MR. JONATHAN LYON WHITNEY (Orcid ID : 0000-0001-6455-0640)

Article type : Original Article

**Flickers of speciation? Sympatric color morphs of the arc-eye hawkfish,
Paracirrhites arcatus, reveal key elements of divergence-with-gene-flow**

Jonathan L. Whitney^{1,2,*}, Brian W. Bowen¹, and Stephen A. Karl¹

¹Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa
P.O. Box 1346, Kāne'ohe, Hawaii 96744, USA

* Corresponding author: Jonathan.Lyon.Whitney@gmail.com (JLW)

Running Head: Flickers of speciation in hawkfish

Key Words: local adaptation, color polymorphism, reproductive isolation, incipient speciation,
divergence with gene flow, coral reef fish

Abstract

² Present address: Joint Institute for Marine and Atmospheric Research, University of Hawai'i at
Mānoa, 1000 Pope Road, Marine Sciences Building Rm 312, Honolulu, Hawai'i, 96822, USA

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/mec.14527](https://doi.org/10.1111/mec.14527)

This article is protected by copyright. All rights reserved

One of the primary challenges of evolutionary research is to identify ecological factors that favor reproductive isolation. Therefore, studying partially isolated taxa has the potential to provide novel insight into the mechanisms of evolutionary divergence. Our study utilizes an adaptive color polymorphism in the arc-eye hawkfish (*Paracirrhites arcatus*) to explore the evolution of reproductive barriers in the absence of geographic isolation. Dark and light morphs are ecologically partitioned into basaltic and coral microhabitats a few meters apart. To test whether ecological barriers have reduced gene flow among dark and light phenotypes, we evaluated genetic variation at 30 microsatellite loci and a nuclear exon (*Mc1r*) associated with melanistic coloration. We report low, but significant microsatellite differentiation among color morphs and stronger divergence in the coding region of *Mc1r* indicating signatures of selection. Critically, we observed greater genetic divergence between color morphs on the same reefs than between the same morphs in different geographic locations. We hypothesize that adaptation to the contrasting microhabitats is overriding gene flow and is responsible for the partial reproductive isolation observed between sympatric color morphs. Combined with complementary studies of hawkfish ecology and behavior, these genetic results indicate an ecological barrier to gene flow initiated by habitat selection and enhanced by assortative mating. Hence the arc-eye hawkfish fulfill theoretical expectations for the earliest phase of speciation-with-gene-flow.

Introduction

One of the current challenges of evolutionary research is to investigate ecological factors that promote the early stages of reproductive isolation (Nosil and Feder 2012; Marie Curie Speciation Network *et al.* 2012; Shaw & Mullen 2014). Understanding the extent to which ecological barriers can limit gene flow within species can shed light on the relative role of selection during the early stages of population divergence and speciation. Therefore, study of partially isolated taxa has the potential to provide novel insight into the ecological factors that drive the initiation of reproductive barriers.

56 There is a growing appreciation that ecological adaptation may be contributing to the
57 incredible radiation of coral reef fishes in tropical seas (Briggs 2005; Rocha & Bowen 2008;
58 Bernardi 2013; Bowen *et al.* 2013). The case for ecology driving reef fish speciation was initially
59 supported by phylogenetic studies showing sister lineages separated by ecological factors in
60 parrotfishes (Scaridae; Streelman *et al.* 2002), puffers (Tetraodontidae; Alfaro *et al.* 2007), and
61 gobies (Gobiidae; Rüber *et al.* 2003; Taylor & Hellberg 2005). Rocha *et al.* (2005) offered a
62 noteworthy case of parapatric ecological speciation in Atlantic wrasses (Labridae; *Halichoeres*
63 spp.) by showing strong genetic partitions between adjacent but ecologically distinct habitats.

64 Studies combining ecology, morphology and genetics have shown that divergent
65 selection in contrasting marine environments can create barriers to gene flow even in sympatry
66 (Munday *et al.* 2004; Puebla *et al.* 2007; Bongaerts *et al.* 2011; Bird *et al.* 2011a; Puebla *et al.*
67 2014). The critical ingredients needed for ecological speciation-with-gene-flow are divergent
68 selection, a mechanism of reproductive isolation and a link between the two (Nosil 2012).
69 Comprehensive studies that have identified all three ingredients and the ecological trait under
70 selection are extremely limited. Two of the more compelling marine fish examples involve host
71 shift in coral-dwelling *Gobiodon* gobies (Munday *et al.* 2004) and aggressive mimicry in color
72 polymorphic *Hypoplectrus* hamlets (Puebla *et al.* 2007). These studies highlight the potential
73 mechanisms by which reproductive isolation can result from divergent selection on key
74 ecological traits. As exemplified by the coral gobies, when individuals show strong habitat
75 preference, and mating takes place in that habitat, reproductive isolation can result (Munday *et*
76 *al.* 2004). As seen in the hamlets, color is a potentially important pleiotropic trait as it can be
77 both under disruptive selection and used as a cue in assortative mating (Puebla *et al.* 2007).
78 Scenarios where the trait under selection is linked directly with mate choice are considered the
79 most plausible paths to ecological speciation (Maynard-Smith 1966; Gavrilets 2004).

80 The few examples of speciation-with-gene-flow in marine systems may not be due to
81 their rarity, but rather the difficulty in diagnosing a mode of speciation that has already occurred.
82 Identifying modes of speciation in wild populations is inherently challenging, as the processes
83 that led to the formation of most species are often obscured by time. Here, we use the natural
84 color polymorphism in the arc-eye hawkfish (*Paracirrhites arcatus*) to investigate the evolution
85 of initial reproductive isolation during a potential scenario of incipient ecological speciation-
86 with-gene-flow. *P. arcatus* is a small (<20 cm) shallow reef predator with two recognized color

morphs (Randall 1963) co-occurring throughout their wide distribution (East Africa to the Central Pacific; Fig. 1). Melanistic (MEL) and pink white-striped (PWS) morphs and occasional intermediates (Fig. 1) are sympatric throughout their range, yet segregate across an ecological cline on shallow coastal reefs (DeMartini & Donaldson 1996; Whitney *et al.* 2018). Color morphs are partitioned into distinct basalt and coral microhabitats, while phenotypic intermediates are mostly restricted to transition habitats (Whitney *et al.* 2018). The strong correlation between phenotype and environment suggests that morphs have fitness advantages in their respective habitats. Melanistic morphs appear more cryptic against exposed basalt rock in shallow surge zones, whereas PWS morphs are more cryptic against *Porites* corals in deeper in areas of higher coral cover (Whitney *et al.* 2018). We speculate that disruptive natural selection on color pattern has led to divergence in habitat preferences.

Adaptation to different microhabitats in association with assortative mating by color could provide the barriers necessary to initiate reproductive isolation. The combination of small home ranges, strong territoriality and harem mating (Donaldson 1990; DeMartini 1996) decreases the probability of mating with individuals in different microhabitats (Whitney 2016; Whitney *et al.* 2018). Therefore, habitat isolation among color morphs limits opportunities for intermorph mating, which is reinforced by behavioral preference for like individuals. Field surveys revealed that females were over five times more likely to pair with like-colored males, and mate choice experiments confirmed this (Whitney 2016). Thus, the relatively low proportion of pairings between phenotypes on reefs provides a mechanism of premating isolation among sympatric color morphs.

The goal of this study is to test for genetic evidence of reproductive isolation among color morphs. If ecological barriers are reducing gene flow between sympatric phenotypes then we would expect to find patterns of isolation-by-adaptation (Nosil *et al.* 2008), including (1) genetic differentiation at neutral loci, (2) genetic differentiation greater between contrasting environments than between similar environments, and (3) genetic differentiation at non-neutral loci relevant to the adaptive trait (i.e., color genes).

Melanistic coloration is widespread in fish and is reported to have several adaptive functions such as greater protection from UV light (Ahmed & Setlow 1993), thermoregulation (Trullas *et al.* 2007), disease resistance (Chakarov *et al.* 2008), cryptic coloration (Marjerus 1998), and reduced vulnerability to stress (Kittilsen *et al.* 2009). The melanocortin receptor 1

gene (*Mclr*) encodes a G-protein coupled transmembrane receptor that regulates melanin synthesis in vertebrates (Barsh 1996). Mutations in this gene have been linked to color polymorphisms in a wide range of vertebrates (reviewed in Hoesktra 2006) including birds (Mundy 2005), mammals (Ritland *et al.* 2001; Eizirik *et al.* 2003; Mundy & Kelly 2003; Nachman *et al.* 2003), reptiles (Rosenblum *et al.* 2004) and fish (Protas *et al.* 2006; Selz *et al.* 2007; Richardson *et al.* 2008; Henning *et al.* 2010). Given its widespread association in color evolution and role in melanin synthesis, we targeted *Mclr* as a promising candidate gene to investigate the underlying genetic basis of color polymorphism in *P. arcatus*.

