MS. KRISTEN ALYSSA BEHRENS (Orcid ID : 0000-0003-1895-3523)

Article type : Original Article

Regions of genetic divergence in depth-separated *Sebastes* rockfish species pairs: Depth as a potential driver of speciation

Kristen A. Behrens[1], Quinn L. Girasek[1], Alex Sickler[2], John Hyde[3], Vincent P. Buonaccorsi[1]

[1] Department of Biology, Juniata College, Huntingdon, PA 16652, USA

[2] Center for Data Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[3] Fisheries Resources Division, Southwest Fisheries Science Center, NOAA Fisheries, 8901 La Jolla Shores Dr., La Jolla, CA 92037, USA

Corresponding Author: Kristen A. Behrens, kristenbehrens14@gmail.com

Running Title: Divergence genetics in depth-separated *Sebastes*

Abstract

Depth separation is a proposed driver of speciation in marine fishes, with marine rockfish (genus *Sebastes*) providing a potentially informative study system. *Sebastes* rockfishes are commercially and ecologically important. This genus encompasses more than one hundred species and the ecological and morphological variance between these species provides opportunity for identifying speciation-driving adaptations, particularly along a depth gradient. A reduced-representation sequencing method (ddRADseq) was used to compare 95 individuals encompassing six *Sebastes* species. In this study, we sought to identify regions of divergence

between species that were indicative of divergent adaptation and reproductive barriers leading to speciation. A pairwise comparison of *S. chrysomelas* (black-and-yellow rockfish) and *S. carnatus* (gopher rockfish) F_{ST} values revealed three major regions of elevated genomic divergence, two of which were also present in the *S. miniatus* (vermilion rockfish) and *S. crocotulus* (sunset rockfish) comparison. These corresponded with regions of both elevated D_{XY} values and reduced nucleotide diversity in two cases, suggesting a speciation-with-gene-flow evolutionary model followed by post-speciation selective sweeps within each species. Limited whole genome re-sequencing was also performed to identify mutations with predicted effects between *S. chrysomelas* and *S. carnatus*. Within these islands, we identified important SNPs in genes involved in immune function and vision. This supports their potential role in speciation, as these are adaptive vectors noted in other organisms. Additionally, changes to genes involved in pigment expression and mate recognition shed light on how *S. chrysomelas* and *S. carnatus* may have become reproductively isolated.

Key words: genomic islands of divergence, depth separation, speciation, *Sebastes*

Introduction

During allopatric speciation, genomes diverge over time by mutation, genetic drift, and natural selection (Via and West 2008; Via 2009). By contrast, speciation-with-gene-flow is theorized to occur based on differences in niche exploitation, where species diverge based on macro habitats or environmental gradients (β-niches), or due to micro-level relationships involving partitioned resources such as food (α-niches; Ingram, 2011). Speciation-with-gene-flow may involve primary or secondary contact, which can alter genome structure and influence factors such as genome hitchhiking (Nosil & Feder, 2012). Genomic islands of divergence, which are regions of the genome that are diverged between species, may form and contain genes that are important to speciation (Turner et al., 2005; Nosil et al., 2009). Many studies have identified genomic islands of divergence in species comparisons (Han et al., 2017; Hess et al., 2020; Hohenlohe et al., 2010).

Islands of divergence may be characterized by elevated relative divergence measures such as fixation indexes (F_{ST}) flanked by regions of lower F_{ST}. However, re-examination of several studies showed that calculating absolute divergence measures not impacted by intraspecific variation, such as $D_{XY,}$ at these elevated regions is critical to determining whether

regions arose pre- or post-speciation, as F_{ST} can be influenced by intraspecific variation (Cruickshank & Hahn, 2014; Nei, 1987). Genomic islands of divergence that arise pre-speciation should exhibit high absolute and relative measures of divergence concurrently, while surrounding loci remain homogenized as a result of gene flow. Lower nucleotide diversity within these islands and between species would suggest continued selection or recent selective sweeps (Cruickshank & Hahn, 2014; Han et al., 2017).

Multiple models attempt to predict the conditions that form genomic islands of divergence. Divergent natural selection resulting in reduced gene flow was initially favored (Wu, 2001), however divergence with gene flow via divergence hitchhiking (Feder & Nosil, 2010; Via, 2009, 2012; Via & West, 2008) and other causes occurring in the absence of gene flow such as resistance to introgression in these islands (Nosil & Feder, 2012) have offered alternate explanations (Cruickshank & Hahn, 2014). Four non-exclusive paths to islands of divergence have been proposed: differential gene flow between loci, lineage sorting of ancestral polymorphisms, adaptation without differential gene flow, and the effects of divergence hitchhiking and background selection (Han et al., 2017). Modeling has found that formation of genomic islands of divergence relies on multiple interacting factors such as gene flow, divergent selection, linkage, drift, and time since divergence (Quilodrán et al., 2020). Differential gene flow is not necessary for these regions to arise; however, limited gene flow can extend the number of generations that the region persists, as increasing gene flow is predicted to be correlated with decreasing instances of genomic islands (Quilodrán et al., 2020). Often, genomic islands are eliminated over time by consistently elevated genome-wide divergence and are more apparent during earlier periods of divergence (Quilodrán et al., 2020). Due to this, speciation-with-gene-flow is easier to identify in younger or slowly diverging species and harder to characterize in later stage speciation (Ravinet et al., 2018).

Depth separation is a recognized driver of speciation in marine fishes, including species in the genus *Sebastes* (Ingram, 2011; Sivasundar & Palumbi, 2010; Shum et al., 2014). *Sebastes* features cases of likely parapatric speciation between species at similar latitudes based on α-niche diversity involving divergence along a depth gradient (Hyde & Vetter, 2007; Ingram, 2011; Stefánsson et al., 2009). For example, *S. miniatus* and *S. crocotulus* are sister species strongly separated by depth, where adult *S. miniatus* are found mainly in shallower depths (<100m) and *S. crocotulus* at >100m (Hyde et al., 2008). Associations have been drawn between

the *rhodopsin* gene, which produces an extremely light-sensitive retinal pigment, and depth preference in North Pacific *Sebastes* species (Shum et al., 2014). North Atlantic *Sebastes* species share the same amino acid changes in the *rhodopsin* gene. These results are potentially indicative of adaptation based on divergent environmental conditions as a result of positive selection (Sivasundar & Palumbi, 2010; Shum et al., 2014). An analysis comparing *Sebastes* RNA-seq data found that the hemoglobin subunit alpha gene (*HBA2*) is under strong positive selection associated with species living at different depths (Heras & Aguilar, 2019). The association of hemoglobin with depth lends additional credibility to the role of depth in species diversification due to the variation in oxygen levels based on depth. Additionally, two immune-related genes, *E-D α chain* (*HA21*) and *membrane cofactor protein* (*CD46*), were under positive selection, implicating a role of immune genes in depth-based speciation (Heras & Aguilar, 2019).

Sebastes are commercially and ecologically important, with approximately 110 known species. Extensive variability among species, a broad range of divergence times, and cases of potential divergence with gene flow provide an opportunity to investigate islands of genomic divergence related to speciation (Hyde & Vetter, 2007). Rockfishes in the Northeastern Pacific are found from intertidal ocean waters to depths greater than 1500m in almost every habitat type from Mexico to Alaska (Love et. al. 2002). Worldwide speciation mechanisms in rockfishes have likely involved a variety of evolutionary processes, such as rare long-distance migrations followed by geographic isolation (Love et al. 2002; Hyde and Vetter 2007). In ecological speciation, prezygotic differentiation may be facilitated in *Sebastes* as a result of their internal fertilization strategy, which provides the opportunity for mate choice, courtship rituals, and assortative mating.

Several incipient sister species have recently been discovered within the genus, where the species occupy overlapping ranges and are morphologically similar, with only slight differences in habitat preference (Narum et al. 2004; Hyde and Vetter 2007; Hyde et al. 2008) or latitudinal occurrence (Burford & Bernardi, 2008). The genetic mechanism of speciation in these sister species is unknown, although truncation in ontogenetic migration from shallow juvenile settlement to deep adult habitat may have been important in developing a genetic basis for niche preference for *S. miniatus* and *S. crocotulus* (Hyde et al. 2008). In *S. chrysomelas* and *S. carnatus*, depth segregation is initiated by preferential settlement of larvae at different depths and reinforced through intense interspecific defense of highly contested territories (Larson, 1980).

The *S. chrysomelas* and *S. carnatus* pair was shown to have an exceptionally low background allele frequency divergence (F_{ST}) and some evidence of recent introgressive hybridization (Buonaccorsi et al., 2011; Narum et al., 2004). Additionally, a diverged microsatellite region was characterized between them. These studies suggest that with the recent high-quality assembly of the Korean rockfish/black rockfish *Sebastes schlegelii* genome as a reference, there is potential to achieve high resolution of regions of divergence in sister rockfishes (He et al., 2019).

In this study, we aim to use the phylogenetic replication offered within the genus *Sebastes* to examine the genomic architecture of depth-related speciation and identify functional variants in regions implicated in the speciation event. Three species pairs were selected for this study, *S. chrysomelas* (black-and-yellow rockfish) and *S. carnatus* (gopher rockfish)*, S. miniatus* (vermilion rockfish) and *S. crocotulus* (sunset rockfish), and *S. mentella* (deep-water redfish) and *S. alutus* (Pacific Ocean perch). *S. chrysomelas* and *S. carnatus* are depth separated, are the most recently diverged rockfishes (~0.5 MYA; Hyde and Vetter 2007), and have broadly overlapping latitudinal ranges. *S. miniatus* and *S. crocotulus* are less recently diverged (~2.3 MYA; Hyde and Vetter 2007) depth-segregated species. *S. mentella* and *S. alutus* inhabit the North Atlantic and North Pacific respectively. Invasion and subsequent allopatry of Atlantic from Pacific taxa are hypothesized to have occurred 4.8-5.5 MYA with the opening of the Bering Strait (Marincovich & Galdenkov, 2001), after which speciation presumably occurred allopatrically. This example of ancient and allopatric divergence will provide context for divergence patterns in the focal comparisons. We first used ddRADseq (Peterson et al., 2012) data to test whether individuals of *S. miniatus* and *S. crocotulus* showed evidence of recently mixed ancestry as was noted between *S. chrysomelas* and *S. carnatus* previously (Buonaccorsi et al. 2011). We then identified regions of divergence between paired species that were indicative of reproductive barriers leading to speciation. Furthermore, we performed whole genome resequencing of a single individual each from *S. chrysomelas* and *S. carnatus* to help eliminate genes with no inter-specific variation as candidates for selection, and to identify possible target genes containing SNPs with functional effects. Under ecological speciation, we expected a combination of genes primarily involved in pre-zygotic reproductive barriers including resource partitioning, habitat preference, and mate recognition system, from a small number of large (>1Mbp) divergence islands. While reduced recombination associated with a centromere, sex chromosome, or inversion may facilitate combinations of genes evolving in linkage disequilibrium, in the model of ecological speciation,

they aren't necessary. The identification of these regions will be key to understanding the genes underlying speciation in *Sebastes*. It may also provide insight into how sympatric marine fish species diverge in general, a long-standing question due to the lack of geographical barriers in marine environments.

Methods

Sampling

Samples of *S. carnatus* ($N = 20$) and *S. chrysomelas* ($N = 20$) used in the present study were collected from the Southern California Bight and previously reported in Buonaccorsi et al. (2011). *S. miniatus* ($N = 15$) and *S. crocotulus* ($N = 16$) individuals were available from Hyde (2008). *S. mentella* ($N = 8$) and *S. alutus* ($N = 16$) individuals were additionally available from NOAA Southwest Fisheries Science Center collections for a total of 95 samples.

