

DETERMINING THE EFFECT OF VARIABILITY IN HABITAT QUALITY ON
DISPERSAL

A Dissertation

by

JASON D SELWYN

BS, Northeastern University, 2011
MS, Texas A&M University – Corpus Christi, 2015

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

MARINE BIOLOGY

Texas A&M University-Corpus Christi
Corpus Christi, Texas

May 2022

© Jason D. Selwyn

All Rights Reserved

May 2022

DETERMINING THE EFFECT OF VARIABILITY IN HABITAT QUALITY ON
DISPERSAL

A Dissertation

by

JASON D SELWYN

This dissertation meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Christopher E. Bird, Ph.D.
Chair

J. Derek Hogan, Ph.D.
Co-Chair

David S. Portnoy, Ph.D.
Committee Member

Blair Sterba-Boatwright, Ph.D.
Committee Member

Ryan O'Malley, M.F.A.
Graduate Faculty Representative

May 2022

ABSTRACT

The dispersal of individuals between populations is a foundational process to understand at the interface of ecology and evolution. The natal habitat is theorized to strongly influence the degree of dispersal expected. However, understanding the interaction between habitat and dispersal is difficult to study empirically, particularly in a single location where other environmental factors are held constant. Understanding how habitats influence dispersal is important not only for the foundational understanding of ecological and evolutionary processes but also as they relate to the design of marine protected area networks. Here I seek to understand how heterogeneity in habitat quality influences the dispersal dynamics of the common Caribbean reef goby *Coryphopterus hyalinus* as a model for other species with similar life histories in different systems. To determine how variation in habitat quality influences dispersal first I had to establish what topographical features of the reef equate to greater habitat quality from the perspective of the previously presumed habitat generalist *C. hyalinus*. I found that as adults *C. hyalinus* live in mixed species shoals with their congener *C. personatus* and are distributed across shallow coral reef ecosystems tending to be found in greater densities in more complex, deeper reef areas at the margin of large sand patches. In Turneffe Atoll, *C. hyalinus* has an average dispersal distance of 3.1 ± 0.3 km with 95% of individuals dispersing less than 7.7 ± 0.65 km. However, spatially heterogeneous habitats are characterized by shorter mean dispersal distances, smaller dispersal spreads, and higher propensity for long-distance dispersal events. This observation likely has strong conservation implications for the design and futureproofing of network-based conservation designs which depend upon dispersal between individual nodes of the network for proper functioning. As anthropogenic climate change alters habitats and in the short-term leads

to increasingly fragmented and heterogeneous landscapes these networks may no longer be sustainable given the shrinking of the dispersal spread of the species these networks are designed to protect.

DEDICATION

My parents

ACKNOWLEDGEMENTS

Sitting on Padre Island National Seashore writing these acknowledgements it's hard to begin to think about how many people have helped me in countless ways over the years to get to the point of finishing this dissertation. Here is my best effort, but I'm sure I will forget to mention important people along the way. For that I must first apologize and hope you believe my assurance that I am grateful for how you've helped me, despite my momentary lapse of memory.

About a decade ago I was living in New Zealand sitting on my bed in the basement of the flat I was sharing in Kohimarama, I came across an ad in the coral-list listserv which, unbeknownst to me at the time, would fundamentally change my life. In that ad Dr. Derek Hogan was looking for graduate students to come join his newly established lab to study dispersal and the effect of lionfish on native fishes in Belize. I, knowing nothing about dispersal but a thing or two about lionfish, decided I'd send in an application. Back then applying for a master's degree, passionate about wanting to understand more about the scourge of lionfish on native species and wanting to continue SCUBA diving as a feature of my career. Luckily for me Derek decided whatever I had written in my application was interesting enough and he wanted to interview me. I felt like we hit it off in that first interview despite it being the middle of the night, such are the costs of being on the other side of the world, and apparently Derek felt the same as he offered me a position at Texas A&M University – Corpus Christi, a place I'd never heard of, much less considered living in, as his master's student. So, I quit my job and booked a flight home where I needed to buy a car (my trusty Blueberry), pack everything I owned into it, and drive down to Texas which I knew nothing about. When I made it to Corpus, I moved into the first place I found that I could move into that day, the next day I meet Alan Downey-Wall, my fellow master's student, and the

day after that I met Derek and Sharon Magnuson. The following day Alan, Derek and I flew down to Belize, where I meet Dr. Paolo Usseglio and went to Turneffe Atoll for my first month of field work as a graduate student. And so began my time as a graduate student.

After getting back from the field I found that the Gulf and Caribbean Fisheries Institute annual international meeting was going to be held in a few months in Corpus Christi, and we had data that could be presented at the special symposium about lionfish. Derek gave me the chance to analyze that data, sending me off to try and figure out how to tell if the community of reef fishes was different now than it was when he had done the same surveys during his PhD about a decade prior, thus beginning our relationship with me as Derek's "free-range grad student". I am eternally grateful for Derek treating me this way as it allowed me to explore my curiosity about the subject in whatever direction that led, always being able to come back and meet with him if I got too lost or had some exciting new finding.

At the same time despite only having been his student for a couple of weeks and still very much struggling to learn R I met with Dr. Blair Sterba-Boatwright to ask for advice about how best to interpret the analysis I had come up with, particularly how to understand which species were driving the changes we'd observed. Learning from Blair in that meeting and then through all the various classes I was able to take with him as both a masters and PhD student and in the many informal questions he's answered for me has been invaluable for me growing into the now somewhat "promiscuous user of methodology" I've become. At some point during my time in graduate school Blair referred a request, like the one I had made of him right at the very start, to me. That led to a cool analysis and paper I was able to contribute to, to understand more about humpback whale singing behavior and unlocked for me the potential for statistical consulting as something I truly enjoyed. After that initial project Dean Pezold was generous enough to allow

me to use a portion of my required teaching assistantship hours to dedicate to consulting with various graduate students and faculty on statistical questions and get paid for it. Ultimately, when COVID struck that led to my being invited to join the County/City/University taskforce to use those statistical skills I had learned and been given the opportunity to grow from Blair and the Dean to help the city I lived in, particularly in the beginning when everything about this new plague was so uncertain.

Now dear reader beyond simply taking a chance on hiring me in the first place Derek, along with future committee member Dr. David Portnoy, was also the source of my master's project which subsequently spawned the idea for my dissertation work on gobies. In what I can only describe as the hubris and optimism of youth I had decided initially to replicate a study I had worked on in the past that involved controlling the densities of lionfish on a reef. Unfortunately, those lionfish decided they wanted to move around between my locations that were far too close together on continuous rather than patch reef and completely ruin my experiment. Fortunately, as part of a class project for Dave I, along with Alan and Lauren Gurski, had analyzed some of Derek's old goby data and found a cool pattern that we decided to do some additional collections for, thus saving my master's degree after a quick pivot, and leading to a frankly much more interesting dissertation about dispersal that I hope you're about to read. Beyond being the impetus for what eventually became both my masters and dissertation research Dave has been a wonderful font of knowledge to learn from about ecology/evolution/ichthyology and has always encouraged me to delve deeper into my research to find how it ties into deeper processes of evolution.

Last and certainly not least of my committee without whom my dissertation would have been entirely different, and worse, is Dr. Christopher Bird. Chris has consistently impressed upon me

the need to tell a story with my analyses. This was most strikingly pressed upon me during our time working together on the COVID taskforce where regardless of the quality of the analysis I performed the primary issues were first communicating the story the analysis conveyed to the rest of the taskforce and then to the public. Beyond emphasizing the importance of this basic need for storytelling Chris has helped me to see the value diving deep into the details of the data to understand any strange features rather than merely relying on data summaries, particularly when it comes to understanding why particular SNP loci are removed by various filters and determining if that's the behavior you want to see. Most practically I must thank Chris for supporting my summer salary whenever I wasn't on other funding which allowed me to play in a variety of other people's data and learn a variety of bioinformatic techniques and how to best approach new data to answer the questions of interest.

Over the years so many people have helped me that these acknowledgements nearly got even more ridiculous than they already are. First, I must thank Paolo, Alan, Emily Anderson, BanDula, Nacho, Miss T and all the folks at Calabash Caye Field Station for their help with all aspects of my fieldwork. Particular thanks to Paolo for coming up with the initial idea for us to try using photogrammetry to map the habitat and taking all the pictures as our human underwater drone. Also requiring my heartfelt thanks is Brett Dodson, despite the many arguments we had about whether the AAUS rules as written required an infinite number of safety divers on the boat, you made sure we did everything safely and the only injuries I got from diving were lionfish stings, rest in peace.

While I may wish all my time could have been spent in the field sadly that is not the way of things. Fortunately, I was lucky enough to be a member of the HoBi lab whom I owe for a huge range of growth I've experienced through graduate school. First my deepest thanks must go to

Sharon, the person in Corpus whom I met earliest that's also still here. I can't begin to describe the ways I owe you both professionally, for teaching me how to do all the lab work and answering any questions or crazy lab-based hypotheses I may have had, and personally with all the long conversations at your house after parties and when I was lucky enough to share an office with both you and Dr. Rebecca Hamner. Nicolette Beeken, thank you for coming along to study for your masters and taking on a project about a pair of species no-one on Earth cares about and doing it with the absolute diligence required to uncover a great deal of the natural history of *Coryphopterus hyalinus/personatus* without which I would not have been able to understand the foundations for their life history upon which all the conclusions of this dissertation are built upon. One of the best experiences of my graduate school career is having the opportunity to mentor Ashley Hamilton in her studies and post-undergraduate path. What started out as merely helping make some code more efficient has turned into the thing I'm most proud of as you pursue your own PhD.

My thanks for the many engaging and thought-provoking discussions and conversations I was able to have as an adopted/honorary member of the Marine Genomics Lab. In particular I should thank Lizz Hunt for always being able to answer whatever questions I may have had about the secrets of fish known only to the most dedicated of ichthyologists and always being up for bouncing ideas around over a beer (or wine as it may be). Thanks must go to Dr. Andrew Fields for being a thoughtful and constructively critical presence to bounce both analyses and conclusions around with. I also must thank the members of the Early Life History Lab for showing me how to use the fancy microscopes to take high quality pictures of the head-pores of the gobies to show just how difficult a character it is to use.

Often overlooked and under-thanked I'm sure I wouldn't have managed to get to this point with the help of all the various staff at the university who have helped me. In particular, thanks to Tom Merrick for his help installing the various spaghetti code programs I always wanted to try out on the HPC. Thanks to Sue Burgess and Patricia Rodriguez for making sure all the money went where it needed to go with all the paperwork it needed. Ronnie Emmanuel, I'm finally graduating, thank you for making sure I had all the forms and paperwork in on time and my coursework was all in order.

I'd also like to thank my funding sources the Carl R. Beaver Memorial Scholarship, the Karen Koester Dodson Memorial Fund Grant and the Texas Sea Grant for taking a chance on my research. Without their support none of this would have been possible.

Finally, I would be lying if I said I got here without the support and help of my friends and family, many of whom are mentioned above and helped directly. I would argue more importantly than the direct help is the indirect support that has kept me sane over the years whether that be going to Recess on Thursday (or really most) nights to long nights at Matt and Sharon's house to the D&D crews to COVID zoom hangouts and everything in between. Without those times in between the working none of the work would have been possible or worthwhile and while things will always change, I deeply value those times together and look forward to the ways we'll continue our friendships onwards. Dr. Lynette Strickland, I'm so thankful for having met you when I did you've been an amazing emotional anchor and source of encouragement as well as being someone I can ask to read my writing or talk to and dissect the scientific ideas within my research with curiosity and interest. And of course, I must thank my parents, who never stopped asking the forbidden question of PhD students, "when are you going to finish?". Your support over my entire life has led to me being able to get here.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iv
DEDICATION.....	vi
ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS.....	xiii
LIST OF FIGURES.....	xix
LIST OF TABLES.....	xxii
CHAPTER I: INTRODUCTION.....	1
CHAPTER II: Maintenance of species boundaries within social aggregations of ecologically similar goby sister species.....	6
Abstract.....	6
Introduction.....	7
Materials and Methods.....	11
Sample Collection.....	11
Figure 2.1. Map of Turneffe Atoll, Belize.....	12
Morphology.....	13
Mitochondria.....	13
Microsatellites.....	14
Joint Species Identification.....	16
Figure 2.2. Venn diagram showing the methods of identification used for the samples..	17
Characterizing Genetic Diversity.....	17
Hybridization.....	18

Shoal Composition.....	19
Results.....	20
Morphology.....	20
Mitochondria.....	21
Figure 2.3. Haplotype network showing COI haplotypes.....	21
Microsatellites.....	21
Figure 2.4. Structure plot showing cluster assignments	22
Joint Species Identification.....	23
Characterizing Genetic Diversity.....	23
Figure 2.5. Principal component analysis of morphological, mitochondrial, and microsatellite data	25
Hybridization	25
Shoal Composition.....	26
Figure 2.6. Percentage of each shoal (A-M) which was identified as <i>Coryphopterus</i> <i>personatus</i>	26
Discussion.....	27
Acknowledgements.....	31
 CHAPTER III: Photogrammetry derived habitat model enables characterization of fine-scale habitat use in a pair of coral reef fishes	 33
Abstract.....	33
Introduction.....	33
Methods.....	37

Population and habitat surveys	37
Figure 3.1. Sampling locations	38
Photo processing and photogrammetry.....	38
Figure 3.2. Panels showing site orthomosaics overlaid with <i>Coryphopterus hyalinus</i> and <i>C. personatus</i> shoals	40
Topographic Measurement	40
Distribution modelling.....	42
Shoal composition modelling	43
Habitat Quality Metric Example.....	44
Results.....	45
Site & Shoal Description	45
Structure-from-Motion model statistics.....	46
Topography.....	46
Figure 3.3. Panels showing the mean estimated number of <i>C. hyalinus</i> and <i>C. personatus</i> m ⁻²	48
Distributional Model.....	50
Figure 3.4. Model goodness-of-fit plot.....	50
Figure 3.5. Modelled parameter estimates	52
Shoal composition model.....	52
Habitat Quality Metric Example.....	53
Figure 3.6. Relationship between habitat quality and number of fish	54
Discussion.....	54
Acknowledgements.....	57

CHAPTER IV: Effect of habitat heterogeneity on dispersal	59
Abstract	59
Introduction.....	59
Methods.....	64
Sampling	64
Figure 4.1: Map showing sampling locations on the windward side of Turneffe Atoll, Belize.	65
Genomic Library Preparation and Sequencing	65
Bioinformatics.....	68
Species Identification.....	70
Relatedness	71
Dispersal Estimation	73
Habitat Quality	75
Analysis of Habitat Quality & Dispersal Distribution	76
Results.....	77
Bioinformatics.....	77
Species Identification.....	78
Relatedness	78
Figure 4.2: Probabilistic Families.	80
Dispersal Estimation	81
Figure 4.3: Dispersal Kernel.	81
Figure 4.4: Familial Dispersal Kernels.	82
Dispersal Habitat Relationship	82

Figure 4.5: Effect Heterogeneity and Quality on Dispersal.....	83
Discussion.....	84
Appendix.....	87
Sampling Distribution.....	87
Weighting.....	88
Dispersal Distribution.....	88
Acknowledgements.....	88
CHAPTER V: CONCLUSIONS	90
REFERENCES	92
APPENDIX A.....	128
Supplement for Chapter 2.....	128
APPENDIX B.....	132
Supplement for Chapter 3.....	132
Supplemental Methods.....	132
Habitat Classification.....	132
Specimen collection and species identification	133
Supplemental Results.....	134
Habitat Classification.....	134
Figure 3.S1. Composition of shoals at each site.....	138
Figure 3.S2. Topography of Site A.....	139
Figure 3.S3. Topography of Site B.....	140
Figure 3.S4. Topography of Site C.....	141

Figure 3.S5. Topography of Site D.....	142
Figure 3.S6. Topography of Site E.	143
Figure 3.S7. Topography of Site F.	144
Figure 3.S8. Topography of Site G.....	145
Figure 3.S9. Topography of Site H.....	146
Figure 3.S10. Topography of Site I.	147
Figure 3.S11. Topography of Site J.	148
Figure 3.S12. Topography of Site K.....	149
Figure 3.S13. Topography of Site L.	150
Figure 3.S14. Fine-scale complexity principal component analysis.....	151
Figure 3.S15. Observed Fish Density with predicted 95% credible interval.....	152
Figure 3.S16. Modelled and observed pairwise distances between individuals for each site.....	153
APPENDIX C	154
Supplement for Chapter 4.....	154
Figure 4.S1: Heterogeneity and Quality Comparison.....	154
Figure 4.S2: Ten-fold cross-validation error of ADMIXTURE assignments.....	155
Figure 4.S3: Admixture plot	155
Figure 4.S4: Traceplot of parameter estimates	156
Figure 4.S5: Assignment plot using NewHybrids	156
Figure 4.S6: True relatedness vs Simulated relatedness.....	157
Figure 4.S7: Misclassification of Unrelated Dyads.....	158

LIST OF FIGURES

	Page
Figure 2.1. Map of Turneffe Atoll, Belize	12
Figure 2.2. Venn diagram showing the methods of identification used for the samples.	17
Figure 2.3. Haplotype network showing COI haplotypes.....	21
Figure 2.4. Structure plot showing cluster assignments	22
Figure 2.5. Principal component analysis of morphological, mitochondrial, and microsatellite data.....	25
Figure 2.6. Percentage of each shoal (A-M) which was identified as <i>Coryphopterus personatus</i>	26
Figure 3.1. Sampling locations	38
Figure 3.2. Panels showing site orthomosaics overlaid with <i>Coryphopterus hyalinus</i> and <i>C.</i> <i>personatus</i> shoals	40
Figure 3.3. Panels showing the mean estimated number of <i>C. hyalinus</i> and <i>C. personatus</i> m ⁻² ..	48
Figure 3.4. Model goodness-of-fit plot.....	50
Figure 3.5. Modelled parameter estimates	52
Figure 3.6. Relationship between habitat quality and number of fish	54
Figure 4.1: Map showing sampling locations on the windward side of Turneffe Atoll, Belize..	65
Figure 4.2: Probabilistic Families.	80
Figure 4.3: Dispersal Kernel.	81
Figure 4.4: Familial Dispersal Kernels.	82
Figure 4.5: Effect Heterogeneity and Quality on Dispersal.....	83
Figure 2.S1. STRUCTURE Trace Plot	129

Figure 2.S2. NEWHYBRIDS Trace Plot	130
Figure 2.S3. Evanno Method Cluster Identification	130
Figure 2.S4. PCA Loadings Plot.....	131
Habitat Classification.....	132
Figure 3.S1. Composition of shoals at each site.	138
Figure 3.S2. Topography of Site A.....	139
Figure 3.S3. Topography of Site B.	140
Figure 3.S4. Topography of Site C.	141
Figure 3.S5. Topography of Site D.....	142
Figure 3.S6. Topography of Site E.	143
Figure 3.S7. Topography of Site F.	144
Figure 3.S8. Topography of Site G.....	145
Figure 3.S9. Topography of Site H.....	146
Figure 3.S10. Topography of Site I.	147
Figure 3.S11. Topography of Site J.	148
Figure 3.S12. Topography of Site K.....	149
Figure 3.S13. Topography of Site L.	150
Figure 3.S14. Fine-scale complexity principal component analysis.....	151
Figure 3.S15. Observed Fish Density with predicted 95% credible interval.....	152
Figure 3.S16. Modelled and observed pairwise distances between individuals for each site	153
Figure 4.S1: Heterogeneity and Quality Comparison.....	154
Figure 4.S2: Ten-fold cross-validation error of ADMIXTURE assignments.....	155
Figure 4.S3: Admixture plot	155

Figure 4.S4: Traceplot of parameter estimates 156

Figure 4.S5: Assignment plot using NewHybrids 156

Figure 4.S6: True relatedness vs Simulated relatedness. 157

Figure 4.S7: Misclassification of Unrelated Dyads. 158

LIST OF TABLES

	Page
Table 2.1. Identification of each of the 18 specimens with disagreements between identification methods	19
Table 2.2. Per locus population summary statistics for each species	24
Table 3.1. Site level summary statistics.....	49
Table 4.1: Minimum number of Loci.	79
Table 2.S1. PCR Conditions for each locus.....	128
Table 3.S1. Habitat classification model quality statistics	135
Table 3.S2. Summary of model frameworks	136
Table 3.S3. Classification quality metrics across all sites	137

CHAPTER I: INTRODUCTION

Demographic processes define a population's ability to persist in an environment. In completely isolated populations the primary drivers of demography are the birth and death rates. However, such isolated populations are rare in nature and as such further depend on immigration and emigration to define the long-term trend of the population. These metapopulations form complex interconnected networks, linked together by the exchange of migrants (Hanski 1991). Evolutionarily-relevant migration between populations is the result of movement of an individual and subsequent reproduction within the new population (Pineda et al. 2007). This can, and does, occur at any number of different life stages and with varying degrees of frequency during the life of an individual.

The decision to migrate is often one of the most potentially costly decisions that may occur in the life of an organism (Bonte et al. 2012). This cost is a result of the combined effects of opportunity costs associated with altered activity states during migration, as well as direct risks which may be accrued while migrating (Bonte et al. 2012). One of the great unknowns that makes migration such a risky prospect is the uncertainty of the destination. For many organisms the act of migration is akin to buying a one-way plane ticket to an unknown destination, with no ability to return later in life. This uncertainty, along with the risk of death on the journey is what makes migration into a new population such a high-stakes proposition. Yet, across all taxa dispersal among populations is the rule rather than the exception. As such to have been universally selected for across all species, there must be benefits for the individual dispersers making the risks worthwhile, and plausibly methods to mitigate some of these risks (Bonte et al. 2012).

Some general methods have evolved to mitigate some of the risks associated with the act of dispersal. The first class of mitigation strategies can be thought of as “risk management” or “risk avoidance” strategies, while the second class is more accurately described as information acquisition and use strategies. The most extreme risk management strategy is the evolution of a complete lack of dispersal (Duffy 1996, Planes and Doherty 1997, Bernardi and Vagelli 2004). Less extreme variants of this basic mitigation strategy involve minimizing the length of time spent dispersing (D’Aloia et al. 2013, Salles et al. 2016, D’Aloia and Neubert 2018). These strategies are particularly important in organisms like fishes that experience extremely high rates of mortality during their larval dispersive phase (Houde 1997, Buston et al. 2012). Another risk management strategy common across taxa is to disperse as cohesive packet of individuals, where a cohort remains spatially coherent throughout dispersal (Riquet et al. 2017, Burgess et al. 2017, Berenshtein et al. 2018). This has been observed in species as diverse as the humbug damselfish, *Dascyllus aruanus*, and the sub-social spider, *Stegodyphus lineatus* (Johannesen and Lubin 2001, Buston et al. 2009). Selective group formation during dispersal is likely of greatest importance when predation is a key threat, and/or food resources are patchily distributed (Hamilton 1971, Shapiro 1983, Landeau and Terborgh 1986). The second class of mitigation strategies involves gathering and utilizing information to inform the process of dispersal. One method of doing this, particularly useful in organisms with relatively large home ranges and long lifespans, could involve active surveys of new areas and only subsequently moving after determining it to be a more suitable location (Isbell 2004, Selonen and Hanski 2006). In organisms which are sedentary some general heuristics could be used to determine the optimal dispersal syndrome. For example, if the habitat surrounding a sedentary individual is of highly variable quality and/or the habitat quality changes very slowly over time, then local dispersal is more selectively advantageous as

there is risk to leaving the natal habitat as the nearby habitat is likely of lower quality (Holt 1985, Baker and Rao 2004, Massol and Débarre 2015). Finally, organisms can use a variety of senses (Gerlach et al. 2007, Leis et al. 2011) to investigate locations during dispersal and steer themselves towards optimal locations (Atema et al. 2002, Hogan and Mora 2005, Burgess et al. 2022).

The evolution of mitigation strategies to minimize risks associated with dispersal implies that there are selective benefits for individuals if the various costs can be mitigated/avoided. One of these benefits is that dispersers can potentially colonize new areas and take advantage of resources with fewer competitors than are present at their natal site (Hanski and Thomas 1994, Bowler and Benton 2005). A corollary of this initial benefit is that dispersing away from the natal site results in the the dilution of related individuals across a larger area (Bowler and Benton 2005). As such the risk of inbreeding and kin competition is dramatically reduced as a result of dispersal away from the natal site. At the population level dispersal leads to increased gene flow, which can introduce novel alleles into a population conferring greater genetic diversity and thereby an increased resistance to perturbation (Bowler and Benton 2005, Ronce 2007, Duputié and Massol 2013).

The process of dispersal takes place within a dispersal matrix. For marine and aquatic organisms which generally have a bipartite life-history with a dispersive larval stage, the dispersal matrix is the water where currents and other oceanographic processes distribute propagules, some of which promote the aggregation of dispersing particles (Berenshtein et al. 2018). The duration of the larval period is generally thought to be coupled with the dispersal distance, with longer durations leading to greater dispersal potential (Levin 2006, Bradbury et al. 2008). However, many biotic and physical processes can act to reduce dispersal distances

(Sponaugle et al. 2002, Cowen and Sponaugle 2009). In marine organisms, species that lay benthic eggs exhibit higher levels of genetic differentiation than those that broadcast spawn into the water column, suggesting lower levels of dispersal (Riginos et al. 2014). Additionally, the developmental competence of the larvae can influence dispersal distance by affecting larval swimming performance (Stobutzki and Bellwood 1997), sensory abilities (Kingsford et al. 2002) and the timing of vertical migrations (Munk et al. 1989). Highly competent larvae can control their position in the water column and orient towards food sources and suitable habitat using auditory and olfactory cues, and possibly prevent advection by currents (Kingsford et al. 2002, Simpson et al. 2004, Dixson et al. 2014).

Local dispersal, where dispersal occurs but is spatially restricted to the area surrounding the natal site, has been observed in many reef fishes which exhibit some degree of self-recruitment (Jones et al. 1999, Planes et al. 2002, D'Aloia et al. 2013, 2018, Salles et al. 2016). Local dispersal is likely favoured by selection when spatial heterogeneity in patch quality is high, because there is an increased risk of leaving the “known” natal patch. Local dispersal is also favoured when temporal heterogeneity in patch quality is low, like many hard-bottom marine habitats including coral reefs; this is because the quality of the natal site is expected to remain stable over generations (Holt 1985, Baker and Rao 2004, Massol and Débarre 2015).

Additionally, local dispersal serves to minimize the length of the pelagic larval stage, thereby reducing the mortality associated with this costly life stage (Houde 1997, Buston et al. 2012).

Here I empirically assess theoretical hypotheses about how spatial habitat heterogeneity influences the dispersal of a common Caribbean reef goby, *Coryphopterus hyalinus* (Selwyn et al. in prep, Chapter 4). To accomplish this, we first need an understanding on how reef topography equates to habitat quality from the perspective of *C. hyalinus* and its sister species, *C.*

personatus (Selwyn et al. in review, Chapter 3). Finally, both *C. hyalinus* and *C. personatus* are found throughout the Caribbean with largely overlapping depth distributions (Baldwin and Robertson 2015, Robertson and Van Tassell 2019) which begs the question as to how these ecologically very similar species maintain a species boundary. As such we test the hypothesis that these species spatially segregate into species specific shoals on the reef (Selwyn et al. 2022, Chapter 2).

CHAPTER II: Maintenance of species boundaries within social aggregations of ecologically similar goby sister species

Abstract

The maintenance of species boundaries when opportunities for admixture are abundant, is a poorly understood phenomenon for many taxa. While many mechanisms for maintaining species boundaries have been described their relative importance depends largely on the particulars of the system in question. Aggregating social behavior can be a means to keep sympatric sister species distinct if it leads to segregation during reproduction. The widespread Caribbean reef gobies *Coryphopterus personatus* and *C. hyalinus* are sympatric sister species with nearly identical morphology that spend their entire adult lives in shoals in which reproduction occurs. To date no studies have investigated whether shoals are species-specific, which would be expected if aggregating behavior helps to maintain species boundaries. To address this, the species of individual fishes collected from 16 shoals were identified using morphology, mitochondrial sequence data, and microsatellite allele frequencies. Levels of admixture between the species were also assessed. Shoals were generally composed of both species in similar proportions to their relative abundances on the reef where the shoals were found, indicating that the species are not behaviorally segregating. For most specimens, morphological, mitochondrial, and nuclear data were congruent with a single species, but 18 individuals showed disagreements with microsatellite genotypes of 16 suggesting some level of historic/contemporary admixture. Of these, two were identified as likely first- or second-generation hybrids or backcrosses. Despite co-occurrence and evidence of some gene flow, the two species show little admixture overall suggesting that microscale differences in breeding site selection, allochrony, and/or

cryptic mate choice may play an important role in the maintenance of species boundaries despite cooccurrence well within the range typically thought of as sympatry.

Introduction

Understanding processes that maintain species boundaries in sympatry is a major focus of evolutionary ecology. Many ecological and evolutionary mechanisms have been identified in the maintenance of species boundaries, with their relative importance depending on the specifics of the study system (Harrison et al. 2017). This is particularly true amongst closely related species living in sympatry, which can lead to frequent opportunities for hybridization. Research has found that allochronic, ecological, and behavioral isolation, as well as gametic incompatibility, are frequently important (Coyne and Orr 2004, Harrison et al. 2017). Allochronic isolation occurs when gamete release and production are offset in time, leading to a reduced opportunity for interaction between gametes of sympatrically distributed species (Levitan et al. 2011, Bouwmeester et al. 2021) and can occur on scales from hours to years (Knowlton et al. 1997, Rosser 2015, Tarpey et al. 2017). Ecological isolation occurs when species utilize different ecological niches which subsequently minimizes opportunities for mating interactions (Bovbjerg 1970). Behavioral isolation occurs when differences in behavior develop that impact the likelihood of heterospecific mating and can include differences in mating behavior (Parchman et al. 2013) and/or formation of spatially segregated social groups (Gerhardt 1974, Diabaté et al. 2009). Gametic incompatibility occurs when a viable zygote is not formed during fertilization (Rawson et al. 2003).

Sister species need not be isolated completely, and a continuum of states exists between the homogenization of once distinct gene pools to complete reproductive isolation, as levels of gene flow between species decrease and larger portions of genomes become isolated (Kopp and Frank

2005, Harrison and Larson 2014). Furthermore, as incipient species form, recurrent gene flow can lead to further diversification (through reinforcement or differential gene flow) or homogenization (Mallet 2005, Abbott et al. 2013). If species potentially interact across large geographic areas, there may be differences in levels of contemporary gene flow related to local conditions (Muhlfeld et al. 2009, Gagnaire et al. 2013). For example, in colonial nesting waterbirds, such as gulls, reproductive barriers are incomplete in areas of overlapping breeding habitat, and species complexes are known to form (Liebers et al. 2004). Alternatively, other isolating mechanisms (e.g., gametic isolation, mate recognition) may maintain species boundaries despite close association. For example, sub-social colonial spiders in the genus *Chikunia* form mixed-species colonies and have been observed to indiscriminately provide care for heterospecific broods and yet appear to maintain evolutionary independence (Grinsted, Agnarsson, & Bilde, 2012; Smith et al., 2019). Because multiple mechanisms likely operate simultaneously, careful consideration of individual processes and their relative contribution to the cessation/interruption of gene flow is required to gain a holistic view of species boundaries (Coyne and Orr 2004).

For marine species with external fertilization gamete incompatibility and temporal/spatial offsets in gamete release, are thought to be particularly important (Levitan et al. 2004, Ohki et al. 2015) because gametes are released into a dispersive environment (Babcock et al. 1994) and may remain viable for hours to days (Williams and Bentley 2002). Social behavior seen in many mobile marine species including formation of spawning aggregations, schooling, and/or monogamous pairing (Domeier and Colin 1997, Pavlov and Kasumyan 2000, Whiteman and Côté 2004) can further decrease the opportunity for interaction between heterospecific gametes. However, many social units feature heterospecifics and the observation of incomplete isolation

of interacting marine species is becoming increasingly common (Miranda et al. 2010, Montanari et al. 2012).

Gobies (Gobioidei) are small bodied short-lived fishes and comprise one of the most diverse sub-orders of vertebrates (Nelson et al. 2016). Habitat differentiation appears to be a primary driving force behind diversification and speciation (Thacker 2009) within the gobies, and sister taxa often segregate based on microhabitat (Brandl et al. 2018). Strong natural selection associated with habitat specialization can maintain species boundaries (Rice 1987, Teske et al. 2019, Öhlund et al. 2020) and is hypothesized to be important within the taxon (Brandl et al. 2018). For example, mudskipper diversification appears to have proceeded by differentiation into ecological guilds characterized by differences in salinity and water quality tolerance, as well as degree of terrestriality (Polgar et al. 2010). On coral reefs, microhabitat utilization is thought to be a contributing factor to the degree of diversification seen in the genus *Elacatinus*, which initially differentiated into sponge and coral-dwelling groups, with later diversification based on other ecological characteristics (Taylor and Hellberg 2005, Colin 2010). In a similar manner, diversification in the genus *Eviota* seems to be related to the degree of association with specific microhabitats including coral, rubble, or sand (Tornabene et al. 2013).

Most species of Caribbean gobies are solitary, however, in the genus *Coryphopterus* two sister species, *C. personatus* and *C. hyalinus* occupy social aggregations throughout juvenile and adult life stages (Allsop and West 2004). The genus *Coryphopterus* is a relatively recent radiation of fourteen species, one from the eastern Pacific with the remaining thirteen in the western Atlantic, arising within the last 30 million years (Baldwin and Robertson 2015, Thacker 2015). The two species, *C. personatus* and *C. hyalinus*, were initially split based on counts and positions of anterior head pores of the cephalic lateralis sensory system, with *C. hyalinus* having

laterally paired pores and *C. personatus* having a single median pore (Böhlke and Robins 1962). Phylogenetic analysis consistently resolves *C. personatus* and *C. hyalinus* as sister species with a relatively recent common ancestor (Baldwin et al. 2009). Additionally, their status as species is supported by 7.14% (6.79% – 7.65%) sequence divergence between the species at the mitochondrially-encoded COI gene, as compared to 0.06% and 0.14% sequence divergence within each species respectively (Baldwin et al. 2009). Despite these genetic differences the ecology of these two species is similar and their geographic distribution is nearly completely overlapping throughout the entirety of the Greater Caribbean (Robertson and Van Tassell 2019), leading many researchers to lump them together in ecological studies (e.g. Serna Rodríguez et al. 2016, Chagaris et al. 2017). While *C. hyalinus* is generally collected from slightly deeper depths, the overall depth range of the two species shows near complete overlap. At Turneffe Atoll both species are frequently observed between 0 and 27 m with *C. hyalinus* occurring at depths greater than 30 m (Greenfield and Johnson 1999). Caribbean wide depth ranges for each are between 1 and 52 m, with *C. personatus* being observed to 70 m depths (Baldwin and Robertson 2015).

Additionally, the species are reproductively similar, as both are protogynous hermaphrodites that lay and fertilize eggs within the reef structure and form large shoals of up to tens to thousands of individuals, with shoals in the current study area typically composed of fewer than 100 individuals (Böhlke and Robins 1962, Robertson and Justines 1982, Cole and Robertson 1988, Selwyn et al. 2021). These shoals are spatially discrete, temporally stable aggregations on the reef, which reform rapidly when disturbed and exist in the same location for multiple days (J. Selwyn pers. obs.). Shoals serve a number of purposes, including reproduction, with the ratio of males to females within a shoal influenced by shoal density and the ability of males to monopolize mates (Allsop and West 2004). In both species mating occurs between pairs of

individuals within reef crevices, with males guarding and aerating the fertilized nest (Thresher 1984, Gardner 2000). Unfortunately, little else is known about the reproductive behaviors of these species in nature. The reproductive life span is short (~100 days; Beeken et al. 2021) and spawning is frequent (every 7 – 10 days; Gardner 2000). Because opportunity for adult dispersal is limited as adults are small and not highly mobile; individuals likely spend their entire reproductive life within a single shoal (Selwyn et al. 2016).

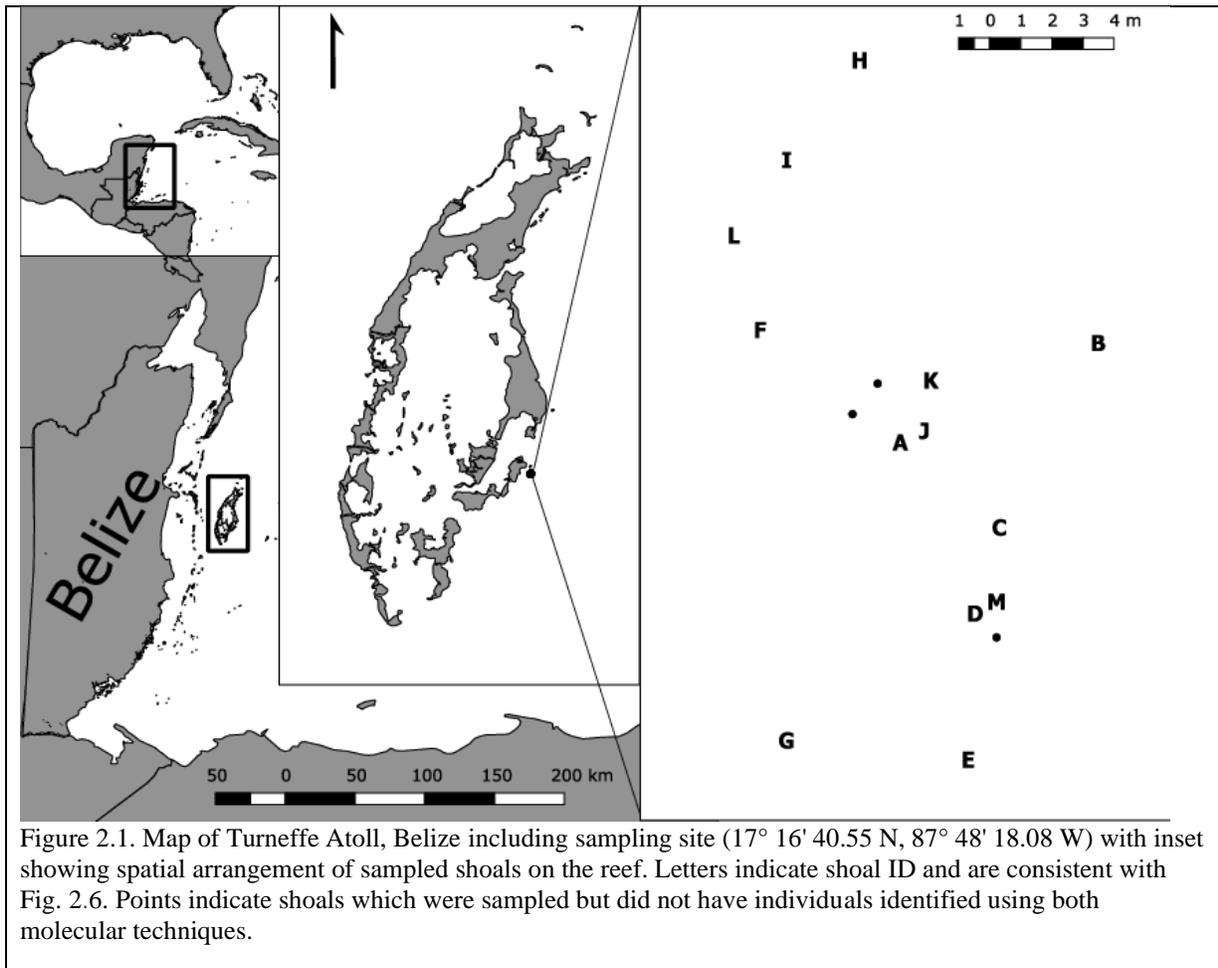
The main objective of this study was to identify whether species-specific aggregation was the mechanism maintaining species boundaries between the sympatric gobies, *C. personatus* and *C. hyalinus*. We hypothesized that *C. personatus* and *C. hyalinus* would form spatially segregated shoals composed of a single species or both species but in ratios significantly different than the background. If shoals contain both species, there is the potential for admixture which can be assessed using genetic techniques. Therefore, samples of individuals were taken from multiple, spatially explicit shoals spread across a single reef system, and analyzed using molecular and morphological characters to determine individual species identity, shoal composition, and degree of admixture.

Materials and Methods

Sample Collection

Individual *Coryphopterus hyalinus/personatus* were collected by divers on SCUBA using hand nets from a single reef (17° 16' 40.55" N, 87° 48' 18.08" W, Fig. 2.1) in Turneffe Atoll, Belize, in August 2014. Turneffe Atoll is composed of numerous mangrove islands approximately 9 – 23 km offshore from the main Belize Barrier Reef. This area has records of both *C. hyalinus* and *C. personatus* and is composed of suitable forereef habitat in depths where both species are commonly found (Greenfield and Johnson 1999). The site habitat composition is

typical of shallow, windward forereef locations in Turneffe Atoll (for more detailed description see: Garcia and Holtermann 1998). This reef was selected for study of small spatial scale interaction of *C. personatus* and *C. hyalinus*, as the depth (16 m) of this reef is where the maximum number of both *C. hyalinus* and *C. personatus* have been observed at Turneffe Atoll previously (Greenfield and Johnson 1999), located between the shallow water coral heads preferred by *C. personatus*, and the deeper coral walls preferred by *C. hyalinus* (Victor 2019).



A total of 428 individual *Coryphopterus sp.* were collected from 16 shoals. Fish were humanely euthanized using buffered MS222 and stored in 95% non-denatured ethanol. Individual shoals were kept separate during and after collection. All collections were performed

in accordance with the ethical guidelines of Texas A&M University – Corpus Christi (TAMUCC-AUP-05-14) and in compliance with standards outlined in the US National Research Council's Guide for the Care and Use of Laboratory Animals. Collections were made with the express permission of the government of Belize (Aquatic Scientific Research Permit 000044-13).

Morphology

To distinguish between *C. personatus* and *C. hyalinus* using morphology, anterior interorbital cephalic (AIC) pores were counted using a dissecting microscope; *Coryphopterus personatus* have one pore, while *C. hyalinus* have two (Böhlke and Robins 1962). Because these pores only develop in larger individuals (> 10 mm SL), morphological identification was attempted only with individuals larger than this threshold (Victor 2019) and a total of 134 (31.3%) specimens were examined.

Mitochondria

The mitochondrial gene cytochrome c oxidase subunit 1 (COI) was amplified from DNA extracted from fin and muscle tissue from the caudal end of each fish using either E.Z.N.A.® DNA extraction kit (Omega Bio-tek) or Chelex extraction using a multiplex reaction with four universal fish primers (de Lamballerie et al. 1992, Ward et al. 2005): FishF1 (5'-TCAACCAACCACAAAGAGATTGGCAC-3'), FishF2 (5'-TCGACTAATCATAAAGATATCGGCAC-3') and FishR1 (5'-TAGACTTCTGGGTGGCCAAAGAATCA-3'), FishR2 (5'-ACTTCAGGGTGACCGAAGAATCAGAA-3'). Each 30µl reaction contained 1 X buffer (pH 8.5), 1.5 mM MgCl₂, 0.20 mM dNTPs each, 0.04% Tween, 250 nM forward and reverse primers each (F1 and F2, R1 and R2), 1.0 U *Taq* polymerase and 1.0 µl of template. PCR amplification was run with initial denaturing at 95 °C for 2 minutes, followed by 35 cycles of denaturation at

95 °C for 1 minute, annealing at 50-55 °C for 1 minute, and elongation at 72 °C for 1 minute. A final round of elongation was run at 72 °C for 10 minutes. Amplicons were cleaned using 0.7X Mag-Bind® Total Pure NGS beads (Omega Bio-Tek) and sequenced at the Genomics Core Lab at Texas A&M University – Corpus Christi or at Retrogen, Inc. (San Diego, CA) on 96-capillary ABI 3730xl Genetic Analyzer (Applied Biosystems Inc.). Chromatographs were edited by eye using SEQUENCHER v.5.4.6 (GeneCodes Corporation). All COI sequences from this study can be found on GenBank (accession numbers MT784949 - MT785286).

To determine species identity from COI sequences, a haplotype network was created using individuals sequenced from this sampling along with 11 sequences of each species downloaded from GenBank, which had been morphologically identified by Baldwin et al (2009; Accession numbers: GQ367313 - GQ367334). Sequences from this study and those from Baldwin et al. (2009) were then aligned using CLUSTAL W implemented in MSA (Thompson et al. 1994, Bodenhofer et al. 2015) and trimmed to retain a core region of 547 bp for all individuals. A median joining network was created using POPART (Bandelt et al. 1999, Leigh and Bryant 2015) and used to assign individuals to species. For individuals where morphological and the genetic identity were incongruent, specimens were reexamined microscopically without prior knowledge of whether COI sequences were consistent with *C. hyalinus* or *C. personatus*. Net genetic divergence between species and average genetic distance within species was calculated using STRATAG (Nei and Kumar 2000, Archer et al. 2017).

