVAs and how to do them in R made sense." [B9]	1	0
G - Like the tutorials	3	G1 - Like the tutorials	Tutorial	"The tutorials were very helpful because the teaching was more personal and more go-at-your-own-pace." [B10]	3	0
H - Like the lectures	3	H1 - Like the lectures	lectures, notes	"the class lectures and diagrams were very (underlined) helpful" [B20]	2	0
		H2 - Like the diagrams in lecture	diagrams, lecture	"the class lectures and diagrams were very (underlined) helpful" [B20]	1	0
D - Liked the labs	2	D1 - Liked the labs	lab, useful	"Lab sessions were useful" [B23]	2	0
B - Questions were useful	1	B1 - Question selection and order was useful	Selection, questions, order	"A lot of thought was obviously put into selection of questions and order. Generally [] this was done well in a way that help build an understanding." [D9]	0	1
J - Assignments were a good level of difficulty	1	J1 - Assignments were a good level of difficulty	difficulty, level, assignments	"Stats assignments were pretty helpful and good level of difficulty" [D12]	0	1

Biol 404 - Question: If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

Theme	Total responses	Child nodes	Keywords	Example answer	Control responses	Treatment responses
B - More R instruction on functions and packages	28	B1 - More instructions	functions, teach, code, statistical analysis	"Go into the details of how to actually write the codes and what items/functions mean. When introducing concepts such as regression or ordination plots have an R lab to accompany it." [B6]	3	1
		B2 - Instructions on the code/output	code, output, feedback, explanation, write, structure	"I think the lectures had too much random code without adequate explanation about what parts of the code were referring to or even how to interpret the results. Most of what I learned about R was reading blogs online or watching videos outside of class to teach myself." [B9]	7	2
		B3 - Provide csv files/ R files to follow in class	R data, csv files, example, class	"Please also give us the companion csv files so we can follow along lectures." [B10]	3	0
		B4 - Have a referece/ faq/guide	overview, reference guide, document, FAQs, resources	"However, it would be nice if the explanations of the R inpus were put together in a short manual to reference throughout the term opposed to get at the beggining of the R quizzes or in the lab manuals" [D28]	3	7
		B5 - More practical examples	practical, apply, code	"More practical use [of R] during labs and course" [B4]	1	1
E - More synchrony between the assignments, the lectures and labs	13	E1 - Spend time doing/explaining R in lectures/labs	teaching, R, labs, lecture, more time	"Include more R components to the lecture as well, in addition to the labs" [D3]	1	4
		E2 - Teachers/TA need to know assignments in greater detail	Instructors, TA, assignments	"Instructors and TA need to understand assignment better (they know how to code, but were often unprepared to answer questions specific to quizzes)" [D6]	0	1
		E3 - More synchrony between R and lectures	R, lectures, concepts, code, theory	"Do better job of linking the theory with the statistics produced in R." [B22]	5	1
		E4 - More synchrony between labs and assignments	codes, lab,	"I think that basic graphs and codes should be introduced early to assist with lab reports and there should be a reference page or something to refer to." [D23]	0	1
C - More help outside of class	11	C1 - More R homework	more, assignments, exercises, R	"Please give us exercises for R; the stats assignments were overwhelming." [B10]	3	0
		C2 - More tutorials	tutorials, code, more, workshop	"More explanations for tutorials, go through [R] with template data together." [B5]	7	0
		C3 - R specific office hours	Office hours, R	"Add R office hours" [D8]	0	1
A - More R	9	A1 - Generally more R	More, code	"I would add more on how to make the code faster to type. i.e. loops, learn plyr. -These could just be provided as examples and then people can figure out how they work if their interested" [D10]	0	1
		A2 - Want R earlier in the course	more, R, introduction, start,	"A more thorough introduction of the basics of the course/R would have been valuable at the start" [B8]	2	0
		A3 - Want more teaching in R generally	Overview, teach, R, code	"I would have liked to given lesson on how to use R and told what the different codes were" [D25]	4	2
G - R teaching in general not good enough	8	G1 - Stats assignments were overwhelming	stats assignments, overwhelming, challenging	"Change the stats assignments. They were too challenging and I didn't feel like I learned anything from them." [B19]	2	0
		G2 - Explanations unclear and insufficient	explanation, unclear, unhelpful, code, discouraging	"The explanations given in lecture were unclear, insufficient and, simply put, frustrating." [B17]	3	1
		G3 - Need to break down steps	Break it down, step by step, smaller sections	"Break it down more, step by step. Rather than give lines of code for a concept, break down how to write the code (what does a comma, or bracket mean?)." [B18]	2	0
K - Assignments were not clear	7	K1 - Assignments generally not clear	Assignments, do not make sense, questions	"Assignments need to be more clarified. Some questions do not make sense" [D2]	0	1
		K2 - R help file questions not clear	help file, question	"Simpler concepts, more breakdowns of what exactly each function is doing instead of "read the help file"." [D14]	0	5
		K3 - Multiple choice questions not clear	multiple choice, questions	"A lot of the multiple choice questions felt more like careful reading and interpretation exercises and not R competence questions, they weren't super clear and were way too wordy" [D5]	0	2
F - Too fast	6	F1 - Generally too fast	quickly, slower	"Maybe go slower because I had no previous experience and was lost" [D27]	1	1
		F2 - Professor was too fast	too fast,	"Please just slow down the teaching pace so everything can be digested" B10	1	0
		F3 - R workshops were too fast	rushed, slower, workshop	"Some assignments felt a bit rushed, and because of that felt like was unable to comprehend all the material. Slower pace would be great" [D8]	2	1