Here, we combine data from (putative) neutral microsatellite markers and the *Mclr* exon to investigate the roles of natural selection and nonrandom mating in driving genetic divergence between sympatric color morphs. To test the hypothesis that *P. arcatus* color morphs are reproductively isolated, we first estimate neutral levels of divergence between phenotypes using 30 microsatellite loci. Next, we combine microsatellite and habitat data to test for the relationship between genetic and environmental distance as expected under scenarios of isolation-by-adaptation. Finally, we test for genetic signatures of selection by comparing levels of differentiation in the *Mclr* gene and putative neutral markers.

Methods

Field collections

A collection of 264 adult *P. arcatus* individuals from west Hawai'i Island was made during July-August 2010 and September 2012 (Fig. 1c; Table S1). Fish were collected using micropole spears while on SCUBA along specific transects described in Whitney *et al.* (2018). The procedures employed were ethically reviewed and approved by the University of Hawai'i IACUC (protocol 10-816-3). Upon collection, each fish was measured, photographed, and dissected; during which a sample of white muscle tissue and caudal fin were removed and preserved in either 95% ethanol or saturated salt (NaCl) 20% DMSO solution for genetic analysis.

Each individual's color phenotype was scored immediately after collection using two qualitative traits (1) body color: melanistic (uniform brown to black), pink (light pink to red with lighter ventral) or intermediate (light brown to dark red), and (2) size of white-stripe (absent, faint or full). To simplify analyses, the nine possible phenotypic groupings were collapsed into

three discrete color morphs (Fig. 1): melanistic (MEL; dark body, no-stripe), pink-white-stripe (PWS; light pink/red body, full white stripe), and intermediate (INT), which included all fish with intermediate body color, faint-stripes, or other combinations (i.e., dark body with stripe, or light body with no-stripe).

Diversity & Differentiation at Microsatellite Loci

We successfully genotyped 264 individuals at 30 microsatellite loci designed for *P. arcatus* (Whitney & Karl 2012). Genomic DNA (gDNA) was extracted from muscle tissue following the HotSHOT protocol (Meeker *et al.* 2007). Loci were amplified using four PCR multiplexes in 5 µl reactions following the protocol of Whitney & Karl (2012). Fluorescently-labeled PCR products were resolved using an ABI 3130 Genetic Analyzer (Applied Biosystems). Alleles were called manually using GENEMAPPER v4.0 with GS500LIZ size standards (Applied Biosystems), and binned using TANDEM v1.08 (Matschiner & Salzburger 2009).

Multiple quality control procedures were applied to ensure robust microsatellite analyses following the recommendations of Selkoe & Toonen (2006) and are described in Supplemental File 1. Individuals missing more than 15% of data were excluded from analyses. PDGSPIDER v2.0.0.0 (Lischer & Excoffier 2012) was used to convert genotype matrices into various formats for analyses.

Measures of genetic diversity were calculated using GENEPOP v4.2 (Raymond & Rousset 1995; Rousset 2008), GENALEX v6.5 (Peakall & Smouse 2006; 2012) and ARLEQUIN v3.5 (Excoffier & Lischer 2010). We tested for conformance to Hardy-Weinberg equilibrium (HWE) in GENEPOP using the Markov chain method (Guo & Thompson 1992) with 10,000 permutations. All pairs of loci were tested for linkage disequilibrium using GENEPOP. The false discovery approach of (Benjamini & Hochberg 1995) was used to adjust significant *P*-values. We estimated the proportion of null alleles at each locus separately for each morphotype using FRENA (Chapuis & Estoup 2007), according to the expectation maximization algorithm of Dempster *et al.* (1977).

Genetic differences between the two predominant color morphs were analyzed in several ways. First we tested for significant differences in allele frequencies using the genic differentiation option in GENEPOP, which provides an unbiased estimate of the *P*-value with a *G* log-likelihood-based exact test (Goudet *et al.* 1996). We adjusted significance thresholds for

multiple comparisons by controlling the false discovery rate ($FDR < 0.05$) using the q -value method (Storey *et al.* 2004) implemented in the R package qvalue v2.2.2 (Dabney *et al.* 2010).

Second, we evaluated the partitioning of genetic variance both between color morphs and locations using a hierarchical allelic distance-based Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992) using GENODIVE v2.0b27 (Meirmans & van Tienderen 2004) assuming an infinite allele model (10,000 permutations). We grouped individuals into two populations based on location of collection site: Kohala in the north ($N=112$: 57 MEL, 55 PWS) and Kona in the south ($N=99$: 50 MEL, 49 PWS). Collection sites were binned into these geographic regions because they were the most geographically separated groups that contained sufficient sample sizes ($N \sim 50$ per morph per region).

Third, we calculated locus-specific and multilocus estimates of genetic divergence among phenotypes using F -statistics (F_{ST}) in GENALEX as per Weir & Cockerham (1984) and Michalakis & Excoffier (1996), assuming an infinite allele model. Fixation indices (e.g., F_{ST}) may be poor estimators of population divergence when using markers with high allelic diversity because F_{ST} is reduced proportionately to within-population levels of heterozygosity (Jost 2008; Meirmans & Hedrick 2010; Bird *et al.* 2011b). We employed GENALEX, GenoDive, and SMOGD v1.2.5 (Crawford 2010) to calculate the standardized allele-frequency based estimators of population differentiation, F'_{ST} , G'_{ST} , and Jost's D respectively (Hedrick 2005; Jost 2008) following the recommendations of Meirmans & Hedrick (2010). Significance for all estimates of genetic differentiation were tested by comparing observed data with null distributions generated from 10,000 random permutations, and variances were estimated via jackknifing and bootstrapping over loci. Despite the limitations of F_{ST} , we report it along with standardized metrics to allow for a direct comparison with older literature (Weersing & Toonen 2009).

Assignment Tests and Clustering

Multilocus genotypes from 107 MEL, 104 PWS and 53 INT morphotypes (13 no-stripe, INS; 16 faint-stripe, IFS; and 24 white-stripe, IWS) were used to assess population structuring using both STRUCTURE v2.3 (Pritchard *et al.* 2000) and discriminant analysis of principal components (DAPC) in Adegnet v1.4-2 for R 3.1.0 (Jombart 2008; R Core Development Team). For STRUCTURE we used a model assuming admixture and correlated allele frequencies, as we expect current gene flow and many shared alleles among morphotypes due to common ancestry (Falush

et al. 2003). Phenotype was used as prior information to assist the structuring as recommended for weak signals of structuring (Hubisz *et al.* 2009). We ran the model using values of K from 1 to 5, with 5 replicate runs for each K and a burn-in of 500,000 steps, followed by 2,000,000 MCMC iterations. We used the delta K method of Evanno *et al.* (2005) implemented in STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt 2012) to determine the most likely number of clusters. Membership scores from five replicate runs were aligned and averaged using CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007), and visualized with DISTRUCT v1.1 (Rosenberg 2004).

DAPC yields information similar to STRUCTURE, but does not rely on a particular population genetics model, and is thus free of assumptions about Hardy-Weinberg and linkage equilibrium (Jombart *et al.* 2010). We imposed the number of clusters (K) of two and ran analyses first without any priors and then by using phenotype as a prior group assignment. The optimal number of principal components (PCs) retained for the DAPC was chosen using an optimization procedure (Jombart 2008). The DAPC was performed using 30 PCs (representing ~50% of the total genetic variation) and one discriminant function, which is the maximum when $K = 2$. For both STRUCTURE and DAPC results, membership probabilities in each of the two clusters were arcsine square root transformed, and a t -test was used to compare values between the two putative clusters for each color morph.