Microsatellites

Nine microsatellite loci were amplified for 384 specimens (89.7%; Table S1; Hepburn et al. 2005, Hogan et al. 2010). Each 10 µl reaction contained 1X buffer (pH 8.5), 3.0-4.5 mM MgCl₂, 0.8 mM each dNTPs, 100-500 nM fluorescent labelled forward and unlabelled reverse primers,

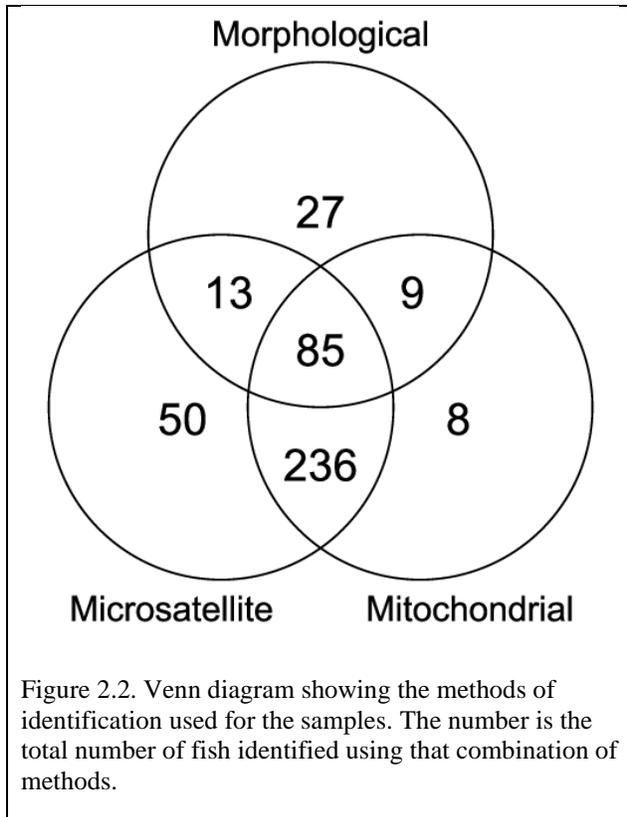
0.5 U *Taq* polymerase, and 1.0 µl of template (Table S1). PCR amplification was run using a touchdown protocol with an initial denaturation of 94 °C for 3 min, followed by a 40-50 cycles (Table 2.S1) of denaturation for 15 sec at 94 °C, annealing for 45 sec at 68 to 64 °C - 58 to 52 °C (Table 2.S1), and elongation for 30 sec at 72 °C, followed by 5 min elongation at 72 °C.

Amplicons were analysed at the Texas A&M University – Corpus Christi Genomics core lab on a 96-capillary ABI 3730xl Genetic Analyzer with the Liz 600® (Applied Biosystems Inc.) size standard. Size polymorphisms were scored by eye using GENEMARKER 2.6.4 software (SoftGenetics Inc.).

To determine species identity of individuals based on the microsatellite loci, STRUCTURE was run using the admixture model with correlated allele frequencies (Pritchard et al. 2000). Prior to running any models two loci (COPE 10 and CPER52) were removed due to failure to amplify in more than 20% of individuals. The models were run for all values of K (number of distinct clusters) between 1 and 15, using 1,000,000 burn-in iterations followed by 10,000,000 sampling iterations with a thinning interval of 100 and 10 replicate runs. Mixing, proper exploration of parameter space, and chain convergence were confirmed by visually inspecting trace plots, which show parameter values across MCMC steps, and ensuring the \hat{R} value, a measure of chain convergence, equaled one (Fig. 2.S1; Vehtari et al. 2019). The optimal value of K was then determined using the Evanno method (Evanno et al. 2005). After performing STRUCTURE analysis, species identity as determined from COI was matched with each individual's cluster assignments, and clusters were associated with one species or the other, based on the majority of individuals in a cluster having the same COI species identity.

Joint Species Identification

Because DNA quality of some specimens was poor, not all fish could be identified using both molecular methods (Fig. 2.2). Therefore, a final species identification was made only when both COI and microsatellite data were available ($n = 321$, number of shoals = 13). To visualize differences among specimens based on all three identification methods, a principal components analysis (PCA) combining morphological, mitochondrial, and microsatellite data was performed. Morphological identification was not required as the informative character does not develop until the fish are above 10 mm standard length (Victor 2019). The size at which AIC pores become informative characters is coincidentally similar to the minimum size observed for the transition from female to male (~ 13 mm TL, Cole and Robertson 1988, $TL = -1.5 + 1.3 * SL$, 10 mm SL = 11.1 mm TL; 95% prediction interval = $10.1 - 12.2$, unpublished data) though it should be noted there are many mature females larger than this size (Cole and Robertson 1988). Individuals for which microsatellite and mtDNA-based identification disagreed or one of the two marker types failed to amplify, were excluded from analysis of shoal composition ($n = 107$).



Characterizing Genetic Diversity

After making final species identifications, microsatellites were used to characterize variation within and between *C. hyalinus* and *C. personatus*, excluding all samples with uncertain species identity. Within species diversity measures included, per locus rarefied allelic richness (A_R , Hurlbert 1971), number of private alleles, corrected expected heterozygosity (H_e , Nei and Chesser 1983), and the inbreeding coefficient (F_{IS} , Nei and Chesser 1983). Each locus within each species was tested for Hardy-Weinberg Equilibrium with the p -values corrected for familywise error using the sequential Bonferroni correction (Holm 1979). Additionally, both overall and per locus fixation indices (F_{ST} , Nei 1973) were calculated with significance tested using 10,000 permutations. The maximum F_{ST} given the genetic diversity of these loci was also calculated (Hedrick 2005). These metrics were calculated using in R v3.5.1 using the packages

ADEGENET, PEGAS, POPPR, and HIERFSTAT (Goudet 2005, Jombart 2008, Paradis 2010, Kamvar et al. 2014, R Core Team 2018).

Hybridization

To discriminate between historic and contemporary gene flow a Bayesian analysis of hybridization, based upon the microsatellite data, was performed using NEWHYBRIDS (Anderson and Thompson 2002). This analysis was done to identify admixed individuals, including first- and second-generation (the offspring of two first-generation hybrids) hybrids and first-generation backcrosses. For the analysis, 5 independent MCMC chains, with 100,000 burnin iterations and a subsequent 1,000,000 sampling iterations, were run. Uniform priors were used for both mixing proportions and allele frequencies to minimize the influence of rare alleles on the classifications. Proper mixing, exploration of parameter space, and convergence were confirmed by visually inspecting trace plots and confirming that the \hat{R} value equaled one (Fig. 2.S2; Vehtari et al. 2019). Due to skew in sample sizes between the species, a general lack of fixed alleles, and the relatively small number of loci, combined with evidence of strong genetic differentiation between individuals confidently assigned to one species or the other, only specimens with mismatching species identifications and/or low STRUCTURE assignment probabilities were assessed as potential hybrids, to minimize the rate of Type I error (False Positives, Table 2.1). To further mitigate misclassification errors, all hybrid categories were merged into a single category (hereafter admixed).

Table 2.1. Identification of each of the 18 specimens with disagreements between identification methods and the 16 specimens with less than 0.9 STRUCTURE assignment probability. Specimens in the first section showed disagreement between the morphological and molecular identification methods. Specimens in the middle section have disagreements between the nuclear and mitochondrial identification methods. Specimens in the final section have assignment probabilities < 0.9. STRUCTURE assignment probability shows the probability of assignment to species indicated in the Microsatellite column from the STRUCTURE analysis. The STRUCTURE assignment probabilities for both species cluster is shown for each specimen with microsatellite data. The posterior probability of each specimen being a pure *C. hyalinus*/*personatus* or a first- or second-generation hybrid based on the NEWHYBRIDS analysis is also shown for each specimen with microsatellite data. Individuals in bold are those with > 50% probability of being a first- or second-generation hybrid.

ID	Shoal	Morphological	Mitochondrial	Microsatellite	STRUCTURE		NEWHYBRIDS		
					<i>C. hyalinus</i>	<i>C. personatus</i>	<i>C. hyalinus</i>	Admixed	<i>C. personatus</i>
0002	A	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.977	0.023	0.995	0.005	0
0032	J	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.990	0.010	0.993	0.007	0
0090	F	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.993	0.007	0.997	0.003	0
0103	F	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.988	0.012	0.992	0.008	0
0104	F	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.991	0.009	0.996	0.004	0
0119	L	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.990	0.010	0.994	0.006	0
0130	L	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.991	0.009	0.995	0.005	0
0151	B	<i>C. personatus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.993	0.007	0.994	0.006	0
0158	B	<i>C. personatus</i>	-	<i>C. hyalinus</i>	0.993	0.007	0.997	0.003	0
0382	N	<i>C. personatus</i>	-	<i>C. hyalinus</i>	0.995	0.005	0.998	0.002	0
1284	O	<i>C. hyalinus</i>	<i>C. personatus</i>	-	-	-	-	-	-
1313	O	<i>C. personatus</i>	<i>C. hyalinus</i>	-	-	-	-	-	-
1314	O	<i>C. personatus</i>	<i>C. hyalinus</i>	-	-	-	-	-	-
0039	J	-	<i>C. hyalinus</i>	<i>C. personatus</i>	0.200	0.800	0.693	0.200	0.108
0084	K	-	<i>C. hyalinus</i>	<i>C. personatus</i>	0.272	0.728	0.281	0.636	0.083
0094	F	<i>C. hyalinus</i>	<i>C. hyalinus</i>	<i>C. personatus</i>	0.432	0.568	0.927	0.061	0.012
0216	M	-	<i>C. hyalinus</i>	<i>C. personatus</i>	0.098	0.902	0.508	0.301	0.191
0302	I	-	<i>C. hyalinus</i>	<i>C. personatus</i>	0.185	0.815	0.553	0.197	0.249
0025	A	<i>C. hyalinus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.899	0.101	0.907	0.093	0.001
0040	J	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.889	0.111	0.922	0.076	0.002
0051	K	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.859	0.141	0.787	0.213	0
0081	K	-	<i>C. personatus</i>	<i>C. personatus</i>	0.186	0.814	0.102	0.671	0.227
0145	L	-	<i>C. personatus</i>	<i>C. personatus</i>	0.138	0.862	0	0.050	0.95
0152	B	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.848	0.152	0.780	0.212	0.008
0157	B	-	<i>C. personatus</i>	<i>C. personatus</i>	0.196	0.804	0.035	0.365	0.600
0252	C	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.897	0.103	0.986	0.013	0
0262	C	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.858	0.142	0.976	0.024	0
0287	I	<i>C. hyalinus</i>	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.638	0.362	0.971	0.028	0.001
0305	I	-	<i>C. hyalinus</i>	<i>C. hyalinus</i>	0.897	0.103	0.964	0.036	0
0383	N	-	-	<i>C. personatus</i>	0.202	0.798	0.870	0.110	0.019

Shoal Composition

Because the individuals collected represent a random sample taken from each shoal that are composed of less than ~100 individuals, a 95% credible interval around the observed proportion of *C. personatus/hyalinus* present in each shoal, and overall on the reef, was estimated using an algebraic solution of the binomial distribution using the beta distribution as the conjugate prior (Gelman et al. 2013), for all models a uninformative conjugate prior was used ($\beta(1, 1)$). To

determine if species proportions in shoals differed from species proportions at the collection site, an algebraic solution to the difference in two proportions was calculated in R v3.5.1 (Pham-Gia et al. 1993, R Core Team 2018). If zero was contained within the 95% credible interval of the difference between the site and shoal proportions then the shoal was determined to be composed of a random mixture of species. If zero was not contained within the 95% credible interval, then the shoal was determined to contain a biased mixture with either more *C. personatus* (greater than zero) or more *C. hyalinus* (less than zero) than would be expected from a random sample taken across shoals at the site.

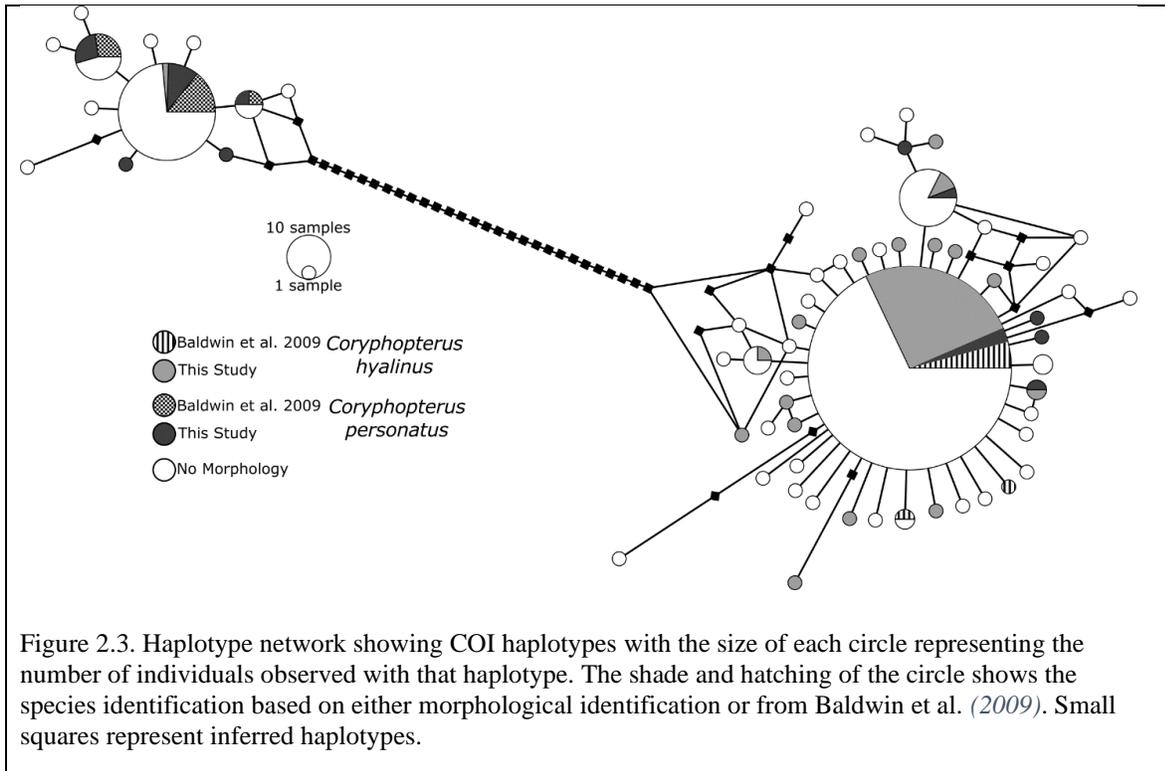
Results

Morphology

The collected specimens ranged from 5 – 22 mm standard length (SL). This distribution of lengths was not normal and skewed to smaller individuals with ~78% of the specimens likely being new recruits (Beeken et al. 2021). Thirty (22.4%) individuals had a single anterior interorbital cephalic (AIC) pore, consistent with *C. personatus* and 104 (77.6%) individuals had two AIC pores, consistent with *C. hyalinus*. Specimens which were able to be identified morphologically were significantly larger than those which could not be identified morphologically (Kruskal-Wallis's $\chi^2_{(1)} = 176.8, p < 0.0001$). Moreover, specimens which were misidentified based on morphology tended to be larger than those which were correctly identified (Kruskal-Wallis's $\chi^2_{(1)} = 6.04, p = 0.014$). There was a significant difference in SL between *C. hyalinus* and *C. personatus* based on morphological identification (Kruskal-Wallis's $\chi^2_{(1)} = 9.38, p = 0.0022$) with no differences observed when using all other methods of species identification (see below; COI: Kruskal-Wallis's $\chi^2_{(1)} = 1.23, p = 0.269$; Microsatellite: Kruskal-Wallis's $\chi^2_{(1)} = 0.81, p = 0.397$; Joint Method: Kruskal-Wallis's $\chi^2_{(1)} = 0.72, p = 0.370$).

Mitochondria

The COI locus was successfully amplified for 338 (79.0%) individuals. After trimming sequences to contain only a shared core of 547 bp, there were 34 alternately fixed sites in the samples from Baldwin et al. (2009). The net sequence divergence between the two species in this sampling was 6.0%, while mean within species divergence was 0.12% in *C. hyalinus* and 0.13% in *C. personatus*. Consistent with Baldwin et al. (2009), there were two distinct groups of haplotypes separated by 27 mutations (Fig. 2.3). Based on COI alone 62 (18.3%) individuals were identified as *C. personatus* and 276 (81.7%) individuals as *C. hyalinus*.



Microsatellites

STRUCTURE identified two distinct genetic clusters following the Evanno method (Fig. 2.S3). Most individuals fully assigned (>90% assignment probability) to either one cluster or the other, with only 16 (4.2%) individuals showing evidence of more than 10% admixture (Fig. 2.4, Table

2.1). All individuals assigned to cluster 1 were identified as *C. hyalinus* using COI data, and cluster 1 was therefore assumed to represent *C. hyalinus*. Ninety-one percent of individuals in cluster 2 were identified as *C. personatus* using the COI data, and cluster 2 was therefore assumed to represent *C. personatus*. The rate of disagreement between nuclear and mitochondrial markers was 1.6%. Based on STRUCTURE analysis alone, 55 (14.3%) individuals were identified as *C. personatus* and 329 (85.7%) as *C. hyalinus*.

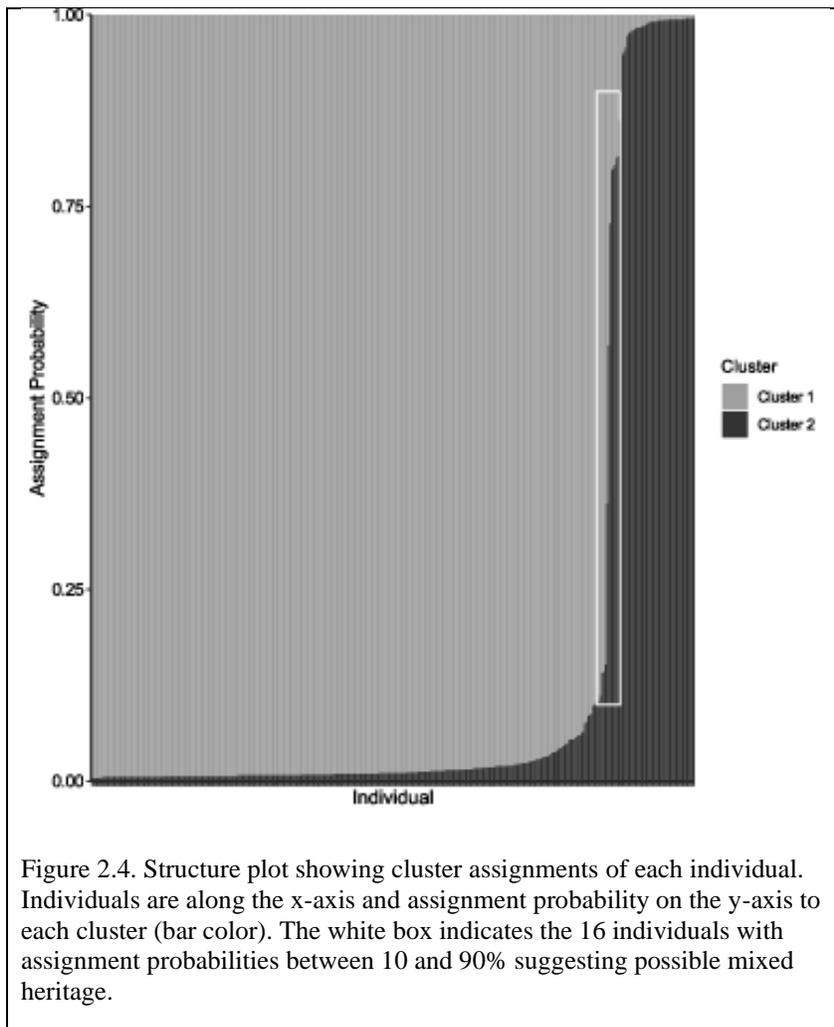


Figure 2.4. Structure plot showing cluster assignments of each individual. Individuals are along the x-axis and assignment probability on the y-axis to each cluster (bar color). The white box indicates the 16 individuals with assignment probabilities between 10 and 90% suggesting possible mixed heritage.

Joint Species Identification

Generally, species identification methods agreed but there were several potentially interesting examples of disagreements between the various methods (Table 2.1). A total of 13 specimens showed disagreements between morphological and molecular species identification, an error rate of 12.1%. Twelve of these samples were morphologically identified as *C. personatus* (one AIC pore) and one as *C. hyalinus* (two AIC pores). All 13 individuals were reexamined for morphology, blind to the original species identification, and the original morphological identification was confirmed in all cases. Five individuals with a COI species identification of *C. hyalinus* had a high STRUCTURE assignment probability < 0.9 to cluster 2 (*C. personatus*, Table 2.1).

Characterizing Genetic Diversity

A total of 107 (25.0%) individuals were only identified using one method or using only morphology and one molecular method and were excluded as a result. Of the remaining 321 specimens, 308 (96.0%) showed agreement among all three methods, or both molecular techniques. Of these, 15.6% were identified as *C. personatus* and 84.4% as *C. hyalinus*. Both species tend to exhibit homozygote excess across the same loci and overwhelmingly had elevated inbreeding coefficients. The high level of inbreeding and ubiquity of homozygote excess across loci likely results from the reproductive strategy and relatively short distance of larval dispersal leading to a heightened frequency of inbreeding rather than genotyping artifacts (Waples 2015, Selwyn 2015, Selwyn et al. 2016). The two species were significantly differentiated from each other ($F_{ST} = 0.19$, $p < 0.0001$, $F_{STmax} = 0.21$, Table 2.2).

Individuals definitively classified as either *C. personatus* or *C. hyalinus* fell into two well-formed clusters separating along PC1 (Fig. 2.5). Specimens which did not assign clearly to one

species or the other in STRUCTURE, and/or showed disagreements among identification techniques generally fell between the two clusters. The variables most associated with PC1 were alleles of microsatellite loci that had the highest frequency differences between the two species (Fig. 2.S4, Table 2.2). While PC2 is strongly associated with specimen morphology, matching with the observation that 12 out of the 13 morphological/molecular disagreements involved a specimen observed with a single AIC pore (Fig. 2.S4, Table 2.1). Specimens where disagreement was observed between morphology and molecular identification were significantly shifted to the right on PC1 and higher on PC2 (MANOVA $\eta_p^2 = 0.32$, Pillai's trace statistic = 0.64, $F_{(4,54)} = 6.4$, $p = 0.00027$, Fig. 5), relative to specimens with low assignment probabilities and specimens with mitochondrial-nuclear discordance. The observed placement of specimens with molecular-morphological discordance on PC1 seems to indicate that genetically these specimens are *C. hyalinus*. Meanwhile, specimens with either nuclear-mitochondrial discordance or low STRUCTURE assignment probabilities were generally found between the two main species clusters on PC1 (Fig. 2.5), and closer to zero on PC2.

Table 2.2. Per locus population summary statistics for each species (value in parenthesis after species name shows number of individuals). Summary statistics included are number of alleles (A), number of private alleles (A_P), rarefied allelic richness (A_R), expected (H_e) heterozygosity, and the inbreeding coefficient (F_{IS}). Underlined H_e indicates significant homozygote excess (all $p < 0.0028$). Finally, locus specific fixation indices (F_{ST}) are included with underlined F_{ST} values indicating significant differentiation. The p -value (p) indicates if the F_{ST} is significantly different from zero based on 10,000 permutations and has been corrected for familywise error using sequential Bonferroni (Holm 1979).

Locus	<i>Coryphopterus hyalinus</i> (260)					<i>Coryphopterus personatus</i> (48)					Fixation Index	
	A	A_P	A_R	H_e	F_{IS}	A	A_P	A_R	H_e	F_{IS}	F_{ST}	p
COPE5	69	57	32.36	<u>0.98</u>	0.21	14	4	12.49	0.60	0.18	0.20	0.0007
COPE9	10	2	7.75	<u>0.73</u>	0.21	9	1	8.80	<u>0.77</u>	0.43	0.01	0.0068
CPER26	10	4	5.87	<u>0.59</u>	0.15	8	2	7.21	0.79	0.26	0.29	0.0007
CPER92	12	2	6.01	<u>0.50</u>	0.41	9	0	10.00	<u>0.70</u>	0.45	0.05	0.0226
CPER99	4	0	2.48	0.11	-0.05	5	0	4.78	0.63	-0.07	0.77	0.0007
CPER119	20	2	13.61	<u>0.90</u>	0.46	19	2	17.98	<u>0.95</u>	0.56	0.01	0.0008
CPER188	12	7	6.68	0.51	0.03	7	3	5.72	0.32	-0.06	0.03	0.0008

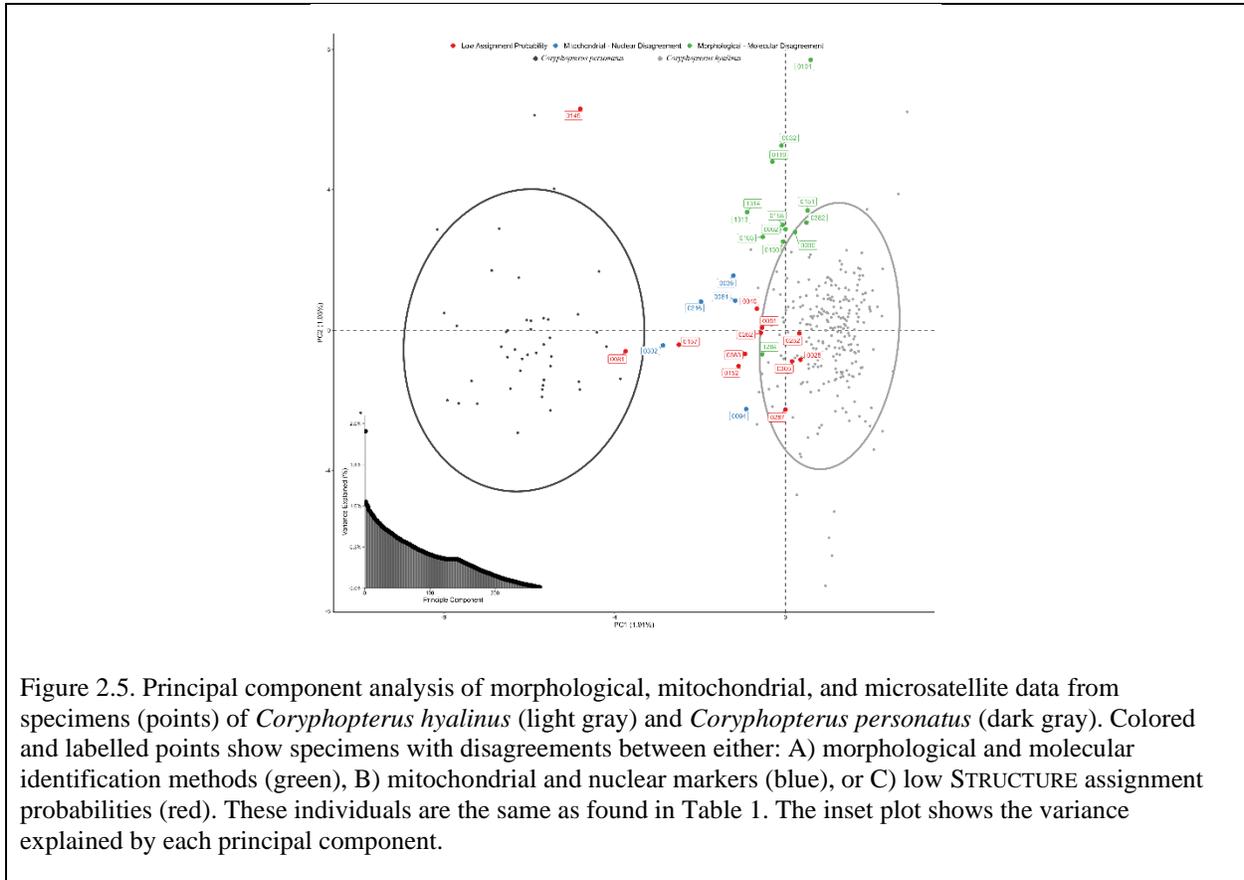


Figure 2.5. Principal component analysis of morphological, mitochondrial, and microsatellite data from specimens (points) of *Coryphopterus hyalinus* (light gray) and *Coryphopterus personatus* (dark gray). Colored and labelled points show specimens with disagreements between either: A) morphological and molecular identification methods (green), B) mitochondrial and nuclear markers (blue), or C) low STRUCTURE assignment probabilities (red). These individuals are the same as found in Table 1. The inset plot shows the variance explained by each principal component.

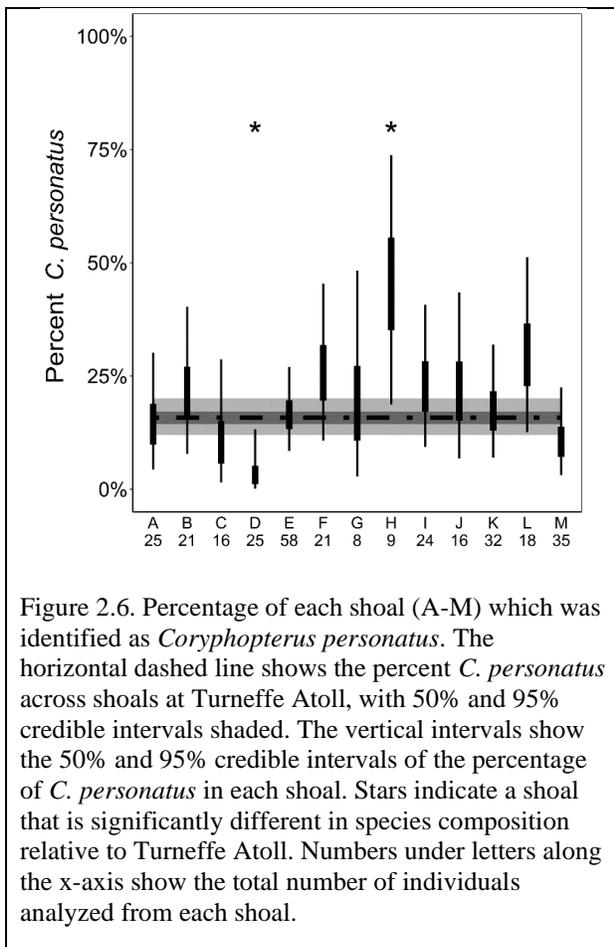
Hybridization

All specimens which showed disagreement between morphological and molecular identification methods were confidently classified as pure *C. hyalinus* (minimum posterior probability *C. hyalinus* assignment = 0.992). For all other specimens there was greater uncertainty in the delineation between admixed and pure species, with only two specimens identified as admixed with >50% posterior probability (0081 & 0084; Table 2.1). Based on body lengths (SL = 6 & 7 mm respectively) both individuals were likely new recruits (Beeken et al. 2021). Specimen 0084 was positioned centrally between the two species clusters in the PCA, while specimen 0081 was located near the *C. personatus* cluster. This is likely a manifestation of the agreement between nuclear and mitochondrial markers in 0081 and the disagreement

between these markers in 0084 (Table 2.1). For both specimens the second highest posterior probability was associated with species identified using mtDNA.

Shoal Composition

Overall, the reef was estimated to be composed of 12.0% – 20.1% (95% CI) *C. personatus* and 79.9% – 88.0% (95% CI) *C. hyalinus*. Only two shoals (out of 13) differed significantly in the proportion of the two species present compared to the site-level proportion (Fig. 2.6). One of these shoals had significantly more *C. personatus* than expected (45.5%, 95% CI: 18.7% - 73.8%; Difference from site 95% CI: 0.019 – 0.581) while the other had significantly fewer *C. personatus* (3.7%, 95% CI: 0.1% - 13.2%; Difference from site 95% CI: -0.202 – -0.0402).



Discussion

Coryphopterus personatus and *C. hyalinus* are sympatric sister-taxa that live in social aggregations throughout their reproductive lifespans. The general agreement between nuclear and mitochondrial identification (98.4%) and the paucity of admixed individuals (4.2%) suggests that the boundary between the two species is generally well maintained. However, gene flow has not ceased entirely as there was evidence for ongoing hybridization (two putative hybrids were observed) and historic introgression (admixture and disagreement between mitochondrial and nuclear loci). Contrary to the hypothesis that species boundaries are maintained by the formation of species-specific shoals, all shoals were composed of a mixture of *C. personatus* and *C. hyalinus*. Further, the proportions of the species in each shoal generally conformed to the proportion of these species on the entire reef, indicating that the species are not segregating due to social behavior. Additionally, there was disagreement between molecular and morphological methods in 12.1% of individuals and between the two molecular methods in 1.6% of the individuals, suggesting that gene flow between the species is not completely interrupted.

Mismatches in species identification between methods occurred but only for a small percentage of samples. Disagreements between morphological and molecular methods are likely attributable to one or two sources of error and showed a bias of morphologically misidentifying larger individuals as *C. personatus*. First morphological characters may be difficult to accurately assess following preservation (Kristoffersen and Salvanes 1998, Martinez et al. 2013). In *Coryphopterus*, anterior interorbital cephalic pores are difficult to see prior to ethanol preservation, which causes the pores to dilate, making them more readily visible (Baldwin et al. 2009). The preservation process could act asymmetrically, causing one pore to become more easily visible, appear to be more centrally located, and/or tear the tissue dividing pores, forming

what appears to be a single pore and lead to misidentification. Alternatively, there may be natural variation in the character state within species (one or two interorbital cephalic pores) leading to misidentification. Overlap in the distribution of meristic characters is a common problem in ichthyology (Hubbs 1922, Tåning 1952, McKay and Miller 1997) and variation in the number and arrangement of sensory pores within species is a common phenomenon (Ahnelt et al. 2004, Vanderpham et al. 2013, Ito et al. 2017). Further, because this character develops in larger individuals (> 10 mm SL), it may not be fully developed in some smaller individuals (Victor 2019). While morphological identification is possible and, in this study, seemed to be reasonably accurate (~87.9%), species identification using molecular methods is likely more reliable. By contrast, disagreements between mitochondrial and nuclear markers are likely the result of historic introgression or contemporary gene flow, including hybridization (Toews and Brelsford 2012). The presence of recently admixed individuals suggests that barriers to reproduction may be incomplete. However, the loci available, while capable of distinguishing between species (Table 2.2), do not offer sufficient resolution to distinguish between hybrid categories.

Despite indications of low levels of recurrent gene flow and regular co-occurrence in social aggregations, the two species remain genetically distinct with an estimated pairwise F_{ST} (0.19) at ~90% of its maximum value (0.21). The benefits of aggregating in this system seem clear, since larger aggregations make predators less efficient and decrease the probability of any individual being depredated (Hamilton 1971, Landeau and Terborgh 1986). Small reef fishes are constantly under high predation risk and both *C. personatus* and *C. hyalinus* spend the entirety of their life with elevated predation risk due to their small maximum body sizes, likely heightened by their behavior of hovering above the reef structure (Goatley and Bellwood 2016). In heterospecific aggregations the individuals of the less numerous species may benefit from allying with the more

numerous species, if co-aggregation increases aggregation sizes beyond that which the minor species could achieve on its own (Parrish 1989, Gibson et al. 2002, Wood and Ackland 2007). The idea that these heterospecific shoals form in part to offer protection is further supported by the presence of a third phylogenetically distant species of similar size within *Coryphopterus* shoals, the arrow blenny (*Lucyablennius zingaro*; Greenfield 1972).

Spending a significant proportion of the reproductive life span interacting with congeners comes with the risk of a reduction in fitness caused by production of inviable or less viable hybrids (Dobzhansky 1940, Coyne 1974, Friberg et al. 2013). Hybrid inviability can be induced through epistatic gene interactions (Dobzhansky 1936, Goodnight 2000), often resulting in viable, fertile F1 hybrids which are unable to produce viable F2 hybrids, but can backcross with parental lineages (Stelkens et al. 2015). As species diverge the number of incompatibilities tends to increase, further reinforcing isolation (Bolnick and Near 2005). Hybrid inviability in turn can reinforce pre-mating reproductive isolation and prevent the formation of hybrid swarms (Liou and Price 1994, Sadedin and Littlejohn 2003). For example, species boundaries are maintained between sympatric darters (family: Percidae) because of epistatic incompatibilities leading to elevated mortality in backcrossed individuals (Moran et al. 2019). Reproduction is energetically costly and as such the production of inviable hybrids represents a disproportionately large energetic cost (Dobzhansky 1940, Wootton 1985) that would be borne more heavily in species with short reproductive lifespans, such as *C. personatus* and *C. hyalinus*. Selection might quickly cause pre-mating isolation to develop when the probability of wasted energy via hybridization is high (Ortiz-Barrientos et al. 2004). Consistent with this idea, contemporary geneflow between the two species of *Coryphopterus* appears to be minimal, despite opportunity for frequent hybridization, suggesting that another mechanism may be maintaining species boundaries.

Unidirectional hybridization is a commonly observed pattern and occurs for a variety of reasons (Wirtz 1999). Generally, when females are the choosy sex, hybridization occurs between females of the rare species and males of the common species (Wirtz 1999). However, in this case when there was mito-nuclear discordance within individuals the maternal lineage was the more common species, *C. hyalinus* (Table 2.1). This observation could be explained by several different mechanisms. First, females of the less common species, *C. personatus* may be more discriminatory in choosing mates than female *C. hyalinus*. It could be selectively advantageous for the less common species to be more discerning due to the increased probability of heterospecific mating (Cooley 2007), especially if genetic incompatibilities have developed between the species leading to high fitness costs (i.e., less viable hybrid offspring) that outweigh potential costs associated with mate discrimination (Milinski and Bakker 1992, Wong and Jennions 2003). Under this mechanism locations where the relative abundances of the species are reversed, should result in female *C. hyalinus* being more discriminatory than *C. personatus*. The observed pattern could also result from female-biased sex ratios within shoals. Both *C. personatus* and *C. hyalinus* are protogynous (i.e. change sex from female to male) and research has demonstrated associated female skew in sex ratios (Cole and Robertson 1988, Cole and Shapiro 1990, Allsop and West 2004). This could result in larger dominant males that defend nesting sites being the choosy sex, rather than females, with the less common males (*C. personatus*) involved in more interspecies matings due to the relative lack of intraspecies females (Thresher 1984, Kramer et al. 2009).

While mate choice may be an isolating mechanism these data do not preclude the possibility of ultra-fine scale spatial segregation or asynchronous reproduction as alternative plausible mechanisms. Sympatric species of triplefin blennies (Family Tripterygiidae) in New Zealand,

utilize distinct nesting micro-habitats leading to reduced reproductive encounters between species (Wellenreuther and Clements 2007). However, triplefin nests are spread over the expanse of rocky reefs (Feary and Clements 2006). *Coryphopterus personatus/hyalinus* are thought to nest in reef crevices in the immediate vicinity of the shoal, suggesting that any differentiation in nest site preference between the species is occurring on at a much finer spatial scale.

Asynchronous reproduction, a common mechanism of reproductive isolation seen across taxa (Aspinwall 1974, Palumbi 1994), does not seem to be a likely for *C. personatus* and *C. hyalinus*, because they only live ~100 days post-settlement (Beeken et al. 2021) and congeners reproduce continuously through the lunar cycle (Kramer et al. 2009).

Despite ample opportunity and evidence of ongoing/past hybridization, *C. hyalinus* and *C. personatus* remain distinct but the exact mechanism(s) keeping them distinct remain unclear. Whatever mechanism(s) are at work, they likely occur at within-shoal scales and are mechanisms not often explored as first principles when seeking to explain the maintenance of species boundaries. Breeding site selection within shoals and/or cryptic mate recognition may play a large role in reducing gene flow between the species, as may be the case in other social species that co-aggregate with closely related taxa.

Acknowledgements

This research would not have been possible without the field assistance of Emily Anderson and Alan Downey-Wall. Field logistics were provided by the University of Belize Calabash Caye Field Station. We would also like to thank Sharon Magnuson and Christopher Bird for their helpful comments and suggestions on the manuscript. We would like to thank funding sources: the Carl R. Beaver Memorial Scholarship, the Karen Koester Dodson Memorial Fund Grant, and Grants-in-Aid of Graduate Student Research Award by Texas Sea Grant College Program

awarded to JDS, the Texas Research Development Fund award to JDH and funds provided by TAMUCC to JDH.

CHAPTER III: Photogrammetry derived habitat model enables characterization of fine-scale habitat use in a pair of coral reef fishes

Abstract

Coral reefs are composed of diverse microhabitats of varying quality, which support a diverse assemblage of species. Identifying which microhabitats are utilized by which species and their relative importance is critical to understanding the evolution and ecology of reef species. Here we develop a method for assessing the fine-scale habitat usage of cryptobenthic reef fishes by studying comingled shoals of *Coryphopterus hyalinus* and *Coryphopterus personatus* as an informative case-study. Structure-from-motion photogrammetry was used to develop high precision three-dimensional bathymetric models of the reef to measure fine-scale habitat characteristics (1 cm). A log-Gaussian Cox process spatial model was used to develop an interpretable model of the distribution of *C. hyalinus/personatus* and how the bathymetric features influence this distribution. The most important variables explaining the distribution of *C. hyalinus/personatus* on the reef were depth and distance from the nearest sand-reef ecotone, with a higher density of individuals found in deeper parts the reef and on portions closest to a sand-reef margin. The methods developed here can be used to render a continuous measure of habitat quality/importance that can be applied to eco-evolutionary studies of species-habitat relationships.

Introduction

Understanding the relationship between habitat and the associated fauna has been a foundational focus of ecological and evolutionary study (e.g. ideal free distribution, Fretwell and Lucas 1969). Habitat quality can affect fitness of individuals and the abundance of populations (Tregenza 1995). Moreover, the variability in and distribution of habitat quality influences

metapopulation dynamics, affecting gene flow across a landscape (Moilanen and Hanski 1998). Coral reefs are one of the most diverse and structurally complex ecosystems on the planet and provide habitat for a diverse assemblage of fishes that exhibit varying degrees of habitat specialization. Specialization is thought to drive species diversification, and reef fish biodiversity has been linked to reef complexity and coral cover (Schluter 2001, Losos 2010, Komyakova et al. 2013). Specialization is not merely a binary category but a continuum on which a species may exhibit some degree of specialization ranging between both extremes (Dapporto and Dennis 2013).

The scale at which habitat quality and variation are observed and measured is important for the accurate understanding of the species-habitat relationships (Levin 1992) And can influence our perception of whether a species is a generalist or specialist (Ainsworth and Drake 2020). For example, an adult of a given species is likely to perceive potential habitat much differently than a juvenile of that same species (Wilson et al. 2008a). Moreover, scales of human perception cannot be assumed to be relevant to the diverse array of organisms inhabiting the reef environment (Sale 1998). Observing the relationship between a species and its environment at an overly coarse spatial scale can lead to categorization of a species into a generalist or specialist rather than a more nuanced understanding of where a species falls on the generalist specialist spectrum (Ainsworth and Drake 2020). As such, it is important to perform analyses of habitat usage at a scale relevant to the organism in question.

Often overlooked are inconspicuous or imperceptible nano-scale relationships between species and habitat (Anderson 2007). For example, if a species spends the entirety of its life in an area no larger than 1 m², then differences at the scale of the entire reef are unlikely to directly influence growth or fitness, but rather act as an indirect influence through its effect on the habitat

at a small scale. Historically, methods of studying reef fish-habitat relationships have been limited to meso- or micro-habitat scales due to technical and logistical constraints. Beyond categorizing faunal communities utilizing different reef zones (e.g., fore-reef vs back-reef and lagoonal vs reef crest), one widely used approach consists of characterizing a reef site according to habitat metrics of interest (e.g., live coral cover, rugosity, habitat classes etc.) and site level differences in the means of these metrics are analyzed in relation to the site level variable of interest (e.g., species richness, abundance of a particular species, etc.; Greenfield and Johnson 1999, Burt et al. 2009, Biggs and Olden 2011, Komyakova et al. 2013). A second approach is to measure various habitat metrics in the area where a species of interest is found and compare that with similar, equally sized areas where the species of interest is absent, to identify differences between occupied and unoccupied habitats (Appeldoorn et al. 2003). While both methods provide insight into the link between habitat composition and habitat usage of specific species on coral reefs, they are unable to examine nano-scale habitat relationships. This is an important knowledge-gap because nano-scale is the most appropriate scale of inference for crypto-benthic reef fishes, and cryptobenthic fishes make up the plurality of reef fish diversity and majority of reef fish abundance (Brandl et al. 2018).