H - Not enough support	5	H1 - Not enough support	teach, support, R, guidance	"Actually teach it! I felt like we were not taught a thing in class. Any skills I gained were from the online workshops or google. I didn't feel like I had any support." [B16]	4	1
N - Layout in canvas not great	5	N1 - Layout in canvas not great	scroll, assignment instructions, canvas, layout	"The assignment on canvas made us scroll up to see the question and down to answer it which got confusing" [D18]	0	5
L - Assignments not challenging enough	4	L1 - Assignments not challenging enough	assignments, independent thought, figure out	"Let students figure out how to do things on their own more without always giving step by step instructions. Id learn more if I was forced to figure things out on my own" [D17]	0	3
		L2 - Not sure if retained info	retained, assignments	"Some assignments walke us through it but I'm not sure how much I retained" [D18]	0	1
M - Labs unclear	4	M1 - Need to have correct code for labs	incorrect code, labs, assignments	"Ensure that all the code given to students works, sometimes we received incorrect code which made assignments and labs difficult to understand" [D20]	0	3
		M2 - Labs overwhelming	labs, overwhelming	"More intro- the beginning was overwhelming because there wasnt enough time to figure things out (ie labs took too long)." [D21]	0	1
I - More lecture material	3	I1 - More stats explanation	clarification, method, statistics	"Learn exactly what each statistics means better" [B27]	2	1
D - Wanted R in prereq course (BIOL300)	1	D1 - Wanted R in prereq course (BIOL300)	learn, earlier, R	"As someone who had no programming knowledge before this, basic programming skills and logic would have been helpful to learn earlier on." [D16]	0	1
J - More feedback	1	J1 - More feedback	feedback	"Provide more feedback (which stats when following assignment)" [B3]	1	0

## E.4 SUPPORTING RESULTS

TABLE E.1: The perceived difficulty of the material did not change based on the format (application or concept), nor the treatment, nor the concepts learnt.  $X^2$  and P-value of the  $X^2$  test for each main effect and interaction.

Perceived difficulty of the material	Biostatistics			Eco-Methods		
	$X^2$	DF	P	$X^2$	DF	P
Format	0.24	1	0.62	0.006	1	0.93
Course	0.74	1	0.39	0.00	1	1.00
Material	0.32	3	0.95	1.60	3	0.66
Format:Course	0.007	1	0.93	0.11	1	0.73
Format:Material	0.21	3	0.97	0.24	3	0.97
Course:Material	2.68	3	0.44	0.30	3	0.96
Format:Course:Material	0.22	3	0.97	0.14	3	0.99