

Macroevolutionary Consequences of Trophic Specialization

**A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

**Samuel Robert Borstein
May 2019**

Copyright © 2019 by Samuel Robert Borstein
All rights reserved.

DEDICATION

I dedicate this work to my friends and family who have supported me all these years.

ACKNOWLEDGEMENTS

I am grateful for the many people who have directly and indirectly aided in my growth as both a person and researcher. I would first like to thank my advisor, Brian O'Meara, who has provided excellent guidance in preparing me for a career in research. I would also like to thank my committee members, Ben Fitzpatrick, Jim Fordyce, and Ben Keck for their support throughout my doctoral career and providing me with many interesting and enlightening conversations. In addition to these people, I want to acknowledge Jeremy Beaulieu, Sandy Kawano, Xingli Giam, Joe Bailey, Jenn Schweitzer, David Etnier, and Nina Fefferman for providing advice and intriguing discussions during my graduate career. I am indebted to my collaborators who have aided me in my research over the years. I want to thank Peter Wainwright and Ron Coleman for providing me initial research experience. It is also important to acknowledge Matthew McGee, Jonathan Chang, Mike Alfaro, Ole Seehausen, and Liam Mueller.

I would like to thank the various funders of my research, including the University of Tennessee, National Science Foundation, American Museum of Natural History, Society of Systematic Biologists, Society for Integrative and Comparative Biology, the American Cichlid Association, and the Ohio Cichlid Association.

Last but hardly least, I want to thank my family and friends who have supported me during my graduate career and made my experience in Knoxville so much more enjoyable. I am especially grateful to (in no particular order) Chloe Lash, Katie Kulhavy, Zachary Marion, Greg Alibone, John Reese, Christopher Peterson, Sharon Munoz, Chelsae Miller, Orlando Schwery, Todd Pierson, Tyson Paulson, Erica Casey, Franklin Finney, Harmony Yomai, Athma Senthilnathan, Lacy Chick, Michael Wilson, Leigh Moorehead, Taylor Woods, Wade Wilmoth, Liam Mueller, Jessica Dryer, Ben Holt, Domo Hatton, Rachel Wooliver, & Rachel Fovargue.

ABSTRACT

One of the central issues limiting progress towards a generalized theory of biological organization involves integrating the interplay of current ecological conditions with long-term macroevolutionary dynamics. Trophic ecology has long been known to be a strong force in shaping biological diversity for all organisms, with its central tenet being the procurement of resources for survival. Though trophic ecology is well known to effect evolutionary trajectories, few studies have explicitly tested hypotheses related to the effects of diet on macroevolutionary patterns at deep phylogenetic scales. Here, I investigate the interaction between trophic ecology and macroevolution using acanthomorph fishes as a model.

My first two chapters introduce R packages that were used to conduct the research in chapters three and four. In my first chapter I introduce a new R, *AnnotationBustR*, which extracts sequences from GenBank annotations. The second chapter highlights *dietr*, an R package that calculates fractional trophic levels from quantitative and qualitative diet data. Chapter three investigates the effect of trophic level and diet breadth in promoting or constraining phenotypic evolution and functional diversity in coral reef acanthomorphs. My final chapter test the macroevolutionary consequences of an evolutionary innovation related to trophic resource use, pharyngognathy, testing if it does in fact increase diversification rates and promote trophic evolution as previously hypothesized. Overall, my dissertation highlights how trophic specialists may not be constrained in their morphology and the importance of adequately testing proposed evolutionary innovations using phylogenetic comparative methods.

TABLE OF CONTENTS

Introduction.....	1
Background.....	1
Acanthomorph fishes and rift lake cichlids as a system to study the consequences of trophic ecology.....	2
The use of trophic ecomorphology of fishes to study macroevolutionary patterns.....	2
Evolutionary Innovations as Drivers of Diversity	3
References.....	4
Chapter 1 <i>AnnotationBustR</i> : an R package to extract subsequences from GenBank annotations..	8
Abstract.....	9
Introduction.....	9
Implementation	11
Examples.....	15
Conclusions.....	17
Citation.....	18
Acknowledgements.....	18
References.....	19
Chapter 2 <i>dietr</i> : An R package for calculating fractional trophic levels from quantitative diet and qualitative food item data.....	23
Abstract.....	24
Introduction.....	24
Implementation	25
Methods.....	27
Results.....	28
Discussion.....	31
References.....	32
Chapter 3 Reef fish functional traits evolve fastest at trophic extremes.....	34
Abstract.....	35
Introduction.....	35
Methods.....	37
Trophic Level Data Acquisition.....	37
Phylogenetic Reconstruction	37
Morphological Data Acquisition.....	37
Measuring Rates of Trait Evolution.....	38
Diet Breadth Analysis	38
Functional Diversity Analysis.....	39
Results.....	40
Trophic Level Distribution	40
Morphospace.....	40
Rates of Morphological Evolution.....	40
Diet Breadth	44
Multivariate Functional Diversity.....	44
Discussion.....	47

Data Availability	50
References	51
Chapter 4 A classic evolutionary innovation, pharyngognathy, does not lead to increased rates of diversification	58
Abstract	59
Introduction	59
Methods	61
Phylogenetic Reconstruction	61
Diversification Analyses	62
Dietary Data Acquisition and Rates of Trophic Level and Diet Breadth Evolution	63
Transitions to Piscivorous and Processing Intensive Diets	63
Results & Discussion	64
References	70
Conclusion	74
Vita	75

LIST OF TABLES

Table 1.1: Functions and data included in the package <i>AnnotationBustR</i>	13
Table 2.1: Functions and datasets available in the <i>diatr</i> R package.	26
Table 3.1: Rates of morphological evolution for five trophic levels of 1545 species of coral reef fishes.	43

LIST OF FIGURES

Figure 1.1: Flow chart of functions for a complete usage of <i>AnnotationBustR</i>	13
Figure 1.2: Timings of subsequence extraction using AnnotationBust for thirteen metazoan mitochondrial coding sequences (black), thirteen chloroplast coding sequences (green), and five metazoan rDNA subsequence (purple).....	14
Figure 1.3: Phylogeny of 60 North American leuciscine minnows from sequences extracted using <i>AnnotationBustR</i> (left) and from individual GenBank sequences (right).	16
Figure 2.1: Correlation between trophic levels obtained from FishBase (x-axis in both plots) and those estimated by <i>dietr</i>	29
Figure 2.2: Trophic levels of <i>Herichthys minckleyi</i> calculated using <i>dietr</i> from data in Magalhaes et al. (2015).....	30
Figure 3.1: Stochastic character mapping reveals over six hundred trophic level transitions across the phylogeny of 1545 acanthomorph reef taxa.....	41
Figure 3.2: Morphospace by trophic level of 1545 reef acanthomorphs based on eight phenotypic functional traits..	42
Figure 3.3: Violin plot of diet breadth (left) and phenotypic functional diversity (right) at each trophic level..	45
Figure 3.4: Summary of multiple pairwise comparisons for rates of morphological evolution, functional diversity, and diet breadth amongst trophic levels. Trophic levels are denoted by a single letter: H=herbivore, O=omnivore, L= low-level predator, M=mid-level predator, and T=top-level predator.	46
Figure 4.1: Net Diversification rates (speciation-extinction) across a phylogeny of 1,184 percomorph fishes.....	65
Figure 4.2: Turnover (speciation+extinction) across a phylogeny of 1,184 percomorph fishes..	66
Figure 4.3: Continuous character maps of trophic level (left) and diet breadth (right) across 2859 percomorph fishes.....	67
Figure 4.4: Transition network map of processing intensive diets (left) and piscivorous diets (right).	68

LIST OF ATTACHMENTS

Chapter 1 Supplemental Material. Supplemental data.....	Chapter_1_Supplement.zip
Chapter 3 Supplemental Material. Supplemental data, methods, results, figures, and tables.....	Chapter_3_Supplement.zip
Chapter 4 Supplemental Material. Supplemental data, methods, results, and tables.....	Chapter_4_Supplement.zip

INTRODUCTION

Background

Determining the factors that generate and maintain the diversity of life is a key question in biology. Ecology shapes divergence in a variety of traits and behaviors, acting as a strong selective force (Dawkins & Krebs 1979; Schluter 2000b). The link between ecology and morphological form and function has been noted extensively in biology (Macarthur & Levins 1964; Bock & Von Wahlert 1965; Fryer 1968; Goodman *et al.* 2008). As performance is the basis upon which selection acts, morphology is important for understanding selective forces that shape biological diversity, acting as the point of interaction between an organism and its surrounding environment, directly impacting functional performance (Arnold 1983; Norberg & Rayner 1987; Wainwright 2007; Langerhans & Reznick 2010). For instance, it is quite common for changes in the performance of an organism to coincide with a morphological shift, which can be driven by a variety of ecological factors such as competition, a change in habitat with different conditions, or a transition to a novel niche (Streelman & Danley 2003; Goodman *et al.* 2008; Nosil 2012). However, certain morphological shifts to an adaptive peak may have consequences and constrain possible evolutionary trajectories (Collar *et al.* 2009; Collar *et al.* 2010; Smith *et al.* 2015).

Trophic ecology has long been known to be a strong force in shaping biological diversity for all organisms, with its central tenet being the procurement of resources for survival (Bentzen & McPhail 1984; Friman *et al.* 2008). Diet variation is common at nearly every biological level and can vary among species, populations, and even individuals. Understanding patterns of diet variation is fundamental to elucidating patterns of macroevolution that may be due to ecological speciation (Futuyma & Moreno 1988; Bolnick *et al.* 2003; Cohen *et al.* 2005; Devictor *et al.* 2010). Aspects of trophic ecology, such as predator-prey interactions can influence behavioral and morphological traits and change evolutionary trajectories (Vermeij 1977; Palkovacs *et al.* 2011). For example, the morphological traits and foraging behaviors of a species directly impact how it interacts with the surrounding environment. These interactions are important in governing a species performance, playing a major role in resource allocation (Schluter 2000a; Wainwright 2007).

For my dissertation, I investigated the macroevolutionary consequences of trophic ecology within one of the most diverse vertebrate radiations, acanthomorph fishes. Specifically, I investigate how trophic specialization effects rates of phenotypic evolution and how a proposed evolutionary innovation, pharyngognathy effects both the diversification and rates of trophic evolution in percomorphs. In addition to this, I discuss R packages I developed to perform this research that extract sequence data from GenBank and calculate fractional trophic levels. My research highlights the use of using phylogenetic comparative methods in conjunction with trophic data to test ecological hypotheses and the need for adequate testing of evolutionary innovations

Acanthomorph fishes and rift lake cichlids as a system to study the consequences of trophic ecology

Acanthomorph fishes represent one of the most speciose vertebrate radiations and includes over 14,000 species (Near *et al.* 2013). They are extremely diverse in form and function as well as ecology. Coral reefs are often dominated by acanthomorph fishes (Nelson *et al.* 2016). The dominance on coral reefs is likely due in part to these being the most productive nutrient rich environments on the planet which boast some of the most diverse assemblages of fishes (Fraser & Currie 1996). Coral reefs are widely known to promote ecological and morphological novelty in acanthomorph fishes and increase rates of both diversification and morphological disparity (Bellwood *et al.* 2006; Cowman *et al.* 2009; Price *et al.* 2011). This is likely due to the highly complex structure of coral reefs produced by the calcium carbonate skeletons of corals which allow a vast array of microhabitats, supporting a diverse array of prey species (Gratwicke & Speight 2005a, b). This habitat complexity is extremely important in creating heterogeneous microhabitats and resource patchiness which provide opportunities for ecological specialization and the promotion of morphological novelty (Price *et al.* 2011; Price *et al.* 2013; Price *et al.* 2015).

Additional acanthomorph radiations are also model systems for the study of evolution. These include families such as the cichlid fishes of East Africa (Seehausen 2006), the ice fishes (Notothenoidei) of the Antarctic Ocean, Lake Matano Silversides (Telmatherinidae) (Pfaender *et al.* 2010), and many others. These radiations show a dazzling array of ecological novelties that have evolved rapidly in short periods of time.

The use of trophic ecomorphology of fishes to study macroevolutionary patterns

There is a close relationship between morphology and function of fishes. Fish morphology is regularly used to predict ecological function and performance (Wainwright & Richard 1995; Carroll *et al.* 2004; Wainwright *et al.* 2007). While some traits, like craniofacial morphology, are directly responsible for the ingestion of prey and have been extensively researched, most other morphological aspects of a fish's body are directly implicated in locomotion, which impacts habitat choice and prey selection (Langerhans 2008; Holzman *et al.* 2012; Higham *et al.* 2015). Almost all prey acquisition strategies by fishes can be broadly categorized into three functional categories: biting, suction feeding, or ram feeding. Biting is when fish use their oral jaws directly in contact with prey to dislodge and ingest them while suction feeders generate suction pressure by buccal expansion to obtain prey, and ram feeders propel themselves into the prey for capture (Wainwright *et al.* 2015; Longo *et al.* 2016; McGee *et al.* 2016). These strategies make comparing trophic ecomorphology applicable and comparable at a wide level for fishes, such as acanthomorph fishes, as phenotypic traits for these strategies have been documented extensively.

Evolutionary Innovations as Drivers of Diversity

Evolutionary innovations are traits that provide species access to novel niches. These traits, sometimes referred to in the literature as “key innovations” are thought to impart an advantage in the clades which possess them and drive diversification (Simpson 1944; Mayr 1963). While the exact definition of an evolutionary innovation is debated, the vast consensus across biologists is that they are thought to be integral to the evolutionary success of organisms (Hunter 1998). While historically viewed in the context of increasing speciation or decreasing extinction risk, evolutionary innovations may promote diversity in other ways without directly effecting net diversification, such as through promoting ecological diversification (Futuyma & Moreno 1988; Rabosky 2017). While this would be beneficial to test, it is rarely done and much literature focuses on the effects of innovations on net diversification alone (Rabosky 2017). One difficulty to testing the impact of evolutionary innovations is that singular origins of an innovation prevent issues in a phylogenetic context as ideally multiple independent transitions into the innovation provide ways around issues of phylogenetic pseudoreplication (Maddison & FitzJohn 2015). While some traits have historic usage in the literature as an evolutionary innovation (Wainwright *et al.* 2012), they remain to be tested using modern day phylogenetic comparative methods.

References

- Albertson, R.C., Markert, J.A., Danley, P.D. & Kocher, T.D. (1999). Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 5107-5110.
- Arnegard, M.E., Markert, J.A., Danley, P.D., Stauffer, J.R., Ambali, A.J. & Kocher, T.D. (1999). Population Structure and Colour Variation of the Cichlid Fish *Labeotropheus fuelleborni* Ahl along a Recently Formed Archipelago of Rocky Habitat Patches in Southern Lake Malawi. *Proceedings of the Royal Society B: Biological Sciences* 266, 119-130.
- Arnold, S.J. (1983). Morphology, performance and fitness. *Am Zool*, 23, 347-361.
- Bellwood, D.R., Wainwright, P.C., Fulton, C.J. & Hoey, A.S. (2006). Functional versatility supports coral reef biodiversity. *P Roy Soc B-Biol Sci*, 273, 101-107.
- Bentzen, P. & McPhail, J. (1984). Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): specialization for alternative trophic niches in the Enos Lake species pair. *Canadian Journal of Zoology*, 62, 2280-2286.
- Bock, W.J. & Von Wahlert, G. (1965). Adaptation and the form-function complex. *Evolution; international journal of organic evolution*, 269-299.
- Bolnick, D.I., Svanback, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D. *et al.* (2003). The Ecology of Individuals: Incidence and Implications of Individual Specialization. *American Naturalist*, 161, 1-28.
- Burress, E.D. (2014). Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia*, 748, 7-27.
- Carroll, A.M., Wainwright, P.C., Huskey, S.H., Collar, D.C. & Turingan, R.G. (2004). Morphology predicts suction feeding performance in centrarchid fishes. *The Journal of experimental biology*, 207, 3873-3881.
- Cohen, A.E., Hendrickson, D.A., Parmesan, C. & Marks, J.C. (2005). Habitat segregation among trophic morphs of the Cuatro Ciénegas Cichlid (*Herichthys minckleyi*). *Hidrobiologica*, 15, 169-181.
- Collar, D., Schulte, J., O'Meara, B. & Losos, J. (2010). Habitat use affects morphological diversification in dragon lizards. *Journal of evolutionary biology*, 23, 1033-1049.
- Collar, D.C., O'Meara, B.C., Wainwright, P.C. & Near, T.J. (2009). Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution; international journal of organic evolution*, 63, 1557-1573.
- Cowman, P.F., Bellwood, D.R. & van Herwerden, L. (2009). Dating the evolutionary origins of wrasse lineages (Labridae) and the rise of trophic novelty on coral reefs. *Molecular phylogenetics and evolution*, 52, 621-631.
- Dawkins, R. & Krebs, J.R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London B: Biological Sciences*, 205, 489-511.
- Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W. *et al.* (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology*, 47, 15-25.
- Fraser, R.H. & Currie, D.J. (1996). The species richness-energy hypothesis in a system where historical factors are thought to prevail: Coral reefs. *American Naturalist*, 148, 138-159.

- Friman, V.-P., Hiltunen, T., Laakso, J. & Kaitala, V. (2008). Availability of prey resources drives evolution of predator–prey interaction. *Proceedings of the Royal Society of London B: Biological Sciences*, 275, 1625-1633.
- Fryer, G. (1968). Evolution and adaptive radiation in the Chydoridae (Crustacea: Cladocera): a study in comparative functional morphology and ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 254, 221-384.
- Fryer, G. & Iles, T.D. (1972). *The cichlid fishes of the great lakes of Africa : their biology and evolution*. Oliver and Boyd, Edinburgh.
- Futuyma, D.J. & Moreno, G. (1988). The Evolution of Ecological Specialization. *Annu Rev Ecol Syst*, 19, 207-233.
- Goodman, B.A., Miles, D.B. & Schwarzkopf, L. (2008). Life on the rocks: habitat use drives morphological and performance evolution in lizards. *Ecology*, 89, 3462-3471.
- Gratwicke, B. & Speight, M.R. (2005a). Effects of habitat complexity on Caribbean marine fish assemblages. *Marine Ecology Progress Series*, 292, 301-310.
- Gratwicke, B. & Speight, M.R. (2005b). The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of Fish Biology*, 66, 650-667.
- Higham, T.E., Stewart, W.J. & Wainwright, P.C. (2015). Turbulence, temperature, and turbidity: the ecomechanics of predator–prey interactions in fishes. *Integr Comp Biol*, 55, 6-20.
- Holzman, R., Collar, D.C., Mehta, R.S. & Wainwright, P.C. (2012). An integrative modeling approach to elucidate suction-feeding performance. *The Journal of experimental biology*, 215, 1-13.
- Hulsey, C.D., Mims, M.C., Parnell, N.F. & Streelman, J.T. (2010). Comparative rates of lower jaw diversification in cichlid adaptive radiations. *Journal of evolutionary biology*, 23, 1456-1467.
- Joyce, D.A., Lunt, D.H., Genner, M.J., Turner, G.F., Bills, R. & Seehausen, O. (2011). Repeated colonization and hybridization in Lake Malawi cichlids. *Current biology : CB*, 21, R108-109.
- Kocher, T.D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. *Nat Rev Genet*, 5, 288-298.
- Konings, A. (2015). *Tanganyika cichlids in their natural habitat*. 3rd edn. Cichlid Press, El Paso, TX.
- Konings, A. (2016). *Malawi cichlids in their natural habitat*. 5th edn. Cichlid Press, El Paso, TX.
- Langerhans, R.B. (2008). Predictability of phenotypic differentiation across flow regimes in fishes. *Integr Comp Biol*, 48, 750-768.
- Langerhans, R.B. & Reznick, D.N. (2010). Ecology and evolution of swimming performance in fishes: predicting evolution with biomechanics. *Fish locomotion: An etho-ecological perspective*, 200-248.
- Longo, S.J., McGee, M.D., Oufiero, C.E., Waltzek, T.B. & Wainwright, P.C. (2016). Body ram, not suction, is the primary axis of suction-feeding diversity in spiny-rayed fishes. *The Journal of experimental biology*, 219, 119-128.

- Macarthur, R.H. & Levins, R. (1964). Competition, habitat selection, and character displacement in a patchy environment. *Proceedings of the National Academy of Sciences of the United States of America*, 51, 1207-1210.
- McGee, M.D., Faircloth, B.C., Borstein, S.R., Zheng, J., Darrin Hulsey, C., Wainwright, P.C. *et al.* (2016). Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. *Proceedings of the Royal Society B: Biological Sciences*, 283.
- Meyer, B.S., Matschiner, M. & Salzburger, W. (2016). Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic biology*.
- Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L. *et al.* (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 12738-12743.
- Nelson, J.S., Grande, T.C. & Wilson, M.V. (2016). *Fishes of the World*. John Wiley & Sons.
- Norberg, U.M. & Rayner, J.M. (1987). Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 316, 335-427.
- Nosil, P. (2012). *Ecological speciation*. Oxford University Press.
- Palkovacs, E.P., Wasserman, B.A. & Kinnison, M.T. (2011). Eco-evolutionary trophic dynamics: loss of top predators drives trophic evolution and ecology of prey. *PloS one*, 6, e18879.
- Price, S.A., Claverie, T., Near, T.J. & Wainwright, P.C. (2015). Phylogenetic insights into the history and diversification of fishes on reefs. *Coral Reefs*, 34, 997-1009.
- Price, S.A., Holzman, R., Near, T.J. & Wainwright, P.C. (2011). Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecol Lett*, 14, 462-469.
- Price, S.A., Tavera, J.J., Near, T.J. & Wainwright, P.C. (2013). Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. *Evolution; international journal of organic evolution*, 67, 417-428.
- Schluter, D. (2000a). Ecological character displacement in adaptive radiation. *The American naturalist*, 156, S4-S16.
- Schluter, D. (2000b). *The ecology of adaptive radiation*.
- Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, 273, 1987-1998.
- Smith, A.J., Nelson-Maney, N., Parsons, K.J., James Cooper, W. & Craig Albertson, R. (2015). Body Shape Evolution in Sunfishes: Divergent Paths to Accelerated Rates of Speciation in the Centrarchidae. *Evolutionary Biology*, 42, 283-295.
- Streelman, J.T., Albertson, R.C. & Kocher, T.D. (2007). Variation in body size and trophic morphology within and among genetically differentiated populations of the cichlid fish, *Metriaclicha zebra*, from Lake Malawi. *Freshwater Biol*, 52, 525-538.
- Streelman, J.T. & Danley, P.D. (2003). The stages of vertebrate evolutionary radiation. *Trends Ecol Evol*, 18.
- Takahashi, T. & Sota, T. (2016). A robust phylogeny among major lineages of the East African cichlids. *Molecular phylogenetics and evolution*.

- Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J. & Robinson, R.L. (2001). How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, 10, 793-806.
- Vermeij, G.J. (1977). The Mesozoic marine revolution: evidence from snails, predators and grazers. *Paleobiology*, 3, 245-258.
- Wainwright, P., Carroll, A.M., Collar, D.C., Day, S.W., Higham, T.E. & Holzman, R.A. (2007). Suction feeding mechanics, performance, and diversity in fishes. *Integr Comp Biol*, 47, 96-106.
- Wainwright, P.C. (2007). Functional versus morphological diversity in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, 381-401.
- Wainwright, P.C., McGee, M.D., Longo, S.J. & Hernandez, L.P. (2015). Origins, Innovations, and Diversification of Suction Feeding in Vertebrates. *Integr Comp Biol*, 55, 134-145.
- Wainwright, P.C. & Richard, B.A. (1995). Predicting patterns of prey use from morphology of fishes. *Environmental Biology of Fishes*, 44, 97-113.