Although intuitively one might expect that more loci should be more informative about population structure, more loci can actually cause reduced performance in individual assignments in some cases (Guinand *et al.* 2004). Assignment accuracy is dependent on the number and variability of markers used, the levels of co-ancestry and degree of genetic differentiation among groups (Bjørnstad & Røed 2002; Guinand *et al.* 2006). Smouse & Chevillon (1998) recommend using a moderate number of informative loci when estimating mixture composition. In some cases, the selection of informative microsatellite markers is required for improving individual assignment efficiency (Li *et al.* 2014). Therefore, our goal was to identify a subset of microsatellite loci with high discriminatory power for informing assignments to phenotypic clusters.

We used backward elimination locus selection (BELS; Bromaghin 2008) to evaluate the power of each microsatellite locus to improve individual assignment estimation, by excluding loci (one by one) that cause the least reduction in performance. A test set (baseline) of 100 randomly chosen individuals classified into two groups based on color phenotype (50 MEL, 50

PWS) was resampled 1000 times to produce 200 randomly generated genotypes from each group. The average individual assignment accuracy across both groups (0.80) was used as the performance measure to evaluate locus utility. The set of loci required to meet the performance measure of 80% individual assignment accuracy was used in STRUCTURE (using parameters described above) to test if individuals could be assigned to phenotype groupings and assess the degree of admixture of individuals of intermediate phenotypes. As a comparison, we executed the same BELS procedure using groupings defined by geographic location (north versus south) to determine whether individuals could be assigned to geographic clusters.

Distance-based Redundancy Analysis

To test whether habitat-related variables explained some of the genetic variation observed among *P. arcatus* color morphs, we partitioned a genetic distance matrix using distance-based redundancy analysis (dbRDA; Legendre & Anderson 1999; McArdle & Anderson 2001). This procedure examines the extent to which each predictor variable explains significant genetic variability among individuals by regressing a response matrix of genetic distances on various explanatory matrices including geographic distance, phenotype and habitat. This analysis was restricted to 100 individuals (50 MEL and 50 PWS) that were collected on transects with high resolution habitat survey data (see Whitney *et al.* 2018). Genetic distances among individuals were calculated using Euclidean distance in GENODIVE based on multilocus microsatellite genotypes. Geographic distances between individuals were calculated as linear Euclidean distances from the approximate latitude and longitude of collection site. Phenotype was categorical with two levels (MEL and PWS), and habitat type had three levels corresponding to reef zones: (1) surge zone dominated by *Pocillopora* corals, (2) sub-surge zone co-dominated by *Pocillopora* and *Porites* corals, and (3) reef slope dominated by *Porites* corals. These habitat types encapsulate the multivariate environmental gradient along which hawkfish color morphs are partitioning (Whitney *et al.* 2018). The relationship between the response matrix (genetic distance) and each predictor variable was analyzed separately using dbRDA with Vegan (Oksanen *et al.* 2015) in R. Statistical significance was determined using 10000 permutations.

*Diversity and Differentiation of *Mc1r**

An 850-bp protein-coding fragment of the *Mc1r* gene was amplified using the conserved primers

MC1RFor and MC1RRev of Eytan *et al.* (2012). PCRs were performed in 15 μ l reactions of 7.5 μ l BioMixRed (Bioline), 0.3 μ l of each primer (10 μ M), 5.9 μ l deionized water, and 1 μ l of gDNA with the following cycles: 1 cycle of 94 °C for 45s, 50 °C for 45s, 72 °C for 90s followed by 35 cycles of 95 °C for 1min, 63 °C for 1min, and 72 °C for 90s, and a final cycle of 94°C for 40s, 50°C for 45s, and 72°C for 10min. We purified PCR amplicons with FastAP (ThermoFisher) and sequenced fragments in both directions using ABI 3730xl Sequencers (Applied Biosystems). Individuals with poor sequence quality were re-sequenced. Sequences were edited and aligned using GENEIOUS v7.0.3 (Biomatters). Nucleotide and amino acid polymorphisms were viewed in MEGA v6 (Tamura *et al.* 2013). Double peaks of approximately equal height present in the same sequence from both directions were scored as heterozygotes. Haplotype phases were inferred using PHASE v2.1.1 (Stephens *et al.* 2001).

We estimated diversity of the *Mclr* gene as haplotype number and diversity (h ; Nei 1987) overall and for each phenotype separately. Nucleotide diversity was estimated with Watterson's Θ (Watterson 1975) and Nei's π (Tajima 1983, Nei 1987). We calculated Tajima's D (Tajima 1989) to test for departures from neutral expectations. Diversity metrics were all calculated using the software DNASP v5 (Librado & Rozas 2009). To assess the proportion of genetic variation explained by color phenotype, we ran AMOVA tests in GENALEX and GENODIVE using both allele frequencies (F_{ST}) and allele sequence divergence (Φ_{ST}). Statistical significance was estimated using randomization tests with 10,000 random permutations. To estimate the degree of genetic differentiation between color morphs at this locus, we also calculated Jost's D in GENODIVE.

F_{ST} Outlier Test

Microsatellite loci are generally noncoding and assumed to act as a neutral marker in population genetics. Divergent selection at a locus physically linked to a neutral microsatellite, however, can inflate estimates of population genetic differentiation (Nielsen *et al.* 2006; Feder *et al.* 2012). To assess if any microsatellite loci or the *Mclr* locus were more divergent than expected under neutral conditions, we used the FDIST2 outlier analysis in Beaumont & Nichols (1996) and implemented in LOSITAN v1.0 (Antao *et al.* 2008). The FDIST method uses simulations to generate a null distribution of neutral F_{ST} , based on the assumption that all F_{ST} values are equal among loci, and then uses the relationship between F_{ST} and the expected heterozygosity to

identify loci that have excessively high or low F_{ST} compared to neutral expectations. We ran 50,000 simulations using the neutral F_{ST} and force mean F_{ST} options (both recommended), which increase the reliability of the mean F_{ST} by running an initial simulation that removes potential outliers. We ran simulations under the infinite allele (IAM) model with a false discovery rate of 0.05 and 99.9% confidence intervals. For the *Mc1r* gene, we used the single nucleotide polymorphism (SNP) with the highest F_{ST} (*Mc1r_023*) to represent the gene.

Results

Diversity & Differentiation at Microsatellite Loci

Thirty microsatellite loci were genotyped in 264 *Paracirrhites arcatus* of three color morphs (107 MEL, 104 PWS, 53 INT). All loci were polymorphic with the number of alleles ranging from 6 to 34 (mean = 16.3). Microsatellite quality control procedures and summaries of locus-by-locus measures of genetic diversity are described in Supplementary File 1. Analyses of the two predominant color morphs revealed high levels of genetic variability within subpopulations: mean number of alleles per locus ($\bar{A}_R$) were 14.7 (MEL) and 15.0 (PWS) and heterozygosity (H_E) was 0.83 for both MEL and PWS (Table S2). Estimates of unbiased heterozygosity were consistently high ($H_S = 0.83$) across morphotypes and ranged from 0.50 – 0.95 across loci (Tables S3-S4). Levels of genetic diversity were not significantly different between morphotypes, either in terms of allelic richness ($P = 0.85$) or heterozygosity ($P = 0.59$; Table S4). Each morphotype had private alleles, however none had frequency $> 0.05\%$ (21 of 422 alleles; Table S4). Five of 440 pairwise tests for linkage disequilibrium (1.1%) were significant after applying Bonferroni adjustments (Table S5), however, none of these pairs were detected as being linked when each morphotype was tested separately. Therefore the lack of a repeatable pattern across subpopulations indicates that linkage disequilibria are not due to physical linkage of loci. All subsequent results are reported using the full set of 30 microsatellite loci.

Microsatellite variation shows low levels of genetic differentiation among morphotypes (global mean $F'_{ST} = 0.014$, $F_{ST} = 0.003$). Multilocus exact G log-likelihood tests revealed significant differences in allelic ($P = 0.0003$), and genotypic distributions ($P = 0.034$) between color morphs. F'_{ST} estimates were highly heterogeneous across the genome, ranging from -0.035 to 0.072 among loci. Locus-by-locus comparisons revealed significant allelic differentiation in six loci ($P < 0.05$; Table S6). If we assume that all the null hypotheses are true ($F'_{ST} = 0$), given

a criterion of $\alpha = 0.05$, we would expect that up to 5% of loci could show significant differences by chance alone. Using exact G tests, 20% of loci (6/30) revealed significant differentiation at $P < 0.05$, and four remained significant after accounting for possible false positives. Together these four loci (Paa55, Paa60, Paa72, Paa88) had a mean $F'_{ST} = 0.044$, whereas the remaining 26 loci had a mean $F'_{ST} = 0.004$.