With advances in habitat mapping technology, we are now able to investigate smaller spatial scales at finer resolution (1 cm) across relatively large areas, which is more appropriate when studying nano-scale habitat usage. Recent technological advances in Structure-from-Motion (SfM) photogrammetry has led to high-resolution 3D topographic maps of coral reefs which are both relatively inexpensive and logistically feasible (Burns et al. 2015). This technology involves stitching together many photographs from above the substratum using uniquely identifiable landmarks to create a high-resolution 3D model of the landscape (Westoby et al. 2012). From

these models, it is possible to measure ultra-fine scale topographic metrics continuously across the reefscape, including traditional metrics (e.g., rugosity) as well as metrics previously impossible to characterize on reefs for logistical reasons (e.g. viewshed, the area of reef visible from a given location, Burns et al. 2015, Young et al. 2017, Oakley-Cogan et al. 2020). This technological advancement changes the scope and scale of questions which can be asked, particularly in relation to how species use habitat, and has been increasingly embraced by the scientific community (e.g., González-Rivero et al. 2017, Oakley-Cogan et al. 2020, Bongaerts et al. 2021). Here we use a case-study to explore the use of SfM in characterizing the relationship between nano-scale topography and the distribution of two sympatric cryptobenthic reef fishes on Caribbean fishes.

Coryphopterus personatus and *C. hyalinus* are reef-associated, planktivorous, shoal-forming, cryptobenthic gobies, common on shallow Caribbean coral reefs (Böhlke and Robins 1962, Greenfield and Johnson 1999). These species are often considered habitat generalists, living in shoals distributed across a large depth distribution with only weak micro-habitat associations compared to specialist species like *Elacatinus lori* which live in a single species of sponge (Greenfield and Johnson 1999, Garcia-Sais 2010, Lesneski et al. 2019). Mixed shoals of *C. personatus/hyalinus* are composed of tens to thousands of individuals which exist as temporally stable (pers. obs.), spatially distinct groups hovering slightly above the reef structure (Robertson and Justines 1982, Selwyn et al. 2022). Due to their small body (~48 mm SL; Beeken et al. 2021) and their abundance on reefs, these species are commonly consumed by a wide array of piscivorous fishes (Randall 1967, Opitz 1996). When exposed to a predator, shoals will either disappear into the reef structure or temporarily move and potentially blend with adjacent shoals before reforming after the predator has left (pers. obs.).

Given the life history and ecology of these species, we can form some general hypotheses about how these species likely use reef structure. Typically, on coral reefs, planktonic food resources are more available at the margins of the reef (Glynn 1973, Hamner et al. 1988). As such, we hypothesize *C. personatus/hyalinus* to be more concentrated at the sand/reef ecotone where there is a greater supply of plankton (Hamner et al. 1988). Further, as *C. personatus/hyalinus* utilize complex reef structure as both a refuge from predators and a nesting location, we hypothesize that their densities will be more concentrated in areas of greater habitat complexity with a greater selection of shelter and nest sites (Thresher 1984, Cole and Robertson 1988, Kramer et al. 2009). Finally, if *C. personatus/hyalinus* avoid predation primarily by avoiding predator detection, they will use areas which are less visible to the surrounding reef (low viewshed; Sansom et al. 2009). Alternatively, if they avoid predation by group vigilance and secondarily hiding within the reef structure, we hypothesize they will prefer areas that provide greater visibility of the reef around them (high viewshed; Sansom et al. 2009).

Methods

Population and habitat surveys

Surveys were performed at twelve sites (approximately 10 m x 20 m) along the windward face of Turneffe atoll (17.3638° N, 87.8581° W), Belize, Central America in January 2017 (Fig. 3.1). Sites were located along the forereef at 15-20 m depth and named A-L from north to south. Each site was surveyed by three divers. One diver estimated the number of individuals in each shoal, which are composed of both *C. personatus/hyalinus* at this depth range (Selwyn et al. 2022), and marked the location of each shoal with fluorescent flagging tape. The other two divers placed a 93.9 cm PVC pipe and measured the depth, distance, and angle of up to ten, evenly distributed, small rectangular ground control points to an anchor location, generally in the

northeasterly corner, for which a GPS coordinate was taken. Latitude and longitude of each of these ground control points were calculated, using the anchor location as a reference and a compass and transect tape, and used to geo-reference the resulting habitat models (see below). After geo-referencing and fish surveys were complete, a single diver took between 150 and 400 photographs using a Canon PowerShot S110 camera (Canon Inc, Tokyo, Japan) in an Ikelite housing (Ikelite Underwater Systems, Indianapolis, Indiana, USA) from roughly four meters above the substratum following the methods described by (Burns et al. 2015).

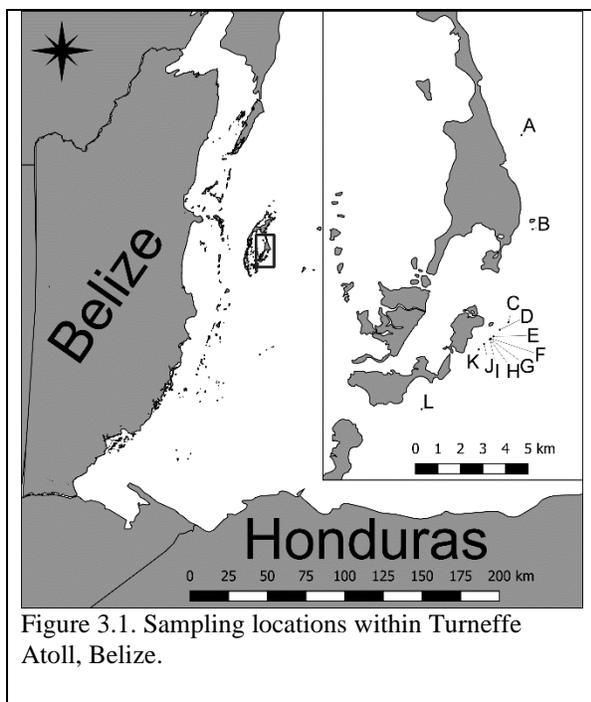


Photo processing and photogrammetry

All images were downloaded onto a laptop computer and digitally corrected for white balance before being aligned. The photogrammetry workflow to create the digital models included photo alignment, followed by geometry building, and lastly texture building. Additionally, the site point clouds were used to construct both an orthomosaic, a photographic reconstruction of the complete site (Fig. 3.2), and a digital elevation model (DEM, Fig. 3.3) of

each site. The orthomosaic and DEM were then used to determine reef cover (by measuring reef RGB color values) and metrics derived from depth (e.g., rugosity) respectively (see “*Habitat classification*” in the supplement and “*Topographic Measurement*” below). For each site, the root-mean squared error (RMSE) of the digital elevation model was measured to assess the quality of the 3D reconstruction of the reef, based upon the known length of the PVC pipe and the length and widths of the ground control points compared to the length/width measured from the DEM model (Table 3.1). Photogrammetry was conducted using the software Agisoft Photoscan Pro (Agisoft LLC., St. Petersburg, Russia).

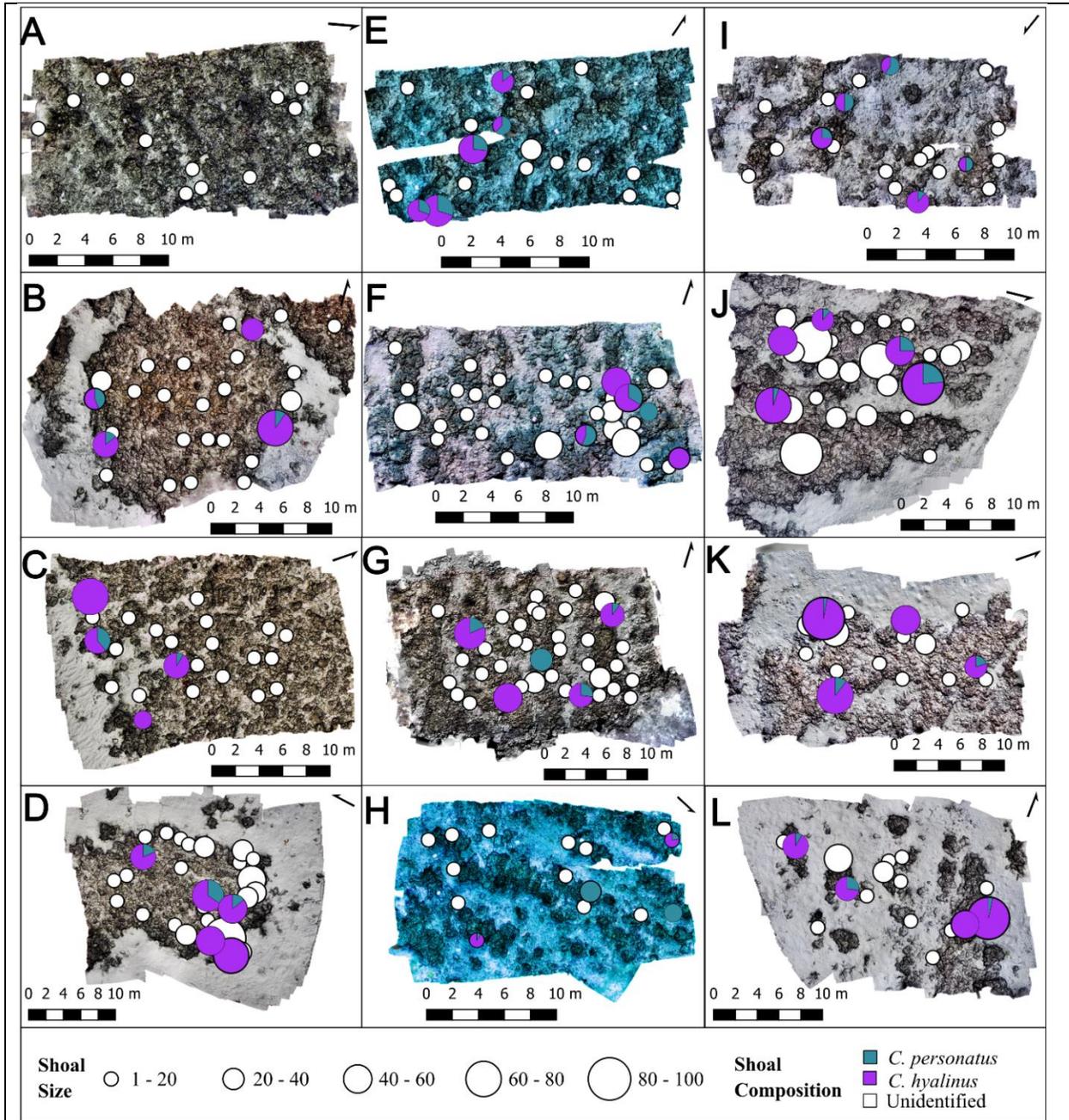


Figure 3.2. Panels showing site orthomosaics overlaid with *Coryphopterus hyalinus* and *C. personatus* shoals showing shoal size (size of point) and shoal composition (pie charts). Empty circles mark unsampled shoals with unknown composition of *C. hyalinus* and *C. personatus*.

Topographic Measurement

A variety of components of the topography of the reef habitat were calculated from the digital elevation model (DEM): depth (m), relief (cm), rugosity, vector dispersion, Moran's I

(Moran 1950), slope ($^{\circ}$), viewshed (%), and distance to the sand/reef boundary (m). These topographical measures were selected to represent major aspects of the local topography in ways that are relevant to fish living there (Wilson et al. 2008a, Pittman et al. 2009). Relief, rugosity, slope, and vector dispersion all measure different aspects of the fine-scale structural complexity of the reef and as such were combined into a single metric using principal components analysis from which the first principal component was retained. While the previous group of variables measure fine-scale structural complexity, Moran's I, generally a measure of spatial autocorrelation, can be recontextualized as a coarse-scale measure of structural complexity. For ease of interpretation, Moran's I was multiplied by -1 so that larger values indicate more structurally complex regions while low values are areas of reduced complexity. All metrics of structural complexity were calculated using the RASTER and TERRA R packages with the PCA calculated using the RSTOOLBOX package (Leutner et al. 2019, Hijmans et al. 2021b, 2021a).

Viewshed is the geographical area visible from a specific location, excluding areas that are beyond a visual horizon or obstructed by an object. Viewshed has been historically used to determine ideal placement of human structures, for example, fire watch towers, and in our implementation measures the proportion of the rest of the site visible to a fish hovering one meter above the substratum. In terms of potential prey species like *C. personatus*, viewshed is indicative of how much of their surroundings can be observed visually and how exposed they are to visual detection by predators in the water column. Viewshed analysis was conducted using the WHITEBOXTOOLS R package (Lindsay 2019, Qiusheng Wu 2020).

Edges between different habitat regimes (i.e., forest/field boundaries) are often some of the most productive and diverse parts of any given ecosystem (Harris 1988). To determine if *Coryphopterus* make use of the ecotone between sand and reef habitat regimes within a site, we

calculated the distance (m) from every point on the reef to the nearest patch of sand with an area $> 0.25 \text{ m}^2$ or the largest sand patch on the site if none were larger than 0.25 m^2 . This was done to reduce the influence of habitat misclassification of individual, or small clusters of pixels.

Furthermore, extremely small patches of sand within the larger reef structure likely do not act as an ecotone between two distinct habitat regimes. This cutoff retained 83% of the sand area across all sites. Distance to the sand/reef ecotone was measured using custom code using the TERRA and SF R packages (Hijmans et al. 2021a, Pebesma et al. 2021).

Distribution modelling

To understand how topography influences the distribution of *Coryphopterus* on the reef we used a log-Gaussian Cox process model (LGCP, Møller et al. 1998). The LGCP models the location of individual fish as a point process influenced by the measured habitat metrics (see above) as well as a component of spatial autocorrelation (Beguin et al. 2012). To fit this model the number of fish observed in each shoal were first randomly distributed around the shoal centroid using a multivariate normal distribution with the shoal diameter being estimated as ~ 25 cm (pers. obs.). The coordinates of the fish were used as the dependent variable in the point process to estimate the number of *Coryphopterus* per m^2 as a result of a spatial autocorrelation using a Matérn covariance function and influenced by the measured habitat variables (Beguin et al. 2012). Prior to inclusion in the model, habitat, the distance to the sand/reef boundary, and Moran's I were log plus one transformed and all metrics were z-score transformed to simplify prior settings and interpretation. Habitat metrics were included as linear additive effects influencing the density of *Coryphopterus* and their effect was modelled using weakly regularizing normal priors ($N(0, 10)$). In addition to the continuous habitat metrics, we included

the categorical effect of habitat type (reef vs. sand, uninformative prior: $\text{Loggamma}(1, 0.00005)$) and an effect of site identity as hierarchical effects (weakly informative prior: $\text{Gamma}(2, 1)$).

Due to the randomized distribution of fish around the shoal centroid, we fit this model 1,000 times with different realizations of locations of each specific fish and then merged the posterior distributions to a single joint posterior integrating across the specific randomized positions of the fish. To confirm the model appropriately represents the observed data, we performed posterior predictive checks which consisted of sampling the joint posterior distribution 1,000 times and determining if both the total number of observed fish and the observed number of fish per site are generally well represented in the posterior distribution (Gelman et al. 2020). Additionally, to determine if the model is accurately representing second order effects of the point process (i.e., distribution of distances between fish) we sampled 100 posterior point processes from the joint posterior and calculated Ripley's K, a measure of the observed pairwise distances between points, and compared the posterior distribution of K with the distribution in the observed point process calculated using the SPATSTAT R package (Ripley 1977, Baddeley et al. 2021). Finally, we calculated the credible intervals and posterior probability of the directionality of each topographic parameter included in the model. The log-Gaussian Cox process model was fit using the INLABRU interface to INLA (Rue et al. 2009, Bachl et al. 2019).

Shoal composition modelling

In addition to creating digital reef models, a total of 729 individual fish were collected for species identification from between 4 and 7 haphazardly chosen shoals from each site (except site A). Specimens were genetically identified, see supplement for details, and we employed a logistic regression model to identify relationships between reef topography and the composition of shoals. We modelled the effects of all measured habitat metrics and the size of the shoal on

the percentage of *C. personatus* found in the sampled shoals using weakly regularizing normal priors ($N(0, 10)$). Additionally, we included hierarchical effects of site and shoal nested within site to account for both among site and among shoal variation in shoal composition unrelated to habitat metrics. To model the hierarchical effects, we used gamma priors ($G(2, 2)$). This model was created using the R package BRMS (Bürkner 2018) and fit using an HMC sampler in STAN (Carpenter et al. 2017). The model was run for 2,000 iterations which includes 1,000 warmup iterations on four independent chains. Mixing of chains and proper exploration of parameter space was confirmed by visually inspecting trace plots and ensuring all \hat{R} values equal one (Vehtari et al. 2019). After fitting the model, we confirmed that there were no divergent transitions and that the model reasonably represents the observed data with a posterior predictive check (Gelman et al. 2020). Finally, we calculated the r^2 (Gelman et al. 2019) and the posterior probabilities of the direction of the topographical effects.

Habitat Quality Metric Example

As an example of how these methods can be used to produce a habitat quality metric which can be applied to new sites to predict the number/location of *C. hyalinus* and *C. personatus* we develop this metric for the present study sites. The habitat quality can be calculated as the linear prediction of the estimated number of fish in an area scaled to be between 0 and 1. This measure is inherently continuous and spread across the reefscape in the same way as the topography it is measuring. Depending on the specific research question researchers can summarize the habitat quality across sites, or regions within sites, to have a measure of the quality of the specific area. The benefit of this technique is that it integrates the relevant habitat features into a single measure of habitat quality from the perspective of the species of interest. In this example we calculated the average and standard deviation of habitat quality across a site and

used both metrics, and their interaction to examine how both effect the density of fish present on the site using a Bayesian generalized linear model with a Poisson likelihood function and a hierarchical effect of site to account for overdispersion. The model was run for 2,000 iterations which includes 1,000 warmup iterations on four independent chains. Model convergence and fit was assessed as above.

All analyses were performed using R v 3.5.1 (R Core Team 2018) and the packages in the TIDYVERSE suite of packages (Wickham et al. 2019). The code used in this analysis can be found at: github.com/jdselwyn/Habitat_Usage and all habitat data can be found in doi:10.5281/zenodo.5348484 (Selwyn et al. 2021) and molecular data can be accessed with GenBank accession numbers SAMN23384469 – SAMN23385266.

Results

Site & Shoal Description

Coryphopterus hyalinus and *C. personatus* were found at all sites, living singly or in shoals of up to ~100 individuals. Shoals were characterized by an average of 19.2 fish (± 1.24 SE), a median of 10 fish, and a mode of 5 fish. The average shoal size varied considerably across sites (generalized linear regression model, negative binomial with log link function, likelihood ratio test $\chi^2_{(11)} = 94.56$, $p \ll 0.0001$; Table 1) with the largest mean shoal size (37.8 ± 5.39 SE) found at site J and the smallest at site A (3.38 ± 0.53 SE, Fig. 3.2). All sites contained a mix of both *C. hyalinus* and *C. personatus*, with an overall composition of 81.5% *C. hyalinus* (78.6 – 84.3% 95% CI), although the specific composition varied by site (Table 3.1, Fig. 3.2). Further, in concordance with previous findings (Selwyn et al. 2022), species compositions of the shoals were found to match background frequencies of each species found on the reefs with 42 out of the 50 sampled shoals showing no difference in composition from the background frequency

(Fig. 3.2, Fig. 3.S1). Of the eight shoals with different compositions than the background reef, five were composed exclusively of a single species, the other three were composed of a mix of both species different from the background composition. All but one of the shoals containing a single species had fewer than 10 fish collected, suggesting a likely sampling error.

Structure-from-Motion model statistics

The DEMs were highly precise with an average RMSE of 1.9 cm (± 0.28 SE). Precision varied among sites; however, all sites had a RMSE less than 4 cm (Table 3.1). Sampling sites were between 16 and 23 m at their deepest points (18.3 ± 0.584 m SE) with overall relief (depth difference between shallowest and deepest parts) of the sites ranging from 2.9 to 5.9 m (4.37 ± 0.31 m SE; Table 3.1). The final mapped and surveyed cross-sectional areas of the sites ranged from 180 to 581 m², with an average area of 345.96 m² (± 34.49 SE). The average site density of *Coryphopterus hyalinus/personatus* was 1.22 fish/m² (± 0.178 SE), although densities ranged among sites by an order of magnitude from 2.29 fish/m² (Site F) to 0.175 fish/m² (Site A). All sampled sites were composed of a mixture of reef and sand, with most sites being predominantly reef (mean percent reef = 70.7% $\pm 4.53\%$ SE, Table 3.1, see supplement for details).

Topography

Habitat metrics varied considerably both among and within surveyed sites (Table 3.1, Figs. 3.S2 – 3.S13). The average depth across sites was 16.1 m (± 0.45 SE) with an average standard deviation within sites of 0.87 m (± 0.09 SE). The average distance to the nearest 0.25 m² sand patch was 2.17 m (± 0.87 SE), which varied within sites by 1.62 m (± 0.46 SE). In general, from a height of 1 m above the substratum, most of the sites are visible and unobstructed (average viewshed = 63.8% ± 2.7 SE) with an average within site standard deviation of 2.8% (± 0.4 SE). The average Moran's I (course scale complexity) was consistent across sites (0.999 ± 0.0001 SE)

as depth was highly spatially autocorrelated, however this masks within site variability in coarse complexity (mean standard deviation = 1.21 ± 0.07 SE). The metrics which were used in PCA to create a fine-scale complexity measure generally showed a high degree of within site variability. The average relief within a 3x3 cm window was 3.66 cm (± 0.298 SE) with considerable variability within sites (mean standard deviation = 7.10 cm ± 0.53 SE). Similarly, rugosity averaged 2.04 (± 0.11 SE) across sites and varied substantially within sites (1.87 ± 0.15 SE). Across sites the average slope in a 3x3 cm window was 35.6° (± 1.8 SE) and varied substantially within sites (average standard deviation = $22.1^\circ \pm 0.35$ SE). The final metric used in the creation of the fine-scale complexity measure was vector dispersion which averaged 0.145 across sites (± 0.012 SE) and varied substantially within sites (mean standard deviation 0.118 ± 0.005 SE). The primary axis of the PCA of the relief, rugosity, slope, and vector dispersion explained 70.06% of the variance in these four metrics with all four metrics loading positively, indicating increasing complexity (Figure 3.S14). The second axis of the PCA discriminates relief and rugosity from slope and vector dispersion (explaining 13.62% of the overall variance). Only the first complexity axis was retained for subsequent analyses.

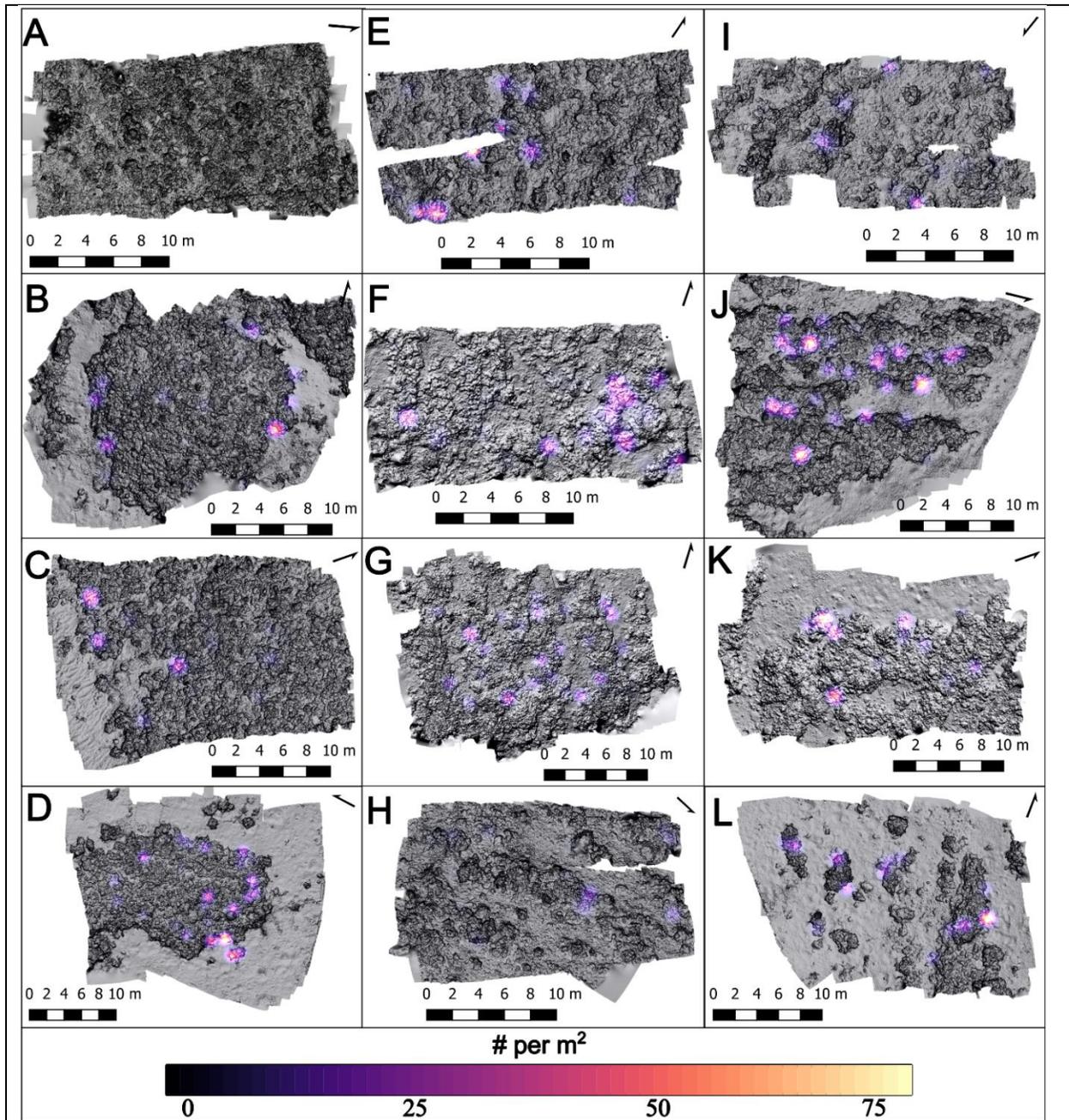


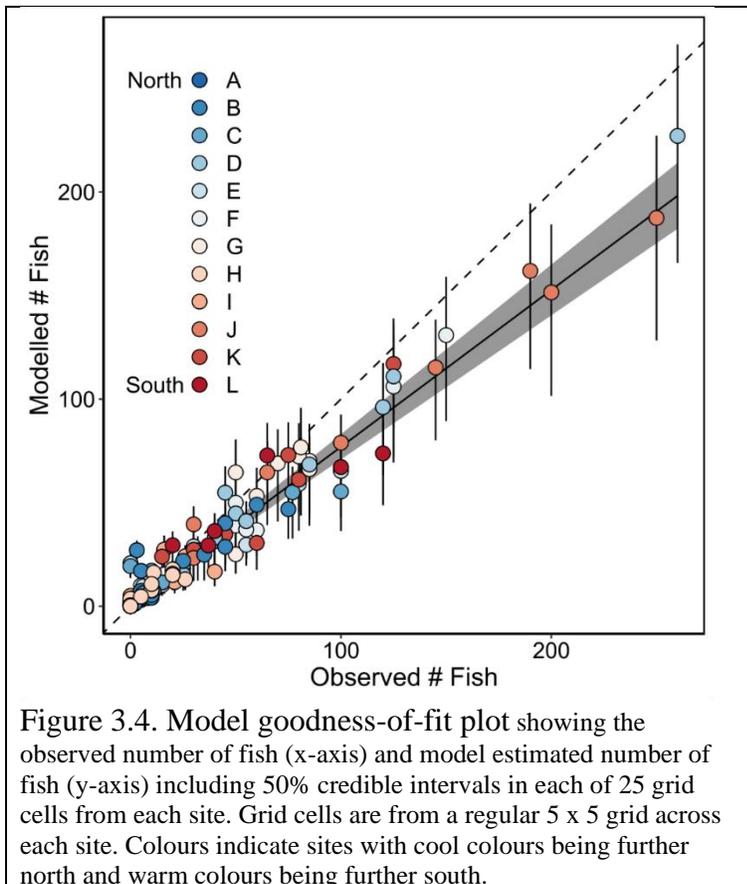
Figure 3.3. Panels showing the mean estimated number of *C. hyalinus* and *C. personatus* m^{-2} from the model joint posterior is shown overlaid on the digital elevation model. Warmer colours show locations with larger predicted average numbers of *C. hyalinus* and *C. personatus* m^{-2} while grey areas show locations with few predicted individuals m^{-2} .

Table 3.1. Site level summary statistics (arranged north-south) including the number of *Coryphopterus* on the site (N), summaries of shoal size and composition, photogrammetry precision, overarching site topography and habitat composition, and moving window topography metrics and standard deviations within sites. Site composition of *C. hyalinus* and *C. personatus* are estimated based on the number of each species sampled from each site using a binomial likelihood and an uninformative ($\beta(1, 1)$) beta conjugate prior.

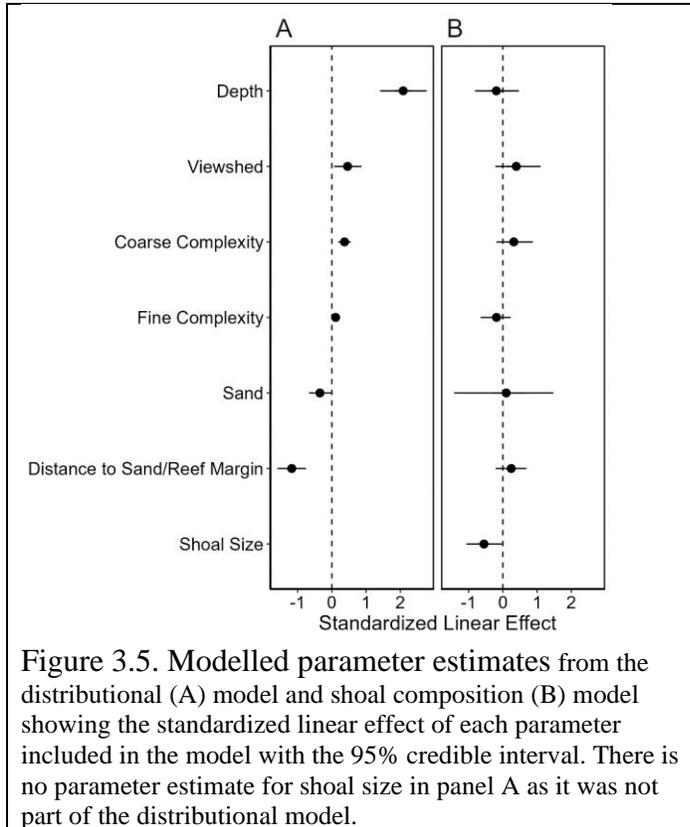
				Site Measures					Moving Window Summaries								
Site	Shoal Size	<i>C. hyalinus</i> (% 95CI)	RMSE (cm)	Depth (m)	Relief (m)	Area (m ²)	Density (#/m ²)	Reef (%)	Depth (m)	Sand Distance (m)	Viewshed (%)	Coarse Complexity	Relief (cm)	Rugosity	Slope (°)	Vector Dispersion	Fine-Scale Complexity
A	3.4 ± 0.5		1.04	17.1	4.8	252	0.17	95.5	14.9 ± 0.8	11.3 ± 6.0	45.9 ± 3.3	1.0 ± 1.2	5.5 ± 9.1	2.7 ± 2.6	45.9 ± 22.3	0.23 ± 0.14	0.53 ± 1.12
B	13.0 ± 3.2	80 (70 - 89)	1.53	17.1	5.9	400	0.84	71.4	14.1 ± 1.4	1.9 ± 1.8	66.8 ± 1.3	1.0 ± 0.9	4.2 ± 8.4	2.2 ± 2.3	38.0 ± 23.2	0.16 ± 0.13	0.13 ± 1.07
C	13.3 ± 3.6	83 (73 - 92)	1.14	22.8	5.8	369	0.87	77.7	19.5 ± 1.4	2.7 ± 3.0	59.6 ± 2.3	1.0 ± 1.0	3.6 ± 5.6	2.0 ± 1.4	37.9 ± 22.0	0.15 ± 0.11	0.05 ± 0.80
D	29.8 ± 4.3	82 (74 - 90)	2.67	20.1	4.7	581	1.39	47.9	17.9 ± 1.1	1.5 ± 1.4	64.7 ± 2.0	1.0 ± 0.9	2.6 ± 6.2	1.7 ± 1.6	27.3 ± 22.6	0.10 ± 0.12	-0.26 ± 0.90
E	16.3 ± 4.1	69 (57 - 80)	1.41	16.9	2.9	194	1.60	90.5	15.4 ± 0.5	3.6 ± 2.2	75.5 ± 3.4	1.0 ± 1.2	3.8 ± 6.2	2.1 ± 1.6	39.5 ± 21.3	0.17 ± 0.11	0.15 ± 0.85
F	18.4 ± 2.9	61 (45 - 77)	3.39	17.8	3.3	249	2.29	83.8	16.2 ± 0.6	0.9 ± 0.7	66.7 ± 2.3	1.0 ± 1.3	3.6 ± 6.3	2.0 ± 1.6	37.8 ± 21.9	0.14 ± 0.11	0.04 ± 0.84
G	16.6 ± 2.1	80 (70 - 89)	1.27	20.0	5.5	411	1.54	69.1	16.9 ± 1.1	0.9 ± 1.1	63.1 ± 1.6	1.0 ± 1.5	5.5 ± 11.0	2.8 ± 2.9	41.0 ± 23.9	0.19 ± 0.15	0.38 ± 1.30
H	9.5 ± 1.8	70 (55 - 83)	2.83	16.8	4.2	270	0.53	65.9	14.7 ± 0.7	0.4 ± 0.4	73.0 ± 5.5	1.0 ± 1.3	3.3 ± 6.0	1.9 ± 1.6	35.6 ± 21.5	0.14 ± 0.11	-0.02 ± 0.83
I	9.1 ± 1.9	62 (48 - 74)	3.58	16.0	3.2	181	1.00	68.5	14.6 ± 0.6	0.4 ± 0.4	76.9 ± 5.1	1.0 ± 1.2	2.6 ± 4.5	1.6 ± 1.2	33.6 ± 19.7	0.12 ± 0.10	-0.15 ± 0.68
J	37.8 ± 5.4	87 (79 - 92)	1.51	20.2	5.0	473	2.16	71.7	17.3 ± 0.9	0.9 ± 0.8	49.9 ± 1.8	1.0 ± 1.8	3.5 ± 7.4	2.0 ± 2.0	34.8 ± 22.2	0.13 ± 0.11	-0.05 ± 0.94
K	26.0 ± 6.9	92 (87 - 96)	1.57	17.1	3.6	379	1.17	65.6	15.9 ± 0.7	1.2 ± 1.2	59.3 ± 2.6	1.0 ± 1.0	3.6 ± 8.6	2.1 ± 2.3	33.5 ± 23.5	0.13 ± 0.12	-0.04 ± 1.06
L	27.6 ± 5.9	90 (81 - 96)	0.70	17.6	3.4	395	1.12	40.5	16.3 ± 0.5	0.4 ± 0.4	63.8 ± 2.6	1.0 ± 1.3	2.1 ± 5.9	1.5 ± 1.6	22.6 ± 20.5	0.07 ± 0.11	-0.45 ± 0.84

Distributional Model

The joint posterior prediction contains the observed density of fish in each site within the 95% credible interval (Fig. 3.S15). Furthermore, the joint model performed well modelling distances between pairs of *Coryphopterus* (Fig. 3.S16). The model underfits nearby individuals and therefore slightly underestimates the number of individuals within a single shoal (Fig. 3.S16), resulting in a slight underestimation of the total number of fish in subregions of the sites (Monte Carlo χ^2 goodness-of-fit = 421.16, $p = 0.026$, Fig. 3.3, Fig. 3.4, Hope 1968). The mean range of the Matérn spatial correlation function was 8.19 m (95% credible interval: 6.55 – 9.9) meaning that the range of spatial dependence among fish habitats was ~6.5 – 10 m with fish located further apart being seen as practically uncorrelated (Beguin et al. 2012).



The effects of depth and distance from the sand/reef margin had the strongest influence on the distribution of *Coryphopterus* (Fig. 3.5A). Across average habitat metrics, the estimated density of *Coryphopterus* was 7.1 - 417.2 / km² (95% credible interval) reflecting the extreme patchiness of their distribution and restriction of living within large shoals. Depth had the strongest effect on *Coryphopterus* density with an increase of one meter in depth leading to a density increase of 555 – 2,679% (95% CI, posterior probability ~ 1.00) within the depth range of the site boundaries. The effects of increasing one unit of viewshed (0.57 – 11.7% / 1% viewshed), fine complexity (0.814 – 23.4% / fine complexity unit), and coarse complexity (25.0 – 91.5% / coarse complexity unit) all led to smaller but generally positive increases in the density of aggregations (posterior probabilities of positive effect = 0.989, 0.982, and ~1.00 respectively, evidence ratios for positive effect = 86, 54, and 38 thousand respectively). Finally, shoals were 171 – 258% denser (# / m² / m) closer to the sand/reef margin (posterior probability ~ 1.00) and 1.5 – 48.6% less dense over sand areas than reef areas (posterior probability of negative effect: 0.990, evidence ratio 103).



Shoal composition model

We found that while shoals were composed according to the background species frequencies (above), there were subtle but important differences in shoal composition based on local topography ($r^2 = 0.76 \pm 0.060$ SE, Fig. 3.5B). Specifically, we found a strong negative relationship between the size of the shoal and the proportion of *C. personatus*, with larger shoals having smaller fractions of *C. personatus* (posterior probability = 0.98, evidence ratio = 43.94). We also found substantial evidence that shoals are composed of greater proportions of *C. personatus* further from the sand/reef boundary (posterior probability = 0.86, evidence ratio = 6.31), in areas of greater visibility (i.e., viewshed; posterior probability = 0.89, evidence ratio = 8.07), and in areas with reduced fine-scale complexity (posterior probability = 0.80, evidence ratio = 4.03) but higher coarse-scale complexity (posterior probability = 0.88, evidence ratio =

7.30). These results qualitatively suggest differences in habitat use between the two species at this micro-scale mirroring macro-scale differences previously observed, with *C. hyalinus* being more concentrated on reef walls and edges, while *C. personatus* is more concentrated internally on smaller reef patches (Victor 2019).

Habitat Quality Metric Example

Mean site-scale habitat quality and variance, along with a hierarchical effect of site, does a good job explaining the density of fish present on the surveyed sites ($r^2 = 0.99 \pm 0.003$). Both increasing average quality and variation in quality result in sites having more *C.*

hyalinus/personatus (mean 95% CI: 5.39 – 16.34, posterior probability ~ 1.00 , evidence ratio = 284.71, Fig 3.6A; standard deviation 95% CI: 12.48 – 41.24, posterior probability = 0.99, evidence ratio – 180.82, Fig 3.6B). Further we found a negative interaction between mean habitat quality and the standard deviation in habitat quality indicating the effect of increases in one are tempered by the other measure (95% CI: -320.26 – 42.18, posterior probability = 0.90, evidence ratio – 9.20, Fig. 3.6C).

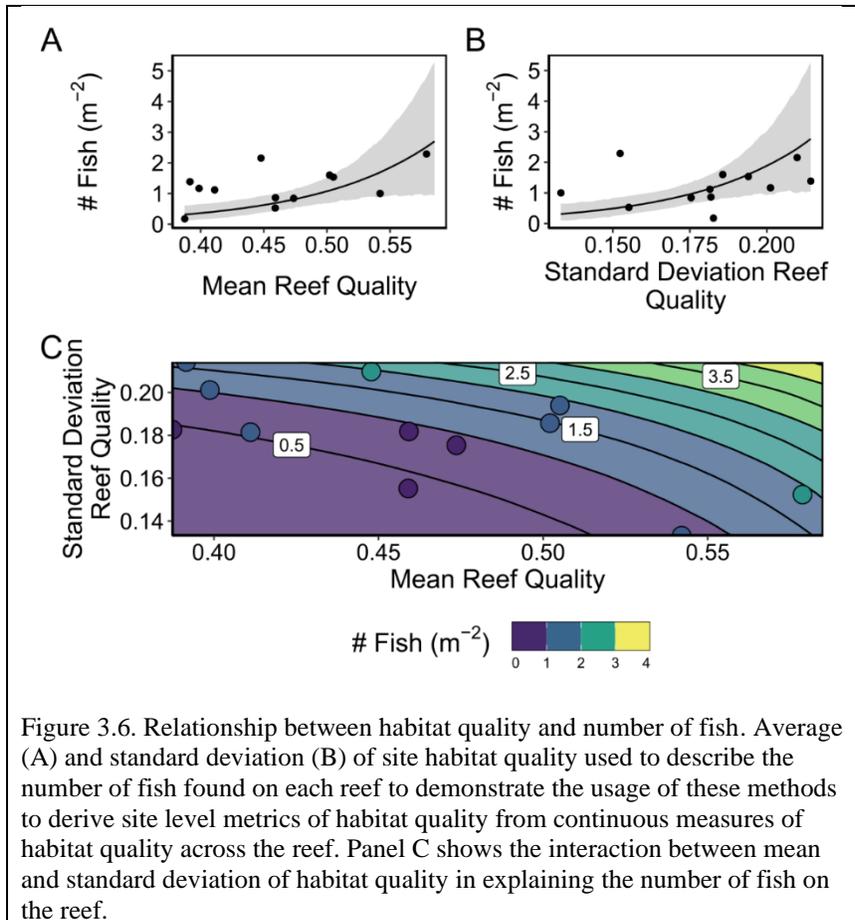


Figure 3.6. Relationship between habitat quality and number of fish. Average (A) and standard deviation (B) of site habitat quality used to describe the number of fish found on each reef to demonstrate the usage of these methods to derive site level metrics of habitat quality from continuous measures of habitat quality across the reef. Panel C shows the interaction between mean and standard deviation of habitat quality in explaining the number of fish on the reef.

Discussion

We found strong evidence that there are elements of reef topography, existing at meso- and micro-spatial scales, which led to greater densities of *Coryphopterus hyalinus* and *C. personatus* despite being widespread and often considered habitat generalists among reef fishes.

Specifically, increasing depth led to the greatest increase in fish density, followed by proximity to the sand/reef ecotone. The effect of increased depth may be related to the fact that travelling deeper within the reef increases proximity to the reef drop-off, which is present at all of our sites and is the preferred habitat of *C. hyalinus* (the more abundant species in this study; Victor 2019). Additionally, being closer to the reef drop-off will lead to an increase in available planktonic resources since there is less distance between the pelagic and reef ecosystems, reducing potential

competition for plankton resources from other planktivores (Hamner et al. 1988). Understanding the differing influences of various topographical metrics on the distributions of these species requires a holistic view of the reefscape at a scale relevant and appropriate to the existence of these species (Sale 1998).

These results support the hypothesis that there will be greater densities of *Coryphopterus hyalinus/personatus* at the sand/reef ecotone. We propose that living close to the sand/reef ecotone allows for greater densities of these species for two reasons. First, on the majority of these sites (ex. Site D, Fig. 3.1) the largest area of sand is facing the reef drop-off, indicating a greater availability of planktonic resources and allowing for an increased number of fish to be supported in a shoal (Hamner et al. 1988, Skinner et al. 2021). Second, being on the sand/reef margin allows easy access to both habitat types, which may beneficially allow increased access into the reef structure for fishes to hide in or use as nest sites.

Of the habitat features which were of less importance to describing the distribution of *C. hyalinus* and *C. personatus* on the reef, the most surprising is the effect of sand. That sand habitats are correlated with a reduction in the density of these fishes is itself unsurprising as they are generally reef associated. However, the relative effect of the habitat being sand (only ~half as influential as distance to the ecotone) is a surprise. This is likely an effect of the survey design where flagging tape used to mark a shoal was put onto the sand under the shoal, if possible, to ensure it was visible from multiple angles to the camera taking SfM imagery. The distribution of fish around each shoal centroid may then be slightly biased in placing more fish onto sand habitats than would be the case in reality. With more refinement of this survey methodology, the effect of exact placement of the flagging tape in marking shoals can be further elucidated and controlled for.