CHAPTER 1
***ANNOTATIONBUSTR*: AN R PACKAGE TO EXTRACT SUBSEQUENCES
FROM GENBANK ANNOTATIONS**

A version of this chapter was originally published by Samuel Borstein & Brian O'Meara:

Borstein, S.R. & O'Meara, B.C. (2018). AnnotationBustR: an R package to extract subsequences from GenBank annotations. *PeerJ*, 6, e5179.

Samuel R. Borstein conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft. Brian C. O'Meara authored or reviewed drafts of the paper, approved the final draft.

Abstract

Background. DNA sequences are pivotal for a wide array of research in biology. Large sequence databases, like GenBank, provide an amazing resource to utilize DNA sequences for large scale analyses. However, many sequence records on GenBank contain more than one gene or are portions of genomes. Inconsistencies in the way genes are annotated and the numerous synonyms a single gene may be listed under provide major challenges for extracting large numbers of subsequences for comparative analysis across taxa. At present, there is no easy way to extract portions from many GenBank accessions based on annotations where gene names may vary extensively.

Results. The R package *AnnotationBustR* allows users to extract sequences based on GenBank annotations through the ACNUC retrieval system given search terms of gene synonyms and accession numbers. *AnnotationBustR* extracts subsequences of interest and then writes them to a FASTA file for users to employ in their research endeavors.

Conclusion. FASTA files of extracted subsequences and accession tables generated by *AnnotationBustR* allow users to quickly find and extract subsequences from GenBank accessions. These sequences can then be incorporated in various analyses, like the construction of phylogenies to test a wide range of ecological and evolutionary hypotheses.

Introduction

The use of DNA sequence data is vital for a wide variety of research in evolutionary biology and ecology. Molecular phylogenies, which rely on DNA sequences for their construction, are extremely prevalent in biological research. Whether being used to correct for shared ancestry among organisms (Felsenstein, 1985), or to test hypotheses related phylogeography (Avice et al., 1987), diversification (Hey, 1992; Maddison, 2006), and trait evolution (Hansen, 1997; Bollback, 2006), phylogenies are required. Additionally, the use of phylogenies is important in community ecology to place systems into an evolutionary framework (Webb et al., 2002; Cavender-Bares et al., 2009). The construction of molecular phylogenies for systematic purposes is also a popular tool for taxonomists to identify new taxa and classify organisms (De Queiroz & Gauthier, 1994; Tautz et al., 2003). Some DNA sequences, like the mitochondrial gene cytochrome oxidase subunit I (COI), have utility as a method to identify and catalog species using DNA barcoding (Hebert et al., 2003; Ratnasingham & Hebert, 2007; Ratnasingham & Hebert, 2013).

Sequence databases like GenBank provide a valuable resource for using DNA sequence data to test evolutionary and ecological hypotheses. With the reduction in cost of DNA sequencing and the advancement of methods to analyze sequence data, the amount of sequence data available for use is growing at a rapid pace. Given that GenBank has over one-trillion sequences from over 370,000 species (Benson et al., 2017) and recent advances in methods to create massive phylogenies using either super-matrix (Driskell et al., 2004; Ciccarelli et al., 2006) or mega-phylogeny approaches (Smith et al., 2009; Izquierdo-Carrasco et al., 2014), the ability to generate large DNA sequence data sets for comparative analyses has become fairly common (Leslie et al., 2012; Rabosky et al., 2013; Spriggs et al., 2014; Zanne et al., 2014; Shi & Rabosky, 2015). Additionally, sequence retrieval with command line utilities like National Center for Biotechnology Information's eutils (NCBI Resource Coordinators, 2017) as well as within common scripting environments for biological analyses, like R (R Development Core Team, 2018), Perl (Perl Development Team, 2017), and Python (Python Software Foundation, 2016), are made possible with packages like *ape* (Paradis et al., 2004), *rentrez* (Winter, 2016), *reutils* (Schofl, 2015), *seqinr* (Charif & Lobry, 2007), *Bioperl* (Stajich et al., 2002), and *Biopython* (Chapman & Chang, 2000).

While GenBank provides a wealth of sequence data for researchers to use, some of it is rather difficult to manipulate into a useful form. For example, some sequences may be concatenated together, or the only gene sequence available for a species for the locus of interest may be within a mitochondrial or chloroplast genome. At the time of writing, GenBank has 70,048 complete or partial mitochondrial genomes, 4,698 chloroplast genomes, and a simple search for sequences containing concatenated coding sequences and tRNAs resulted in 286,538 additional sequences exclusive of the previously mentioned sequences. Although GenBank's annotation system provides a means to see where a locus of interest is in a genome or concatenated sequence and provides the ability to download it manually, this is extremely time consuming when many accessions are involved and not a feasible way to extract mass amounts of sequence data for use in research. While alternative sequence databases exist, especially for popular loci utilized for DNA barcoding and microbial community identification (ex. COI, 16S, etc.), these databases typically are extremely focused on just a few loci (DeSantis et al., 2006; Ratnasingham & Hebert, 2007; Cole et al., 2013). While these databases typically house data from GenBank, they may not have complete overlap with sequence data on GenBank, such as those from complete organelle sequences (ex. [GenBank KR150862.1](#) for the mitochondrial genome of *Bujurquina mariae*, which is a species not included in the Barcode of Life Database (BOLD) even though the GenBank accession has a record for cytochrome oxidase subunit I).

Another major challenge to obtaining large amounts of sequence data is the highly variable nomenclature of gene names. Most genes have several alternative names and symbols that are present in sequence databases. Among distant taxa, it is common for homologous genes to vary considerably in nomenclature (Tuason et al., 2003). Even within a group of closely related taxa or within a single taxon itself, how genes are annotated may differ substantially from record to record and a wide variety of alternative gene names may be found for a single gene (Morgan et al., 2004; Fundel & Zimmer, 2006). This poses serious problems when searching

through databases for molecular sequence data (Mitchell et al., 2003; Tamames & Valencia, 2006).

Here we present the R package *AnnotationBustR* to address the issues discussed above. *AnnotationBustR* reads GenBank annotations in R and pulls out the gene(s) of interest given a set of search terms and a vector of taxon accession numbers supplied by the user. It then writes the sequence for the gene(s) of interest to FASTA formatted files for each locus that users can then use in further analyses. For a more in-depth introduction to using *AnnotationBustR* users should consult the vignette in R through `vignette("AnnotationBustR-vignette")`, which provides instructions on how to use the different functions and their respective options. Other details about the package can be accessed through the documentation via `help("AnnotationBustR")`.

Implementation

AnnotationBustR is written in R (R Development Core Team, 2017), a popular language for analyzing biological data, and requires R version 3.4 or higher. It uses the existing R packages *ape* (Paradis et al., 2004) and *seqinr* (Charif & Lobry, 2007). *AnnotationBustR* uses *seqinr*'s interface to the online ACNUC database to extract gene regions of interest from concatenated gene sequences or genomes (Gouy et al., 1985; Gouy & Delmotte, 2008). ACNUC is a database and retrieval system for molecular sequence data maintained by the Pôle Bio-Informatique Lyonnais, which provides access to GenBank data and allows for easy access and manipulation of complex sequences, such as trans-spliced genes that may be on opposite strands of DNA. A list of the currently implemented commands is given in Table 1.1 and a flow chart of function usage is shown in Figure 1.1.

The main function of *AnnotationBustR*, `AnnotationBust`, takes a vector of accession numbers and a data frame of synonym search terms to extract loci of interest and writes them to a FASTA formatted file. This function also returns an accession table of all the loci of interest and the corresponding accession numbers the loci were extracted from for each species that can then be written to a csv file. Users can specify that duplicate genes should be extracted as well. If extracting coding sequences, users can also specify if they would like to translate the sequence into the corresponding peptides by specifying the appropriate GenBank numerical translation code.

We have included pre-made data frames with search terms in *AnnotationBustR* for animal and plant mitochondrial genomes, chloroplast genomes, and rDNA. These can be used to easily extract DNA barcodes, like cytochrome oxidase subunit I (COI) for animals in mitochondrial genomes (Hebert et al., 2003), the internal transcribed spacers (ITS) in rDNA for fungi and plants (Kress et al., 2005; Schoch et al., 2012), and maturase K (matK) and ribulose-bisphosphate carboxylase (rbcL) genes in the chloroplast genome of plants (Hollingsworth et al., 2009). These pre-made data frames consist of three columns with the column Locus containing the output file name, Type containing the type of sequence it is (i.e. CDS, tRNA, rRNA, misc_RNA, D-loop, etc.), and the third column, Name, containing a possible synonym of the loci to search for. For example, for cytochrome oxidase subunit I, GenBank includes gene names of

COI, CO1, COX1, cox1, COXI, cytochrome c oxidase subunit I, and COX-I. An additional column, IntronExonNumber, is used to specify the intron or exon number for extracting introns and exons. These search terms can be loaded into the workspace using the `data()` function. Annotation files for each accession are read in through *seqinr* and regular expressions matching of the synonyms provided by the user to the feature annotations are performed to identify the subsequence to extract. As certain loci may have numerous synonymous listings in GenBank feature tables that may not be included in the pre-made data frames of search terms, *AnnotationBustR* has the function `MergeSearchTerms` which allows users to add additional search terms to a pre-existing data frame of search terms if users follow the basic column formatting stated above. An additional feature of *AnnotationBustR* is the function `FindLongestSeq` which finds the longest sequence for each species in a set of GenBank accessions.

To demonstrate the performance of *AnnotationBustR*, we timed how long it took to extract thirteen popular coding sequences from 100 chloroplast genomes, the thirteen coding sequences from 100 metazoan mitochondrial genomes, and the three ribosomal RNA genes and internal transcribed spacers 1 and 2 from 100 metazoan rDNA sequences (Figure 1.2, see code in Supplemental Data S1). Timings were performed for each accession number starting with the extraction of a single locus and progressively adding an additional locus until all targeted loci were extracted for that accession. Timing trials were performed on a Windows desktop with a 3.8 GHz Intel Core i7 processor and 64 GB of RAM. For each accession, we timed how long it took to extract one through the full number of subsequences sought. Our timings indicate that *AnnotationBustR* can efficiently extract these loci into FASTA files and that performance scales well as the number of loci to extract increases. Our timings also scale well in terms of sequence input size, as there is over variation in mean input sequence size between the rDNA (1,140 bp), mtDNA (16,560 bp), and cpDNA (149,971 bp) sequences. From profiling the code, it appears that the variation in the mean times to extract sequences is related to speed of the ACNUC server in returning responses to requests. While we did perform timing trials on a machine with a decent amount of RAM, *AnnotationBustR* is not very taxing on memory. The peak RAM usage of the `AnnotationBust` function during extraction for the largest dataset, the chloroplast genomes (average size of 149,971 bp), was only 87.35 megabytes when profiled using the R package `peakRAM` (Quinn, 2017).

AnnotationBustR is available through CRAN (<https://cran.r-project.org/package=AnnotationBustR>) and is developed on GitHub (<https://github.com/sborstein/AnnotationBustR>). New extensions in development and fixes can be seen under the issues section on the packages GitHub page. Active lists of synonyms to be implemented in the next version are available on our github site in the issues section, as we gather more from users. While ACNUC currently only provides access to RefSeq virus sequences from the RefSeq database, we plan to add compatibility in the future for RefSeq accessions if access to other RefSeq databases is made available in *seqinr*.

Table 1.1: Functions and data included in the package *AnnotationBustR*.

Function/Data Name	Description
AnnotationBust	Writes found subsequences for loci of interest to a FASTA file for a vector of GenBank accessions and writes a corresponding accession table.
data(cpDNAterms)	Loads a data frame of search terms for chloroplast genes.
data(mtDNAterms)	Loads a data frame of search terms for animal mitochondrial genes.
data(mtDNAtermsPlants)	Loads a data frame of search terms for plant mitochondrial genes.
data(rDNAterms)	Loads a data frame of search terms for ribosomal DNA genes and spacers.
FindLongestSeq	Finds the longest sequence for each species in a set of GenBank accession numbers.
MergeSearchTerms	Merges two or more data frames containing search terms of features to extract into a single data frame.

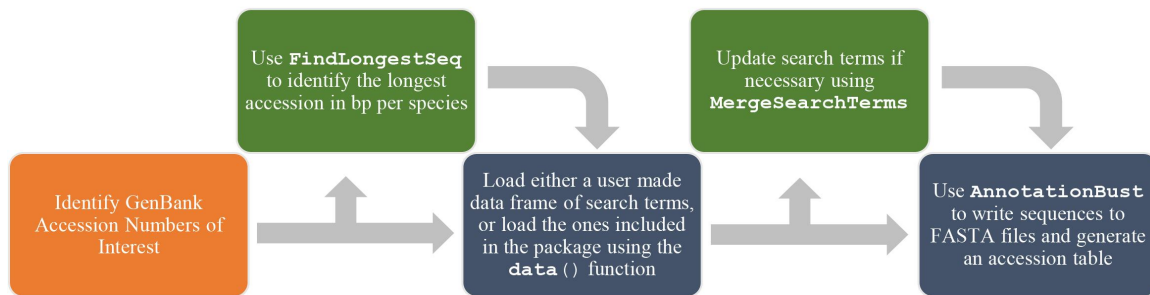


Figure 1.1: Flow chart of functions for a complete usage of *AnnotationBustR*. Blue boxes indicate a step using the package *AnnotationBustR* while orange boxes represent steps that need to be completed outside of *AnnotationBustR*. Boxes in green represent optional steps in the *AnnotationBustR* pipeline.

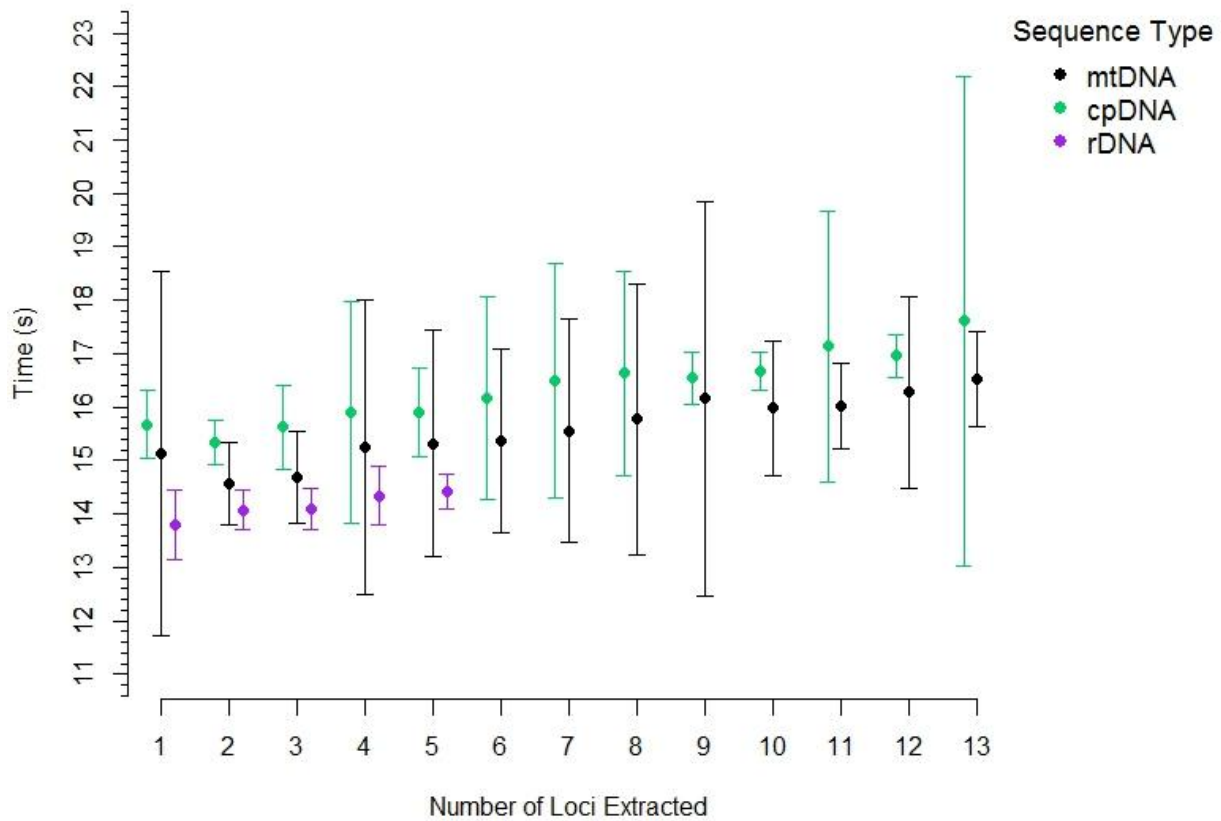


Figure 1.2: Timings of subsequence extraction using AnnotationBust for thirteen metazoan mitochondrial coding sequences (black), thirteen chloroplast coding sequences (green), and five metazoan rDNA subsequence (purple). Points represent the mean time in seconds with bars representing +/- one standard deviation.

Examples

We provide an example of the utility of sequences obtained using *AnnotationBustR* relative to single sequences on GenBank in North American leuciscine minnows. This group was chosen as they have been widely sequenced, both for whole mitochondrial genomes as well as single mitochondrial genes. Specifically, we extracted 12S rRNA, 16S rRNA, COI, CYTB, and ND2 genes from sixty species of leuciscine minnows and one outgroup, the zebrafish (*Danio rerio*). These five genes were chosen as individual gene sequences have also been regularly sequenced in this clade for phylogenetic studies (Bufalino & Mayden, 2010; April et al., 2011; Schoenhuth et al., 2012). We used PHLAWD v. 3.3 (Smith et al., 2009) to identify GenBank accessions to download for single genes that had the best coverage and identity for each species. PHLAWD uses BLAST (Altschul et al., 1990) restricted to a specified clade of interest to identify the best sequence in terms of coverage and identity for each species relative to a set of known bait sequences. We specified sequences had to have at minimum of at least 20 percent identity and coverage to be kept in the dataset of sequences accessed from GenBank on April 16th, 2018. Multiple sequence alignment of sequences obtained from mitochondrial genomes using *AnnotationBustR* and of individual GenBank genes was performed using MAFFT v. 7.402 (Katoh & Standley, 2013). Phylogenetic reconstruction and bootstrap analysis was performed in RAxML v. 8.2 under the GTRGAMMA model of sequence evolution with the alignment partitioned by gene (Stamatakis, 2014)(supplemental data S1).

The alignment constructed using sequences extracted from mitochondrial genomes using *AnnotationBustR* was 6,537 bp and slightly longer than the 6,508 bp alignment constructed from individual GenBank sequences. The total number of sequences was 305 (extracted from the sixty-one mitogenomes) and 198, in the *AnnotationBustR* and GenBank alignments respectively. The most notable difference between the two alignments is the amount of missing data. The GenBank alignment was only 37.72% complete relative to the 97.59% complete *AnnotationBustR* alignment. The disparity in number of sequences and completeness of alignments is due to some species only having sequence data for a locus within a mitochondrial genome or other concatenated sequence and not as a single sequence on GenBank. While we do recover slightly different topologies between the datasets, we do find that the phylogeny reconstructed from the *AnnotationBustR* extracted sequences has higher bootstrap support among shared bipartitions relative to the tree constructed from individual GenBank sequences, with shared bipartitions amongst the phylogenies having an average of 93.69% and 81.56% bootstrap support in the *AnnotationBustR* and GenBank trees respectively (Figure 1.3). Both tree topologies are similar to those recovered in other phylogenetic studies of leuciscine minnows (Bufalino & Mayden, 2010; Hollingsworth et al., 2013; Martin & Bonett, 2015). Our results may not necessarily reflect any aspect of quality of sequence data used, but rather quantity. For example, all the sequences extracted using *AnnotationBustR* are the complete sequence for that gene. This is not necessarily true of the single gene sequences obtained from GenBank, where some species only have a partial sequence for a gene.

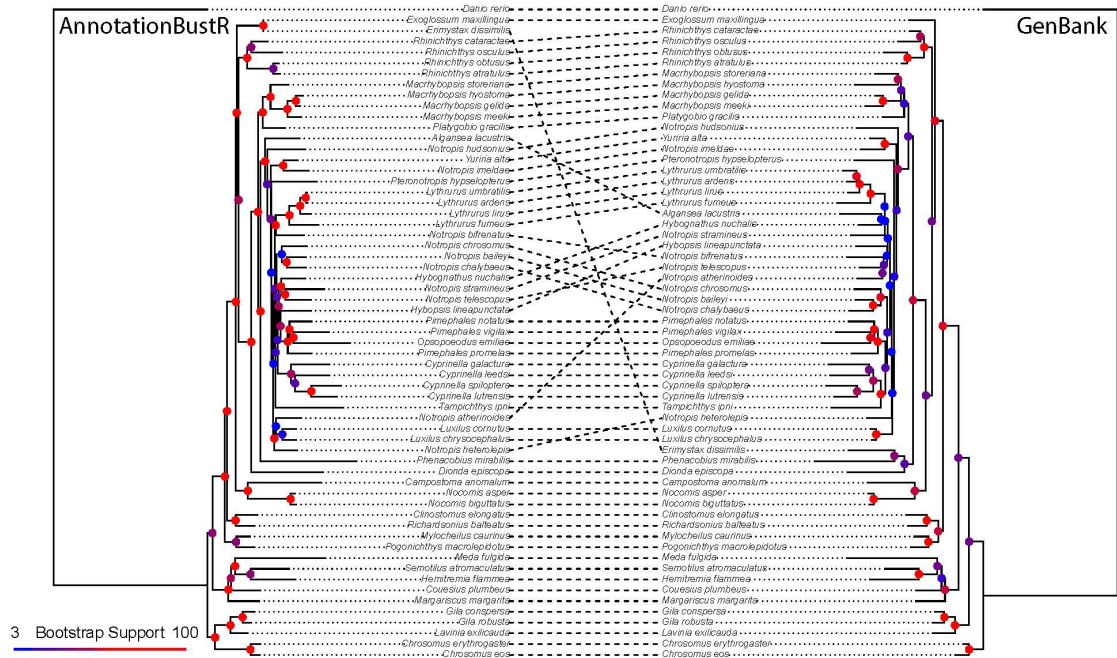


Figure 1.3: Phylogeny of 60 North American leuciscine minnows from sequences extracted using *AnnotationBustR* (left) and from individual GenBank sequences (right). Circles at nodes represent bootstrap support, with warmer colors representing higher bootstrap support and cooler colors representing lower bootstrap support.

Additionally, some species only have sequence data for certain genes as a mitochondrial genome accession and not necessarily have any other accession for that gene.

Conclusions

AnnotationBustR provides a way for users to extract subsequences from concatenated sequences, plastid, and mitochondrial genomes where gene names for subsequences may vary substantially. The major limitation to the functionality of *AnnotationBustR* is that it is only as good as the annotations in the features table it is using for extraction. For instance, some concatenated sequences do not have the individual gene positions annotated for the record and just state that it contains the genes, therefore making it impossible to extract a gene from it (ex. [GenBank KM260685.1](#), [GenBank KT216295.1](#)). Additionally, some loci may be present in the sequence yet missing from the features table completely (ex. mitochondrial D-loop missing in [GenBank KU308536.1](#)). Another limitation is that some popular loci such as intergenic spacers and some introns are not always annotated in the features table, making them impossible to extract. A good example of this is the trnH-psbA intergenic spacer, a proposed locus for plant DNA barcodes (Kress et al., 2005). As *AnnotationBustR* will extract all sequences within the annotations it finds that fit the supplied search terms, it is not a guarantee that these sequences are properly annotated, as annotation errors in sequence data are a known problem (Ben-Shitrit et al., 2012). Users will have to use due diligence on the sequences to check for incorrect annotations, mislabeled taxa, paralogs, chimeric sequences, and so forth. Alternative methods for extracting concatenated sequences could involve aligning them to known bait sequences of interest using a program like BLAST (Altschul et al., 1990). This can be automated by writing scripts to use BLAST locally or through NCBI. This represents a different approach from ours, relying on sequence similarity rather than annotations (PHLAWD builds on this but adds additional checks). While this can be useful, it potentially could be problematic as sequences could potentially align to non-orthologous sequences, which may cause issues with downstream analyses (Lassmann & Sonnhammer, 2005). At large phylogenetic scales, bait sequences may blast to paralogous copies rather than to orthologous sequences of more distantly related taxa in the clade. Nonetheless, BLAST and tools building on it represent historically useful approaches for creating a matrix of sequences that may be quite complementary to use of *AnnotationBustR* as another quality control check in the curation of large molecular sequence datasets.