Library Prep and Sequencing

Double-digest restriction associated digestion and sequencing of genomic DNA was performed (ddRADseq; Peterson et al. 2012). Genomic DNA from each individual was cut with a rare cutting (6 bp; e.g. *Sph*I) enzyme and a frequent cutting enzyme (4 bp; *MluC*I), following manufacturer's protocols and libraries were constructed at Indiana University (IU), following Peterson et al. (2012). Sixteen internal barcodes and six indices were used. The Qubit 2.0 high sensitivity DNA assay (Life Technologies, Carlsbad, CA) was used for double-stranded DNA quantification and the IU genome core used gel electrophoresis for fragment size extraction with an average insert size of 300bp. The library was then subject to a single lane of 2x100bp paired-end Illumina sequencing on a HiSeq 2000. To obtain more exhaustive SNP information from two target species, genomic DNA from a single individual each of *S. carnatus* (GW23) and *S. chrysomelas* (GW1) were sequenced using a single lane of 75BP single-end reads from a high-output Illumina NextSeq protocol from IU. Genomic DNA was prepared using Bio Rapid DNA libraries.

Bioinformatic Analysis

The STACKS v2.52 ref_map.pl pipeline (Catchen et al., 2013) was used to process the data and generate population statistics. Raw data were demultiplexed using default settings of the STACKS v1.36 process_radtags script. Reads were aligned to the *S. schlegelii* genome using BOWTIE2 (Langmead & Salzberg, 2013) at default settings, and then were filtered with SAMTOOLS (Li et al., 2009) to remove reads that may have aligned to multiple locations

(mapping quality score less than ten). Construction of the STACKS catalog, SNP calling, and genotype construction was performed using the ref_map.pl pipeline at default settings. The populations sub-pipeline of STACKS v2.52 was run with all six species included in analysis, with the following optional flags: F_{ST} calculations enabled for pairwise comparison, minimum percentage of individuals in a population (i.e., species) required to process a locus (0.5), minimum number of populations a locus must be present in to process a locus (2), minimum minor allele frequency (0.05), and maximum heterozygosity (0.7). Smoothing was enabled and bootstrap resampling for smoothed statistics was set to 1000 reps. Plots were generated using the R packages tidyverse (R Core Development Team, 2019; Wickham et al., 2019) and qqman (Turner, 2018). Loci were assessed for deviations from Hardy-Weinberg equilibrium within species using the probability test of GENEPOP v4.4.3. Loci were filtered for sites in Hardy-Weinberg disequilibrium within a species (F_{ST} and D_{XY} by site, pi by locus, $p < 0.001$). Loci were also assessed for genotypic association with sequencing library pools within species using the exact test of GENEPOP v4.4.3 and removed if $p < 0.001$. Nucleotide diversity was based on individual sites that were variable in at least one species in a species pair (Nei and Li 1979). In order to determine if the sex-determining region was involved in species divergence, the previously characterized y-specific loci from Fowler & Buonaccorsi (2016) were located in the *S. schlegelii* reference genome using blastn. Blastn was also used to align the *Sra*.7-2 microsatellite region previously identified as an F_{ST} divergence island by Buonaccorsi et al. (2011) against the *S. schlegelii* genome assembly (e-value = $1e^{-6}$, max target sequences = 1).

The program STRUCTURE (Pritchard et al. 2000) was used to evaluate whether there was evidence of recent hybridization between *S. miniatus* and *S. crocotulus*. For this analysis, one SNP each from the first 30 loci that had a minor allele frequency of at least 0.1 in at least one species, and no more than 10% missing data in either species were used. STRUCTURE was run with two assumed populations, the admixture and correlated allele frequency model, 200,000 burn-in steps and 1,000,000 repetitions of the Markov Chains, alpha inferred, and without using prior information on species of origin. Six iterations were run and processed using CLUMPAK (Kopelman et al. 2015) to visualize results over all runs.

We explored peaks of genomic divergence and diversity in order to better understand speciation mechanisms. To define peaks, we used a sliding 500 KBp window to calculate ZF_{ST} (standardized F_{ST}) values following Han et al. (2017). The mean F_{ST} was taken from SNPs in the

window, subtracted from the mean of the whole chromosome, and divided by the standard deviation of the whole chromosome. Values greater than four standard deviations from the mean were considered outliers, corresponding to the 99.3% percentile of mean F_{ST} values from windows for all the comparisons. Window size of 500 KBp was chosen to ensure that a substantial number of SNPs were covered (~20) in most windows. Intervals with fewer than three SNPs were not considered in ZF_{ST} calculation due to low precision. Adjacent significant windows were merged. Significance of D_{XY} was then assessed for the ZF_{ST} intervals using Monte Carlo simulations to assess whether ZF_{ST} peaks had greater D_{XY} values (1-tailed test) than the chromosome background D_{XY} level (Manly, 2006). Genome-wide patterns of D_{XY} were examined as well. Significance of nucleotide diversity (pi) values was then assessed for the ZF_{ST} intervals using Monte Carlo simulations to assess whether peaks values differed (in either direction) from chromosome background level. Paired t-tests were performed to determine if paired species differed in diversity for peak or background regions. To determine if any polymorphisms were shared between the relatively shallow-water species (*S. chrysomelas* and *S. miniatus*), or between deep-water species (*S. crocotulus* and *S. carnatus*), corresponding pairwise comparisons of ZF_{ST} were calculated and examined for negative outliers.

For exhaustive SNP identification in regions of interest, single individuals from *S. chrysomelas* and *S. carnatus* were aligned against the *S. schlegelii* reference genome using BOWTIE2 (Langmead & Salzberg, 2013) at default settings. They were then filtered with SAMTOOLS (Li et al., 2009) to remove reads with a mapping quality score less than ten. SAMTOOLS was used to convert bam files to mpileup format. VarScan2 (Koboldt et al., 2012) was used to call both SNPs and indels ($p < 0.01$). The variant effect predictor SNPEFF was used to predict function of SNPs using the *S. schlegelii* genome and annotation as reference (Cingolani et al., 2012). The Unix tool GREP was used to pull SNPs from the diverged regions for high or moderate effects and that were homozygous and different between *S. chrysomelas* and *S. carnatus*. This involved identifying cases where SNPs in one species matched the reference while the other had the alternate allele, or where SNPs in both species of interest differed from the reference allele and each other. Prediction of functional effects of non-synonymous SNPs was further assessed using SNAP2 (Hecht et al., 2016). Because SNAP2 functional predictions are based on single amino acid substitutions only, we considered genes with multiple amino acid substitutions between the two focal species to be of potential functional

247 importance even if each individual mutation was considered neutral by SNAP2. Substitutions
248 where both focal species differed from the reference but were identical to each other were
249 disregarded. While some SNPs in divergence islands will represent within-species
250 polymorphisms, our approach will capture fixed differences between species. This will filter out
251 most genes in the divergence regions from further consideration. After stringent effect filtering,
252 we then identify for further exploration a finite pool of candidate SNPs that may have important
253 functional differences between species.
254 **Results**
255 An average of 1.41M paired reads were obtained per individual after quality filtering (26.98Gbp
256 total sequence), with an average overall alignment rate of 85% to the *S. schlegelii* genome. A
257 total of 62569 variable sites were found between *S. chrysomelas* and *S. carnatus (Sch-Sca)*, a
258 total of 47500 variable sites between *S. miniatus* and *S. crocotulus (Smi-Scr)*, and a total of
259 105195 variable sites between *S. alutus* and *S. mentella (Sal-Sme)* prior to filtering (Table 1).
260 Filtering for pool bias removed a total of 17 sites for *Sch-Sca*, three for *Smi-Scr*, and three for
261 *Sal-Sme*. Filtering for HW disequilibrium resulted in removal of a total of 1455 sites from *Sch-*
262 *Sca*, 571 from *Smi-Scr*, and 550 from *Sal-Sme*. STRUCTURE results showed evidence for recent
263 introgressive hybridization between *Smi-Scr* in two of the 32 individuals. Individual MY6 was
264 likely an F1 hybrid with mean ancestry to *S. miniatus* at 0.48 (mean 90% credibility interval 0.38
265 to 0.62). VMZ4 was likely a hybrid backcross to *S. crocotulus*, with mean ancestry to *S. miniatus*
266 of 0.186 (mean 90% credibility interval 0.071 to 0.325). The F1 hybrid was removed from
267 further population analysis. Backcrossed individuals were not excluded as they were considered
268 part of the gene pool for their respective (majority) species. Average F_{ST} values for each pair
269 were 0.0248 (*Sch-Sca*), 0.2297 (*Smi-Scr*), and 0.7321 (*Sal-Sme*; Table 1). Overall, average
270 nucleotide diversity was higher in *S. chrysomelas* (0.2549) than *S. carnatus* (0.2505; paired
271 $t_{60,509df} = 8.84$; p <2.2e-16), higher in *S. miniatus* (0.2144) than *S. crocotulus* (0.1837; paired
272 $t_{45,624df} = 22.0$; p <2.2e-16), and higher in *S. alutus* (0.2079) than *S. mentella* (0.0765; paired
273 $t_{48,043df} = 102.6$; p <2.2e-16). Pairwise comparisons of ZF_{ST} between the two shallow-water and
274 two deep-water species did not reveal negative outliers that might indicate shared inter-specific
275 polymorphisms.
276 A number of divergence islands were detected between *Sch-Sca*. ZF_{ST} analysis showed
277 seven significant peaks above four standard deviations (Table 2). The most pronounced

278 divergence peaks were apparent from smoothed F_{ST} graphs and ZF_{ST} analysis on chromosomes
279 four, nine, and 13 (Figures 1-3, Figure S1). D_{XY} divergence was also significantly greater than
280 chromosomal background levels for only these three ZF_{ST} peaks of the seven (Table 2). Adjacent
281 significant windows were merged so peak sizes were 1.5 M, 0.5 Mbp and 2.5 Mbp for
282 chromosomes four, nine, and 13, respectively (Table 2). D_{XY} values similarly transformed to a
283 ZD_{XY} scale showed no outliers in any species comparison (Figure S2). Nucleotide diversity was
284 lower in *S. chrysomelas* and *S. carnatus* for ZF_{ST} peaks than chromosomal background for
285 chromosomes four and nine. For chromosome 13, diversity was lower than the background in *S.*
286 *chrysomelas* but higher than background in *S. carnatus* (Table 2). In comparisons between
287 species, nucleotide diversity was also lower in *S. chrysomelas* than *S. carnatus* for each ZF_{ST}
288 peak (significant at chromosomes nine and 13) notwithstanding that chromosome background
289 diversity was higher in *S. chrysomelas* (Table 2).
290 The ZF_{ST} peaks with significant D_{XY} between *Sch-Sca* were also examined between *Smi-*
291 *Scr.* For the peaks at chromosomes four and 13, smoothed F_{ST} and ZF_{ST} analysis also showed
292 peaks between *Smi-Scr*, but these were not significant for ZF_{ST} or D_{XY}. Diversity at peaks for
293 chromosomes four and 13 in *Smi-Scr* were lower than chromosomal background and did not
294 differ between species (Table 2). The previously characterized *Sra*.7-2 microsatellite region from
295 the *Sch-Sca* study (Buonaccorsi et al., 2011) was identified within the central peak of the
296 smoothed F_{ST} defined divergence region on chromosome nine (9711013-9711162 bp). The sex-
297 determining regions from Fowler & Buonaccorsi (2016) were found via BLAST to have multiple
298 radtag hits against chromosome 17 (6958394-7237580 and 22513546-30244549 bp).
299 Six significant ZF_{ST} peaks were detected for the *Smi-Scr* comparison and none between
300 *Sal-Sme.* Of the six ZF_{ST} peaks between *Smi-Scr,* two had significant D_{XY} divergence compared
301 to background, and one overlapped the chromosome 13 peak observed for *Sch-Sca* (Table S1).
302 The chromosome four peak was also visible in both ZF_{ST} and smoothed F_{ST} graphs between *Smi-*
303 *Scr*, but was not significant (Figure 2, Figure S1). Diversity did not differ between peaks and
304 background for any peak, and diversity did not differ between species for peak regions (Table
305 S1). No distinct ZF_{ST} peaks of divergence were evident in the *Sal-Sme* comparison (Figure S1,
306 Figure 3).
307 Divergence islands did not have a higher relative gene density than background levels,
308 nor were they larger or more numerous in the most anciently diverged species pair. For the *Sca-*

Sch comparison there were three F_{ST}/D_{XY} islands with a mean size of 1.5Mbp, representing 0.56% of the genome. Islands contained only 0.43% of genes, which was lower than the average gene density (X^2 GOF test; $P = 0.0019$). For the *Smi-Scr* comparison there were three significant F_{ST}/D_{XY} islands with mean size of 0.667 Mbp, representing 0.25% of genome and 0.2% of genes. This gene density was not different from expected (X^2 GOF test; $P = 0.07$).

