GENOMIC SIGNATURES OF RECENT ADAPTIVE DIVERGENCE IN THE SWAMP SPARROW (*MELOSPIZA GEORGIANA*)

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Petra Elizabeth Deane

December 2017

GENOMIC SIGNATURES OF RECENT ADAPTIVE DIVERGENCE IN THE SWAMP SPARROW (*MELOSPIZA GEORGIANA*)

Petra Elizabeth Deane, Ph. D.

Cornell University 2017

Populations that have recently diverged across sharp environmental gradients provide an opportunity to study the mechanisms by which natural selection drives adaptive divergence. Inland and coastal populations of the North American swamp sparrow (*Melospiza georgiana*) have become an emerging model system for studies of natural selection because they are morphologically and behaviorally distinct despite a very recent divergence time (<15,000 years), yet common garden experiments have demonstrated a genetic basis for their differences. I characterized genomic patterns of variation within and between inland and coastal swamp sparrows via reduced representation sequencing and demonstrated that background genomic differentiation (F_{ST}=0.02) and divergence (Φ_{ST}=0.05) between these populations is very low, rendering signatures of natural selection highly detectable (max F_{ST}=0.8). I then sequenced and assembled a *de novo* reference genome for the species and conducted a scan for genes involved in coastal adaptation, particularly the evolution of a deeper bill, darker plumage, and tolerance for salinity. I recovered a multigenic snapshot of adaptation via robust signatures of selection at 31 genes. As in Darwin's finches, bone morphogenetic protein (BMP) signaling appears responsible for changes in bill depth, a putative magic trait for ecological speciation. Genes for salinity tolerance constituted the majority of candidates (23/31), including genes involved in regulating osmotic

balance via vasoconstriction and intracellular vesicle trafficking. I then quantified genotype-phenotype associations in a naturally occurring hybrid zone between inland and coastal swamp sparrows and demonstrated that melanism in coastal swamp sparrows is a product of both direct positive selection and molecular "spandrel" effects. Black plumage patches are the product of intrasexual selection on a melanin-specific transcription factor (BNC2), whereas black legs are a pleiotropic consequence of selection on a vesicle trafficking gene involved in salt tolerance (BLOC1S2). Pleiotropic effects of vesicle trafficking may therefore explain why other salt marsh birds, snakes and mammals have also evolved darker coloration in a phenomenon known as "salt marsh melanism". Natural selection has therefore driven genetic, ecological and phenotypic divergence in the swamp sparrow, and possibly other locally adapted salt marsh lineages, through a combination of direct and indirect molecular mechanisms.

BIOGRAPHICAL SKETCH

Prior to her PhD research on evolutionary genomics in sparrows with Drs. Richard Harrison and Irby Lovette in Ecology and Evolutionary Biology at Cornell, Petra Elizabeth Deane (a Canadian citizen) completed her MSc research on population genetics and speciation in seabirds in the lab of Dr. Vicki Friesen in Ecology and Evolutionary Biology at Queen's University in Kingston, Ontario. Her motivation to understand biological diversity at the molecular level stems from her background in natural history, including prior positions as a park ranger and natural history educator for the Ontario provincial park system and private conservation companies. Outside of academics, Petra is a committed equestrian athlete and is an advocate for the conservation of traditional Irish bloodlines in modern sport horses.

To Richard G. Harrison (November 19, 1945 – April 12, 2016),

who provided an incredible example of what is good and true in the world:

in wine, in gardening, in marriage, in politics, and in speciation research.

An irreplaceable touchstone, mentor, scientist and friend.

We lost you too soon. Miss you every day.

ACKNOWLEDGMENTS

Sincere thanks to my co-advisors Irby Lovette and Rick Harrison, and committee members Jeremy Searle and Nina Overgaard-Therkildsen, for engaging with this research, sharing their expertise and providing thoughtful feedback throughout. I'm also grateful to the following colleagues for specific conversations and discussions that improved one or more of the chapters contained herein: R. Fleischer, M. Sorenson, H. Kunerth, L. Campagna, S. Taylor, D. Toews, J. Walsh. The Smithsonian Conservation Biology Institute's Center for Conservation and Evolutionary Genetics (R. Fleischer), and Cornell University Museum of Vertebrates (C. Dardia, V. Rohwer) contributed samples collected by N. Perlut, N. Mason and G. Seeholzer, among others. Edwin B. Forsythe National Wildlife Refuge (P. Castelli, V. Turner, R. Albers, M. Sieges), Blackwater National Wildlife Refuge (M. Whitbeck) and Fishing Bay Wildlife Management Area (J. Moulis) provided logistical support, and A. Bessler, K. Bostwick, K. Deane-Coe, S. Aguillon, J. Swonger and M. Molloy provided field assistance, waist deep in a marsh at uncivilized hours. S. Bogdanowicz, B. Butcher and L. Campagna generously shared their expertise during several phases of genomic library construction, sequencing and assembly. John Anderton graciously permitted reuse of his beautiful and nuanced illustrations of inland and coastal swamp sparrows. This research was supported by the Athena Fund at the Laboratory of Ornithology, Cornell Vertebrate Genomics, the Frank M. Chapman Fund at the American Museum of Natural History, the Andrew W. Mellon Student Research fund and Doctoral Dissertation Improvement Grant No. 1501471 from the National Science Foundation. Fellowship support for P.D. was provided by Cornell University, the Cornell Lab of Ornithology and the Natural Sciences and Engineering Research Council of Canada (PGSD3 405451-11), the David and Sandra Junkin / Walter E. Benning Graduate

Fellowship at the Cornell Laboratory of Ornithology and the Cornell University Andrew and Margaret Paul Graduate Fellowship. Sampling was conducted under appropriate state and federal permits and all procedures conformed to protocols approved by Cornell's Institutional Animal Care and Use Committee (IACUC). My collaborator Russ Greenberg and co-advisor Rick Harrison passed away before the completion of this dissertation, but both were heavily involved in discussion of the ecological (R.G.) and evolutionary (R.H.) aspects of this system and of this work. They left an indelible mark on their respective fields, and are deeply missed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

LOW GENOMIC DIVERGENCE AND HIGH GENE FLOW BETWEEN LOCALLY ADAPTED POPULATIONS OF THE SWAMP SPARROW

Abstract

Populations that have recently diverged across sharp environmental gradients provide an opportunity to study the mechanisms by which natural selection drives adaptive divergence. Inland and coastal populations of the North American swamp sparrow have become an emerging model system for studies of natural selection because they are morphologically and behaviourally distinct despite a very recent divergence time (<15,000 years), yet common garden experiments have demonstrated a genetic basis for their phenotypic differences. We characterized genomic patterns of variation within and between inland and coastal swamp sparrows via reduced representation sequencing in order to reconstruct the contributions of demography, gene flow and selection to this case of recent adaptive divergence. Compared to inland swamp sparrows, coastal swamp sparrows exhibited fewer polymorphic sites and reduced nucleotide diversity at those sites, indicating that a bottleneck and/or recent selective sweeps occurred in that population during coastal colonization and local adaptation. Estimates of genome-wide differentiation (F_{ST}=0.02) and sequence divergence (Φ_{ST}=0.05) between inland and coastal populations were very low, consistent with postglacial divergence. A small number of SNPs were strongly differentiated (max F_{ST}=0.8) suggesting selection at linked sites. Swamp sparrows sampled from breeding sites at the habitat transition between freshwater and brackish marshes exhibited high levels of genetic admixture. Such evidence of active contemporary gene flow makes the evolution and maintenance of local adaptation in these two populations even more

notable. We summarize several features of the swamp sparrow system that may facilitate the maintenance of adaptive diversity despite gene flow, including the presence of a magic trait.