Of the other three habitat variables (viewshed, fine-, and coarse-complexity) included in this analysis, all had strong evidence to support a positive association with fish density. The positive association of both fine- and coarse-scale habitat complexity with greater densities of fishes seems to be based on the fact that these fish use the complex habitats for shelter from predators and reproduction (Almany 2004). That increased viewshed leads to an increase in the fish density suggests that despite being more visible to predators, this cost is outweighed by either the availability of food, the ability to perceive predators and thus avoid predation, or that sight-based pelagic predators (e.g., jacks, snappers, etc.) are less of a risk compared to more cryptic, benthic predators (e.g., groupers, scorpionfish, etc. Randall 1967, Opitz 1996, Roopnarine and Hertog 2013, Cirtwill and Eklöf 2018).

While this study has substantially less power to understand differences in the distributions of *C. personatus* and *C. hyalinus* compared to the distribution of both species in combination, we did find subtle differences in species contribution to each shoal that are attributable to topographic characteristics. Qualitatively the micro-habitat features influencing shoal composition (e.g. more *C. personatus* in smaller shoals within the structure of the reef rather than on the edges) appear to recapitulate the coarse scale differences in species distribution seen across large spatial scales (Greenfield and Johnson 1999, Victor 2019). While there is some evidence from this analysis which resolves this pattern, only the pattern of larger shoals being composed of relatively fewer *C. personatus* is strongly supported (evidence ratio 74.47).

Here we provide a framework for using modern techniques to evaluate the habitat quality of a reef from the perspective of the species of interest. The relationship between mean site quality and the variation in quality on a site and the number of fish reflects that *C. hyalinus/personatus* primarily utilize areas along the sand/reef ecotone (Fig. 3.6). In this study the two opposite

extremes of sites can be seen in sites A and L (Fig. 3.2). Site A has very few sand patches and a generally consistent quality while site L is mostly sand which results in a low average quality, but the patches of coral provide a relatively large ecotone on which *C. hyalinus/personatus* can live resulting in a high variance in quality. Finally, site F balances the presence of sand patches mixed with complex reef habitat to provide the ideal habitat in this study and supports the densest population of *C. hyalinus/personatus*. One key aspect of this understanding of habitat quality is that it is inherently a continuum of habitat quality rather than a binary choice, directly informed by how individuals of the species utilize the habitat itself, at a spatial scale relevant to the individual. Metrics like this, developed for other species of interest, can unlock theoretically grounded questions fundamental to the understanding of ecology and evolution for empirical study and/or study in non-model organisms.

When combined with regular recreation of the SfM habitat models, a logistically straightforward task, models developed for species of ecological/economic concern could be used in monitoring efforts. As reefs continue to degrade and become less complex, and efforts are being made at reef restoration a monitoring program could identify the specific ways changes in the reefscape are affecting species of interest and could target restoration efforts (Alvarez-Filip et al. 2009). For example, in a marine protected area, habitat use metrics could be developed for any/all species of economic and ecological interest. Then routine mapping efforts requiring only a few diver-hours of time (depending on survey areas) could be regularly performed to identify locations of maximum change in the reef over relevant timescales. This approach allows for precise refinement of the goals of reef protection or restoration that are likely to lead to greater or more direct effects benefiting the fauna of interest.

Acknowledgements

We would like to thank the people at the University of Belize, Calabash Caye Field Station, for providing field logistics without which this research would not have been possible. We would also like to thank Sharon Magnuson and Maili Allende for assistance with lab work. Finally, we would like to thank Drs. Chris Bird, David Portnoy, Blair Sterba-Boatwright, and Lynette Strickland for helpful comments to improve the manuscript. We would like to thank funding sources: the Carl R. Beaver Memorial Scholarship, the Karen Koester Dodson Memorial Fund Grant, and Grants-in-Aid of Graduate Student Research Award by Texas Sea Grant College Program awarded to JDS, the Texas Research Development Fund award to JDH and funds provided by TAMUCC to JDH.

CHAPTER IV: Effect of habitat heterogeneity on dispersal

Abstract

Dispersal and its associated costs and benefits have led to the evolution of bet-hedging strategies which reduce variation in fitness at the expense of mean fitness. One environmental factor which theory predicts to play an important role in the dispersal dynamics of a species in a particular place is the spatial heterogeneity of the habitat quality. Spatially heterogeneous habitat is predicted to result in shorter mean dispersal distances, smaller dispersal spreads, and greater propensity for long-distance dispersal events. Here we study how local habitat quality and heterogeneity influence the dispersal dynamics of a common Caribbean reef goby, *Coryphopterus hyalinus*. We find that *C. hyalinus* has an average dispersal distance of 3.1 ± 0.3 km with 95% of individuals dispersing less than 7.7 ± 0.65 km. We empirically test the hypothesis that spatial habitat heterogeneity results in shorter mean dispersal distances and long tails of the dispersal kernel. We observed that families of *C. hyalinus* living in more heterogeneous habitats exhibited shorter mean dispersal distances, smaller dispersal spreads, and higher propensity for rare, long-distance dispersal events. This observation likely has implications for the design of marine reserve networks and for understanding how changes to marine habitats can affect population dynamics.

Introduction

Dispersal is associated with both benefits and costs that manifest at the individual and population levels. At the individual level dispersal allows for the colonization of lower density areas, thereby reducing costs of competition faced by an individual and reducing the likelihood of inbreeding and kin competition (Hanski and Thomas 1994, Bowler and Benton 2005). At the population level dispersal leads to increased gene flow which can introduce novel alleles into a

population conferring greater genetic diversity, and thereby an increased resistance and resilience to perturbation (Bowler and Benton 2005, Ronce 2007, Duputié and Massol 2013). Despite these benefits dispersing individuals face both opportunity and risk costs associated with dispersal (for review see Bonte et al. 2012). These costs can include the opportunity cost of deferring reproduction while dispersing (Part 1991, Hinsley 2000). Additionally, dispersal is often associated with high levels of mortality, either during the process of dispersing, or by failing to find suitable habitat after dispersing (Greig 1993, Leggett and Deblois 1994, Houde 1997, Cheptou et al. 2008).

One consequence of the balancing of costs and benefits of dispersal has been the evolution of bet-hedging strategies (see also: “drift-retention dichotomy”, Hannah et al. 2000, “dispersal plasticity”, Clobert et al. 2001, “dispersal polymorphism”, Nanninga and Berumen 2014) where some offspring disperse (higher risk, higher reward) while others either don’t disperse at all, or do so on a reduced scale (lower risk, lower reward; Toonen and Pawlik 1994, 2001, Nathan and Muller-Landau 2000, Rousset and Gandon 2002, Nanninga and Berumen 2014). Short dispersal distances may be favored evolutionarily because of reduced individual costs associated with local dispersal such as the lower chance of dispersing away from any suitable habitat (e.g., into the middle of the ocean), and the increased likelihood of finding high quality habitat close to the natal site, despite the increase in competition. The natal site is by definition a sufficiently high quality patch, as the individual’s parents were able to survive to reproduce in that location (Hastings 1983, Holt 1985). Additionally, patch quality is often not uniformly distributed in space. In the case of spatial heterogeneity, any movement away from a quality patch increases the probability of settling on a lower-quality patch, potentially reducing reproductive output and leading to a reduction in fitness (Hastings 1983). Theoretical work has found that spatial

heterogeneity and temporal homogeneity in patch quality select for increased peakedness of the distribution of dispersal distances (a.k.a., the dispersal kernel) leading to reduced mean and variance in dispersal distances while selecting for an increased propensity for long distance dispersal (Birnbaum 1948, Proschan 1965, McPeck and Holt 1992, Baker and Rao 2004, Massol and Débarre 2015).

Spatial heterogeneity in habitat quality selects not only for a reduction of the mean dispersal distance, but also for dispersal polymorphisms favoring both short and long-distance dispersers. Short mean dispersal distances increase the probability of inbreeding and kin competition which will also reduce fitness and so selection can favor individuals that successfully disperse long distances (Hamilton and May 1977). Through these competing selective pressures the mean distance of dispersal is greatly reduced, with reduced dispersal spread, while simultaneously extending the tails of the dispersal kernel, leading to a distribution characterized by increased propensity for rare long-distance dispersal (Birnbaum 1948, McPeck and Holt 1992, Westfall 2014, Lampert and Tlustý 2016). This leads to an evolutionarily stable state where a few individuals widely disperse while the majority of individuals remain close to the natal site, possibly dispersing locally within the area (Hastings 1983, McPeck and Holt 1992).

Despite the abundance of theoretical evidence few studies have attempted to empirically evaluate these assertions, particularly in the context of a single species. Empirical research on dispersal, particularly in a marine environment, is logistically and technically difficult, and tends to produce only a single estimate of dispersal for a species in a particular region making it difficult to test hypotheses about how a species will respond within a region (e.g. D'Aloia et al. 2015, Pinsky et al. 2017). *Coryphopterus hyalinus*, a common Caribbean reef goby, provides an excellent model system to empirically investigate dispersal dynamics. Larvae of cryptobenthic

reef fishes, an ecomorphological category containing a diversity of small bodied, benthic associated fishes, tend to be more abundant in near-reef environments compared with pelagic environments which are dominated by larvae of larger-bodied reef fishes, suggesting short-distance dispersal predominates in this group (Brandl et al. 2019). The dispersal kernel for *C. hyalinus* is thought to be relatively small based on the presence of many, multi-generational relatives within a single site in the ecologically similar species *C. personatus* (Selwyn 2015, Selwyn et al. 2016, 2022, Chapter 2). In addition to the apparent generally restricted dispersal, this species shows little genetic differentiation across relatively large spatial scales (Selwyn et al. 2016) and is thought to be a single species throughout the Caribbean basin (Baldwin et al. 2009). These two pieces of evidence suggest that the dispersal kernel of *C. hyalinus* will be generally characterized by short mean dispersal distances with some fraction of individuals dispersing long distances (Wright 1942, Spieth 1974).

Coryphopterus hyalinus are nearly ubiquitous throughout the Caribbean, and likely are one of the most numerically abundant fishes found in the region (Böhlke and Robins 1962). Ecologically, *C. hyalinus* tend to be habitat generalists, living in shoals of tens to hundreds of individuals, patchily distributed across the reef (Robertson and Justines 1982). Despite broad generality in habitat use, particular topographical features of the reef (e.g. boundaries between sand and reef areas) tend to attract larger shoals suggesting that there are areas of higher and lower quality habitat across the reef-scape (Selwyn et al. in review, Chapter 3). Once an individual has settled in a shoal they appear to not leave or change shoals as adults and live in that location their entire ~90 day post-settlement life (Dominici-Arosemena and Wolff 2005, Beeken et al. 2021).

Reproductively *C. hyalinus* is a protogynous hermaphrodite with males forming harems at low population density where mate monopolization is possible (Cole and Robertson 1988, Allsop and West 2004). Eggs are fertilized continuously throughout the year with eggs laid within the reef structure and guarded by the male (Cole and Robertson 1988, Gardner 2000). The distribution of larval durations of *C. hyalinus* is fairly broad between 21 and 50 days (32 ± 5 days SD), with larval growth rate strongly negatively associated with larval duration, suggesting larvae are actively involved in habitat selection (Montgomery et al. 2001, Elkin and Marshall 2007, Beeken et al. 2021). Coral reef goby larvae are born with well-developed pigmented eyes and well developed vertical and pectoral fin folds (Thresher 1984). They emerge with a small or absent yolk sac, well developed jaws and digestive tracts and need to feed within the first two days post-hatch (Thresher 1984, Gardner 2000). *Coryphopterus hyalinus/personatus* larvae were observed to be negatively buoyant at hatching (JD Hogan pers. obs.); swim bladders inflate within a few hours of hatching (Thresher 1984). The ecologically and visually similar *C. personatus* is also observed far more frequently in light-traps near the reef rather than near the surface (Hendriks et al. 2001). All of this suggests that the larvae of *C. hyalinus* may be capable of reducing their advection very soon after hatching which could subsequently lead to reduced average dispersal distances.

Using *C. hyalinus* as a model we test the hypothesis that habitat heterogeneity and habitat quality, independent of each other, influence the shape of the dispersal kernel. We estimate dispersal kernels for individual families of *C. hyalinus*, based on genetic kinship analysis, and utilize these familial dispersal kernels to test the hypothesis that habitat heterogeneity and habitat quality influence the dispersal kernels. We expect that families residing on more heterogeneous habitats will exhibit dispersal distributions characterized by an increasingly leptokurtic

distribution, with a short mean dispersal distance and small neighborhood size as an expression of the common bet-hedging strategy, with most larvae remaining near the natal reef, while a few disperse a long distance. Additionally, we hypothesize that families residing in more homogenous habitats will exhibit a relatively flatter dispersal distribution, exhibiting longer mean dispersal distances with fewer rare long-distance dispersers.

Methods

Sampling

Samples were collected from eleven reef sites (approximately 10 m x 20 m) on the windward face of Turneffe Atoll (centered around 17° 16' 40.55" N, 87° 48' 18.08" W), Belize, Central America in January 2017 (Fig 4.1). Turneffe Atoll consists of mangrove islands surrounding a shallow lagoon on a submarine ridge approximately 9 – 23 km offshore from the main Belize Barrier Reef. The forereef habitat consists of spur and groove macrohabitat with generally southerly currents (Garcia and Holtermann 1998, Ezer et al. 2005). All sites were located along the forereef at 15-20 m maximum depth and spaced approximately evenly apart on a logarithmic scale radiating from a central collection site centered on Calabash Caye. At these depths, the major habitat forming benthic community consists mainly of *Orbicella*, *Siderastrea*, *Porites*, and *Agaricia* corals, various sponge species and, to a lesser degree, Gorgonian species interspersed amongst live rock, coral rubble, and sand. Seven hundred ninety-eight specimens were collected using hand nets from multiple, distinct, mixed-species shoals (4 to 7) haphazardly selected within each site.

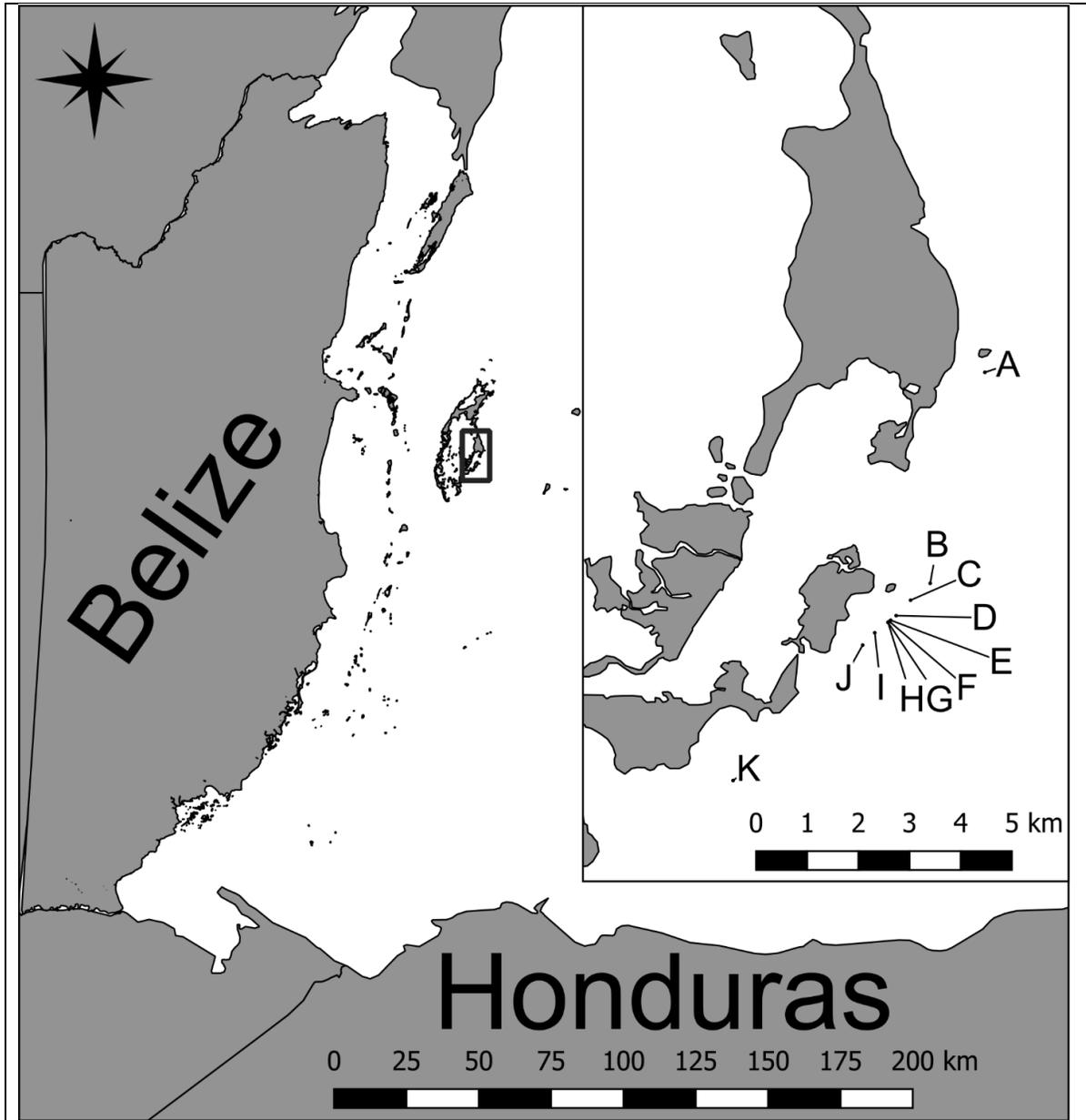


Figure 4.1: Map showing sampling locations on the windward side of Turneffe Atoll, Belize. Sampling sites are labelled A – K based on position along the North-South axis of the atoll.

Genomic Library Preparation and Sequencing

After collection, samples were humanely euthanized using buffered MS222 and stored in 95% non-denatured ethanol. All collections were performed under approval of Texas A&M University-Corpus Christi IACUC protocol (#05-14) and Belize Fisheries Department Aquatic

Scientific Research Permit (#000002-17). Genomic DNA was extracted from fin and muscle tissue from the caudal end of each fish using E.Z.N.A.® DNA extraction kit (Omega Bio-tek). Extracted DNA was used to make double digest restriction associated digest libraries (ddRAD) following a modified Peterson et al. (2012) protocol. For each sample, 150 ng of high-quality, high-molecular weight DNA were cleaned with AMPureXP beads (Beckman Coulter A63882) at a 2X bead to DNA ratio, following the manufacturers protocol. The beads were left in with the samples for bead-in library prep for all steps, except the first binding of the bead size selection. Samples were digested with 2.5U each *MspI* (NEB R0106) and *EcoRI*-HF (R3101) at 37°C for 1 hour. To stop the digestion and reactivate the AMPure beads, a 1.5X ratio of 3M NaCl, 20% PEG solution (Fisher et al. 2011, Faircloth and Glenn 2014) was added to the reaction and the bead clean-up was performed following the AMPure protocol. Barcoded adapters with an overhang complementary to the *EcoRI* overhang and non-barcoded *MspI* complementary adapters were ligated to the samples using 3.75U T4 ligase (ThermoScientific FEREL0011). There were 48 uniquely barcoded adapters used. The ligation reactions were incubated at 20°C for 1 hour, followed by 65°C for 10 minutes to inactive the ligase, and then a controlled cool at a rate of 2C/90 seconds until the samples reached room temperature. Ligation reactions were cleaned with 1.5X ratio PEG solution (Fisher et al. 2011, Faircloth and Glenn 2014) to remove ligation buffer that could interfere with the bead size selection and any unincorporated adapters.

The cleaned ligation products underwent a double-sided bead size selection using a 0.3X ratio PEG for the right-side size selection and 0.6X ratio SPRIselect beads (Beckman Coulter B23319) for the left-side size selection. A sample size check was performed by eluting the beads from the first binding step with 10µl water and gelling 3µl of this elution and 3µl of the supernatant from the second binding step on a 1% agarose gel. If there was a gap in sample

presence in the desired size range between the two clean-up discards, then the size-selection product was PCR amplified. The PCR amplification was performed in a 20 μ l reaction consisting of 10 μ l cleaned DNA, 3.666 μ l sterile dH₂O, 4 μ l 5X Phusion Buffer, 0.5 μ l MgCl₂, 0.4 μ l 10mM dNTPs (Thermo Scientific R0192), 0.667 10 μ M indexed primer 1, 0.667 10 μ M indexed primer 2, and 0.1 μ l Thermo Scientific™ Phusion™ High-Fidelity DNA Polymerase (F530). The thermal profile was 98°C 1 minute, 12 cycles of 98°C 10 seconds, 62°C 30 seconds, 72°C 30 seconds, and a final extension step at 72°C for 1 minute. Unique dual indexes were used for each group of 48 barcodes. Purification of the amplification product was done with two consecutive bead clean-ups using a 1X PEG to DNA ratio, with a final elution in 25 μ l water.

Amplification products were checked using 1.5 μ l product on a 1% agarose gel, double quantified using AccuBlue High Sensitivity dsDNA solution (Biotium) with an eight-point standard curve on a SpectraMax M3 plate reader and pooled in equal ng quantities by index. A final size selection targeting a 450-575bp insert size was performed on a BluePippin (Sage Science). Size selected library pools were checked for size on a Fragment Analyzer (Agilent) and qPCR quantified using a KAPA library quantification kit (Roche 07960204001) on an ABI StepOnePlus. The libraries were pooled together in an equimolar manner, and the final library was sequenced on a single NovaSeq S4 lane using 150-bp paired-end reads. To minimize the risk of under/over splitting of loci, a random set of eight samples were secondarily sequenced on a MiSeq 2x300-bp lane and used to create the *de novo* genome on which the NovaSeq reads were mapped. This additional sequencing step was performed to ensure all fragments have overlapping forward and reverse reads in the *de novo* reference genome.

Bioinformatics

After sequencing both NovaSeq and MiSeq reads were separately demultiplexed using the STACKS function `process_radtags` (Catchen et al. 2011). Following this, reads were quality filtered with adapters removed using FASTP (Chen et al. 2018) to only contain paired reads which were longer than 140 bp (NovaSeq) or 280 bp (MiSeq), composed of more than 40% bases with a PHRED quality greater than 20, and have greater than 30% sequence complexity. Additionally, all reads were filtered for possible contaminants using FASTQ SCREEN (Wingett and Andrews 2018). After quality filtering one MiSeq sample and 20 NovaSeq samples were excluded from subsequent steps as they contained fewer than 10,000 total reads. All raw sequences used in this study were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project accession number PRJNA782562.

Assembly of the *de novo* reference genome, read mapping, and genotyping was performed using the DDOCENT pipeline (Puritz et al. 2014) modified for use on a high performance computer (Biesack et al. 2020, Bird 2020). To assemble the *de novo* reference genome from the MiSeq reads, first PEAR (Zhang et al. 2014) was used to join forward and reverse reads together into a single overlapping read. These reads were then clustered and collapsed into unique reads using CD-HIT (Li and Godzik 2006, Fu et al. 2012) to have a similarity of greater than 90%. Following this, RAINBOW (Chong et al. 2012) was used with a similarity threshold of 90% to create reference contigs from the unique reads. Contigs with less than 2x coverage were filtered from the *de novo* assembly. Following reference assembly, NovaSeq reads were mapped to the reference genome using the MEM (Li 2013) algorithm in BWA (Li and Durbin 2009, 2010). To map the reads, a match value of 1 was used with a mismatch value of 6, a gap opening penalty of 10, and a clipping penalty of 30 and 5 for 5' and 3' clipping respectively. Reads were only

mapped if they had a minimum alignment score of 50. Mapped reads were further filtered to only retain reads where both pairs were successfully mapped with a minimum mapping quality of 20, without secondary alignments. Finally, genotyping was performed using FREEBAYES (Garrison and Marth 2012). Genotypes were only called if they had a read depth per individual of at least two, mapping quality of at least 30 and a minimum base quality of 20. To call an alternate allele, the number of alternate reads at a locus was required to be greater than 20 with the sum of the alternate read base quality greater than 600. Further, across individuals to be genotyped, a locus was required to have a minimum coverage of 10, and at least 20 alternate reads in the population with at least one individual having more than 37.5% alternate reads at the locus.

Genotypes were filtered using FLTRVCF (Biesack et al. 2020, Bird and Selwyn 2021) with the general filtering principles based on O’Leary et al. (2018). Because in the region both *C. hyalinus* and *C. personatus* exist in shoals composed of both species and are difficult to visually distinguish (Selwyn et al. 2022, Chapter 2) an initial genotype filtering was performed on the combined dataset to separate the species (see below). After extracting the *C. hyalinus* individuals from the dataset, a second more thorough genotype filtering was performed for subsequent analyses. In the round of filtering to distinguish species, MNPs and indels were removed to retain a dataset containing only SNPs which were filtered to have a minimum PHRED quality of 40, minimum mean depth of coverage across individuals of 10, a minor allele frequency in the population of at least 0.5% and present in at least 75% of the samples. After this initial filtering, a single SNP per contig was randomly selected to use in the analysis required to distinguish between the two species. After subsetting the data to only include pure *C. hyalinus* samples, SNPs were filtered again to include only loci with properly paired reads, a minimum mean PHRED quality of 200, minimum mean depth of coverage across individuals of 20, maximum

mean depth of coverage of 150, minimum depth in an individual of three, minimum alternate allele count of three, minor allele frequency at least 0.5%, allele balance of 37.5%, present in at least 70% of samples with samples with more than 50% missing data excluded from the dataset.

Species Identification

Using the initially filtered dataset, we split the samples into *C. hyalinus* and *C. personatus*. To distinguish between these species we first used ADMIXTURE (Alexander et al. 2009). ADMIXTURE was run with 10-fold cross validation to test for K between 1 and 25, where we hypothesized that a K of 2 would minimize the cross-validation error, given the presence of two species in the sample. From the two clusters observed by ADMIXTURE based upon the observed allele frequencies, individuals with an assignment probability great than 0.9999 were considered pure specimens of that cluster and used to initialize NEWHYBRIDS (Anderson and Thompson 2002) to determine if individuals of less certain assignment were first- or second-generation hybrids, or first-generation backcrosses between the two groups. Due to constraints of the software, only the 200 loci most differentiated (i.e., greatest F_{ST} estimates) between the two “pure” clusters were used in the NEWHYBRIDS analysis. NEWHYBRIDS was run using 5 independent MCMC chains, with 100,000 burnin iterations and a subsequent 1,000,000 sampling iterations which were thinned by 100. Jeffreys priors were used for both mixing proportions and allele frequencies to allow rare alleles to influence the classifications more fully. Proper mixing, exploration of parameter space, and convergence were confirmed by visually inspecting trace plots and confirming that the \hat{R} value equaled one (Vehtari et al. 2019). Finally, to determine the species identities of each cluster we first mapped the trimmed NovaSeq reads to the mitochondrial genome of *Bathygobius cocosensis* (Evans et al. 2018) to exclude reads from the nuclear genome using the MEM (Li 2013) algorithm in BWA (Li and Durbin 2009, 2010). Next we

used BLAST in GenBank (Altschul et al. 1990) to identify the best hit for each mitochondrial read in each sample. The species identity of the clusters was determined based on the species identity of the majority of individuals with mapped mitochondrial reads in the cluster.

Relatedness

To identify the best estimator of pairwise relatedness we simulated 1,000 pedigrees to create dyads with known relationships based on the observed allele frequencies in the sampled *C. hyalinus*. The relationships simulated were parent-offspring ($r = 0.5$), full-sibling ($r = 0.5$), half-sibling ($r = 0.25$), grandparent-grandchild ($r = 0.25$), avuncular ($r = 0.25$), double-first cousin ($r = 0.25$), first cousin ($r = 0.125$), second cousin ($r = 0.03125$) and unrelated ($r = 0$). Due to half-siblings, grandparent-grandchild, avuncular, and double-first cousins being indistinguishable with the genetic data available these are pooled in later analyses as secondary relationships (Jacquard 1974). Pairwise relatedness for simulated dyads was calculated using both the dyadic likelihood (Milligan 2003) and the PLINK method-of-moment technique (Purcell et al. 2007, Morrison 2013) as implemented in the SNPRELATE R package (Zheng et al. 2012). The best estimator was chosen based on which was most closely correlated according to the Pearson's correlation coefficient to the true relatedness across a range of shared loci between dyads (Taylor 2015). The best estimator of relatedness from this data was the dyadic likelihood method which was used in all subsequent analyses.

To determine how many unlinked loci were needed to distinguish an unrelated dyad from any of the simulated relationship types, we calculated the percentage of the simulated unrelated dyads with relatedness values greater than the lower 95% confidence interval of each of the other simulated relationship classes. We then used a generalized additive spline model, with a beta family distribution, to model the relationship between the percentage of misclassified unrelated

dyads and the number of loci for each relationship type and method of calculation. Finally, bootstrap resampled dyads (see below) were removed if they did not share sufficient successfully genotyped loci to exclude 95% of simulated unrelated pairs, based on the most likely relationship type of the observed dyad.

For each observed dyad we calculated the observed relatedness using the subset of independent loci with no missing data shared by each member of the pair. Additionally, we simulated 1,000 unrelated dyads using these loci to generate a null distribution representing an unrelated pair. Only dyads with an estimated relatedness greater than the 99.9% quantile of the simulated unrelated dyads were included in subsequent analyses as related pairs. Finally, we used 1,000 bootstrap resamples across loci to calculate relatedness confidence intervals. For each bootstrap resample in the process of calculating relatedness we calculated the Jacquard (1974) coefficients of identity to categorize the most likely relationship class, given that bootstrapping of the loci. Bootstrapped relationship categories were binned into parent-offspring, full-sib, secondary (including half-sibs, avuncular, grandparent-grandoffspring, and double-first cousin), cousin, and unrelated (including unrelated and second cousins). For each dyad the probability of the existence of a relationship, along with the probability of each type of relationship was calculated based on the bootstrap resampling of the loci. An undirected weighted graph with nodes representing individual fish, edges indicating related dyads, and edge weights indicating the probability of the dyad being related. Louvain clustering was used to identify familial groups through maximization of within group modularity (Blondel et al. 2008). Modularity is a measure of the density of edges connecting nodes within a cluster, compared to the density of edges connecting nodes found in different clusters (Clauset et al. 2004).

Dispersal Estimation

The dispersal kernel was estimated for each family group following the methods of Filipović et al. (2020). Briefly this technique utilizes the information encoded within the familial relationship between pairs of individuals and knowledge of the life-history of the study species to identify the set of possible dispersal events which occurred, ultimately resulting in the observed spatial distance between the pair. This familial relationship and the observation of their current physical proximity is used to create a set of effective dispersal distances for each dyad from which the effective dispersal kernel is estimated. For example, between pairs of full- or half-siblings there are two dispersal events. From these sets of possible effective dispersal events, we created for each pair the set of effective dispersal distances by dividing the observed spatial distance between the pair by each of the set of effective dispersal events. To avoid pseudo-replication in the case when the set of dispersal events separating a dyad is greater than one (cousins) we averaged across the effective dispersal distances.

It is likely that the method employed here resulted in consistent underestimates of the dispersal distance (Jasper et al. 2021), but there is no reason to believe that bias in the method employed here is different among habitats or would influence conclusions. Recently this method of calculating dispersal from close-kin relationships has been criticized for consistently underestimating dispersal distances (Filipović et al. 2020, Jasper et al. 2021). Some of these critiques are specific to the life history of the mosquitoes upon which the method was developed, such as not accounting the movement of females between oviposition events (Jasper et al. 2021). This criticism is not applicable to *C. hyalinus* which has a small home range and is thought not to leave the shoal post-settlement, certainly not distances on the scale of kilometers as the dispersal is likely to be (Dominici-Arosemena and Wolff 2005). Despite this, there are substantive

critiques regarding the validity of this estimator on theoretical grounds as it pseudo-replicates distances by creating a set of distances per dyad and fails to correctly decompose distances between dyads as variances (Jasper et al. 2021). We account for the problem of pseudo-replication by using the average of the decomposed set of spatial distances for each dyad (Jasper et al. 2021). However, the more fundamental problem of how the distances between dyads is decomposed is left unaddressed which constantly downwardly biases the dispersal estimate to approximately 80% of the true dispersal (Jasper et al. 2021).

To adjust for the fact that the distribution of distances between all pairs of individuals is not uniform and includes far more proximal pairs than distant pairs, we randomly assigned a set of dispersal events to each unrelated dyad to calculate the sampling distribution in the same manner as the dispersal distribution (Filipović et al. 2020). This sampling distribution was then used to weight observations of related dyads based on the likelihood of observing the pair that far apart solely based on how sampling was performed (Appendix 1).

The effective dispersal distribution was calculated using a hierarchical gamma model (Nathan et al. 2012) with both the shape and rate parameters allowed to vary from an overall “related” value for each family group with more than three dyads (see Appendix 1 for full model). Due to the uncertainty in the assigned relationships between pairs we simulated 1,000 datasets of dyadic relationships and family structure based on the bootstrapped probability of each relationship existing. By fitting the model to each simulated dataset, we integrate across the uncertainty in relationship assignments making this uncertainty inherent to the results. We fit the model to each of these simulated datasets using an HMC sampler in STAN (Carpenter et al. 2017). For each simulation the model was run for 2,000 iterations including 1,000 warmup iterations on four independent chains. Mixing of chains within each simulation was confirmed by

ensuring all \hat{R} values equal one (Vehtari et al. 2019). The posterior distribution for all simulations were then combined into a joint posterior accounting for the uncertainty in the relationship assignments. Because each probabilistic family is not represented in each imputed dataset, familial dispersal parameters were combined only across simulations which contained that probabilistic family. After fitting the models, we confirmed that the model reasonably represents the observed and simulated data with a posterior predictive check (Gelman et al. 2020). To account for the bias in the dispersal estimator and make our dispersal estimates comparable with other studies we multiplied posterior mean and quantile dispersal estimates by 1.2 (Jasper et al. 2021).

Habitat Quality

For each sampled site we also measured the habitat quality as a function of the amount of reef habitat along with depth, distance to the sand/reef margin, fine and coarse scale complexity, and viewshed at a cm scale resolution. Briefly, to measure the habitat quality we generated 3D habitat models using structure-from-motion photogrammetry which were then analyzed to determine the habitat characteristics associated with denser shoals of *C. hyalinus/personatus*, using a Bayesian log-Gaussian Cox process spatial model (for data see: Selwyn et al. 2021, for details see: Selwyn et al. in review, Chapter 3). We converted the predicted density estimates into a metric of habitat quality, integrating across topographical metrics by first linearizing the predictions using a log transformation and then scaling it to be between 0 (low quality) and 1 (high quality). For each site we calculated the mean and standard deviation of the habitat quality to determine how habitat quality and heterogeneity respectively explain the observed variation in the mean dispersal distances across families (Fig 4.S1).

Analysis of Habitat Quality & Dispersal Distribution

To understand the relationship between habitat quality, as measured using the habitat usage model above, and dispersal characteristics we used a multivariate Bayesian mixed model to explain variation in estimated familial dispersal parameters using average site habitat quality and heterogeneity. The dispersal parameters used as the dependent variables relating to the theoretically grounded hypotheses are the mean dispersal distance (μ), dispersal standard deviation (σ , i.e. the dispersal spread, Siegel et al. 2003), and excess kurtosis (κ , i.e. the propensity for long-distance dispersal events, Westfall 2014). To incorporate uncertainty in both dependent and independent variables we included measurement error in the estimation of habitat quality and heterogeneity. To confirm that the number of dyads present in a family is not driving the observed relationship we additionally included the total number of dyads as an explanatory variable in the model. All dependent variables were log transformed and modelled using a gaussian link function to allow for the incorporation of uncertainty in the dependent variable, through the inclusion of the posterior standard error of each dispersal parameter. This model was built using BRMS and fit using an HMC sampler in STAN (Carpenter et al. 2017, Bürkner 2018). As before the model was run for 2,000 iterations including 1,000 warmup iterations on four independent chains. Mixing of chains and proper exploration of parameter space was confirmed by visually inspecting trace plots and ensuring all \hat{R} values equal one (Vehtari et al. 2019). After fitting the model, we confirmed that there were no divergent transitions and that the model reasonably represents the observed data with a posterior predictive check (Gelman et al. 2020). Finally, the posterior probability of the hypothesized relationships between mean dispersal, variation in dispersal, and propensity for long-distance dispersal and the habitat quality and

heterogeneity were calculated to determine the support for and strength of the observed relationships.

Results

Bioinformatics

The ddRAD library preparation resulted in 1.9 billion total NovaSeq read pairs and 12 million MiSeq read pairs, of which 93.3% and 53.8% respectively passed quality filtering and decontamination steps. After removal of a single individual from the MiSeq sample for having fewer than 10,000 reads, the average number of reads per individual used to construct the reference genome was $953,436 \pm 209,720$ (SE). The constructed reference genome was composed of 88,540 contigs with a mean contig length of 470 ± 65 bp (SD). A total of 20 individuals were removed from the NovaSeq samples due to having fewer than 10,000 reads with the remaining 778 containing an average of $2,254,302 \pm 95,568$ (SE) reads per individual. After mapping NovaSeq reads to the reference genome and filtering out poorly mapped reads, an average of $1,850,867 \pm 81,223$ (SE) reads per individual were successfully mapped to the reference genome. Genotyping of individuals initially led to 1,020,147 putative SNPs spread across 22,217 contigs (46 ± 25 SD SNPs/contig). This was then filtered to a reduced set used for species identification of 1,726 SNPs which had been thinned to one random SNP per locus. After species identification (see below) and subsetting to only include *Coryphopterus hyalinus*, there were 625 individuals initially genotyped at 802,220 putative SNPs across 22,196 contigs (36 ± 19 SNPs/contig) which was subsequently filtered to 479 individuals with 59,966 SNPs on 6,345 loci (8.98 ± 5.35 SNPs/locus).

Species Identification

Admixture, hybridization, barcoding analysis, and secondary SNP filtering were used to isolate 479 *C. hyalinus* for further analysis. The admixture analysis found two distinct clusters which >70% of specimens were identified as pure representatives of one cluster (assignment probability >0.9999, Figs 4.S2 & 4.S3). Mitochondrial DNA was used to determine the species identity of each cluster with all pure specimens in cluster one that had successfully identified mtDNA being *C. personatus* and 98% of pure specimens in cluster two that had successfully identified mtDNA being *C. hyalinus*. Similar to previous work (Selwyn et al. 2022, Chapter 2), we found the reefs to harbor ~82% *C. hyalinus* (95% CI: 80 – 85%). NEWHYBRIDS found that a total of 18 specimens (2%, 95% CI 1.5 – 3.6%, Figs 4.S4 & 4.S5) were putative first- or second-generation hybrids, including F1 and F2 hybrids and both forms of backcrossed individuals. Of the 625 specimens identified as pure *C. hyalinus* by both ADMIXTURE and NEWHYBRIDS 479 remained after a secondary round of SNP filtering designed to retain individuals with fewer missing loci.

Relatedness

The simulation analysis found that the dyadic likelihood relatedness estimator was better correlated with the true expected relatedness than the moments-based estimator for the numbers of loci analyzed in this study (Fig 4.S6). Given at least 700 unlinked shared loci in the dyad, the dyadic likelihood estimator is extremely well correlated with the true expected relatedness value ($\rho \geq 0.96$, $p < 0.01$). We found that even fewer shared loci were required to confidently classify the relationship of more closely related dyads based on the analysis of the proportion of simulated unrelated pairs incorrectly classified into various relationship types. For example, to be distinguished from unrelated pairs, full sibling and parent-offspring dyads required at least 122

unlinked shared loci while first cousins require at least 1,525 unlinked shared loci, and second cousins require more loci than available in this study (Fig 4.S7, Table 4.1). Dyads were included in subsequent analyses only if they shared enough loci to confidently determine the dyad is not unrelated given the estimated relationship class.

Table 4.1: Minimum number of Loci. Table showing the minimum number of loci required to exclude at least 95% of simulated unrelated pairs.

Relationship Class	Relationship	Minimum number of loci	
		Maximum Likelihood	Method of Moments
Parent-Offspring	Parent-Offspring	123	612
Full Sibling	Full Sibling	123	605
	Half Sibling	466	2,821
	Avuncular	459	2,764
Secondary	Grandparent-Grandchild	472	2,833
	Double First Cousin	472	2,802
	First Cousin	1,526	>6,345
Cousin	Second Cousin	>6,345	>6,345

After filtering to retain only high-confidence, related dyads, we found 1,267 probabilistically related dyads after bootstrapping across loci. Across all simulations these dyads formed 13 families composed of more than three dyads containing an average of 8.9 (± 1.4 SE) members (Fig 4.2). All other probabilistic families were too small (or too unlikely to include multiple members in each simulation) that they were excluded from estimating individual familial dispersal kernels and only retained for the overall species dispersal kernel. After simulating likely relationships 1,000 times, we found an average of 38.7 ± 4.1 related dyads across simulations made up of 6.4 ± 0.5 full-sib pairs, 7.0 ± 1.0 half-sib (or other second-degree relation) pairs, and 25.4 ± 4.1 cousin pairs (Figure 2). The lengths of individuals making up related pairs runs across the range of collected sizes with an average of 5.2 pairs of cousins in each simulation likely of different generations based on a difference in standard length greater

than 10 mm (Beeken et al. 2021) with 0 intergenerational full-sibs and 0.1 intergenerational half-sibs (or other second-degree relation).