According to the hierarchical AMOVA, nearly all genetic diversity (99.9%) was partitioned among individuals within morphotypes, indicating high genetic similarity overall (Table 1). Global F'_{ST} estimates revealed very slight, but significant overall genetic structure among morphotypes ($F'_{ST} = 0.006$; $P = 0.025$), yet no detectable structure between regions (north versus south) either when nested in morphotype ($F'_{ST} = -0.012$, $P = 0.97$) or when pooling both morphotypes ($F'_{ST} = -0.003$, $P = 0.82$). Comparisons of sampling year also revealed no structure between 2010 and 2012 ($F'_{ST} = -0.001$, $P = 0.66$).

Pairwise multilocus comparisons of allelic distribution revealed higher divergences among color morphs, even within a geographic cluster, than between geographic clusters (Table 2). There were no significant differences between north and south regions when pooling all individuals ($P = 0.13$) or when each morph was tested separately (MEL: $P = 0.35$; PWS: $P = 0.18$; Table 2). Jost's D also indicated more differentiation among color morphs ($D_{EST} = 0.002$, $P = 0.025$) than among locations ($D_{EST} = -0.001$, $P = 0.13$), however this difference is subtle.

Assignment Tests and Clustering

Results from the STRUCTURE analysis indicate that color morphs form discrete clusters when considered in the context of differentiation at a subset of microsatellite loci, whereas they do not group when all loci are included (Fig. 2). Simulations using all 30 microsatellite loci, indicated that $K = 2$ had the highest likelihood ($\ln L = -21,747$). However, the probability of assignment to either of the two clusters shows no distinguishable pattern (Fig. 2A) and there is no significant difference in membership among phenotypes regardless of whether we use the mean posterior scores of five simulation runs ($F_{1,205} = 1.86$, $P = 0.17$) or the model with the highest likelihood ($F_{1,205} = 0.89$, $P = 0.35$). Therefore, we conclude that when using all 30 loci STRUCTURE does not detect distinct genetic clusters.

Using the panel of the seven loci selected by BELS (Table S7), STRUCTURE was able to correctly assign individuals to their phenotypic cluster. Mean posterior membership scores

differed significantly among MEL and PWS color morphs ($F_{1,205} = 865.6$, $P < 0.001$; Fig. 2B). Intermediate morphotypes clustered together, and cluster membership did not vary significantly among the three intermediate subtypes ($F_{2,42} = 2.85$, $P = 0.92$; Fig. 2B). Though we only expected differentiation between MEL and PWS, the intermediates showed mixed ancestry between the two clusters with mean membership in cluster 1 (0.60 of individuals; associated with MEL morphs) significantly higher ($t = 9.26$, $P < 0.001$) than in cluster 2 (0.40 of individuals; associated with PWS morphs). For comparison, using the panel of seven loci selected by BELS based on the power to discriminate locations (Table S7), mean posterior membership scores did not vary significantly among fish collected in northern versus southern sites ($F_{1,205} = 2.85$, $P = 0.092$; Fig. 2D).

In the DAPC, most likely number of clusters was $K = 2$ (based on the lowest BIC score). When using phenotype as a prior, the mean posterior probabilities of assignment of PWS and MEL individuals to their own cluster was significantly higher (0.72 ± 0.22) than the alternate cluster ($F_{1,205} = 57.75$, $P < 0.001$; $t = 10.5$, $P < 0.001$). However, when no priors were used clusters did not correspond well to color morphs, and there was no significant difference in membership probability in the two clusters ($t = 0.586$, $P = 0.56$).

Distance-based Redundancy Analysis

We detected significant isolation by habitat in the distance-based redundancy analysis. Habitat type defined as reef zone (i.e., surge, sub-surge, or reef slope) based on coral community and topography had a significant relationship with genetic distance ($F_{2,93} = 1.34$, $P = 0.025$), explaining 5.9% of the genetic variability among individuals. Concordant with the results of pairwise allelic differentiation, geographical distance did not significantly influence the patterns of genetic distance ($P = 0.33$). Color morph explained only 1.5% of genetic variability, but was highly significant ($P = 0.007$), and had significant interaction terms with both habitat type (% var = 3.2, $P = 0.001$) and geographic distance (% var = 1.6, $P = 0.019$; Table 3). When the spatial variation was taken into account by using distance as a conditional variable in a partial dbRDA, a significant relationship between genetic variation and habitat was still detected ($P = 0.001$). No other environmental variables (depth, slope, wave action, and percent coral cover estimates) showed a significant relationship with genetic distance ($P > 0.05$).

398 We sequenced 742 basepairs of the coding region of the nuclear gene *Mc1r* in 85 individuals (42
399 MEL, 43 PWS; Genbank accession numbers: MG877306-MG877390). Six of the 742
400 nucleotides were biallelic. Haplotype diversity was high ($h = 0.816 \pm 0.035$) with 15 haplotypes
401 estimated using PHASE. Nucleotide diversity also was high ($\pi = 0.0026$, $\Theta = 0.008$), and
402 maximum sequence divergence between any two haplotypes was $d = 1.26\%$. All segregating
403 sites were synonymous (5 transitions, 1 transversion) and therefore do not affect protein coding
404 (Tajima's $D = 0.58$, $P > 0.10$). Nevertheless, *Mc1r* shows significant differentiation between
405 morphotypes ($F_{ST} = 0.032$, $P = 0.022$; $D_{EST} = 0.012$). An analysis of molecular variance
406 indicated that 5.1% of the variation at *Mc1r* was partitioned between color phenotypes and that
407 this was significantly greater than with random permutations ($\Phi_{ST} = 0.051$, $P = 0.031$; Table 4).
408 Individual SNP estimates revealed significant differences in allele frequencies between color
409 morphs at a single variable site (*Mc1r*-023: $F'_{ST} = 0.216$, $F_{ST} = 0.175$; $P < 0.001$), whereas the
410 remaining five SNPs were uninformative (mean $F'_{ST} = -0.004$; $P > 0.40$). Substitutions at *Mc1r*-
411 23 are in the third codon position for glutamic acid (GAA/GAG) and do not change the amino
412 acid coding. Therefore, *Mc1r* does show a correlation with color phenotype, but the relationship
413 is not directly mechanistic. There could, however, be tRNA quantity limitations (codon bias) that
414 make one codon more favorable than the other (Quax *et al.* 2015).

415

416 *F_{ST} Outlier Test*

417 The FDIST analysis conducted with LOSITAN identified the *Mc1r* gene as the only significant
418 outlier ($P < 0.001$) and thus a candidate for positive selection. The most strongly segregating
419 SNP in *Mc1r* had a conspicuously large F_{ST} (0.175) compared to the neutral microsatellite loci
420 (0.003) and exceeded the 99.9% confidence interval (Fig. 3). Although there was variation in
421 uncorrected F_{ST} among microsatellite loci (-0.004 to 0.007), all were well within neutral
422 expectations based on simulated null distributions.

423

424 **Discussion**

425 Variation at 30 putatively neutral microsatellite loci and a melanin pigment gene (*Mc1r*) was
426 used to evaluate genetic divergence among sympatric color morphs of the arc-eye hawkfish.
427 First, we report weak but significant genetic differentiation of microsatellites that provides

evidence of partial reproductive isolation among ecologically-differentiated color morphs. Second, the finding that color is genetically correlated with ecological divergence provides support for a pattern of isolation-by-adaptation. Third, divergence in the Mc1r gene offers evidence of selection relevant to this presumably adaptive color polymorphism. Together, the genetic pattern of divergence in both neutral and non-neutral loci suggests that ecological specialization is creating a generalized barrier to gene flow among adaptive ecotypes. We hypothesize that adaptation to the contrasting microhabitats overrides gene flow and is responsible for the partial reproductive isolation observed between sympatric ecomorphs. We discuss these results in the context of adaptation and incipient ecological speciation.