Introduction

Birds offer many striking examples of local adaptation, particularly in terms of two evolutionarily labile traits: bill shape and plumage. Both traits are capable of substantial phenotypic change in response to natural selection, over both shallow and deep phylogenetic timescales (eg. Kusmierski et al. 1997; Lovette et al. 2002; Burns et al. 2002). In many cases, bill or plumage traits likely evolve via post-speciation selection, when allopatric taxa that are already reproductively isolated experience different forms of selection on these phenotypes depending on their environment (Dobzhansky 1940). Alternatively, when natural selection itself is responsible for driving genetic divergence between lineages during the evolution of local adaptation, we refer to the process as adaptive divergence. Since adaptive divergence may lead to the evolution of reproductive isolation ("ecological speciation"; Mayr 1963; Harrison 1991; Rice and Hostert 1993; Funk 1998; Schluter 2000, 2001; Coyne and Orr 2004; Nosil et al. 2005; Rundle and Nosil 2005; Nosil 2012), case studies of the early stages of adaptive divergence offer an opportunity to understand the mechanisms by which natural selection drives diversification over time. How does selection create and maintain locally adapted lineages before reproductive isolation has evolved between them?

Ecologically distinct populations of the North American swamp sparrow (*Melospiza georgiana*) have served as a natural laboratory for studies of adaptive divergence for decades. In the eastern USA there is both a widespread inland form of the swamp sparrow (*M. g. georgiana*), and a range-restricted coastal form (*M. g.*

nigrescens) that breeds in brackish tidal marshes of the Delaware and Chesapeake Bays (Figure 1.1; Greenberg and Droege 1990; Beadell et al. 2003). "Coastal Plain" swamp sparrows were first recognized as distinct from inland swamp sparrows more than 60 years ago (Bond and Stewart 1951), and various phenotypes that distinguish the two forms have since been studied extensively. The coastal population differs from inland populations in at least ten traits (Table 1), including a deeper bill and melanic plumage (Figure 1.2; Greenberg and Droege 1990). Field studies have demonstrated that at least eight of the ten derived traits that distinguish coastal swamp sparrows are adaptive in coastal environments (Table 1.1; Goldstein et al. 2004; Grenier and Greenberg 2005; Olsen et al. 2008, 2013; Peele et al. 2009; Greenberg et al. 2012), and common garden rearing experiments have confirmed that there is a strong genetic basis to most trait differences between the populations, including bill and plumage (Table 1.1; Ballentine and Greenberg 2010). Natural selection for deeper bills in coastal habitats has, in turn, biomechanically constrained the range of possible bandwidths at which coastal males can trill (Ballentine 2006), driving further population divergence through nonrandom mating. Inland swamp sparrow females prefer to mate with males that sing broad-bandwidth songs and maximize vocal performance, but coastal swamp sparrow females prefer the songs of coastal males, who compensate for a loss of bandwidth by increasing trill rate (Ballentine et al. 2013a, 2013b). Swamp sparrows are the third empirical example of bill shape as a "magic trait" (Gavrilets 2004): natural selection on the bills of Darwin's finches and crossbills also imposes biomechanical constraints on song, facilitating divergence in this mating signal and driving the evolution of reproductive isolation (Podos 2001; Podos et al. 2004; Huber and Podos 2006; Huber et al. 2007; Smith and Benkman 2007).

Although many of the adaptive traits in coastal swamp sparrows have a heritable genetic basis, coastal birds have not had much time to evolve them. Coastal tidal marshes did not exist in the northeastern or mid-atlantic regions during the Last Glacial Maximum (LGM) because northeastern coasts were under the ice sheet, and mid-atlantic coasts were hydrologically unstable due to glacial outflow from rivers and streams (Malamud-Roam et al. 2006). Tidal marsh habitats also require sediment accretion from rising sea levels in order to establish (Pethick 1984; Warren and Niering 1993). Coastal swamp sparrows have therefore likely colonized and adapted to this novel habitat rapidly, within the last 15,000 years. Consistent with a recent divergence, inland and coastal swamp sparrows have proven indistinguishable using allozymes and mitochondrial DNA (Balaban 1988; Greenberg et al. 1998). A recent study using microsatellite markers was the first to successfully resolve subtle genetic differentiation between coastal and inland swamp sparrows, and a genetic contact zone in northern New Jersey coincident with the ecotone between brackish and freshwater marshes (Greenberg et al. 2016).

To reconstruct the evolutionary processes responsible for adaptive divergence between inland and coastal swamp sparrows, we quantify genome-wide molecular variation for the species and extract patterns of historical demography, gene flow and selection. We also test for an association between phenotypic and genomic divergence, and characterize levels of admixture in the putative inland/coastal hybrid zone with genome-level sampling.