Three high effect SNPs and 36 moderate effect SNPs and indels in 22 genes were identified across all three *Sch-Sca* divergence regions and were 1) homozygous in both species and 2) different between *S. chrysomelas* and *S. carnatus*. From the location of these SNPs, genes of interest were identified. Seven genes had SNAP2 predicted effects. Two genes had start or stop loss mutations (*igsf11* and an unidentified locus), and an additional two had frameshift (*CYP2F2*) or in-frame deletion mutations (*peroxisome assembly protein 12*; Table 3). These genes were further examined for potential involvement in speciation based on function as described in the literature (Table 4). Two genes that had SNPs of biochemical significance but that failed to analyze in SNAP2 were *ryanodine receptor 2* (*RYR2,* Val2728Ala) and *pecanex-like protein 3* (*PCNX3*, Ile200Thr). The effect of this amino acid change on *RYR2* is likely neutral as both amino acids are non-polar, however the effect on *PCNX3* cannot be determined with the available data.

Discussion

Divergence Regions

The aim of this study was to identify genomic islands of divergence in depth-separated *Sebastes* species pairs, infer evolutionary models of divergence, and characterize functional variation in those regions. We present genes that may play a role in speciation or adaptation, as well as address the absence of genes implicated in depth-related divergence from other rockfishes, *rhodopsin* and *HBA2,* in our genomic islands of divergence. Our data reflected the expected relative degrees of baseline divergence between species pairs based on previous phylogeny (Hyde & Vetter, 2007). For the *Sch-Sca* comparison, major trends from Buonaccorsi et al. (2011) were reinforced, including low overall divergence, existence of genomic islands of high F_{ST} values, and lower diversity within divergence islands, in particular within *S. chrysomelas*. Furthermore, our findings are largely concordant with the theory that divergence islands reflect a speciation-level event followed by recurrent selection, and that gene flow between species varies among regions of the genome.

In the *Sch-Sca* comparison, three significant divergence regions were identified and supported by both ZF_{ST} and D_{XY} analyses, consistent with divergence in those regions dating to the speciation event. The regions on chromosomes four and 13 corresponded with two evident divergence regions in the *Smi-Scr* comparison, one of which made the four-standard deviation ZF_{ST} cutoff and the D_{XY} cutoff. This suggests that these regions may play a role in depth-based divergence as they are differentiated between two species pairs that are depth-separated. It is also possible that a common mate recognition mechanism is supported in the region. A common region of genomic divergence providing a mechanism for variation in plumage coloration was detected between pairs of warblers and within a species of finch (Kim et al., 2019). Other examples of parallel evolution have been detected. For example, Ravinet et al. (2021) described parallel introgression genomic sites between two species of stickleback in two different regions of the species range that were associated with three QTLs conveying heterozygote advantage for juvenile survival. Together with our findings, it may be common for certain regions of the genome to have broad evolutionary consequences for different taxa. Two other divergence islands in the *Smi-Scr* comparison were also significant at both ZF_{ST} and D_{XY} cutoffs. Divergence was high between *Sal-Sme*, species that likely diverged allopatrically and did not show signs of divergence islands.

Elevated F_{ST} and D_{XY} at the identified regions for both depth-segregated species pairs suggests that these genomic islands of divergence may be the result of either reduced gene flow at certain regions of the genome during speciation, or lineage sorting of ancient, diverged haplotypes (Ma et al., 2017; Han et al., 2017). The high divergence at these islands has developed despite evidence of recent introgressive gene flow detected between *Smi-Scr* (this study) and between *Sch-Sca* species pairs (Buonaccorsi et al. 2011), reinforcing the notion of reduced effective gene flow at divergence islands. Inter-specific gene combinations at divergence islands must be selected against in order for us to observe nearly fixed differences at some locations despite ongoing gene flow between species. Based on this evidence, in addition to F_{ST} and D_{XY} divergence, we suggest that a speciation-with-gene-flow evolutionary model is more likely than a model of solely post-speciation selective sweeps in these species (Ma et al., 2017; Han et al., 2017).

In the three peak regions between *Sch-Sca,* nucleotide diversity was low compared to background. This implicates the influence of post-speciation selection such as selective sweeps

and recurrent selection based on variation in these regions for *S. chrysomelas* and *S. carnatus*. Between species, *S. carnatus* displayed higher diversity in the peak regions than *S. chrysomelas*, notwithstanding that genomic average diversity was slightly higher in *S. chrysomelas*. This suggests that selective pressures in the region may have continued to act in *S. chrysomelas* after speciation in a stronger way than *S. carnatus*, more intensely reducing the effective population size of that region. In our study, four of seven ZF_{ST} peaks between *Sch-Sca* also did not have higher D_{XY} and thus more likely represented the results of post-speciation selection, although nucleotide diversity did not differ from background or between species consistently. Peaks in chromosomes four and 13 in *Smi-Scr* showed lower diversity than background when defined using the ZF_{ST} peak coordinates from *Sch-Sca*. Since there was additional evidence of elevated D_{XY} divergence for the chromosome 13 peak between *Smi-Scr* ($p < 0.1$), this region was likely important for speciation of both species pairs, suggesting common mechanisms. Either a common genomic feature (e.g., region of low recombination) and/or a common genetic mechanism underlying the organism's response to environmental or evolutionary challenges is implicated.

Candidate Genes For Adaptation Or Speciation

Here we suggest hypotheses for potential roles of genes of interest in adaptation or speciation. Ion transport differences between *Sch-Sca* may be related to fine tuning of the mate recognition system. Rockfish are internal fertilizers and exhibit male courtship display before copulation. Male rockfish are known to urinate near female's snouts before mating in some species, and this is thought to send a message to females that makes coupling more probable (Helvey, 1982; Shinomiya & Ezaki, 1991). Male blue rockfish (*S. mystinus*) were also shown to have larger urinary bladders than females (Helvey, 1982). Here, a nonsynonymous substitution I297V in *Ssc_10012345* of predicted moderate effect between *Sch-Sca* was identified as in *OAT1* (Table 4). This study raises the possibility of a non-olfactory receptor mechanism for generating and detecting species-specific odorants, thereby facilitating mate recognition. A "remote sensing hypothesis" has suggested that *OATs* are involved in non-olfactory receptor inter-organismal communication (Nigam, 2018; Nigam et al., 2015; Wu et al., 2011). *OAT* expression is also regulated by sex hormones, and sex-specific expression patterns have been detected (Ljubojević et al., 2004). Further, the *OAT* gene family is found clustered in the genomes of humans and

rodents, as well as in *S. schlegelii*, indicating that other polymorphisms in this region may also play a role in species divergence.

Odc1 (Table 4) features two SNPs predicted as moderate effect by SnpEff and neutral by SNAP2. This gene is potentially important for species-specific ability to adapt vision to varying light amounts and wavelengths. *Igsf11* (Table 4), involved in pigmentation, has a high effect (start codon loss) and a second SNP predicted as moderate effect by SnpEff and neutral by SNAP2. As depth increases, available light decreases and the wavelength of light able to penetrate becomes shorter (Smyth, 2011). Adaptations for color vision based on niche depth and water clarity have been noted extensively in cichlids as a result of changes to opsins and opsin expression (Carleton & Kocher, 2001; Carleton & Yourick, 2020; Dalton et al., 2015). *S. chrysomelas* and *S. carnatus* are morphologically similar except for distinct color patterns (e.g., Orr et al., 2000; Orr and Blackburn, 2004) so vision-related adaptations may be essential to reproduction and color assortative mate-selection (Elmer et al., 2009). While the *odc1* gene is not noted in the well-documented role of opsins in color and light adaptations, we hypothesize that its function in photoreceptor development may contribute to the underlying pathway and warrants further investigation. Changes to pigmentation in combination with potential visual tuning may be a mechanism for reproductive isolation (RI) in *Sebastes*.

NLRP12 (Table 4), an immune-related gene, featured four SNPS called as moderate effect by SnpEff, one of which was also determined to have an effect by SNAP2. A transcriptomic survey of adaptive evolution associated with depth and age found the immune-related genes *HA21* and *CD46* to be under positive selection in *Sebastes* inhabiting different depths (Heras & Aguilar, 2019). Environmental temperature has been noted to impact immune response in fish, as immune system response can become adapted to specific temperatures depending on when immunological memory is established (Le Morvan et al., 1998). Interactions between host and pathogen are impacted by environmental temperature and in marine environments, changes in temperature can increase pathogen virulence and range, as well as induce stress in the host organism (Cohen et al., 2018). It is possible that these genes are differentiated between *Sch-Sca* due to differences in temperature or variance in pathogen identity and density in their depth ranges, leading to necessary immune adaptation by the organism.

Several genes have more uncertain roles in adaptation or speciation, so while they featured functional SNPs, we will not provide hypotheses for them. *PHLDB2* (Table 4)

contained four SNPs predicted as moderate effect by SnpEff and neutral by SNAP2. In the Heras & Aguilar (2019) study, pleckstrin homology domain was an enriched functional gene cluster in both the age and age/depth overlap analyses. Currently, it is unclear how this gene plays a role in depth-based adaptation, but its occurrence in two studies warrants further investigation.

Some genes that have been identified as important to divergence of rockfishes in previous studies were not detected in the three focal divergence islands between *Sch-Sca*. The *rhodopsin* gene from Sivasundar & Palumbi (2010) was not identified in the divergence regions in our study. This is likely due to minimal divergence at this gene between *Sch-Sca*, which were grouped into the same clade based on the *rhodopsin* gene (Sivasundar & Palumbi 2010). We also did not observe the *HBA2* gene from Heras & Aguilar (2019). It is possible that the *HBA2* gene was in a divergence island too small to be detected given the resolution of ddRADseq. However, there may also be no significant divergence of the *HBA2* gene between these two species due to lack of sufficient difference in depth to drive selection on this gene.