Interpreting Low Levels of Genetic Differentiation

Estimates of genetic differentiation between color morphs were highly heterogeneous across loci, F'_{ST} ranges from -0.035 to 0.072 (mean $F'_{ST} = 0.014$, $F_{ST} = 0.003$). Whereas the majority of microsatellites showed little to no structure between color morphs, 20% showed significant differences in allele frequencies. The comparison of STRUCTURE results when using all loci versus a subset of the most diverse loci (selected by BELS) illustrates the variability in estimates of gene flow across the genome (Fig. 2). Heterogeneity in gene flow estimates across loci is expected because differentiation at neutral loci is a product of drift, and therefore should vary stochastically (Schluter & Conte 2009). Among-locus variation could also be influenced by differences in mutation rates, however, mutation rate is not predicted to strongly affect patterns of genetic differentiation (Beaumont & Nichols 1996; Balloux & Moulin 2002; Hedrick 2005). Quantifying the current magnitude of gene flow is difficult at such low levels of differentiation (Faubet *et al.* 2007), as reliable estimates of Nm (number of migrants per generation) based on F_{ST} requires moderate to large values ($F_{ST} > 0.05$ –0.10; Allendorf *et al.* 2013).

The relatively low genetic differentiation among color morphs could be a consequence of recent divergence and/or ongoing, but restricted gene flow. It is certainly feasible that reproductive isolation has occurred, but too recently to allow genome-wide differentiation via drift. This neutral process may require hundreds or thousands of generations, even if selection is acting directly on parts of the genome. The alternative explanation, ongoing gene flow (incomplete reproductive isolation), is also likely given that mixed morph pairs and intermediate phenotypes are observed in zones of overlap, albeit at low frequency. Currently the lack of

related phylogenetic data precludes any inference of age of the species or divergence from closest relatives that would help to assess the age of the polymorphism. Knowledge of the phylogeographic history of this species, color morphs, and its closest relatives will be useful for estimating the age of this presumably incipient divergence.

Detecting any level of significant differentiation on such a fine-spatial scale is surprising. Most marine fishes only show significant population genetic differences over much greater geographical distances (Waples 1998; Waples & Gaggiotti 2006; Eble *et al.* 2015). The anomalous finding here, significant population structure among phenotypes in the absence of geographic isolation, indicates that genetic differentiation is likely driven by an ecological barrier (Nosil *et al.* 2008). The relatively weak signal of genetic divergence at neutral loci that emerges from our data implies that hawkfish color morphs are partially reproductively isolated. An alternative hypothesis could be that random divergence among sample groups has led to the expected signature. When genetic differentiation is low, confounding factors such as nonrandom sampling (Allendorf & Phelps 1981), temporal fluctuations (Turner *et al.* 2002) and genotyping errors (Bonin *et al.* 2004) can lead to false conclusions (Waples 1998). Here we address these concerns and provide multiple lines of evidence contradicting a false conclusion. First, the absence of genetic differentiation between samples collected in replicate years (2010, 2012) and locations (northern and southern sites), demonstrates that we did not detect significant genetic differences in comparisons where they were not expected. Second, by using temporal and geographic replicate sampling, we show that differences among phenotypes are consistent over time and location, and should therefore indicate that the genetic signal is real.

Adaptive Divergence

Evolutionary genetic divergence among ecotypes is difficult to detect under a scenario of ongoing gene flow. When gene flow among ecotypes is high, divergent selection may be overwhelmed by gene flow, to the point where divergence at neutral genetic loci (drift) cannot proceed (Thibert-Plante & Hendry 2010). When gene flow is low or absent, a spurious signal of divergent selection is more likely, because differentiation between populations could be due to neutral drift. Using simulations, Thibert-Plante & Hendry (2010) demonstrated that when gene flow is ongoing and ecological differences are strong, finding signatures of isolation usually indicates true adaptive divergence. The persistence of population structure under gene flow

indicates that environmental differences (strong divergent selection) or sexual isolation are inducing a generalized barrier to gene flow.

The observed pattern of genetic differentiation among hawkfish habitats is in agreement with predictions of isolation-by-adaptation (Nosil *et al.* 2009; Mendez *et al.* 2010; Nosil 2012). Results from the distance-based redundancy analysis (dbRDA) indicated that habitat and color phenotype, but not geographic distance, were significantly correlated with genetic distance among individuals and together explained ~11% of genetic variation in microsatellite allele frequencies. Hence, we detected greater genetic differences between adjacent microhabitats and phenotypes than between the same habitats in different geographic locations. This pattern of reduced gene flow between ecologically diverged phenotypes provides compelling evidence for ecological barriers restricting gene flow on a very fine spatial scale.

Nonrandom mating could explain the overall genetic differentiation at unlinked neutral loci and corresponds well to patterns of ecological and behavioral divergence observed among color morphs. First, the strong phenotype-environment correlation demonstrated in Whitney *et al.* (2018) indicates that phenotypic variation has evolved in response to divergent selection among habitats, as ecological divergence is often used as a proxy for the presence of selection (Nosil *et al.* 2009). Second, hawkfish mating occurs in their preferred microhabitat, so habitat preferences are capable of curtailing random mating, irrespective of mate choice. Field surveys demonstrate that fish pair with like morphs 6x more frequently than with dissimilar morphs (Whitney 2016). The strong relationship between genetic differentiation and habitat type indicates that divergent habitat preferences are acting as an ecological barrier constraining gene flow. Third, color morphs show a tendency to mate assortatively with respect to color both in the field and in the laboratory (Whitney 2016). Therefore, even when divergent phenotypes co-occur in the same habitat, divergence is reinforced by premating isolation. Overall, habitat choice and assortative mating are synergistically forming a potent premating barrier.

We hypothesize that divergent natural selection acting along an environmental gradient (Whitney *et al.* 2018) is overriding the homogenizing effect of gene flow – with a subsequent shift in mating preferences (Whitney 2016), in turn yielding a correlation of genetic and ecological differentiation. Disentangling the proportion of reproductive isolation that is due to premating versus postmating isolation (selection against migrants or hybrids) is beyond the scope of this study. Further experimental studies that directly measure fitness differences in each

habitat would be necessary to evaluate the relative strength of isolating mechanisms.

Our findings on the genetic and phenotypic divergence in *P. arcatus* are consistent with patterns of local adaptation as well as early phases of speciation-with-gene-flow. A correlation between neutral genetic differentiation and ecological divergence is generally recognized as a precursor of incipient ecological speciation (Nosil 2012; Shafer & Wolf 2013). Whether complete reproductive isolation will develop between hawkfish color morphs can only be speculated. The likelihood of speciation versus persistence of a stable color polymorphism is influenced by the balance between selection and gene flow. On one hand, ecologically-induced population differentiation may never result in complete reproductive isolation or the formation of new species (Nosil 2012). The polymorphism could be maintained in the system by local variation in microhabitat and local adaptation to divergent habitats. Indeed, ecological speciation is a continuum and is often stalled at intermediate points at some selection-gene-flow balance (Räsänen & Hendry 2008; Schafer & Wolf 2013). Conversely, speciation in the presence of gene flow can occur rapidly when ecological traits under selection are genetically correlated and affect mate choice (Gavrilets 2004; Hendry *et al.* 2007; Langerhans *et al.* 2007; Jiggins 2008). When color morphs mate assortatively, speciation is greatly facilitated because any ecologically-derived reproductive barrier will be reinforced by the isolating effects of assortative mating (Dieckmann and Doebeli 1999; Gavrilets 2004; Nosil 2012). Ultimately, the strengths of selection and assortative mating will determine whether color morphs speciate or remain in a stable polymorphic state (Matessi *et al.* 2001; Bolnick 2006).

Regardless of its evolutionary fate, this system provides an empirical example of ecological specialization overriding gene flow in the absence of physical barriers. By identifying an ecological driver of genetic differentiation we show that adaptive divergence to contrasting microhabitats can initiate evolutionary divergence even in the face of high gene flow.

The Coloration Connection

Our results reveal low F_{ST} across the genome in neutral microsatellite loci, but high differentiation at the pigmentation pathway gene *Mc1r*, a non-neutral locus relevant to eco-phenotypic differentiation. Using simulation-based methods for detecting selection, we confirmed that *Mc1r* is likely under the influence of positive selection (Fig. 3), either directly or via physical linkage. As no polymorphism had a detectable change in the function of this

melanocyte receptor, color differences between hawkfish morphs may not be attributable to variation in this locus. Substitutions elsewhere in the gene or epistasis with another gene may explain results. Noncoding flanking regions of the *Mc1r* gene regulate gene expression. For example, Stahl & Gross (2015) recently reported that cave dwelling fishes (genus *Astyanax*) with no coding sequence variation in *Mc1r* still demonstrate reduced expression due to mutations in the 5' regulatory region. Similarly, mutations in regulatory elements of two other pigment genes, yellow in *Drosophila* (Jeong *et al.* 2006) and agouti in *Peromyscus* deer mice (Kingsley *et al.* 2009), can affect pigmentation via altered protein expression. Therefore, the color polymorphism in hawkfish could be controlled by variation in *Mc1r* regulatory regions and/or other gene sequences involved with melanin synthesis.