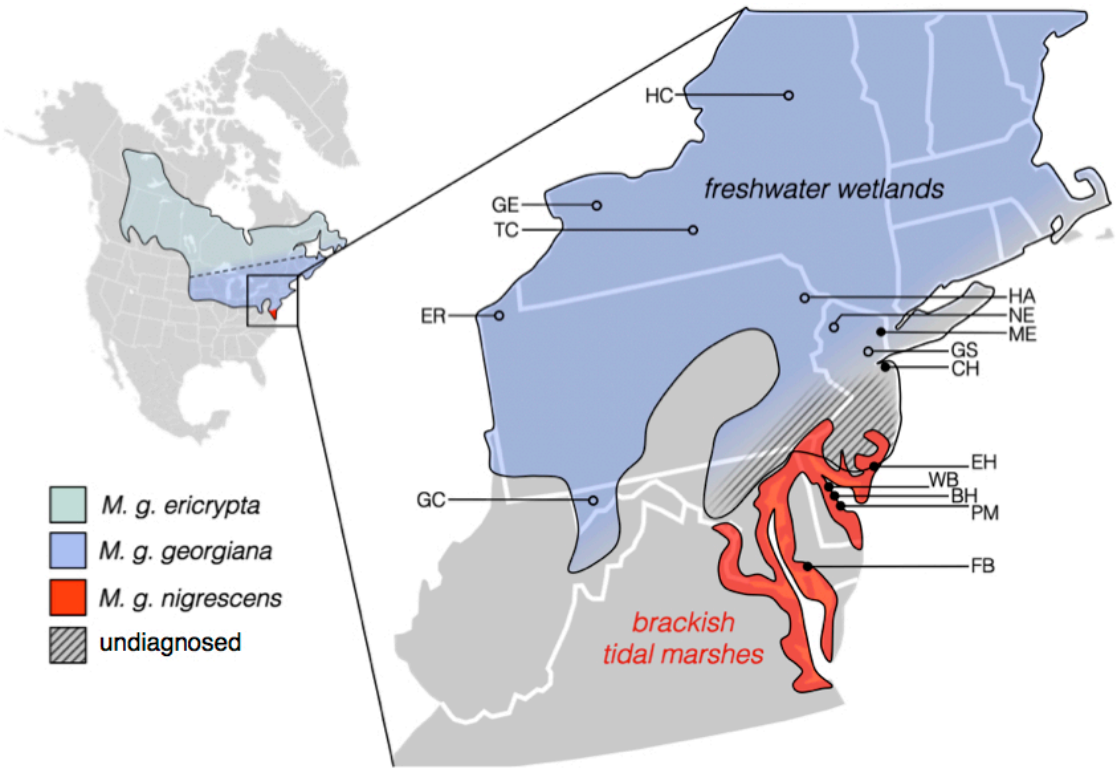


Figure 1.1 Breeding range of Swamp Sparrow subspecies (Greenberg and Droege 1990; Beadell et al. 2003). Freshwater (∘) and brackish (•) sampling sites are shown with two-letter site codes (Table 1.2). Hatching denotes an area of very low breeding density in which the phenotypes of birds have not been studied. Dashed line denotes the approximate southern extent of the boreal subspecies (*ericrypta*), not included in this study.

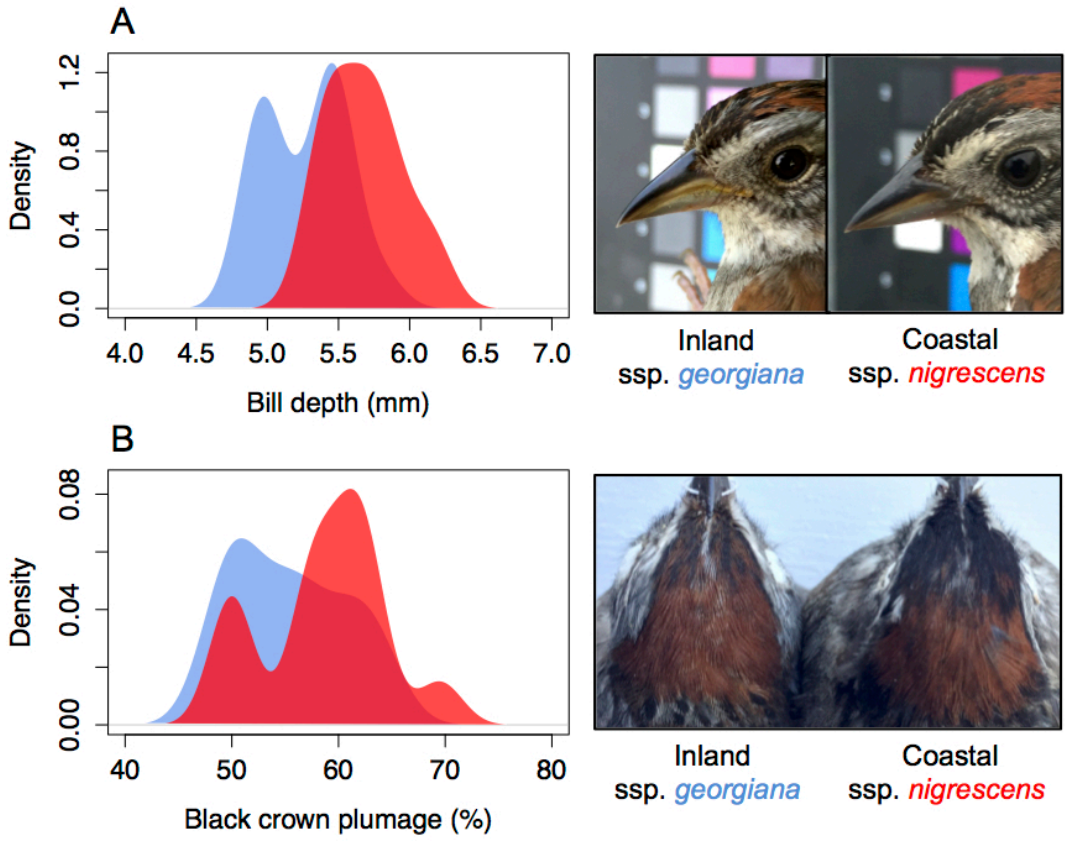


Figure 1.2 Bill depth (A) and the extent of black crown plumage (B) are two of many traits that subtly distinguish inland (blue) and coastal (red) swamp sparrows. Morphological data include only birds from allopatric sites (N=29) and are from Greenberg and Droege (1990). Photos are of breeding males (inland: Hamilton County NY and Assunpink NJ, coastal: Fishing Bay MD).

Table 1.1. Proposed adaptive hypotheses for eight traits that distinguish coastal Swamp Sparrows from inland forms, and their empirical support from field studies. Trait differences indicated by an asterisk (*) persist when coastal and inland birds are reared in a common garden. Clutch size could not be measured in captivity (Ballentine and Greenberg 2010).