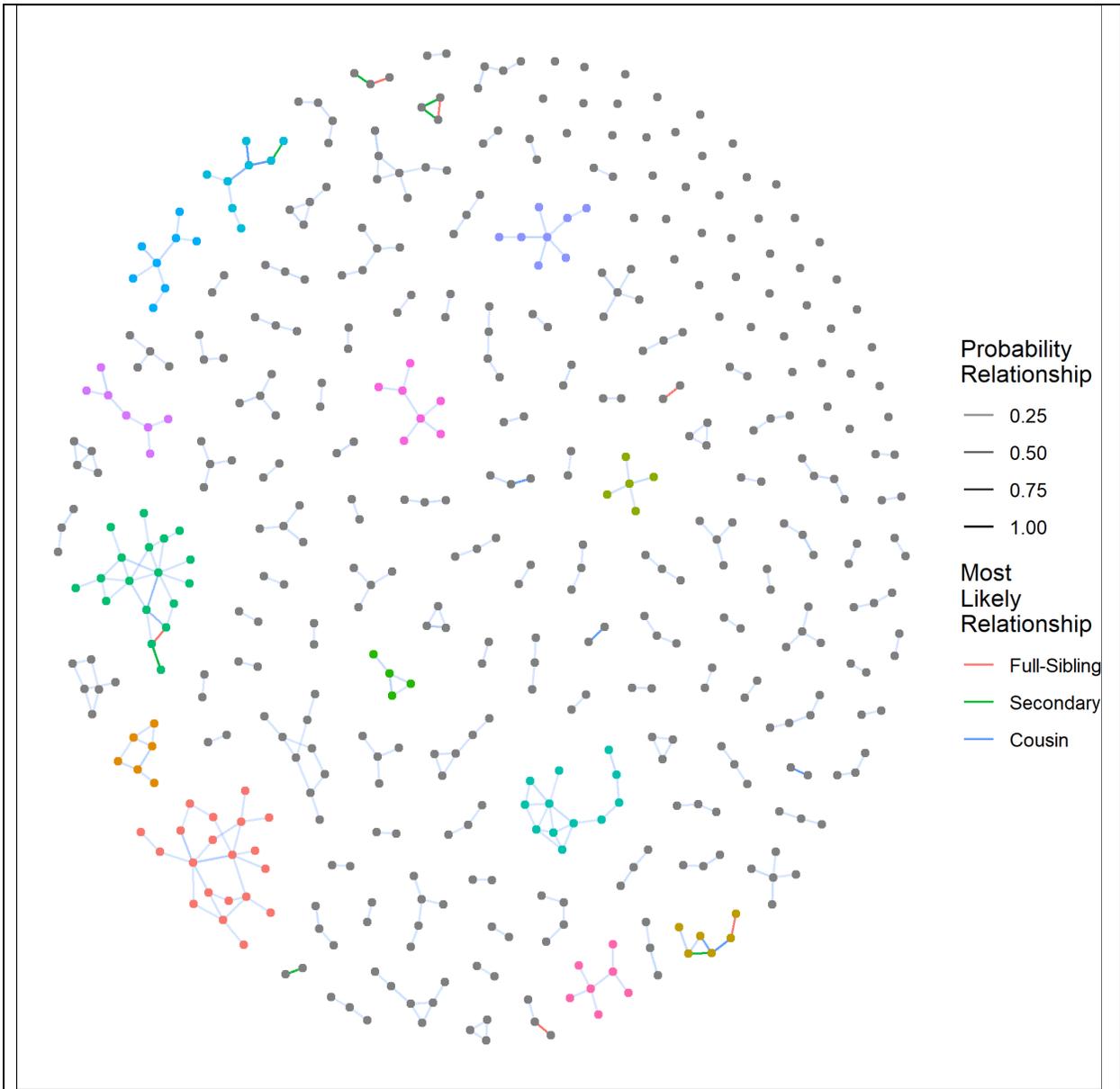
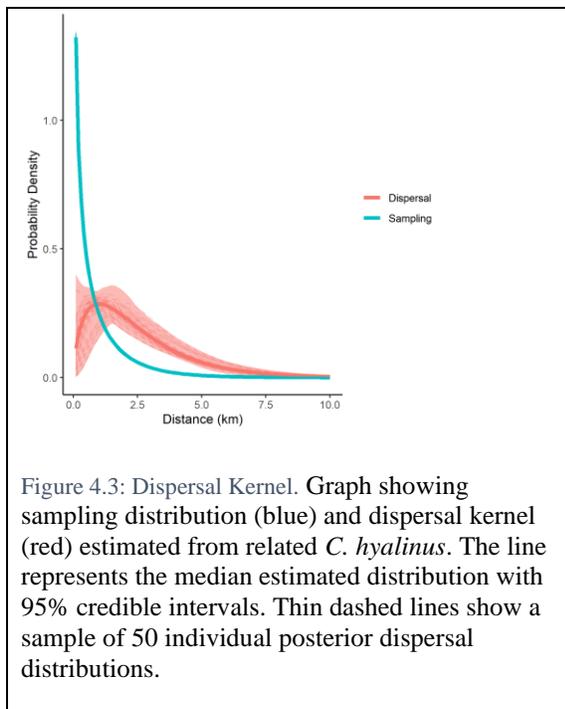
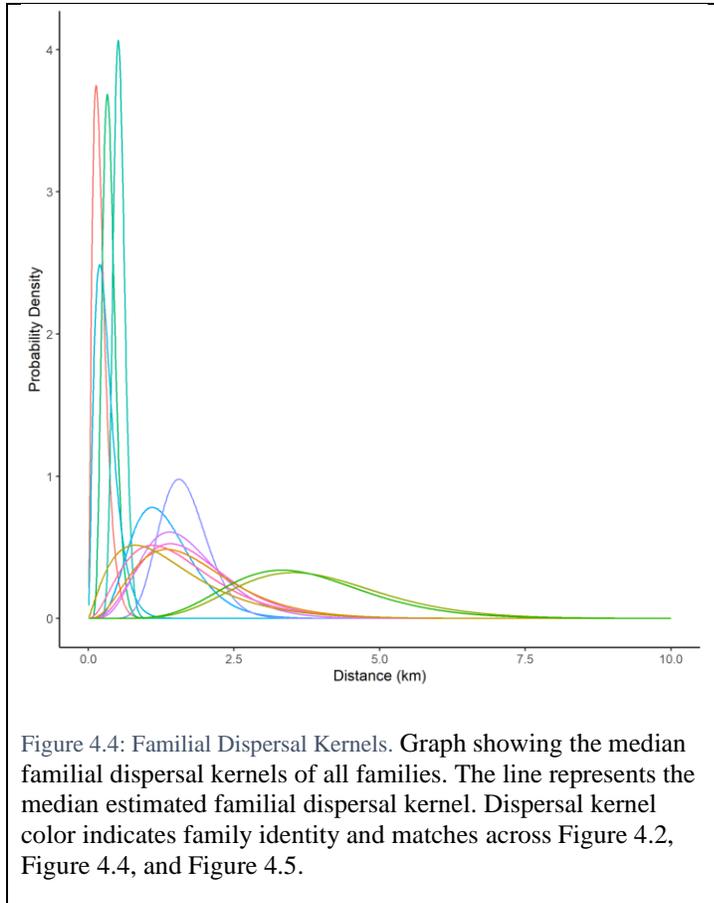


Figure 4.2: Probabilistic Families. Graph showing interconnections of all probabilistically related dyads. Points represent individual fish with colors family membership in families with familial dispersal kernels calculated. Family color matches across Figure 4.2, Figure 4.4, and Figure 4.5. Edges between points show the probability of a pair being related given the observed loci with darker edges indicating more probable relationships. Edge color indicates the most likely relationship of a dyad given they are related.

Dispersal Estimation

From the probabilistic resampling of related pairs, we calculated a dispersal kernel for the species, accounting for the sampling distribution, as well as estimating hierarchical components showing variation among families from the species mean kernel. The average distance between a pair of unrelated fish was 0.8 ± 0.015 km. After accounting for the distance between unrelated pairs, we estimated a mean dispersal distance of 3.1 ± 0.3 km with a σ of 2.0 ± 0.19 and κ of 3.5 ± 0.67 (Fig 4.3). Fifty percent of individuals had dispersal estimates less than 2.5 ± 0.31 km with 95% dispersing less than 7.7 ± 0.65 km. Across families with more than three dyads average intra-familial dispersal distances ranged between 0.31 ± 0.08 km to 4.48 ± 0.24 km (Fig 4.4). Further the range of variation in dispersal (σ) measured across families was between 0.095 ± 0.081 and 1.27 ± 0.14 and propensity for long distance dispersal (κ) between 0.47 ± 0.5 and 2.89 ± 0.44 .

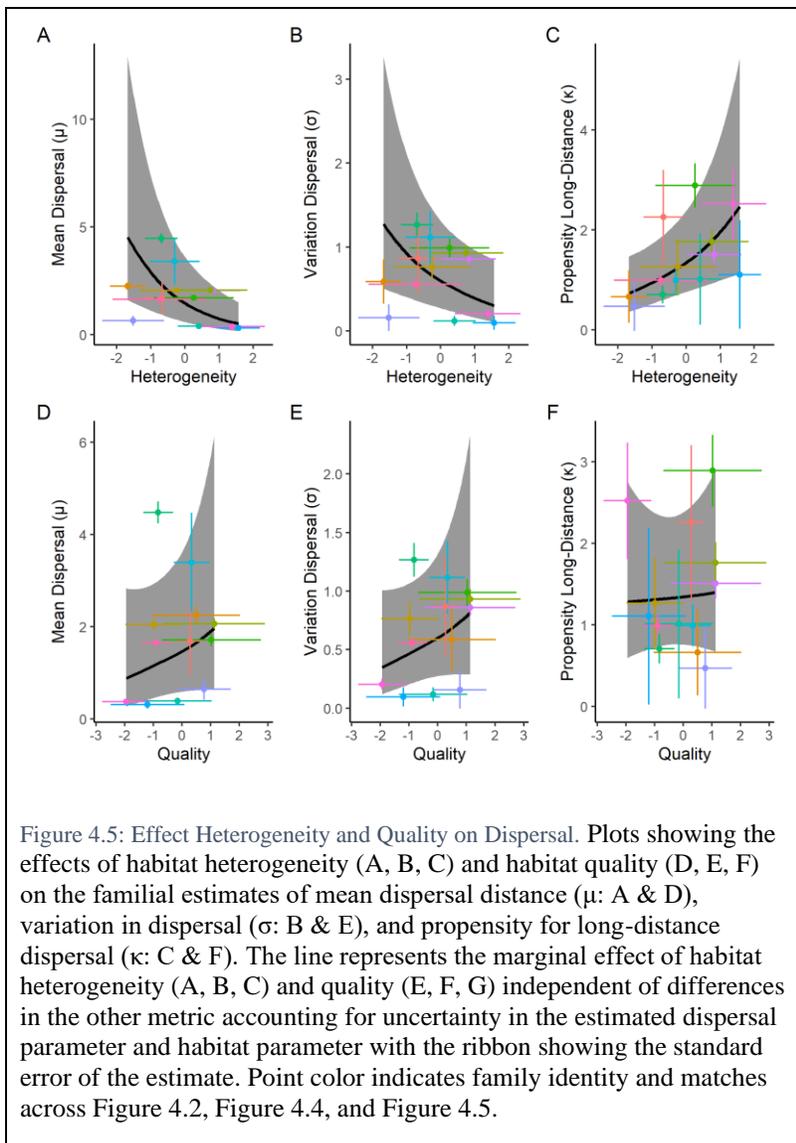




Dispersal Habitat Relationship

The mean estimated habitat quality and heterogeneity that a family is distributed across explains a significant portion of the variation in mean estimated dispersal distance ($r^2 = 0.74 \pm 0.25$; Fig 4.5A/D), variation in dispersal distance ($r^2 = 0.43 \pm 0.23$; Fig 4.5B/E) and propensity for long distance dispersal events ($r^2 = 0.15 \pm 0.07$; Fig 4.5C/F). The number of dyads contained in a family was found to have no influence on the mean familial dispersal (μ 95% CI: -0.05 – 0.06), the familial dispersal spread (σ 95% CI: -0.06 – 0.05) or the propensity for long distance dispersal (κ 95% CI: -0.06 – 0.03). Generally, we found that habitat heterogeneity and not habitat quality influenced the familial dispersal parameters. We found strong evidence (Jeffreys 1961) supporting the *a priori* hypotheses that mean dispersal distance (μ : -0.70 \pm 0.35, posterior

probability = 0.97, evidence ratio = 32.9, Fig 4.5A) decreases while propensity for long distance dispersal increases (κ : 0.41 ± 0.31 , posterior probability = 0.91, evidence ratio = 10.30, Fig 4.5C) with increasing habitat heterogeneity. We also found substantial evidence that variation in dispersal (σ : -0.47 ± 0.37 , posterior probability = 0.90, evidence ratio = 8.55, Fig 4.5B) decreases with increasing habitat heterogeneity. Meanwhile, we found no evidence for habitat quality influencing any of the estimated dispersal parameters with all estimated slopes being centered around 0 (μ 95% CI: $-0.78 - 1.17$; σ 95% CI: $-0.84 - 1.16$; κ 95% CI: $-0.83 - 0.88$, Figs 4.5D-F).



Discussion

Coryphopterus hyalinus exhibits restricted dispersal with an average estimated dispersal distance of 3.1 ± 0.3 km and 95% of fish dispersing less than 7.7 ± 0.65 km. At the family level, mean dispersal distance varied considerably (range 0.31 – 4.48 km). As hypothesized, the variation among familial dispersal kernels was explained by the average heterogeneity of the habitat upon which the family resided. Specifically, we found that families living on heterogeneous habitats had smaller neighborhood sizes (σ) and shorter mean dispersal distances (μ), but also with greater potential for long range dispersal events (κ). Meanwhile, homogenous habitats were characterized by larger neighborhood sizes and longer mean dispersal distances but have a low propensity for long distance dispersal events. These results empirically support the hypothesis that, independent of habitat quality, in heterogeneous habitats, bet-hedging strategies are selected for which result in the majority of individuals remaining relatively close to the natal site with some rare long dispersing individuals (Krug 2009, Scheiner 2014).

While the specific biological mechanisms underlying the relationship between habitat and dispersal are nebulous and as yet unknown, maternal effects and larval traits are likely to be primary determinants (Krug 2009). First, differential provisioning of eggs facilitates dispersal variation and could be tied to habitat quality, linking the two processes (Krug 2001). The degree of habitat heterogeneity could easily influence the degree of maternal provisioning of eggs, either by changing the average amount of provisioning or the degree of variation among eggs (Dziminski et al. 2009, Leal et al. 2013). A slow initial growth rate due to low quality eggs leads to increased mortality and an increase in the amount of time spent as a passive particle in the plankton due to slower development (Green and McCormick 2005). Despite *C. hyalinus* exhibiting a wide range of dispersal durations there is a consistent size at settlement and strong

correlation between dispersal duration and larval growth rate suggesting settlement timing is constrained by larval development (Beeken et al. 2021).

A competing and potentially complementary mechanism to describe the effect of habitat on larval dispersal is the behavior of larvae themselves (Armsworth et al. 2001). Larvae are no longer thought of as merely passive particles whose destination is simply at the mercy of oceanic currents. Larvae are active agents, able to detect suitable settlement areas (Gerlach et al. 2007, Leis et al. 2011), orient (Atema et al. 2002), and swim (Hogan and Mora 2005) permitting larvae to exert choice over settlement habitats (Burgess et al. 2022). *C. hyalinus* spawn in crevices within the reef and the larvae exhibit a relatively high degree of larval competence at hatch which might explain the observed restricted dispersal distances. The degree to which the habitat effects on dispersal kernels observed here are the result of pre-hatch maternal effects and/or post-hatch larval behavior is unknown.

Both maternal effects and larval behavior are heritable traits upon which selection can act (Raimondi and Keough 1990, Schroeder et al. 2012). This provides the link between the increased fitness resulting from the bet-hedging strategy observed in dispersal on heterogeneous habitats and the biological mechanisms leading to the bet-hedging (McAdam et al. 2002, Crean and Marshall 2009). This feedback of the biological mechanisms leading to individuals staying near home and thereby increasing their fitness allows for selection to directly act upon dispersal as a result of the selective pressures on larval behavior and maternal effects.

The average dispersal distance of *C. hyalinus* (3.1 ± 0.3 km) is slightly larger than the average dispersal measured in the same region for *Elacatinus lori*, a sponge dwelling goby (2.8 km, D'Aloia et al. 2015). Notably while both species show habitat preferences *E. lori* is a more extreme habitat specialist, requiring a particular species of host sponge (D'Aloia et al. 2011,

Selwyn et al. in review, Chapter 3). By the very nature of its habitat specificity, *E. lori* inherently lives in a more heterogeneous habitat landscape than *C. hyalinus* due to the extreme differences (more or less binary) in habitat quality between locations with vacant tube sponges and those without. Given this similarity, theory predicts that *E. lori* should have a shorter mean dispersal distance than *C. hyalinus* which is supported by the findings of this study and D'Aloia et al. (2015).

Scaling up from the microscale assessed in this study to a mesoscale there are analogous patterns of the effects of habitat patchiness on dispersal (Pinsky et al. 2012). The result of habitat patchiness at scales of tens of kilometers is relatively closed populations, particularly of species with a dispersal spread (σ) <10 km (Pinsky et al. 2012). Further, the patchiness of the habitat at mesoscales results in reductions in dispersal spread due to the lack of appropriate habitat outside these patches (Pinsky et al. 2012). To extend these concepts to a microscale we see a similar effect of habitat heterogeneity, on the scale of meters, on the dispersal spread which when scaled through generations will influence population dynamics, such as immigration, and could provide a key link to furthering our understanding about how interactions across scale influence population processes such as a population's openness. Indeed, we can see the effect of the short-distance dispersal at the metapopulation level in the observation of significant population differentiation of *C. personatus* between sites as little as 5.4 km apart in the same reef system, assuming the ecological similarity of these species extends to dispersal syndromes (Selwyn et al. 2016).

An emergent property that develops as a consequence of the effect of habitat heterogeneity on the intergenerational familial dispersal kernel is an intergenerational search strategy which progressively reduces the average heterogeneity on which the family lives. If a family lives in a

heterogeneous landscape, then the reduction of the mean dispersal and smaller neighborhood size of the resulting dispersal kernel serves to keep the family living generally on similar habitat. Meanwhile, increasing the propensity for rare long distance dispersal events to potentially take advantage of habitat further afield. However, in relatively homogenous habitats dispersal encourages the family to become more diffusely spread across the landscape. However, these families have a reduced propensity for long distance dispersal events.

Habitats worldwide are becoming more degraded and fragmented with increased human impacts (Hoegh-Guldberg 1999, Haddad et al. 2015). Coral reefs in particular are gradually dying off and those that remain are becoming less complex, less suitable habitats for the populations they sustain (Alvarez-Filip et al. 2009). The species which dwell on these habitats have evolved to utilize the habitats as they existed in the past. As such these anthropogenic effects will serve to reduce the overall habitat quality as well as making habitats much more heterogeneous and patchier (Wilson et al. 2008b, Fontúrbel et al. 2015). Our findings suggest that as the short-term environment trends toward creating more heterogeneous habitats there will be a concomitant decline in the average distance organisms tend to disperse with overall smaller dispersal spreads. Conservation designs which rely on interconnected networks of habitats (e.g. marine protected area networks; Planes et al. 2009) which were initially placed close enough together to exchange migrants within the network may become increasingly isolated and less efficient at serving their conservation purpose (Green et al. 2014, Beltrán et al. 2017).

Appendix

Sampling Distribution

$$d_u \sim G(\alpha_u, \beta_u)$$

$$\alpha_u = \phi_u^{-1}$$

$$\beta_u = \frac{\phi_u^{-1}}{e^{\mu_u}}$$

$$\mu_u \sim N(0,1)$$

$$\phi_u^{-1} \sim G(2,1)$$

Weighting

$$w_k = \frac{1}{G(d_{r_k} | \alpha_u, \beta_u)}$$

$$w_k = W \frac{w_k}{\sum w_k}$$

$$W = \frac{N_u}{N_r}$$

Dispersal Distribution

$$d_{r_{f_k}} \sim w_k * G(\alpha_{r_{f_k}}, \beta_{r_{f_k}})$$

$$\alpha_{r_f} = \phi_f^{-1}$$

$$\beta_{r_f} = \frac{\phi_f^{-1}}{e^{\mu_f}}$$

$$\mu_{r_f} \sim N(\mu_r, \sigma_\mu)$$

$$\log(\phi_{r_f}^{-1}) \sim N((\log \phi_r^{-1}), \sigma_{\phi^{-1}})$$

$$\mu_r \sim N(0,1)$$

$$\phi_r^{-1} \sim G(2,1)$$

$$\sigma_\mu \sim G(1,3)$$

$$\sigma_{\phi^{-1}} \sim G(1,3)$$

Acknowledgements

We would like to thank the people at the University of Belize, Calabash Caye Field Station, for providing field logistics without which this research would not have been possible. We would also like to thank Maili Allende for assistance with lab work. Finally, we would like to thank Drs. David Portnoy, Blair Sterba-Boatwright, and Lynette Strickland for helpful comments to

improve the manuscript. We would like to thank funding sources: the Carl R. Beaver Memorial Scholarship, the Karen Koester Dodson Memorial Fund Grant, and Grants-in-Aid of Graduate Student Research Award by Texas Sea Grant College Program awarded to JDS, the Texas Research Development Fund award to JDH and funds provided by TAMUCC to JDH.

CHAPTER V: CONCLUSIONS

The sister species of *Coryphopterus hyalinus* and *C. personatus* provide an interesting study system for gaining insight into the biological functions at the interface of evolutionary and ecological processes. When considering evolutionary processes, we see two distinct species which maintain this species boundary despite living together in shoals (Böhlke and Robins 1962, Baldwin et al. 2009, Selwyn et al. 2022, Chapter 2). These species live well within the range of sympatric association despite the extremely small neighborhood sizes characterized by their dispersal kernels (Wright 1946, Selwyn et al. 2016, in prep, Chapter 4). While there is shared occupancy of individual shoals, we find evidence of diversifying niche space, in the context of differential habitat usage, between these two sister species which recapitulates previous macro-scale observations of *C. hyalinus* associating with deeper reef walls, while *C. personatus* is typically associated with shallower reef flats and back-reefs (Victor 2019, Selwyn et al. in review, Chapter 3).

Scale has always been an important component at the heart of both evolution and ecology (Sale 1998). Traditionally, evolutionary ecology has treated independently evolving populations as the finest scale of study, with variation within populations treated as noise from which the signal of different populations is extracted (Wright 1931). However, populations are composed of many interconnected families (not in the Linnean sense) of individuals, upon which selection acts. Through interbreeding these families are not evolutionarily independent but are the direct link between the evolutionary processes acting upon individuals (for example driving changes in individual dispersal behavior) and the populations which integrates the evolutionary effects across all families within the population.

The goal of this dissertation has been to address questions on the ecological and evolutionary processes of two sister species at the finest scale possible. Chapter 2 seeks to understand the fine-scale distribution of two sister species within a single reef, and how they maintain species boundaries, despite being extremely similar ecologically and living well within the range of what is typically thought of as sympatry. Meanwhile, chapter 3 endeavors to view the reef at a scale relevant to *C. hyalinus/personatus* and explain what features of the reef provide better habitat. Finally, chapter 4 uses the individual relationships between pairs of individuals, combining these into families, and an understanding of what habitat variability means from the perspective of the study species, to develop our understanding of why individuals disperse across the landscape into unknown habitats.

The history of the gains made in science can be, at least in part, attributed to changes in, and developments of technology. The findings of this dissertation are no exception. Seeking to address questions at the scale I have been interested in, relevant to the species of study, and truly at the interface of ecological and evolutionary processes would have been impossible without recent technological developments in fields as wide ranging as genomic sequencing, bioinformatics, geospatial science, photogrammetry, and statistics.

REFERENCES

- Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26:229–246.
- Ahnelt, H., M. N. Dawson, and D. K. Jacobs. 2004. Geographical variation in the cephalic lateral line canals of *Eucyclogobius newberryi* (Teleostei, Gobiidae) and its comparison with molecular phylogeography. *Folia Zoologica* 53:358–398.
- Ainsworth, A., and D. R. Drake. 2020. Classifying Hawaiian plant species along a habitat generalist-specialist continuum: Implications for species conservation under climate change. *PLOS ONE* 15:e0228573.
- Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655–1664.
- Allsop, D. J., and S. A. West. 2004. Sex allocation in the sex-changing marine goby, *Coryphopterus personatus*, on atoll-fringing reefs. *Evolutionary Ecology Research* 6:843–855.
- Almany, G. R. 2004. Does increased habitat complexity reduce predation and competition in coral reef fish assemblages? *Oikos* 106:275–284.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.

- Alvarez-Filip, L., N. K. Dulvy, J. A. Gill, I. M. Côté, and A. R. Watkinson. 2009. Flattening of Caribbean coral reefs: region-wide declines in architectural complexity. *Proceedings of the Royal Society of London B: Biological Sciences* 276:3019–3025.
- Anderson, E. C., and E. A. Thompson. 2002. A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics* 160:1217–1229.
- Anderson, R. A. 2007. Food acquisition modes and habitat use in lizards: questions from an integrative perspective. Pages 450–490 in S. M. Reilly, L. B. McBrayer, and D. B. Miles, editors. *Lizard Ecology*. Cambridge University Press, Cambridge.
- Appeldoorn, R. S., A. Friedlander, J. S. Nowlis, P. Usseglio, and A. Mitchell-Chui. 2003. Habitat connectivity in reef fish communities and marine reserve design in Old Providence-Santa Catalina, Colombia. *Gulf and Caribbean Research* 14:61–77.
- Archer, F. I., P. E. Adams, and B. B. Schneiders. 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* 17:5–11.
- Armstrong, P. R., M. K. James, and L. Bode. 2001. When to Press On or Turn Back: Dispersal Strategies for Reef Fish Larvae. *The American Naturalist* 157:434–450.
- Aspinwall, N. 1974. Genetic Analysis of North American Populations of the Pink Salmon, *Oncorhynchus gorbuscha*, Possible Evidence for the Neutral Mutation-Random Drift Hypothesis. *Evolution* 28:295–305.
- Atema, J., M. J. Kingsford, and G. Gerlach. 2002. Larval reef fish could use odour for detection, retention and orientation to reefs. *Marine Ecology Progress Series* 241:151–160.

- Babcock, R. C., C. N. Mundy, and D. Whitehead. 1994. Sperm Diffusion Models and In Situ Confirmation of Long-Distance Fertilization in the Free-Spawning Asteroid *Acanthaster planci*. *The Biological Bulletin* 186:17–28.
- Bachl, F. E., F. Lindgren, D. L. Borchers, and J. B. Illian. 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution* 10:760–766.
- Baddeley, A., R. Turner, and E. Rubak. 2021. spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests.
- Baker, M. B., and S. Rao. 2004. Incremental Costs and Benefits Shape Natal Dispersal: Theory and Example with *Hemilepistus reaumuri*. *Ecology* 85:1039–1051.
- Baldwin, C. C., and D. R. Robertson. 2015. A new, mesophotic *Coryphopterus* goby (Teleostei, Gobiidae) from the southern Caribbean, with comments on relationships and depth distributions within the genus. *ZooKeys* 513:123–142.
- Baldwin, C. C., L. A. Weigt, D. G. Smith, and J. H. Mounts. 2009. Reconciling genetic lineages with species in western Atlantic *Coryphopterus* (Teleostei: Gobiidae). *Smithsonian Contributions to the Marine Sciences* 38:111–138.
- Bandelt, H.-J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37–48.
- Beeken, N. S., J. D. Selwyn, and J. D. Hogan. 2021. Determining the life history strategy of the cryptobenthic reef gobies *Coryphopterus hyalinus* and *C. personatus*. *Marine Ecology Progress Series* 659:161–173.

- Beguin, J., S. Martino, H. Rue, and S. G. Cumming. 2012. Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution* 3:921–929.
- Beltrán, D. M., N. V. Schizas, R. S. Appeldoorn, and C. Prada. 2017. Effective Dispersal of Caribbean Reef Fish is Smaller than Current Spacing Among Marine Protected Areas. *Scientific Reports* 7:4689.
- Berenshtein, I., C. B. Paris, H. Gildor, E. Fredj, Y. Amitai, and M. Kiflawi. 2018. Biophysical Simulations Support Schooling Behavior of Fish Larvae Throughout Ontogeny. *Frontiers in Marine Science* 5.
- Bernardi, G., and A. Vagelli. 2004. Population structure in Banggai cardinalfish, *Pterapogon kauderni*, a coral reef species lacking a pelagic larval phase. *Marine Biology* 145:803–810.
- Biesack, E. E., B. T. Dang, A. S. Ackiss, C. E. Bird, P. Chheng, L. Phounvisouk, O. T. Truong, and K. E. Carpenter. 2020. Evidence for population genetic structure in two exploited Mekong River fishes across a natural riverine barrier. *Journal of Fish Biology* 97:696–707.
- Biggs, C. R., and J. D. Olden. 2011. Multi-scale habitat occupancy of invasive lionfish (*Pterois volitans*) in coral reef environments of Roatan, Honduras. *Aquatic Invasions* 6:347–353.
- Bird, C. E. 2020. dDocentHPC. A pipeline to trim, assemble, map, and genotype reduced representation genomic data. <https://github.com/cbirdlab/dDocentHPC>.
- Bird, C. E., and J. D. Selwyn. 2021. fltrVCF v4.4: A script to filter VCF files and simplify reporting and reproduction of filters. <https://github.com/cbirdlab/fltrVCF>.

- Birnbaum, Z. W. 1948. On Random Variables with Comparable Peakedness. *The Annals of Mathematical Statistics* 19:76–81.
- Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008:P10008.
- Bodenhofer, U., E. Bonatesta, C. Horejš-Kainrath, and S. Hochreiter. 2015. msa: an R package for multiple sequence alignment. *Bioinformatics* 31:3997–3999.
- Böhlke, J. E., and C. R. Robins. 1962. The Taxonomic Position of the West Atlantic Goby, *Eviota personata*, with Descriptions of Two New Related Species. *Proceedings of the Academy of Natural Sciences of Philadelphia* 114:175–189.
- Bolnick, D. I., and T. J. Near. 2005. Tempo of Hybrid Inviability in Centrarchid Fishes (Teleostei: Centrarchidae). *Evolution* 59:1754–1767.
- Bongaerts, P., C. E. Dubé, K. E. Prata, J. C. Gijsbers, M. Achlatis, and A. Hernandez-Agreda. 2021. Reefscape Genomics: Leveraging Advances in 3D Imaging to Assess Fine-Scale Patterns of Genomic Variation on Coral Reefs. *Frontiers in Marine Science* 8.
- Bonte, D., H. Van Dyck, J. M. Bullock, A. Coulon, M. Delgado, M. Gibbs, V. Lehouck, E. Matthysen, K. Mustin, M. Saastamoinen, N. Schtickzelle, V. M. Stevens, S. Vandewoestijne, M. Baguette, K. Barton, T. G. Benton, A. Chaput-Bardy, J. Clobert, C. Dytham, T. Hovestadt, C. M. Meier, S. C. F. Palmer, C. Turlure, and J. M. J. Travis. 2012. Costs of dispersal. *Biological Reviews* 87:290–312.
- Bouwmeester, J., A. J. Edwards, J. R. Guest, A. G. Bauman, M. L. Berumen, and A. H. Baird. 2021. Latitudinal variation in monthly-scale reproductive synchrony among *Acropora* coral assemblages in the Indo-Pacific. *Coral Reefs*.

- Bovbjerg, R. V. 1970. Ecological Isolation and Competitive Exclusion in Two Crayfish (*Orconectes virilis* and *Orconectes immunis*). *Ecology* 51:225–236.
- Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews* 80:205–225.
- Bradbury, I. R., B. Laurel, P. V. Snelgrove, P. Bentzen, and S. E. Campana. 2008. Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society B: Biological Sciences* 275:1803–1809.
- Brandl, S. J., C. H. R. Goatley, D. R. Bellwood, and L. Tornabene. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. *Biological Reviews* 93:1846–1873.
- Brandl, S. J., L. Tornabene, C. H. R. Goatley, J. M. Casey, R. A. Morais, I. M. Côté, C. C. Baldwin, V. Parravicini, N. M. D. Schiettekatte, and D. R. Bellwood. 2019. Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science*.
- Burgess, S. C., M. Bode, J. M. Leis, and L. B. Mason. 2022. Individual variation in marine larval-fish swimming speed and the emergence of dispersal kernels. *Oikos* 2022:e08896.
- Burgess, S. C., R. E. Snyder, and B. Rountree. 2017. Collective Dispersal Leads to Variance in Fitness and Maintains Offspring Size Variation within Marine Populations. *The American Naturalist*:000–000.
- Bürkner, P.-C. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10:395–411.

- Burns, J. H. R., D. Delparte, R. D. Gates, and M. Takabayashi. 2015. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. *PeerJ* 3:e1077.
- Burt, J., A. Bartholomew, P. Usseglio, A. Bauman, and P. F. Sale. 2009. Are artificial reefs surrogates of natural habitats for corals and fish in Dubai, United Arab Emirates? *Coral Reefs* 28:663–675.
- Buston, P. M., C. Fauvelot, M. Y. Wong, and S. Planes. 2009. Genetic relatedness in groups of the humbug damselfish *Dascyllus aruanus*: small, similar-sized individuals may be close kin. *Molecular Ecology* 18:4707–4715.
- Buston, P. M., G. P. Jones, S. Planes, and S. R. Thorrold. 2012. Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proceedings of the Royal Society B: Biological Sciences* 279:1883–1888.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan : A Probabilistic Programming Language. *Journal of Statistical Software* 76.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* 1:171–182.
- Chagaris, D., S. Binion-Rock, A. Bogdanoff, K. Dahl, J. Granneman, H. Harris, J. Mohan, M. B. Rudd, M. K. Swenarton, R. Ahrens, W. F. Patterson III, J. A. Morris Jr, and M. Allen. 2017. An Ecosystem-Based Approach to Evaluating Impacts and Management of Invasive Lionfish. *Fisheries* 42:421–431.

- Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.
- Cheptou, P.-O., O. Carrue, S. Rouifed, and A. Cantarel. 2008. Rapid evolution of seed dispersal in an urban environment in the weed *Crepis sancta*. *Proceedings of the National Academy of Sciences* 105:3796–3799.
- Chong, Z., J. Ruan, and C.-I. Wu. 2012. Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* 28:2732–2737.
- Cirtwill, A. R., and A. Eklöf. 2018. Data from: Feeding environment and other traits shape species' roles in marine food webs. Dryad.
- Clauset, A., M. E. J. Newman, and C. Moore. 2004. Finding community structure in very large networks. *Physical Review E* 70:066111.
- Clobert, J., E. Danchin, A. A. Dhondt, and J. D. Nichols. 2001. *Dispersal*. 1 edition. Oxford University Press, Oxford.
- Cole, K. S., and D. R. Robertson. 1988. Protogyny in the Caribbean reef goby, *Coryphopterus personatus*: gonad ontogeny and social influences on sex-change. *Bulletin of marine science* 42:317–333.
- Cole, K. S., and D. Y. Shapiro. 1990. Gonad Structure and Hermaphroditism in the Gobiid Genus *Coryphopterus* (Teleostei: Gobiidae). *Copeia* 1990:996–1003.
- Colin, P. L. 2010. Fishes as living tracers of connectivity in the tropical western North Atlantic: I. Distribution of the neon gobies, genus *Elacatinus* (Pisces: Gobiidae). *Zootaxa* 2370:36–52.
- Cooley, J. R. 2007. Decoding Asymmetries in Reproductive Character Displacement. *Proceedings of the Academy of Natural Sciences of Philadelphia* 156:89–96.

- Cowen, R. K., and S. Sponaugle. 2009. Larval dispersal and marine population connectivity. *Annual Review of Marine Science* 1:443–466.
- Coyne, J. A. 1974. The Evolutionary Origin of Hybrid Inviability. *Evolution* 28:505–506.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. 1 edition. Sinauer Associates, Sunderland, Massachusetts, USA.
- Crean, A. J., and D. J. Marshall. 2009. Coping with environmental uncertainty: dynamic bet hedging as a maternal effect. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364:1087–1096.
- D’Aloia, C. C., S. M. Bogdanowicz, R. K. Francis, J. E. Majoris, R. G. Harrison, and P. M. Buston. 2015. Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences* 112:13940–13945.
- D’Aloia, C. C., S. M. Bogdanowicz, J. E. Majoris, R. G. Harrison, and P. M. Buston. 2013. Self-recruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape. *Molecular Ecology* 22:2563–2572.
- D’Aloia, C. C., J. E. Majoris, and P. M. Buston. 2011. Predictors of the distribution and abundance of a tube sponge and its resident goby. *Coral Reefs* 30:777–786.
- D’Aloia, C. C., and M. G. Neubert. 2018. The formation of marine kin structure: effects of dispersal, larval cohesion, and variable reproductive success. *Ecology* 99:2374–2384.
- D’Aloia, C. C., A. Xuereb, M.-J. Fortin, S. M. Bogdanowicz, and P. M. Buston. 2018. Limited dispersal explains the spatial distribution of siblings in a reef fish population. *Marine Ecology Progress Series* 607:143–154.

- Dapporto, L., and R. L. H. Dennis. 2013. The generalist–specialist continuum: Testing predictions for distribution and trends in British butterflies. *Biological Conservation* 157:229–236.
- Diabaté, A., A. Dao, A. S. Yaro, A. Adamou, R. Gonzalez, N. C. Manoukis, S. F. Traoré, R. W. Gwadz, and T. Lehmann. 2009. Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proceedings of the Royal Society B: Biological Sciences* 276:4215–4222.
- Dixson, D. L., D. Abrego, and M. E. Hay. 2014. Chemically mediated behavior of recruiting corals and fishes: a tipping point that may limit reef recovery. *Science* 345:892–897.
- Dobzhansky, T. G. 1936. Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila pseudoobscura* Hybrids. *Genetics* 21:113–135.
- Dobzhansky, T. G. 1940. Speciation as a Stage in Evolutionary Divergence. *The American Naturalist* 74:312–321.
- Domeier, M. L., and P. L. Colin. 1997. Tropical Reef Fish Spawning Aggregations: Defined and Reviewed. *Bulletin of Marine Science* 60:698–726.
- Dominici-Arosemena, A., and M. Wolff. 2005. Reef fish community structure in Bocas del Toro (Caribbean, Panama): gradients in habitat complexity and exposure. *Caribbean Journal of Science* 41:613–637.
- Duffy, J. E. 1996. Eusociality in a coral-reef shrimp. *Nature* 381:512–514.
- Duputié, A., and F. Massol. 2013. An empiricist’s guide to theoretical predictions on the evolution of dispersal. *Interface Focus* 3:20130028.
- Dziminski, M. A., P. E. Vercoe, and J. D. Roberts. 2009. Variable Offspring Provisioning and Fitness: A Direct Test in the Field. *Functional Ecology* 23:164–171.

- Elkin, C., and D. J. Marshall. 2007. Desperate larvae: influence of deferred costs and habitat requirements on habitat selection. *Marine Ecology Progress Series* 335:143–153.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611–2620.
- Evans, J. L., J. A. Thia, C. Riginos, and J. P. Hereward. 2018. The complete mitochondrial genome of *Bathygobius cocosensis* (Perciformes, Gobiidae). *Mitochondrial DNA Part B* 3:217–219.
- Ezer, T., D. V. Thattai, B. Kjerfve, and W. D. Heyman. 2005. On the variability of the flow along the Meso-American Barrier Reef system: a numerical model study of the influence of the Caribbean current and eddies. *Ocean Dynamics* 55:458–475.
- Faircloth, B. C., and T. C. Glenn. 2014. Protocol: preparation of an AMPure XP substitute (AKA Serapure). Online at 10:j9mw2f26.
- Feary, D. A., and K. D. Clements. 2006. Habitat use by triplefin species (*Tripterygiidae*) on rocky reefs in New Zealand. *Journal of Fish Biology* 69:1031–1046.
- Filipović, I., H. C. Hapuarachchi, W.-P. Tien, M. A. B. A. Razak, C. Lee, C. H. Tan, G. J. Devine, and G. Rašić. 2020. Using spatial genetics to quantify mosquito dispersal for control programs. *BMC Biology* 18:104.
- Fisher, S., A. Barry, J. Abreu, B. Minie, J. Nolan, T. M. Delorey, G. Young, T. J. Fennell, A. Allen, and L. Ambrogio. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome biology* 12:R1.
- Fontúrbel, F. E., A. B. Candia, J. Malebrán, D. A. Salazar, C. González-Browne, and R. Medel. 2015. Meta-analysis of anthropogenic habitat disturbance effects on animal-mediated seed dispersal. *Global Change Biology* 21:3951–3960.

- Fretwell, S. D., and H. L. Lucas. 1969. On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica* 19:16–36.
- Friberg, M., O. Leimar, and C. Wiklund. 2013. Heterospecific courtship, minority effects and niche separation between cryptic butterfly species. *Journal of Evolutionary Biology* 26:971–979.
- Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Gagnaire, P.-A., S. A. Pavey, E. Normandeau, and L. Bernatchez. 2013. The Genetic Architecture of Reproductive Isolation During Speciation-with-Gene-Flow in Lake Whitefish Species Pairs Assessed by Rad Sequencing. *Evolution* 67:2483–2497.
- Garcia, E., and K. Holtermann. 1998. Calabash Caye, Turneffe Islands Atoll, Belize. Caricomp: Caribbean coral reef, seagrass and mangrove sites. Coastal region and small islands unit. Paris: UNESCO 347.
- Garcia-Sais, J. R. 2010. Reef habitats and associated sessile-benthic and fish assemblages across a euphotic–mesophotic depth gradient in Isla Desecheo, Puerto Rico. *Coral Reefs* 29:277–288.
- Gardner, T. 2000. Breeding the marine goby *Coryphopterus personatus* in Aquaria. http://gobiidae.com/breeding_c_pers.htm.
- Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian Data Analysis, Third Edition. 3 edition. Chapman and Hall/CRC, Boca Raton.

- Gelman, A., B. Goodrich, J. Gabry, and A. Vehtari. 2019. R-squared for Bayesian Regression Models. *The American Statistician* 73:307–309.
- Gelman, A., A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. 2020. Bayesian Workflow. arXiv:2011.01808 [stat].
- Gerhardt, H. C. 1974. Behavioral Isolation of the Tree Frogs, *Hyla cinerea* and *Hyla andersonii*. *The American Midland Naturalist* 91:424–433.
- Gerlach, G., J. Atema, M. J. Kingsford, K. P. Black, and V. Miller-Sims. 2007. Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the National Academy of Sciences* 104:858–863.
- Gibson, R. M., A. S. Aspbury, and L. L. McDaniel. 2002. Active formation of mixed-species grouse leks: a role for predation in lek evolution? *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269:2503–2507.
- Glynn, P. W. 1973. Ecology of a Caribbean coral reef. The *Porites* reef-flat biotope: Part II. Plankton community with evidence for depletion. *Marine Biology* 22:1–21.
- Goatley, C. H. R., and D. R. Bellwood. 2016. Body size and mortality rates in coral reef fishes: a three-phase relationship. *Proceedings of the Royal Society B: Biological Sciences* 283:20161858.
- González-Rivero, M., A. R. Harborne, A. Herrera-Reveles, Y.-M. Bozec, A. Rogers, A. Friedman, A. Ganase, and O. Hoegh-Guldberg. 2017. Linking fishes to multiple metrics of coral reef structural complexity using three-dimensional technology. *Scientific Reports* 7:13965.

- Goodnight, C. J. 2000. Modeling gene interaction in structured populations. Pages 129–145 in J. B. Wolf, E. D. Brodie, and M. J. Wade, editors. *Epistasis and the evolutionary process*. Oxford University Press, Oxford, UK.
- Goudet, J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184–186.
- Green, A. L., A. P. Maypa, G. R. Almany, K. L. Rhodes, R. Weeks, R. A. Abesamis, M. G. Gleason, P. J. Mumby, and A. T. White. 2014. Larval dispersal and movement patterns of coral reef fishes, and implications for marine reserve network design. *Biological Reviews*.
- Green, B. S., and M. I. McCormick. 2005. Maternal and paternal effects determine size, growth and performance in larvae of a tropical reef fish. *Marine Ecology Progress Series* 289:263–272.
- Greenfield, D. W. 1972. Notes on the Biology of the Arrow Blenny, *Lucayablennius zingaro* (Böhlke) from British Honduras. *Copeia* 1972:590–592.
- Greenfield, D. W., and R. K. Johnson. 1999. Assemblage Structure and Habitat Associations of Western Caribbean Gobies (Teleostei: Gobiidae). *Copeia* 1999:251–266.
- Greig, N. 1993. Predispersal seed predation on five Piper species in tropical rainforest. *Oecologia* 93:412–420.
- Grinsted, L., I. Agnarsson, and T. Bilde. 2012. Subsocial behaviour and brood adoption in mixed-species colonies of two theridiid spiders. *Naturwissenschaften* 99:1021–1030.
- Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, W. M. Cook, E. I. Damschen, R. M. Ewers, B. L. Foster, C. N. Jenkins, A. J. King, W. F. Laurance, D. J. Levey, C. R. Margules, B. A.

- Melbourne, A. O. Nicholls, J. L. Orrock, D.-X. Song, and J. R. Townshend. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances* 1:e1500052.
- Hamilton, W. D. 1971. Geometry for the selfish herd. *Journal of Theoretical Biology* 31:295–311.
- Hamilton, W. D., and R. M. May. 1977. Dispersal in stable habitats. *Nature* 269:578–581.
- Hamner, W. M., M. S. Jones, J. H. Carleton, I. R. Hauri, and D. McB. Williams. 1988. Zooplankton, Planktivorous Fish, and Water Currents on a Windward Reef Face: Great Barrier Reef, Australia. *Bulletin of Marine Science* 42:459–479.
- Hannah, C. G., J. A. Shore, and J. W. Loder. 2000. The drift-retention dichotomy on Browns Bank: a model study of interannual variability. *Canadian Journal of Fisheries and Aquatic Sciences* 57:2506–2518.
- Hanski, I. 1991. Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society* 42:17–38.
- Hanski, I., and C. D. Thomas. 1994. Metapopulation dynamics and conservation: A spatially explicit model applied to butterflies. *Biological Conservation* 68:167–180.
- Harris, L. D. 1988. Edge Effects and Conservation of Biotic Diversity. *Conservation Biology* 2:330–332.
- Harrison, H. B., M. L. Berumen, P. Saenz-Agudelo, E. Salas, D. H. Williamson, and G. P. Jones. 2017. Widespread hybridization and bidirectional introgression in sympatric species of coral reef fish. *Molecular Ecology* 26:5692–5704.
- Harrison, R. G., and E. L. Larson. 2014. Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity* 105:795–809.