While *utils* and packages in popular programming language can provide access to sequence data, and in some cases access to subsequences, they require complex query language. The R package we have developed provides a simple way for users to extract subsequences that have variable annotations by supplying a vector of accessions and search terms, either the ones included in the package or their own curated set. While R may not be as widely used as Python or Perl as a bioinformatics platform, it is a widely popular scripting language and we feel that this package fills a need in the R community.

Citation

Researchers publishing a paper that has used *AnnotationBustR* should cite this article and indicate the version of the package they are using. Package citation information can be obtained using `citation("AnnotationBustR")`.

Acknowledgements

We thank members of the O'Meara lab, Cedric Landerer, and Christopher Peterson for helpful discussions while developing the package and Orlando Schwery and Frankie West for beta testing. We thank the *seqinr* package authors for technical help related to their package. We thank three anonymous reviewers for their helpful feedback on improvements for the paper and the software.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- April, J., Mayden, R.L., Hanner, R.H. & Bernatchez, L. (2011). Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10602-10607.
- Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E. *et al.* (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst*, 489-522.
- Ben-Shitrit, T., Yosef, N., Shemesh, K., Sharan, R., Ruppín, E. & Kupiec, M. (2012). Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nature methods*, 9, 373-378.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. *et al.* (2017). GenBank. *Nucleic Acids Res*, 45, D37-D42.
- Bollback, J.P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, 7, 88.
- Bufalino, A.P. & Mayden, R.L. (2010). Phylogenetic evaluation of North American Leuciscidae (Actinopterygii: Cypriniformes: Cyprinoidea) as inferred from analyses of mitochondrial and nuclear DNA sequences. *Systematics and Biodiversity*, 8, 493-505.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol Lett*, 12, 693-715.
- Chapman, B. & Chang, J. (2000). Biopython: Python tools for computational biology. *ACM Sigbio Newsletter*, 20, 15-19.
- Charif, D. & Lobry, J.R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: *Structural approaches to sequence evolution*. Springer, pp. 207-232.
- Ciccarelli, F.D., Doerks, T., Von Mering, C., Creevey, C.J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311, 1283-1287.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y. *et al.* (2013). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic acids research*, 42, D633-D642.
- De Queiroz, K. & Gauthier, J. (1994). Toward a phylogenetic system of biological nomenclature. *Trends Ecol Evol*, 9, 27-31.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72, 5069-5072.
- Driskell, A.C., Ané, C., Burleigh, J.G., McMahon, M.M., O'Meara, B.C. & Sanderson, M.J. (2004). Prospects for building the tree of life from large sequence databases. *Science*, 306, 1172-1174.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, 125, 1-15.
- Fundel, K. & Zimmer, R. (2006). Gene and protein nomenclature in public databases. *Bmc Bioinformatics*, 7, 1.

- Gouy, M. & Delmotte, S. (2008). Remote access to ACNUC nucleotide and protein sequence databases at PBIL. *Biochimie*, 90, 555-562.
- Gouy, M., Gautier, C., Attimonelli, M., Lanave, C. & Di Paola, G. (1985). ACNUC—a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage. *Bioinformatics*, 1, 167-172.
- Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution; international journal of organic evolution*, 51, 1341-1351.
- Hebert, P.D., Cywinska, A. & Ball, S.L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270, 313-321.
- Hey, J. (1992). Using phylogenetic trees to study speciation and extinction. *Evolution; international journal of organic evolution*, 46, 627-640.
- Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M. *et al.* (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, 106, 12794-12797.
- Hollingsworth, P.R., Simons, A.M., Fordyce, J.A. & Hulsey, C.D. (2013). Explosive diversification following a benthic to pelagic shift in freshwater fishes. *BMC evolutionary biology*, 13.
- Izquierdo-Carrasco, F., Cazes, J., Smith, S.A. & Stamatakis, A. (2014). PUmPER: phylogenies updated perpetually. *Bioinformatics*, 30, 1476-1477.
- Katoh, K. & Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution*, 30, 772-780.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8369-8374.
- Lassmann, T. & Sonnhammer, E.L. (2005). Automatic assessment of alignment quality. *Nucleic acids research*, 33, 7120-7128.
- Leslie, A.B., Beaulieu, J.M., Rai, H.S., Crane, P.R., Donoghue, M.J. & Mathews, S. (2012). Hemisphere-scale differences in conifer evolutionary dynamics. *Proceedings of the National Academy of Sciences*, 109, 16217-16221.
- Maddison, W.P. (2006). Confounding asymmetries in evolutionary diversification and character change. *Evolution*, 60, 1743-1746.
- Martin, S.D. & Bonett, R.M. (2015). Biogeography and divergent patterns of body size disparification in North American minnows. *Molecular phylogenetics and evolution*, 93, 17-28.
- Mitchell, J., McCray, A. & Bodenreider, O. (2003). From phenotype to genotype: issues in navigating the available information resources. *Methods of information in medicine*, 42, 557-563.
- Morgan, A.A., Hirschman, L., Colosimo, M., Yeh, A.S. & Colombe, J.B. (2004). Gene name identification and normalization using a model organism database. *Journal of biomedical informatics*, 37, 396-410.
- NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic acids research*, 45, D12.

- Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20, 289-290.
- Perl Development Team (2017). The Perl directory. Available at: <https://www.perl.org/>.
- Python Software Foundation (2016). Python programming language. Available at: <https://www.python.org>.
- Quinn, T. (2017). peakRAM: Monitor the Total and Peak RAM Used by an Expression or Function. R package version 1.0.2. <https://CRAN.R-project.org/package=peakRAM>
- R Development Core Team (2018). R: A language and environment for statistical computing. Available at: <http://www.R-project.org/>.
- Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J. *et al.* (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature communications*, 4, 1958.
- Ratnasingham, S. & Hebert, P.D. (2013). A DNA-based registry for all animal species: the barcode index number (BIN) system. *PloS one*, 8, e66213.
- Ratnasingham, S. & Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes*, 7, 355-364.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A. *et al.* (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109, 6241-6246.
- Schoenhuth, S., Shiozawa, D.K., Dowling, T.E. & Mayden, R.L. (2012). Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). *Zootaxa*, 3586, 281-303.
- Schofl, G. (2015). reutils: Talk to the NCBI EUtils. R package version 0.2.2. <https://CRAN.R-project.org/package=reutils>
- Shi, J.J. & Rabosky, D.L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution; international journal of organic evolution*, 69, 1528-1545.
- Smith, S.A., Beaulieu, J.M. & Donoghue, M.J. (2009). Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC evolutionary biology*, 9, 37.
- Spriggs, E.L., Christin, P.-A. & Edwards, E.J. (2014). C 4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PloS one*, 9, e97722.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C. *et al.* (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome research*, 12, 1611-1618.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.
- Tamames, J. & Valencia, A. (2006). The success (or not) of HUGO nomenclature. *Genome biology*, 7, 1.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2003). A plea for DNA taxonomy. *Trends Ecol Evol*, 18, 70-74.
- Tuason, O., Chen, L., Liu, H., Blake, J.A. & Friedman, C. (2003). Biological nomenclatures: a source of lexical knowledge and ambiguity. *Proceedings of the Pacific Symposium of Biocomputing*, 9, 238-249.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu Rev Ecol Syst*, 33, 475-505.

Winter, D. (2016). rentrez: Entrez in R. 0.2.4. <https://cran.r-project.org/package=rentrez>

Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G. *et al.* (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, 506, 89-92.

CHAPTER 2
DIETR: AN R PACKAGE FOR CALCULATING FRACTIONAL TROPHIC
LEVELS FROM QUANTITATIVE DIET AND QUALITATIVE FOOD
ITEM DATA

A version of this is available through Samuel Borstein's GitHub (<https://github.com/sborstein>) and a manuscript is to be submitted to *Hydrobiologia* as:

Borstein, S.R. *dietr*: An R package for calculating fractional trophic levels from quantitative diet and qualitative food item data.

Author contributions: Samuel R. Borstein performed all aspects of the study.

Abstract

This article introduces an R package, *dietr*, which implements the routines used to calculate fractional trophic levels from quantitative diet items and qualitative food item data implemented in TrophLab within the open source R environment. *dietr* is easy to use and can quickly calculate trophic levels for many records. Unlike TrophLab, users can also specify a taxonomic hierarchy and estimate trophic levels at multiple taxonomic levels all at once. Additionally, *dietr* works well with FishBase data read into R using *rfishbase* and comes with pre-made databases of prey trophic levels that users can use for estimating trophic levels. We provide information on *dietr*'s performance and provide a use case example of how *dietr* can be used on an empirical dataset.

Introduction

The study of trophic ecology is vital to biological research. One such aspect of trophic ecology, the trophic level, is important in understanding various processes shaping biological diversity (Post 2002a). Trophic level data are employed in a variety of research ranging from studies of macroevolution (Rojas *et al.* 2018; Borstein *et al.* 2019), ontogenetic changes in life-history (Wilbur *et al.* 1974; Mittelbach *et al.* 1988), ecological assembly (Leibold *et al.* 2004; Duffy *et al.* 2007; Marczak *et al.* 2007), food-web structure (Pimm *et al.* 1991; Williams & Martinez 2004), assessing fisheries management and impact (Pauly *et al.* 1998; Essington *et al.* 2006), assessing impacts of invasive species (Vander Zanden *et al.* 1999; Grosholz 2002), and trophic modelling of ecosystems (Christensen & Pauly 1992, 1993; Christensen & Walters 2004). The main methods used to estimate trophic level are stomach contents analysis (Hyslop 1980) or stable isotopes analysis (Post 2002b). While both methods have their pros and cons, of which a discussion falls outside the scope of this paper, stomach contents are still commonly utilized to estimate trophic levels, either by themselves or in conjunction with stable isotope analyses (Post 2002b; Rybczynski *et al.* 2008; Polito *et al.* 2011; Mancinelli *et al.* 2013; Cresson *et al.* 2014). Diet data is also widely available for many species through several literature and database resources (Jones *et al.* 2009; Wilman *et al.* 2014; Palomares & Pauly 2018; Froese & Pauly 2019).

While some databases, such as FishBase (Froese & Pauly 2019) and SeaLifeBase (Palomares & Pauly 2018) provide a wealth of diet and trophic level data and are easily accessible through R using *rfishbase* (Boettiger *et al.* 2012), some trophic level values one can obtain from functions using these databases can be problematic. For example, multiple life-history stages may be included for an estimate of a species (ex. *Plectropomus laevis*), prey

trophic levels may be incorrect due entry errors causing a spurious trophic level calculation (ex. *Macropharyngodon geoffroy* being artificially high) or may have diet data but no calculated trophic level (ex. *Deuterodon langei*). In the most extreme cases, some trophic levels for heterotrophs are incorrectly listed as equal to those of autotrophs in these databases (e.x. *Cottocomephorus grewinkii*, DietCode 2169)! Thus, users may want to re-calculate trophic levels for a species while removing certain life-history stages or re-calculate trophic levels when new data on the trophic levels of prey items becomes available. Additionally, users may want to calculate trophic levels for taxa from newly collected diet data or from a study not incorporated in these databases.

Here, I introduce *dietr*, a package that calculates trophic level from diet or food item data implemented in the open source environment of R. *dietr* uses the TROPH routines described in FishBase and used in *TrophLab*, which is distributed only as a Microsoft Access program (Pauly *et al.* 2000; Froese & Pauly 2019). For a more in-depth introduction to using *dietr* users should consult the vignette in R through vignette(“dietr-vignette”, which provides in-depth instructions and a tutorial on how to use the functions and their respective options. Other details about the package can be accessed through the documentation via help(“dietr”).

Implementation

dietr is written in R (R Development Core Team 2018), a popular language for analyzing biological data, and requires R version 3.5 or higher. It uses the existing R packages *mgcv* (Wood 2018) and *rfishbase* (Boettiger *et al.* 2012). For ease of use and fast vectorized calculations of trophic levels, the main inputs for *dietr* functions are data frames.

dietr has two main function types (Table 2.1). The first type is a pair of functions that convert FishBase/SeaLifeBase data obtained from *rfishbase* into a format that *dietr* can use. The functions *ConvertFishbaseDiet* and *ConvertFishbaseFood* take as input the data frame produced by *rfishbase*’s *diet* and *fooditems* functions respectively, with an additional argument *ExcludeStage*, for which users can specify life-history stages they would like to exclude from the dataset.

The second set of functions perform the actual trophic level calculations. The equations used to calculate trophic level are not different from those implemented in *TrophLab*, though for sake of clarity, they will be discussed below. The function *DietTroph* calculates trophic level from quantitative diet data where the percent contributions of prey are known (i.e. volumetric, weight, etc.). The calculation of the trophic level is defined by the equation 1. Here the trophic level of the focal species, $Troph_i$, is calculated by adding 1 to the sum of the trophic level of the prey items weighted by their contribution in the diet, where $Troph_j$ is the trophic level of prey item j , DC_{ij} is the proportion of item j in the diet of species i and G is the number of prey items consumed.

Table 2.1: Functions and datasets available in the *diatr* R package.

Function/Data Name	Description
ConvertFishbaseDiet	Converts diet data from FishBase/SeaLifeBase into a usable format for diatr
ConvertFishbaseFood	Converts food item data from FishBase/SeaLifeBase into a usable format for diatr
CortesPreyVals	Trophic levels of prey items following Cortes, 1999, which can be used in trophic level calculations
DietTroph	Calculates trophic level from diet data
FishbasePreyVals	Trophic levels of prey items following FishBase and TrophLab which can be used in trophic level calculations
Electivity	Calculates forage index and several electivity indices.
diatr-vignette	In depth tutorial on how to use diatr

$$Troph_i = 1 + \sum_{j=1}^G DC_{ij} \times Troph_j$$

Equation 1

As estimates of prey trophic levels may not be exact, the standard error around the estimate of the trophic level can be calculated with equation 2. Here, the standard error of the focal species i , s.e. _{i} , is calculated as the square root of the sum of the product of the standard errors of the prey items s.e. _{j} to the second weighted by their respective contribution in the diet, DC _{ij} over 100.

$$s.e._i = \sqrt{\frac{\sum_{j=1}^G DC_{ij} \times s.e._j^2}{100}}$$

Equation 2

For qualitative data where only a list of food items appearing in the diet is known without knowledge of their percent contribution in the diet, a different procedure is performed for calculating the trophic level and standard error and is implemented in the function *FoodTroph*. In the simplest of cases, where a species only consumes one type of prey, the trophic level is simply 1 added to the trophic level of the prey item and the standard error is equal to the standard error of the prey item. In cases where more than one type of prey is consumed, trophic level is calculated as follows. First, prey items are randomly selected and ordered, and their simulated contribution to a diet, P , is calculated by equation 3. Here, R denotes the ordered rank of the prey item and G represents the total number of prey items consumed.

$$\log_{10}P = 2 - 1.9\log_{10}R - 0.16\log_{10}G$$

Equation 3

The simulated contribution of a diet items, P , calculated from equation three are then used to calculate the trophic level in equation four, which is the sum of the products of the simulated contribution of prey items P_i and their respective trophic levels divided by the sum of P_i .

$$Troph = \sum (P_i * Troph_i) / \sum P_i$$

Equation 4

The estimate of the standard error around the trophic level from food item data is defined by equation five. Here, the standard error is the square root of the sum of the product of the standard error of a prey item squared and the contribution of the prey item minus 1 divided by the sum of the contribution of the prey items, P , minus the total number of prey items, G . For food item data, the random sampling routine and calculation to estimate trophic level and standard error is repeated 100 times, with the final trophic level and standard error being the mean of these 100 calculations.

$$s.e. = \sqrt{\frac{(s.e._1)^2 \times (P_1 - 1) + (s.e._2)^2 \times (P_2 - 1) \dots (s.e._G)^2 \times (P_G - 1)}{100}}$$

Equation 5

As input, the *DietTroph* and *FoodTroph* functions require either a data frame of diet or food item data and a data frame containing the trophic levels of the prey items. A vector of shared column names that contain the classification of the prey items also needs to be specified. The most unique feature implemented in *diatr* and distinct from TrophLab is the option for users to specify a data frame containing a taxonomy to calculate trophic levels at different hierarchies. This data frame should increase from least inclusive to most inclusive from left to right. By specifying a taxonomy, users can calculate the trophic level at numerous levels (i.e. Individual, population, species, etc.), all in a single run.

diatr is developed on GitHub (<https://github.com/sborstein/tRophic>). New extensions in development and bug fixes can be seen under the issues section on the packages GitHub page.

Methods

To test the accuracy of trophic levels estimated with *diatr* we randomly sampled 1000 diets from FishBase obtained using the *rfishbase* package. We then used the *DietTroph* function to estimate trophic levels using a database of prey trophic level values from FishBase/TrophLab available in the *diatr* package data object *FishbasePreyVals*. We then measured the correlation between the trophic levels calculated using *diatr* to the trophic levels reported by FishBase for

those diets. We would expect a strong correlation between the two datasets if *dietr* is performing accurately. To assess speed, we used R's `system.time` function to measure the time in seconds it took *dietr* to perform.

We performed a similar analysis to the one above comparing trophic levels from qualitative food item data on FishBase to those calculated using *dietr*. To do this, we selected 1000 random species which had food item data and used the *dietr* function `FoodTroph` with prey trophic levels from the `FishbasereyVals` data object in the *dietr* package. Speed of the function was assessed using R's `system.time` function.

To highlight a use case of *dietr*, we obtained data from a study by (Magalhaes et al. 2015) on *Herichthys minckleyi*, a cichlid native to the Cuatro Ciénegas, Mexico. This dataset reported dietary items for *Herichthys minckleyi* by lake and year collected. We used *dietr* to calculate trophic levels in a hierarchy for 422 individuals by lake and year (2011/2012), lake regardless of year, and finally across all samples for the species. In cases where portions of the stomach contents were non-identifiable, the percentage dietary contribution was re-calculated from the identified items.

Results

We found trophic levels estimated with *dietr* to strongly correlate with trophic levels from FishBase ($r = 0.888$, $df = 998$, $p < 0.001$, Figure 2.1). We note that we do calculate some values that are different than those listed on FishBase. This appears due to two factors. The first appears due to database entry errors on FishBase. An example would be the diet entry 2454 for *Clupea pallasii*, which has a trophic level of 1.35. This value would be impossible as it would suggest that this species is partially autotrophic. We estimate a trophic level of 3.09 for this species using the data obtained for this record. Another example of an entry error is record 22526 for *Merlangius merlangus*. For this record, only a single prey item is consumed, which should be calculated as 1+trophic level of the prey item. In this case, the trophic level of the prey item according to the database in TrophLab should be 1.5, yielding a trophic level of 2.5. While *dietr* calculated a trophic level for this record as 2.5, this record has a trophic level of 3.5 on FishBase. An additional factor that may be generating differences in the trophic levels estimated by *dietr* and those stored on FishBase may be the use of unique prey trophic levels, which are listed in the `diet_items` table from *rfishbase* as `preyTroph`. As the database of prey trophic levels used in *dietr* are the same as those used in TrophLab, deviations from these will cause differences in estimated trophic levels. Code execution time to measure trophic levels for these 1000 studies was 2.76 seconds.

Trophic levels estimated using the *FoodTroph* function were highly correlated with the values reported on FishBase ($r = 0.989$, $df = 998$, $p < 0.001$, Figure 2.1). While slight differences do occur in the trophic level reported by FishBase and *dietr* this is not necessarily surprising given the routine to calculate trophic levels from food item data randomly samples food items, ranks them, and calculates trophic levels, with the final trophic level being the mean across 100 iterations. Code execution time to measure trophic levels for 1000 species was 11.08 seconds.

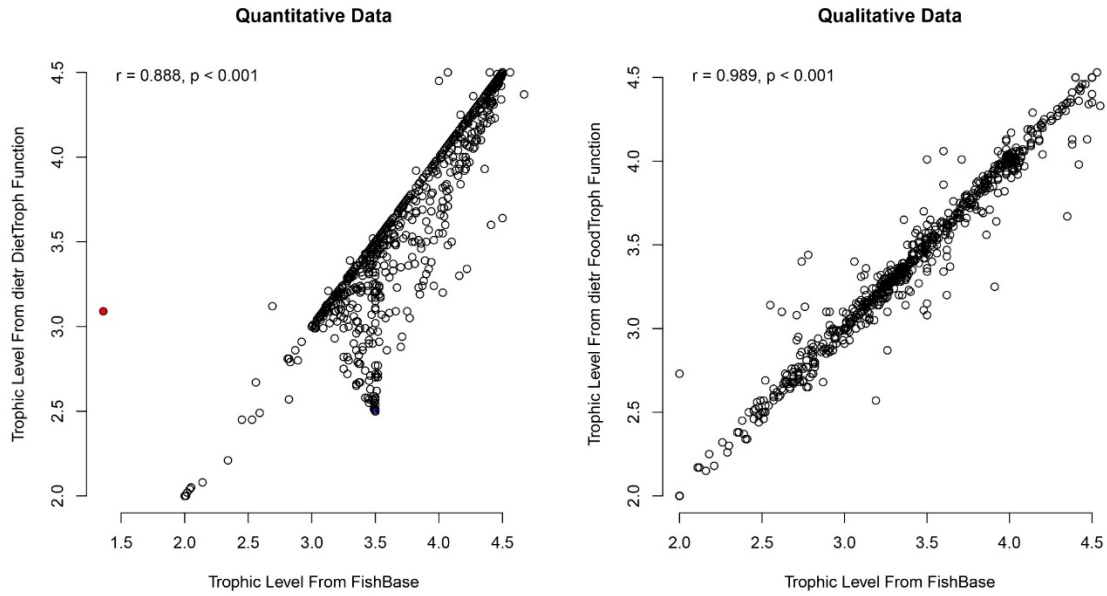


Figure 2.1: Correlation between trophic levels obtained from FishBase (x-axis in both plots) and those estimated by *dietr*. Trophic levels estimated from quantitative diet data are shown on the left while trophic levels estimated from qualitative food item data appear on the right. Trophic levels calculated by *dietr* were comparable and strongly correlated to those on FishBase. For the quantitative data, outlier taxa with substantial difference between trophic levels calculated with *dietr* and FishBase are colored. The red point highlights a species with an incorrect trophic level that would be equal to that of an autotroph (record 2454, *Clupea pallasii*), while blue points represent incorrect calculations where an incorrect trophic level for the prey item was used in calculations, leading to an incorrectly low trophic level (record 22526, *Merlangius merlangus*).

