5-2017

Hybridization dynamics of a newly discovered parrotfish swarm in the Tropical Eastern Pacific

Robert Barron
rlbarron15@gmail.com

Follow this and additional works at: https://digitalcommons.bowdoin.edu/honorsprojects

Part of the Computational Biology Commons, Genetics Commons, Genomics Commons, and the Molecular Genetics Commons

Recommended Citation
https://digitalcommons.bowdoin.edu/honorsprojects/58

This Open Access Thesis is brought to you for free and open access by the Student Scholarship and Creative Work at Bowdoin Digital Commons. It has been accepted for inclusion in Honors Projects by an authorized administrator of Bowdoin Digital Commons. For more information, please contact mmcderm2@bowdoin.edu.
Hybridization dynamics of a newly discovered parrotfish swarm in the Tropical Eastern Pacific

An Honors Paper for the Department of Biology

By Robert Louis Barron

Bowdoin College, 2017

©2017 Robert Louis Barron
TABLE OF CONTENTS

Abstract ................................................................................................................................. 3
Introduction .......................................................................................................................... 4
Methods ................................................................................................................................. 10
  Sites, sampling, and sequencing ....................................................................................... 10
  Genetic analysis ............................................................................................................... 12
Results ................................................................................................................................. 14
  Hardy-Weinberg expectations and linkage disequilibrium ........................................ 14
  Trees and networks ......................................................................................................... 15
  Assignment of individuals ............................................................................................... 15
  The nature and geographic distribution of hybrids ....................................................... 16
Discussion .......................................................................................................................... 21
Acknowledgements ............................................................................................................. 29
Works Cited ......................................................................................................................... 30
Appendices .......................................................................................................................... 35
  Appendix A ...................................................................................................................... 35
  Appendix B ...................................................................................................................... 36
  Appendix C ...................................................................................................................... 37
  Appendix D ...................................................................................................................... 37
  Appendix E ...................................................................................................................... 38
ABSTRACT

Hybrid zones and their dynamics are important in the understanding of the genetic basis of reproductive isolation and speciation. This study seeks to investigate the hybridization dynamics of a *Scarus* hybrid swarm within the Tropical Eastern Pacific (TEP) that includes four phenotypically distinct species: *S. perrico*, *S. ghobban*, *S. rubroviolaceus*, and *S. compressus*. Genetic and population structure analyses of four nuclear loci and a mitochondrial locus revealed that one of the four species, *S. compressus*, was the result of two different hybrid crosses: *S. perrico* × *S. rubroviolaceus* and *S. perrico* × *S. ghobban*. A NewHybrids model indicated that most of the *S. compressus* samples were F1 hybrids, but 21% of the *S. compressus* sample was classified as “parentals” which could also be explained by the presence of either F2 hybrids or backcrosses with *S. compressus* phenotypes, given the relatively low power of the nuclear data set (4 loci) to resolve complex hybrid genotypes. Significant mito-nuclear discordance in all three non-hybrid species is consistent with an evolutionary effect of backcrossing between F1 hybrids and “pure” species. This study reveals a relative ease of hybridization between parrotfish taxa separated by an estimated 4.5 million years of isolation and opens the door to further studies on the potential effects of gene flow across old species boundaries and perhaps the formation of new species by hybrid speciation in a diverse clade of tropical reef fish. Elucidating the nature of potentially “deep” F2 crosses and backcrosses within the TEP *Scarus* hybrid system will allow us to better understand the effects of hybridization on evolution and speciation on both a micro- and macro-ecological scale.
INTRODUCTION

Hybrid zones, the geographic regions where different species meet and interbreed, are invaluable biological systems for ecologists and evolutionary biologists (Harrison 1993). These regions have been studied to better understand a variety of concepts, including the evolution of reproductive isolation, how ecological factors promote or breakdown species boundaries, and how interbreeding between species affects the formation of new species (Endler 1977; Barton and Hewitt 1985, 1989; Seehausen 2004; Mallet 2005, 2007). Hybrid zones, therefore, play a central role in understanding micro- and macro-evolutionary processes.

As species meet and interbreed within hybrid zones, the divergent species or populations exchange genes. Consequently, hybrid zones are often characterized by changes in a variety of genetically–determined characters and consist of clusters of parallel gradients in gene frequencies, otherwise known as clines (Barton & Hewitt 1989). The width of genetic clines may vary greatly, depending on the biological system and ecological context. Hybrid zones are typically only a few hundred meters wide but can reach up to several hundred kilometers in length (Barton & Hewitt 1985, 1989). While the shape and width of hybrid zones can vary, they are theoretically thought to be maintained by an equilibrium between the dispersal of organisms away from the hybrid zone center and natural selection that may be acting against hybrids (Barton and Hewitt 1985; Barton and Bengtsson 1986). Two types of hybrid zones vary in how the environment affects the fitness of progeny from hybrid crosses. Extrinsic zones are sustained by spatially varying natural selection, while intrinsic zones are maintained by hybrid inviability or sterility that is independent of environmental
conditions, enabling them to move spatially and in a manner that minimizes their length (Barton & Hewitt 1985; Barton and Bengtsson 1986).

The potential for selection against hybrid offspring is a key factor in determining the dynamics and persistence of hybrid zones. Interspecific hybrids can be sterile if parental species that meet and interbreed have a different number of chromosomes, as in many terrestrial plant systems. Progeny with odd numbered chromosome structures exhibit disrupted meiotic processes as appropriate pairing and segregation of chromosomes is prevented and viable gametes cannot be produced (Rieseberg 2001; Mallet 2007). However, hybrids exhibit a range of reproductive potential and some are biologically successful; successful hybrids can even colonize unoccupied ecological niches or adaptive peaks (Mallet 2007). Hybrid zones consequently exhibit variability in both fitness and morphology (Endler 1977; Barton & Hewitt 1985).