- Hastings, A. 1983. Can spatial variation alone lead to selection for dispersal? *Theoretical Population Biology* 24:244–251.
- Hedrick, P. W. 2005. A standardized genetic differentiation measure. *Evolution; International Journal of Organic Evolution* 59:1633–1638.
- Hendriks, I., D. Wilson, and M. Meekan. 2001. Vertical distributions of late stage larval fishes in the nearshore waters of the San Blas Archipelago, Caribbean Panama. *Coral Reefs* 20:77–84.
- Hepburn, R. I., E. P. Mottillo, P. Bentzen, and D. D. Heath. 2005. Polymorphic microsatellite loci for the masked goby, *Coryphopterus personatus* (Gobiidae). *Conservation Genetics* 6:1059–1062.
- Hijmans, R. J., R. Bivand, K. Forner, J. Ooms, and E. Pebesma. 2021a. terra: Spatial Data Analysis.
- Hijmans, R. J., J. van Etten, M. Sumner, J. Cheng, D. Baston, A. Bevan, R. Bivand, L. Busetto, M. Canty, B. Fasoli, D. Forrest, A. Ghosh, D. Golicher, J. Gray, J. A. Greenberg, P. Hiemstra, K. Hingee, I. for M. A. Geosciences, C. Karney, M. Mattiuzzi, S. Mosher, B. Naimi, J. Nowosad, E. Pebesma, O. P. Lamigueiro, E. B. Racine, B. Rowlingson, A. Shortridge, B. Venables, and R. Wueest. 2021b. raster: Geographic Data Analysis and Modeling.
- Hinsley, S. A. 2000. The costs of multiple patch use by birds. *Landscape Ecology* 15:765–775.
- Hoegh-Guldberg, O. 1999. Climate change, coral bleaching and the future of the world's coral reefs. *Marine and freshwater research* 50:839–866.
- Hogan, J. D., J. A. Galarza, R. P. Walter, and D. D. Heath. 2010. Bayesian analysis of population structure and the characterization of nine novel microsatellite markers for the study of a

- Caribbean coral reef Gobiid (*Coryphopterus personatus*) and related taxa. *Molecular Ecology Resources* 10:1–11.
- Hogan, J. D., and C. Mora. 2005. Experimental analysis of the contribution of swimming and drifting to the displacement of reef fish larvae. *Marine Biology* 147:1213–1220.
- Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6:65–70.
- Holt, R. D. 1985. Population dynamics in two-patch environments: Some anomalous consequences of an optimal habitat distribution. *Theoretical Population Biology* 28:181–208.
- Hope, A. C. A. 1968. A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society: Series B (Methodological)* 30:582–598.
- Houde, E. D. 1997. Patterns and trends in larval-stage growth and mortality of teleost fish*. *Journal of Fish Biology* 51:52–83.
- Hubbs, C. L. 1922. Variations in the Number of Vertebrae and Other Meristic Characters of Fishes Correlated with the Temperature of Water during Development. *The American Naturalist* 56:360–372.
- Hurlbert, S. H. 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* 52:577–586.
- Isbell, L. A. 2004. Is there no place like home? Ecological bases of female dispersal and philopatry and their consequences for the formation of kin groups. *Kinship and behavior in primates*:71–108.

- Ito, T., T. Fukuda, T. Morimune, and K. Hosoya. 2017. Evolution of the connection patterns of the cephalic lateral line canal system and its use to diagnose opsariichthyin cyprinid fishes (Teleostei, Cyprinidae). *ZooKeys*:115–131.
- Jacquard, A. 1974. Genetic Information Given by a Relative. *Biometrics* 28:1101–1114.
- Jasper, M. E., A. A. Hoffmann, and T. L. Schmidt. 2021. Estimating dispersal using close kin dyads: The kindisperse R package. *Molecular Ecology Resources* n/a.
- Jeffreys, H. 1961. *Theory of Probability*. 3 edition. Oxford University Press, Oxford.
- Johannesen, J., and Y. Lubin. 2001. Evidence for kin-structured group founding and limited juvenile dispersal in the sub-social spider *Stegodyphus lineatus* (Araneae, Eresidae). *Journal of Arachnology* 29:413–422.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.
- Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11.
- Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. *Nature* 402:802–804.
- Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.
- Kingsford, M. J., J. M. Leis, A. Shanks, K. C. Lindeman, S. G. Morgan, and J. Pineda. 2002. Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science* 70:309–340.

- Knowlton, N., J. L. Maté, H. M. Guzmán, R. Rowan, and J. Jara. 1997. Direct evidence for reproductive isolation among the three species of the *Montastraea annularis* complex in Central America (Panamá and Honduras). *Marine Biology* 127:705–711.
- Komyakova, V., P. L. Munday, and G. P. Jones. 2013. Relative Importance of Coral Cover, Habitat Complexity and Diversity in Determining the Structure of Reef Fish Communities. *PLOS ONE* 8:e83178.
- Kopp, A., and A. K. Frank. 2005. Speciation in Progress? A Continuum of Reproductive Isolation in *Drosophila Bipectinata*. *Genetica* 125:55–68.
- Kramer, A., J. L. Van Tassell, and R. A. Patzner. 2009. Aspects of Spawning Behaviour in Five Gobiids of the Genus *Coryphopterus* (Pisces: Gobiidae) in the Caribbean Sea. *The Open Fish Science Journal* 2:50–54.
- Kristoffersen, J. B., and A. G. V. Salvanes. 1998. Effects of formaldehyde and ethanol preservation on body and otoliths of *Maurolicus muelleri* and *Benthosema glaciale*. *Sarsia* 83:95–102.
- Krug, P. J. 2001. Bet-hedging dispersal strategy of a specialist marine herbivore: a settlement dimorphism among sibling larvae of *Alderia modesta*. *Marine Ecology Progress Series* 213:177–192.
- Krug, P. J. 2009. Not My “Type”: Larval Dispersal Dimorphisms and Bet-Hedging in Opisthobranch Life Histories. *The Biological Bulletin* 216:355–372.
- Kuhn, M., and H. Wickham. 2020. Tidymodels: Easily Install and Load the ‘Tidymodels’ Packages. R package version 0.1. 0.

- de Lamballerie, X., C. Zandotti, C. Vignoli, C. Bollet, and P. de Micco. 1992. A one-step microbial DNA extraction method using “Chelex 100” suitable for gene amplification. *Research in Microbiology* 143:785–790.
- Lampert, A., and T. Tlustý. 2016. Where Two Are Fighting, the Third Wins: Stronger Selection Facilitates Greater Polymorphism in Traits Conferring Competition-Dispersal Tradeoffs. *PLOS ONE* 11:e0147970.
- Landeau, L., and J. Terborgh. 1986. Oddity and the “confusion effect” in predation. *Animal Behaviour* 34:1372–1380.
- Leal, M. C., P. N. Pochelon, T. L. da Silva, A. Reis, R. Rosa, and R. Calado. 2013. Variable within-brood maternal provisioning in newly extruded embryos of *Homarus gammarus*. *Marine Biology* 160:763–772.
- Leggett, W. C., and E. Deblois. 1994. Recruitment in marine fishes: Is it regulated by starvation and predation in the egg and larval stages? *Netherlands Journal of Sea Research* 32:119–134.
- Leigh, J. W., and D. Bryant. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6:1110–1116.
- Leis, J. M., U. Siebeck, and D. L. Dixon. 2011. How Nemo Finds Home: The Neuroecology of Dispersal and of Population Connectivity in Larvae of Marine Fishes. *Integrative and Comparative Biology* 51:826–843.
- Lesneski, K. C., C. C. D’Aloia, M.-J. Fortin, and P. M. Buston. 2019. Disentangling the spatial distributions of a sponge-dwelling fish and its host sponge. *Marine Biology* 166:66.
- Leutner, B., N. Horning, J. Schwalb-Willmann, and R. J. Hijmans. 2019. RStoolbox: Tools for Remote Sensing Data Analysis.

- Levin, L. A. 2006. Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology* 46:282–297.
- Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology* 73:1943–1967.
- Levitan, D. R., N. D. Fogarty, J. Jara, K. E. Lotterhos, and N. Knowlton. 2011. Genetic, Spatial, and Temporal Components of Precise Spawning Synchrony in Reef Building Corals of the *Montastraea annularis* Species Complex. *Evolution* 65:1254–1270.
- Levitan, D. R., H. Fukami, J. Jara, D. Kline, T. M. McGovern, K. E. McGhee, C. A. Swanson, and N. Knowlton. 2004. Mechanisms of Reproductive Isolation Among Sympatric Broadcast-Spawning Corals of the *Montastraea annularis* Species Complex. *Evolution* 58:308–323.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595.
- Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
- Liebers, D., P. de Knijff, and A. J. Helbig. 2004. The herring gull complex is not a ring species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271:893–901.
- Lindsay, J. B. 2019. WhiteboxTools User Manual.

- Liou, L. W., and T. D. Price. 1994. Speciation by Reinforcement of Premating Isolation. *Evolution* 48:1451–1459.
- Losos, J. B. 2010. Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism. *The American Naturalist* 175:623–639.
- Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20:229–237.
- Martinez, P. A., W. M. Berbel-Filho, and U. P. Jacobina. 2013. Is formalin fixation and ethanol preservation able to influence in geometric morphometric analysis? Fishes as a case study. *Zoomorphology* 132:87–93.
- Massol, F., and F. Débarre. 2015. Evolution of dispersal in spatially and temporally variable environments: The importance of life cycles. *Evolution* 69:1925–1937.
- McAdam, A. G., S. Boutin, D. Réale, and D. Berteaux. 2002. Maternal Effects and the Potential for Evolution in a Natural Population of Animals. *Evolution* 56:846–851.
- McKay, S. I., and P. J. Miller. 1997. The affinities of European sand gobies (Teleostei: Gobiidae). *Journal of Natural History* 31:1457–1482.
- McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12:153–157.
- McPeck, M. A., and R. D. Holt. 1992. The Evolution of Dispersal in Spatially and Temporally Varying Environments. *The American Naturalist* 140:1010–1027.
- Milinski, M., and T. C. M. Bakker. 1992. Costs influences sequential mate choice in sticklebacks, *Gasterosteus aculeatus*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 250:229–233.
- Milligan, B. G. 2003. Maximum-likelihood estimation of relatedness. *Genetics* 163:1153–1167.

- Miranda, M. B. B., D. J. Innes, and R. J. Thompson. 2010. Incomplete Reproductive Isolation in the Blue Mussel (*Mytilus edulis* and *M. trossulus*) Hybrid Zone in the Northwest Atlantic: Role of Gamete Interactions and Larval Viability. *The Biological Bulletin* 218:266–281.
- Moilanen, A., and I. Hanski. 1998. Metapopulation Dynamics: Effects of Habitat Quality and Landscape Structure. *Ecology* 79:2503–2515.
- Møller, J., A. R. Syversveen, and R. P. Waagepetersen. 1998. Log Gaussian Cox Processes. *Scandinavian Journal of Statistics* 25:451–482.
- Montanari, S. R., L. van Herwerden, M. S. Pratchett, J.-P. A. Hobbs, and A. Fugedi. 2012. Reef fish hybridization: lessons learnt from butterflyfishes (genus *Chaetodon*). *Ecology and Evolution* 2:310–328.
- Montgomery, J. C., N. Tolimieri, and O. S. Haine. 2001. Active habitat selection by pre-settlement reef fishes. *Fish and Fisheries* 2:261–277.
- Moran, P. A. P. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika* 37:17–23.
- Moran, R. L., J. M. Catchen, and R. C. Fuller. 2019. Genomic Resources for Darters (Percidae: Etheostominae) Provide Insight into Postzygotic Barriers Implicated in Speciation. *Molecular Biology and Evolution*.
- Morrison, J. 2013. Characterization and Correction of Error in Genome-Wide IBD Estimation for Samples with Population Structure. *Genetic epidemiology* 37:635–641.
- Muhlfeld, C. C., T. E. McMahon, M. C. Boyer, and R. E. Gresswell. 2009. Local Habitat, Watershed, and Biotic Factors Influencing the Spread of Hybridization between Native Westslope Cutthroat Trout and Introduced Rainbow Trout. *Transactions of the American Fisheries Society* 138:1036–1051.

- Munk, P., T. Kiørboe, and V. Christensen. 1989. Vertical migrations of herring, *Clupea harengus*, larvae in relation to light and prey distribution. *Environmental Biology of Fishes* 26:87–96.
- Nanninga, G. B., and M. L. Berumen. 2014. The role of individual variation in marine larval dispersal. *Frontiers in Marine Science* 1:1–17.
- Nathan, R., E. Klein, J. J. Robledo-Arnuncio, and E. Revilla. 2012. Dispersal kernels: review. *Dispersal Ecology and Evolution*. Oxford University Press. Oxford. pp:187–210.
- Nathan, R., and H. C. Muller-Landau. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution* 15:278–285.
- Nei, M. 1973. Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences* 70:3321–3323.
- Nei, M., and R. K. Chesser. 1983. Estimation of fixation indices and gene diversities. *Annals of Human Genetics* 47:253–259.
- Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. 1 edition. Oxford University Press, Oxford, UK.
- Nelson, J. S., T. C. Grande, and M. V. H. Wilson. 2016. *Fishes of the World*. John Wiley & Sons.
- Oakley-Cogan, A., S. B. Tebbett, and D. R. Bellwood. 2020. Habitat zonation on coral reefs: Structural complexity, nutritional resources and herbivorous fish distributions. *PLOS ONE* 15:e0233498.
- Ohki, S., R. K. Kowalski, S. Kitanobo, and M. Morita. 2015. Changes in spawning time led to the speciation of the broadcast spawning corals *Acropora digitifera* and the cryptic

- species *Acropora* sp. 1 with similar gamete recognition systems. *Coral Reefs* 34:1189–1198.
- Öhlund, G., M. Bodin, K. A. Nilsson, S.-O. Öhlund, K. B. Mobley, A. G. Hudson, M. Peedu, Å. Brännström, P. Bartels, K. Præbel, C. L. Hein, P. Johansson, and G. Englund. 2020. Ecological speciation in European whitefish is driven by a large-gaped predator. *Evolution Letters* 4:243–256.
- O’Leary, S. J., J. B. Puritz, S. C. Willis, C. M. Hollenbeck, and D. S. Portnoy. 2018. These aren’t the loci you’re looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology* 27:3193–3206.
- Opitz, S. 1996. Trophic interactions in Caribbean coral reefs. *WorldFish*.
- Ortiz-Barrientos, D., B. A. Counterman, and M. A. F. Noor. 2004. The Genetics of Speciation by Reinforcement. *PLoS Biology* 2:e416.
- Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics* 25:547–572.
- Paradis, E. 2010. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26:419–420.
- Parchman, T. L., Z. Gompert, M. J. Braun, R. T. Brumfield, D. B. McDonald, J. a. C. Uy, G. Zhang, E. D. Jarvis, B. A. Schlinger, and C. A. Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology* 22:3304–3317.
- Parrish, J. K. 1989. Layering with depth in a heterospecific fish aggregation. *Environmental Biology of Fishes* 26:79–85.

- Part, T. 1991. Philopatry Pays: A Comparison between Collared Flycatcher Sisters. *The American Naturalist* 138:790–796.
- Pavlov, D. S., and A. O. Kasumyan. 2000. Patterns and Mechanisms of Schooling Behavior in Fish: A Review. *Journal of Ichthyology* 40:S163–S231.
- Pebesma, E., R. Bivand, E. Racine, M. Sumner, I. Cook, T. Keitt, R. Lovelace, H. Wickham, J. Ooms, K. Müller, T. L. Pedersen, D. Baston, and D. Dunnington. 2021. *sf: Simple Features for R*.
- Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLOS One* 7:e37135.
- Pham-Gia, T., N. Turkkan, and P. Eng. 1993. Bayesian analysis of the difference of two proportions. *Communications in Statistics - Theory and Methods* 22:1755–1771.
- Pineda, J., J. Hare, and S. Sponaugle. 2007. Larval Transport and Dispersal in the Coastal Ocean and Consequences for Population Connectivity. *Oceanography* 20:22–39.
- Pinsky, M. L., S. R. Palumbi, S. Andréfouët, and S. J. Purkis. 2012. Open and closed seascapes: Where does habitat patchiness create populations with high fractions of self-recruitment? *Ecological Applications* 22:1257–1267.
- Pinsky, M. L., P. Saenz-Agudelo, O. C. Salles, G. R. Almany, M. Bode, M. L. Berumen, S. Andréfouët, S. R. Thorrold, G. P. Jones, and S. Planes. 2017. Marine Dispersal Scales Are Congruent over Evolutionary and Ecological Time. *Current Biology* 27:149–154.
- Pittman, S. J., B. M. Costa, and T. A. Battista. 2009. Using Lidar Bathymetry and Boosted Regression Trees to Predict the Diversity and Abundance of Fish and Corals. *Journal of Coastal Research*:27–38.

- Planes, S., and P. J. Doherty. 1997. Genetic and color interactions at a contact zone of *Acanthochromis polyacanthus*: a marine fish lacking pelagic larvae. *Evolution* 51:1232–1243.
- Planes, S., G. P. Jones, and S. R. Thorrold. 2009. Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences* 106:5693–5697.
- Planes, S., G. Lecaillon, P. Lenfant, and M. Meekan. 2002. Genetic and demographic variation in new recruits of *Naso unicornis*. *Journal of Fish Biology* 61:1033–1049.
- Polgar, G., A. Sacchetti, and P. Galli. 2010. Differentiation and adaptive radiation of amphibious gobies (Gobiidae: Oxudercinae) in semi-terrestrial habitats. *Journal of Fish Biology* 77:1645–1664.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945–959.
- Proschan, F. 1965. Peakedness of distributions of convex combinations. *The Annals of Mathematical Statistics* 36:1703–1706.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* 81:559–575.
- Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2:e431.
- Qiusheng Wu. 2020. whitebox: “WhiteboxTools” R Frontend.

- R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raimondi, P. T., and M. J. Keough. 1990. Behavioural variability in marine larvae. *Australian Journal of Ecology* 15:427–437.
- Randall, J. E. 1967. Food habits of reef fishes of the West Indies. Institute of Marine Sciences, University of Miami.
- Rawson, P. D., C. Slaughter, and P. O. Yund. 2003. Patterns of gamete incompatibility between the blue mussels *Mytilus edulis* and *M. trossulus*. *Marine Biology* 143:317–325.
- Rice, W. R. 1987. Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology* 1:301–314.
- Riginos, C., Y. M. Buckley, S. P. Blomberg, and E. A. Treml. 2014. Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *The American Naturalist* 184:52–64.
- Ripley, B. D. 1977. Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)* 39:172–192.
- Riquet, F., T. Comtet, T. Broquet, and F. Viard. 2017. Unexpected collective larval dispersal but little support for sweepstakes reproductive success in the highly dispersive brooding mollusc *Crepidula fornicata*. *Molecular Ecology* 26:5467–5483.
- Robertson, D. R., and G. Justines. 1982. Protogynous hermaphroditism and gonochorism in four Caribbean reef gobies. *Environmental Biology of Fishes* 7:137–142.
- Robertson, D. R., and J. Van Tassell. 2019. Shorefishes of the Greater Caribbean: online information system. Version 2.0 Smithsonian Tropical Research Institute, Balboa, Panamá.

- Ronce, O. 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*:231–253.
- Roopnarine, P. D., and R. Hertog. 2013. Detailed Food Web Networks of Three Greater Antillean Coral Reef Systems: The Cayman Islands, Cuba, and Jamaica. *Dataset Papers in Ecology* 2013:1–9.
- Rosser, N. L. 2015. Asynchronous spawning in sympatric populations of a hard coral reveals cryptic species and ancient genetic lineages. *Molecular Ecology* 24:5006–5019.
- Rousset, F., and S. Gandon. 2002. Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. *Journal of Evolutionary Biology* 15:515–523.
- Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71:319–392.
- Sadedin, S., and M. J. Littlejohn. 2003. A Spatially Explicit Individual-Based Model of Reinforcement in Hybrid Zones. *Evolution* 57:962–970.
- Sale, P. F. 1998. Appropriate spatial scales for studies of reef-fish ecology. *Australian Journal of Ecology* 23:202–208.
- Salles, O. C., B. Pujol, J. A. Maynard, G. R. Almany, M. L. Berumen, G. P. Jones, P. Saenz-Agudelo, M. Srinivasan, S. R. Thorrold, and S. Planes. 2016. First genealogy for a wild marine fish population reveals multigenerational philopatry. *Proceedings of the National Academy of Sciences* 113:13245–13250.
- Sansom, A., J. Lind, and W. Cresswell. 2009. Individual behavior and survival: the roles of predator avoidance, foraging success, and vigilance. *Behavioral Ecology* 20:1168–1174.

- Scheiner, S. M. 2014. Bet-hedging as a complex interaction among developmental instability, environmental heterogeneity, dispersal, and life-history strategy. *Ecology and Evolution* 4:505–515.
- Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16:372–380.
- Schroeder, J., T. Burke, M.-E. Mannarelli, D. A. Dawson, and S. Nakagawa. 2012. Maternal effects and heritability of annual productivity. *Journal of Evolutionary Biology* 25:149–156.
- Selonen, V., and I. K. Hanski. 2006. Habitat exploration and use in dispersing juvenile flying squirrels. *Journal of Animal Ecology* 75:1440–1449.
- Selwyn, J. D. 2015, July. How does Dispersal affect Chaos: Viscous Populations and Families in a Common Caribbean Goby (*Coryphopterus personatus*). Texas A&M University - Corpus Christi, Corpus Christi, Texas, USA.
- Selwyn, J. D., J. D. Hogan, A. M. Downey-Wall, L. M. Gurski, D. S. Portnoy, and D. D. Heath. 2016. Kin-Aggregations Explain Chaotic Genetic Patchiness, a Commonly Observed Genetic Pattern, in a Marine Fish. *PLOS ONE* 11:e0153381.
- Selwyn, J. D., E. P. Hunt, D. S. Portnoy, and J. D. Hogan. 2022. Maintenance of species boundaries within social aggregations of ecologically similar goby sister species. *Marine Biology* 169:32.
- Selwyn, J. D., S. F. Magnuson, C. E. Bird, and J. D. Hogan. in prep. Effect of habitat heterogeneity on dispersal.
- Selwyn, J. D., P. Usseglio, and J. D. Hogan. 2021. Spatial Distribution and Habitat Usage of *Coryphopterus personatus* and *C. hyalinus* in Turneffe Atoll, Belize. Zenodo.

- Selwyn, J. D., P. Usseglio, and J. D. Hogan. in review. Photogrammetry derived habitat model enables characterization of fine-scale habitat use in a pair of coral reef fishes (CORE-S-21-00271). *Coral Reefs*.
- Serna Rodríguez, K. M., F. A. Zapata, and L. M. Mejía-Ladino. 2016. Diversity and distribution of fishes along the depth gradient of a coral reef wall at San Andrés Island, Colombian Caribbean. *Boletín de Investigaciones Marinas y Costeras-INVEMAR* 45:15–39.
- Shapiro, D. Y. 1983. On the possibility of kin groups in coral reef fishes. *Ecology of Deep and Shallow Coral Reefs* 1:39–45.
- Siegel, D. A., B. P. Kinlan, B. Gaylord, and S. D. Gaines. 2003. Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* 260:83–96.
- Simpson, S. D., M. G. Meekan, R. D. McCauley, A. Jeffs, and others. 2004. Attraction of settlement-stage coral reef fishes to reef noise. *Marine Ecology Progress Series* 276:263–268.
- Skinner, C., A. C. Mill, M. D. Fox, S. P. Newman, Y. Zhu, A. Kuhl, and N. V. C. Polunin. 2021. Offshore pelagic subsidies dominate carbon inputs to coral reef predators. *Science Advances* 7:eabf3792.
- Smith, C., A. Cotter, L. Grinsted, A. Bowolaksono, N. L. Watiniasih, and I. Agnarsson. 2019. In a relationship: sister species in mixed colonies, with a description of new *Chikunia* species (Theridiidae). *Zoological Journal of the Linnean Society* 186:337–352.
- Spieth, P. T. 1974. Gene Flow and Genetic Differentiation. *Genetics* 78:961–965.
- Sponaugle, S., R. K. Cowen, A. Shanks, S. G. Morgan, J. M. Leis, J. Pineda, G. W. Boehlert, M. J. Kingsford, K. C. Lindeman, and C. Grimes. 2002. Predicting self-recruitment in marine

- populations: biophysical correlates and mechanisms. *Bulletin of Marine Science* 70:341–375.
- Stelkens, R. B., C. Schmid, and O. Seehausen. 2015. Hybrid Breakdown in Cichlid Fish. *PLOS ONE* 10:e0127207.
- Stobutzki, I. C., and D. R. Bellwood. 1997. Sustained swimming abilities of the late pelagic stages of coral reef fishes. *Marine Ecology Progress Series* 149:35–41.
- Tåning, Å. V. 1952. Experimental Study of Meristic Characters in Fishes. *Biological Reviews* 27:169–193.
- Tarpey, C. M., J. E. Seeb, G. J. McKinney, W. D. Templin, A. Bugaev, S. Sato, and L. W. Seeb. 2017. Single-nucleotide polymorphism data describe contemporary population structure and diversity in allochronic lineages of pink salmon (*Oncorhynchus gorbuscha*). *Canadian Journal of Fisheries and Aquatic Sciences*.
- Taylor, H. R. 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution* 5:3140–3150.
- Taylor, M. S., and M. E. Hellberg. 2005. Marine Radiations at Small Geographic Scales: Speciation in Neotropical Reef Gobies (*Elacatinus*). *Evolution* 59:374–385.
- Teske, P. R., J. Sandoval-Castillo, T. R. Golla, A. Emami-Khoyi, M. Tine, S. von der Heyden, and L. B. Beheregaray. 2019. Thermal selection as a driver of marine ecological speciation. *Proceedings of the Royal Society B: Biological Sciences* 286:20182023.
- Thacker, C. E. 2009. Phylogeny of Gobioidae and placement within Acanthomorpha, with a new classification and investigation of diversification and character evolution. *Copeia* 2009:93–104.

- Thacker, C. E. 2015. Biogeography of goby lineages (Gobiiformes: Gobioidae): origin, invasions and extinction throughout the Cenozoic. *Journal of Biogeography* 42:1615–1625.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673–4680.
- Thresher, R. E. 1984. *Reproduction in reef fishes*. T.F.H. Publications, Neptune City, New Jersey.
- Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* 21:3907–3930.
- Toonen, R. J., and J. R. Pawlik. 1994. Foundations of gregariousness. *Nature* 370:511.
- Toonen, R. J., and J. R. Pawlik. 2001. Foundations of Gregariousness: A Dispersal Polymorphism Among the Planktonic Larvae of a Marine Invertebrate. *Evolution* 55:2439–2454.
- Tornabene, L., G. N. Ahmadi, M. L. Berumen, D. J. Smith, J. Jompa, and F. Pezold. 2013. Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). *Molecular Phylogenetics and Evolution* 66:391–400.
- Tregenza, T. 1995. Building on the Ideal Free Distribution. Pages 253–307 in M. Begon and A. H. Fitter, editors. *Advances in Ecological Research*. Academic Press.
- Vanderpham, J. P., S. Nakagawa, and G. P. Closs. 2013. Habitat-related patterns in phenotypic variation in a New Zealand freshwater generalist fish, and comparisons with a closely related specialist. *Freshwater Biology* 58:396–408.

- Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. 2019. Rank-normalization, folding, and localization: An improved \widehat{R} for assessing convergence of MCMC. arXiv:1903.08008 [stat].
- Victor, B. C. 2019, September 22. *Coryphopterus* and *Lythrypnus* gobies of the Western Atlantic. <http://www.coralreeffish.com/gobiidae2adult.html>.
- Waples, R. S. 2015. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity* 106:1–19.
- Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia’s fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360:1847–1857.
- Wellenreuther, M., and K. D. Clements. 2007. Reproductive isolation in temperate reef fishes. *Marine Biology* 152:619–630.
- Westfall, P. H. 2014. Kurtosis as Peakedness, 1905 – 2014. R.I.P. The American statistician 68:191–195.
- Westoby, M. J., J. Brasington, N. F. Glasser, M. J. Hambrey, and J. M. Reynolds. 2012. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* 179:300–314.
- Whiteman, E. A., and I. M. Côté. 2004. Monogamy in marine fishes. *Biological Reviews* 79:351–375.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo,

- and H. Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4:1686.
- Williams, M. E., and M. G. Bentley. 2002. Fertilization Success in Marine Invertebrates: The Influence of Gamete Age. *The Biological Bulletin* 202:34–42.
- Wilson, S. K., S. C. Burgess, A. J. Cheal, M. Emslie, R. Fisher, I. Miller, N. V. C. Polunin, and H. P. A. Sweatman. 2008a. Habitat utilization by coral reef fish: implications for specialists vs. generalists in a changing environment. *Journal of Animal Ecology* 77:220–228.
- Wilson, S. K., R. Fisher, M. S. Pratchett, N. a. J. Graham, N. K. Dulvy, R. A. Turner, A. Cakacaka, N. V. C. Polunin, and S. P. Rushton. 2008b. Exploitation and habitat degradation as agents of change within coral reef fish communities. *Global Change Biology* 14:2796–2809.
- Wingett, S. W., and S. Andrews. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* 7:1338.
- Wirtz, P. 1999. Mother species–father species: unidirectional hybridization in animals with female choice. *Animal Behaviour* 58:1–12.
- Wong, B. B. M., and M. D. Jennions. 2003. Costs influence male mate choice in a freshwater fish. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.
- Wood, A. J., and G. J. Ackland. 2007. Evolving the selfish herd: emergence of distinct aggregating strategies in an individual-based model. *Proceedings of the Royal Society B: Biological Sciences* 274:1637–1642.
- Wootton, R. J. 1985. Energetics of Reproduction. Pages 231–254 *in* P. Tytler and P. Calow, editors. *Fish Energetics: New Perspectives*. Springer Netherlands, Dordrecht.

- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
- Wright, S. 1942. Isolation by distance. *Genetics* 28:114.
- Wright, S. 1946. Isolation by Distance under Diverse Systems of Mating. *Genetics* 31:39–59.
- Yeo, I.-K., and R. A. Johnson. 2000. A new family of power transformations to improve normality or symmetry. *Biometrika* 87:954–959.
- Youden, W. J. 1950. Index for rating diagnostic tests. *Cancer* 3:32–35.
- Young, G. C., S. Dey, A. D. Rogers, and D. Exton. 2017. Cost and time-effective method for multi-scale measures of rugosity, fractal dimension, and vector dispersion from coral reef 3D models. *PLOS ONE* 12:e0175341.
- Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614–620.
- Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.

APPENDIX A

SUPPLEMENT FOR CHAPTER 2

Table 2.S1. PCR Conditions for each locus including concentrations of MgCl₂ and primer used, the number of cycles and touchdown annealing temperature for each reaction. Missing indicates the proportion of individuals where microsatellites were unable to be detected following PCR. COPE10 and CPER52 were excluded from subsequent analysis due to high rates of missingness.

Locus	MgCl ₂ (mM)	Primer (nM)	Anneal (°C)	Cycles	Missing
COPE5	3.0	200	62-54	50	0.005
COPE9	3.0	200	62-54	50	0.084
COPE10	3.0	200	62-54	50	0.278
CPER26	3.0	500	62-54	50	0.026
CPER52	3.0	200	62-54	50	0.391
CPER92	4.5	100	58-52	40	0.068
CPER99	4.5	200	68-64	40	0.011
CPER119	4.5	100	58-52	40	0.055
CER188	4.5	100	58-52	40	0.016

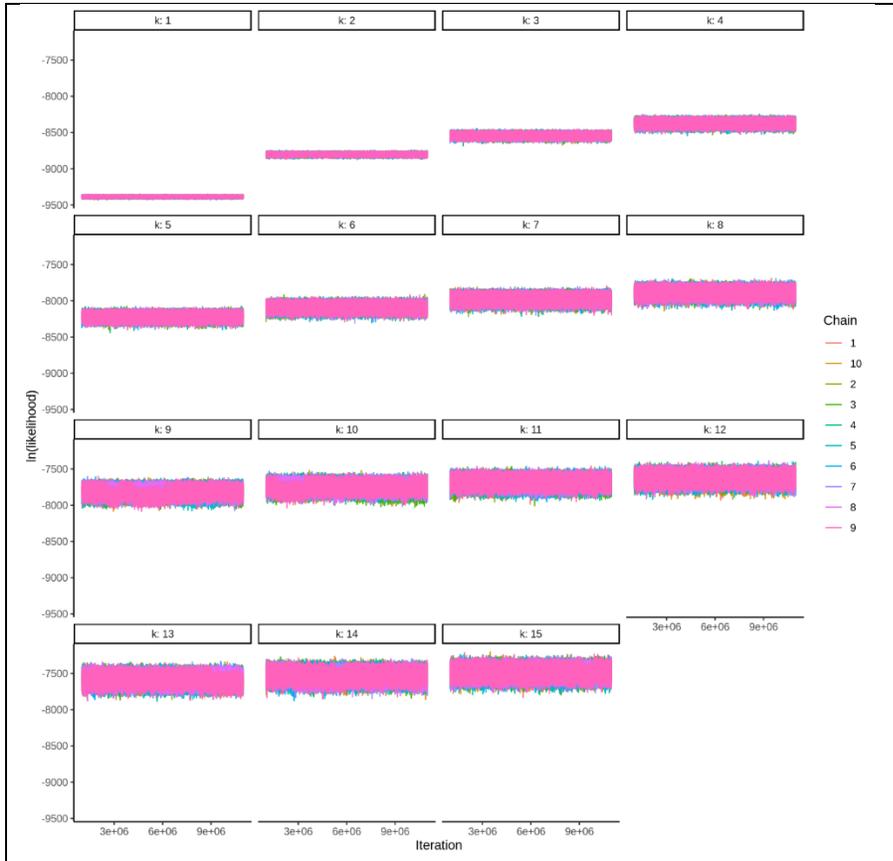
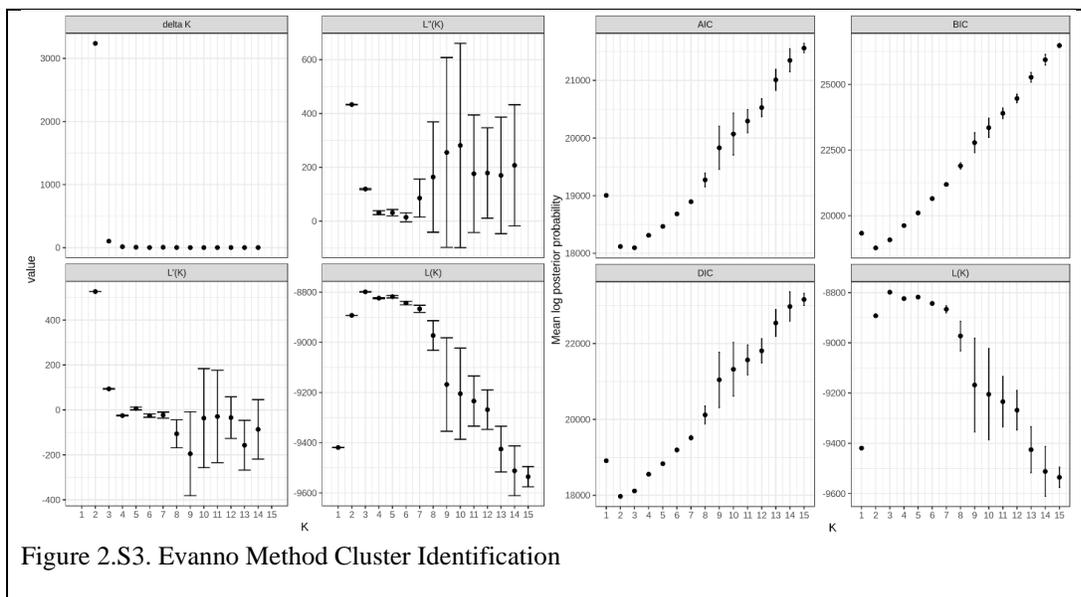
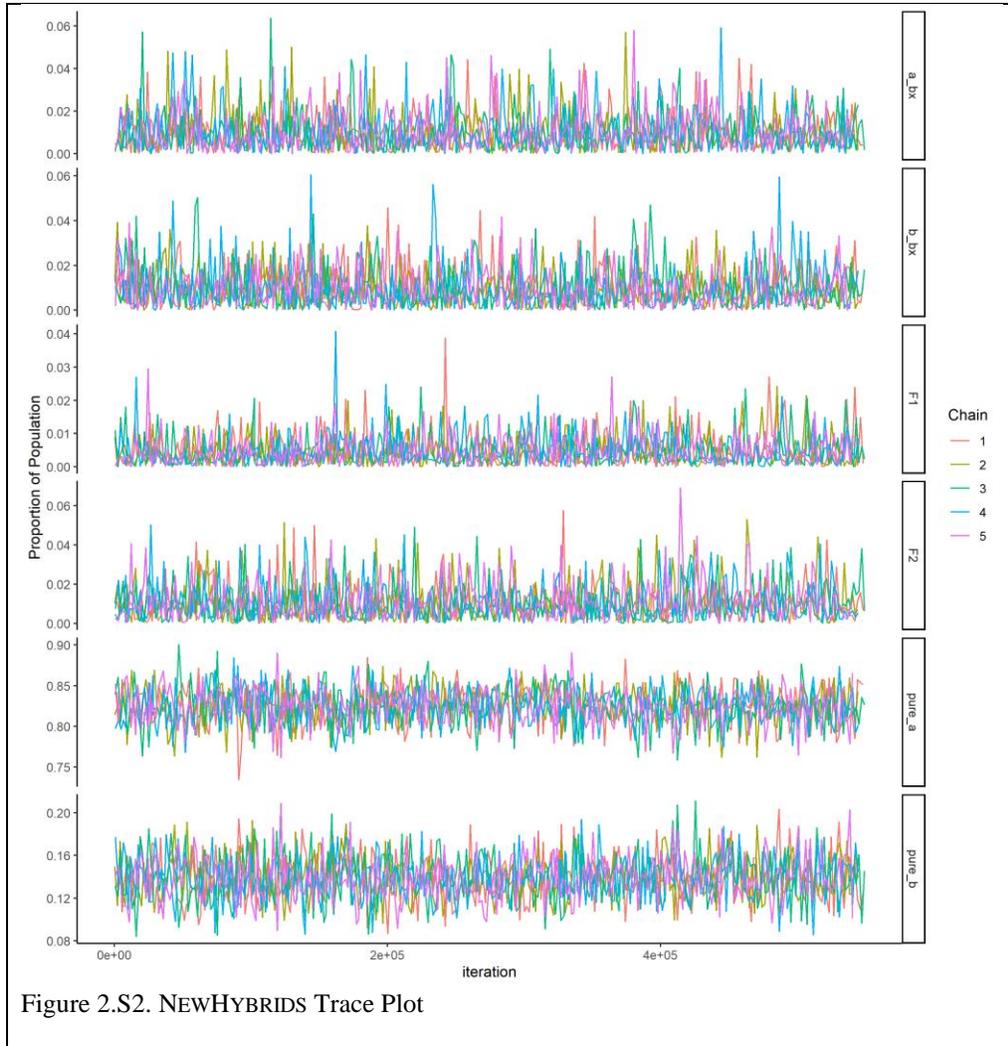
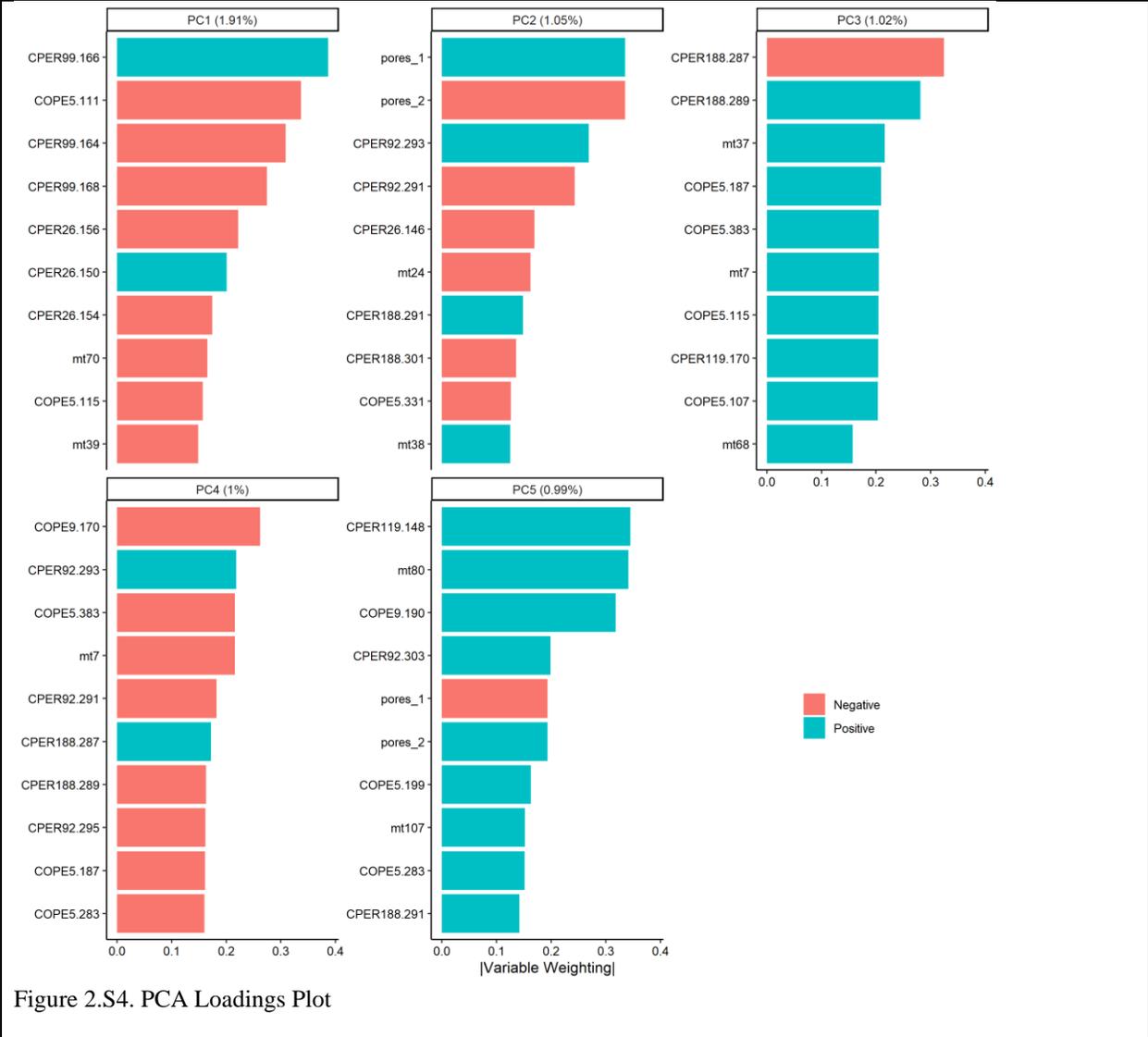


Figure 2.S1. STRUCTURE Trace Plot





APPENDIX B

SUPPLEMENT FOR CHAPTER 3

Supplemental Methods

Habitat Classification

The site-level orthomosaic model was used to classify pixels as either sand or reef habitat. To classify the habitat, we manually created an evaluation dataset by marking 5,133 polygons (427.75 ± 112.09 SD per site) spread haphazardly across sites as either “sand” or “reef” habitat. These polygons ranged in area from 0.25 cm^2 to 19.7 m^2 with a mean area of $596 \pm 4,840 \text{ cm}^2$ (SD). To account for variation in both the number of overall pixels corresponding to sand versus reef habitat, and the variation in the number of pixels within each polygon (i.e., sand polygons are generally larger than reef polygons), the data was downsampled to have equal representation of sand and reef pixels from each site in proportion to the relative area of the polygon to ensure all individual polygons were represented in the final, human-classified dataset. After downsampling, the dataset was separated into a set of training (70%, 19,993 pixels) and testing pixels (30%, 8,568 pixels) which were stratified by site to ensure all sites are represented in both sets. The training set was used to select optimal model parameters. We used the testing set of pixels to calculate model accuracy metrics and select the best modelling strategy to avoid overfitting. Pixels were classified into either reef or sand habitat based on the RGB color value of the pixel as well as the site, to account for differences in orthomosaic quality and lighting on the day data was gathered. Additionally, RGB color data was normalized using Yeo-Johnson transformation and z-score transformed (Yeo and Johnson 2000).

As we had no *a priori* reason to suspect one modelling strategy may be better than another, we trained models using twelve different classification algorithms (Table 3.S1). To identify the

best parameter setting for these models, we used 10-fold cross validation repeated five times and initially fit a space-filling random selection of parameter values. The parameter values were sequentially tuned using an iterative Bayesian algorithm to search parameter space and maximize Youden's J, a 0-1 metric where a value of 1 indicates no false-positive and no false-negative values (Youden 1950). The tuned parameters were: whether to PCA transform RGB color data, the number of components to retain, and model specific tuning parameters (Table 3.S2). Model fitting and parameter tuning were performed using the R packages PARSNIP and TUNE in the TIDYMODELS suite of packages (R Core Team 2018, Kuhn and Wickham 2020). After identifying optimal parameter settings for each model type, models were fit to the full training dataset using the chosen parameters and several quality metrics (described below) were calculated using the testing data.