Paths To Speciation

Models of speciation with gene flow have proposed general patterns of divergence between species pairs. Feder et al. (2012) proposed four phases of species divergence. The *Sca-Sch* species pair would likely be in phase 2, divergence hitchhiking (DH), characterized by new mutations that begin to diverge surrounding the locus of direct selection. The *Smi-Scr* pair would likely be phase 3, genomic hitchhiking, characterized by a rise in background divergence level and multiple loci under divergent selection. Finally, *Sal-Sme* would be more likely to be in phase 3 or phase 4, post speciation divergence, where introgression is unfavored altogether and divergence becomes widespread. Many cases of heterogeneous gene flow among different genomic sites in species pairs with intermediate levels of divergence have been documented (Roux et al. 2016). A newer model of speciation focuses on the degree to which species are reproductively isolated along a continuum (Stankowski and Ravinet, 2021). While the present study was not focused on quantifying the strength of barriers to reproduction, for both the *Smi-Scr* pair and *Sca-Sch* pair (Buonaccorsi et al., 2011) hybridization still occurs, and F1s were both viable and capable of backcrossing to some extent. This demonstrates that reproductive isolation is not complete for either species pair. Authors have noted that RI is essentially complete in species pairs with at least 2% net nucleotide divergence (Roux et al., 2016). Levels in the species pairs from this study were below 2%, although the *Sal-Sme* comparison was close (at 1.54%).

The others were lower (*Smi-Scr* 0.72%; *Sca-Sch* 0.46%), concordant with our observations that RI was not complete for these comparisons. While the *Sal-Sme* pair live in different oceans, the high divergence suggests they are approaching complete RI if they were they to come into contact.

Some predictions of the genomic architecture of the divergence hitchhiking model of speciation were not met. Speciation with gene flow may occur more easily if divergence islands were gene-rich, allowing for a greater chance of linked selection on adaptive polymorphisms, and evidence for this pattern has been found (Schreiber and Pfenninger, 2021). However, lack of a higher gene density in divergence islands in the present study points to the importance of the particular genes in the region, rather than an abundance of random evolutionary targets. The fact that the divergence islands were shared among species pairs supports the notion that these certain locations may be particularly significant for speciation. Pre-zygotic isolating mechanisms are expected to evolve more quickly in cases of ecological speciation, and of our candidate genes, *OAT1*, *odc1*, and *igsf11* are likely pre-mating barriers between species due to their anticipated roles in mate recognition. *PEX12* is likely also a pre-mating barrier due to a role in niche exploitation (Table 4). *NRFKB*, *CYP2F2*, *B4galnt2*, and *NLRP12* are likely post-mating/post-zygotic barriers (Table 4).

We expected the number and size of divergence islands to be greater in the species pair with greater divergence time, following the theory of divergence hitchhiking (Feder et al., 2012). While we did not detect that pattern, F_{ST} outliers are statistically harder to detect as background divergence increases, which may have limited resolution. Also, islands may not be in the process of growing, but rather have reached different equilibrium divergence levels reflecting factors specific for the species comparison, like strength of selection, recombination, and gene flow. Alternatively, islands may also be in the process of growing, but the pairs are approaching RI at different rates due to variation in demographic history or environmental challenges.

Speciation of marine organisms has been historically difficult to characterize (reviewed in Faria et al., 2021), however, our analysis provides valuable insight into patterns of genomic divergence. In the most recently diverged species *Sca-Sch* (~0.5 MYA; Hyde and Vetter 2007), F_{ST} was low across the genome except for the peaks we note to be potentially involved in divergence. Peak size was consistent with expectations of ecological divergence of a small number of large islands, allowing multiple traits to diverge simultaneously. The evidence for

recent gene flow, along with low diversity of some divergence peaks implies their continued importance in maintaining species divergence. By contrast, the less recently diverged *Smi-Scr* (~2.6 MYA; Hyde and Vetter 2007) has higher average F_{ST} and fewer discernible peaks. *Sme-Sal* (~4.8-5 .5MYA; Marincovich & Galdenkov, 2001) has the highest average F_{ST} and no discernible peaks due to high overall divergence and a likely allopatric speciation mechanism. Therefore, there is likely a divergence time at which F_{ST} becomes unusable for this purpose, which is described by Quilodrán et al. (2020) and the difficulty of identifying divergence islands in more diverged species is noted by Ravinet et al. (2018). This study shows how parapatric speciation might arise across a depth gradient and shape the division of marine organisms. Our candidate gene hypotheses provide groundwork for further investigation of depth-separation as a driver of speciation. Previously, metabolic adaptations have been proposed in deep-sea organisms, however adaptations related to mate recognition, vision, and immune response may characterize important facets of differentiation between species (Childress, 1995).

Study Limitations

A limitation of this study is detecting and estimating size of genomic islands due to the resolution of the sequencing technology. Here, with approximately 50,000 SNPs to cover an 800 Mbp genome, an average of one SNP per 16,000 bp is obtained. Windows of 500 Kbp would have an average of 31 SNPs if randomly distributed. Larger windows than in some previous studies were used here in order to ensure a minimum sample size to statistically compare mean ZF_{ST}, D_{XY} and Pi values. Another limitation was related to the whole genome sequencing, as only a single individual from each species was used to evaluate SNPs in divergent regions. While we excluded heterozygous SNPs, it is possible that the SNPs we detected are normally segregating polymorphic loci within species rather than divergent between species, and this is impossible to differentiate with our sample size. RADseq, as a reduced representation method does not capture the entire genome, so some loci and smaller regions that are significant to speciation may not have been captured in this study.

Conclusions

This study provides an example of genomic islands of divergence in recently diverged species and contributes to the existing knowledge of depth-driven adaptive evolution in *Sebastes* and other marine fishes. The genomic regions of divergence and the genes within them present a convincing argument for speciation as a result of depth separation, which may be applicable to a

broader range of marine organisms. These distinct islands represent in-depth insight into a recent speciation event, prior to genome-wide divergence rendering them undetectable over time. While it is not possible to completely elucidate the cause of divergence islands in these species, gene flow noted between *Sca-Sch* supports the hypothesis modeled by Quilodrán et al. (2020) that low gene flow between diverging populations may extend how long divergence islands persist. This provides insight for other studies examining similar species, where RI is not complete. *Smi-Scr* and *Sal-Sme* provide context for the potential future landscape of the *Sch-Sca* genomes as the species become more diverged. Within the islands, we identified genes involved in immune function and vision featuring significant SNPs. This supports their potential role in speciation, as these are adaptive vectors noted in other organisms. Additionally, changes to genes involved in pigment expression and mate recognition shed light on how *S. chrysomelas* and *S. carnatus* may have become reproductively isolated. Whole genome re-sequencing on additional *S. chrysomelas* and *S. carnatus* individuals would help validate that SNPs were not within-species polymorphisms. RNAseq gene expression studies and functional molecular studies such as CRISPR modification of the genes containing interesting SNPs would provide insight into the role of these potentially adaptive genes and those whose function is unknown. In summary, the presence of divergence islands and the genes they contain provide insight into the speciation process of organisms.

Acknowledgements

This project was supported in part by a grant to Juniata College from the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program, and in part by NSF Award # DBI-1248096. We are grateful to Christine Walls for NCBI data uploads, as well as assistance with software and hardware. We thank the reviewers for improving this manuscript.

Literature Cited

Aricescu, A. R., & Jones, E. Y. (2007). Immunoglobulin superfamily cell adhesion molecules: zippers and signals. In *Current Opinion in Cell Biology. 19*(5), 543-550. doi:10.1016/j.ceb.2007.09.010

Audard, V., Pawlak, A., Candelier, M., Lang, P., & Sahali, D. (2012). Upregulation of nuclear factor-related kappa b suggests a disorder of transcriptional regulation in minimal change nephrotic syndrome. *PLoS ONE*, *7*(1), 1–9. https://doi.org/10.1371/journal.pone.0030523

Braverman, N. E., D'Agostino, M. D., & MacLean, G. E. (2013). Peroxisome biogenesis disorders: Biological,

clinical and pathophysiological perspectives. *Developmental Disabilities Research Reviews*. https://doi.org/10.1002/ddrr.1113

Burford, M. O., & Bernardi, G. (2008). Incipient speciation within a subgenus of rockfish (Sebastosomus) provides evidence of recent radiations within an ancient species flock. *Marine Biology*, *154*(4), 701–717. https://doi.org/10.1007/s00227-008-0963-6

Buonaccorsi, V. P., Narum, S. R., Karkoska, K. A., Gregory, S., Deptola, T., & Weimer, A. B. (2011). Characterization of a genomic divergence island between black-and-yellow and gopher Sebastes rockfishes. *Molecular Ecology*, *20*(12), 2603–2618. https://doi.org/10.1111/j.1365-294X.2011.05119.x

Burckhardt, G. (2012). Drug transport by Organic Anion Transporters (OATs). *Pharmacology and Therapeutics*, 136(1), 106–130. https://doi.org/10.1016/j.pharmthera.2012.07.010

Carleton, K. L., & Kocher, T. D. (2001). Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Molecular Biology and Evolution*, *18*(8), 1540–1550. https://doi.org/10.1093/oxfordjournals.molbev.a003940

Carleton, K. L., & Yourick, M. R. (2020). Axes of visual adaptation in the ecologically diverse family Cichlidae. *Seminars in Cell & Developmental Biology*, *106*, 43-52. https://doi.org/10.1016/j.semcdb.2020.04.015.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Childress, J. J. (1995). Are there physiological and biochemical adaptations of metabolism in deep-sea animals? *Trends in Ecology & Evolution*, *10*(1), 30–36. https://doi.org/10.1016/S0169-5347(00)88957-0

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*. https://doi.org/10.4161/fly.19695

Cohen, R. E., James, C. C., Lee, A., Martinelli, M. M., Muraoka, W. T., Ortega, M., Sadowski, R., Starkey, L., Szesciorka, A. R., Timko, S. E., Weiss, E. L., & Franks, P. J. S. (2018). Marine host-pathogen dynamics influences of global climate change. *Oceanography*. https://doi.org/10.5670/oceanog.2018.201

Conaway, R. C., & Conaway, J. W. (2009). The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends in Biochemical Sciences*, *34*(2), 71–77. doi:10.1016/j.tibs.2008.10.010

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*. https://doi.org/10.1111/mec.12796

Dalton, B. E., Lu, J., Leips, J., Cronin, T. W., & Carleton, K. L. (2015). Variable light environments induce plastic spectral tuning by regional opsin coexpression in the African cichlid fish, Metriaclima zebra. *Molecular Ecology*. https://doi.org/10.1111/mec.1331

Dubreuil, M. M., Morgens, D. W., Okumoto, K., Honsho, M., Contrepois, K., Lee-McMullen, B., Traber, G. M. A., Sood, R. S., Dixon, S. J., Snyder, M. P., Fujiki, Y., & Bassik, M. C. (2020). Systematic Identification of Regulators of Oxidative Stress Reveals Non-canonical Roles for Peroxisomal Import and the Pentose Phosphate Pathway. *Cell Reports*. https://doi.org/10.1016/j.celrep.2020.01.01

Elmer, K. R., Lehtonen, T. K., & Meyer, A. (2009). Color assortative mating contributes to sympatric divergence of

neotropical cichlid fish. *Evolution*. https://doi.org/10.1111/j.1558-5646.2009.00736.x

Eom, D. S., Inoue, S., Patterson, L. B., Gordon, T. N., Slingwine, R., Kondo, S., Watanabe, M., & Parichy, D. M. (2012). Melanophore Migration and Survival during Zebrafish Adult Pigment Stripe Development Require the Immunoglobulin Superfamily Adhesion Molecule Igsf11. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002899

Faria, R., Johannesson, K., & Stankowski, S. (2021). Speciation in marine environments: Diving under the surface. *Journal of Evolutionary Biology*, *34*(1), 4–15. https://doi.org/10.1111/jeb.13756

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with- gene-flow. *Trends in Genetics*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Fowler, B. L. S., & Buonaccorsi, V. P. (2016). Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. *Molecular Ecology*, *25*(10), 2165–2175. https://doi.org/10.1111/mec.13594

Goldstone, J. V., McArthur, A. G., Kubota, A., Zanette, J., Parente, T., Jönsson, M. E., Nelson, D. R., & Stegeman, J. J. (2010). Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. *BMC Genomics*. https://doi.org/10.1186/1471-2164-11-643

Han, F., Lamichhaney, S., Rosemary Grant, B., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*. https://doi.org/10.1101/gr.212522.116

Hecht, M., Bromberg, Y., & Rost, B. (2016). Better prediction of functional effects for sequence variants From VarI-SIG 2014: Identification and annotation of genetic variants in the context of structure, function and disease. *BMC Genomics*, 16(Suppl 8), 1–12.