Further, rather than thinking of hybrid zones as places that reduce biological diversity by potentially blurring species boundaries, they may also be thought of as geographic regions that create evolutionary novelty. With a greater variance in genotypic and phenotypic frequencies, hybrid zones give rise to increasing functional diversity (Seehausen 2004). Given the variation in functional traits associated with hybrid zones, hybrids will often exhibit either novel or extreme phenotypes compared to the parental taxa, otherwise referred to as transgressive segregation (Seehausen 2004; Mallet 2007). Transgressive segregation, commonly observed in interspecific hybridization, is mostly a result of segregation variance. This variance is caused by the complementary effects of different genetic loci fixed for alleles that act in opposite directions in the parental taxa but exhibit additive effects when recombined in their hybrids (Rieseberg et al. 1999; Seehausen 2004). The resulting novel
genotypes allow hybrid species to occupy different spatial, temporal, or behavioral niches. This ecological niche partitioning enables hybrids to become both genetically stabilized and reproductively isolated from respective parental taxa, leading to hybrid speciation (Seahausen 2004; Mallet 2005, 2007).

In marine systems, hybrid zones are of particular interest because of two common life history traits. First, many marine macroalgae, invertebrates, and fish have highly disperse spores or larvae, remaining in the planktonic stage for weeks or months before returning to the benthos to settle. Perhaps not surprisingly, hybrid zones in marine systems have been observed to extend over 1000s of kilometers (reviewed by Sokta & Palumbi 2006). However, some marine hybrid zones are narrower than predicted by planktonic larval durations (Nielsen et al. 2003; Reginos & Cunningham 2005) or even exhibit a mosaic, patch-like structure on scales of 10s or 100s of meters (Bierne et al. 2003), suggesting that spatially varying selection can play a key role in marine hybrid zone examples. Thus, if we assume that the spatial structure of marine hybrid zones is maintained by a balance between dispersal and natural selection, then wide hybrid zones suggest that dispersal is high and either: (i) selective gradients are fairly weak, in the case of hybrid zones maintained by extrinsic processes, or (ii) hybrid offspring have comparable fitness to non-hybrid phenotypes, in the case of hybrid zones maintained by intrinsic mechanisms.

The second life-history trait of many marine organisms is external fertilization in the sea. The process of ejecting both or one gamete type into the external environment allows for the possibility of heterospecific gametes to mix freely during spawning events that involve multiple species, or “mass spawning”. Mass spawning occurs in tropical reef corals (Carlon 1999), tropical reef fish (Claydon 2004), and tropical macroalgae (Clifton 1997). Not
surprisingly, some species have systems of gamete recognition (Palumbi 2008) that apparently prevent extensive hybridization during mass or group spawning events, but the existence of gamete recognition is unknown in all but a few model systems. Further, the ecological effectiveness of gamete recognition is an area of active research (Bierne et al. 2002). Mass or group spawning in tropical marine systems clearly opens up opportunity for fertilization among different species, many of which may be relatively closely related.

Until very recently, hybridization in tropical reef fish was thought to be rare. Hybrid species often went undetected due to the similar morphologies of closely related species (DiBattista et al. 2016). Consequently, hybridization frequency was once thought to be inversely correlated with the number of species in a given area. The high level of diversity of reef systems was presumed to give organisms enhanced species recognition capabilities and specialized reproductive responses and behaviors, leading to a low hybridizing frequency (Hubbs 1955; DiBattista et al. 2016). However, novel molecular techniques have shown otherwise, and hybridization has been observed in systems of closely related reef fish species, including surgeonfish (DiBattista et al. 2016), clownfish (Gainsford et al. 2014), and butterflyfish (Montanari et al. 2014).

My honors work is focusing on a newly described system of hybridization among three parrotfish species that live in the Tropical Eastern Pacific (TEP). Because more than two species are involved, I use the term “hybrid swarm” to refer to this specific system. The TEP, defined by the WWF and The Nature Conservancy, is one of 12 marine realms that cover coastal shallows and shelves of the world. This region extends along the Pacific Coast of the Americas, from the central Gulf of California, southward to Ecuador, containing offshore island groups including the Galapagos (Spalding et al. 2007). Due to the distance
between the TEP and the islands of the Central Pacific, the TEP is known for its exceptional marine endemism (Robertson & Kramer 2009). In fact, in his “Origin of Species” Darwin (1872) recognized that the 4,000-mile distance of deep water between the Central Pacific and the TEP is one of the largest barriers to marine organisms on Earth. Thus, it is an interesting marine region because it contains both endemic fauna, as well as “trans-pacific” species that occasionally cross this massive barrier (Lessios & Robertson 2006).

The parrotfishes (Family, Scaridae, Subclade, Scarinae) are a diverse and functionally important ecological group that live on coral reefs and associated shallow water habitats, such as seagrass beds and soft sediments. Their feeding activities include the processing of reef carbonates and the grazing back of fast growing macroalgae, which is associated with the maintenance of coral dominated reefs (Bonaldo et al. 2014). They are a relatively recent player on coral reefs, diversifying into two major clades during the Miocene, Scarus and Chlorurus, followed by punctuated speciation during the Pliocene (Choat et al. 2012). There are over 90 described species in these two genera living on coral reefs today.