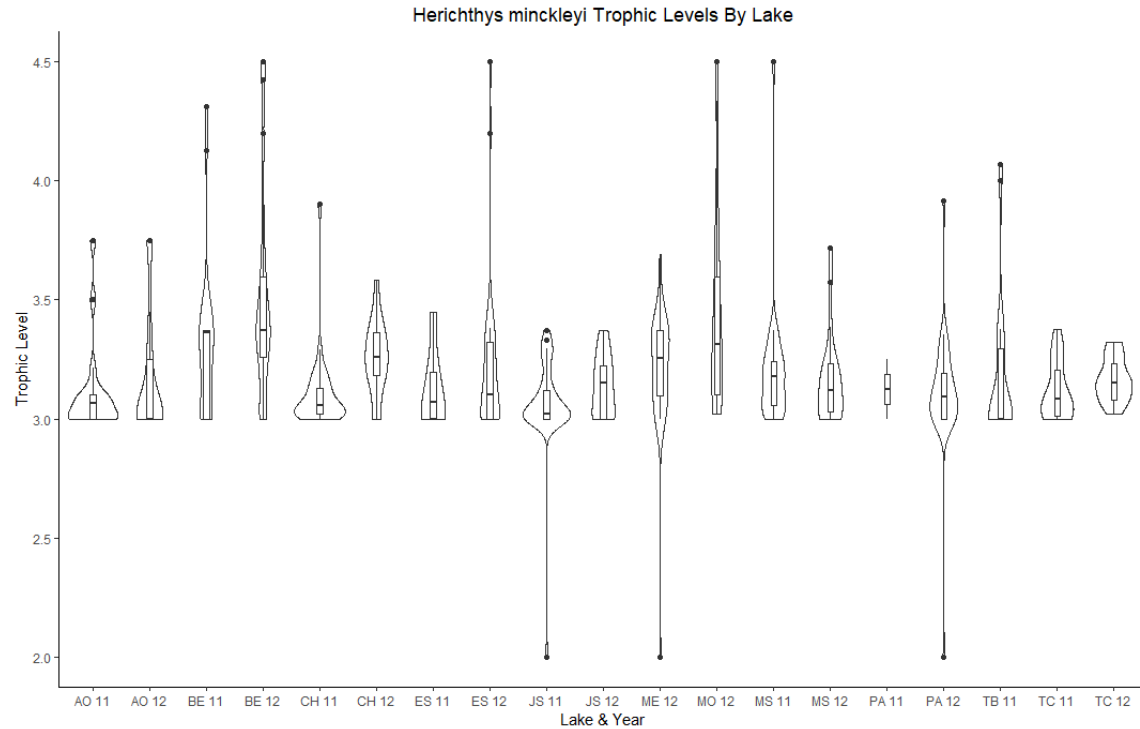


Figure 2.2: Trophic levels of *Herichthyes minckleyi* calculated using dietr from data in Magalhaes et al. (2015). Violin plots show the distribution of trophic levels calculated from different lakes and different sampling years. Lakes are represented by two letter abbreviations while the sampling year (11 or 12) represents sampling in 2011 or 2012. Overlaid box plots represent the mean and quantiles ranges.

For our use case, we found individual trophic levels to range from 2.0 to 4.5 (Figure 2.2). By lake by year, trophic levels ranged from 3.05-3.49 and by lake regardless of year 3.07-3.42. The overall mean trophic level for this species across all individuals in the dataset was 3.21. Time to perform trophic level calculations for 422 individuals at 4 hierarchical grouping took a total of 0.33 seconds.

Discussion

Our results show that `dietr` accurately calculates trophic levels from quantitative diet and qualitative food item data. Our use case demonstrates the flexibility of `dietr` and highlights how it allows users to submit a hierarchy to calculate trophic levels across multiple groupings in a single call of the program. This feature provides an easy method for users to follow a taxonomy and measure trophic levels at individual, population, and species level simultaneously, or could be used in other ways, like calculating trophic levels for species by year in the same fishing ground for studies looking at the mean trophic change within ecosystems over time. `dietr` is also extremely quick in regard to runtime, taking mere seconds for datasets of 1,000 records. While we do find the `FoodTroph` function takes roughly four times as long to perform as the `DietTroph` function, this is not surprising given the random sampling, ranking, and trophic level calculation occurring 100 times for each record. Therefore, the code is calculating 10,000 trophic levels in this amount of time (100x1,000). While we did not directly compare the timings of `dietr` to `TrophLab`, given users must manually enter data into `TrophLab` for each record, `dietr` provides a much faster alternative to calculate trophic levels for many records.

As `dietr` is implemented in an open source language and has a relatively simple structure for input, I hope that this contribution will help aid the reproducibility of trophic ecological studies measuring trophic levels. Researchers using `dietr` could easily supply their code, diet data matrix, and prey trophic level values used for calculating trophic levels as supplemental data files supporting their research, aiding in reproducibility.

References

- Boettiger, C., Lang, D.T. & Wainwright, P.C. (2012). rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81, 2030-2039.
- Borstein, S.R., Fordyce, J.A., O'Meara, B.C., Wainwright, P.C. & McGee, M.D. (2019). Reef fish functional traits evolve fastest at trophic extremes. *Nature Ecology & Evolution*, 3, 191-199.
- Chesson, J. (1983). The estimation and analysis of preference and its relationship to foraging models. *Ecology*, 64, 1297-1304.
- Christensen, V. & Pauly, D. (1992). ECOPATH II—a software for balancing steady-state ecosystem models and calculating network characteristics. *Ecological modelling*, 61, 169-185.
- Christensen, V. & Pauly, D. (1993). *Trophic models of aquatic ecosystems*. WorldFish.
- Christensen, V. & Walters, C.J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. *Ecological modelling*, 172, 109-139.
- Cresson, P., Ruitton, S., Ourgaud, M. & Harmelin-Vivien, M. (2014). Contrasting perception of fish trophic level from stomach content and stable isotope analyses: a Mediterranean artificial reef experience. *Journal of experimental marine biology and ecology*, 452, 54-62.
- Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thébault, E. & Loreau, M. (2007). The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol Lett*, 10, 522-538.
- Essington, T.E., Beaudreau, A.H. & Wiedenmann, J. (2006). Fishing through marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3171-3175.
- Froese, R. & Pauly, D. (2019). FishBase. Available at: <http://www.fishbase.org/2018>.
- Grosholz, E. (2002). Ecological and evolutionary consequences of coastal invasions. *Trends Ecol Evol*, 17, 22-27.
- Hyslop, E. (1980). Stomach contents analysis—a review of methods and their application. *Journal of fish biology*, 17, 411-429.
- Ivlev, U. (1961). Experimental ecology of the feeding of fish. Yale University Press, New Haven.
- Jacobs, J. (1974). Quantitative measurement of food selection. *Oecologia*, 14, 413-417.
- Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L. *et al.* (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90, 2648-2648.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett*, 7, 601-613.
- Magalhaes, I.S., Ornelas-Garcia, C.P., Leal-Cardin, M., Ramirez, T. & Barluenga, M. (2015). Untangling the evolutionary history of a highly polymorphic species: introgressive hybridization and high genetic structure in the desert cichlid fish *Herichthys minckleyi*. *Mol Ecol*, 24, 4505-4520.

- Mancinelli, G., Vizzini, S., Mazzola, A., Maci, S. & Basset, A. (2013). Cross-validation of $\delta^{15}\text{N}$ and FishBase estimates of fish trophic position in a Mediterranean lagoon: The importance of the isotopic baseline. *Estuarine, Coastal and Shelf Science*, 135, 77-85.
- Marczak, L.B., Thompson, R.M. & Richardson, J.S. (2007). Meta-analysis: trophic level, habitat, and productivity shape the food web effects of resource subsidies. *Ecology*, 88, 140-148.
- Mittelbach, G., Osenberg, C. & Leibold, M. (1988). Trophic relations and ontogenetic niche shifts in aquatic ecosystems. In: *Size-structured populations*. Springer, pp. 219-235.
- Palomares, M. & Pauly, D. (2018). SeaLifeBase. Available at: <https://www.sealifebase.org/>.
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. (1998). Fishing down marine food webs. *Science*, 279, 860-863.
- Pauly, D., Froese, R., Sa-a, P., Palomares, M., Christensen, V. & Rius, J. (2000). TrophLab manual. *ICLARM, Manila, Philippines*.
- Pimm, S.L., Lawton, J.H. & Cohen, J.E. (1991). Food web patterns and their consequences. *Nature*, 350, 669.
- Polito, M.J., Trivelpiece, W.Z., Karnovsky, N.J., Ng, E., Patterson, W.P. & Emslie, S.D. (2011). Integrating stomach content and stable isotope analyses to quantify the diets of pygoscelid penguins. *PloS one*, 6, e26642.
- Post, D.M. (2002a). The long and short of food-chain length. *Trends Ecol Evol*, 17, 269-277.
- Post, D.M. (2002b). Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology*, 83, 703-718.
- R Development Core Team (2018). R: A language and environment for statistical computing. Available at: <http://www.R-project.org/>.
- Rojas, D., Ramos Pereira, M.J., Fonseca, C. & Dávalos, L.M. (2018). Eating down the food chain: generalism is not an evolutionary dead end for herbivores. *Ecol Lett*.
- Rybczynski, S., Walters, D., Fritz, K. & Johnson, B. (2008). Comparing trophic position of stream fishes using stable isotope and gut contents analyses. *Ecol Freshw Fish*, 17, 199-206.
- Vander Zanden, M.J., Casselman, J.M. & Rasmussen, J.B. (1999). Stable isotope evidence for the food web consequences of species invasions in lakes. *Nature*, 401, 464.
- Wilbur, H.M., Tinkle, D.W. & Collins, J.P. (1974). Environmental certainty, trophic level, and resource availability in life history evolution. *The American naturalist*, 108, 805-817.
- Williams, R.J. & Martinez, N.D. (2004). Limits to trophic levels and omnivory in complex food webs: theory and data. *The American naturalist*, 163, 458-468.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M.M. & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, 95, 2027-2027.
- Wood, S. (2018). Package ‘mgcv’. 1.8-23

CHAPTER 3
REEF FISH FUNCTIONAL TRAITS EVOLVE FASTEST AT TROPHIC
EXTREMES

A version of this chapter was originally published by Samuel Borstein, James Fordyce, Brian O'Meara, Peter Wainwright, and Matthew McGee.:

Borstein, S.R., Fordyce, J.A., O'Meara, B.C., Wainwright, P.C. & McGee, M.D. (2019). Reef fish functional traits evolve fastest at trophic extremes. *Nature Ecology & Evolution*, 3, 191-199.

Author contributions: S.R.B. and M.D.M. designed the study. S.R.B. and J.A.F. performed the analyses. S.R.B., J.A.F. and M.D.M. wrote the manuscript with substantial comments from B.C.O. and P.C.W.

Abstract

Trophic ecology is thought to exert a profound influence on biodiversity, but the specifics of the process are rarely examined at large spatial and evolutionary scales. We investigate how trophic position and diet breadth influences functional trait evolution in one of the most species-rich and complex vertebrate assemblages, coral reef fishes, within a large-scale phylogenetic framework. We show that, contrary to established theory, functional traits evolve fastest in trophic specialists with narrow diet breadths at both very low and high trophic position. Top trophic level specialists exhibit the most functional diversity, while omnivorous taxa with intermediate trophic position and wide diet breadth have the least functional diversity. Our results reveal the importance of trophic position in shaping evolutionary dynamics while simultaneously highlighting the incredible trophic and functional diversity present in coral reef fish assemblages.

Introduction

One of the central issues limiting progress towards a generalized theory of biological organization involves integrating the interplay of current ecological conditions with long-term macroevolutionary dynamics. Trophic interactions are critically important for understanding how variation in resource acquisition shapes diversity, food webs, and community assembly (Vermeij 1977; Vamosi 2003). While studies have provided valuable insights into how trophic ecology effects community assembly, trophic cascades, and ecosystem-level processes (Hector *et al.* 1999; Cardinale *et al.* 2006; Cadotte *et al.* 2009), there remains a need to quantify how trophic interactions might shape evolutionary dynamics over macroevolutionary timescales using phylogenetic comparative methods (Webb *et al.* 2002; Duffy *et al.* 2007; Cavender-Bares *et al.* 2009). Large comparative studies of trophic ecology on phylogenies have provided valuable insights into the evolutionary dynamics of families (de Graaf *et al.* 2004; Cooper & Westneat 2009; Martin & Wainwright 2011; Fedosov *et al.* 2014; Frederich *et al.* 2014; Davis *et al.* 2016; McGee *et al.* 2016), yet further comparative studies investigating the impact of trophic ecology on global assemblage wide patterns remains crucial.

Studies focused on single species or small clades suggest that the evolution of high trophic position should constrain rates of functional trait evolution due to the constraints imposed

by feeding on evasive energy-rich prey items relative to lower trophic levels (Estes & Arnold 2007; Collar *et al.* 2009; Smith *et al.* 2015a; Svanbäck *et al.* 2015). We term this the height constraint hypothesis (HCH). The niche variation hypothesis (NVH) predicts a direct positive relationship between niche breadth and phenotypic variation (Van Valen 1965). While the NVH has been investigated at the individual (Bolnick *et al.* 2003; Araújo *et al.* 2011), population (Bolnick *et al.* 2007; Hsu *et al.* 2014), and species level (Findley & Black 1983), it is less commonly examined at a macroevolutionary scale (Galeotti & Rubolini 2003), where we predict that guilds composed of species with wider niches might also exhibit greater phenotypic variation. Thus, guilds occupying intermediate trophic positions, which exploit a more diverse set of resources compared to species feeding very high or very low in a food web, might therefore be expected to exhibit faster rates of functional trait evolution as there would be more phenotypic variance at the tips of the phylogeny amongst members of a guild. A third possibility, which we term the ‘trophic extremes’ hypothesis (TEH), posits that species with extremely high or low trophic position experience elevated rates of evolution compared to species at intermediate trophic levels. Increased rates of functional trait evolution may also be associated with the difficulty of accessing resources at very low or very high trophic levels. In the case of herbivory, this may require functional traits related to crushing tough cell walls, as well as adaptations to a diet containing fewer proteins and fats (Hay & Fenical 1988; Paré & Tumlinson 1999; Agrawal 2007; Barton & Koricheva 2010). At the highest trophic level, predators experience physical and energetic constraints related to the capture of highly evasive, large, and/or defended prey items (Lundvall *et al.* 1999; Price *et al.* 2015; Wainwright *et al.* 2015; Mihalitsis & Bellwood 2017).

Here, we examine how trophic position and diet breadth affect the functional dynamics of trait evolution, utilizing one of the most species-rich vertebrate assemblages -- coral reef acanthomorph fishes. Coral reefs are known biodiversity hotspots that promote elevated rates of diversification, morphological diversity, and ecological novelty due to the high amount of habitat partitioning and available trophic niches (Alfaro *et al.* 2007; Cowman *et al.* 2009; Price *et al.* 2011; Price *et al.* 2013; Santini *et al.* 2013; Floeter *et al.* 2018). Reef acanthomorph fishes dominate this habitat and occupy numerous trophic levels, ranging from strict herbivores to specialized apex predators, and these feeding modes have evolved not once but many times (Wainwright *et al.* 2015; Floeter *et al.* 2018). Unfortunately, coral reefs are highly at-risk habitats due to various anthropogenic changes to their environment (Bellwood *et al.* 2004; Carpenter *et al.* 2008; Stuart-Smith *et al.* 2013). Understanding the ecological processes governing the evolutionary dynamics of reef assemblages is critical to successful conservation efforts.

We examine how trophic position affects rate of functional trait evolution in a comparative context that accounts for evolutionary history, as well as the effect of trophic position on diet breadth. If the HCH drives evolution, we would expect to see high trophic level species -- i.e. top predators -- evolving more slowly than other trophic levels. However, if the NVH is the primary driver, we would expect the highest rates of functional trait evolution in intermediate trophic levels. If the TEH is operating, we should see that the trophic extremes --

species with very low or very high trophic levels -- experience faster rates of functional trait evolution than species with intermediate positions.

Methods

Trophic Level Data Acquisition

We used the R interface to FishBase, *rfishbase* (Boettiger *et al.* 2012; Froese & Pauly 2018; R Development Core Team 2018), to identify species of reef-associated acanthomorph fishes for which diet data was available. For the reef-associated acanthomorphs identified we obtained volumetric diet data and food item data using the *rfishbase* diet and ecology functions. To avoid bias in ontogenetic changes in trophic ecology through development, diet data was filtered to remove entries that were entirely based on recruits and juveniles. Trophic levels based on volumetric diet data and from food items were calculated following the routines described by Froese & Pauly (2000). Preference was given to records where trophic level could be calculated from volumetric stomach contents as opposed to estimates based on food items, although estimates based on food items have been shown to be a rough estimator of trophic level (Froese & Pauly 2000; Mancinelli *et al.* 2013). Species were placed in five discrete trophic categories following Froese (2005) with trophic levels less than 2.20 being classified as herbivores, 2.21-2.80 omnivores, 2.81-3.80 low-level predators, 3.81-4.20 mid-level predators, and greater than 4.2 as top-level predators (Supplementary Table 1 in Chapter 3 Supplemental Material). We provide more detail on trophic level data acquisition in the supplementary material (Supplementary Information).

Phylogenetic Reconstruction

DNA alignments were obtained for three mitochondrial and twelve nuclear genes for the 1545 species of reef acanthomorphs from 92 families for which trophic level data were obtained and an outgroup taxa, *Megalops atlanticus*, using the Phylogeny Assembly with Databases pipeline (PHLAWD) (Smith *et al.* 2009). Additional sequences were extracted from whole mitochondrial genomes using the R package *AnnotationBustR* (Borstein & O'Meara 2018b, a) to supplement sequences obtained using PHLAWD (Supplementary Table 2 & Supplementary file 1 in Chapter 3 Supplemental Material). Phylogenetic analyses were performed using RAxML 8.2.8 (Stamatakis 2014) under the GTR+gamma model of sequence evolution partitioned by gene and codon position. For use in comparative analyses, we converted the phylogeny to an ultrametric tree using treePL with twelve fossil calibrations (Smith & O'Meara 2012) (Supplementary Table 5 in Chapter 3 Supplemental Material). We provide more detail on the phylogenetic reconstruction methods used in this study in the supplementary material Supplemental Methods (Chapter 3 Supplemental Material).

Morphological Data Acquisition

We obtained a representative lateral image of each species in our diet dataset from online fish image databases or other available resources (Supplementary Table 1 in Chapter 3

Supplemental Material). A set of thirteen landmarks were digitized on each image using tpsDIG 2.26 (Rohlf 2016) to measure seven functional traits: head length (HL), orbit size (EYE), lower jaw length (LJL), body depth (BD), caudal peduncle length (PL), caudal peduncle depth (PD), and the caudal fin aspect ratio (AR) (Supplementary Figure S1, Supplementary Table 3 in Chapter 3 Supplemental Material). All functional traits except caudal fin aspect ratio were size-corrected to account for allometry by taking the residuals from a $\log_{10}$ - $\log_{10}$ regression on standard length. In addition to the six size-corrected $\log_{10}$ -transformed functional traits and $\log_{10}$ -transformed caudal fin aspect ratio, we included the $\log_{10}$ -transformed maximum standard length of each species for our analysis obtained using the *rfishbase* species function. These traits are important ecomorphological traits for feeding performance as they relate to the trophic apparatus, prey detection, and locomotion associated with prey capture (Webb 1984; Wainwright & Richard 1995; Myrberg & Fuiman 2002; Schmitz & Wainwright 2011). These eight $\log_{10}$ -transformed traits were subjected to principal components analysis using the *prcomp* function in R to generate a morphospace. We provide a detailed description of the digitization process for generating our phenotypic dataset in the supplementary material Supplemental Methods (Supplementary Information).

Measuring Rates of Trait Evolution

We used the R package *geomorph* to measure the multivariate rate of morphological evolution for each trophic level for the eight measured traits using the *compare.evol.rates* function under 10000 simulations to assess significance (Adams & Otárola-Castillo 2013). We also ran the *compare.evol.rates* function on two subsets of traits, one set of traits for prey acquisition morphology composed of head length, lower jaw length, eye diameter, body depth, and maximum standard length, and another set composed of traits important for locomotion associated with prey capture composed of head length, body depth, caudal peduncle length, caudal peduncle depth, and caudal fin aspect ratio. In addition to the multivariate models described above, we ran each of the eight traits individually to measure the rate of evolution for each trait. As the definition of what constitutes a reef fish is of debate (Bellwood 1998; Robertson 1998), we also ran the above models on a subset of 1306 taxa from 57 families that are found on and characteristic of coral reefs according to Bellwood & Wainwright (2002) as well as on a data set of 1038 taxa from 32 families where at least 50% of extant taxa occur on coral reefs, to limit potential biases in results caused by the inclusion of fishes from families not typically associated with reef environments. To assess if similar patterns occur in smaller clades, we also measured rates of trait evolution in all families containing more than one trophic level category and more than five species (n=29).

Diet Breadth Analysis

We measured diet breadth for each species using the R package *ordiBreadth* (Fordyce *et al.* 2016). This method calculates ordinated diet breadth by subjecting a dissimilarity matrix, in our case a Jaccard dissimilarity matrix, of dietary items to principal coordinates analysis (PCoA) and after ordination, measures ordinated diet breadth as the sum of the distances of a given consumer's multivariate centroid to its consumed prey items. This method has advantages over

taxonomic and phylogenetic measures of diet breadth as it tends to group shared suites of prey that have functionally similar demands (Fordyce *et al.* 2016).

For easier interpretability, we report the scaled diet breadth, which is the scaled measure of occupation a species has in relative diet space. This value is bounded between 0 and 1, representing specialist species with diets consisting solely of one prey type to a hypothetical generalist consumer feeding on all prey items respectively. For consistency among food records, we used the “FoodII” column in the FishBase diet records to generate a binary matrix of eighteen prey items. For species that had volumetric diet data, we removed prey items that comprised less than five percent of the diet to remove prey items that were likely incidentally consumed or play a minor role in the overall diet of the species. The analysis was also repeated with a 1% cutoff (Supplementary Information).

To test if certain trophic levels had narrower diet breadths we performed a phylogenetic ANOVA using the R package *geiger* (Harmon *et al.* 2008) and performed Tukey’s honest significance test to determine if there were significant differences in diet breadth among trophic levels. We would expect that taxa at the extremes of trophic levels would have significantly lower diet breadths compared to taxa at intermediate trophic levels, as to reach the extremes of trophic levels the diets of herbivores and top-level predators must be composed predominantly of plant and nekton prey items respectively while intermediate trophic levels can be obtained by feeding on a variety of prey types, some of which are highly disparate in processing requirements. We also measured ordinated diet breadth as described above for the subset of taxa consisting of reef families according to B&W and 50% of extant taxa occurring on reefs data sets.

Functional Diversity Analysis

To determine if certain trophic levels had more phenotypic functional diversity than other trophic levels we measured the multivariate dispersion at each trophic level, following Anderson *et al.* (2006). We measured the mean multivariate dispersion of each trophic level from its centroid in multivariate trait space using the *betadisper* function on the morphospace described above using the R package *vegan* (Oksanen *et al.* 2015), while correcting for sample size biases following Stier *et al.* (2013). This measure of dispersion has been recommended as a functional diversity index (Laliberté & Legendre 2010). We then performed a phylogenetic ANOVA (Garland Jr *et al.* 1993) in *geiger* (Harmon *et al.* 2008) and performed Tukey’s honest significance test to determine if there were significant differences in functional diversity among trophic levels. We would expect that trophic levels with wider overall diet breadths would have greater multivariate phenotypic dispersion relative to trophic levels with narrower diet breadths following the NVH and that trophic extremes would have reduced multivariate functional dispersion following the HCH. We also measured functional phenotypic diversity as described above for the subset of taxa consisting of reef families according to B&W and 50% of extant taxa occurring on reefs data sets. To assess if similar patterns occur in smaller clades, we also measured functional diversity in all families containing more than one trophic level category and more than five species (n=29).