Trait	Hypotheses	Proposed selective agent	Proposed adaptive benefit	Support
↑ Bill depth*	*Trophic adaptation*	More abundant invertebrate prey in coastal habitat	Improves prey capture[1,2,3]	some
	Abiotic adaptation	Heat stress due to exposed, water-limited habitats	Convective radiative heat loss[4]	Y
↑ Bill dimorphism	*Thermal budgets*	Territorial defense causes heat stress in males	Convective radiative heat loss[4]	some
	Intraspecific divergence	High breeding densities, competition for resources	Sex-based diet specialization[5]	N
	Intrasexual selection	High breeding densities, territorial competition	Competitive advantage[5]	unclear
	Female choice	High breeding densities, competition for paternity	Females prefer larger bills[3]	Y
↑ Plumage melanism*	*Predation adaptation*	Anoxia and iron sulphides cause darker sediments	Crypsis[1]	untested
	Feather preservation	Feather degrading bacteria thrive in saline habitats	Confers resistance to bacteria[6]	Y
	Intrasexual selection	High breeding densities, territorial competition	Larger black plumage badges[7]	Y
↓ Plumage rustiness*	*Predation adaptation*	Anoxia and iron sulphides cause darker sediments	Crypsis[1]	untested
	Evolutionary constraint	Selection for plumage melanism	None, females prefer rusty crowns[7,8]	Y
Delayed molt, extended breeding*	*Predation adaptation*	High predation on coast causes nest failure	Allows re-nesting[9]	some
	Abiotic adaptation	Coastal habitats experience late spring flooding	Facilitates delayed breeding[1]	some
↓ Clutch size	*Predation adaptation*	Nest predation rates are higher in coastal habitats	Reduces provisioning, detectibility[9]	Y
	Longevity trade-off	Lifetime breeding effort favored over current	Improves annual survivorship[9]	N
	Abiotic adaptation	Temperature extremes, egg loss during laying	Shorter laying avoids extreme events[9]	Y
↓ Song bandwidth*	*Evolutionary constraint*	Larger bills constrain the bandwidth of male trills	None, broad bandwidth favored[10,11,12]	Y
↑ Song trill rate	*Acoustic adaptation*	Faster trills transmit better in open habitats	Improves signal transmission[13,14]	some
	Female choice	Faster songs are honest indicators of quality	Females prefer fast trills[10,12,15,16]	Y
	Signal compensation	Bill size constrains bandwidth, males trill faster	Females prefer challenging songs[10,11]	some

1. Greenberg and Droege 1990 2. Grenier and Greenberg 2005 3. Olsen et al. 2013 4. Greenberg et al. 2012 5. Greenberg and Olsen 2010 6. Peele et al. 2009 7. Olsen et al. 2008a 8. Olsen et al. 2010 9. Olsen et al. 2008b 10. Ballentine 2006 11. Ballentine et al. 2013a 12. Liu et al. 2008 13. Morton 1975 14. Wiley and Richards 1978 15. Ballentine 2009 16. Ballentine et al. 2013b

Methods

Sample collection and phenotypic measurements. We compiled 92 DNA samples from male and female swamp sparrows from 15 sites distributed across the breeding range of inland and coastal forms in the northeastern US, including samples from the putative contact zone in northern New Jersey (Figure 1.1; Tables 1.2, A1.1). Most were blood samples collected from birds banded in the 2001 breeding season (N=66), and the remainder were blood samples from banded birds or tissue samples from vouchered specimens, collected by P. Deane-Coe (P.D.) or other collectors from the Cornell University Museum of Vertebrates during the 2007-2014 breeding seasons (N=26). We collected four voucher specimens and six blood samples from a previously unstudied coastal population at Fishing Bay Wildlife Management Area (MD) on the eastern shore of the Chesapeake Bay, representing the most southern breeding location sampled to date (Table A1.1). We also took a complete set of high resolution photographs for all birds banded in 2012-2014, using a Canon DSLR camera against a standardized color reference target (Xrite Digital SG Colorchecker), and collected standard morphological data for all banded birds.

Library preparation and sequencing. We extracted DNA using the DNeasy kit (Qiagen), quantified concentrations on a Qubit (Invitrogen, Carlsbad, CA, USA), and balanced concentrations across samples using either dilution or vacuum filtration. We digested 175ng of each sample with enzymes SBf1 and Msp1. We followed standard ddRAD protocols for digestion, Illumina TruSeq primer ligation and clean-up as described by Peterson et al. (2012). We chose a combination of eight index groups and 12 barcodes, each 5 or 6 bases in length (Table A1.2). Prior to barcode ligation, we selected fragments from 400-600bp in size using the Pippen protocol (Sage Science, MA, USA). Barcode sequences were semi-randomized by systematically distributing

individuals from each sampling location across index groups and barcode identities (Table A1.2). Within index groups, we ligated barcodes with either 14 or 16 cycles of PCR and visually inspected the products of these reactions on a gel to confirm that amplification error had not distorted the size distribution of fragments. We balanced DNA concentration to 2nM across index groups following quantification on a Bioanalyzer, and pooled index groups. We used two lanes of an Illumina HiSeq in Rapid Run mode for sequencing (yielding over 200 million 150bp reads), and we performed quality filtering and individual genotyping using a custom modification of the STACKS pipeline (Catchen et al. 2011). In brief, we executed process_radtags commands to demultiplex reads and used fastX tools for custom processing. We clipped reads with insert sizes short enough to contain adapter sequence, and trimmed all reads to 145bp. The first eight base pairs of each read did not meet quality standards due to low complexity across index sequences, and were also trimmed. We determined that trimming the sequence containing the enzyme cut site did not negatively affect the performance of the STACKS assembly by performing multiple parallel processing runs. In fact, this step improved the quality of the assembly by removing sections of low quality sequence and preserving the high quality remainder of each read. We caution, however, that this modified assembly approach may not work as efficiently in comparisons across a deeper divergence time than in our system, since successful assembly of reads from multiple individuals will rely on sequence similarity downstream of the cut site.

SNP genotyping. We quality-filtered reads, retaining those with Illumina quality scores above 24, and the resulting 137bp fragments were assembled into a catalogue of 72,246 unique loci according to a minimum depth of five reads per stack, two locus mismatches allowed within individuals, and four locus mismatches allowed in the

catalog. We dropped two individuals from the northern part of the inland range (TCm4, TCu1; both sampled far from the hybrid zone) prior to SNP discovery because they possessed unusual and divergent haplotypes at several loci, representing either sequencing error or gene flow from the unsampled boreal subspecies *M. g. ericrypta*. We grouped remaining individuals (N=90) into three populations for SNP discovery, according to whether they originated from the allopatric range of either the inland (sites HC, GE, TC, ER, GC) or coastal (WB, BH, PM, FB) populations, or from sites near the contact zone between the two (HA, NE, ME, GS, CH, EH). After extensive exploration of the impact of missing data on the accuracy of population genetic estimates from our data, we conservatively required that a given SNP be present in at least 50% of the individuals in each of the three populations. A total of 29,058 single-nucleotide sites on 4,256 stacks that met these criteria were polymorphic across allopatric inland, allopatric coastal, and sites near the contact zone. We selected the first SNP from each remaining locus to define a panel of independent SNPs for subsequent analyses of genome-wide variation and divergence (N=4,238 after additional SNP quality filtering). We performed an Analysis of Molecular Variance (AMOVA; Meirmans 2006) in the program ARLEQUIN (Excoffier and Lischer 2010) to assess the distribution of total genetic variation across different hierarchical categories of population structure. We extracted a range of different population genetic metrics and locus-specific allele frequencies from the output of STACKS, including F_{ST} (a relative measure of differentiation in allele frequencies; Wright 1965; Weir and Cockerham 1984) and Φ_{ST} (a relative measure of mutational distance between populations; Meirmans and Hedrick 2010). We also tested for F_{ST} outliers using Bayescan (Foll and Gaggiotti 2008).