In the TEP, preliminary genetic data collected by my advisor, David Carlon, has shown that the three dominant parrotfish species in the genus Scarus (Fig. 1A – C) are hybridizing in different proportions to produce a fourth species (Fig. 1D). These data indicate that two different crosses are occurring between S. perrico (Fig. 1A) and either S. ghobban (Fig. 1B) or S. rubroviolaceus (Fig. 1C). These two crosses: S. perrico × S. ghobban and S. perrico × S. rubroviolaceus - produce variants of what was previously thought to be a fourth biological species in this region: S. compressus (Fig. 1D) that was described early in the 20th Century by Osburn et al. (1916). The four “species” (including the hybrid form) have a broad
distribution across the TEP: from the Galapagos Islands in the south to Baja California Sur in the north (Allen & Robertson 1994).

Figure 1. Four parrotfish species in the genus *Scarus* from the Tropical Eastern Pacific. A. *S. perrico*, B. *S. ghobban*, C. *S. rubroviolaceus*, D. *S. compressus*. Terminal phase coloration is in the top row of photos, while intermediate phase coloration is in the bottom row. *S. perrico* (A) and *S. compressus* (D) are TEP endemics, while *S. ghobban* (B) and *S. rubroviolaceus* (C) span the Indian and Pacific oceans: from the West coast of Africa to the East coast of Central America and the Galapagos.

My honors thesis examines the dynamics of this hybrid swarm across three sites in the TEP ranging from Baja California to Panama. With robust sample sizes, and data from a mitochondrial gene and four nuclear introns, I address three principal questions:

1. What is the evolutionary “depth” of this hybrid swarm, in terms of the frequency of F1 hybrids, F2 hybrids, and backcrossing into the parental populations?
2. Does the structure of the hybrid swarm vary among the three geographic sites?
3. Does hybridization depend on evolutionary divergence?

With combined mitochondrial and nuclear data, I show that hybridization is more frequent between divergent species than between the two closely related species. While there is evidence for possible F2 hybrids and backcrosses occurring within this zone, I lack the statistical power to properly predict the evolutionary “depth” of this system.
METHODS

Sites, sampling, and sequencing

Between 2014-2016, whole fish were collected by spear gun from three different locations in the TEP, spanning Baja California to Panama (Fig. 2). At La Ventana and Pixvae, all four species occurred over shallow rocky reefs, but at the Perlas Islands in Panama, *S. rubroviolescit* was very rare (Carlon, unpublished data). At each location, the abundance of the hybrid phenotype *S. compressus* was about 10% that of the other species. Collecting and export permits were obtained from the government agencies of Mexico and Panama. From each fish, a fin clip or liver sample was taken for genetic analysis, and additional morphometric and reproductive sampling were completed on the La Ventana samples. In a few cases, scales were used from unsuccessful spearing. Fin clips, scales, and liver samples were stored in either 95% ethanol or DMSO for DNA extraction. DNA from liver samples was obtained using a phenol chloroform extraction, according to the protocol used by Sambrook & Russell, 2001. DNA from both fin clips and scales were obtained using a Qiagen DNeasy tissue extraction kit (Qiagen, Valencia, CA). DNA was quantified with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) and normalized to 5µg/ml.
Four nuclear introns and the mitochondrial control region were amplified via PCR with the primers listed in Table 1. PCR reactions were carried out in 12.5µl OneTaq MasterMix (New England Biolabs, Ipswich, MA), 5µl of both forward and reverse primers, 10.5µl nanopure water, and 1µl of concentration-normalized DNA extract. PCR was performed in a C1000Touch Thermal Cycler (Bio-Rad, Hercules, CA) with the following cycling parameters: 3 min ramp at 94°C followed by 30 cycles of 30 sec at 94°C, 1 min at 55°C, 1 min at 68°C. For the mtCR gene, the annealing temperature was 52°C. PCR products were verified on a Lonza FlashGel System (Lonza, Rockland, ME) and/or via agarose gel electrophoresis, and samples were purified with an Exo-SAP protocol (Affymetrix) before Sanger sequencing (GeneWiz, South Plainfield, NJ).
Table 1. Primer identities for sequences of interest for *Scarus* sp. PCR amplification.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rag2-38F</td>
<td>Forward: GAAAAGAGTGTTTGAAATGA</td>
<td>715 bp</td>
<td>Smith <em>et al.</em>, 2008</td>
</tr>
<tr>
<td>Rag2-355R</td>
<td>Reverse: CATCGTGCTCCTGGGTGACAAAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tmo-f1-6</td>
<td>Forward: GAAAAGAGTGTTTGAAATGA</td>
<td>485 bp</td>
<td>Smith <em>et al.</em>, 2008</td>
</tr>
<tr>
<td>Tmo-r1-3</td>
<td>Reverse: CATCGTGCTCCTGGGTGACAAAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dlx2-F760</td>
<td>Forward: GAAGAGAGYGGAGCCAGAAATC</td>
<td>522 bp</td>
<td>Smith <em>et al.</em>, 2008</td>
</tr>
<tr>
<td>Dlx2-R2</td>
<td>Reverse: AGTTTGCCAAAAACGACGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bmp4-2F34</td>
<td>Forward: CACACCTCTTCGCTTCCCTGT</td>
<td>488 bp</td>
<td>Smith <em>et al.</em>, 2008</td>
</tr>
<tr>
<td>Bmp4-2R375</td>
<td>Reverse: TGGTGCGGTGAAGTCTTGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mtCR-L15995</td>
<td>Forward: AATTCTCACCCTAGCTCCCAAA</td>
<td>350-400 bp</td>
<td>Lee <em>et al.</em>, 1995</td>
</tr>
<tr>
<td>mtCR-H16498</td>
<td>Reverse: CCTGAAGTAGGAACCAGATG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Genetic analysis**

All sequences were edited and aligned using Geneious software (Biomatters). For the nuclear introns, heterozygous sites were called with ambiguity codes using the “find heterozygotes” tool and confirmed visually. The resulting alignments were then phased using PHASE 2.1.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003). At each locus, individual alleles were called by grouping phased sequences by exact identity. Exact tests for Hardy-Weinberg equilibrium and for linkage disequilibrium were estimated for the nuclear loci for each species less the hybrid *S. compressus* using Genepop (Raymond & Rousset 1995). For both mitochondrial and nuclear sequences, maximum likelihood trees were constructed with PHYML and a GTR model (Guindon 2010). PopART software was used to build a TCS haplotype network using the mitochondrial data (popart.otago.ac.nz).