Hess, J., Smith, J., Timoshevskaya, N., Baker, C., Caudill, C., Graves, D., Keefer, M., Kinziger, A., Moser, M., Porter, L., Silver, G., Whitlock, S., & Narum, S. (2020). Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey. *Molecular Ecology*, 29(20), 3841-3856. https://doi.org/10.22541/au.158584219.94023029

He, Y., Chang, Y., Bao, L., Yu, M., Li, R., Niu, J., Fan, G., Song, W., Seim, I., Qin, Y., Li, X., Liu, J., Kong, X., Peng, M., Sun, M., Wang, M., Qu, J., Wang, X., Liu, X., … Qi, J. (2019). A chromosome-level genome of black rockfish, Sebastes schlegelii, provides insights into the evolution of live birth. *Molecular Ecology Resources*, *19*(5), 1309–1321. https://doi.org/10.1111/1755-0998.13034

Helvey, M. (1982). First Observations of Courtship Behavior in Rockfish, Genus Sebastes. *Copeia*, 4, 763–770.

Heras, J., & Aguilar, A. (2019). Comparative transcriptomics reveals patterns of adaptive evolution associated with depth and age within marine rockfishes (sebastes). *Journal of Heredity*, *110*(3), 340–350. https://doi.org/10.1093/jhered/esy070

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2).

https://doi.org/10.1371/journal.pgen.1000862

Hotta, A., Kawakatsu, T., Nakatani, T., Sato, T., Matsui, C., Sukezane, T., Akagi, T., Hamaji, T., Grigoriev, I., Akhmanova, A., Takai, Y., & Mimori-Kiyosue, Y. (2010). Laminin-based cell adhesion anchors microtubule plus ends to the epithelial cell basal cortex through LL5α/β. *Journal of Cell Biology*. https://doi.org/10.1083/jcb.200910095

Houbrechts, A. M., Vergauwen, L., Bagci, E., Van houcke, J., Heijlen, M., Kulemeka, B., Hyde, D. R., Knapen, D., & Darras, V. M. (2016). Deiodinase knockdown affects zebrafish eye development at the level of gene expression, morphology and function. *Molecular and Cellular Endocrinology*. https://doi.org/10.1016/j.mce.2016.01.018

Hyde, J. R., Kimbrell, C. A., Budrick, J. E., Lynn, E. A., & Vetter, R. D. (2008). Cryptic speciation in the vermilion rockfish (Sebastes miniatus) and the role of bathymetry in the speciation process. *Molecular Ecology*, 17(4), 1122–1136. https://doi.org/10.1111/j.1365-294X.2007.03653.x

Hyde, J. R., & Vetter, R. D. (2007). The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier). *Molecular Phylogenetics and Evolution*, *44*(2), 790–811. https://doi.org/10.1016/j.ympev.2006.12.026

Ingram, T. (2011). Speciation along a depth gradient in a marine adaptive radiation. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2010.1127

Kim, K. W., Jackson, B. C., Zhang, H., Toews, D. P. L., Taylor, S. A., Greig, E. I., Lovette, I. J., Liu, M. M., Davison, A., Griffith, S. C., Zeng, K., & Burke, T. (2019). Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nature Communications*, 10(1). https://doi.org/10.1038/s41467-019-09806-6

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., & Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. https://doi.org/10.1101/gr.129684.111

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5). https://doi.org/1179–1191. 10.1111/1755-0998.12387

Langmead, B., & Salzberg, S. (2013). Bowtie2. *Nature Methods*.

Larson, R. J. (1980). Competition, Habitat Selection, and the Bathymetric Segregation of Two Rockfish (Sebastes) Species. *Ecological Monographs*. https://doi.org/10.2307/1942480

Le Morvan, C., Troutaud, D., & Deschaux, P. (1998). Differential effects of temperature on specific and nonspecific immune defences in fish. In *Journal of Experimental Biology*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Wei, Y., Van Winkle, L., Zhang, Q. Y., Zhou, X., Hu, J., Xie, F., Kluetzman, K., & Ding, X. (2011). Generation and characterization of a Cyp2f2-null mouse and studies on the role of CYP2F2 in naphthalene-

induced toxicity in the lung and nasal olfactory mucosa. *Journal of Pharmacology and Experimental Therapeutics*. https://doi.org/10.1124/jpet.111.184671

Ljubojević, M., Herak-Kramberger, C. M., Hagos, Y., Bahn, A., Endou, H., Burckhardt, G., & Sabolić, I. (2004). Rat renal cortical OAT1 and OAT3 exhibit gender differences determined by both androgen stimulation and estrogen inhibition. *American Journal of Physiology - Renal Physiology*, 287(1 56-1), F124–F138. https://doi.org/10.1152/ajprenal.00029.2004

Love MS, Yoklavich M, Thorsteinson L (2002) The Rockfishes of the Northeast Pacific. University of California Press, Berkeley, California.

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R. J., Lascoux, M., Nevo, E., & Liu, J. (2017). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1713288114

Manly, B. F. (2006). *Randomization, bootstrap and Monte Carlo methods in biology* (Vol. 70). CRC press.

Marincovich, L., Gladenkov, A.Y. (2001). New evidence for the age of Bering Strait. *Quaternary Science Reviews*. 20(1-3), 329-335. https://doi.org/10.1016/S0277-3791(00)00113-X.

Narum, S. R., Buonaccorsi, V. P., Kimbrell, C. A., & Vetter, R. D. (2004). Genetic divergence between gopher rockfish (Sebastes carnatus) and black and yellow rockfish (Sebastes chrysomelas). Copeia, 2004(4), 926–931. https://doi.org/10.1643/CG-02-061R2

Nei, M. (1987). Molecular Evolutionary Genetics. Columbia University Press. https://doi.org/https://doi.org/10.7312/nei-92038

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America, 76(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nigam, S. K. (2018). The SLC22 Transporter Family: A Paradigm for the Impact of Drug Transporters on Metabolic Pathways, Signaling, and Disease. *Annual Review of Pharmacology and Toxicology*, 58(Figure 1), 663–687. https://doi.org/10.1146/annurev-pharmtox-010617-052713

Nigam, S. K., Bush, K. T., Martovetsky, G., Ahn, S. Y., Liu, H. C., Richard, E., Bhatnagar, V., & Wu, W. (2015). The organic anion transporter (OAT) family: A systems biology perspective. *Physiological Reviews*, 95(1), 83–123. https://doi.org/10.1152/physrev.00025.2013

Nishi, R., Wijnhoven, P., Sage, C., Tjeertes, J., & Galanty, Y. (2015). Europe PMC Funders Group Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nature Cell Biology*, *16*(10), 1016–1026. https://doi.org/10.1038/ncb3028.Systematic

Normand, S., Waldschmitt, N., Neerincx, A., Martinez-Torres, R. J., Chauvin, C., Couturier-Maillard, A., Boulard, O., Cobret, L., Awad, F., Huot, L., Ribeiro-Ribeiro, A., Lautz, K., Ruez, R., Delacre, M., Bondu, C., Guilliams, M., Scott, C., Segal, A., Amselem, S., … Chamaillard, M. (2018). Proteasomal degradation of NOD2 by NLRP12 in monocytes promotes bacterial tolerance and colonization by enteropathogens. *Nature Communications*. https://doi.org/10.1038/s41467-018-07750-5

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. In *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2008.10.011

Orr, J. W., & Blackburn, J. E. (2004). The dusky rockfishes (Teleostei: Socrpaeniformes) of the North Pacific Ocean: resurrection of Sebastes variabilis (Pallas, 1814) and a redescription of Sebastes ciliatus (Tilesius, 1813). *Fishery Bulletin*, *102*(2), 328-348.

Orr, J. W., Brown, M. A., & Baker, D. C. (2000). Guide to rockfishes (Scorpaenidae) of the genera Sebastes, Sebastolobus, and Adelosebastes of the Northeast Pacific Ocean. *NOAA Tech. Mem*. NMFS-AFSC-117, Seattle, WA.

Okumoto, K., Shimozawa, N., Kawai, A., Tamura, S., Tsukamoto, T., Osumi, T., Moser, H., Wanders, R. J. A., Suzuki, Y., Kondo, N., & Fujiki, Y. (1998). PEX12, the Pathogenic Gene of Group III Zellweger Syndrome: cDNA Cloning by Functional Complementation on a CHO Cell Mutant, Patient Analysis, and Characterization of Pex12p. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.18.7.4324

Paranavitane, V., John Coadwell, W., Eguinoa, A., Hawkins, P. T., & Stephens, L. (2003). LL5β is a phosphatidylinositol (3,4,5)-trisphosphate sensor that can bind the cytoskeletal adaptor, γ-filamin. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M208352200

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Posner, M., McDonald, M. S., Murray, K. L., & Kiss, A. J. (2019). Why does the zebrafish cloche mutant develop lens cataract? *PLoS ONE*. https://doi.org/10.1371/journal.pone.0211399

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. https://doi.org/10.1093/genetics/155.2.945

Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, *11*(1), 6–21. https://doi.org/10.1111/2041-210X.13324

R Core Development Team. (2019). R: A language and environment for statistical computing. *Vienna, Austria*.

Rausch, P., Steck, N., Suwandi, A., Seidel, J. A., Künzel, S., Bhullar, K., Basic, M., Bleich, A., Johnsen, J. M., Vallance, B. A., Baines, J. F., & Grassl, G. A. (2015). Expression of the Blood-Group-Related Gene B4galnt2 Alters Susceptibility to Salmonella Infection. PLoS pathogens, 11(7), e1005008. https://doi.org/10.1371/journal.ppat.1005008

Ravinet, M., Kume, M., Ishikawa, A., & Kitano, J. (2021). Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *Journal of Evolutionary Biology*, *34*(1), 114–127. https://doi.org/10.1111/jeb.13664

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., & Kitano, J. (2018). The genomic landscape at