Characterizing genome-wide patterns of variation and divergence. To determine whether genome-wide SNP variation reflected evolutionary distinctness of swamp sparrows from freshwater and coastal marshes, we performed k-means clustering on the results of a Principle Coordinates Analysis (PCA) using Discriminant Analysis of Principle Components (DAPC) (Adegenet; Jombart et al. 2010), and performed bayesian assignment tests using STRUCTURE (Pritchard et al. 2000). We also performed k-means clustering manually in R, to confirm the results of the DAPC algorithm. To resolve the location and extent of admixture, we estimated the probability of assignment for all individuals in the dataset to the significant genetic clusters determined in the previous step, comparing the assignments generated by both DAPC and STRUCTURE.

Testing morphological associations. We tested for significant associations between phenotypic scores for adaptive traits (bill depth, length, width, volume and extent of black crown plumage) and discriminant scores based on genome-wide SNP variation using Spearman Rank correlation (correlation coefficient = ϱ). For consistency, only birds sampled in 2001 and measured by R. Greenberg were used in these analyses. We performed analyses of covariance (ANCOVA) to test for significant associations between morphology and discriminant score for any traits that significantly scaled with weight. We also performed T-tests to assess the degree of morphological differentiation between distinct genetic clusters inferred from DAPC.

Table 1.2 Locations and dates for Swamp Sparrows sampled in this study (N=92).

Code	Sampling locations	N	Date
HC	Hamilton County, NY	3	2013
TC	Tompkins County, NY	12	2008-2014
GE	Genesee County, NY	1	2007
ER	Erie National Wildlife Refuge, PA	8	2001
GC	Garrett County, MD	8	2001
HA	Hawley, PA	3	2001
NE	Newton, NJ	6	2001
ME	Meadowlands, NJ	6	2001
GS	Great Swamp National Wildlife Refuge, NJ	10	2001
CH	Cheesequake State Park, NJ	5	2001
EH	Egg Harbor, NJ	1	2001
WB	Woodland Beach, DE	2	2001
BH	Bombay Hook, DE	8	2001
PM	Port Mahon, DE	9	2001
FB	Fishing Bay Wildlife Management Area, MD	10	2014

Results

Genome-wide differentiation and divergence. Out of the 29,058 SNPs in the complete dataset, only 11% (3,126 SNPs) showed significant allele frequency differences (corrected $F_{ST} > 0$) between coastal and inland swamp sparrows. Our panel of 4,238 independent SNPs also reflected overall similarity between coastal and inland genomes (median F_{ST}=0.00, mean F_{ST}=0.02), and sequence divergence was similarly low (median Φ_{ST}=0.03, mean Φ_{ST}=0.05). There was a long tail to the distributions of both F_{ST} and Φ_{ST} (max F_{ST}=0.80; max Φ_{ST}=0.77), representing SNPs with large allele frequency differences (F_{ST}), and loci with large mutational distances between inland and coastal haplotypes (Φ_{ST}), but Bayescan did not identify any of these as significant outliers. An AMOVA including only allopatric coastal or inland sites indicated that most of the total genetic variance is partitioned across individuals (76%) or across individuals within sites (19%), providing further indication that most variable sites within the genome are not structured between coastal and inland swamp sparrows. Higher variation within groups than between groups is commonly observed in investigations of recently diverged taxa (e.g., Campagna et al. 2015). A very low proportion of the variance (1.5%) could be attributed to structure between breeding sites within each population, indicating little to no role for isolation by distance or other geographic or demographic processes operating within coastal or inland ranges. Consistent with the presence of a long tail to the distributions of F_{ST} and Φ_{ST}, however, a small subset of genome-wide variation (3%) was partitioned between coastal and inland habitats.

The program STRUCTURE did not resolve strong support for two distinct genetic groups (best K=1), representing genetic variation within swamp sparrows as a gradual cline in allele frequencies across many individuals of intermediate assignment when K=2 was enforced (Figure 1.3). In contrast, according to DAPC, total genomic

variation within our samples was best represented by a single discriminant axis from a scaled PCA, and the lowest Bayesian Information Criterion (BIC) scores supported two genetic clusters corresponding to birds sampled from inland freshwater vs. coastal brackish marshes (Figures 1.3, A1.1). A conventional PCA with manually applied k-means clustering also robustly supported the genetic distinctness of swamp sparrows from coastal vs. inland habitats.

Molecular variation within each population. As a group, birds sampled from coastal brackish tidal marshes possessed fewer polymorphisms across the genome than those from inland freshwater marshes (2.8% vs. 3.7% of sites), and had fewer private alleles (2,367 vs. 6,473). Average nucleotide diversity was also subtly but significantly lower in coastal birds compared to inland birds (inland $\pi = 0.11 \pm 0.14$, coastal $\pi = 0.10 \pm 0.15$; $p=3.2 \times 10^{-7}$), however, despite a smaller population size coastal birds exhibited more negative values of F_{IS} than inland birds (inland $F_{IS}=-2.0$, coastal $F_{IS}=-3.2$; $p=6.3 \times 10^{-29}$), suggesting that inbreeding is not responsible for this loss of variation.

Location and extent of admixture. The two methods used to infer individual assignment probabilities (STRUCTURE and DAPC) exhibited discordance in terms of the extent of admixture, but both did exhibit concordance with respect to the location of the zone of admixture, which occurs in and around the ecotonal transition between freshwater and brackish habitats in northern NJ (Figure 1.4). Birds from two sites (ME and CH) had highly heterogeneous individual assignments to one cluster or the other according to DAPC, consistent with the wide range of discriminant axis values attributed to individuals from these northern brackish habitats (Figure 1.3, 1.4). Heterogeneity across individuals at these sites was also observed from the assignments calculated in STRUCTURE.