To determine the levels of admixture among species, the nuclear data set (N = 236) was used in STRUCTURE v2.3 (Pritchard *et al.* 2000). Five independent runs were conducted for each model where K = 1 to K = 6, using the admixture model, correlated allele frequencies, 30,000 burn-in steps followed by $10^4$ iterations. The statistic Delta K (Evanno *et al.* 2005) was used to determine the best fitting model, which turned out to be K = 3 (see
Results). A second model was therefore run with K = 3 and the “pop-flag” option that included phenotypic information for all three “non-hybrid” species: *S. perrico*, *S. ghobban*, and *S. rubroviolaceus*. This model with priors was run independently 5 times, with the same options as the non-prior model except allele frequencies were not correlated. To align clusters and average assignments among runs I used the graphical software PopHelper (Francis 2017), which implements the algorithm CLUMPP (Jakobsson and Rosenberg 2007).

The mitochondrial haplotype network and the STRUCTURE models clearly indicated that the phenotype *S. compressus* was the result of hybrid crosses between either: *S. perrico* × *S. ghobban* or *S. perrico* × *S. rubroviolaceus*. To determine the depth of such crosses, e.g. whether they were first generation F1 hybrids, F2 hybrids (F1 hybrids × F1 hybrids), or backcrosses between F1 hybrids and the parental taxa, I used the Bayesian model of NewHybrids (Anderson & Thompson 2002) to identify these specific hybrid classes. Two different models were run, assuming different parental species and the hybrids that were the most likely offspring of those parents. I identified hybrids by classifying each *S. compressus* sample with the STRUCTURE model output (with priors). If a *S. compressus* phenotype had Q values > 0.10 for more than two clusters, it was considered a genetic hybrid. Further, the two clusters with the highest Q values determined whether it was analyzed in the *S. perrico* × *S. ghobban* model vs. the *S. perrico* × *S. rubroviolaceus* model. Note that average assignment for each species was high for each of the three clusters (Fig. 5).
RESULTS

Out of a sample size of N = 280, the number of individuals that had successful DNA amplification and sequencing with high enough quality to compare the results with other sequences for each gene and *Scarus* sp. are shown in Table 2. The analyzed loci exhibit variable degrees of polymorphism among the samples; the most variable locus was *Rag2* with 51 alleles, followed by *mtCR*, *Dlx2*, *Tmo*, and *Bmp4* with 47, 27, 26, and 9 alleles respectively (Table 2).

<table>
<thead>
<tr>
<th></th>
<th><em>mtCR</em></th>
<th><em>Rag2</em></th>
<th><em>Bmp4</em></th>
<th><em>Dlx2</em></th>
<th><em>Tmo</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. perrico</em></td>
<td>63</td>
<td>61</td>
<td>66</td>
<td>64</td>
<td>42</td>
</tr>
<tr>
<td><em>S. ghobban</em></td>
<td>47</td>
<td>40</td>
<td>46</td>
<td>46</td>
<td>32</td>
</tr>
<tr>
<td><em>S. rubroviolaceus</em></td>
<td>51</td>
<td>43</td>
<td>50</td>
<td>51</td>
<td>34</td>
</tr>
<tr>
<td><em>S. compressus</em></td>
<td>42</td>
<td>43</td>
<td>50</td>
<td>50</td>
<td>31</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>203</td>
<td>187</td>
<td>212</td>
<td>211</td>
<td>139</td>
</tr>
<tr>
<td>Number of haplotypes</td>
<td>47</td>
<td>51</td>
<td>9</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

*Hardy-Weinberg expectations and linkage disequilibrium*

Deviations from Hardy-Weinberg equilibrium were examined within each species less the hybrid *S. compressus*, at each of the 4 nuclear loci used for population structure analysis (Appendix A). The $F_{IS}$ values calculated were positive and highly significant for all nuclear loci tested for each species ($p < 0.01$), except for the *Tmo* locus in *S. rubroviolaceus*. Analysis of linkage disequilibrium between the four nuclear loci suggested that linkage disequilibrium was greatest in *S. rubroviolaceus* with 5 of 6 comparisons being significant,
followed by S. perrico and S. ghobban with 4 of 6 and 1 of 6 significant comparisons respectively (Appendix B).

Trees and networks

Maximum likelihood analyses yielded mtDNA and nuclear trees with congruent topologies that all divided the samples into three major clades grouped by dominant species (Fig. 3). All clades share haplotypes, but with greater frequency in the S. ghobban and S. rubroviolaceus clades.

An analysis of the mtCR sequence data identified three different clusters of haplotypes, that I call mitotypes (Fig. 4). The haplotype network indicates that haplotypes sampled from S. perrico, S. ghobban, and S. rubroviolaceus each formed their own dominant clusters, corresponding to the three clades identified in the maximum likelihood phylogeny (Fig. 3B). The S. perrico and S. rubroviolaceus clusters and the S. perrico and S. ghobban clusters are separated by 38 and 52 mutational steps respectively, indicating deep divergence between the variants. Except for one individual, S. compressus samples cluster in both the S. ghobban and S. rubroviolaceus mitotype clusters in relatively similar frequencies.