The genes coding for melanocortins and the hormones that stimulate *Mcr* receptors are also involved with melanistic pigmentation in teleost fishes (Fujii 2000; Richardson *et al.* 2008). For example, in the polychromatic neotropical midas cichlid complex (*Amphilophus citrinellus*), *Mc1r* was found to be upregulated in the skin of amelanic (gold) fish, despite no sequence polymorphism between morphs in coding or noncoding regions (Henning *et al.* 2010). Whole genome scans combined with targeted sequencing of color candidate genes will no doubt help to elucidate the genetic architecture of color polymorphism in *P. arcatus*.

Conclusion

The major ingredients of incipient ecological speciation have been identified in *P. arcatus*, specifically a source of natural divergent selection (crypsis), a mechanism of reproductive isolation (habitat preference and color-based assortative mating) and a link between the two. Here we provide genetic data that these mechanisms have resulted in partial reproductive isolation between color morphs in adjacent habitats. When physical barriers are absent, partial reproductive isolation may mean that divergent selection is acting on phenotypes to reduce gene flow. The combination of ecological (Whitney *et al.* 2018), behavioral (Whitney 2016), and now genetic studies of the arc-eye hawkfish provide compelling evidence for partial reproductive isolation resulting from ecological barriers in the absence of geographic isolation.

Disentangling the proportion of reproductive isolation that is due to premating versus postmating isolation (selection against migrants or hybrids) is beyond the scope of this study. Further experimental studies that directly measure fitness differences in each habitat would be

necessary to evaluate the relative strength of isolating mechanisms. Nevertheless, the genetic pattern that emerges from our data is consistent with the signatures of natural selection. The peak in divergence in the *Mclr* gene, amidst a background signal of low but heterogeneous differentiation across the genome, matches the premises of isolation-by-adaptation (Nosil et al. 2008).

Speciation is a continuum of differentiation, and based on these data, hawkfish color morphs reveal key traits associated with the early stages of evolutionary divergence. When mate choice is linked to an ecologically important trait under selection, rapid divergence may occur in sympatry (Rice 1984; Rice & Salt 1990; Schluter 2000), and this may be the most probable scenario for speciation-with-gene-flow (Maynard-Smith 1966; Gavrillets 2004). Because color influences mate preference, the genetic correlation of color polymorphism to ecological divergence indicates an innate potential for evolutionary divergence. It is also possible that the ecologically-induced divergence among arc-eye color morphs represents a stable equilibrium between the tides of selection and gene flow (Butlin *et al.* 2008; Rougeux *et al.* 2017). Arc-eyes occur on shallow reefs from Africa to Hawaii, with light and dark color morphs co-occurring through much of this range. Why have none of these far-flung populations advanced through the speciation process? Perhaps we are observing ‘flickers of speciation’, sparks of evolutionary differentiation that require yet another ingredient to ignite the process of complete isolation. While terrestrial species can readily radiate into adjacent micro-niches (e.g. Shaw & Gillespie 2016) this process is much rarer in the oceans (Pinheiro *et al.* 2017). Rocha & Bowen (2008) suggest that a parapatric distribution, in addition to ecological differentiation, may be the most fruitful arrangement for speciation in the sea. Intensified selection, stronger assortative mating, or ecosystem disturbances may push the speciation process to completion. Whether complete reproductive isolation will develop between arc-eye color morphs remains speculation. Regardless of the outcome, *P. arcatus* provides a rare case confirming that partial reproductive isolation can evolve in the face of continuous gene flow, bringing us one step closer to understanding the role ecological barriers play in initiating the early stages of speciation.

Acknowledgements

We thank A Estes, J & D Kleinhenz, A Nordschow, A Purves and L Wheeler for assistance conducting fieldwork; C Bird, J Delia, E DeMartini, M Gaither, JP Hobbs, M Iacchei, R Toonen

and B Wainwright for thoughtful discussions that improved the study; M Iacchei and J Zamzow for reviewing a draft manuscript, and the UH Dive Safety Program for diving support. This work was funded by research grants from NSF EPSCoR Hawai'i Program, ARCS Foundation, PADI Foundation, Jessie D Kay Memorial Fellowship, Society for Integrative & Comparative Biology, American Society for Ichthyologists & Herpetologists, and Sigma Xi awarded to JLW. This is contribution #10315 from the School of Ocean and Earth Science and Technology and #1718 from the Hawai'i Institute of Marine Biology.

References

- Ahmed FE, Setlow RB (1993) Ultraviolet radiation-induced DNA damage and its photorepair in the skin of the platyfish *Xiphophorus*. *Cancer Research*, **53**, 2249-2255.
- Alfaro ME, Santini F, Brock CD (2007) Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). *Evolution*, **61**, 2104–2126.
- Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1507–1514.
- Allendorf FW, Luikart G, Aitken SN (2013) *Conservation and the genetics of populations*. Blackwell Publishing, Oxford.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.
- Balloux F, Lugon Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.
- Barsh GS (1996) The genetics of pigmentation: from fancy genes to complex traits. *Trends in Genetics*, **12**, 299–305.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B Methodological*, **57**, 289–300.
- Bernardi G (2013) Speciation in fishes. *Molecular Ecology*, **22**, 5487–5502.

- Bird CE, Holland BS, Bowen BW, Toonen RJ (2011a) Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Molecular Ecology*, **20**, 2128–2141.
- Bird CE, Karl SA, Smouse PE, Toonen RJ (2011b) Detecting and measuring genetic differentiation. *Phylogeography and Population Genetics in Crustacea* **19**, 31–55.
- Bjørnstad G, Røed KH (2002) Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. *Animal genetics*, **33**, 264–270.
- Bolnick DI (2006) Multi-species outcomes in a common model of sympatric speciation. *Journal of Theoretical Biology*, **241**, 734–744.
- Bongaerts P, Riginos C, Hay KB (2011) Adaptive divergence in a scleractinian coral: physiological adaptation of *Seriatopora hystrix* to shallow and deep reef habitats. *BMC Evolutionary Biology*, **11**, 303–317.
- Bonin A, Bellemain E, Bronken-Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.
- Bowen BW, Rocha LA, Toonen RJ, Karl SA, the ToBo Laboratory (2013) The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, **28**, 1–8.
- Briggs JC (2005) The marine East Indies: diversity and speciation. *Journal of Biogeography*, **32**, 1517–1522.
- Bromaghin JF (2008) BELS:backward elimination locus selection for studies of mixture composition or individual assignment. *Molecular Ecology Resources*, **8**, 568–571.
- Butlin, RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B* **363**, 2997–3007.
- Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.
- Chakarov N, Boerner M, Kruger O (2008) Fitness in common buzzards at the cross-point of opposite melanin-parasite interactions. *Functional Ecology* **22**, 1062–1069.
- Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

- Dabney A, Storey JD, Warnes GR (2010) qvalue: Q-value estimation for false discovery rate control. *R package version 2.2.2*, <http://github.com/jdstorey/qvalue>.
- DeMartini EE, Donaldson TJ (1996) Color morph-habitat relations in the arc-eye hawkfish *Paracirrhites arcatus* (Pisces: Cirrhitidae). *Copeia*, **1996**(2), 362-371.
- DeMartini EE (1996) Sheltering and foraging substrate uses of the arc-eye hawkfish (*Paracirrhites arcatus*). *Bulletin of Marine Science* **58**, 826–837.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society Series B Methodological*, **39**, 1–38.
- Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* **400**:354–357.
- Donaldson TJ (1990) Reproductive behavior and social organization of some Pacific hawkfishes (Cirrhitidae). *Japanese Journal of Ichthyology* **36**, 439–458.
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Eble JA, Bowen BW, Bernardi G (2015) Phylogeography of coral reef fishes. Pp. 64 – 75 In C. Mora (ed.) *Ecology of Fishes on Coral Reefs*. University of Hawaii Press, Honolulu.
- Eizirik E, Yuhki N, Johnson WE *et al.* (2003) Molecular genetics and evolution of melanism in the cat family. *Current biology*, **13**, 448–453.
- Elmer KR, Lehtonen TK, Meyer A (2009) Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution* **63**, 2750–2757.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Eytan RI, Hastings PA, Holland BR, Hellberg ME (2012) Reconciling molecules and morphology: Molecular systematics and biogeography of Neotropical blennies

(*Acanthemblemaria*). *Molecular Phylogenetics and Evolution*, **62**, 159–173.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Fujii R (2000) The regulation of motile activity in fish chromatophores. *Pigment Cell Research*, **13**, 300–319.