742 a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of
743 introgression. In *PLoS Genetics* (Vol. 14, Issue 5). https://doi.org/10.1371/journal.pgen.1007358
744 Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding Light on the Grey
745 Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biology*, *14*(12), 1–22.
746 https://doi.org/10.1371/journal.pbio.2000234
747 Schrader, M., & Fahimi, H. D. (2006). Peroxisomes and oxidative stress. In *Biochimica et Biophysica Acta -*
748 *Molecular Cell Research*. https://doi.org/10.1016/j.bbamcr.2006.09.006
749 Schreiber, D., & Pfenninger, M. (2021). Genomic divergence landscape in recurrently hybridizing Chironomus
750 sister taxa suggests stable steady state between mutual gene flow and isolation. *Evolution Letters*, *5*(1), 86–
751 100. https://doi.org/10.1002/evl3.204
752 Shaw, G. (1996). The pleckstrin homology domain: An intriguing multifunctional protein module. *BioEssays*.
753 https://doi.org/10.1002/bies.950180109
754 Shen, S., Li, L., Ding, X., & Zheng, J. (2014). Metabolism of styrene to styrene oxide and vinylphenols in
755 cytochrome P450 2F2- and P450 2E1-knockout mouse liver and lung microsomes. *Chemical Research in*
756 *Toxicology*. https://doi.org/10.1021/tx400305w
757 Shinomiya, A., & Ezaki, O. (1991). Mating habits of the rockfish Sebastes inermis. *Environmental Biology of*
758 *Fishes*, 30(1–2), 15–22. https://doi.org/10.1007/BF02296872
759 Shum, P., Pampoulie, C., Sacchi, C., & Mariani, S. (2014). Divergence by depth in an oceanic fish. *PeerJ*, *2014*(1),
760 1–13. https://doi.org/10.7717/peerj.525
761 Sivasundar, A., & Palumbi, S. R. (2010). Parallel amino acid replacements in the rhodopsins of the rockfishes
762 (Sebastes spp.) associated with shifts in habitat depth. *Journal of Evolutionary Biology*, *23*(6), 1159–1169.
763 https://doi.org/10.1111/j.1420-9101.2010.01977.x
764 Smyth, T. J. (2011). Penetration of UV irradiance into the global ocean. *Journal of Geophysical Research: Oceans*.
765 https://doi.org/10.1029/2011JC007183
766 Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, 1–18.
767 https://doi.org/10.1111/evo.14215
768 Stefánsson, M. Ö., Reinert, J., Sigurosson, O., Kristinsson, K., Nedreaas, K., & Pampoulie, C. (2009). Depth as a
769 potential driver of genetic structure of sebastes mentella across the north Atlantic ocean. *ICES Journal of*
770 *Marine Science*, *66*(4), 680–690. https://doi.org/10.1093/icesjms/fsp059
771 Turner, S.D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of*
772 *Open Source Software*. https://doi.org/10.1101/005165
773 Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS*
774 *Biology,* 3(9), 1572–1578. https://doi.org/10.1371/journal.pbio.0030285
775 Vallier, M., Abou Chakra, M., Hindersin, L., Linnenbrink, M., Traulsen, A., & Baines, J. F. (2017). Evaluating the
776 maintenance of disease-associated variation at the blood group-related gene B4galnt2 in house mice. *BMC*
777 *Evolutionary Biology.* https://doi.org/10.1186/s12862-017-1035-7
778 Via, S. (2009). Natural selection in action during speciation. In *Proceedings of the National Academy of Sciences of*

the United States of America. https://doi.org/10.1073/pnas.0901397106

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2011.0260

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*(19), 4334–4345. https://doi.org/10.1111/j.1365-294X.2008.03921.x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*. https://doi.org/10.21105/joss.01686

Williams, A. F., & Barclay, A. N. (1988). The immunoglobulin superfamily - Domains for cell surface recognition. In *Annual Review of Immunology*. https://doi.org/0.1146/annurev.iy.06.040188.002121

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

Wu, W., Dnyanmote, A. V., & Nigam, S. K. (2011). Remote communication through solute carriers and ATP binding cassette drug transporter pathways: An update on the Remote Sensing and Signaling Hypothesis. *Molecular Pharmacology*, 79(5), 795–805. https://doi.org/10.1124/mol.110.070607

Zhang, J., Yao, J., Wang, R., Zhang, Y., Liu, S., Sun, L., Jiang, Y., Feng, J., Liu, N., Nelson, D., Waldbieser, G., & Liu, Z. (2014). The cytochrome P450 genes of channel catfish: Their involvement in disease defense responses as revealed by meta-analysis of RNA-Seq data sets. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1840(9), 2813-2828. https://doi.org/10.1016/j.bbagen.2014.04.016

Data Accessibility

The raw sequences are available in the NCBI Short Read Archive (NCBI project PRJNA307574) and STACKS output files are available in Dryad (doi:10.5061/dryad.c866t1g6z).

Author Contributions

J.H. collected samples, V.P.B. and A.S. designed the research, K.A.B., V.P.B., and Q.L.G performed analyses, and all authors contributed towards interpreting results and writing the manuscript.

Tables

Table 1. Summary of mean F_{ST}, D_{XY}, variable sites, and unique loci for species pairs based on *Stacks* generated population statistics. HWE and pool bias filtering (p-value < 0.001), means are for HWE/pool bias filtered data, standard error in parentheses.

	chrysomelas/carnatus	miniatus/crocotulus	mentella/alutus
Total variable sites prior to filtering	62569	47500	105195
Total unique loci prior to filtering	40153	32052	54712
Variable sites where $F_{ST} > 0.75$	49	10	43835
Variable sites in HWE	61309	46929	104748
Unique loci in HWE	39630	31809	54601
Variable sites in HWE where $F_{ST} > 0.75$	49	10	43695
Variable sites in HWE where $F_{ST} > 0.5$	94	10	104116
Mean F_{STv}	0.0248 (0.0002)	0.2746 (0.0004)	0.7321 (0.0003)
Mean D_{XY}	0.00641 ($6.6x10^{-6}$)	0.0080 ($8.6x10^{-6}$)	0.01691 ($1.1x10^{-5}$)

Table 2. Analysis of genomic divergence islands within S. carnatus and S. chrysomelas defined using ZFST methodology. Columns include species or species comparison (Sca = *S. carnatus*, Sch = *S. chrysomelas*, Smi = *S. miniatus*, Scr = *S. crocotulus*), chromosome (Chr), starting and ending coordinates, statistic being compared (Pi = nucleotide sequence diversity), mean value of the statistic for the peak or background levels (or for species for pi comparisons), p-value, and sample sizes for comparisons.

Spp	Chr	Start	End	Statistic	Mean: peak	Mean: background	P-value	N:peak	N:background
Sch vs Sca	3	3.90E+07	39500000	D_{XY}	0.0071	0.0069	0.4118	8	1612
Sch vs Sca	4	18500000	2.00E+07	D_{XY}	0.0097	0.0072	**0.0079**	24	1409
Sch vs Sca	6	3.60E+07	36500000	D_{XY}	0.0059	0.0067	0.5176	5	1843
Sch vs Sca	8	3500000	4.00E+06	D_{XY}	0.0060	0.0067	0.5307	8	1812
Sch vs Sca	9	10500000	11000000	D_{XY}	0.0109	0.0065	**0.0047**	8	1602
Sch vs Sca	13	3.00E+07	3.25E+07	D_{XY}	0.0087	0.0068	**0.0134**	34	1426
Sch vs Sca	15	1500000	2.00E+06	D_{XY}	0.0098	0.0067	0.0541	6	1598
Smi vs Scr	4	18500000	20000000	D_{XY}	0.0090	0.0085	0.3052	36	1222
Smi vs Scr	13	30000000	32500000	D_{XY}	0.0090	0.0082	0.2726	17	1401
Sca	4	18500000	2.00E+07	Pi	0.1358	0.2494	0.0013	26	2335
Sch	4	18500000	2.00E+07	Pi	0.0724	0.2530	0.0001	26	2335
Sca	9	10500000	11000000	Pi	0.1745	0.2468	0.0905	11	2493
Sch	9	10500000	11000000	Pi	0.0610	0.2540	0.0042	11	2493
Sca	13	3.00E+07	3.25E+07	Pi	**0.3277**	0.2495	0.0017	46	2305
Sch	13	3.00E+07	3.25E+07	Pi	0.0791	0.2540	0.0000	46	2305
Smi	4	18500000	20000000	Pi	0.0752	0.2312	0.0010	23	1938
Scr	4	18500000	20000000	Pi	0.1112	0.1848	0.0557	23	1938
Smi	13	30000000	32500000	Pi	0.1501	0.2154	0.0874	20	1766

Scr	13	30000000	32500000	Pi	0.0752	0.1768	0.0288	20	1766
					Mean: sca	**Mean: sch**		**N: sca**	**N: sch**
Sca vs Sch: Peak	4	18500000	2.00E+07	Pi	0.1358	0.0724	0.1213	26	26
Sca vs Sch: Background	4	1	3.90E+07	Pi	0.2494	0.2530	0.1846	2335	2335
Sch vs Sca: Peak	9	10500000	11000000	Pi	0.1745	0.0610	0.0058	11	11
Sch vs Sca: Background	9	1	35190670	Pi	0.2468	0.2540	0.0081	2493	2493
Sca vs Sch: Peak	13	3.00E+07	3.25E+07	Pi	0.3277	0.0791	0.0000	46	46
Sca vs Sch: Background	13	1.00E+00	3.23E+07	Pi	0.2495	0.2540	0.0830	2305	2305
					Mean: smi	**Mean: scr**	**P-value**	**N:smi**	**N: scr**
Smi vs Scr: Peak	4	18500000	20000000	Pi	0.0752	0.1112	0.4215	23	23
Smi vs Scr: Background	4	1	38958735	Pi	0.2312	0.1848	0.0464	1938	1938
Smi vs Scr: Peak	13	30000000	32500000	Pi	0.1501	0.0752	0.1574	20	20
Smi vs Scr: Background	13	1	3.23E+07	Pi	0.2154	0.1768	0.0000	1766	1766

852 Table 3. Peak variant summary information for *Sch-Sca*. Entries in bold represent individual SNPs or combinations
853 most likely to have an effect.

Gene symbol	Gene name	BlastHit	Scaffold	BP	CDS change	AA change	SNP Eff Type	SNP Effect Prediction	SNAP Effect, Score, Exp Accuracy
Ssc_10012318	**MAP7 domain-containing protein 2a (MAP7D2A)**	**XP_037604095.1**	**4**	**18606476**	**c.2455C>T**	**p.Arg819Cys**	**Missense_variant**	**Moderate**	**effect, 29, 63%**
Ssc_10012325	**pleckstrin homology-like domain family B member**	**XP_037605714.1**	**4**	**18683830**	**c.1801G>A**	**p.Val601Ile**	**Missense_variant**	**Moderate**	**neutral, -44, 72%**

Ssc_10012325	pleckstrin homology-like domain family B member 2	XP_037605714.1	4	18686254	c.1507A>G	p.Lys503Glu	Missense_variant	Moderate	neutral, -3, 53%
Ssc_10012325	pleckstrin homology-like domain family B member 2	XP_037605714.1	4	18686304	c.1480C>A	p.Pro494Thr	Missense_variant	Moderate	neutral, -29, 61%
Ssc_10012325	pleckstrin homology-like domain family B member 2	XP_037605714.1	4	18698151	c.13T>C	p.Phe5Leu	Missense_variant	Moderate	neutral, -29, 61%
Ssc_10012333	Ryanodine receptor 1	KAF6728934.1	4	18845851	c.505G>A	p.Glu169Lys	Missense_variant	Moderate	neutral, -12, 57%
Ssc_10012334	Ryanodine receptor 2	XP_037610012.1	4	18898683	c.8183T>C	p.Val2728Ala	Missense_variant	Moderate	NC†
Ssc_10012335	**nuclear factor related to kappa-B-binding protein NFRKB**	XP_037605694.1	**4**	**18945673**	**c.3803C>T**	**p.Pro1268Leu**	**Missense_variant**	**Moderate**	**effect,55, 75%**
Ssc_10012345	solute carrier family 22 member 6	XP_037604807.1	4	19264725	c.889A>G	p.Ile297Val	Missense_variant	Moderate	neutral, -97, 97%
Ssc_10012349	neurexin-2-like	XP_037604828.1	4	19366254	c.616G>A	p.Gly206Ser	Missense_variant	Moderate	neutral, -70, 82%
Ssc_10012351	pecanex-like protein 3	XP_037604181.1	4	19507073	c.1199T>C	p.Ile400Thr	Missense_variant	Moderate	NC†
Ssc_10020815	immunoglobulin superfamily member 11	XP_037603604.1	4	19645230	c.577G>A	p.Glu193Lys	Missense_variant	Moderate	neutral, -75, 87%
Ssc_10020815	**immunoglobulin superfamily member 11**	XP_037603604.1	**4**	**19648635**	**c.1A>G**	**p.Met1?**	**Start_lost**	**High**	**N/A**
Ssc_10020814	dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	XP_037604519.1	4	19663337	c.34G>T	p.Ala12Ser	Missense_variant	Moderate	neutral, -90, 93%
Ssc_10020812	**cilia- and flagella-associated protein 54-like**	**XP_037604654.1**	**4**	**19720288**	**c.5957G>A**	**p.Ser1986Asn**	**Missense_variant**	**Moderate**	neutral, -91, 97%
Ssc_10020812	cilia- and flagella-associated protein 54-like	XP_037604654.1	4	19733796	c.3328C>T	p.Leu1110Phe	Missense_variant	Moderate	neutral, -46, 72%
Ssc_10020812	cilia- and flagella-associated protein 54-like	XP_037604654.1	4	19734571	c.2998G>A	p.Val1000Ile	Missense_variant	Moderate	neutral, -89, 93%