Assignment of individuals

Bayesian clustering analyses with STRUCTURE software indicated that there are three genetic populations (K = 3) in our sample (Fig. 5). Evaluating models of K = 1 to K = 6 with Delta K unambiguously identifies K = 3 as the best fitting model (Appendix C). The STRUCTURE models with prior population information and without prior information produced remarkably similar results (Fig. 5; Appendix D). In both of these models, S.
*compressus* individuals were identified as hybrids, with admixed genomes mostly made up of either *S. perrico* and *S. ghobban* or *S. perrico* and *S. rubroviolaceus* genotypes.

**The nature and geographic distribution of hybrids**

Mapping the three mitotypes (Fig. 4) onto the nuclear assignments revealed mitochondrial-nuclear discordance in all three parental species (Fig. 6). Some of these individuals had mixed nuclear ancestry (e.g. *S. perrico* and *S. ghobban*), but others did not, suggesting the effects of relatively old backcrossing events from hybrids back into pure parental genomes. In the hybrid phenotype - *S. compressus*, individuals carried either the *S. ghobban* or *S. rubroviolaceus* mitotype, and the species-specific mitotype matched ½ of the admixed nuclear genome. There was also evidence for pure *S. perrico* (1 individual) and pure *S. ghobban* (4 individuals) in the *S. compressus* sample, suggesting a decoupling of phenotype from these specific nuclear genes.

Bayesian assignments into six different hybrid categories through NewHybrids software identified most of the *S. compressus* samples as F1 hybrids (Fig. 7). Eight individuals in the *S. perrico x S. ghobban* model and four from the *S. perrico x S. rubroviolaceus* model exhibit variable probabilities for parental species assignments, representing 21% of the total *S. compressus* sample. However, the NewHybrids assignments to other hybrid categories excluding F1 hybrids is uncertain due to the relatively few number of loci tested.

Hybrid *S. perrico X S. ghobban* and *S. perrico X S. rubroviolaceus* crosses were more common than *S. ghobban X S. rubroviolaceus* crosses at all three sites (Table 3).
Figure 3. Maximum likelihood trees of the the nuclear Bmp4 gene (A), and the mitochondrial control region, mtCR (B). For both genes, species share haplotypes within clades B and C, but dominant species within clades are listed after clade names in panel B.
Figure 4. Haplotype network of control region sequence data. The three mitotype groups correspond to the three clades in Fig. 3.

Figure 5. Structure plots for nuclear intron data and K = 3. Top panel: the model with no prior information on parental species. Bottom panel: the model with prior information on parental species. Each stacked bar on the x-axis represents an individual fish grouped by morphological species. Y-axis is Q, the assignment probability in a given cluster coded by color.
Figure 6. Discordance between nuclear mitochondrial genomes. Top panel: a STRUCTURE plot for all samples, model of $K = 3$, for the nuclear data. Lower four panels are expanded views of nuclear assignments for each of the four species. Capital letters on expanded plots indicate discordant mitotypes among individuals in the parental species; all other individuals carry the common species-specific mitotype indicated in the key. The *S. compressus* plot shows that the two classes of hybrids with admixed nuclear genomes generally have the mitotype of the non-*S. perrico* parental species. Mitotypes correspond to the three major groups illustrated in the haplotype network in Fig. 4. Y-axis is $Q$, the assignment probability in a given cluster coded by color.
Figure 7. Assignment probabilities of *S. compressus* categorized individuals into six hybrid classes from two NewHybrids models run on all four nuclear loci. Top panel: *S. perrico* × *S. ghobban*. Bottom panel: *S. perrico* × *S. rubroviolaceus*.

Table 3. The inferred number of genetic hybrids resulting from three types of crosses among three sites in the TEP. Species abbreviations: Sp, *S. perrico*, Sg, *S. ghobban*, Sr, *S. rubroviolaceus*. Hybrids have Q values > 0.10 for two of the three species-specific clusters found in the STRUCTURE model, and are assigned to a specific cross by the cluster identity of the two highest Q values.

<table>
<thead>
<tr>
<th></th>
<th>Sp x Sg</th>
<th>Sp x Sr</th>
<th>Sr x Sg</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>La Ventana, BCS</td>
<td>10</td>
<td>21</td>
<td>11</td>
<td>42</td>
</tr>
<tr>
<td>Pixbae, Panama</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Perlas, Panama</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>23</td>
<td>32</td>
<td>12</td>
<td>67</td>
</tr>
<tr>
<td><strong>% of total</strong></td>
<td>34%</td>
<td>48%</td>
<td>18%</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

The purpose of this study was to examine the fine-grained genetic structure and dynamics of the *Scarus* hybrid zone at different geographic locations within the TEP. My results corroborate the preliminary research performed by my advisor, showing that the three dominant parrotfish species within the region, *S. perrico*, *S. ghobban*, and *S. rubroviolaceus*, are hybridizing in different proportions to produce what was previously described as a fourth biological species in the region: *S. compressus*. Analysis of the data provides an in-depth understanding of the genetic structure across the hybrid zone.

The studies of pooled subpopulations show highly significant and positive $F_{IS}$ values for nearly all nuclear loci comparisons and significant deviations from Hardy-Weinberg equilibrium. These data suggest that hybridization is occurring within the sampling range as deviations from Hardy-Weinberg are most likely a result of highly prevalent inbreeding rather than random mating between individuals. Observations of high, positive $F_{IS}$ values are consistent with other studies performed on hybrid zone structure, such as the hybrid zone between the westslope cutthroat trout *Oncorhynchus clarki lewisi* and rainbow trout *O. mykiss* in Canadian rivers (Rubidge & Taylor 2004) and the hybrid zone between domestic dogs and the wild wolf *Canis lupus* in central Italy (Randi & Lucchini 2001).