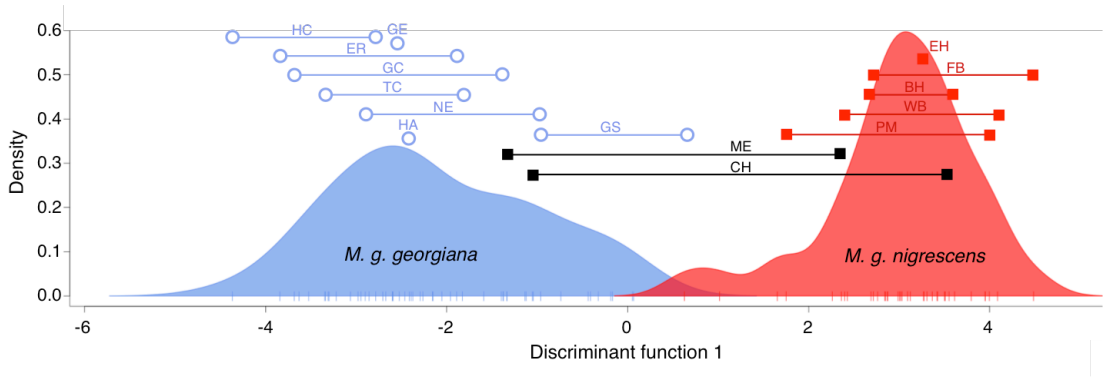


Figure 1.3 Discriminant analysis of 22 principal components, via K-means clustering and model selection, identified two clusters of genetically related swamp sparrows from a panel of 4,238 independent SNPs. A single axis of variation best distinguished the clusters. The x-axis represents the placement of individuals (short vertical lines) along this discriminant axis. The y-axis is a scaled metric representing the smoothed density of individuals corresponding to each discriminant value. Assignment probabilities consistently placed swamp sparrows from inland freshwater sites (blue) and coastal brackish sites (red) in separate clusters, with the exception of birds from two brackish sites in northern NJ (black). Individuals from these sites exhibited highly heterogeneous assignments to one cluster or the other. Brackets represent the range of genetic variation sampled from each site. Refer to Table 1.2 for site codes.

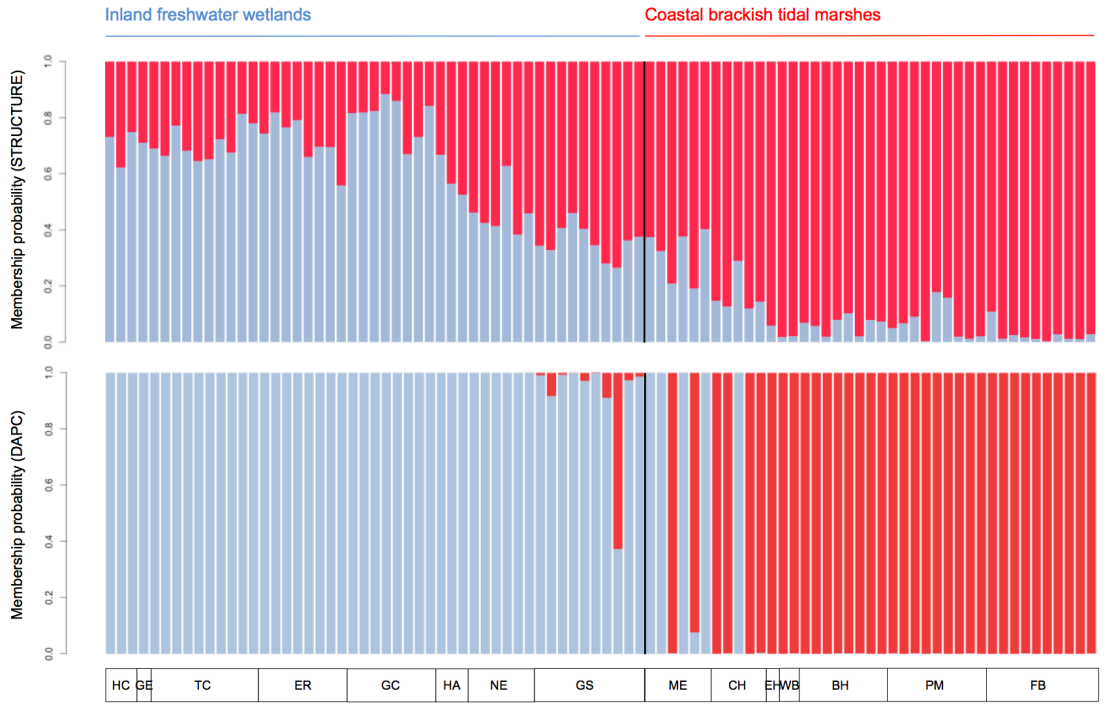


Figure 1.4 Assignment probabilities of individual swamp sparrows to one of two genetic clusters, computed by the programs STRUCTURE (top) and DAPC (bottom). Refer to Figure 1.1 and Table 1.2 for two-letter site codes. Black line indicates the ecotone between freshwater and brackish habitats.

Associations between phenotype scores and genomic assignment. Bill length, bill width and wing chord measures were not significantly correlated with the placement of individuals on the discriminant axis (Figure A1.1). Correlations between DAPC discriminant axis score and morphological score were significant for the following traits: Bill depth (ϱ=0.46, p<0.001), bill volume (ϱ=0.35, p<0.05), extent of black crown plumage (ϱ=0.29, p<0.05), tarsus (ϱ=0.20, p<0.05) and weight (ϱ=0.33, p<0.05). Bill depth and volume did not significantly scale with tarsus length, a common proxy for overall size; however both metrics did scale with weight (Figure A1.2; bill depth, ϱ=0.41, p<0.01; bill volume, ϱ=0.49, p<0.001). We conducted analyses of covariance (ANCOVA) to test for an association between phenotype and discriminant axis score when the confounding effect of weight was removed, and both relationships remained strong and significant (p<0.001). Since variation in bill depth influences bill volume, and bill width and length were not significant, we only show the data for bill depth and black crown plumage (Figure 1.5). Both bill depth and black crown plumage also significantly differed between the two distinct genetic clusters inferred from DAPC (corresponding to inland and coastal sites).