Modelling strategies were compared using a suite of metrics to assess various aspects of model quality. Among these metrics were Youden's J, accuracy, area under the receiver operating curve (AUC), sensitivity (true sand identification rate), specificity (true reef identification rate), and McNemar's χ^2 statistic (a test for random distribution of errors in the model; McNemar 1947). The best model was selected based on a holistic assessment of all these metrics, ordinated with a principal components analysis, to identify the model which performs best across a suite of traits rather than simply maximizing a single metric of model quality. Additionally, we calculated these metrics for each site individually to determine if there were any sites where the classification model either worked exceptionally well, or poorly.

Specimen collection and species identification

In addition to creating digital reef models, a total of 729 individual fish were collected for species identification from between 4 and 7 haphazardly chosen shoals from each site (except A;

14.6 ± 1.31 SE individuals per shoal). Species identity was determined following (Beeken et al. 2021). Briefly, individuals were sequenced using ddRAD (Peterson et al. 2012) which was first mapped to the mitochondrial genome of *Bathygobius cocosensis* (Evans et al. 2018) before using BLAST in GenBank (Altschul et al. 1990) to identify existing sequences with >99% similarity to *C. hyalinus* or *C. personatus*. For individuals which did not map to the reference mitochondrial genome we used k-means clustering of principal components to identify nuclear multi-locus DNA clusters (Jombart et al. 2010) with the resulting clusters associated with the BLAST-based species identification to determine species – cluster correspondence.

Shoal composition was compared to the background of each reef site to determine if shoals are segregated by species or, as previously found, composed of a mixture matching background species frequencies (Selwyn et al. 2022). Briefly, a 95% credible interval for the composition of each shoal and the background reef each shoal was found on was estimated using a binomial distribution with an uninformative ($\beta(1, 1)$) beta conjugate prior (Gelman et al. 2013). To determine if a shoal had a different composition than expected from the site background, we used the algebraic solution to the difference in two proportions (Pham-Gia et al. 1993) with a significantly different shoal composition being present if 0 was not contained in the 95% credible interval.

Supplemental Results

Habitat Classification

Overall, the accuracy of all habitat classification models was quite high (>85.8%) with the best modelling technique across all metrics (except sensitivity) being the C5.0 ruleset using all three components from PCA transformed RGB values and 79 ensemble models requiring splits to contain at least 39 pixels (accuracy = 94.1%; Table 3.S1). The results were marginally skewed

towards falsely classifying reef habitat as sand ($\chi^2_{(1)} = 8.143$; $p = 0.056$) with an overall sensitivity (representing the ability to accurately classify pixels as sand) of 0.952 and a slightly lower specificity (representing the ability to accurately classify pixels as reef) of 0.929. While each site individually had different accuracy values, they mostly behaved in the same way with the models being better able to predict sand than reef points (except for G; Table 3.S3). This disparity is likely due to the nature of the diversity of items classified as “reef” (including rubble, live coral, sponge, dead coral) while only clearly sandy habitat was marked as sand in the training data.

Table 3.S1. Habitat classification model quality statistics based on identifying the best model parameter set and then assessing quality on the test data set. The various modelling frameworks evaluated to classify the habitat quality are ordered by Youden’s J which shows the C5.0 ruleset model as the best framework across all quality metrics, except slightly lower specificity than random forest and boosted random forest models. McNemar p-values have been corrected for multiple comparisons using sequential Bonferroni correction.

Model	Youden's J	Accuracy	ROC/AUC	Sensitivity	Specificity	McNemar $\chi^2_{(1)}$	McNemar p-value
C5.0 Ruleset	0.881	0.941	0.984	0.952	0.929	8.143	0.004
Random Forest	0.878	0.94	0.983	0.959	0.919	37.268	<< 0.0001
Boosted Random Forest	0.875	0.938	0.983	0.957	0.917	36.137	<< 0.0001
Bagged Decision Tree	0.869	0.935	0.98	0.949	0.919	15.612	<< 0.0001
k-Nearest Neighbors	0.872	0.938	0.98	0.963	0.91	69.755	<< 0.0001
Multivariate Adaptive Regression Splines (MARS)	0.834	0.919	0.959	0.953	0.881	100.571	<< 0.0001
Flexible Discriminant Analysis	0.802	0.904	0.956	0.957	0.845	221.582	<< 0.0001
Regularized Discriminant Analysis	0.797	0.902	0.952	0.957	0.84	239.218	<< 0.0001
Logistic Regression	0.796	0.901	0.945	0.945	0.851	143.633	<< 0.0001
Linear Discriminant Analysis	0.767	0.888	0.94	0.968	0.799	459.8	<< 0.0001
Bagged MARS	0.74	0.873	0.924	0.913	0.828	79.839	<< 0.0001
Naïve Bayes	0.707	0.858	0.932	0.933	0.773	304.25	<< 0.0001

Table 3.S2. Summary of model frameworks assessed with the best parameter values identified using the training dataset.

Model	Parameter	Best Value
Decision Tree	trees	73
Decision Tree	min_n	39
Decision Tree	num_comp	3
Random Forest	mtry	7
Random Forest	trees	1036
Random Forest	min_n	13
Boosted Random Forest	mtry	14
Boosted Random Forest	trees	1965
Boosted Random Forest	min_n	8
Boosted Random Forest	tree_depth	13
Boosted Random Forest	learn_rate	0.016314603
Boosted Random Forest	loss_reduction	2.56E-10
Boosted Random Forest	sample_size	0.337691985
Bagged Decision Tree	cost_complexity	3.32E-10
Bagged Decision Tree	tree_depth	15
Bagged Decision Tree	min_n	10
Bagged Decision Tree	class_cost	0.230485588
Bagged Decision Tree	num_comp	3
k-Nearest Neighbors	neighbors	15
k-Nearest Neighbors	weight_func	cos
k-Nearest Neighbors	num_comp	3
Multivariate Adaptive Regression Splines (MARS)	num_terms	12
Multivariate Adaptive Regression Splines (MARS)	prod_degree	1
Multivariate Adaptive Regression Splines (MARS)	prune_method	none
Flexible Discriminant Analysis	num_terms	15
Flexible Discriminant Analysis	prod_degree	1
Flexible Discriminant Analysis	prune_method	exhaustive
Regularized Discriminant Analysis	frac_common_cov	0.006468043
Regularized Discriminant Analysis	frac_identity	0.004812972
Regularized Discriminant Analysis	num_comp	0
Logistic Regression	num_comp	0
Linear Discriminant Analysis	num_comp	0
Bagged MARS	num_terms	5
Bagged MARS	prod_degree	1
Bagged MARS	prune_method	seqrep
Bagged MARS	num_comp	2
Naive Bayes	smoothness	0.500345914
Naive Bayes	Laplace	1.272354675
Naive Bayes	num_comp	3

Table 3.S3. Classification quality metrics across all sites (arranged north-south) independently to show the differences in habitat classification efficacy based on site. Across all metrics the C5.0 model performed worst classifying the habitat of site G. This may be a result of higher surge on the day pictures were taken for photogrammetry leading to increased gorgonian movement which leads to warping of the orthomosaic in the final site model used for classification. McNemar p-values have been corrected for multiple comparisons using sequential Bonferroni correction.

Site	Youden's J	Accuracy	ROC/AUC	Sensitivity	Specificity	McNemar $\chi^2_{(1)}$	McNemar p-value
A	0.836	0.918	0.97	0.919	0.917	0.417	1
B	0.989	0.995	1	0.997	0.991	0.25	1
C	0.881	0.942	0.987	0.949	0.932	0	1
D	0.945	0.973	0.995	0.987	0.958	3.368	0.598
E	0.845	0.924	0.964	0.941	0.903	1.5	1
F	0.902	0.952	0.993	0.969	0.934	2.857	0.728
G	0.62	0.809	0.882	0.787	0.833	2.4	0.849
H	0.898	0.949	0.985	0.98	0.918	12.25	0.006
I	0.911	0.957	0.986	0.971	0.941	1.633	1
J	0.92	0.96	0.989	0.981	0.938	6.759	0.103
K	0.927	0.965	0.993	0.982	0.945	4.654	0.31
L	0.898	0.949	0.988	0.953	0.945	0	1

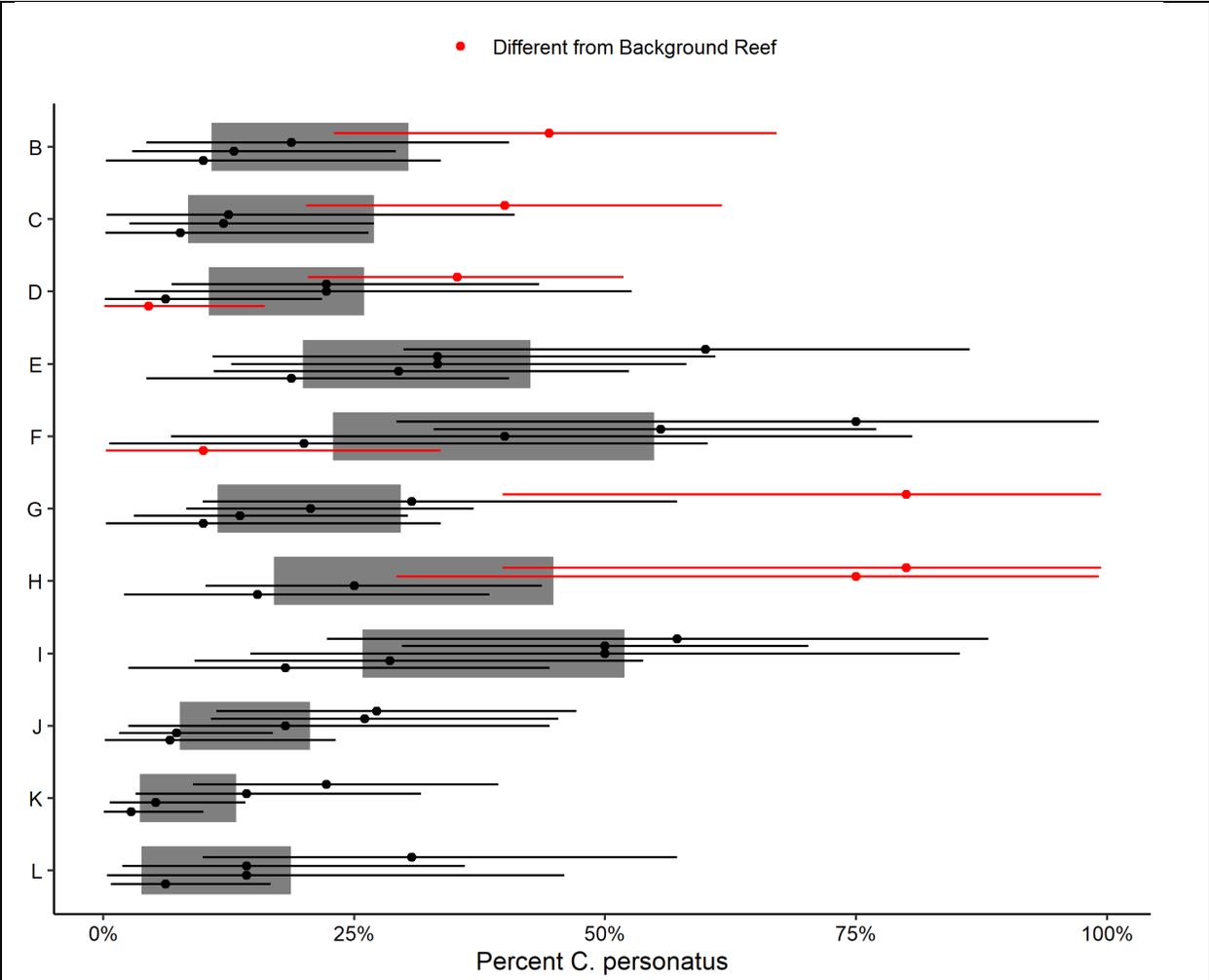


Figure 3.S1. Composition of shoals at each site.

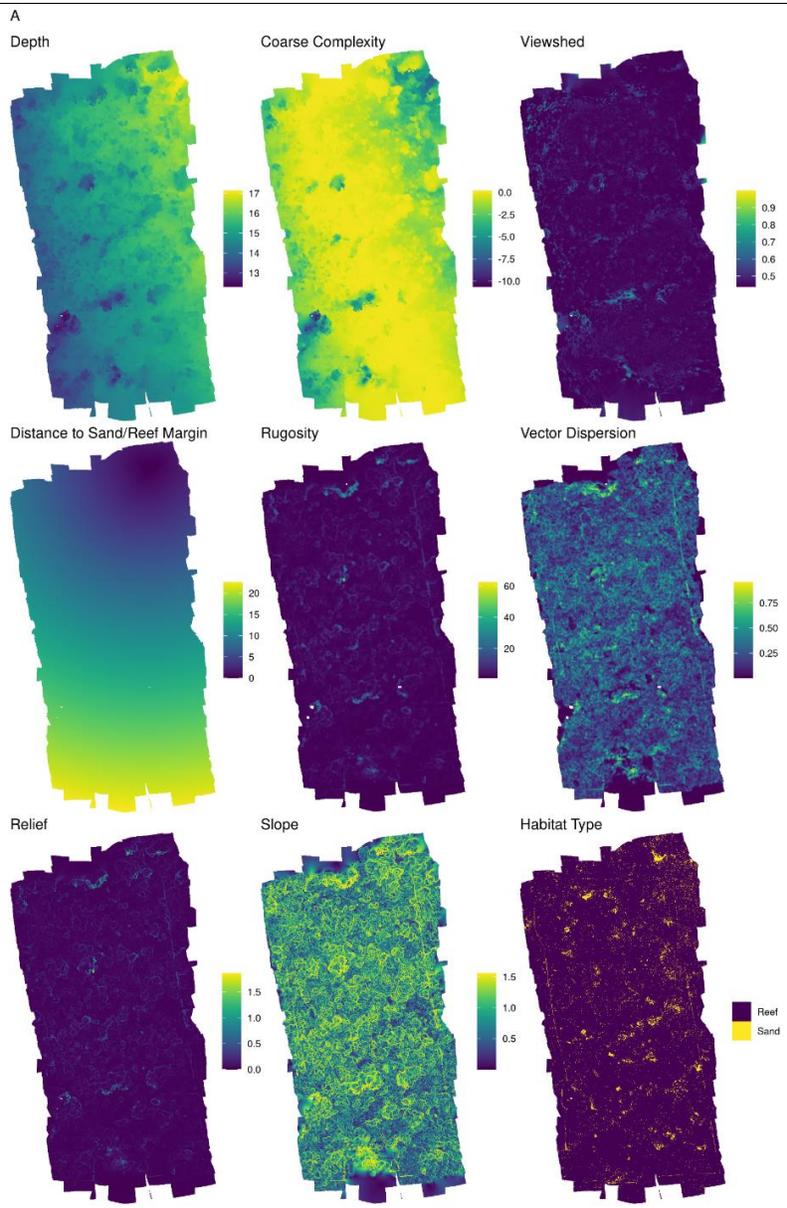
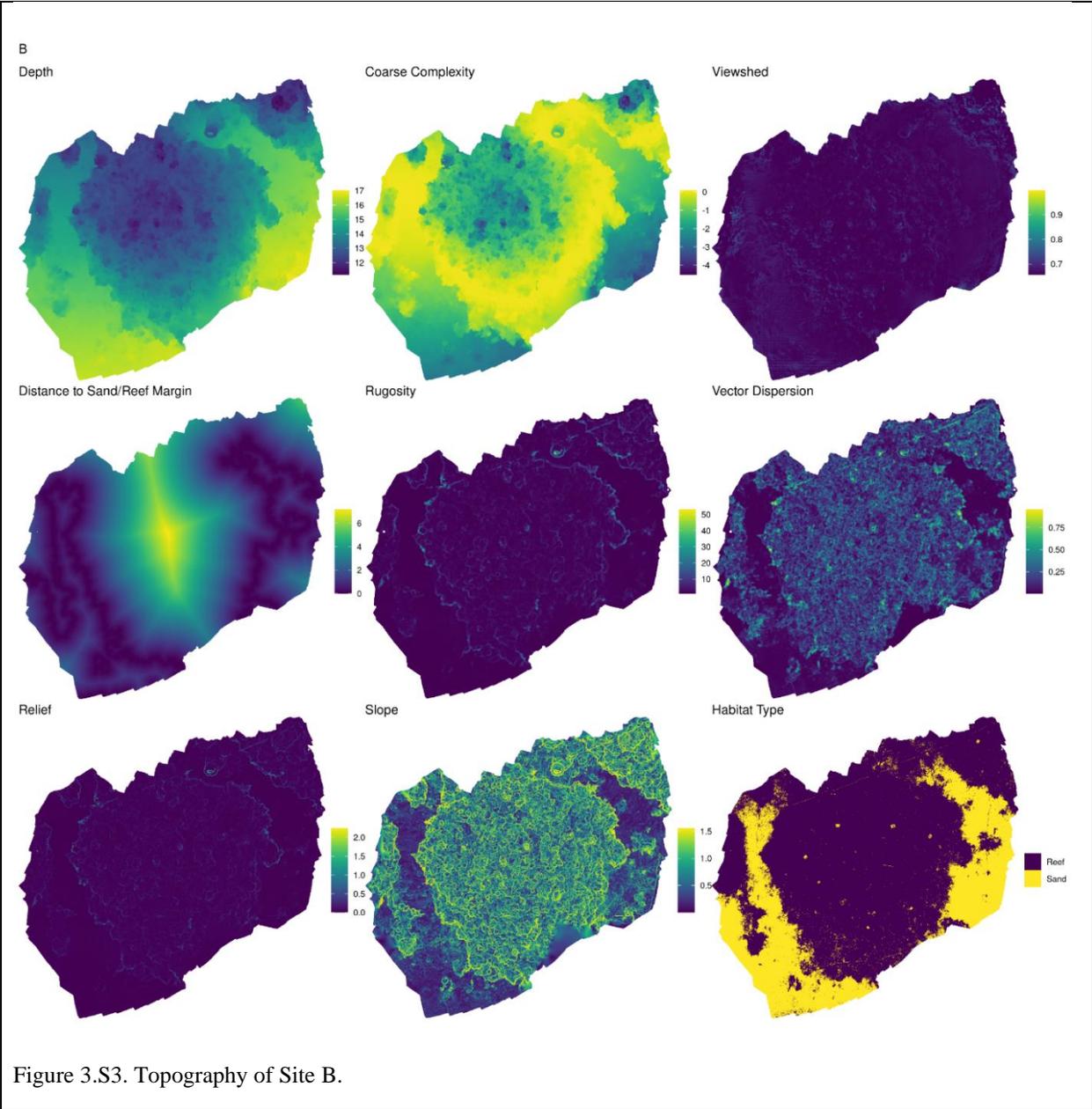


Figure 3.S2. Topography of Site A.



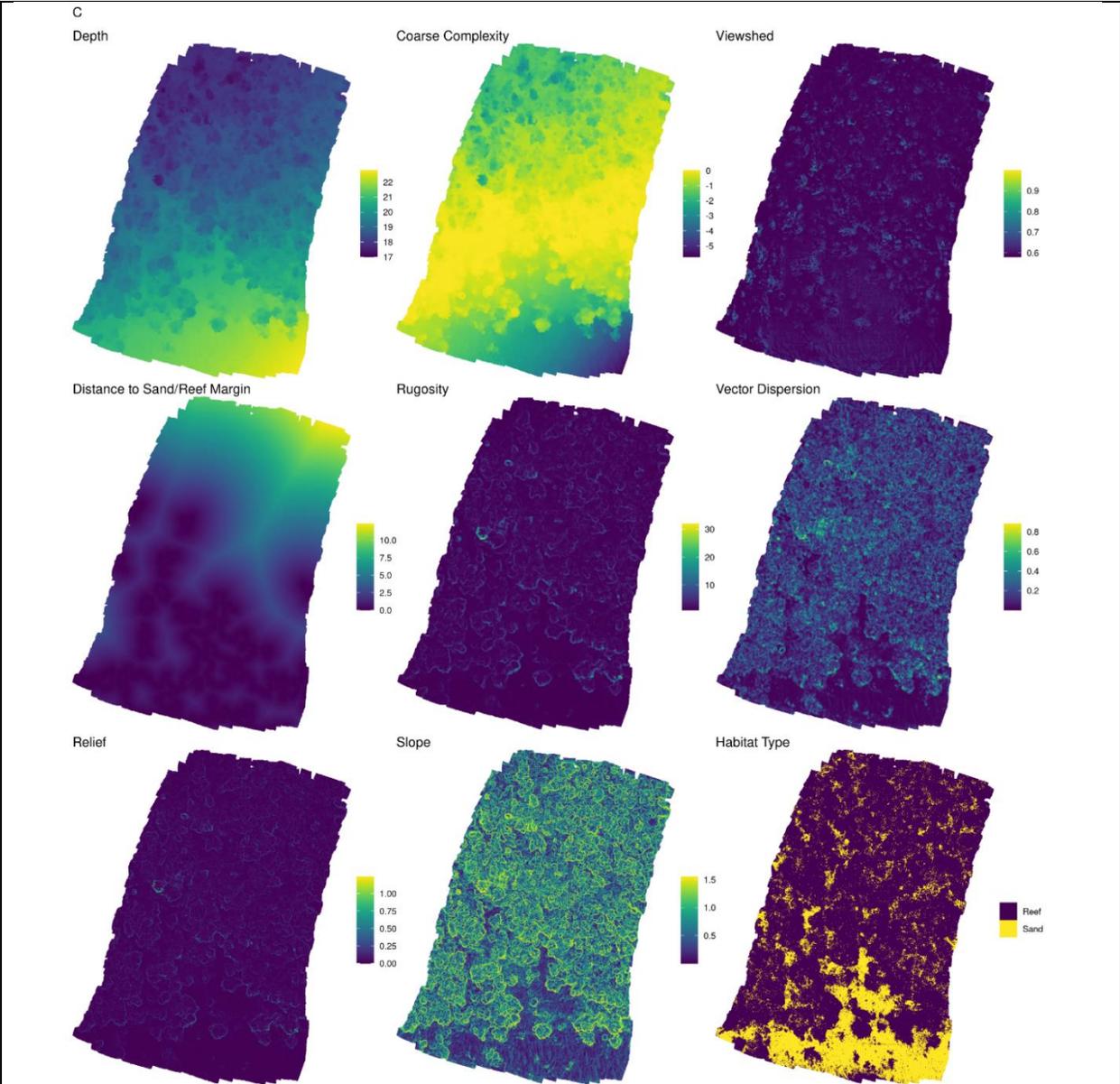


Figure 3.S4. Topography of Site C.

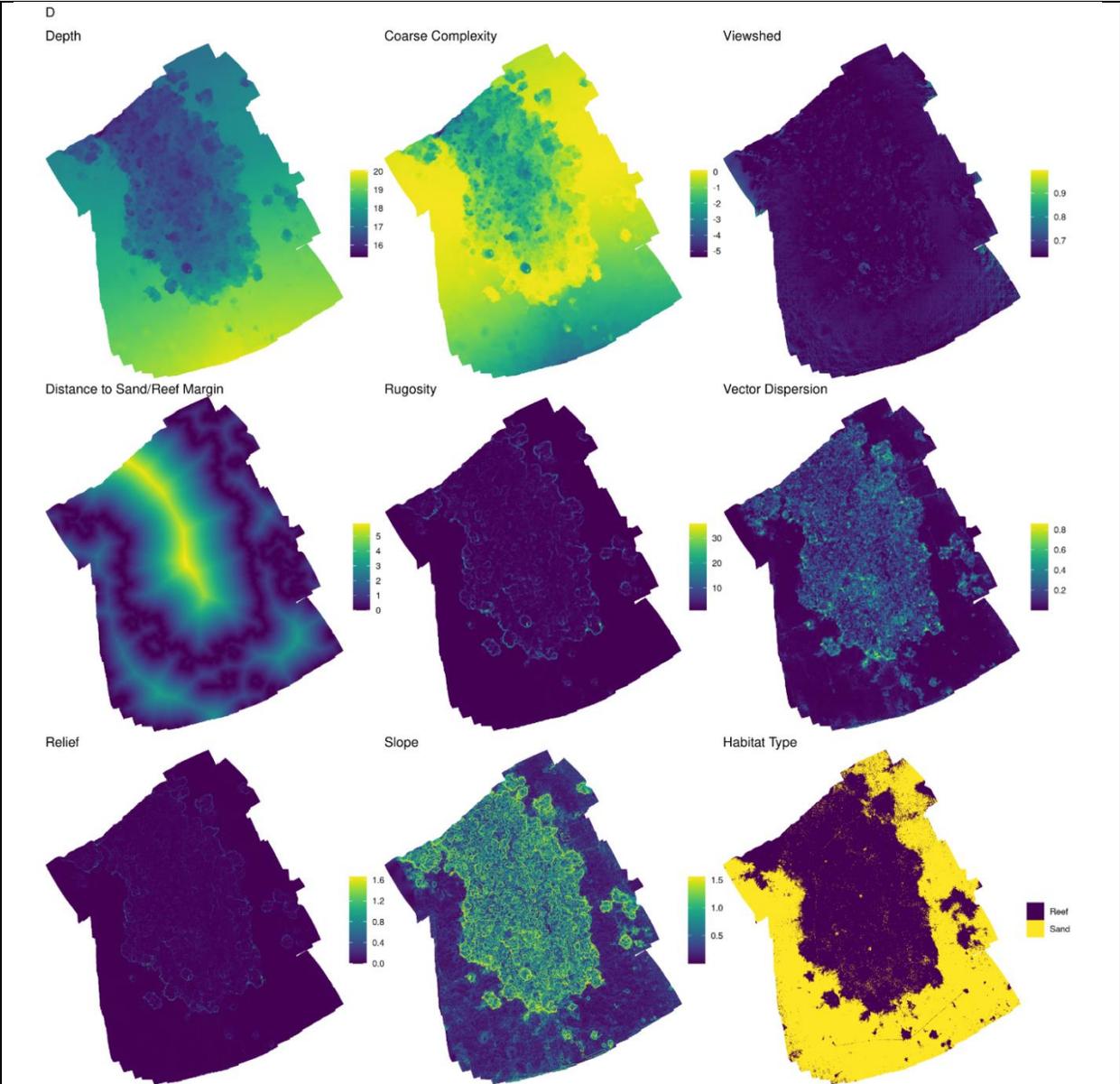
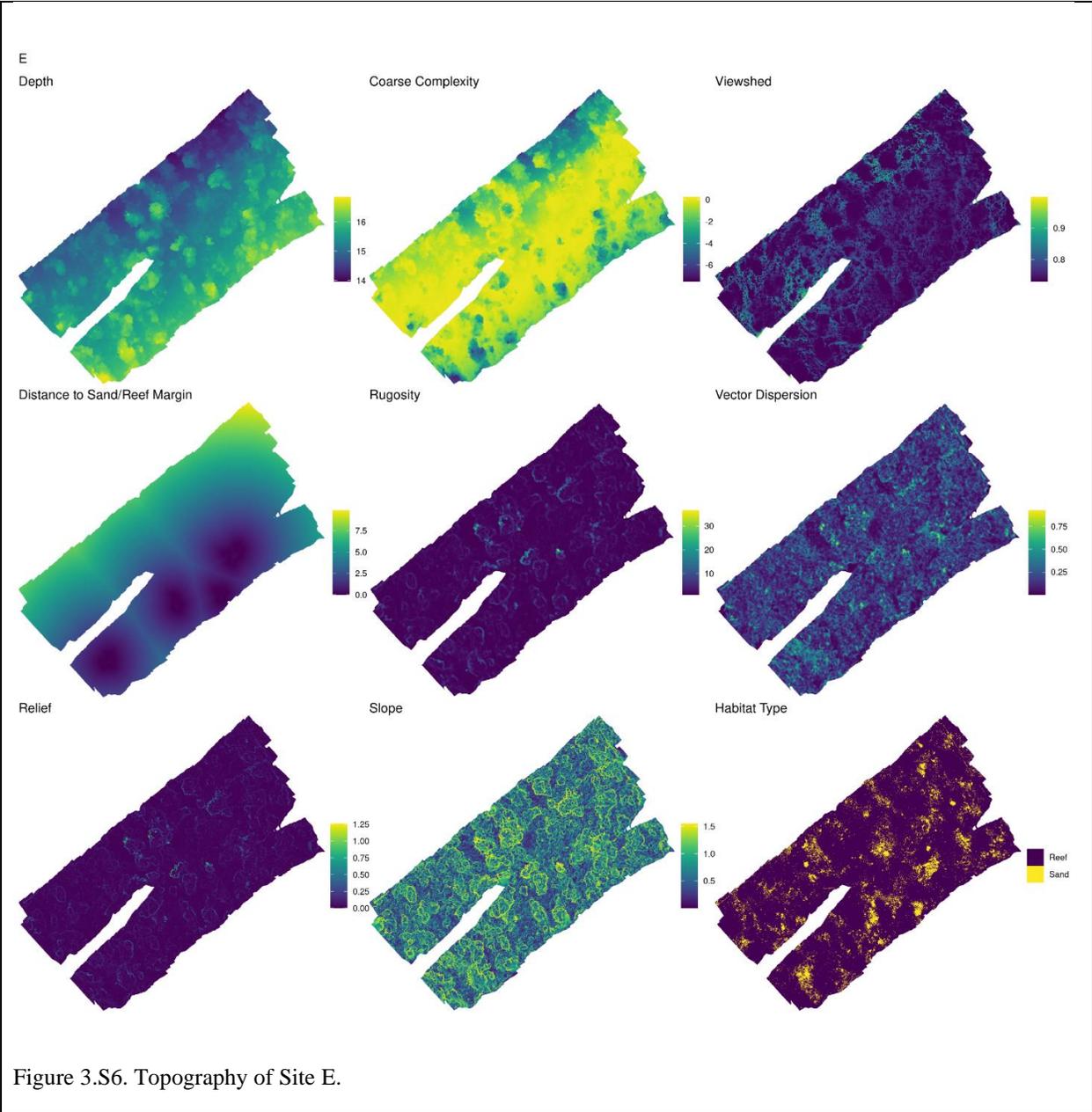
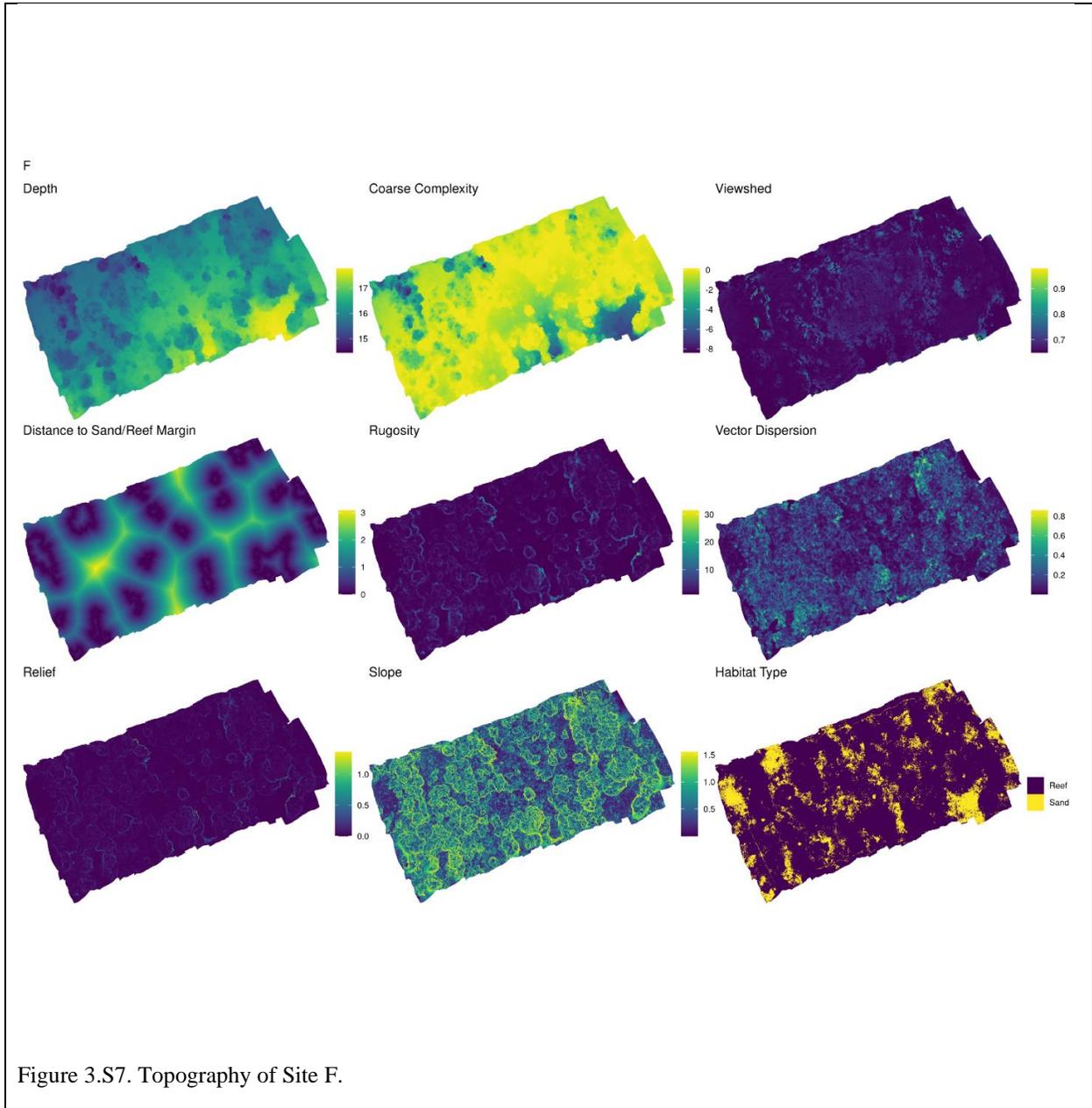
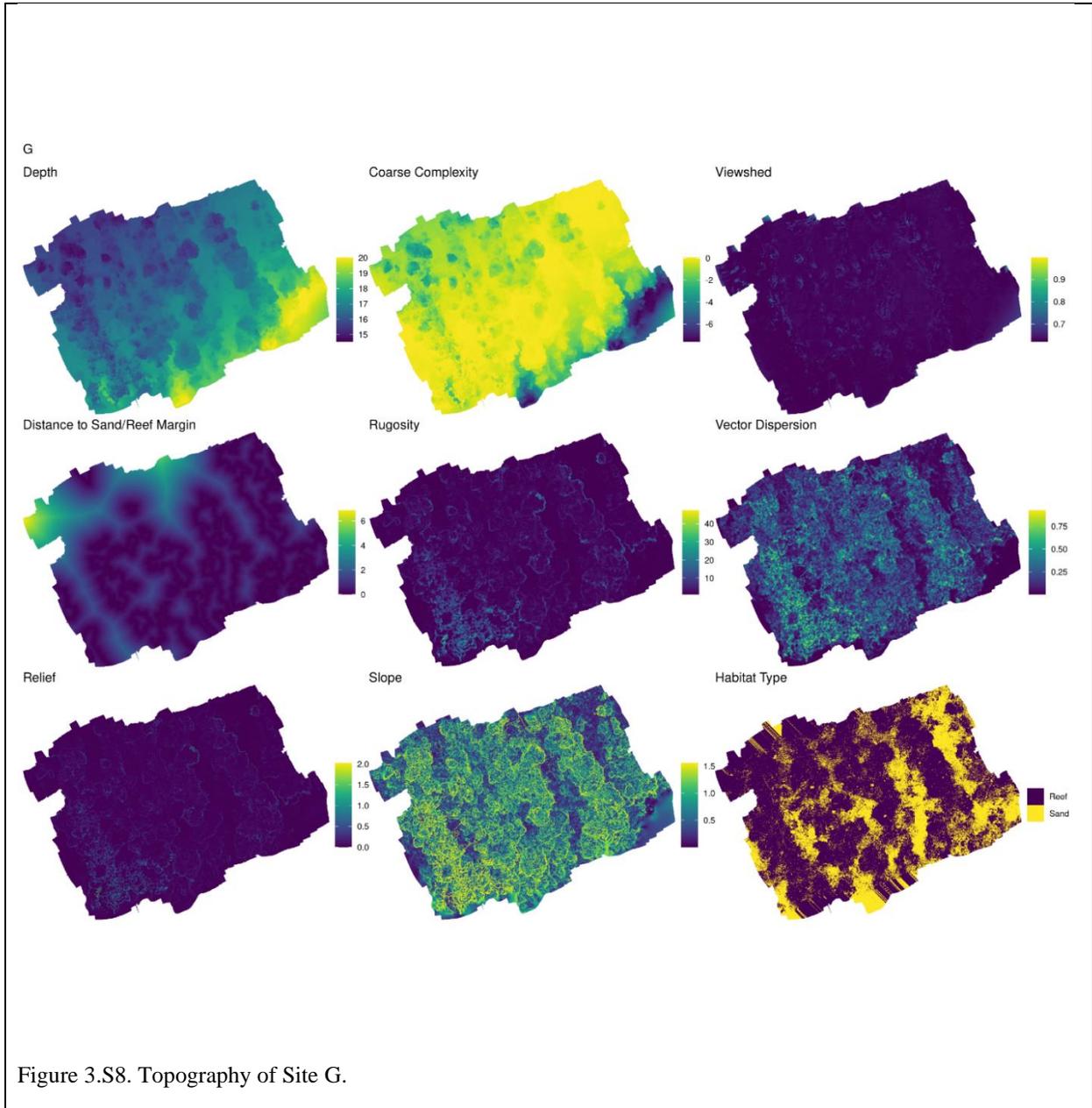
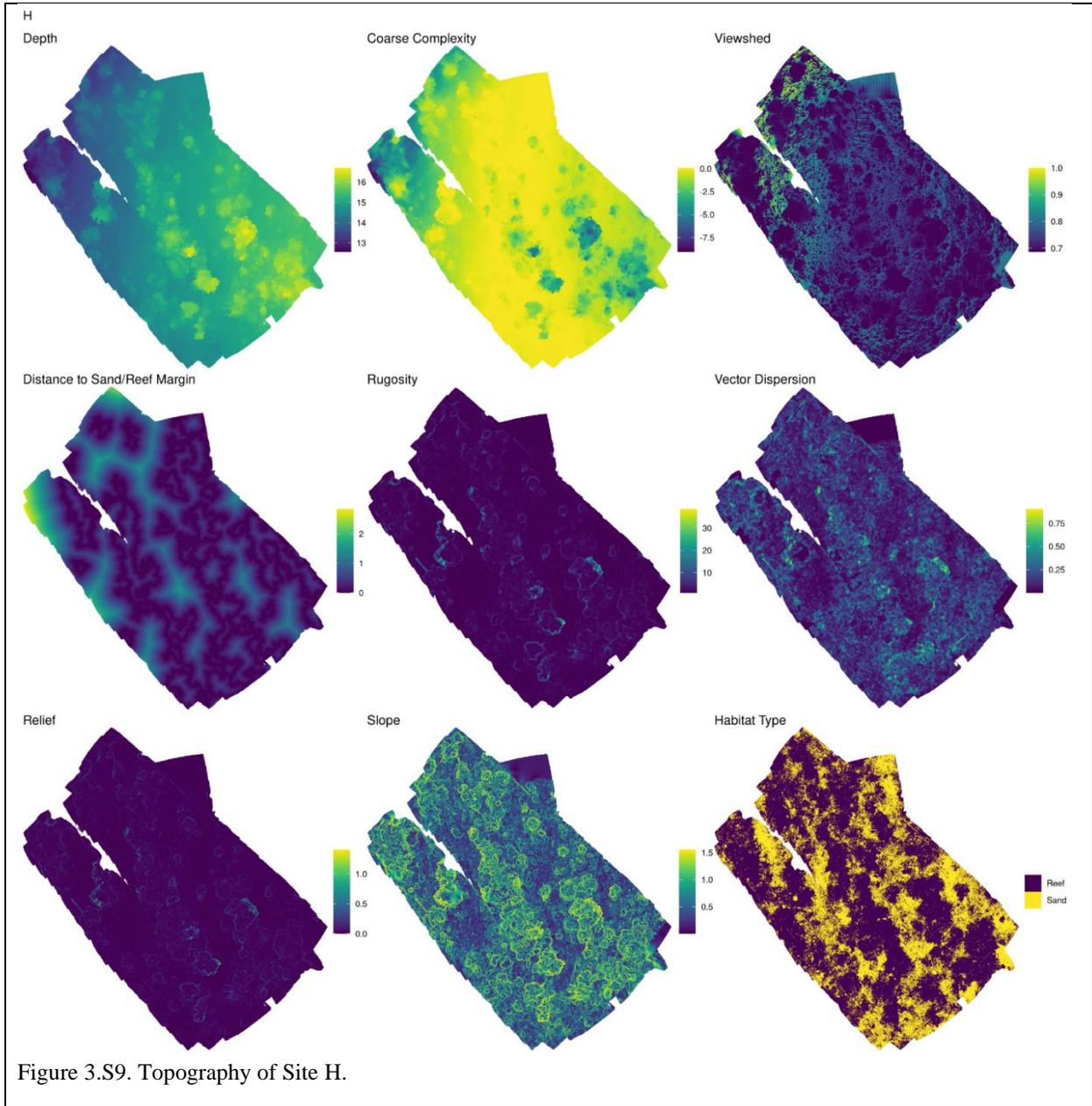


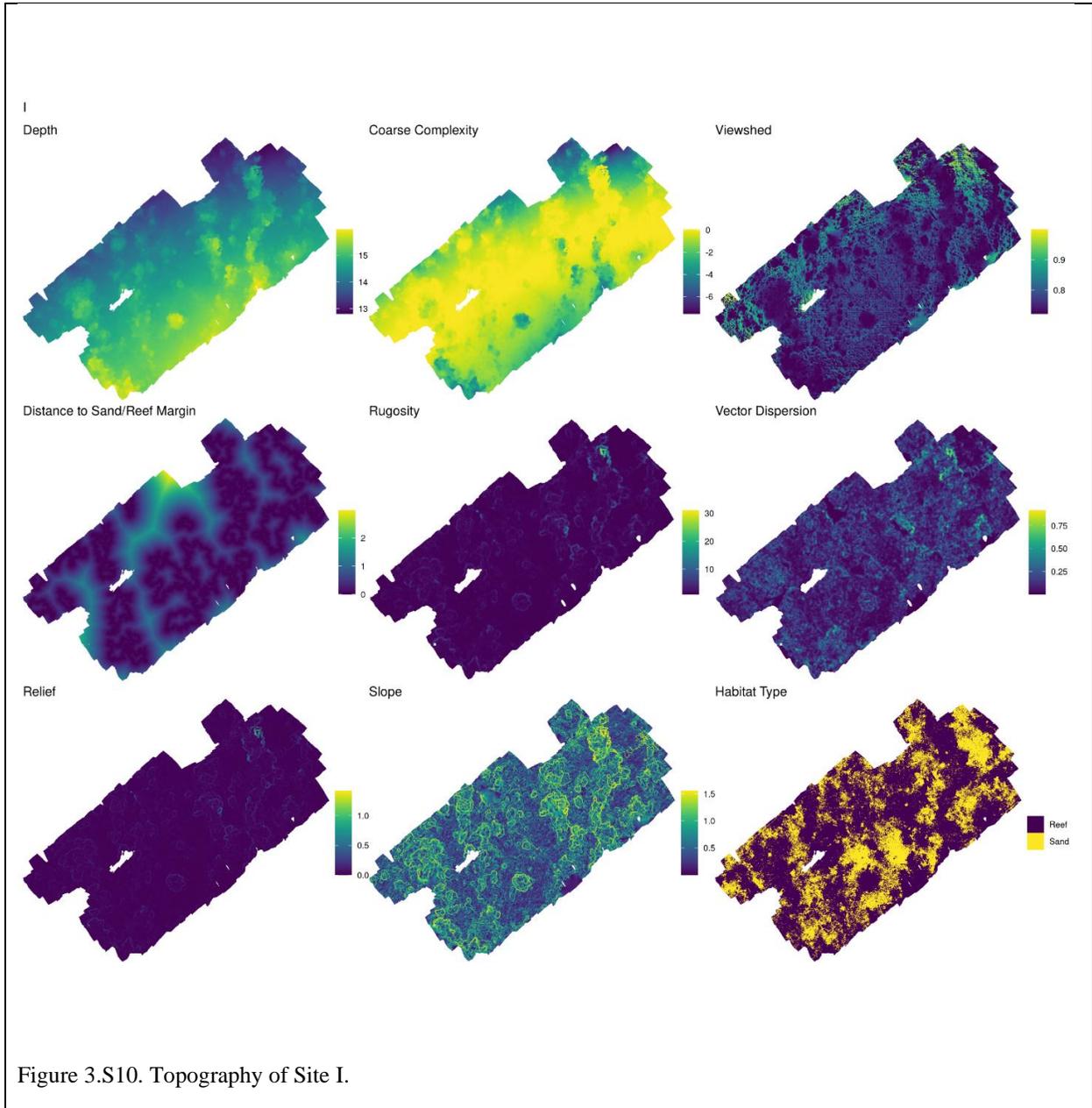
Figure 3.S5. Topography of Site D.