Results

Trophic Level Distribution

Trophic levels for the 1545 species of reef acanthomorphs ranged from 2.0 for species consuming solely autotrophs and/or detritus to 4.5 for species consuming solely higher trophic level prey. Our final data set was composed of 186 species of herbivores, 200 species of omnivores, 844 low-level predators, 217 mid-level predators, and 98 top-level predators (Figure 3.1, Supplementary Table 4 & 6 in Chapter 3 Supplemental Material). 1000 stochastic character maps of trophic level on the phylogeny recovered a range of 548 and 662 trophic levels transitions across the phylogeny with a mean of 602 transitions across all 1000 mappings.

Morphospace

Morphospace generation through principal components analysis recovered four axes that each accounted for at least 10% of the morphological variation (Figure 3.2, Supplementary Table 7 in Chapter 3 Supplemental Material). The first axis was composed of traits representing overall body depth or elongation and accounted for 31.32% of the variation. Axis two represented 18.82% of the variation and was composed of caudal fin aspect ratio and maximum standard length. Species with high aspect ratios and large maximum standard lengths, such as jacks, tend to occupy similar regions of morphospace together while species with smaller maximum standard lengths and smaller aspect ratios, such as gobies, also tend to occupy similar areas of morphospace. The third axis of variation represented a second elongation axis that was driven by species having deep bodies but elongate lower jaws, such as frogfish, representing 17.52% of the variation. However, contrary to the first PCA axis, this axis is predominantly dominated by differences in body shape as they relate to jaw size. Axis four represented variation in caudal peduncle traits between species with short and deep peduncles, such as angelfish, and species with elongate and narrow peduncles, such as pipefish. These results largely conform with morphospaces obtained in other studies as body depth and elongation is commonly found to be the main axis of variation in fishes (Friedman 2010; Mehta *et al.* 2010; Claverie & Wainwright 2014).

Rates of Morphological Evolution

Herbivores were found to have the fastest rates of overall morphological evolution, as well as the fastest rate of locomotor prey acquisition morphology, caudal fin aspect ratio, head length, lower jaw length, and peduncle length (Table 3.1). Top-level predators had the fastest rates of evolution for rate of prey acquisition morphology, body depth, eye diameter, maximum standard length, and caudal peduncle depth. Top-level predators had the slowest rates of lower jaw length evolution. We found that omnivores had the slowest rates of overall morphological evolution as well as for prey acquisition morphology, body depth, eye diameter, maximum standard length, and peduncle depth. Mid-level predators had the slowest rates of morphological evolution for the multivariate rate of locomotor traits associated with prey capture, aspect ratio, head length, and peduncle length.

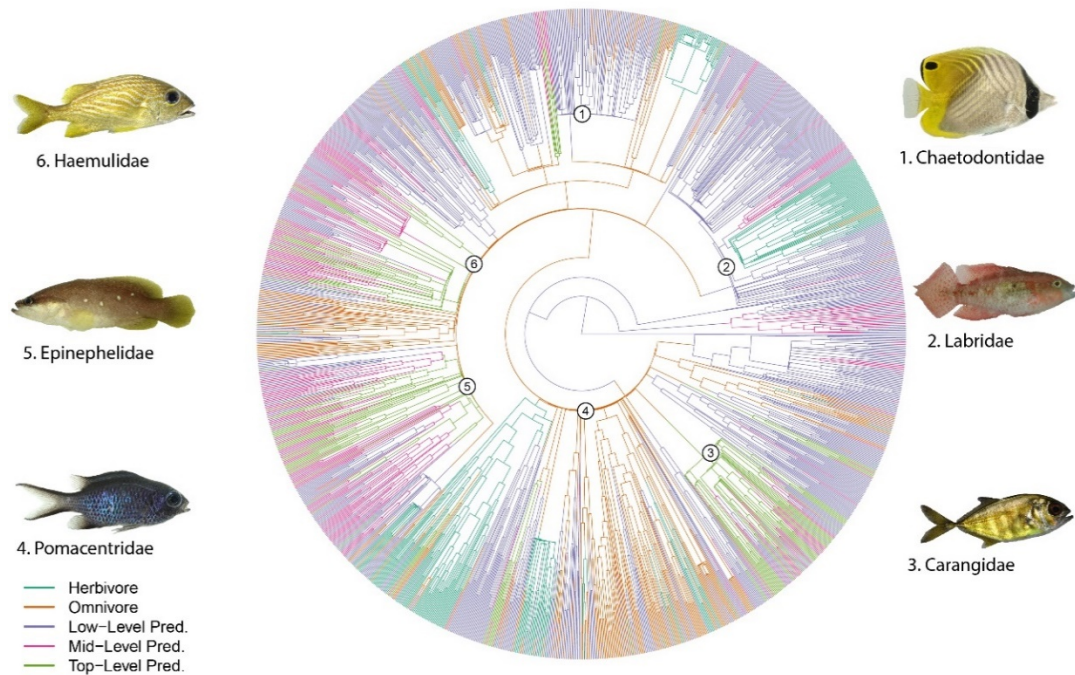


Figure 3.1: Stochastic character mapping reveals over six hundred trophic level transitions across the phylogeny of 1545 acanthomorph reef taxa. Trophic level ranges from 2.0 for species consuming solely autotrophs and detritus to 4.5 for species consuming entirely large nekton. Transitions between the five trophic levels (herbivores, omnivores, low-level predators, mid-level predators and top-level predators), occur frequently throughout the phylogeny. Numbered node labels indicate the phylogenetic position of the families: (1) Chaetodontidae, (2) Labridae, (3) Carangidae, (4) Pomacentridae, (5) Epinephelidae, (6) Haemulidae. Fish photos by M.D. McGee.

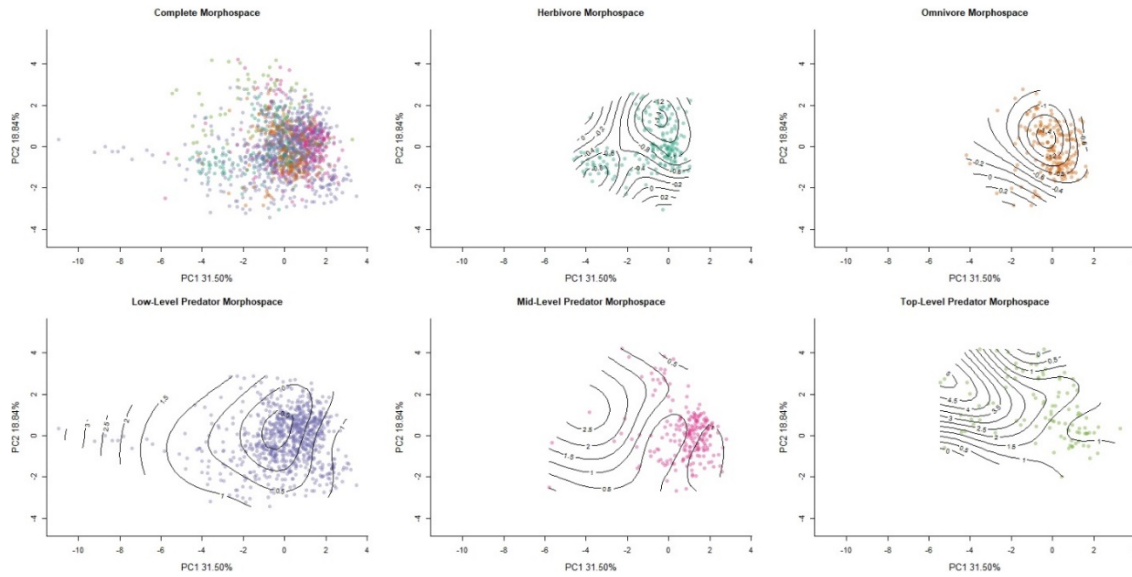


Figure 3.2: Morphospace by trophic level of 1545 reef acanthomorphs based on eight phenotypic functional traits. The morphospace was generated via principle component analysis. PC1 is composed of traits representing differences in body depth and elongation, representing 31.46% of the variation in the morphological data. PC2 represents 18.85% and is driven by changes in body size and caudal fin aspect ratio. Contour lines represent PC3 which accounted for 17.52% of the variation. Each trophic level has been plotted in individual plots scaled to the complete morphospace to show the morphospace occupation at each trophic level. Colors for trophic levels in individual plots are the same as in the complete morphospace of all 1545 reef acanthomorph species.

Table 3.1: Rates of morphological evolution for five trophic levels of 1545 species of coral reef fishes. Results are for eleven different sets of traits, one for the entire data set of eight traits, one composed of five traits for prey acquisition performance, one for five traits important for the locomotion used in prey capture, and for each of the eight individual traits. Rates of each trophic level are presented as the ratio between the rate of evolution for that trophic level relative to lowest calculated rate of evolution.

Trait(s)	Herbivores	Omnivores	Low-level Pred.	Mid-level Pred.	Top-level Pred.	Lowest Rate	p-value
All Traits	2.433	1	1.394	1.383	2.047	0.000268	0.0001
Aspect Ratio	4.982	1.031	1.184	1	1.208	0.000256	0.0001
Body Depth	1.344	1	1.797	1.042	3.044	0.000129	0.0001
Orbit Size	2.135	1	1.824	1.596	2.475	0.00013	0.0001
Head Length	1.952	1.390	1.297	1	1.303	0.0000602	0.0004
Lower Jaw Length	1.647	1.323	1.476	1.437	1	0.00026	0.0119
Locomotor Prey Acq. Traits	3.154	1.091	1.459	1	2.045	0.000175	0.0001
Peduncle Depth	1.786	1	1.391	1.167	3.276	0.000135	0.0001
Peduncle Length	3.573	1.286	1.781	1	2.121	0.000268	0.0001
Prey Acq. Traits	1.961	1	1.442	1.630	2.191	0.000281	0.0001
Max Standard Length	2.449	1	1.525	2.109	2.819	0.000717	0.0001

We found minimal differences between the full data set and the both the Bellwood & Wainwright (2002) (hereinafter B&W) and 50% of extant taxa occurring on reefs data sets (Supplementary Tables 8-9, Supplementary Figures 4-5, in Chapter 3 Supplemental Material). For both data sets herbivores were found to have the fastest rates of morphological evolution for all traits and trait sets except for body depth, eye size, and caudal peduncle depth, for which the fastest rate of evolution was found to be in top-level predators. Omnivores had the slowest overall rate of morphological evolution as well as the slowest rate for prey acquisition feeding morphology, eye diameter, and maximum standard length, while mid-level predators had the slowest rates for locomotor prey acquisition morphology, caudal fin aspect ratio and body depth. Top-level predators had the slowest rates of morphological evolution for lower jaw length and peduncle length in both data sets. The slowest rate of evolution for head length was found to be the mid-level predators and top-level predators for the B&W and 50% of extant taxa on coral reefs respectively.

Analyses limited to the family level exhibited few significant differences in rates of phenotypic evolution, and these differences were not always congruent with the global pattern in our dataset (Supplementary Table 18 in Chapter 3 Supplemental Material). Significant rates within a family were often the result of very few dietary transitions and serve to highlight the issues involved in inferring general patterns from small taxonomic groups with few transitions in the character of interest (Supplementary Information).

Diet Breadth

Herbivores had the narrowest diet breadth with a mean breadth of 0.018 followed by top-level predators, mid-level predators, low-level predators, and omnivores with diet breadths of 0.039, 0.072, 0.094, and 0.101 respectively (Figure 3.3, Supplementary Table 10, Supplementary Figure 6 in Chapter 3 Supplemental Material). Results of the phylogenetic ANOVA ($F_{4,1540}=46.18$, $p = 0.002$) recovered all pairwise comparisons of diet breadth between trophic levels to be significantly different from each other except among top-level predators and herbivores ($p=0.701$) and omnivores and low-level predators ($p=0.203$) (Figure 3.4). Phylogenetic ANOVA performed on the reduced B&W and 50% of extant taxa occurring on reefs data sets yielded similar results to those of the full dataset and are discussed in detail in the supplemental material (Figures S2, S4-S5, Tables S12 & S15 in Chapter 3 Supplemental Material).

Multivariate Functional Diversity

We found that top-level predators had the greatest morphological dispersion, representing a higher amount of phenotypic functional diversity. Top-level predators had a dispersion value from the centroid of 3.115 while low-level predators, herbivores, mid-level predators, and omnivores had multivariate dispersion values of 2.579, 2.207, 2.204 and 1.873 respectively (Figure 3.3, Supplementary Table 11 in Chapter 3 Supplemental Material). Phylogenetic ANOVA ($F_{4,1540}=32.4$, $p = 0.005$) revealed that all trophic levels were significantly different from each other in morphological dispersion, except between herbivores and mid-level predators ($p=0.99$) (Figure 3.4).

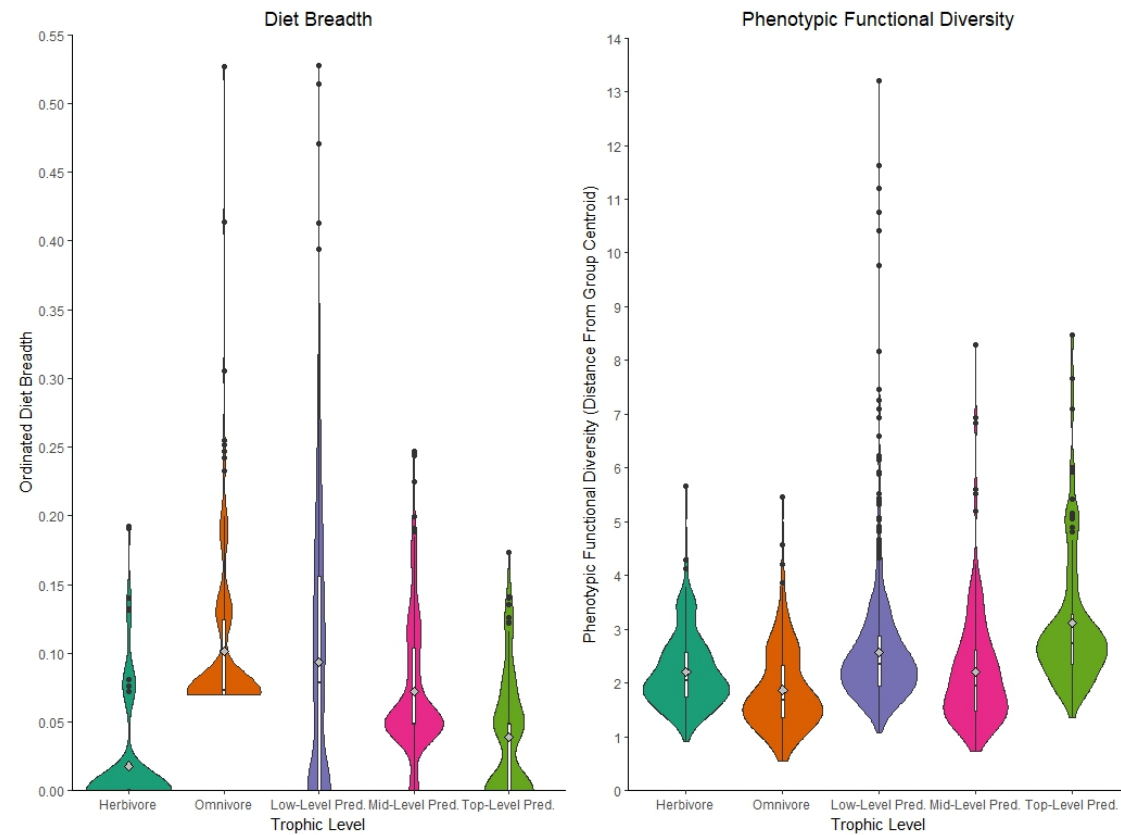


Figure 3.3: Violin plot of diet breadth (left) and phenotypic functional diversity (right) at each trophic level. A boxplot overlaid represents the quantile range while the grey diamond is the mean. Diet breadth measured using the R package *ordiBreadth* for each trophic level show that the extreme trophic levels have the narrowest diet breadths, while omnivores have the widest. Functional diversity values are distances from the centroid measured using the *betadisper* function in *vegan* for eight ecomorphological traits. Phylogenetic ANOVA revealed all trophic levels are significantly different from each other except for between mid-level predators and herbivores.

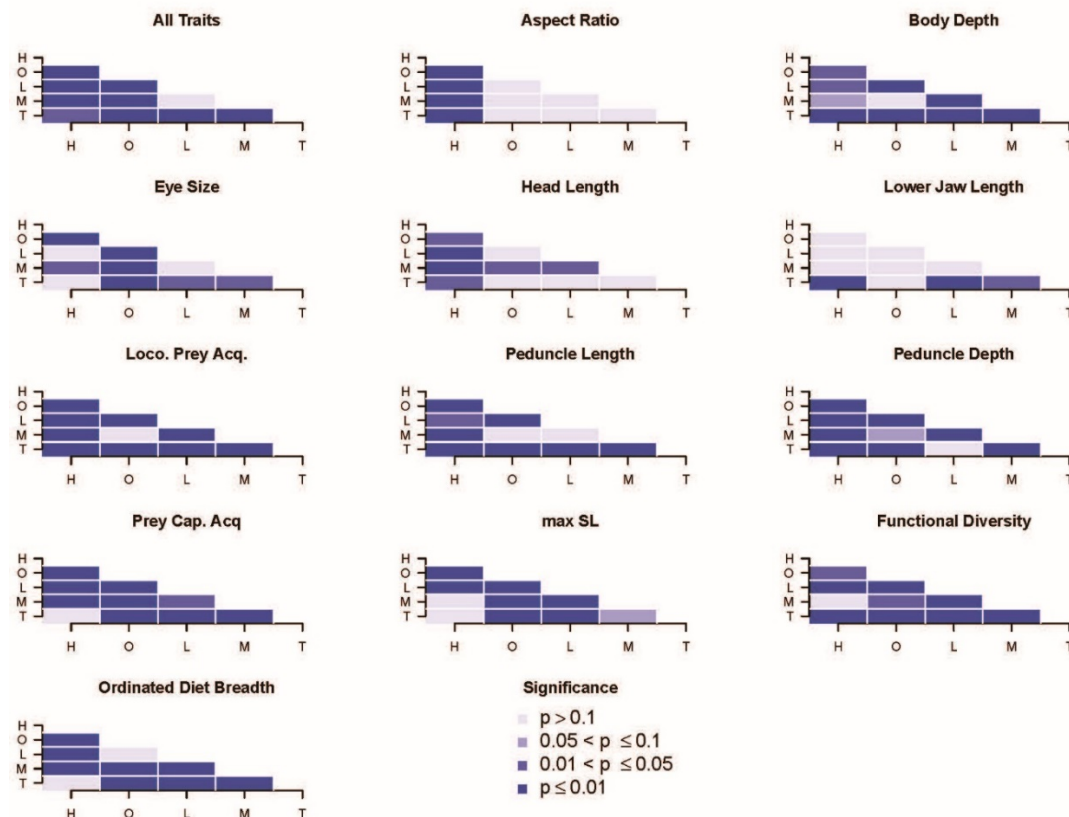


Figure 3.4: Summary of multiple pairwise comparisons for rates of morphological evolution, functional diversity, and diet breadth amongst trophic levels. Trophic levels are denoted by a single letter: H=herbivore, O=omnivore, L= low-level predator, M=mid-level predator, and T=top-level predator. Significance for rates of phenotypic evolution was determined via phylogenetic simulation while phylogenetic ANOVA was used to test for significant differences in functional diversity and diet breadth. Darker values represent lower p-values and significant differences while lighter values represent higher p-values where comparisons are non-significant

We do find slightly different results on the reduced B&W and 50% of extant taxa on reef data sets. While we consistently recover top-level predators as having the highest functional diversity and omnivores the lowest, the functional diversity values for herbivores approach those of omnivores in both datasets. Additionally, the estimated functional diversity values for top-level predators and low-level predators are similar and are not significantly different. We provide more detail on these results in the electronic supplementary material (Supplementary Information, Figures S2, S4-S5, Tables S14 & S16 in Chapter 3 Supplemental Material).

Eighteen families had significant differences in functional diversity (Supplementary Tables 19-20 in Chapter 3 Supplemental Material), but these results are heavily driven by trophic levels occupied by only a single taxon, which produce a functional diversity value of zero (Chapter 3 Supplemental Material).

Discussion

We find that species at the extremes of trophic level, which have reduced diet breadths relative to other trophic levels, have the fastest rates of morphological evolution, and for the top-level predators, more functional phenotypic diversity. This result is contrary to what would be expected by the niche variation hypothesis. In addition, we consistently find that omnivores have the slowest rate of morphological evolution as well as the lowest amount of phenotypic functional diversity. While studies in other organisms find contrasting evidence for omnivory constraining morphological evolution (Herrel *et al.* 2004; Renaud *et al.* 2007), it is possible that functional constraints associated with feeding on functionally disparate prey sources limit morphospace to a generalized morphology in order to exploit a variety of prey (Huey & Hertz 1984).

Our results differ from those found by previous height constraint hypothesis (HCH)-supporting studies that suggested feeding at high trophic levels might constrain morphological evolution in fishes (Collar *et al.* 2009; Smith *et al.* 2015b; Svanbäck *et al.* 2015), as well as non-fish taxa (Holliday & Stepan 2004). One reason could be the broad phylogenetic sampling of our data set in contrast to previous studies. Piscivores consistently have the slowest rate of lower jaw length evolution across all data sets. Lower jaw size is associated with gape size, a key factor governing suction feeding performance on large, elusive prey (Holzman *et al.* 2012) as well as a constraint to the size of prey that is possible for an individual to ingest and is likely under strong selection (Hoyle & Keast 1987; Wainwright & Richard 1995; Holzman *et al.* 2008; Mihalitsis & Bellwood 2017). However, our data reveals that top predators exhibit extreme diversity in other traits. Rates of body depth and peduncle depth evolution in top-level predators are also a significant contribution to them consistently being recovered as having the second-fastest rates of phenotypic evolution throughout all data sets. These two traits are directly related to feeding performance. Body depth is partly composed of the epaxial muscles on the dorsal portion of the fish. These muscles play a crucial role in suction feeding performance, with deeper bodied species having a larger cross-sectional area (Carroll *et al.* 2004; Collar & Wainwright 2006; Holzman *et al.* 2012). Peduncle depth is thought to play a prominent role in the ability to sustain searching for food over long distances in cruising predators by reducing laterally directed drag

incurred by the oscillating tail (Lighthill 1969) or the ability to achieve high acceleration associated with a rapid strike towards an evasive prey item (Webb 1984).

Another explanation for the increased rate of evolution and morphological disparity in top-level predators is that the diversification of fish species at lower trophic levels which form the prey base of top-level predators may have driven ecological and phenotypic diversification in top-level predators, a pattern coined “upward adaptive radiation cascades” (Brodersen *et al.* 2017). While our study does not directly address this question, both theory (Brown & Vincent 1992) and empirical studies (Forbes *et al.* 2009; Hood *et al.* 2015) suggest such a pattern of co-evolution among trophic levels occurs, and we believe that it warrants future study, especially in reef systems, which as biodiversity hotspots harbor a wealth of prey items and have complex trophic interactions.