Ssc_1002 0812	cilia- and flagella-associated protein 54-like	XP_037604 654.1	4	19744 392	c.193A>G	p.Lys65Glu	Missense_variant	Moder ate	neutral, -82, 93%
Ssc_1002 0803	**reverse transcriptase**	**AGO18322 .1**	**4**	**19817 040**	**c.958G>A**	**p.Asp320Asn**	**Missense_variant**	**Moder ate**	**effect, 42, 71%**
Ssc_1002 0803	reverse transcriptase	AGO18322. 1	4	19817 058	c.940C>T	p.Arg314*	Stop_gained	High	N/A
Ssc_1002 0803	reverse transcriptase	AGO18322. 1	4	19818 333	c.436C>T	p.Leu146Phe	Missense_variant	Moder ate	neutral, -42,72%
Ssc_1002 0801	V-set and immunoglobulin domain-containing protein 10-like 2	XP_037605 109.1	4	19836 828	c.1139C>T	p.Thr380Ile	Missense_variant	Moder ate	neutral, -12, 57%
Ssc_1002 0800	**uncharacterized protein, ID'd as "HLA" in one species**		**4**	**19868 037**	**c.479A>G**	**p.Asp160Gly**	**Missense_variant**	**Moder ate**	**effect, 17, 59%**
Ssc_1002 0798	inositol polyphosphate 5-phosphatase K	XP_037603 664.1	4	19912 111	c.680G>A	p.Ser227Asn	Missense_variant	Moder ate	neutral, -83, 93%
Ssc_1001 5597	**cytochrome P450 2F2-like**	XP_037612 192.1	**9**	**10810 706**	**c.1411delT**	**p.Tyr471fs**	**Frameshift_variant**	**High**	**N/A**
Ssc_1000 7501	**ornithine decarboxylase-like**	XP_037649 018.1	13	30148 084	c.114G>T	p.Glu38Asp	Missense_variant	Moder ate	neutral, -84, 93%
Ssc_1000 7501	ornithine decarboxylase-like	XP_037649 018.1	13	30148 701	c.538C>G	p.Leu180Val	Missense_variant	Moder ate	neutral, -29, 61%
Ssc_1000 7506	**peroxisome assembly protein 12 (PEX12)**	**XP_037649 027.1**	**13**	**30391 708**	**c.856_861delCA GCAG**	**p.Gln286_Gln 287del**	**Conservative_infram e_deletion**	**Moder ate**	**N/A**
Ssc_1000 7506	**peroxisome assembly protein 12**	**XP_037649 027.1**	**13**	**30391 894**	**c.676G>A**	**p.Ala226Thr**	**Missense_variant**	**Moder ate**	**effect, 8, 53%**
Ssc_1000 7506	peroxisome assembly protein 12	XP_037649 027.1	13	30393 092	c.601G>A	p.Val201Ile	Missense_variant	Moder ate	neutral, -80, 87%
Ssc_1000 7506	**peroxisome assembly protein 12**	**XP_037649 027.1**	**13**	**30393 186**	**c.507G>C**	**p.Trp169Cys**	**Missense_variant**	**Moder ate**	**effect, 19, 59%**
Ssc_1000 7506	peroxisome assembly protein 12	XP_037649 027.1	13	30394 929	c.64G>A	p.Val22Ile	Missense_variant	Moder ate	neutral, -23, 61%
Ssc_1000 7508	collagen, type I, alpha 1a	XP_037648 990.1	13	30476 101	c.1976A>G	p.Asn659Ser	Missense_variant	Moder ate	neutral, -59, 78%
Ssc_1000 7511	**beta-1,4 N-acetylgalactosaminylt ransferase 2-like**	**XP_037649 015.1**	**13**	**30608 673**	**c.1541G>A**	**p.Gly514Glu**	**Missense_variant**	**Moder ate**	**effect, 31, 66%**
Ssc_1000 7512	**NACHT, LRR and PYD domains-containing protein 12-**	XP_037647 125.1	13	30667 896	c.1729G>A	p.Glu577Lys	Missense_variant	Moder ate	neutral, -51, 78%

	like								
Ssc_10007512	**NACHT, LRR and PYD domains-containing protein 12-like**	**XP_037647125.1**	**13**	**30667905**	**c.1720G>A**	**p.Asp574Asn**	**Missense_variant**	**Moderate**	**effect, 66, 80%**
Ssc_10007512	NACHT, LRR and PYD domains-containing protein 12-like	XP_037647125.1	13	30672446	c.427G>A	p.Val143Ile	Missense_variant	Moderate	neutral, -37, 66%
Ssc_10007512	NACHT, LRR and PYD domains-containing protein 12-like	XP_037647125.1	13	30672586	c.287G>A	p.Gly96Asp	Missense_variant	Moderate	neutral, -49, 72%
Ssc_10007513	alpha-1,3-galactosyltransferase 2-like	XP_037633122.1	13	30695942	c.161T>C	p.Leu54Pro	Missense_variant	Moderate	neutral, -38, 66%

† Not called, SNAP2 unable to process. The amino acid substitution in *pecanex-like protein 3* may be biochemically significant.

Table 4. Functional information for genes with individual SNPs or combinations most likely to have an effect.

Gene Name	Gene ID	Function
organic anion transporter	*OAT1 (SLC22a6)*	• regulates multiple metabolic and signaling pathways (Nigam et al., 2015) • expressed primarily in the basolateral membrane of proximal tubule cells in the kidney, facilitates anions crossing membranes (Burckhardt, 2012)
Nuclear factor related to kappa-B-binding protein	*NFRKB*	• part of the chromatin remodeling complex INO80, involved in transcriptional regulation, DNA replication, and DNA repair (Conaway & Conaway, 2009, Audard et al., 2012; Nishi, 2015)
peroxisome assembly protein 12	*PEX12*	• intermediary lipid metabolism and oxidize long chain fatty acids providing an important energy source (Braverman et al., 2013) • produce and break down reactive oxygen species such as H_2O_2, which can serve as a signaling molecule for autophagy or growth at low concentration (Schrader, 2006; reviewed in Dubreuil, 2020) • important in peroxisome biogenesis, involved in peroxisome matrix protein import (Okumoto et al.,

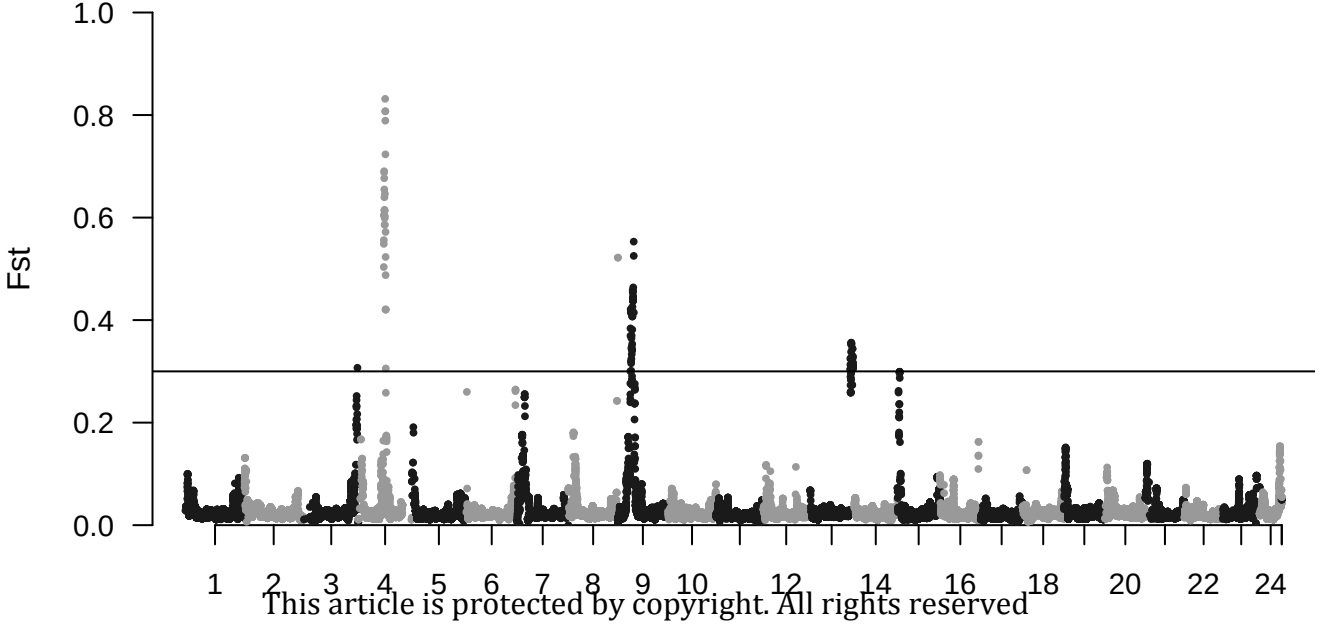
		1998)
ornithine decarboxylase-like	*odc1*	• eye development, specifically in the retina and in photoreceptor development (Posner et al. 2019, Houbrechts et al. 2016)
immunoglobulin superfamily member	*igsf11*	• domains located in the cell surface that are involved in cell surface recognition and cell adhesion (Aricescu & Jones, 2007; Williams & Barclay, 1988) • expressed in melanophores, adult pigment cells, essential to migration and cell adhesion (Eom et al., 2012)
pleckstrin homology-like domain family B member 2	*PHLDB2*	• found in enzymes with regulatory functions, pleckstrin homology domains bind proteins and inositolphosphates (Shaw, 1996) • cytoskeletal function and microtubule organization (Hotta et al., 2010; Paranavitane et al., 2003)
cytochrome P450 2F2-like	*CYP2F2*	• oxidative reactions of internal and external-origin chemicals (Goldstone et al., 2010; Zhang et al., 2014) • metabolism of styrene and naphthalene in the lungs to reduce toxicity (Li et al., 2011; Shen et al., 2014)
beta-1,4 N-acetylgalactosaminyltransferase 2-like	*B4galnt2*	• glycosyltransferase that is related to blood groups and expressed in the gastrointestinal tract, playing a role in host-microbe interactions (Vallier et al., 2017) • influence pathogen resistance, the microbiome in the intestine, and immune responses from the host organism (Rausch et al., 2015)
NACHT, LRR, and PYD domains-containing protein 12-like	*NLRP12*	• involved in multiple pathways, including inflammatory and anti-inflammatory functions (Normand et al., 2018) • inhibit inflammation and promote bacterial tolerance (Normand et al., 2018)