Gavrilets S (2004) *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, New Jersey.

Gavrilets S (2014) Models of speciation: where are we now? *The Journal of heredity*, **105**(Suppl 1), 743–755.

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Guinand B, Scribner KT, Page KS *et al.* (2006) Effects of coancestry on accuracy of individual assignments to population of origin: examples using Great Lakes lake trout (*Salvelinus namaycush*). *Genetica*, **127**, 329–340.

Guinand B, Scribner KT, Topchy A *et al.* (2004) Sampling issues affecting accuracy of likelihood-based classification using genetical data. *Environmental Biology of Fishes*, **69**, 245–259.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hendry AP, Nosil P, Rieseberg LH (2007) The speed of ecological speciation. *Functional Ecology*, **21**, 455–464.

Henning F, Renz AJ, Fukamachi S, Meyer A (2010) Genetic, comparative genomic, and expression analyses of the *Mc1r* locus in the polychromatic Midas cichlid fish (Teleostei, Cichlidae *Amphilophus* sp.) species group. *Journal of Molecular Evolution* **70**, 405–412.

Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, **97**, 222–234.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
- Jeong S, Rokas A, Carroll SB (2006) Regulation of body pigmentation by the abdominal-b hox protein and its gain and loss in *Drosophila* evolution. *Cell* **125**, 1387–1399.
- Jiggins CD (2008) Ecological Speciation in Mimetic Butterflies. *Bioscience*, **58**, 541–548.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94–108.
- Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.
- Kingsley EP, Manceau M, Wiley CD, Hoekstra HE (2009) Melanism in *Peromyscus* is caused by independent mutations in agouti. *PloS one*, **4**, e6435.
- Langerhans RB, Gifford ME, Joseph EO (2007) Ecological speciation in *Gambusia* fishes. *Evolution*, **61**(9), 2056–2074.
- Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.
- Li YH, Chu HP, Jiang YN *et al.* (2014) Empirical selection of informative microsatellite markers within co-ancestry pig populations is required for improving the individual assignment efficiency. *Asian-Australasian Journal of Animal Sciences*, **27**, 616–627.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.
- Marie Curie Speciation Network, Butlin R, Debelle A *et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.
- Marjerus MEN (1998) *Melanism: evolution in action*. Oxford University Press, Oxford.
- Matessi C, Gimelfarb A, Gavrilets S (2001) Long-term buildup of reproductive isolation promoted by disruptive selection: how far does it go? *Selection*, **2**, 41–64.
- Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics

- and genomics workflows. *Bioinformatics*, **25**, 1982–1983.
- Maynard-Smith J (1966) Sympatric speciation. *American Naturalist*, **100**, 637–650.
- McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, **82**, 290–297.
- Meeker N, Hutchinson S, Ho L, Trede N (2007) Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques*, **43**, 610–614.
- Meirmans PG, Hedrick PW (2010) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.
- Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C, Bordino P (2010) Isolation by environmental distance in mobile marine species: molecular ecology of *franciscana* dolphins at their southern range. *Molecular Ecology*, **19**, 2212–2228.
- Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 1061–1064.
- Munday PL, van Herwerden L, Dudgeon CL (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1498–1504.
- Mundy NI (2005) A window on the genetics of evolution: Mc1r and plumage colouration in birds. *Proceedings of the Royal Society: Biological Sciences*, **272**, 1633–1640.
- Mundy NI, Kelly J (2003) Evolution of a pigmentation gene, the melanocortin-1 receptor, in primates. *American Journal of Physical Anthropology*, **121**, 67–80.
- Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *PNAS*, **100**, 5268–5273.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York.
- Nielsen EE, Hansen MM, Meldrup D (2006) Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. *Molecular Ecology*, **15**, 3219–3229.
- Nosil P (2012) *Ecological Speciation*. Oxford University Press, New York, New York.
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: ‘isolation by adaptation’ and multiple roles for divergent selection. *Evolution*, **62**,

316–336.

- Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: community ecology package. R. package. version 2.2-1. <http://CRAN.R-project.org/package=vegan>.
- Kittilsen S, Schjolden J, Beitnes-Johansen I, Shaw JC, Pottinger TG, Sorensen C, Braastad BO, Bakken M, Overli O (2009) Melanin-based skin spots reflect stress responsiveness in salmonid fish. *Hormones and Behavior*, **56**(3), 292–298.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics*, **28**, 2537–2539.
- Pinheiro HT, Bernardi G, Simon T, Joyeux J-C, Macieira RM, Gasparini JL, Rocha C, Rocha LA (2017) Island biogeography of marine organisms. *Nature* **549**, 82–85.
- Protas ME, Hersey C, Kochanek D *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, **38**, 107–111.
- Puebla O, Bermingham E, McMillan WO (2014) Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp., Serranidae). *Molecular Ecology*, **23**, 5291–5303.
- Puebla O, Bermingham E, Guichard F, Whiteman E (2007) Colour pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes? *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1265–1271.
- Quax TEF, Claassens NJ, Söhl D, van der Oost J (2015) Codon bias as a means to fine-tune gene expression. *Molecular Cell*, **59**, 149–161.
- R Development Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Randall JE (1963) Review of the hawkfishes (family Cirrhitidae). *Proceedings of the United States National Museum*, **114**, 389–451.
- Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.
- Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact

- tests and ecumenicism. *The Journal of Heredity*, **86**, 248–249.
- Rice WR (1984) Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution*, **38**, 1251–1260.
- Rice WR, Salt GW (1990) The evolution of reproductive isolation as a correlated character under sympatric conditions: experimental evidence. *Evolution*, **44**, 1140–1152.
- Richardson J, Lundegaard PR, Reynolds NL *et al.* (2008) Mc1r pathway regulation of zebrafish melanosome dispersion. *Zebrafish*, **5**, 289–295.
- Ritland K, Newton C, Marshall HD (2001) Inheritance and population structure of the white-phased “Kermode” black bear. *Current Biology*, **11**, 1468–1472.
- Rocha LA, Bowen BW (2008) Speciation in coral-reef fishes. *Journal of Fish Biology*, **72**, 1101–1121.
- Rocha LA, Robertson DR, Roman J, Bowen BW (2005) Ecological speciation in tropical reef fishes. *Proceedings of the National Academy of Sciences*, **272**, 573–579.
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
- Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution*, **58**, 1794–1808.
- Rousset F (2008) genepop’007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Rüber L, Van Tassell JL, Zardoya R (2003) Rapid speciation and ecological divergence in the american 7-spined Gobies inferred from a molecular phylogeny. *Evolution*, **57**, 1584–1598.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.
- Schluter D, Conte GL (2009) Genetics and ecological speciation. *PNAS*, **106 Suppl 1**, 9955–9962.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615–629.
- Selz Y, Braasch I, Hoffmann C *et al.* (2007) Evolution of melanocortin receptors in teleost fish: the melanocortin type 1 receptor. *Gene*, **401**, 114–122.
- Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.
- Shaw KL, Mullen SP (2014) Speciation continuum. *Journal of Heredity*, **105 Suppl 1**, 741–742.

- Shaw KL, Gillespie RG (2016) Comparative phylogeography of oceanic archipelagos: hotspots for inferences of evolutionary process. *PNAS*, **113**, 7986–7993.
- Smouse PE, Chevillon C (1998) Analytical aspects of population-specific DNA fingerprinting for individuals. *The Journal of heredity*, **89**, 143–150.
- Stahl BA, Gross JB (2015) Alterations in Mc1r gene expression are associated with regressive pigmentation in *Astyanax* cavefish. *Development Genes and Evolution*, **225**, 367–375.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, **68**, 978–989.
- Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **66**, 187–205.
- Streelman JT, Alfaro M, Westneat MW, Bellwood DR, Karl SA (2002) Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution*, **56**, 961–971.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
- Taylor MS, Hellberg ME (2005) Marine radiations at small geographic scales: speciation in neotropical reef gobies (*Elacatinus*). *Evolution*, **59**, 374–385.
- Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.
- Trullas SC, van Wyk JH, Spotila JR (2007) Thermal melanism in ectotherms. *Journal of Thermal Biology*, **32**, 235–245.
- Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics*, **162**, 1329–1339.
- van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*,

893 4, 535–538.