We find mixed evidence for the trophic extremes hypothesis (TEH). While we recover herbivores and top-level predators as having faster rates of phenotypic evolution relative to other trophic levels, supporting the TEH, we find that only top-level predators display high phenotypic functional diversity. While herbivores have the fastest rate of phenotypic evolution in our analyses, we do not recover them as having high functional phenotypic diversity. This result is stronger in the two sub-datasets due to the exclusion of some rather extreme herbivores, such as halfbeaks (family Hemiramphidae) that do not meet the criteria for the B&W or 50% of extant taxa on reef data sets and suggests there are certain constraints on herbivores in relation to possible morphospace occupation

One contributor to the elevated rates of evolution found in both herbivore and top-level predators is body size. Consistently, we recover the trophic extremes as having the fastest rates of body size evolution across all data sets. While there is conflicting evidence looking at broad-scale patterns among body size and trophic level in fish (Layman *et al.* 2005; Romanuk *et al.* 2011), we believe the elevated rates of body size evolution at the extreme trophic levels is best described by microhabitat partitioning among the reef in which large reef cruising phenotypes and smaller phenotypes associating with high levels of reef structure and interstitial spaces are both successful (Ansell *et al.* 1998).

Our results suggest that there are multiple ways to be a successful top-level predator. While our study does not specifically test many-to-one mapping of morphology to trophic level, a many-to-one mapping of morphology to trophic ecology may explain why top-level predators occupy a large area of morphospace and warrants future research (Alfaro *et al.* 2005; Wainwright *et al.* 2005). While we find that low-level predators have the second highest value for phenotypic variation and are significantly different from all other trophic levels, this result is not surprising due to the sheer number of taxa that fall in this group, nearly four times as many taxa than any other trophic level. While the distribution of taxa in each trophic level roughly follows the known distribution for global fish diversity (Froese 2005), the low-level predators encompass a large amount of diversity in terms of ecomorphology and prey use, including some of the most enigmatic and extreme species in our data set, like pipefish. The high rates of phenotypic evolution in herbivores yet low amounts of functional diversity suggest that there are a limited number of ways to consume algae and other plant matter, at least with our measured

functional traits. We would expect that incorporation of other functional morphological traits, such as tooth shape (Jones 1968; Bellwood *et al.* 2014b), shape and size of the digestive tract (Jones 1968; German *et al.* 2015; Konow *et al.* 2017), and the presence of intra-mandibular joints (Konow *et al.* 2008; Gibb *et al.* 2015; Konow *et al.* 2017) would likely elevate the level of functional diversity measured within this group as it relates to partitioning of trophic resources within herbivores (Clements *et al.* 2017). Another possible explanation to the high rates of morphological evolution yet low functional phenotypic diversity in herbivores is that herbivory and detritivory are often relatively recent evolutionary transitions among reef acanthomorphs (Bellwood *et al.* 2014a; Bellwood *et al.* 2017). Under a Brownian motion model of trait evolution, trait variance is expected to increase over evolutionary time. Therefore, recent transitions to herbivory may have led to the fast rates of phenotypic evolution we observed, yet limit possible dispersion into more extreme phenotypes, leading to limited functional phenotypic diversity. Transitions to diets low in protein and fat have also been linked to increased diversification rates, which is likely due to ecological opportunity and exploitation of such resources (Harmelin-Vivien 2002; Lobato *et al.* 2014), and integrating diversification with phenotypic and trophic evolutionary patterns would prove fruitful in elucidating the macroevolutionary factors governing reef diversity.

One criticism of the NVH is that it is generally tied to morphological variation while the hypothesis could pertain to any phenotypic trait, including behavioral traits (Bolnick *et al.* 2007). While we've discussed some possible factors influencing the elevated rates of morphological evolution at extreme trophic levels, foraging behavior likely has a role in shaping the evolution of these traits. For example, one component of the elevated morphological rates measured in herbivorous species might be due to divergence between strategies of active browsing for algae and plants (Puk *et al.* 2016) versus the strategy of algal gardening, in which territorial species defend a patch of reef from other herbivores and forage algae from within this territory (Branch *et al.* 1992). This might explain the elevated rates of caudal peduncle length and caudal fin aspect ratio observed in herbivorous species as it relates to their foraging strategy. Many browsers typically are active swimmers over reefs or in turbulent reef zones, while the more territorial algal gardeners are far more sedentary, where a lower aspect ratio is better suited for protecting their algal garden and renders them capable of quick short bursts chasing intruders away (Bejarano *et al.* 2017). Differences in foraging strategies are also apparent for top predators, which hunt with stalking, ambush, or pursuit strategies (Hobson 1975, 1979; Hixon 1991). Even within these hunting strategy categories, there is considerable variation in body forms. For example, two piscivorous predators, the fishes of the family Aulostomidae and fishes of the family Antennariidae, both employ ambush methods of hunting despite being on opposite extremes of morphospace (Kaufman 1976; Pietsch & Grobecker 1978; Aronson 1983).

The trophic levels experiencing the fastest rates of functional evolution – top-level predators and herbivores – are also heavily targeted by reef fisheries. Many reef fisheries focus on top-level predators, leading to widespread overfishing of iconic coral reef predators like groupers (Rhodes & Tupper 2007). Substantial fishing pressure also affects herbivores, especially large grazers like parrotfish and rabbitfish, and overfishing of these species can substantially affect algal biomass and coral cover (Bellwood *et al.* 2003; Bellwood *et al.* 2004;

Hawkins & Roberts 2004). Unfortunately, because many reef fisheries are at subsistence level and many reef species occur on wide geographic scales, it can be challenging to assess fishing pressure in an integrated way (Roberts 1995). We suggest that a focus on trophic position and functional traits can provide a key avenue for integrating evolutionary ecology with sustainability and conservation, particularly relative to traditional single species-focused approaches.

It is important to note that our study does not address non-trophic selective pressures that may be important in shaping the morphological diversity of reef assemblages. While our main axis of diversification revolves around the depth or elongation of the body, which we have discussed above as being important for prey capture, it is also important for defense from predators (Brönmark & Miner 1992; Price *et al.* 2015). Modes of locomotion are also quite diverse in reef fishes, involving various uses in fins, which are thought to have consequences for swimming performance, body form, and fin design (Sfakiotakis *et al.* 1999; Wainwright *et al.* 2002). Additionally, our study does not contain data on myriad other morphological and physiological aspects that play a role in shaping the trophic ecology of species, and we welcome the addition of increasingly sophisticated sets of functional traits in future studies. At the deep phylogenetic scale of our analysis, interspecific variation is likely far more prevalent than intraspecific variation. However, as our study was performed on a single representative photograph of a species, it does not capture possible intraspecific variation due to factors such as sexual dimorphism or body condition. It is also possible that utilizing more complex models, such as Ornstein-Uhlenbeck processes (Hansen 1997; Beaulieu *et al.* 2012) might reveal other aspects of the trait evolution process. We also note that methods for examining rates of continuous trait evolution, while they do control for phylogeny, may be subject to issues identified by Maddison and FitzJohn (2015): a rate increase in one clade of top-level predators and no rate change in any other top-level predator clade, would still be significant under the tests used here. Our results at the family level highlight the utility of large-scale comparative analyses to adequately account for transitions and understand broad macroevolutionary patterns.

Our results suggest limited diet breadth and the extremes of trophic level do not constrain functional diversity and phenotypic evolution in reef fishes. This result is contrary to that predicted of the NVH, indicating that limited niche breadth does not constrain evolutionary rates and phenotypic diversity. Our results provide the most support for the TEH, as species at the extremes of trophic level have the fastest rates of morphological evolution. We suggest that trophic ecology plays a critical role in assemblage functional trait evolution on macroevolutionary timescales.

Data Availability

Data and scripts used in this study are stored in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.7t3d30c>), which is open access.

References

- Adams, D.C. & Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393-399.
- Agrawal, A.A. (2007). Macroevolution of plant defense strategies. *Trends Ecol Evol*, 22, 103-109.
- Alfaro, M.E., Bolnick, D.I. & Wainwright, P.C. (2005). Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes. *American Naturalist*, 165, E140-E154.
- Alfaro, M.E., Santini, F. & Brock, C.D. (2007). Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). *Evolution; international journal of organic evolution*, 61, 2104-2126.
- Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecol Lett*, 9, 683-693.
- Ansell, A., Gibson, R., Barnes, M. & Press, U. (1998). The ecological implications of small body size among coral-reef fishes. *Oceanography and marine biology: an annual review*, 36, 373-411.
- Araújo, M.S., Bolnick, D.I. & Layman, C.A. (2011). The ecological causes of individual specialisation. *Ecol Lett*, 14, 948-958.
- Aronson, R.B. (1983). Foraging behavior of the west Atlantic trumpetfish, *Aulostomus maculatus*: Use of large, herbivorous reef fishes as camouflage. *Bulletin of Marine Science*, 33, 166-171.
- Barton, Kasey E. & Koricheva, J. (2010). The Ontogeny of Plant Defense and Herbivory: Characterizing General Patterns Using Meta-Analysis. *The American naturalist*, 175, 481-493.
- Beaulieu, J.M., Jhwueng, D.C., Boettiger, C. & O'Meara, B.C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution; international journal of organic evolution*, 66, 2369-2383.
- Bejarano, S., Jouffray, J.-B., Chollett, I., Allen, R., Roff, G., Marshall, A. *et al.* (2017). The shape of success in a turbulent world: wave exposure filtering of coral reef herbivory. *Funct Ecol*, 31, 1312-1324.
- Bellwood, D. (1998). What are reef fishes?-Comment on the report by DR Robertson: Do coral-reef fish faunas have a distinctive taxonomic structure?(Coral Reefs 17: 179-186). *Coral Reefs*, 17, 187-189.
- Bellwood, D.R., Goatley, C.H. & Bellwood, O. (2017). The evolution of fishes and corals on reefs: form, function and interdependence. *Biol Rev*, 92, 878-901.
- Bellwood, D.R., Goatley, C.H., Brandl, S.J. & Bellwood, O. (2014a). Fifty million years of herbivory on coral reefs: fossils, fish and functional innovations. *Proceedings. Biological sciences / The Royal Society*, 281, 20133046.
- Bellwood, D.R., Hoey, A.S., Bellwood, O. & Goatley, C.H. (2014b). Evolution of long-toothed fishes and the changing nature of fish-benthos interactions on coral reefs. *Nature communications*, 5, 3144.
- Bellwood, D.R., Hoey, A.S. & Choat, J.H. (2003). Limited functional redundancy in high diversity systems: resilience and ecosystem function on coral reefs. *Ecol Lett*, 6, 281-285.

- Bellwood, D.R., Hughes, T.P., Folke, C. & Nyström, M. (2004). Confronting the coral reef crisis. *Nature*, 429, 827.
- Bellwood, D.R. & Wainwright, P.C. (2002). *The history and biogeography of fishes on coral reefs*. Academic Press, San Diego, California.
- Boettiger, C., Lang, D.T. & Wainwright, P.C. (2012). rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81, 2030-2039.
- Bolnick, D.I., Svanbäck, R., Araújo, M.S. & Persson, L. (2007). Comparative support for the niche variation hypothesis that more generalized populations also are more heterogeneous. *Proceedings of the National Academy of Sciences*, 104, 10075-10079.
- Bolnick, D.I., Svanback, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D. *et al.* (2003). The Ecology of Individuals: Incidence and Implications of Individual Specialization. *American Naturalist*, 161, 1-28.
- Borstein, S.R. & O'Meara, B.C. (2018a). *AnnotationBustR*: an R package to extract subsequences from GenBank annotations. *PeerJ*, 6.
- Borstein, S.R. & O'Meara, B.C. (2018b). R Package 'AnnotationBustR'. 1.2
- Branch, G., Harris, J., Parkins, C., Bustamante, R. & Eekhout, S. (1992). Algal 'gardening' by grazers: a comparison of the ecological effects of territorial fish and limpets. *Plant-animal interactions in the marine benthos*, 46, 405-423.
- Brodersen, J., Post, D.M. & Seehausen, O. (2017). Upward adaptive radiation cascades: Predator diversification induced by prey diversification. *Trends Ecol Evol*.
- Brönmark, C. & Miner, J.G. (1992). Predator-induced phenotypical change in body morphology in crucian carp. *Science*, 258, 1348-1348.
- Brown, J.S. & Vincent, T.L. (1992). Organization of predator-prey communities as an evolutionary game. *Evolution; international journal of organic evolution*, 46, 1269-1283.
- Cadotte, M.W., Cavender-Bares, J., Tilman, D. & Oakley, T.H. (2009). Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PloS one*, 4, e5695.
- Cardinale, B.J., Srivastava, D.S., Duffy, J.E., Wright, J.P., Downing, A.L., Sankaran, M. *et al.* (2006). Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature*, 443, 989-992.
- Carpenter, K.E., Abrar, M., Aeby, G., Aronson, R.B., Banks, S., Bruckner, A. *et al.* (2008). One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science*, 321, 560-563.
- Carroll, A.M., Wainwright, P.C., Huskey, S.H., Collar, D.C. & Turingan, R.G. (2004). Morphology predicts suction feeding performance in centrarchid fishes. *The Journal of experimental biology*, 207, 3873-3881.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol Lett*, 12, 693-715.
- Claverie, T. & Wainwright, P.C. (2014). A morphospace for reef fishes: elongation is the dominant axis of body shape evolution. *PloS one*, 9, e112732.
- Clements, K.D., German, D.P., Piché, J., Tribollet, A. & Choat, J.H. (2017). Integrating ecological roles and trophic diversification on coral reefs: multiple lines of evidence identify parrotfishes as microphages. *Biol J Linn Soc*, 120, 729-751.

- Collar, D.C., O'Meara, B.C., Wainwright, P.C. & Near, T.J. (2009). Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution; international journal of organic evolution*, 63, 1557-1573.
- Collar, D.C. & Wainwright, P.C. (2006). Discordance between morphological and mechanical diversity in the feeding mechanism of centrarchid fishes. *Evolution; international journal of organic evolution*, 60, 2575-2584.
- Cooper, W.J. & Westneat, M.W. (2009). Form and function of damselfish skulls: rapid and repeated evolution into a limited number of trophic niches. *BMC evolutionary biology*, 9, 24.
- Cowman, P.F., Bellwood, D.R. & van Herwerden, L. (2009). Dating the evolutionary origins of wrasse lineages (Labridae) and the rise of trophic novelty on coral reefs. *Molecular phylogenetics and evolution*, 52, 621-631.
- Davis, A.M., Unmack, P.J., Vari, R.P., Betancur-R, R., Heard, S.B. & Michalakis, Y. (2016). Herbivory Promotes Dental Disparification and Macroevolutionary Dynamics in Grunters (Teleostei: Therapontidae), a Freshwater Adaptive Radiation. *The American naturalist*, 000-000.
- de Graaf, M., Machiels, M.A.M., Wudneh, T. & Sibbing, F.A. (2004). Declining stocks of Lake Tana's endemic *Barbus* species flock (Pisces, Cyprinidae): natural variation or human impact? *Biological Conservation*, 116, 277-287.
- Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thébault, E. & Loreau, M. (2007). The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol Lett*, 10, 522-538.
- Estes, S. & Arnold, S.J. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American naturalist*, 169, 227-244.
- Fedosov, A., Tiunov, A., Kiyashko, S. & Kantor, Y.I. (2014). Trophic diversification in the evolution of predatory marine gastropods of the family Terebridae as inferred from stable isotope data. *Marine Ecology Progress Series*, 497, 143-156.
- Findley, J.S. & Black, H. (1983). Morphological and dietary structuring of a Zambian insectivorous bat community. *Ecology*, 64, 625-630.
- Floeter, S.R., Bender, M.G., Siqueira, A.C. & Cowman, P.F. (2018). Phylogenetic perspectives on reef fish functional traits. *Biol Rev*, 93, 131-151.
- Forbes, A.A., Powell, T.H., Stelinski, L.L., Smith, J.J. & Feder, J.L. (2009). Sequential sympatric speciation across trophic levels. *Science*, 323, 776-779.
- Fordyce, J.A., Nice, C.C., Hamm, C.A. & Forister, M.L. (2016). Quantifying diet breadth through ordination of host association. *Ecology*, 97, 842-849.
- Frederich, B., Olivier, D., Litsios, G., Alfaro, M.E. & Parmentier, E. (2014). Trait decoupling promotes evolutionary diversification of the trophic and acoustic system of damselfishes. *P Roy Soc B-Biol Sci*, 281.
- Friedman, M. (2010). Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20092177.
- Froese, R. (2005). Life-history strategies of recent fishes: a meta-analysis. Christian-Albrechts-Universität.

- Froese, R. & Pauly, D. (2000). *FishBase 2000: Concepts Designs and Data Sources*. ICLARM, Laguna, Philippines.
- Froese, R. & Pauly, D. (2018). FishBase. Available at: <http://www.fishbase.org/2018>.
- Galeotti, P. & Rubolini, D. (2003). The niche variation hypothesis and the evolution of colour polymorphism in birds: a comparative study of owls, nightjars and raptors. *Biol J Linn Soc*, 82, 237-248.
- Garland Jr, T., Dickerman, A.W., Janis, C.M. & Jones, J.A. (1993). Phylogenetic analysis of covariance by computer simulation. *Systematic biology*, 42, 265-292.
- German, D.P., Sung, A., Jhaveri, P. & Agnihotri, R. (2015). More than one way to be an herbivore: convergent evolution of herbivory using different digestive strategies in prickleback fishes (Stichaeidae). *Zoology*, 118, 161-170.
- Gibb, A.C., Staab, K., Moran, C. & Ferry, L.A. (2015). The Teleost Intramandibular Joint: A mechanism That Allows Fish to Obtain Prey Unavailable to Suction Feeders. *Integr Comp Biol*, 55, 85-96.
- Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution; international journal of organic evolution*, 51, 1341-1351.
- Harmelin-Vivien, M.L. (2002). *Energetics and fish diversity*.
- Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24, 129-131.
- Hawkins, J.P. & Roberts, C.M. (2004). Effects of artisanal fishing on Caribbean coral reefs. *Conserv Biol*, 18, 215-226.
- Hay, M.E. & Fenical, W. (1988). Marine plant-herbivore interactions: the ecology of chemical defense. *Annu Rev Ecol Syst*, 19, 111-145.
- Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M., Diemer, M., Dimitrakopoulos, P. *et al.* (1999). Plant diversity and productivity experiments in European grasslands. *Science*, 286, 1123-1127.
- Herrel, A., Vanhooydonck, B. & Van Damme, R. (2004). Omnivory in lacertid lizards: adaptive evolution or constraint? *Journal of evolutionary biology*, 17, 974-984.
- Hixon, M. (1991). Predation as a process structuring coral reef fish communities. In: *The ecology of fishes on coral reefs*. Academic Press, San Diego (ed. Sale, PF).
- Hobson, E.S. (1975). Feeding patterns among tropical reef fishes. *Am Sci*, 63, 382-392.
- Hobson, E.S. (1979). Interactions between piscivorous fishes and their prey. In: *Predator-prey systems in fisheries management* (ed. Clepper, H). Sport Fishing Institute Washington, DC, pp. 231-242.
- Holliday, J.A. & Stepan, S.J. (2004). Evolution of hypercarnivory: the effect of specialization on morphological and taxonomic diversity. *Paleobiology*, 30, 108-128.
- Holzman, R., Collar, D.C., Mehta, R.S. & Wainwright, P.C. (2012). An integrative modeling approach to elucidate suction-feeding performance. *The Journal of experimental biology*, 215, 1-13.
- Holzman, R., Day, S.W., Mehta, R.S. & Wainwright, P.C. (2008). Jaw protrusion enhances forces exerted on prey by suction feeding fishes. *Journal of the Royal Society Interface*, 5, 1445-1457.

- Hood, G.R., Forbes, A.A., Powell, T.H., Egan, S.P., Hamerlinck, G., Smith, J.J. *et al.* (2015). Sequential divergence and the multiplicative origin of community diversity. *Proceedings of the National Academy of Sciences*, 112, E5980-E5989.
- Hoyle, J.A. & Keast, A. (1987). The effect of prey morphology and size on handling time in a piscivore, the largemouth bass (*Micropterus salmoides*) *Can J Zool*, 65, 1972-1977.
- Hsu, Y.C., Shaner, P.J., Chang, C.I., Ke, L. & Kao, S.J. (2014). Trophic niche width increases with bill-size variation in a generalist passerine: a test of niche variation hypothesis. *J Anim Ecol*, 83, 450-459.
- Huey, R.B. & Hertz, P.E. (1984). Is a jack-of-all-temperatures a master of none? *Evolution; international journal of organic evolution*, 38, 441-444.
- Jones, R.S. (1968). Ecological relationships in Hawaiian and Johnston Island Acanthuridae (surgeonfishes). *Micronesica*, 4, 309-361.
- Kaufman, L. (1976). Feeding behavior and functional coloration of the Atlantic trumpetfish, *Aulostomus maculatus*. *Copeia*, 1976, 377-378.
- Konow, N., Bellwood, D.R., Wainwright, P.C. & Kerr, A.M. (2008). Evolution of novel jaw joints promote trophic diversity in coral reef fishes. *Biol J Linn Soc*, 93, 545-555.
- Konow, N., Price, S., Abom, R., Bellwood, D. & Wainwright, P. (2017). Decoupled diversification dynamics of feeding morphology following a major functional innovation in marine butterflyfishes. *Proceedings. Biological sciences / The Royal Society*, 284.
- Laliberté, E. & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91, 299-305.
- Layman, C.A., Winemiller, K.O., Arrington, D.A. & Jepsen, D.B. (2005). Body size and trophic position in a diverse tropical food web. *Ecology*, 86, 2530-2535.
- Lighthill, M. (1969). Hydromechanics of aquatic animal propulsion. *Annual review of fluid mechanics*, 1, 413-446.
- Lobato, F.L., Barneche, D.R., Siqueira, A.C., Liedke, A.M., Lindner, A., Pie, M.R. *et al.* (2014). Diet and diversification in the evolution of coral reef fishes. *PloS one*, 9, e102094.
- Lundvall, D., Svanbäck, R., Persson, L. & Byström, P. (1999). Size-dependent predation in piscivores: interactions between predator foraging and prey avoidance abilities. *Can J Fish Aquat Sci*, 56, 1285-1292.
- Maddison, W.P. & FitzJohn, R.G. (2015). The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic biology*, 64, 127-136.
- Mancinelli, G., Vizzini, S., Mazzola, A., Maci, S. & Basset, A. (2013). Cross-validation of $\delta^{15}\text{N}$ and FishBase estimates of fish trophic position in a Mediterranean lagoon: The importance of the isotopic baseline. *Estuarine, Coastal and Shelf Science*, 135, 77-85.
- Martin, C.H. & Wainwright, P.C. (2011). Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfish. *Evolution; international journal of organic evolution*, 65, 2197-2212.
- McGee, M.D., Faircloth, B.C., Borstein, S.R., Zheng, J., Darrin Hulsey, C., Wainwright, P.C. *et al.* (2016). Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. *Proceedings of the Royal Society B: Biological Sciences*, 283.
- Mehta, R.S., Ward, A.B., Alfaro, M.E. & Wainwright, P.C. (2010). Elongation of the body in eels. *Integr Comp Biol*, 50, 1091-1105.