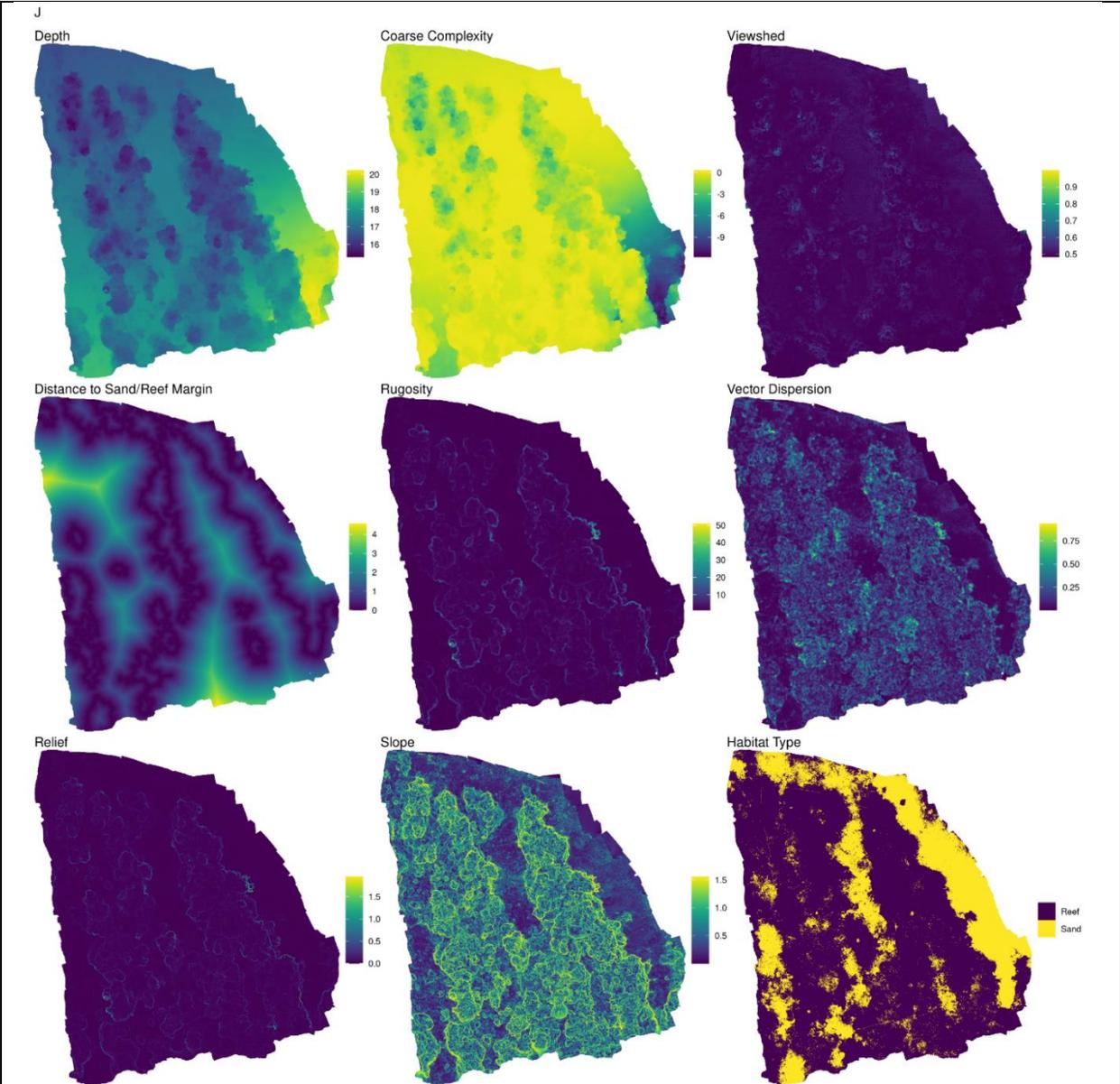


Figure 3.S11. Topography of Site J.

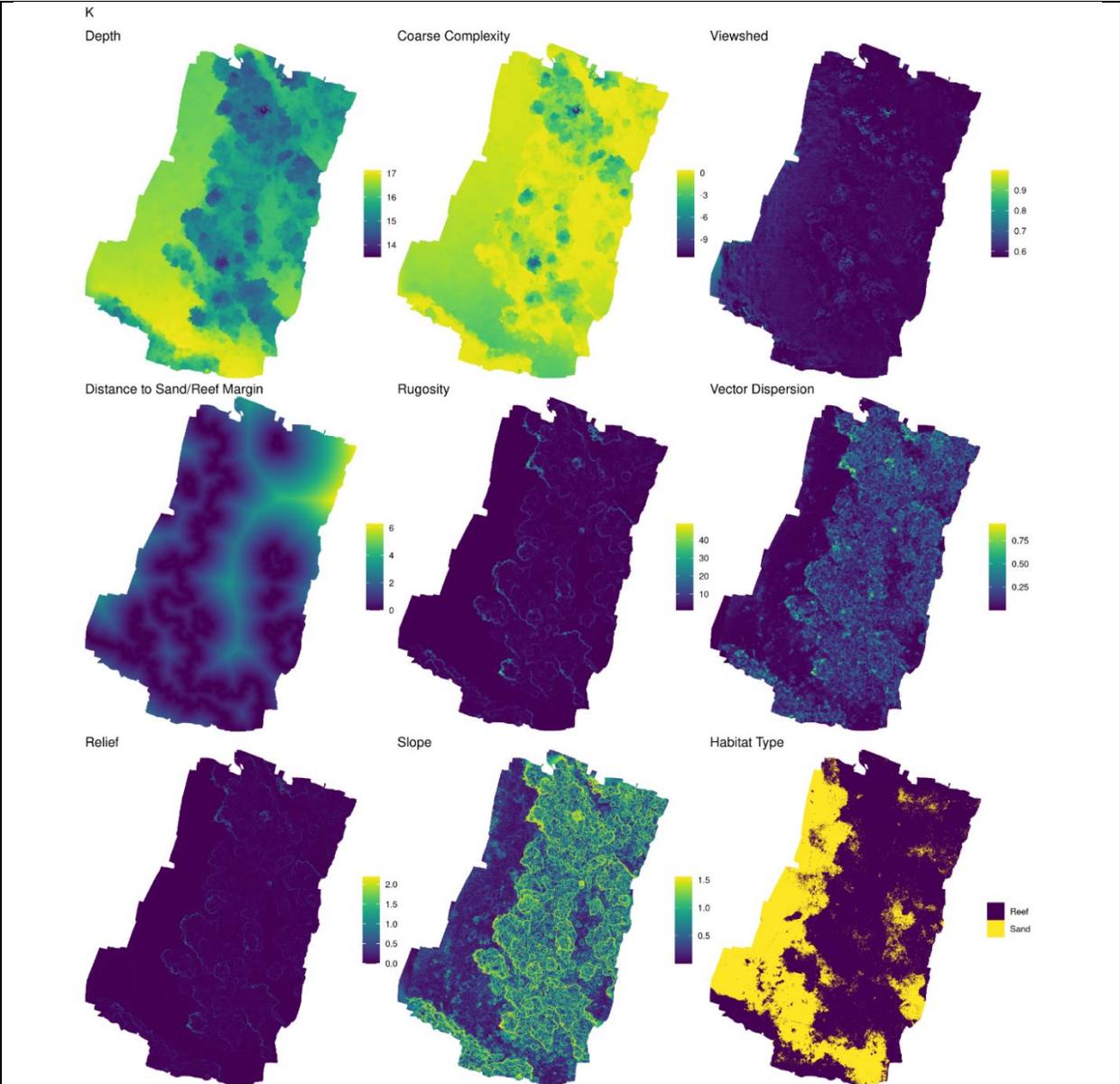
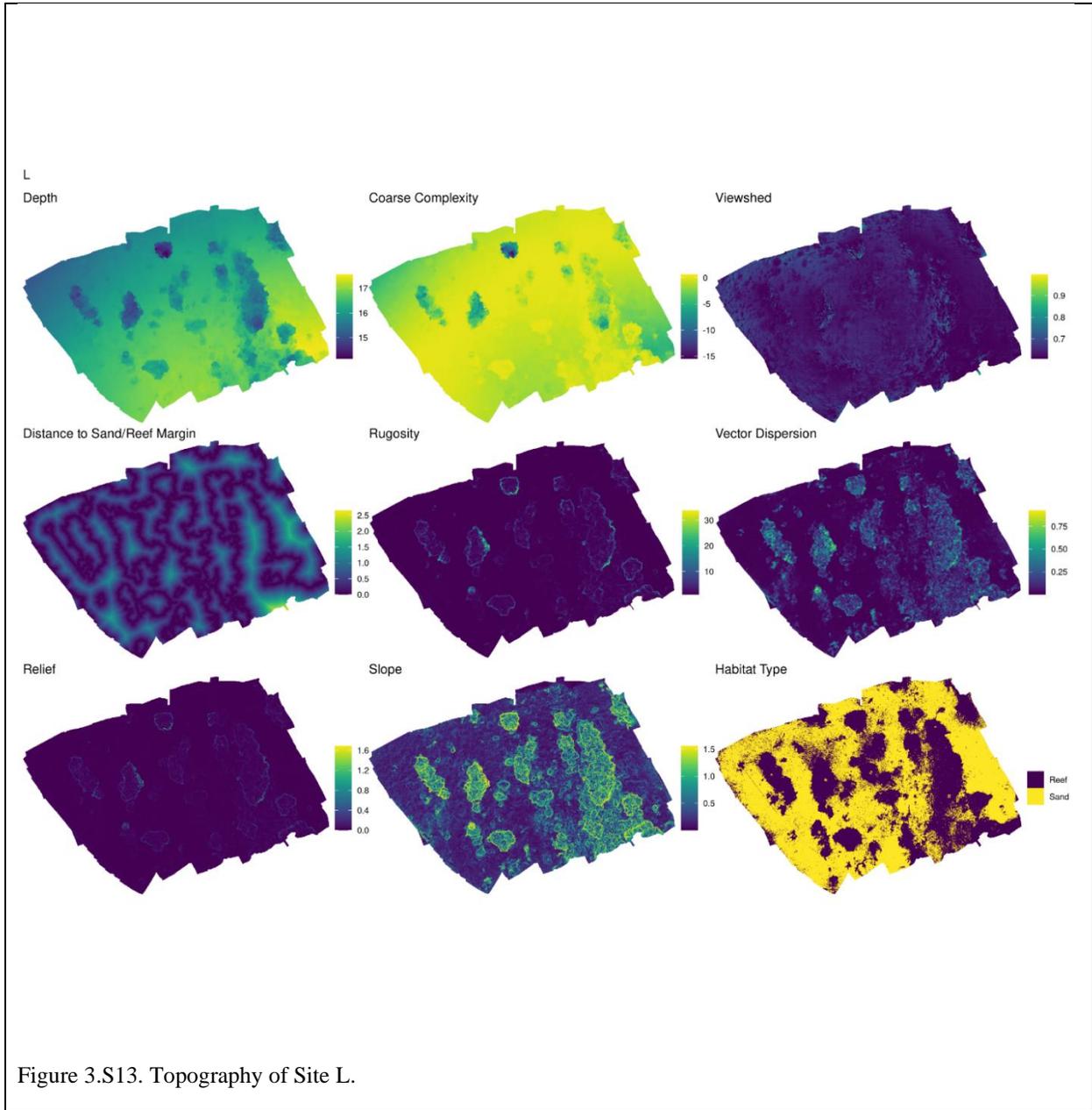


Figure 3.S12. Topography of Site K.



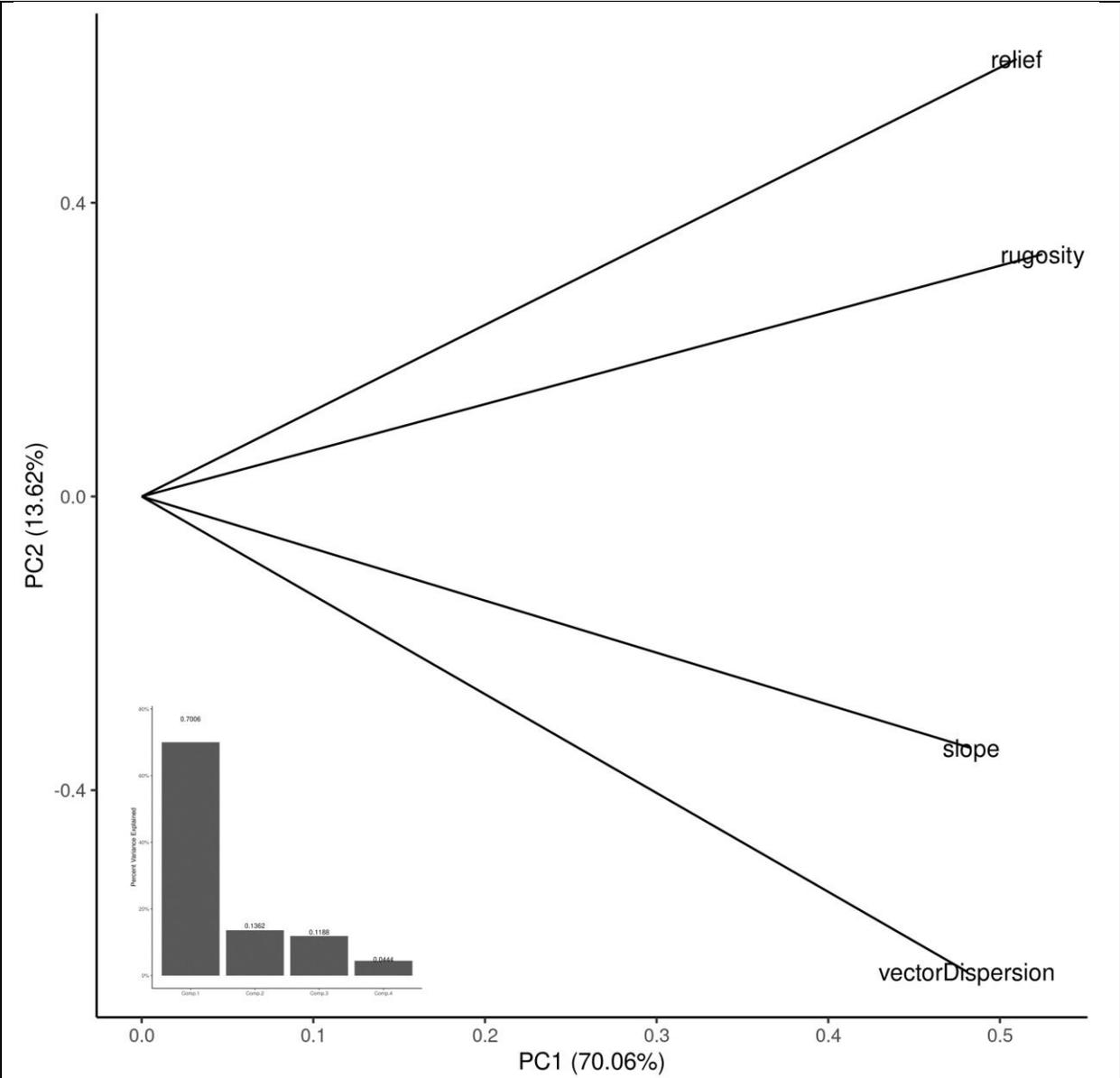


Figure 3.S14. Fine-scale complexity principal component analysis.

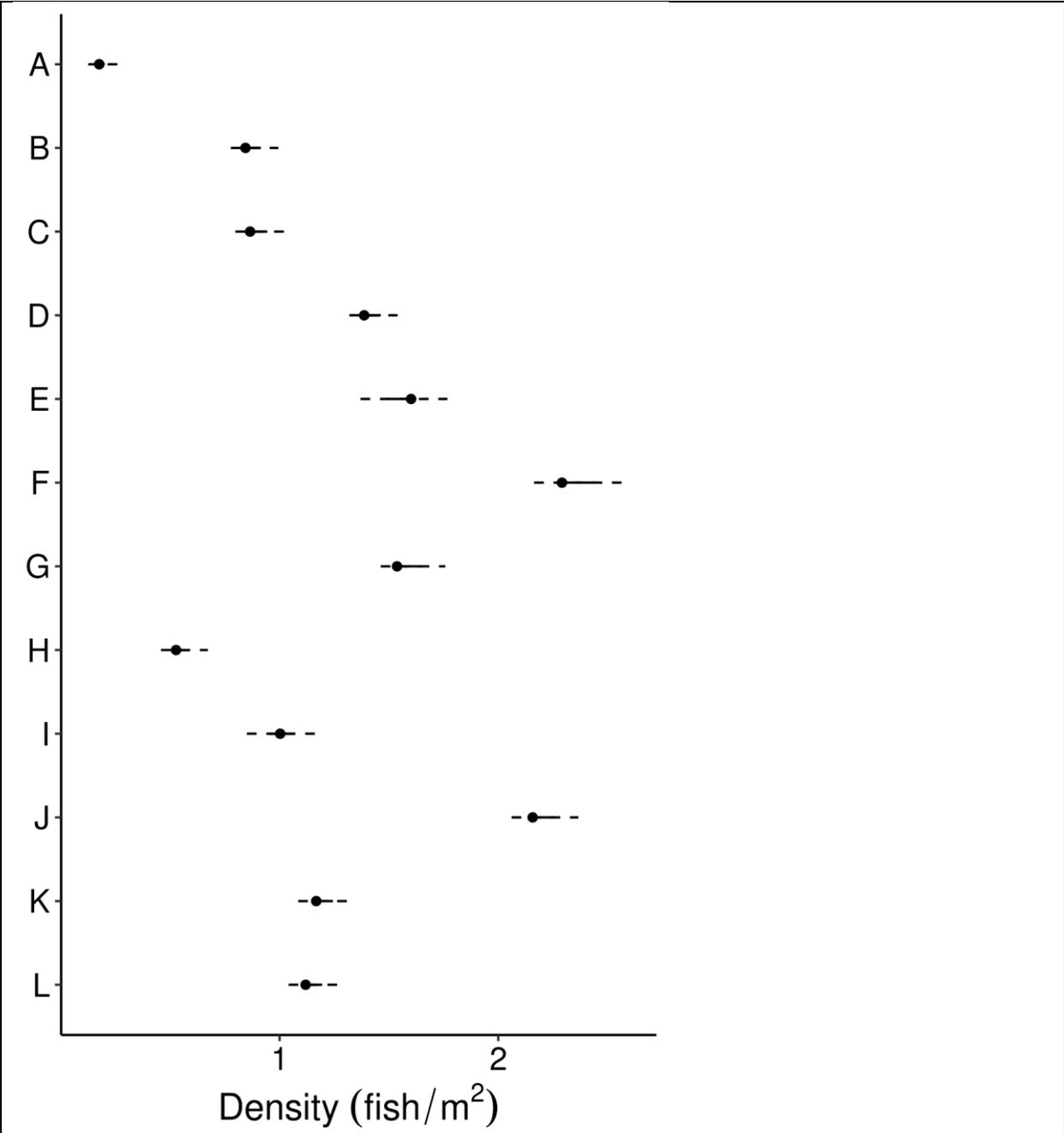
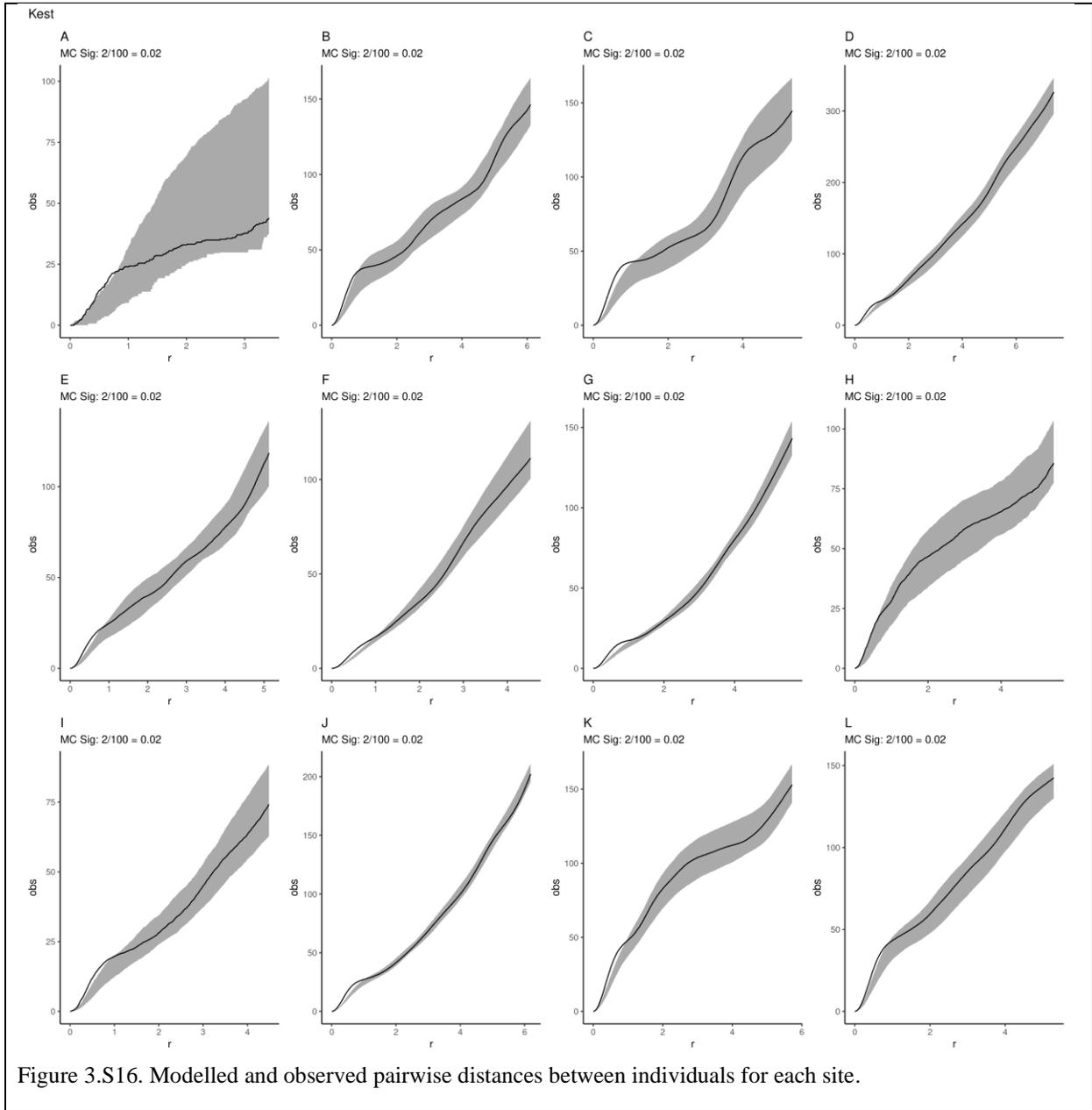
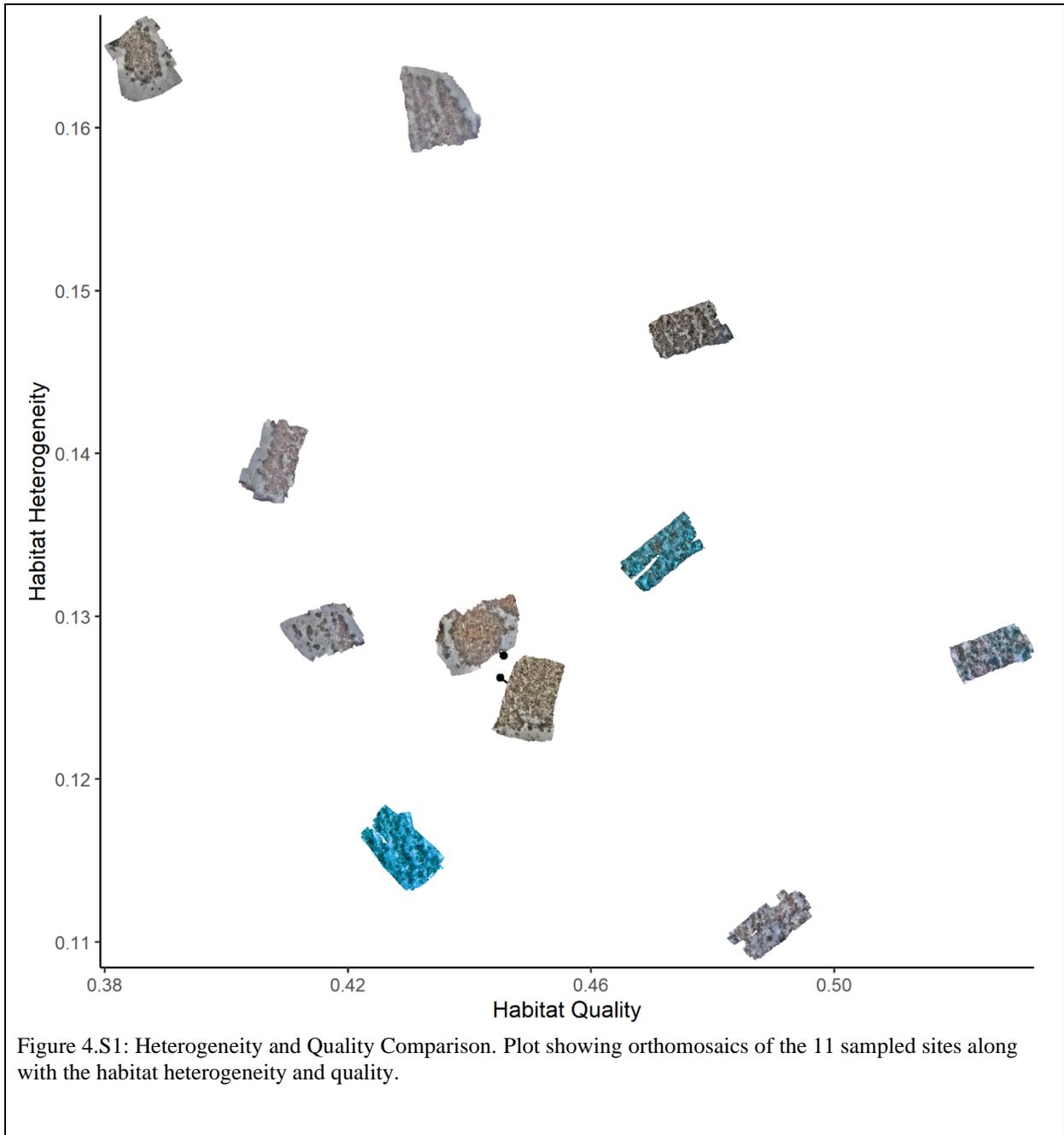


Figure 3.S15. Observed Fish Density with predicted 95% credible interval.



APPENDIX C

SUPPLEMENT FOR CHAPTER 4



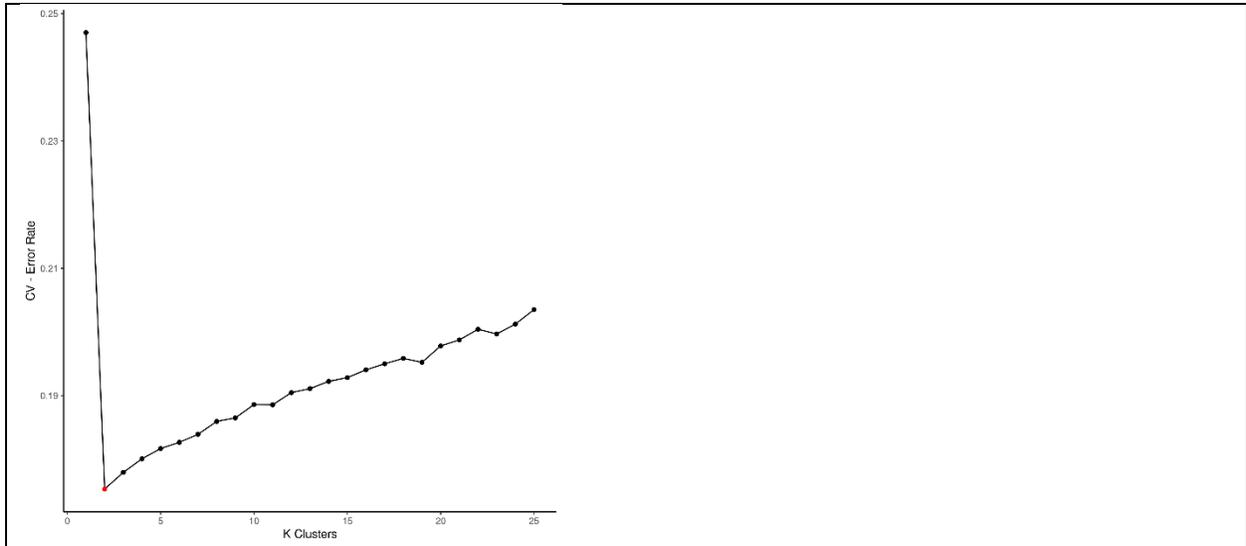


Figure 4.S2: Ten-fold cross-validation error of ADMIXTURE assignments.

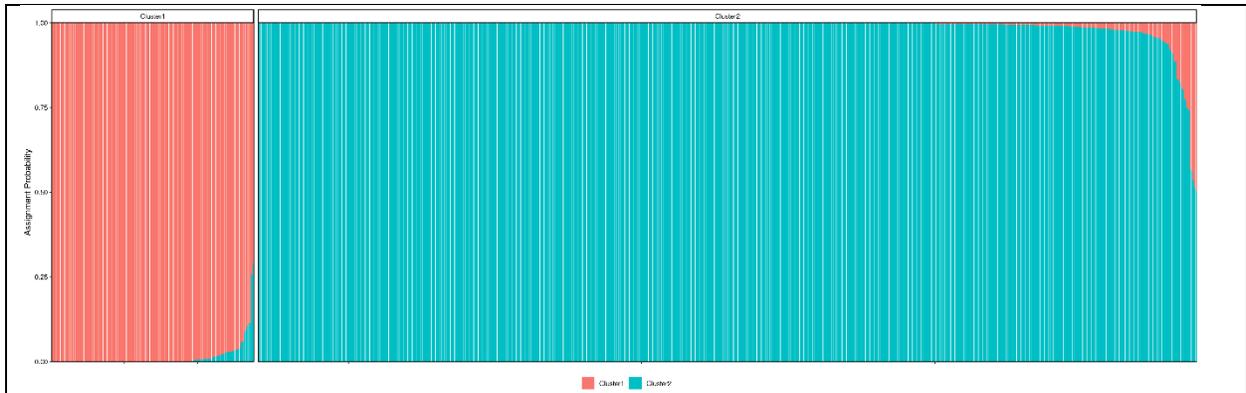


Figure 4.S3: Admixture plot showing assignment probability of all individuals to cluster 1 and cluster 2. Which after mapping to mitochondrial DNA were determined to represent *Coryphopterus personatus* and *Coryphopterus hyalinus* respectively.

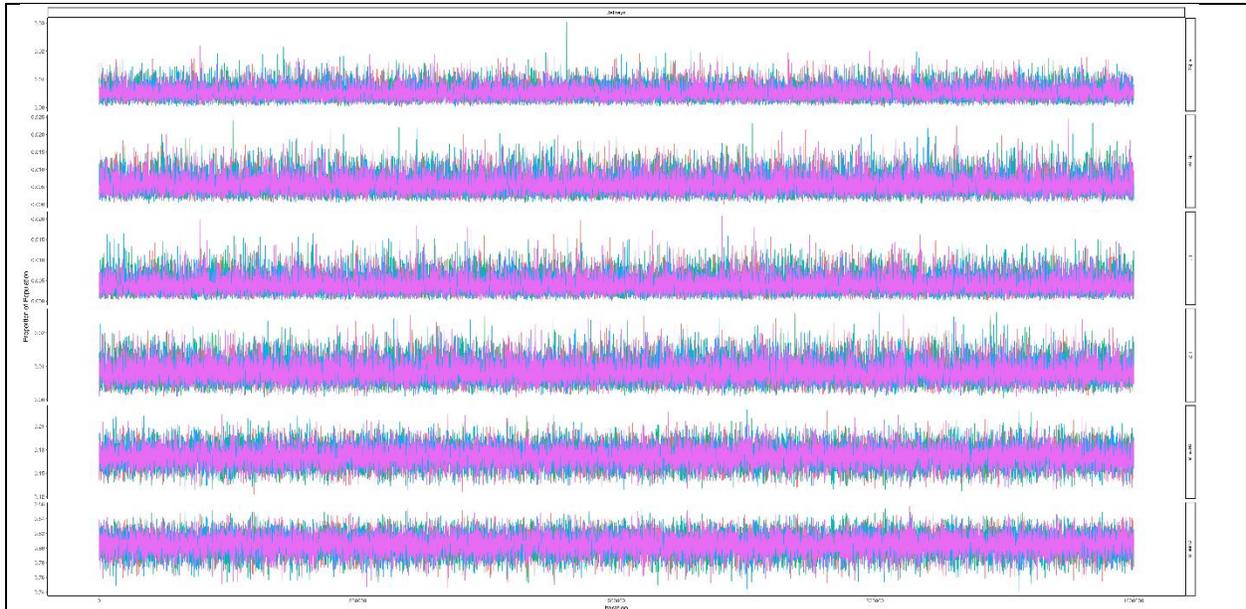


Figure 4.S4: Traceplot of parameter estimates from five independent chains of NEWHYBRIDS.

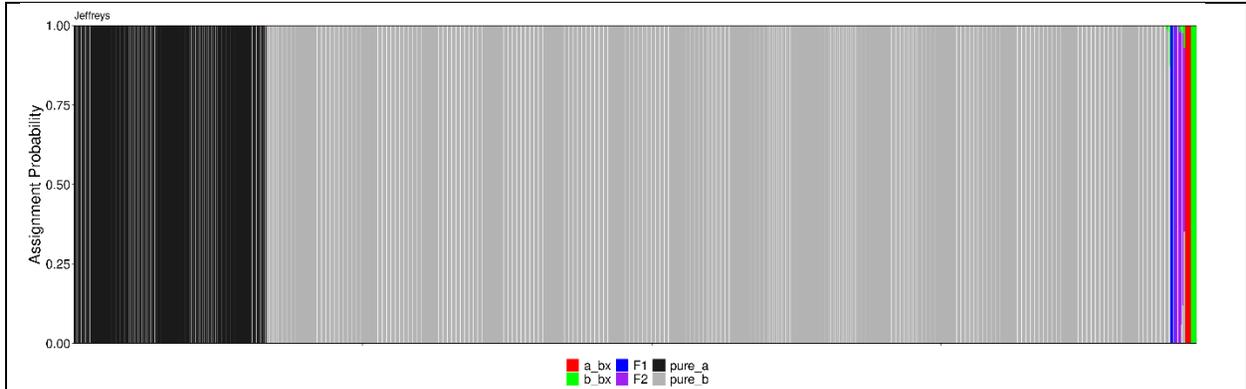


Figure 4.S5: Assignment plot using NewHybrids to identify putative first- and second-generation hybrid individuals. Black bars represent pure *Coryphopterus personatus*, grey bars represent pure *C. hyalinus* assignments, blue indicates F1 hybrids, red shows individuals which are the result of an F1 hybrid backcrossing with *C. personatus*, green shows individuals which are the result of an F1 hybrid backcrossing with *C. hyalinus* and purple show F2 hybrid individuals.

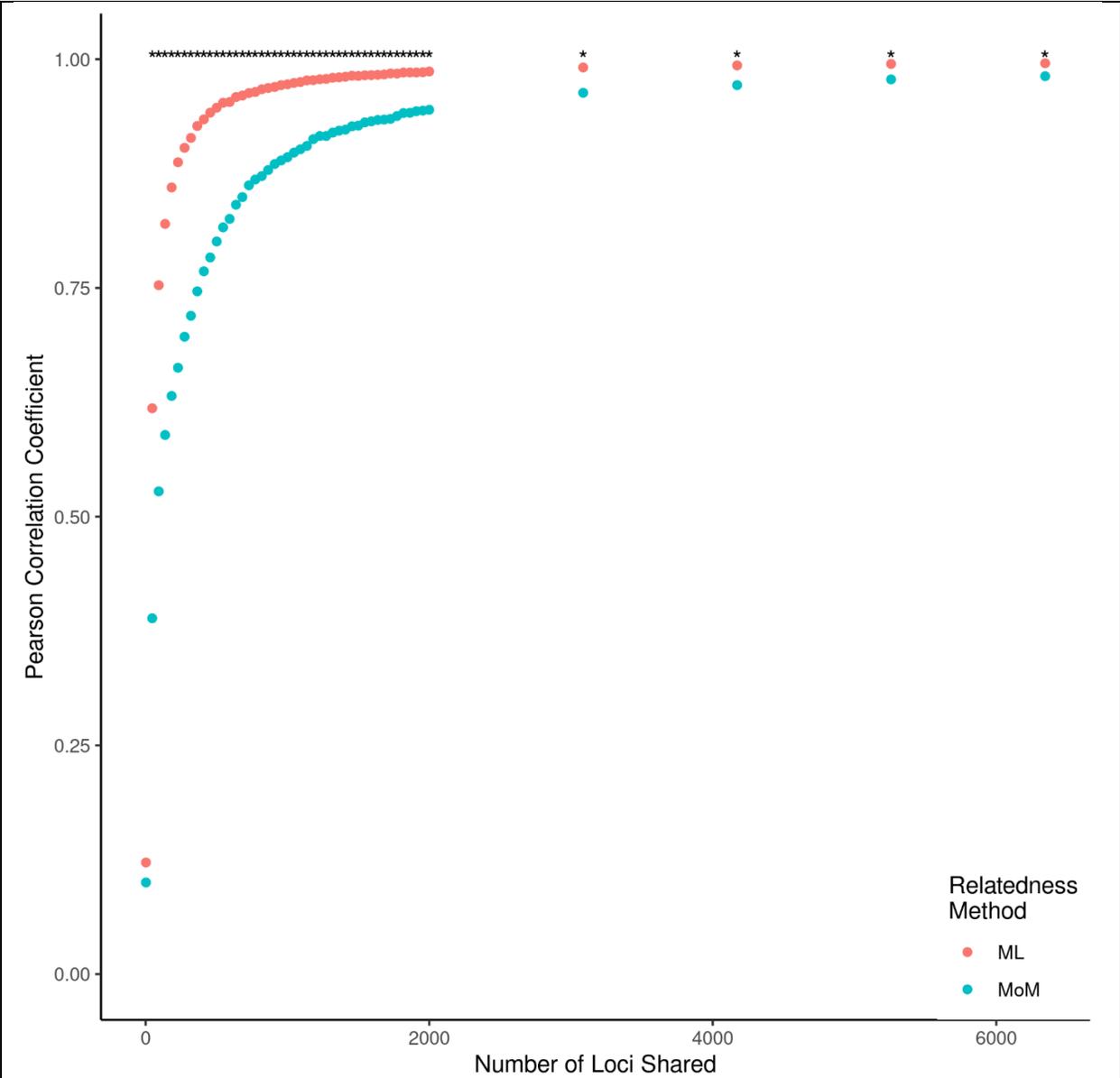


Figure 4.S6: True relatedness vs Simulated relatedness. Plot showing Pearson correlation coefficient of simulated related dyads to the expected true relatedness value using both maximum likelihood (ML, red) and method of moments (MoM, blue) estimation methods given a set number of shared unlinked loci. Asterisks show maximum likelihood – method of moments pairs which significantly differ in their correlation with the expected true relatedness value.

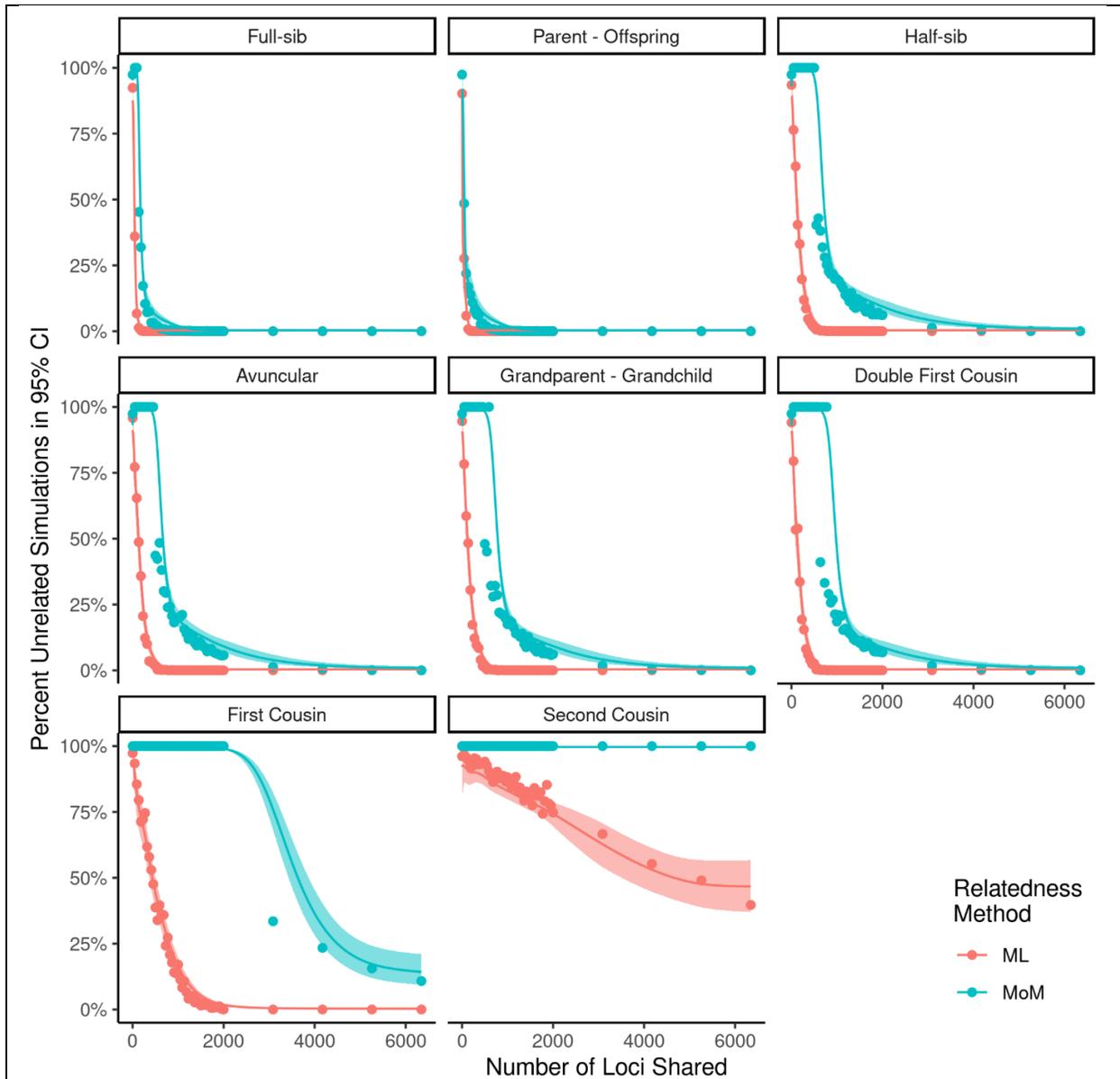


Figure 4.S7: Misclassification of Unrelated Dyads. Plot showing the percentage of unrelated dyads incorrectly included within each relationship class given the relationship class and the number of unlinked shared loci. Both maximum likelihood (ML, red) and method of moments (MoM, blue) estimation methods are shown). Curve is the best fit spline curve fit using a binomial likelihood family.