The maximum likelihood trees for each of the nuclear loci and the mitochondrial locus all reveal similar population structural patterns as their respective topologies show congruence among clade composition. All trees grouped the samples into the same three divergent clades, in which each clade is mainly dominated by individuals of one of the three biological species: *S. perrico*, *S. ghobban*, and *S. rubroviolaceus* (Fig. 3). The hybrid *S. compressus* samples were placed in varying frequencies throughout these three clades, with
most being placed within the *S. ghobban* and *S. rubroviolaceus* groupings. These findings are consistent with systematic studies on hybrids and their phylogenetic assignments. McDade (1990; 1992) found that hybrid species are often not readily identifiable as their own unique taxa, but rather are usually placed as basal members of the most apomorphic parent or as members of the most derived parent.

The haplotype network similarly reveals hybrid population structure that is analogous to that found in the maximum likelihood phylogenies. Individuals in this network are grouped into three major mitotypes that are defined by the three biological *Scarus* species, correlating to the three clades derived from the phylogenies (Fig. 4). These mitotypes are highly divergent from one another, indicating deep ancestral evolutionary splits between the different species. One pattern that is interesting to note is that almost all of the *S. perrico* phenotypes carried the *S. perrico* mitotype. This could be indicative of possible asymmetric reproductive isolation between *S. perrico* and the other two biological species, for example if backcrossing back into *S. perrico* was less likely or if such backcrosses have reduced fitness. This pattern suggests that there is something unique about the *S. perrico* side of the hybridization interaction.

The hybrid *S. compressus* samples were shown to either carry the *S. ghobban* or *S. rubroviolaceus* mitotype, generally matching the same species that composed roughly half of their admixed nuclear genome, with a few discordant exceptions. Assuming maternal inheritance of mitochondria as in most animals, this pattern of asymmetric mitochondria capture of only one of the parental genomes could be consistent with a number of hypotheses. For example, it could be that only females of *S. ghobban* and *S. rubroviolaceus* are mating with male *S. perrico*. Parrotfish species have been observed to participate in both
group- and pair-spawning reproductive behaviors, in which some species specialize in one behavior while others partake in both in varying frequencies (Sadovy de Mitcheson & Colin 2011). It is possible that female *S. ghobban* and *S. rubroviolaceus* participate in group-spawning activities in which upon release of their gametes into the water column, male *S. perrico* are fertilizing those gametes in greater frequency than male *S. ghobban* or *S. rubroviolaceus*.

Conversely, it is also possible that male *S. ghobban* and *S. rubroviolaceus* do attempt to fertilize *S. perrico* eggs. Offspring from these species pairings, however, may be observed less often due to resultant reproductive isolating mechanisms that these species may have developed after co-evolving in sympatry (Choat et al. 2012). This hypothesis is consistent with the process of reinforcement in which pre-zygotic isolation between differentiated taxa is caused by natural selection against maladaptive hybridization (Via 2001). The *S. ghobban* and *S. rubroviolaceus* nuclear and mitochondrial genomes may be incompatible to some extent when combined, leading to either inviable gametes or biologically less fit offspring. Under this same reasoning, there could also be selection against the *S. perrico* mitochondrial genome in nuclear backgrounds of either *S. ghobban* or *S. rubroviolaceus*.

The STRUCTURE analysis for individual population assignment gives us a preliminary view of what the evolutionary “depth” of the TEP *Scar* us hybrid swarm looks like. This analysis identified three genetic populations within our sample, corresponding to the three biological *Scar* us species (Fig. 5). Furthermore, the *S. compressus* individuals were mostly all assigned admixed genomes comprised of the other three species. In the plot, there are individuals belonging to a cluster that are actually assigned an admixed genotype or even a pure genotype of a separate cluster. This pattern is suggestive of relatively old backcrossing
events in which hybrid *S. compressus* individuals breeding with a separate *Scarus* species caused the introgression of genes from their own respective parent, being a different *Scarus* species than the one bred with.

Patterns of mitochondrial-nuclear discordance also support the idea of evolutionarily deep backcrossing events. When combining the STRUCTURE data from the nuclear loci with the mitochondrial data, we see that certain individuals exhibit mito-nuclear discordance (Fig. 6). Mito-nuclear discordance is generally defined as a disparity in patterns of differentiation between mitochondrial and nuclear markers, and is apparent within individuals of our sample as they contain a mitotype that is not consistent with their species-specific nuclear genotype (Toews & Brelsford 2012). Mito-nuclear discordance has been shown to be a result of introgressive hybridization resulting from non-neutral processes and intrinsic differences in the inheritance patterns of nuclear and mitochondrial genomes (Gompert *et al.* 2008; Toews & Brelsford 2012).