894 Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high
895 gene flow species. *The Journal of Heredity*, **89**, 438–450.

896 Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic
897 methods for identifying the number of gene pools and their degree of connectivity.
898 *Molecular Ecology*, **15**, 1419–1439.

899 Watterson GA (1975) On the number of segregating sites in genetical models without
900 recombination. *Theoretical Population Biology*, **7**, 256–276.

901 Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine
902 systems. *Marine Ecology Progress Series*, **393**, 1–12.

903 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure.
904 *Evolution*, **38**, 1358–1370.

905 Whitney JL (2016) A single trait drives incipient ecological speciation in sympatric color morphs
906 of the arceye hawkfish. Doctoral Dissertation, University of Hawaii, Honolulu.

907 Whitney JL, Donahue MJ, Karl SA (2018) Niche divergence along a fine-scale ecological
908 gradient in sympatric color morphs of a coral reef fish. *Ecosphere* 9(1):e02015.

909 Whitney JL, Karl SA (2012) Development of 38 microsatellite loci from the arceye hawkfish,
910 *Paracirrhites arcatus*, using next-generation sequencing and cross-amplification in other
911 Cirrhitid species. *Conservation Genetics Resources*, **4**, 549–553.

912
913
914 **Data Accessibility:**

915 DNA sequences (Mc1r): Genbank accessions MG877306-MG877390. Sampling locations,
916 morphological data, microsatellite genotypes; environmental data, genetic distance matrix:
917 Dryad doi:10.5061/dryad.r51j7

918
919
920 **Author Contributions:** JLW designed the project, developed methodology, conducted field and
921 lab work, analyzed the data, and wrote the manuscript; BWB collaborated in analyses and
922 interpretation of data and assisted with writing and editing the manuscript; SAK assisted with
923 designing the project, collaborated in analysis and interpretation of data and provided editorial

advice.

TABLES

Table 1. Analyses of molecular variance (AMOVA) in *Paracirrhites arcatus* color morphs (and collection site) for 30 microsatellite loci.

Level	% var	F_{ST}	F'_{ST}	P
Within color morph	99.9	—	—	—
Among locations nested in color morph	0.0	-0.002	-0.012	0.977
Among locations	0.0	-0.001	-0.003	0.822
Among color morphs	0.1	0.001	0.006	0.025*

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Table 2. Summary of genic differentiation among *Paracirrhites arcatus* color morphs and geographic regions along west Hawaii Island. Results presented include the overall *P*-value for genic differentiation using all 30 microsatellites, along with the number of individual microsatellite loci displaying significant genic differentiation in each pairwise comparison, as well as the overall D_{EST} with 95% confidence intervals.

Pairwise comparison†	Genic differentiation		Overall <i>D</i> _{EST}	95% CI
	<i>P</i> -value	# of sig loci‡		
Between Regions				
North v. South	0.130	1 (0)	-0.001	(-0.010, 0.008)
North v. South (within MEL)	0.354	1 (0)	-0.001	(-0.013, 0.005)
North v. South (within PWS)	0.187	2 (0)	-0.004	(-0.005, 0.003)
Between Color morphs				
MEL v. PWS	< 0.001*	6 (1)	0.002	(-0.013, 0.005)
MEL v. PWS (within South)	0.116	2 (0)	0.004	(-0.003, 0.014)
MEL v. PWS (within North)	0.087	1 (0)	0.005	(-0.003, 0.008)
Color morph x Region				

MEL North v. PWS South	0.027	4 (1)	-0.003	(-0.011, 0.008)
MEL South v. PWS North	0.006**	6 (0)	0.005	(-0.005, 0.015)

† Sample sizes: (North n = 112: 57 MEL, 55 PWS; South n = 99: 50 MEL, 49 PWS).

‡ number of microsatellite loci (out of 30) with significant differentiation, $P < 0.05$, and after Bonferroni adjustments ($P < 0.01$) in parentheses.

Table 3. Multivariate test for relationship between genetic distance, environmental distance, phenotype and geographic distance using distance-based redundancy analysis (dbRDA) on multilocus microsatellite genotypes of 100 *Paracirrhites arcatus* individuals (50 MEL and 50 PWS morphs).

Variable	d.f.	% var†	<i>F</i>	<i>P</i>
----------	------	--------	----------	----------

Habitat	2	5.9	1.132	0.025*
Color morph	1	1.5	1.442	0.007**
Geographic distance	1	1.1	0.946	0.327
Habitat x Color morph	3	3.2	1.376	< 0.001***
Habitat x Geographic distance	3	1.3	1.100	0.218
Color morph x Geographic distance	1	1.6	1.350	0.019*
Residual†	85	85.4		

† percentage of genetic variation explained by each predictor variable.

‡ habitat type had three levels corresponding to microhabitats: (1) surge zone dominated by *Pocillopora* corals, (2) sub-surge zone co-dominated by *Pocillopora* and *Porites* corals, and (3) reef slope dominated by *Porites* corals.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Table 4 – Summary of analysis of molecular variance (AMOVA) of *Mc1r* gene.

Source of variation	df	SS	MS	% variance	Φ_{ST}	<i>P</i>
Between color morphs	1	6.08	6.08	5%	0.051	0.031
Among individuals within color morphs	83	144.88	2.04	95%	0.284	< 0.001
Total	84	150.9		100%		

Figure Captions

Figure 1. A) *Paracirrhites arcatus* color morphs, from left, pink-white-stripe, melanistic, and intermediate morphs. B) Distributional range of all three color morphs including inset of the main Hawaiian Islands, and C) map of collection sites on Hawai'i Island with piecharts displaying relative abundance of color morphs per site (pink-white-stripe in red; melanistics in black, and intermediates in blue) from survey data presented in Whitney *et al.* (2018).

Figure 2. Bayesian assignment tests of 264 *Paracirrhites arcatus* individuals. (A-B) Clustering results using color morph as a priori grouping, based on (A) all 30 microsatellite loci, and (B) seven loci selected by BELS (backward elimination locus selection) to improve the power of

discrimination among color morphs. Each vertical line represents one individual and its assignment likelihood to the two clusters. Black vertical lines represent the limit between predefined groups based on color morph: melanistic ($n = 107$), pink-white-stripe ($n = 104$), and intermediates ($n = 53$; 13 no-stripe, INS; 16 faint-stripe, IFS; and 24 white-stripe, IWS). (C-D) Clustering results using geographic location as a priori grouping, based on (C) all 30 microsatellites, and (D) seven loci selected by BELS to discriminate among sampling location: Kohala (North; $n = 138$) and Kona (South; $n = 126$). In each barplot, the cluster membership scores represent the average of five runs for each analysis.

Figure 3. Plot of Fdist outlier test showing pairwise F_{ST} estimates of 30 polymorphic microsatellite loci (circles; BELS loci in closed circles) and the *Mc1r* exon (diamond) between MEL and PWS color morphs. Per-locus F_{ST} is plotted against expected heterozygosity to identify potential loci subject to selection. The dashed line corresponds to the median F_{ST} value, and the solid lines indicate upper and lower limits of the 99.9% confidence interval estimated for the neutral model. A single SNP in the *Mc1r* gene (Mc1r-023) was identified as a significant outlier ($P < 0.0001$) as the observed $F_{ST} = 0.175$ was higher than any value in all 50,000 simulations.

mec_14527_f1.tif

mec_14527_f2.tif

mec_14527_f3.tif

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Whitney, JL;Bowen, BW;Karl, SA

Title:

Flickers of speciation: Sympatric colour morphs of the arc-eye hawkfish, *Paracirrhites arcatus*, reveal key elements of divergence with gene flow.

Date:

2018-03

Citation:

Whitney, J. L., Bowen, B. W. & Karl, S. A. (2018). Flickers of speciation: Sympatric colour morphs of the arc-eye hawkfish, *Paracirrhites arcatus*, reveal key elements of divergence with gene flow.. *Mol Ecol*, 27 (6), pp.1479-1493. <https://doi.org/10.1111/mec.14527>.

Persistent Link:

<http://hdl.handle.net/11343/261127>