- Mihalitsis, M. & Bellwood, D.R. (2017). A morphological and functional basis for maximum prey size in piscivorous fishes. *PloS one*, 12, e0184679.
- Myrberg, A. & Fuiman, L.A. (2002). The sensory world of coral reef fishes. *Coral reef fishes: dynamics and diversity in a complex ecosystem. Academic, San Diego*, 123-148.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R. & O'Hara, R.B. (2015). *vegan: Community Ecology Package*. 2.2-1
- Paré, P.W. & Tumlinson, J.H. (1999). Plant volatiles as a defense against insect herbivores. *Plant physiology*, 121, 325-332.
- Pietsch, T.W. & Grobecker, D.B. (1978). The compleat angler: Aggressive mimicry in an antennariid anglerfish. *Science*, 201, 369-370.
- Price, S., Friedman, S. & Wainwright, P. (2015). How predation shaped fish: the impact of fin spines on body form evolution across teleosts. In: *Proc. R. Soc. B. The Royal Society*, p. 20151428.
- Price, S.A., Holzman, R., Near, T.J. & Wainwright, P.C. (2011). Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecol Lett*, 14, 462-469.
- Price, S.A., Tavera, J.J., Near, T.J. & Wainwright, P.C. (2013). Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. *Evolution; international journal of organic evolution*, 67, 417-428.
- Puk, L.D., Ferse, S.C. & Wild, C. (2016). Patterns and trends in coral reef macroalgae browsing: a review of browsing herbivorous fishes of the Indo-Pacific. *Reviews in fish biology and fisheries*, 26, 53-70.
- R Development Core Team (2018). R: A language and environment for statistical computing. Available at: <http://www.R-project.org/>.
- Renaud, S., Chevret, P. & Michaux, J. (2007). Morphological vs. molecular evolution: ecology and phylogeny both shape the mandible of rodents. *Zoologica scripta*, 36, 525-535.
- Rhodes, K.L. & Tupper, M.H. (2007). A preliminary market-based analysis of the Pohnpei, Micronesia, grouper (Serranidae: Epinephelinae) fishery reveals unsustainable fishing practices. *Coral Reefs*, 26, 335-344.
- Roberts, C.M. (1995). Effects of fishing on the ecosystem structure of coral reefs. *Conserv Biol*, 9, 988-995.
- Robertson, D. (1998). Do coral-reef fish faunas have a distinctive taxonomic structure? *Coral Reefs*, 17, 179-186.
- Rohlf, F.J. (2016). tpsDIG 2.26
- Romanuk, T.N., Hayward, A. & Hutchings, J.A. (2011). Trophic level scales positively with body size in fishes. *Global Ecology and Biogeography*, 20, 231-240.
- Santini, F., Nguyen, M.T., Sorenson, L., Waltzek, T.B., Lynch Alfaro, J.W., Eastman, J.M. *et al.* (2013). Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). *Journal of evolutionary biology*, 26, 1003-1018.
- Schmitz, L. & Wainwright, P.C. (2011). Ecomorphology of the eyes and skull in zooplanktivorous labrid fishes. *Coral Reefs*, 30, 415-428.
- Sfakiotakis, M., Lane, D.M. & Davies, J.B.C. (1999). Review of fish swimming modes for aquatic locomotion. *IEEE Journal of oceanic engineering*, 24, 237-252.

- Smith, A.J., Nelson-Maney, N., Parsons, K.J., Cooper, W.J. & Albertson, R.C. (2015a). Body shape evolution in sunfishes: Divergent paths to accelerated rates of speciation in the centrarchidae. *Evolutionary Biology*, 42, 283-295.
- Smith, A.J., Nelson-Maney, N., Parsons, K.J., James Cooper, W. & Craig Albertson, R. (2015b). Body Shape Evolution in Sunfishes: Divergent Paths to Accelerated Rates of Speciation in the Centrarchidae. *Evolutionary Biology*, 42, 283-295.
- Smith, S.A. & O'Meara, B.C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28, 2689-2690.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.
- Stier, A.C., Geange, S.W., Hanson, K.M. & Bolker, B.M. (2013). Predator density and timing of arrival affect reef fish community assembly. *Ecology*, 94, 1057-1068.
- Stuart-Smith, R.D., Bates, A.E., Lefcheck, J.S., Duffy, J.E., Baker, S.C., Thomson, R.J. *et al.* (2013). Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, 501, 539.
- Svanbäck, R., Quevedo, M., Olsson, J. & Eklöv, P. (2015). Individuals in food webs: the relationships between trophic position, omnivory and among-individual diet variation. *Oecologia*, 178, 103-114.
- Vamosi, S.M. (2003). The presence of other fish species affects speciation in threespine sticklebacks. *Evol Ecol Res*, 5, 717-730.
- Van Valen, L. (1965). Morphological variation and width of ecological niche. *The American naturalist*, 99, 377-390.
- Vermeij, G.J. (1977). The Mesozoic marine revolution: evidence from snails, predators and grazers. *Paleobiology*, 3, 245-258.
- Wainwright, P.C., Alfaro, M.E., Bolnick, D.I. & Hulsey, C.D. (2005). Many-to-one mapping of form to function: A general principle in organismal design? *Integr Comp Biol*, 45, 256-262.
- Wainwright, P.C., Bellwood, D.R. & Westneat, M.W. (2002). Ecomorphology of locomotion in labrid fishes. *Environmental Biology of Fishes*, 65, 47-62.
- Wainwright, P.C., McGee, M.D., Longo, S.J. & Hernandez, L.P. (2015). Origins, Innovations, and Diversification of Suction Feeding in Vertebrates. *Integr Comp Biol*, 55, 134-145.
- Wainwright, P.C. & Richard, B.A. (1995). Predicting patterns of prey use from morphology of fishes. *Environmental Biology of Fishes*, 44, 97-113.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu Rev Ecol Syst*, 475-505.
- Webb, P.W. (1984). Body form, locomotion and foraging in aquatic vertebrates. *Am Zool*, 24, 107-120.

CHAPTER 4
A CLASSIC EVOLUTIONARY INNOVATION, PHARYNGOGNATHY,
DOES NOT LEAD TO INCREASED RATES OF DIVERSIFICATION

A version of this chapter is to be submitted by Samuel Borstein, Matthew McGee and Brian O'Meara as a manuscript to *Nature*:

Borstein, S.R., McGee, M.D. & O'Meara, B.C.O. A classic evolutionary innovation, pharyngognathy, does not lead to increased rates of diversification.

Author contributions: S.R.B. designed the study. S.R.B. and B.C.O. performed the analyses. S.R.B. and B.C.O. wrote the manuscript with substantial comments from M.D.M.

Abstract

Evolutionary innovations, traits that provide species access to novel niches, are widely regarded as important drivers of evolutionary success in the lineages that possess them (Mayr 1963). These traits, sometimes referred to as “key innovations,” are thought promote speciation rates and reduce the risk of extinction (Heard & Hauser 1995; Hunter 1998) and have become a major focus of macroevolutionary studies (Rabosky 2014). However, testing evolutionary innovations adequately is difficult and while historically done in the context of speciation rates. However, it is possible that evolutionary innovations may affect other aspects of diversification without impacting speciation rates (Givnish 2015; Rabosky 2017). One of the premier examples of a putative key innovation is a second set of jaws in fishes modified to be more flexible (pharyngognathy), long thought to underlie the spectacular radiation of cichlids (Liem, 1973) and other fishes. The effect of this potential innovation, though long assumed in the literature, has never actually been tested. Here we test the effects of this classic evolutionary innovation on speciation and trophic ecological patterns in percomorph fishes. Our results suggest there is no meaningful difference in state-dependent diversification related to pharyngognathy and we do not recover significant differences in rates of trophic evolution between pharyngognaths and non-pharyngognaths. This evidence suggests that pharyngognathy as originally posited in fact does not affect the net diversification or trophic evolution of the fishes which possess it and we caution future studies to be wary of invoking the evolutionary innovation concept without an explicit test.

Introduction

Pharyngognathy is a unique condition among percomorph fishes in which the pharyngeal jaws, a second set of jaws in the back of the throat, have undergone substantial modification (Liem 1973; Wainwright et al. 2012). Liem (Liem 1973) first proposed pharyngognathy as an evolutionary innovation to explain the diversity of the East African Great Lake cichlid radiations. Liem suggested that the novel jaw structure and unique functional abilities, such as increased bite force and decoupling of the oral and pharyngeal jaws for prey processing and acquisition respectively, provided cichlids (Cichlidae, 1,700 species) access to novel trophic niches and the ability to exploit a diverse array of foods, leading to their explosive diversification. Later, pharyngognathy was used by Liem to explain the diversity of wrasses (Labridae, 660 sp.) (Liem & Sanderson 1986) and has since been suggested to impact the diversification of damselfishes (Pomacentridae, 418 sp.) (Frederich et al. 2014). However, pharyngognathy occurs in an additional three fish families, some of which are either rather depauperate in terms of species

number or ecologically poor: surfperches (Embiotocidae, 23 sp.), flying fish & halfbeaks (Exocoetidae/Hemiramphidae, 120 sp.) and the monotypic Centrogenys. While widely accepted and cited as being an evolutionary innovation due to the proposed entry of new dietary niches enabled by possessing the PJA and associations with speciose adaptive radiations, formal tests of the impacts of pharyngognathy on macroevolution across percomorphs have not been performed.

While pharyngognathous fishes were long thought to be monophyletic and form the clade Labroidei in traditional ichthyology prior to the advent of molecular phylogenetics (Liem & Greenwood 1981; Stiassny & Jensen 1987), molecular phylogenetics has refuted the monophyly of pharyngognathous taxa (Mabuchi et al. 2007), and pharyngognathy has evolved an estimated six times in acanthomorph fishes (Wainwright et al. 2012). The multiple evolutions of pharyngognathy in acanthomorphs also make it an ideal trait to study state dependent diversification as it should be less sensitive to phylogenetic pseudoreplication (Maddison & FitzJohn 2015).

Here, we test for the first time, the effects of pharyngognathy on macroevolutionary patterns in percomorph fishes. Specifically, we test how it affects net diversification rates, the main question for a “key innovation,” but also how it affects rates of dietary evolution, specifically trophic level, diet breadth evolution, and transitions to piscivorous and processing intensive diets. Pharyngognathy has two different predicted effects. These aspects of trophic ecology are of interest as pharyngeal jaws likely have consequences in the trophic niches pharyngognathous species can occupy. Pharyngeal jaws lead to an overall reduced gape and limit the size of prey which can be consumed (Wainwright et al. 2012; McGee et al. 2015), suggesting that pharyngognaths will have difficulty evolving to eat large prey at a high trophic level, such as fish (McGee et al., 2015). On the other hand, the decoupling of oral and pharyngeal jaws can more easily result in evolution of crushing or other processing-intensive feeding, which may allow evolution of specialization (McGee et al. 2015). Our results show that contrary to expectations, this purported evolutionary innovation does not increase diversification rate nor lead to faster rates of dietary evolution as originally predicted. We would expect to detect state dependent diversification and differences in the evolution of trophic ecology if pharyngognathy is acting as an evolutionary innovation.

We tested the effects of pharyngognathy on diversification by generating a time calibrated megaphylogeny of 13,025 percomorphs from 33 genes and 119 fossil constraint (Supplemental File S1). We limited our analysis to the clade containing the most recent common ancestor of all pharyngognathous fishes, which contained 11,184 taxa. We then used hidden state speciation extinction models, HiSSE (Beaulieu & O'Meara 2016), to test for state dependent diversification by generating a set of 32 diversification models. This method is an extension of classic speciation extinction models (Maddison et al. 2007) that incorporates hidden states that can account for rate variation and test for character independent diversification. To investigate the effects of pharyngognathy on diet evolution, we generated a database of quantitative diet data (weight, volumetric, or composite indices involving these metrics) from over 1800 studies on 2858 species. We calculated trophic levels and diet breadths and tested the effects of pharyngognathy on evolution of these trophic traits using geomorph (Adams & Otárola-Castillo

2013). We also used this dataset to estimate transition rates between correlated dietary specializations. Species were categorized as piscivorous if they consumed at least 25% non-larval fish in their diet and species as processing intensive specialists if they consumed 25% or more prey that required substantial processing such as plants, algae, hard shelled mollusks, and echinoderms. We then used corHMM (Beaulieu et al. 2013) to estimate transition rates between these two specialist categories respectively in relation to the presence of pharyngognathy.

Methods

Phylogenetic Reconstruction

We generated a 28,907 bp alignment from seventeen nuclear and sixteen mitochondrial loci for 9,876 species of percomorphs and five outgroup taxa using the Phylogeny Assembly with Databases pipeline (PHLAWD)(Smith *et al.* 2009) (Supplemental File S1 in Chapter 4 Supplemental Material). Additional mitochondrial sequences were extracted from whole mitochondrial genomes using the R package *AnnotationBustR* (Borstein & O'Meara 2018) to supplement sequences obtained using PHLAWD. We inferred a maximum likelihood topology using RAxML 8.2.8 (Stamatakis 2014) under the GTR+CAT approximation of rate heterogeneity partitioned by gene and codon position with the final optimization of the tree and branch lengths scored under GTR+ Γ . We employed topological constraints based on recovered phylogenetic relationships from high-throughput sequencing studies of percomorphs fishes (Jones *et al.* 2013; Eytan *et al.* 2015; Longo & Bernardi 2015; Takahashi & Moreno 2015; Diaz-Arce *et al.* 2016; Harrington *et al.* 2016; McGee *et al.* 2016; Rican *et al.* 2016; Burress *et al.* 2017; Dornburg *et al.* 2017; Ilves *et al.* 2017; Longo *et al.* 2017; Song *et al.* 2017; Alfaro *et al.* 2018; Rodgers *et al.* 2018). Taxa with rogue placement were identified by eye and removed. We converted our phylogeny into a chronogram using treePL with 119 fossil and geological constraints with the optimal smoothing parameter obtained using cross-validation (Smith & O'Meara 2012).

To increase the number of taxa within families represented in the phylogeny above we incorporated meristic data obtained using the R interface to FishBase (Froese & Pauly 2019), rfishbase (Boettiger *et al.* 2012), for anal fin spines, anal fin soft rays, dorsal fin spines and dorsal fin soft rays for 8,703 species, for a final 29,562 site alignment for a total of 13,025 acanthomorph species. Specifically, we obtained the ranges for these traits, which were then converted into a series of binary characters. For example, a species with a dorsal spine count of minimum 5, maximum 7, would have 1s for value 5, value 6, and value 7, but 0s for all other dorsal spine values. We constrained the tree search by imposing constraints on family placement obtained from the molecular tree described above. We used MonoPhy (Schwery & O'Meara 2016) to identify monophyletic genera from the molecular tree, which were then constrained in the tree search that included meristic characters. In cases where species with only meristic data were members of a genus in the molecular tree that were recovered as monophyletic, we included in a monophyletic constraint for that genus. Otherwise, additional meristic taxa were constrained at the family level. Tree search was performed as described above with the addition of binary partitions for each meristic trait. In order to convert the tree constructed from

nucleotide and meristic characters into a chronogram, we used the *congruify.phylo* in the R package *geiger* to impose secondary calibrations using the chronogram described above as a reference (Eastman *et al.* 2013; Pennell *et al.* 2014). We chose a congruification method to limit the potential impact of species with only meristic data from impacting divergence time estimates.

Diversification Analyses

To test for state-dependent diversification, we used the R package *HiSSE* (hidden state speciation and extinction), which employs a hidden markov model that assumes a hidden unobserved state to measure rates of speciation, extinction, and transition rates among states, while also testing for character-independent diversification (Beaulieu & O'Meara 2016). We focused our analysis to the most recent common ancestor of all pharyngognath taxa and ran a set of twenty-four models on both the nucleotide phylogeny and the nucleotide+meristic phylogeny, which left us with a phylogeny of 11,184 species respectively. As transition rates between states can be difficult to estimate in an SSE framework, we constrained transitions to be equal across states (Beaulieu & O'Meara 2016). We mapped pharyngognathy to the six pharyngognathous clades, Centrogeniidae, Cichlidae, Embiotocidae, Exocoetidae/Hemiramphidae, Labridae, and Pomacentridae, following Wainwright *et al.* (2012). We accounted for missing species in our dataset by supplying *HiSSE* with the known sampling fraction of taxa in our phylogeny for each state, which we calculated based on the number of known described species in pharyngognath and non-pharyngognath families listed in the California Academy of Sciences Catalog of Fishes (Eschmeyer *et al.* 2018). Our phylogeny contained 8,418 non-pharyngognaths of 11,337 described non-pharyngognaths and 2,766 pharyngognaths of 2,944 described pharyngognaths representing 65.4% and 93.95% taxon sampling respectively.

We also tested for significant differences in diversification rates between pharyngognaths and non-pharyngognaths using the non-parametric method *FiSSE* (Rabosky & Goldberg 2017). This method measures the mean diversification rate of Jetz *et al.* (2012) for each character state and uses the difference between these as the test statistic. It then computes a null distribution to test the test statistic against by simulating character evolution on the empirical phylogeny under a one-parameter Markov model and assesses significance by comparing the proportion of simulated datasets that are more extreme than the observed test statistic. We ran *FiSSE* for each tree with 1,000 simulations using the *FiSSE.binary* function. In addition to *FiSSE*, we also performed two non-parametric tests comparing the number of taxa in pairs of sister clades, the diversity contrast test (Paradis 2012) and the sign test.

The *HiSSE* and *FiSSE* diversification tests were also performed on the clade Ovalentaria, which contains four of the six transitions to pharyngognathy (Cichlidae, Embiotocidae, Exocoetidae/Hemiramphidae, and Pomacentridae). We did not perform additional tests between Labrids and all other Eupercarian groups due to known problems of singular transitions in diversification methods (Maddison & FitzJohn 2015) and due to the only other transition within Eupercaria, Centrogeniidae, being monotypic.

Dietary Data Acquisition and Rates of Trophic Level and Diet Breadth Evolution

We assembled a database of quantitative diet data from over 1815 published sources for 2858 species of fishes. To limit possible ontogenetic biases, we limited our dataset studies that included adults within their sampling. Dietary studies which were performed outside of a species native range were also excluded to avoid possible shifts in trophic ecology due to invading non-native habitats. We calculated trophic levels using the R package *diatr* (Borstein 2019), which follow the routines implemented by Pauly et al., (Pauly *et al.* 2000). In cases where species had data from multiple studies, trophic levels were calculated for each study and the mean trophic level for that species was used in downstream analyses.

Diet breadth was calculated using the R package *ordiBreadth* (Fordyce *et al.* 2016), which calculates ordinated diet breadth by subjecting a dissimilarity matrix of dietary items to principal coordinates analysis (PCoA) and after ordination, measures ordinated diet breadth as the sum of the distances of a given consumer's multivariate centroid to its consumed prey items. To account for differences in available prey between major habitat types (i.e. no freshwater corals, echinoderms, etc.), we calculated diet breadth for species based on their occurrence in either freshwater, marine water, or brackish waters from data extracted from *rfishbase* (Froese & Pauly 2019). For easier interpretability and to make diet breadth spaces comparable across habitat types, we report the scaled diet breadth, which is the scaled measure of occupation a species has in relative diet space in its respective habitat. This value is bounded between 0 and 1, representing specialist species with diets consisting solely of a single prey type to a hypothetical generalist consumer feeding on all prey items respectively. For consistency among food records, we used the "FoodIII" column in our diet database to generate a binary matrix of 71 prey items. We combined some prey types due to inconsistencies among published studies in how they were reported (i.e. non-phytoplankton algae types, planktonic copepods vs cladocerans, etc.) or known difficulties in identification due to their known habit of being rapidly digested (i.e. gelatinous zooplankton types like jellyfish, siphonophores, etc.) (Supplemental Table S1 & S2 in Chapter 4 Supplemental Material). We removed prey items that comprised less than five percent of the average diet of a species to remove prey items that were likely incidentally consumed or play a minor role in the overall diet of the species.

We used the R package *geomorph* (Adams & Otárola-Castillo 2013) to measure the rate of trophic level and diet breadth evolution respectively using the *compare.evol.rates* function under 10000 simulations to assess differences in rates of evolution between pharyngognaths and non-pharyngognaths.

Transitions to Piscivorous and Processing Intensive Diets

We used the R package *corHMM* (Beaulieu *et al.* 2013) to measure correlated transition rates between pharyngognathy and piscivorous diets and pharyngognathy and processing intensive diets and pharyngognathy. Species were scored as being piscivorous if their diets included on average at least 25% non-larval fishes in their diets. Processing intensive diets were scored in two ways. The first, excluded macrocrustaceans due to the high variability in the

hardness of their shells and fish possibly consuming them during molting when they are soft. For the first way, species were defined as having a processing intensive diets if they consumed 25% or more hard shelled mollusks, brachiopods, diatoms and other autotrophs excluding phytoplankton, foraminiferans, bryozoans, and echinoderms (excluding crinoids and sea cucumbers), and detritus. For the second cutoff, we included species with an average consumption of 25% or more of the above items in addition to non-larval macrocrustaceans like crabs, lobsters, anomurans, and shrimps. To measure transition rates of correlated evolution between pharyngognathy and piscivory or processing intensive diets, we used the corHMM function corDISC. We specifically fit three different models. One in which all rates are equal, a model in which transition rates where symmetrical, and a third in which all transition rates where free. In addition to analyses of correlated transition rates, we also estimated transition rates between trophic states using the rayDISC function using the same three models described above.

Results & Discussion

Contrary to the expectation that pharyngognathy leads to increased diversification rates, we find strong support for a character-independent model of diversification ($AIC_w = 0.999$) (Figures 4.1, Supplemental Table S3-S7 in Chapter 4 Supplemental Material). This suggests other unknown factors explain the diversification rates of pharyngognathous taxa than pharyngognathy itself. Analysis of the same datasets using a non-parametric phylogenetic method FiSSE (Rabosky & Goldberg 2017) and two sister group methods comparing numbers of taxa between pharyngognathous and non-pharyngognath clades (Paradis 2012) also fail to detect significant differences in diversification ($p > 0.47$): only in four of six cases do pharyngognathous clades have more species than their sisters. Interestingly, we find the prime example used for the rapid evolution of pharyngognathous taxa, East African cichlids, have relatively slow rates of net diversification but very high turnover rates: speciation and extinction rates are both overall very high (Figure 4.2).

We also fail to find significant differences in rates of trophic level evolution and diet breadth evolution between pharyngognaths and non-pharyngognaths ($p = 0.98$ for both) with near identical rates of evolution regardless of state, with a rate ratio of 1.01 and 1.03 for diet and trophic level, respectively (Figure 4.3). Analysis of transitions rates between diet specializations revealed that although pharyngognaths had a higher transition rate to piscivorous diets, they also had a substantially high transition rate out of piscivory compared to non-pharyngognaths (Figure 4.4). While this result is different than previous analyses performed by McGee et al. (2015), it is important to note that their analysis was restricted to reef associated taxa, while our analysis contains freshwater and non-reef fishes. For processing intensive diets, pharyngognaths had higher transition rates into processing intensive categories while non-pharyngognaths had higher rates of transitions out of processing intensive diets, in agreement with previous studies (McGee et al. 2015) (Figure 4.4).