Demographic asymmetries including sex-biased dispersal patterns or reproduction have been cited as main reasons for mito-nuclear discordance (Rheindt & Edwards 2011). However, it is possible that selective pressures may be acting upon the mitochondrial genome that are independent of interactions with the nuclear genome (Meiklejohn *et al.* 2007). Research has shown that mitochondrial genes are central to the processes of the electron transport chain and oxidative phosphorylation; variation in mitochondrial function could thus have bioenergetic and phenotypic consequences that determine certain aspects of an organism’s life history traits (Ballard & Melvin 2010). It is possible that the individuals in our sample that exhibit mito-nuclear discordance have been subject to differential selection in which has preferentially driven mitochondrial introgression.
The NewHybrids analysis of our samples reveals a mostly shallow evolutionary “depth” of the *Scarus* hybrid swarm, as the majority of hybrid *S. compressus* individuals were assigned high probabilities for F1 hybridization (Fig. 7). However, it is important to emphasize a lack of statistical power of this dataset to discern more complex hybrid classes. Using simulated data, Anderson and Thompson (2002) have shown that the NewHybrids model has difficulty in robustly assigning true backcrosses and F2 individuals with a relatively small number of loci (20) and with alleles that are not completely fixed between species. In this dataset, it is possible that the parental assignments could actually be backcrosses and the more complex mixtures of genotypic assignments could actually be F2 hybrids. This does raise the possibility that there are offspring of F2 mating or backcrosses in our *S. compressus* sample. The fact that *S. compressus* social groups consisting of a single terminal phase individual schooling with a few intermediate phase individuals have been observed on several occasions in Panama and Baja California (D. Carlon, pers. observation) also leaves open the possibility that *S. compressus* is assortatively mating. This hypothesis could be tested with these samples by using a denser panel of nuclear markers.

Given the population assignments for each individual in our sample, it is apparent that most hybrid crosses are occurring between *S. perrico × S. ghobban* and *S. perrico × S. rubroviolaceus* and that this pattern is consistent across all three of our sampling locations (Table 3). This data suggests that the genetic structure and dynamics of our hybrid zone are similar at the northern and central sites tested within the TEP. To see if these patterns are consistent throughout the entire zone, more testing should be performed at more dispersed locations within the TEP, namely more southern sites and the island systems.
The probability of hybridization in this parrotfish swarm does not depend on evolutionary divergence between species. Choate et al. (2012) constructed a time-calibrated species phylogeny of the two genera Scarus and Chlorurus and found that the clade containing S. perrico diverged from the clade containing both S. rubroviolaceus and S. ghobban around 4.5 million years ago. On the other hand, the split between S. rubroviolaceus and S. ghobban is more recent, dated at 2.75 million years ago. The fact that I found that hybridization is occurring in greater frequency between older species pairs (S. perrico and the other two species) than between the two younger species, S. ghobban and S. rubroviolaceus, suggests that hybridization does not depend upon evolutionary divergence within this system. This suggests that although there is substantial genetic divergence between S. perrico and the other two species, the divergence alone is not enough to inhibit hybridization, consistent with observations that accumulation of genetic change does not necessarily induce reproductive isolation (Dobzhansky 1940).

If we assume that reproductive isolation in parrotfish is achieved primarily by prezygotic mechanisms, such as mating behavior, the history of colonization of the TEP can explain the difference in hybridization rates among species pairs. The age and biogeographic distribution of the species within the clade containing S. perrico suggest this lineage evolved in the Caribbean Sea before the complete closure of the Isthmus of Panama, dated at around 3.5 million years ago (O’Dea et al. 2007). After the isolation of the TEP from the Caribbean Sea, S. perrico is the only species remaining in this clade that occurs in the TEP. In contrast, coalescent modeling of the population history of S. ghobban and S. rubroviolaceus shows that the former species migrated from the Central Pacific to the TEP around 300,000 years ago, while the latter species arrived much more recently, around 20,000 years ago (Lessios &
Robertson 2006; Fitzpatrick et al. 2011). These recent migration events suggest that reproductive isolation mechanisms between species originally evolved in the greater Indo-West or Central Pacific, where they commonly occur in sympatry. Upon immigrating to the rocky reefs in the TEP relatively recently, *S. ghobban* and then *S. rubroviolaceus* would have encountered a new habitat and completely unfamiliar species in the form of *S. perrico*. Thus, the ecological context of the evolution of prezygotic isolation could be extremely important for parrotfish in general. In this case there has simply been too little time for effective reproductive isolation to evolve between ecologically unfamiliar species.

Data from this study has confirmed the hybridization dynamics observed through previous research on the *Scarus* hybrid zone by my advisor, and future steps will continue to parse out the genetic structure of the hybrid zone in greater detail. Despite statistical limitations associated with only analyzing four nuclear loci and one mitochondrial locus, I had the power to infer asymmetrical hybridization dynamics and mito-nuclear discordance patterns, start characterizing the evolutionary depth of this hybrid swarm, and assess the impact of genetic divergence on hybridization frequency. Further analysis of this hybrid zone will include genomic UCE data to examine hundreds of polymorphic loci, enabling us to better define the structure of this hybrid zone with increased certainty. Additionally, this analysis will include both a larger sample size as well as a broader range of sampling locations from across the Indo-Pacific. Given that *S. ghobban* and *S. rubroviolaceus* have geographic ranges extending from the TEP through the Indo-Pacific, the hybrid zone I analyzed has the potential to expand across the Pacific Ocean.

This investigation of *Scarus* hybrid zone structure in the TEP provides insights on not only hybridization dynamics, but also on *Scarus* genetic structure, mito-nuclear discordance,
and evolutionary history. The role of hybridization on speciation is still not fully understood, and understanding the dynamics of this hybrid system will provide insights into the porosity of species boundaries in an enigmatic marine system that has rapidly diversified over the last 4.5 million years. The dynamics of this hybrid system will certainly be relevant to understanding how hybridization affects both micro- and macro-evolutionary processes on a broader scale.
ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, David Carlon, for three years of guidance and support throughout this entire project, as well as organizing sampling efforts with Dr. Ross Robertson and Dr. Carlos Armando Sánchez Ortíz, and being willing to travel the entire Pacific to do so. Thank you to Sarah Kingston for serving as my reader and sharing her biostatistical and laboratory expertise. Lastly, I’d like to thank all my family and friends for their continual support and encouragement. This research was funded by the Bowdoin College Henry L. and Grace Doherty Coastal Studies Summer Research Fellowship and the Grua/O’Connell Research Award.
WORKS CITED


Rubidge, E. & Taylor, E. 2004. Hybrid zone structure and the potential role of selection in hybridizing populations of native westslope cutthroat trout (Oncorhynchus clarki lewisi) and introduced rainbow trout (O. mykiss). Molecular Ecology 13(12): 3735-3749.