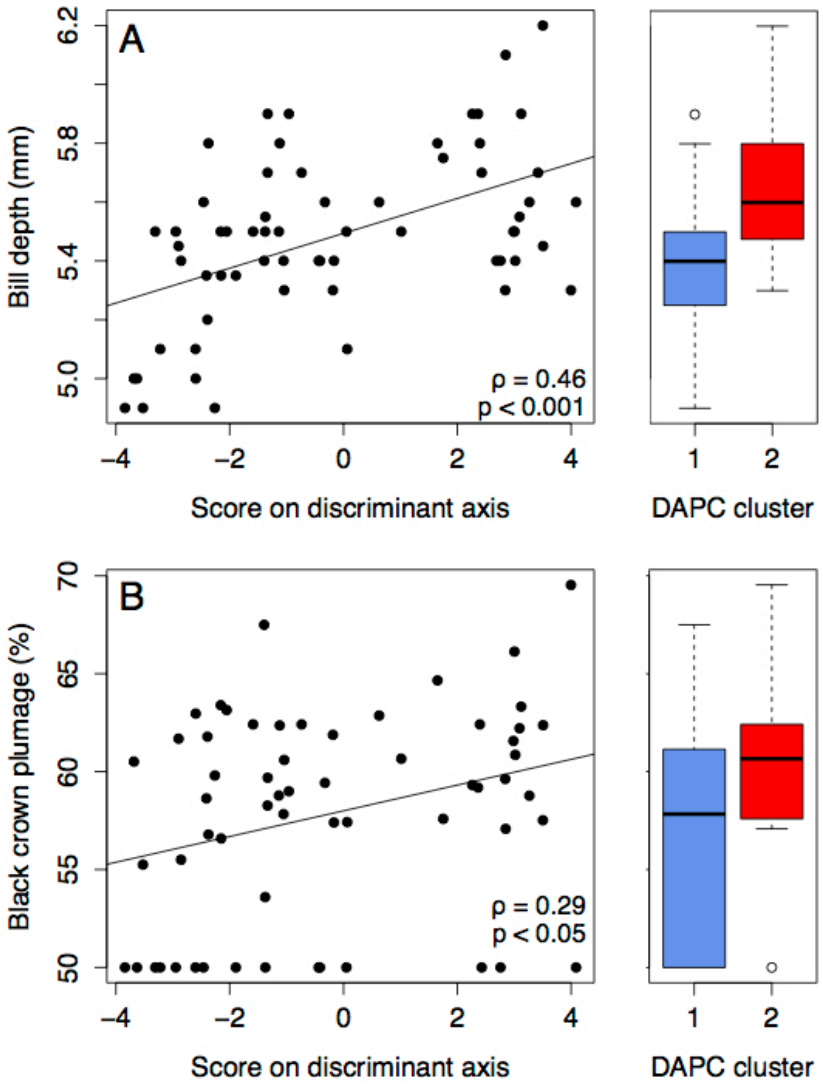


Figure 1.5 Morphological traits distinguishing inland and coastal swamp sparrows, and their association with genomic discrimination scores (left) and cluster assignment (right). Both bill depth (A) and the extent of black crown plumage (B) positively correlate with DAPC score. Birds assigned to the coastal cluster had significantly deeper bills and more extensive black crown plumage than those assigned to the inland cluster ($p < 0.001$ and $p < 0.05$, respectively).

Discussion

Recent adaptive divergence between inland and coastal swamp sparrows. In our comparisons, genome-wide patterns of variation supported previous evidence that inland and coastal swamp sparrows are very closely related, but have diverged very recently and are now on distinct evolutionary trajectories. We found evidence of population genetic structure corresponding to inland and coastal subspecies, but it was sufficiently low to be consistent with divergence during the most recent 15,000 years, and it was only detected by the more sensitive of the two methods of genetic cluster assignment (DAPC inferred two significant clusters but STRUCTURE did not). Across all swamp sparrows sampled, degrees of morphological variation did correlate with degrees of genome-wide differentiation, and both bill depth and black crown plumage were significantly associated with the genomic cluster to which birds were assigned.

Divergent SNPs despite a highly similar genomic background. Inland and coastal swamp sparrow genomes are highly similar, but a long tail in the F_{ST} distribution also suggests selection in regions of the genome linked to these sites. Bayescan detected no significant outliers, however, which is not surprising given the subtle phenotypic divergence in question: Outlier tests scan for sites with reciprocally fixed or nearly fixed haplotype frequencies, and the adaptive and heritable phenotypes that differentiate inland and coastal swamp sparrows are not fixed in either population. Rather, bill depth and plumage melanism in coastal birds is an exaggerated subset of the total trait variation among inland birds (Figure 1.2; Greenberg and Droege 1990). Detection of functional variation responsible for adaptive phenotypes will require much more extensive genomic sampling, both to increase the proportion of the genome scanned and to enhance our ability to detect significant deviations in allele

frequencies due to selection by more tightly estimating the background distribution of F_{ST} across the rest of the genome.

Evidence of a founder effect during coastal colonization. Compared to inland swamp sparrows, coastal swamp sparrows exhibited fewer polymorphic sites, reduced nucleotide diversity at those sites, and fewer private alleles. This pattern of reduced molecular variation is not likely due to post-colonization genetic drift in the smaller coastal population since average F_{IS} is negative in coastal birds, indicating less inbreeding. In fact, coastal populations appear to exhibit even less inbreeding than the large inland population (more negative F_{IS}). Reduced population genetic variation on the coast without evidence of contemporary inbreeding is therefore consistent with an alternative demographic scenario: a recent population bottleneck during coastal colonization. If coastal colonists originated from the same ancestral source population as contemporary inland populations, these founders may have carried only a subset of ancestral variation with them.

Gene flow at the inland/coastal ecotone. Admixture estimates for individuals from across the northeastern US support earlier suggestions that there is a zone of contact between inland and coastal swamp sparrows in northern New Jersey, where inland freshwater marshes transition to brackish coastal marshes. Depending on the assignment algorithm used, this transition between genetically distinct groups is inferred to be either abrupt (DAPC) or gradual (STRUCTURE), reflecting different sensitivities to the subtle signal of divergence present in the data. Finer-scale geographic sampling of sparrows from within this region is ongoing and will facilitate the use of more computationally sophisticated methods to infer ancestry (Introgress,

Gompert and Buerkle 2010; BGC, Gompert and Buerkle 2012; HapMix, Price et al. 2009; RASPberry, Wegmann et al. 2011).

Local adaptation despite gene flow. In the present day, adaptive diversity appears to be maintained by selection despite gene flow since we detected highly admixed individuals at the ecotone between habitats in northern New Jersey. Two aspects of the swamp sparrow system may explain the maintenance of locally adaptive phenotypes despite active gene flow at that contact zone. First, since adaptive trait variation in coastal birds is a subset of the variation present inland (Figure 1.2), inland populations do harbor some coastally adaptive standing variation. Gene flow from the large inland population may therefore seed both adaptive and non-adaptive variants into the coastal gene pool, and if selection were strong, differential survival would mean that only the adaptive variants are maintained on the coast (eg. differential habitat performance; Harrison 1986, 1990). This would be an example of immigrant inviability maintaining adaptive diversity (Nosil et al. 2005). Second, there is robust evidence that divergent natural selection on bill size drives positive assortative mating according to habitat in swamp sparrows (Balletine 2006; Balletine et al. 2013a, 2013b). This process could maintain local adaptation by constraining gene flow to a narrow band of ecotonal or transitional habitat where divergent selection on bill shape is not strong or is not present. Future research in this hybrid zone that includes explicit tests of either mechanism will provide valuable insight into the mechanisms of adaptive evolution.

REFERENCES

Balaban, E. 1988. Cultural and genetic variation in Swamp Sparrows (*Melospiza georgiana*) 1. Song variation, genetic variation, and their relationship. Behaviour. 105:250–291.

Ballentine, B. 2006. Morphological adaptation influences the evolution of a mating signal. Evolution. 60:1936-1944.