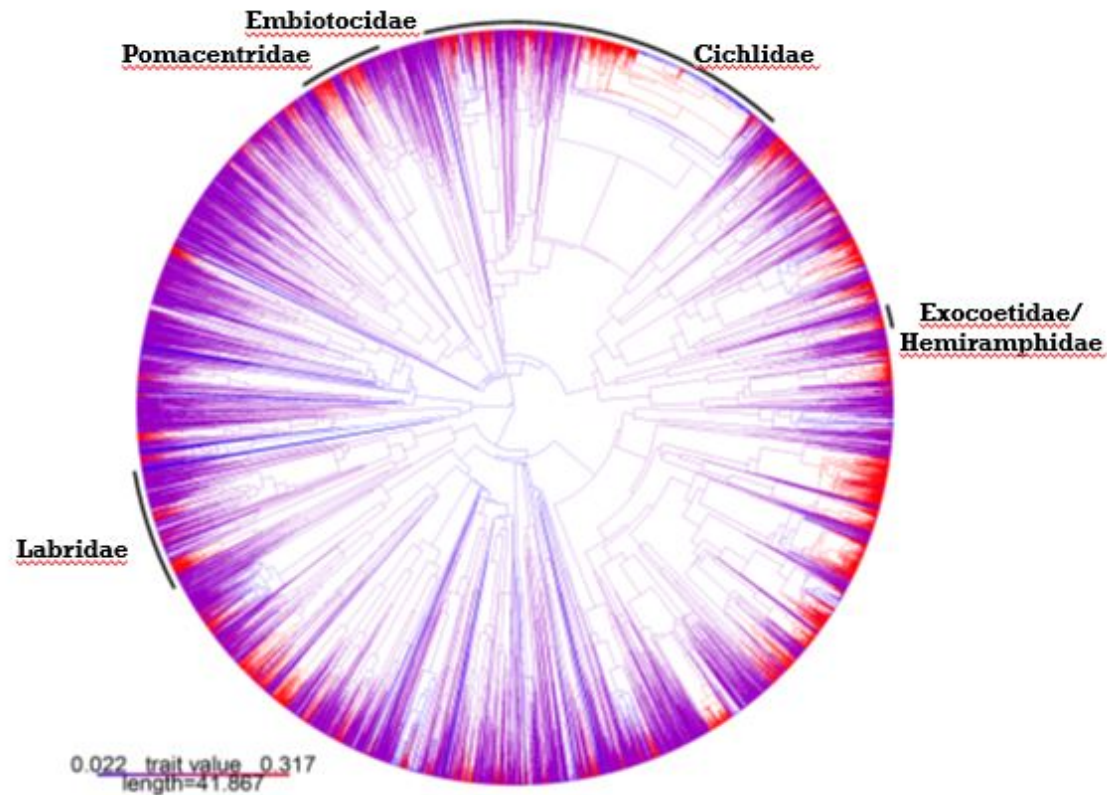


Figure 4.1: Net Diversification rates (speciation-extinction) across a phylogeny of 1,184 percomorph fishes. Pharyngognathous caldes (except the monotypic *Centrogenys*) are highlighted by arched black clade labels. Warmer colors represent faster rates of net diversification while cooler colors represent slower rates. Net diversification rates are rather heterogeneous throughout the tree. Interestingly, a prime example of pharyngognath as an evolutionary innovation, East African cichlids, are found to have low net diversification rates. This is caused by overall high turnover rates (Figure 2), where both rates of speciation and extinction are high.

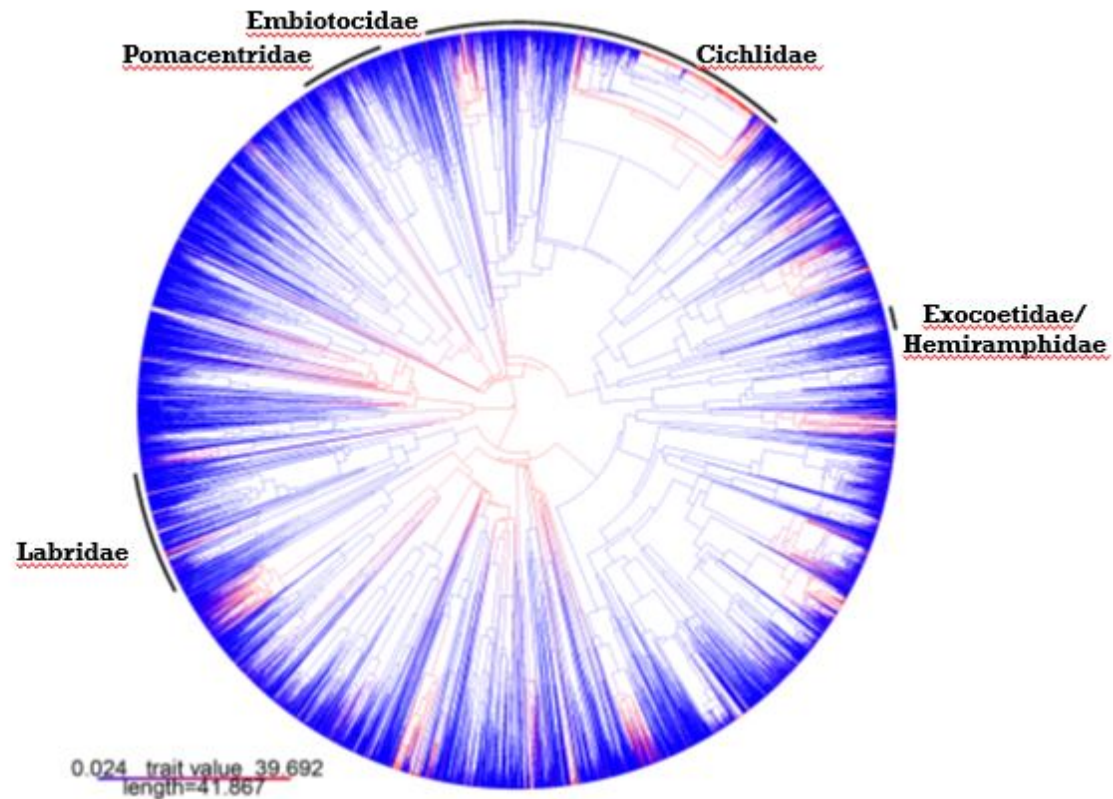


Figure 4.2: Turnover (speciation+extinction) across a phylogeny of 1,184 percomorph fishes. Pharyngognathous caldes (except the monotypic Centrogenys) are highlighted by arched black clade labels. Warmer colors represent faster turnover rates while cooler colors represent slower turnover rates. Turnover rates are rather heterogeneous throughout the tree. East African cichlids, have high turnover rates, reducing their overall net diversification rates (Figure 4.1).

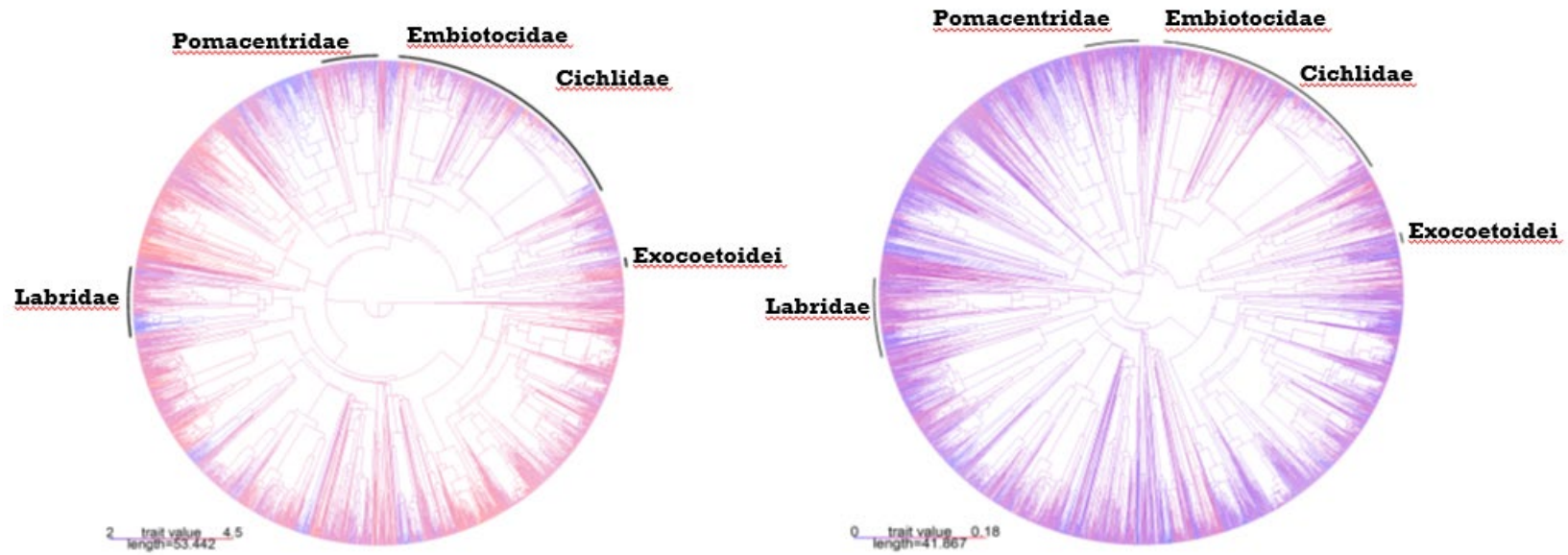


Figure 4.3: Continuous character maps of trophic level (left) and diet breadth (right) across 2859 percomorph fishes. Lower trophic levels and diet breadths are represented by cooler colors while warmer colors represent higher trophic levels and diet breadths respectively. The position of pharyngognathous clades in the phylogeny are represented by arched black clade labels.

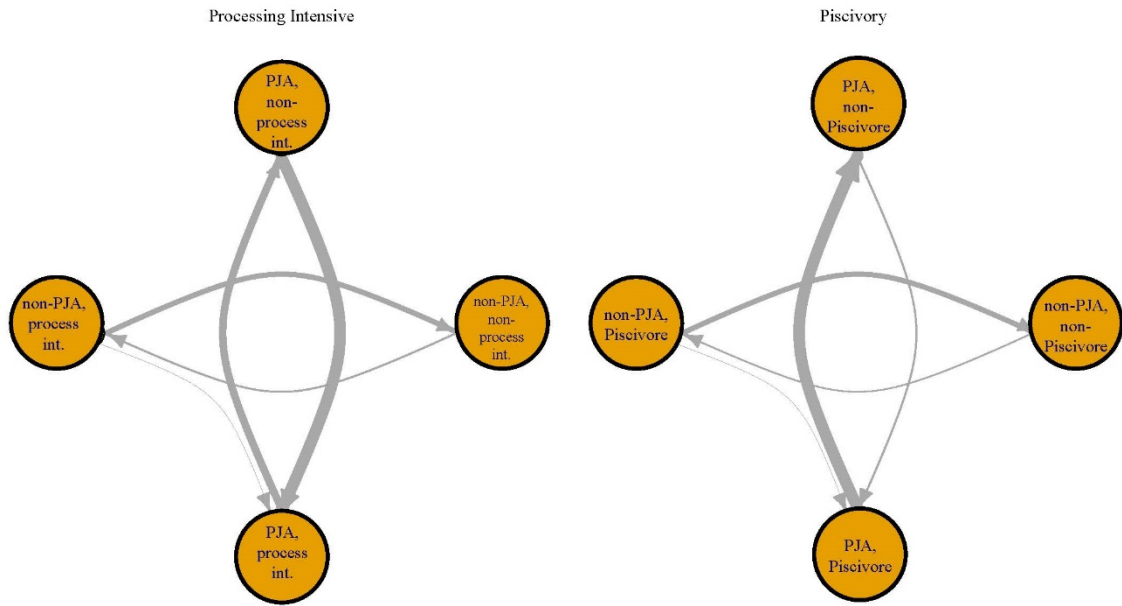


Figure 4.4: Transition network map of processing intensive diets (left) and piscivorous diets (right). PJA non-PJA represents states of pharyngognathy and non-pharyngognathy respectively. Underneath these states is the presence or absence of the state of dietary specialization. Width of arrows represents the relative transition rate, with wider arrows representing faster rates. We recover strong transitions in pharyngognaths into processing intensive prey as well as a fast transition rate out of piscivorous diets.

These results are contrary to the expectation that pharyngognathy has a major impact on diversification in percomorphs. In addition to failing to detect state dependent diversification, we also find limited evidence that pharyngognathy positively impacts trophic evolution. Pharyngognathous taxa tend to evolve processing intensive diets more often than non-pharyngognaths do; they also tend not to evolve piscivory as often. Together, this yields almost no difference in rates of trophic level evolution and diet breadth evolution.

While we have tried to control for sources of error, there are some important caveats to note. One is the difficulty of getting good estimates of extinction rates using modern data only (Rabosky 2016). Another is the assumption of a tree: it is known that hybridization makes it more of a network, especially in groups like East African lake cichlids (Meier *et al.* 2017). Species over- or undersplitting may also affect results.

Our results highlight the need to assess proposed evolutionary innovations. Pharyngognathy has been widely used for 45 years as an example of an evolutionary innovation, yet until now it has never been tested. We encourage the use of modern phylogenetic methods in testing purported evolutionary innovations that have not been adequately assessed, yet we also feel it is necessary to discuss a needed paradigm shift away from evolutionary innovations, especially the idea of key-innovations that are of common focus in biological studies. While the evolutionary innovation concept is simple in that it focuses on a single trait being of key importance to the diversity of a clade that possesses it, reality is more complex. In this study, we found no evidence of an effect of a hypothesized and long-assumed key innovation, but the methods used could also have found that it had only a partial effect, and other, unmeasured, factors could also have played a role. Moving to examine assumptions and look at multiple factors will help shift away from a focus of single key-trait towards a focus on including a wider set of possible biotic and abiotic factor that may be responsible for observed diversity

References

- Adams, D.C. & Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393-399.
- Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E. *et al.* (2018). Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nat Ecol Evol*, 2, 688-696.
- Beaulieu, J.M. & O'Meara, B.C. (2016). Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic biology*, 65, 583-601.
- Beaulieu, J.M., O'Meara, B.C. & Donoghue, M.J. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic biology*, 62, 725-737.
- Boettiger, C., Lang, D.T. & Wainwright, P.C. (2012). rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81, 2030-2039.
- Borstein, S.R. (2019). dietr: Diet Estimated Trophic Levels in R. 1.0.
<https://github.com/sborstein/dietr>
- Borstein, S.R. & O'Meara, B.C. (2018). AnnotationBustR: an R package to extract subsequences from GenBank annotations. *PeerJ*, 6, e5179.
- Burress, E.D., Alda, F., Duarte, A., Loureiro, M., Armbruster, J.W. & Chakrabarty, P. (2017). Phylogenomics of pike cichlids (Cichlidae: *Crenicichla*): the rapid evolution and trophic diversification of an incipient species flock. *Journal of evolutionary biology*.
- Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X. & Rodríguez-Ezpeleta, N. (2016). RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Molecular phylogenetics and evolution*, 102, 202-207.
- Dornburg, A., Townsend, J.P., Brooks, W., Spriggs, E., Eytan, R.I., Moore, J.A. *et al.* (2017). New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. *Molecular phylogenetics and evolution*, 110, 27-38.
- Eastman, J.M., Harmon, L.J., Tank, D.C. & Paradis, E. (2013). Congruification: support for time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4, 688-691.
- Eschmeyer, W., Fricke, R. & Van der Laan, R. (2018). Catalog of fishes: genera, species, references. In: *Electronic version accessed* (ed. Sciences, CAo). California Academy of Sciences.
- Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Lemmon, E.M., Wainwright, P.C. *et al.* (2015). Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. *BMC evolutionary biology*, 15, 113.
- Fordyce, J.A., Nice, C.C., Hamm, C.A. & Forister, M.L. (2016). Quantifying diet breadth through ordination of host association. *Ecology*, 97, 842-849.
- Frederich, B., Olivier, D., Litsios, G., Alfaro, M.E. & Parmentier, E. (2014). Trait decoupling promotes evolutionary diversification of the trophic and acoustic system of damselfishes. *P Roy Soc B-Biol Sci*, 281.
- Froese, R. & Pauly, D. (2019). FishBase. Available at: <http://www.fishbase.org/2018>.
- Givnish, T.J. (2015). Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, 207, 297-303.

- Harrington, R.C., Faircloth, B.C., Eytan, R.I., Smith, W.L., Near, T.J., Alfaro, M.E. *et al.* (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC evolutionary biology*, 16, 224.
- Heard, S.B. & Hauser, D.L. (1995). Key evolutionary innovations and their ecological mechanisms. *Historical Biology*, 10, 151-173.
- Hunter, J.P. (1998). Key innovations and the ecology of macroevolution. *Trends Ecol Evol*, 13, 31-36.
- Ilves, K.L., Torti, D. & Lopez-Fernandez, H. (2017). Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichliformes: Cichlidae: Cichlinae). *Molecular phylogenetics and evolution*, 118, 232-243.
- Jetz, W., Thomas, G., Joy, J., Hartmann, K. & Mooers, A. (2012). The global diversity of birds in space and time. *Nature*, 491, 444.
- Jones, J.C., Fan, S., Franchini, P., Scharl, M. & Meyer, A. (2013). The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, 22, 2986-3001.
- Liem, K.F. (1973). Evolutionary strategies and morphological innovations- cichlid pharyngeal jaws. *Syst Zool*, 22, 425-441.
- Liem, K.F. & Greenwood, P.H. (1981). A functional approach to the phylogeny of the pharyngognath teleosts. *Am Zool*, 21, 83-101.
- Liem, K.F. & Sanderson, S.L. (1986). The pharyngeal jaw apparatus of labrid fishes: A functional morphological perspective. *Journal of morphology*, 187, 143-158.
- Longo, G. & Bernardi, G. (2015). The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular phylogenetics and evolution*, 88, 55-63.
- Longo, S.J., Faircloth, B.C., Meyer, A., Westneat, M.W., Alfaro, M.E. & Wainwright, P.C. (2017). Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. *Molecular phylogenetics and evolution*.
- Mabuchi, K., Miya, M., Azuma, Y. & Nishida, M. (2007). Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC evolutionary biology*, 7.
- Maddison, W.P. & FitzJohn, R.G. (2015). The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic biology*, 64, 127-136.
- Maddison, W.P., Midford, P.E. & Otto, S.P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic biology*, 56, 701-710.
- Mayr, E. (1963). *Animal species and evolution*. Belknap Press of Harvard University Press.
- McGee, M.D., Borstein, S.R., Neches, R.Y., Buescher, H.H., Seehausen, O. & Wainwright, P.C. (2015). A pharyngeal jaw evolutionary innovation facilitated extinction in Lake Victoria cichlids. *Science*, 350, 1077-1079.
- McGee, M.D., Faircloth, B.C., Borstein, S.R., Zheng, J., Darrin Hulsey, C., Wainwright, P.C. *et al.* (2016). Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. *Proceedings of the Royal Society B: Biological Sciences*, 283.

- Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8, 14363.
- Paradis, E. (2012). Shift in diversification in sister-clade comparisons: a more powerful test. *Evolution: International Journal of Organic Evolution*, 66, 288-295.
- Pauly, D., Froese, R., Sa-a, P., Palomares, M., Christensen, V. & Rius, J. (2000). TrophLab manual. *ICLARM, Manila, Philippines*.
- Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G. *et al.* (2014). geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, btu181.
- Rabosky, D.L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS one*, 9, e89543.
- Rabosky, D.L. (2016). Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O'Meara. *Evolution; international journal of organic evolution*, 70, 218-228.
- Rabosky, D.L. (2017). Phylogenetic tests for evolutionary innovation: the problematic link between key innovations and exceptional diversification. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372.
- Rabosky, D.L. & Goldberg, E.E. (2017). FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution; international journal of organic evolution*.
- Rican, O., Pialek, L., Dragova, K. & Novak, J. (2016). Diversity and evolution of the Middle American cichlid fishes (Teleostei: Cichlidae) with revised classification. *Vertebr Zool*, 66, 1-102.
- Rodgers, R., Roach, J., Reid, N., Whitehead, A. & Duvernell, D.D. (2018). Phylogenomic analysis of Fundulidae (Teleostei: Cyprinodontiformes) using RNA-sequencing Data. *Molecular phylogenetics and evolution*, 121, 150-157.
- Schwery, O. & O'Meara, B.C. (2016). MonoPhy: a simple R package to find and visualize monophyly issues. *PeerJ Computer Science*, 2, e56.
- Smith, S.A., Beaulieu, J.M. & Donoghue, M.J. (2009). Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC evolutionary biology*, 9, 37.
- Smith, S.A. & O'Meara, B.C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28, 2689-2690.
- Song, S., Zhao, J. & Li, C. (2017). Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. *Molecular phylogenetics and evolution*, 111, 44-55.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.
- Stiassny, M.L. & Jensen, J.S. (1987). Labroid interrelationships revisited: morphological complexity, key innovations, and the study of comparative diversity. *Bulletin of the Museum of Comparative Zoology*, 151, 269-319.
- Takahashi, T. & Moreno, E. (2015). A RAD-based phylogenetics for *Orestias* fishes from Lake Titicaca. *Molecular phylogenetics and evolution*, 93, 307-317.

Wainwright, P.C., Smith, W.L., Price, S.A., Tang, K.L., Sparks, J.S., Ferry, L.A. *et al.* (2012). The Evolution of Pharyngognathy: A Phylogenetic and Functional Appraisal of the Pharyngeal Jaw Key Innovation in Labroid Fishes and Beyond. *Systematic biology*, 61, 1001-1027.

CONCLUSION

Trophic ecology is of utmost importance to study in general ecology and evolution. As all organisms must procure resources for survival, the selective pressures of successful trophic acquisition have profound effects on shaping biodiversity of macroevolutionary timescales. To elucidate these patterns, my research used fishes as system to study how trophic specialization, both in terms of resource use and traits thought allegedly enhance trophic specialization diversification. I was able to test various hypotheses related to trophic acquisition in a phylogenetic context to identify how it effects macroevolution. My research highlights the utility of using ecological data in conjunction with modern phylogenetic comparative methods to elucidate possible macroevolutionary patterns.

My dissertation has yielded two key results. First, I was able to show that trophic resource specialization does not inhibit phenotypic evolution contrary to previous hypotheses in Chapter 3. Historically, it is thought that populations and species exploiting diverse resources, generalists, displayed more within group phenotypic variation. When scaled to trophic guilds, I have shown the opposite pattern in coral reef fishes, where highly specialized species have faster rates of morphological evolution and display more functional phenotypic diversity. This result is also counter to previous research on smaller clades, in which it has been suggested that highly specialized piscivores have reduced rates of morphological evolution. My results highlight the possible effect of scale and multiple transitions in phylogenetic analyses and further research should investigate these in regard to their general effects on macroevolutionary studies.

In Chapter 4, I was able to show an evolutionary innovation, pharyngognath, which is thought to promote trophic novelty and diversification rates in clades which possess it, does not actually increase diversification nor yield major benefits in occupying novel trophic niches. While pharyngognath has been an oft cited example of an evolutionary innovation, its actual consequences on macroevolution had not been studied in a phylogenetic comparative framework. This result highlights the need to test purported claims in evolutionary biology, but more important, highlight the dangers of focusing on evolutionary innovations, specifically the concept of key innovations, in biology. While evolutionary innovations provide a possible simple mechanistic reason for biological patterns observed, they are also rather unrealistic as we know biology is much more complicated with multiple factors playing pivotal roles in shaping a clades diversity. I hope that my study highlights these issues and a paradigm shift away from simplistic key innovation thinking and a focus on a single factor towards a focus on analyzing a broader set of biotic and abiotic variables that may be responsible for shaping diversity.

VITA

Samuel R. Borstein was born in Illinois. He was raised in Buffalo Grove, Illinois and graduated from Adlai E. Stevenson High School in 2009. Sam attended California State University, Sacramento from 2009 until 2013 when he graduated with a Bachelor of Science in Biology with a concentration in Ecology, Evolution, and Conservation. During this time, he was a curatorial assistant for the California State University, Sacramento fish collection under the guidance of Ron Coleman. He also performed undergraduate research in the Lab of Peter Wainwright at the University of California, Davis. He accepted a graduate position at the University of Tennessee in 2013 to pursue his doctorate under his advisor Brian O'Meara. Sam has received funding through multiple external source for his research, including a National Science Foundation Doctoral Dissertation Improvement Grant and American Museum of Natural History Theodore Roosevelt Memorial Grant. He has authored or co-authored five papers in peer-reviewed journals including *Nature Ecology & Evolution*, *Science*, *Proceedings of the Royal Society B: Biological Science*. Sam has been a reviewer for six journals, including *Molecular Biology & Evolution* and *Methods in Ecology and Evolution*. He has also served as the graduate representative for the University of Tennessee's Biological Collections Committee.