**APPENDICES**

*Appendix A.* Tests for departures from Hardy-Weinberg expectations in three species of parrotfish. The significance of $F_{IS}$ is indicated by ** < 0.01 or ns = nonsignificant. Data for three sites were combined for each species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Locus</th>
<th>$F_{IS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. perrico</em></td>
<td><em>Rag2</em></td>
<td>0.2693**</td>
</tr>
<tr>
<td></td>
<td><em>Bmp4</em></td>
<td>0.4836**</td>
</tr>
<tr>
<td></td>
<td><em>Dlx2</em></td>
<td>0.6473**</td>
</tr>
<tr>
<td></td>
<td><em>Tmo</em></td>
<td>0.7902**</td>
</tr>
<tr>
<td><em>S. ghobban</em></td>
<td><em>Rag2</em></td>
<td>0.2717**</td>
</tr>
<tr>
<td></td>
<td><em>Bmp4</em></td>
<td>0.5556**</td>
</tr>
<tr>
<td></td>
<td><em>Dlx2</em></td>
<td>0.0428**</td>
</tr>
<tr>
<td></td>
<td><em>Tmo</em></td>
<td>0.8311**</td>
</tr>
<tr>
<td><em>S. rubroviolaceus</em></td>
<td><em>Rag2</em></td>
<td>0.2523**</td>
</tr>
<tr>
<td></td>
<td><em>Bmp4</em></td>
<td>0.4759**</td>
</tr>
<tr>
<td></td>
<td><em>Dlx2</em></td>
<td>0.2329**</td>
</tr>
<tr>
<td></td>
<td><em>Tmo</em></td>
<td>0.2492ns</td>
</tr>
</tbody>
</table>
Appendix B. Tests of linkage disequilibrium (LD) among loci for three species of parrotfish. * < 0.05, ** < 0.01. Data for three sites were combined for each species.

<table>
<thead>
<tr>
<th>S. perrico</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rag2</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dlx2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bmp4</td>
<td>**</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Tmo</td>
<td>ns</td>
<td>**</td>
<td>ns</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S. ghobban</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rag2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dlx2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bmp4</td>
<td>**</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Tmo</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

| S. rubroviolaceus |         |       |       |
|                   |         |       |       |
| Rag2              |         |       |       |
| Dlx2              |         |       |       |
| Bmp4              | *       | **    |       |
| Tmo               | ns      | **    | *     |
Appendix C. The delta K statistic for STRUCTURE modeling runs on nuclear loci across samples (N = 236) for runs with prior population information (top) and without prior information (bottom).

Appendix D. Average difference in meanQ population assignments from STRUCTURE model runs with K=3 for with prior information and without prior information.

<table>
<thead>
<tr>
<th></th>
<th>Δ S. perrico</th>
<th>Δ S. ghabban</th>
<th>Δ S. rubroviolaceus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.0008</td>
<td>0.0036</td>
<td>-0.0027</td>
</tr>
</tbody>
</table>
Appendix E. Color images for Figures 1-7 in black & white.

Figure 1. Four parrotfish species in the genus *Scarus* from the Tropical Eastern Pacific. A. *S. perrico*, B. *S. ghobban*, C. *S. rubroviolaceus*, D. *S. compressus*. Terminal phase coloration is in the top row of photos, while intermediate phase coloration is in the bottom row. *S. perrico* (A) and *S. compressus* (D) are TEP endemics, while *S. ghobban* (B) and *S. rubroviolaceus* (C) span the Indian and Pacific oceans: from the West coast of Africa to the East coast of Central America and the Galapagos.

Figure 2. Map of three sampling locations in the Tropical Eastern Pacific.
Figure 3. Maximum likelihood trees of the the nuclear Bmp4 gene (A), and the mitochondrial control region, mtCR (B). For both genes, species share haplotypes within clades B and C, but dominant species within clades are listed after clade names in panel B.
Figure 4. Haplotype network of control region sequence data. The three mitotype groups correspond to the three clades in Fig. 3.

Figure 5. Structure plots for nuclear intron data and K = 3. Top panel: the model with no prior information on parental species. Bottom panel: the model with prior information on parental species. Each stacked bar on the x-axis represents an individual fish grouped by morphological species. Y-axis is Q, the assignment probability in a given cluster coded by color.
Figure 6. Discordance between nuclear mitochondrial genomes. Top panel: a STRUCTURE plot for all samples, model of $K = 3$, for the nuclear data. Lower four panels are expanded views of nuclear assignments for each of the four species. Capital letters on expanded plots indicate discordant mitotypes among individuals in the parental species; all other individuals carry the common species-specific mitotype indicated in the key. The *S. compressus* plot shows that the two classes of hybrids with admixed nuclear genomes generally have the mitotype of the *non-S. perrico* parental species. Mitotypes correspond to the three major groups illustrated in the haplotype network in Fig. 4. Y-axis is $Q$, the assignment probability in a given cluster coded by color.
Figure 7. Assignment probabilities of *S. compressus* categorized individuals into six hybrid classes from two NewHybrids models run on all four nuclear loci. Top panel: *S. perrico* × *S. ghobban*. Bottom panel: *S. perrico* × *S. rubroviolaceus*. 