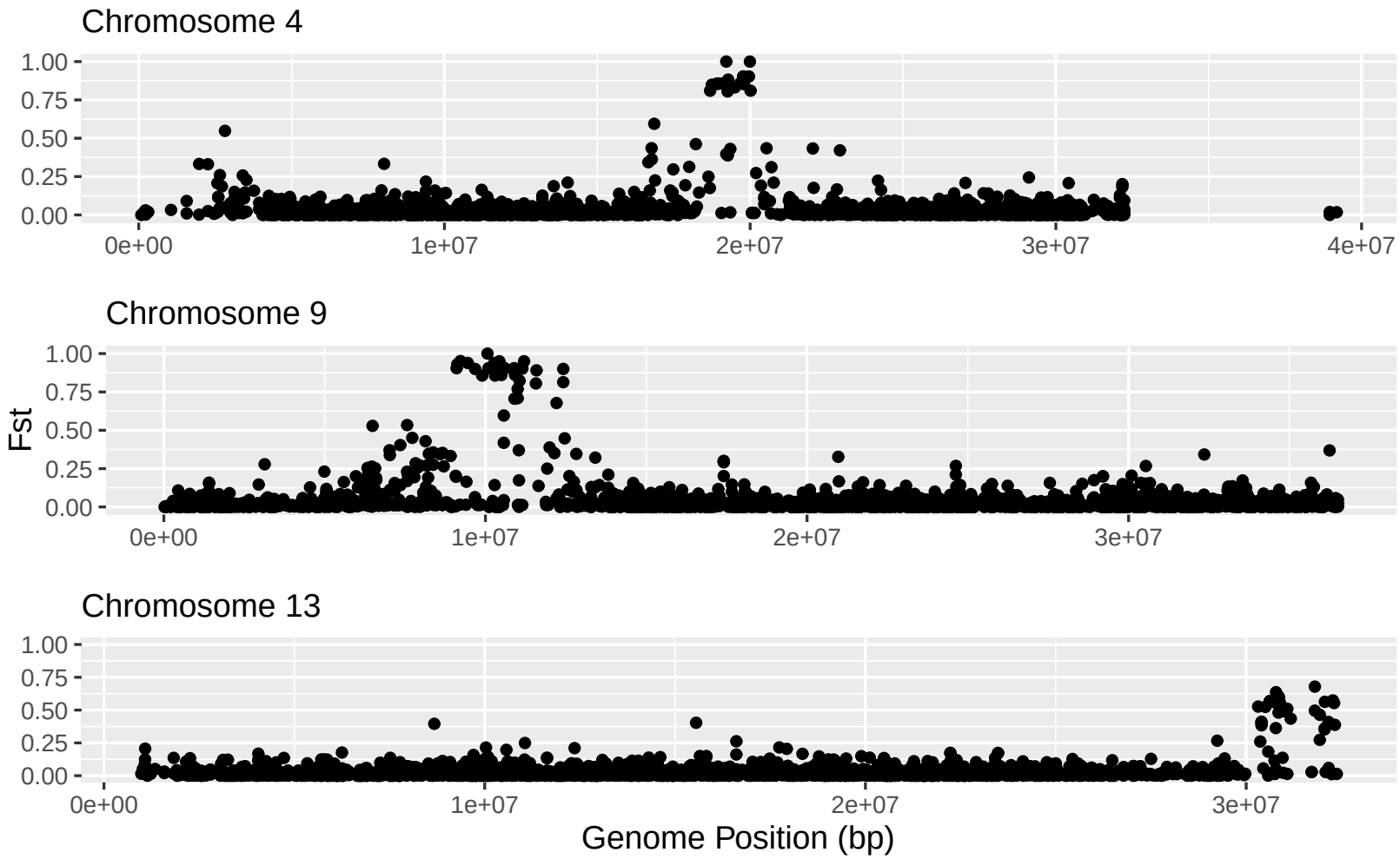
Figure Captions

Figure 1. Smoothed F_{ST} estimates for each variable site between *S. chrysomelas* vs. *S. carnatus*. A) Manhattan plot

862 of variant sites showing smoothed F_{ST} against chromosome position, alternating black and gray shading
863 distinguishes chromosomes. Line indicates cut-off for peaks. B) Detailed scatterplots of smoothed F_{ST} by base pair
864 position for variant sites on three chromosomes that had regions of elevated F_{ST}.
865
866 Figure 2. Smoothed F_{ST} estimates for each variable site between *S. miniatus* vs. *S. crocotulus*. A) Manhattan plot of
867 variant sites plotted by chromosome on the x-axis and smoothed F_{ST} on the y-axis, alternating black and gray
868 shading distinguishes chromosomes. B) Detailed scatterplots of smoothed F_{ST} by base pair position for variant sites
869 on two chromosomes that had regions of elevated F_{ST} corresponding with islands in the *S. chrysomelas* vs. *S.*
870 *carnatus* comparison
871
872 Figure 3. Smoothed F_{ST} estimates for each variable site between *S. alutus* vs. *S. mentella*. Manhattan plot of variant
873 sites plotted by chromosome on the x-axis and smoothed F_{ST} on the y-axis, alternating black and gray shading
874 distinguishes chromosomes.

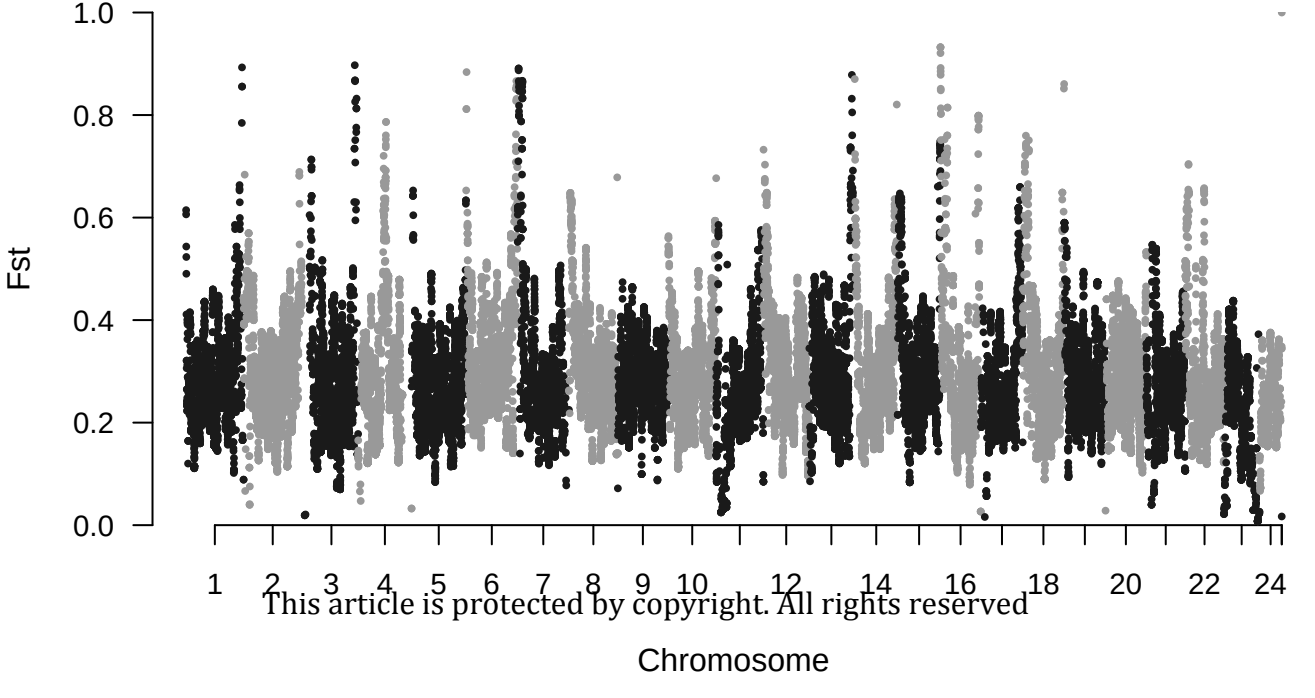
mec_16046_f1a.pdf

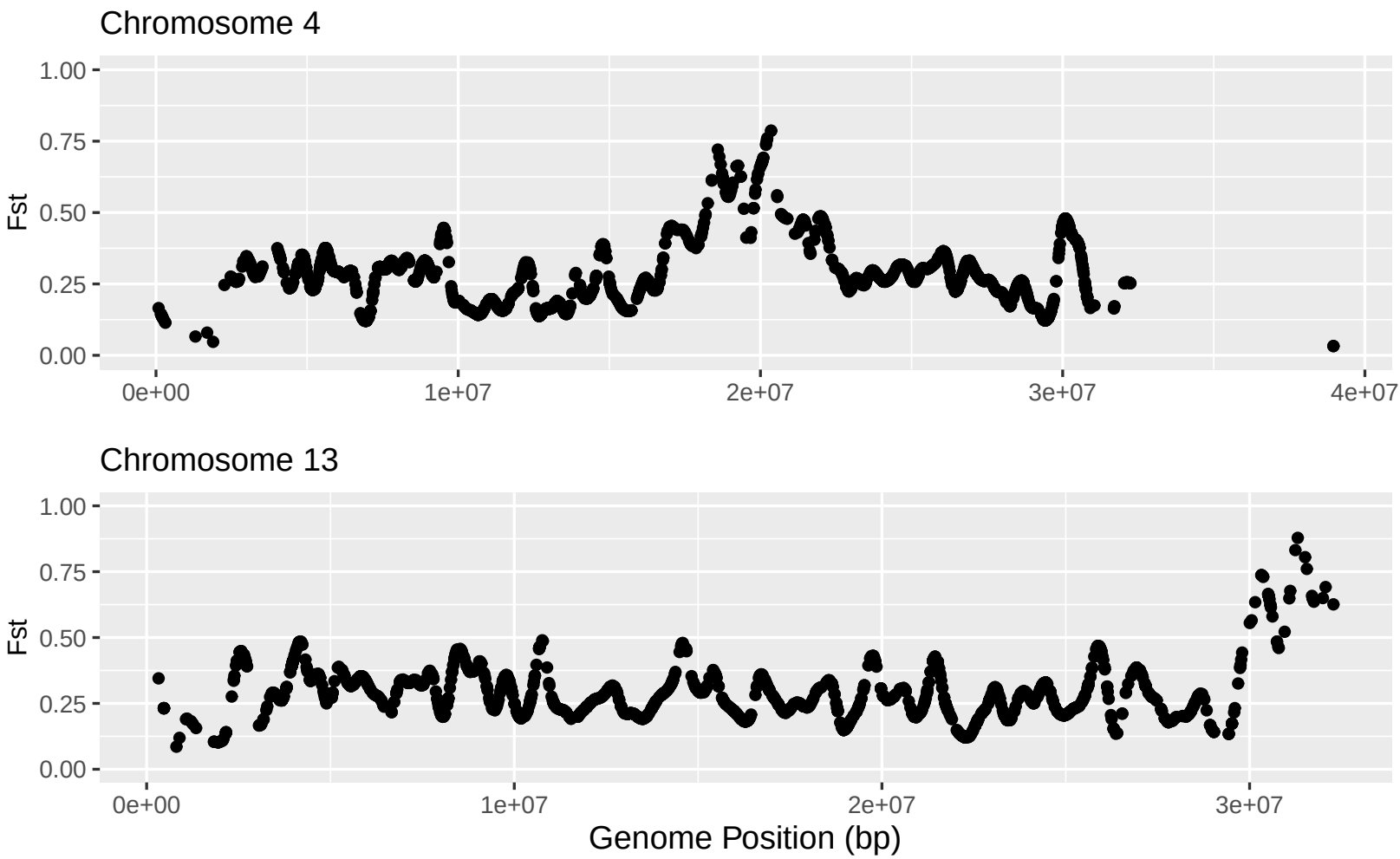




mec_16046_f1b.eps

mec_16046_f2a.pdf





mec_16046_f2b.eps

mec_16046_f3.pdf