Ballentine, B. 2009. The ability to perform physically challenging songs predicts age and size in male swamp sparrows. Animal Behaviour. 77:973-978.

Ballentine, B., and R. Greenberg. 2010. Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. PLoS One. 5:e10229.

Ballentine, B., B. Horton, E. T. Brown and R. Greenberg. 2013a. Divergent selection on bill morphology contributes to nonrandom mating between swamp sparrow subspecies. Animal Behaviour. 86:467-473.

Ballentine, B., K. W. Gkoo and R. Greenberg. 2013b. Mechanisms of song divergence between swamp sparrow subspecies. Behaviour. 150:1165-1181.

Beadell, J., R. Greenberg, S. Droege and J. A. Royle. 2003. Distribution, abundance, and habitat affinities of the Coastal Plain Swamp Sparrow. Wilson Bulletin. 115:38-44.

Bond, G. M., and R. E. Stewart. 1951. A new Swamp Sparrow from the Maryland Coastal Plain. Wilson Bulletin. 63:38-40.

Burns, K. J., S. J. Hackett, and N. K. Klein. 2002. Phylogenetic relationships and morphological diversity in Darwin's finches and their relatives. Evolution. 56(6):1240-52.

Campagna, L., I. Gronau, L. F. Silveira, A. Siepel and I. J. Lovette. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular Ecology. 24(16):4238-4251.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. Genes Genomes and Genetics. 1:171-182.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Dobzhansky, T. Speciation as a stage in evolutionary divergence. 1940. The American Naturalist. 74(753):312-21.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetic analyses under Linux and Windows. Molecular Ecology Resources. 10:564-567.

Excoffier, L., P. E. Smouse and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131(2):479-491.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 180(2):977-93.

Funk, D. J. 1998. Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution. 52(6):1744-1759.

Goldstein, G., K. R. Flory, B. A. Browne, S. Majid, J. M. Ichida and E. H. Burtt Jr. 2004. Bacterial degradation of black and white feathers. The Auk. 121:656-659.

Gompert, Z., and C. A. Buerkle. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Molecular Ecology Resources. 10:378-384.

Gompert, Z., and C. A. Buerkle. 2012. bgc: Software for Bayesian estimation of genomic clines. Molecular Ecology Resources. 12(6):1168-76.

Greenberg, R., and S. Droege. 1990. Adaptation to tidal marshes in breeding populations of the Swamp Sparrow. The Condor. 92:393-404.

Greenberg, R., and J. E. Maldonado. 2006. Diversity and endemism in tidal-marsh vertebrates. Studies in Avian Biology. 32:32-53.

Greenberg, R. and B. Olsen. 2010. Bill size and dimorphism in tidal-marsh sparrows: island-like processes in a continental habitat. Ecology. 91:2428-2436.

Greenberg, R., P. J. Cordero, S. Droege and R. C. Fleischer. 1998. Morphological adaptation with no mitochondrial DNA differentiation in the coastal plain Swamp Sparrow. The Auk. 115:706–712.

Greenberg, R., R. Danner, B. Olsen and D. Luther. 2012. High summer temperature explains bill size variation in salt marsh sparrows. Ecography. 35:146-152.

Greenberg, R., K. M. Cammen, A. G. Wilson, B. J. Olsen, B. Ballentine, N. C. McInerney and R. C. Fleischer. Geographic population structure and subspecific boundaries in a tidal marsh sparrow. Conservation Genetics 17(3):603-613.

Grenier, J. L., and R. Greenberg. 2005. A biogeographic pattern in sparrow bill morphology: parallel adaptation to tidal marshes. Evolution. 59:1588–1595.

Groppe J., J. Greenwald, E. Wiater, J. Rodriguez-Leon, A. N. Economides, W. Kwiatkowski, M. Affolter, W. W. Vale, J. C. Izpisua Belmonte and S. Choe. 2002. Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. Nature. 420(6916):636–42.

Harrison, R. G.. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology. Oxford Univ. Press, New York, NY.

Harrison, R. G. 1991. Molecular changes at speciation. Annual Review of Ecology, Evolution and Systematics. 22:281-308.

Harrison, R. G. 2012. The language of speciation. Evolution. 66(12):3643-3657.

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J Heredity. 105:795-809.

Huber, S. K., and J. Podos. 2006. Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). Biological Journal of the Linnean Society. 88(3):489-498.

Jombart, T., S. Devillard and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics. 11(1):94.

Kane, N. C., M. G. King, M. S. Barker, A. Raduski, S. Karrenberg, Y. Yatabe, S. J. Knapp and L. H. Rieseberg. 2009. Genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. Evolution. 63:2061-2075.

Kusmierski, R., G. Borgia, A. Uy, and R. H. Crozier. 1997. Labile evolution of display traits in bowerbirds indicates reduced effects of phylogenetic constraint. Proceedings of the Royal Society of London B: Biological Sciences. 264(1380):307-13.

Liu, I. A., B. Lohr, B. Olsen and R. Greenberg. 2008. Macrogeographic vocal variation in subspecies of swamp sparrow. The Condor. 110:102-109.

Lovette, I. J., E. Bermingham, and R. E. Ricklefs. 2002. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. Proceedings of the Royal Society of London B: Biological Sciences. 269(1486):37-42.

Malamud-Roam, K. P., F. P. Malamud-Roam, E. B. Watson, J. P. Collins and B. L. Ingram. 2006. The Quaternary geography and biogeography of tidal saltmarshes. Studies in Avian Biology. 32:11-31.

Marshall, J. T. 1948. Ecologic Races of Song Sparrows in the San Francisco Bay Region: Part II. Geographic Variation. The Condor. 50(6):233-256.

Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University Press, Cambridge, MA.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution. 60(11):2399-2402.

Meirmans, P. G., and P. W. Hedrick. 2010. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources. 11(1):5-18.

Morton, E. S. 1975. Ecological sources of selection on avian sounds. The American Naturalist. 109:17-34.

Nosil, P. 2012. Ecological speciation. Oxford Univ. Press, Oxford, U.K.

Nosil, P., T. H. Vines and D. J. Funk. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution. 59(4):705-719.

Olsen, B., R. Greenberg, R. C. Fleischer and J. R. Walters. 2008a.. Extrapair paternity in the the swamp sparrow, *Melospiza georgiana*: male access or female preference? Behavioural Ecology and Sociobiology. 63:285-294.

Olsen, B., R. Greenberg, I. A. Liu, J. M. Fetch and J. R. Walters. 2010. Interactions between sexual and natural selection on the evolution of a plumage badge. Evolutionary Ecology. 24:731-748.

Olsen, B., J. M. Felch, R. Greenberg and J. Walters. 2008b. Causes of reduced clutch size in a tidal marsh endemic. Oecologia. 158:421–435.

Olsen, B. J., R. Greenberg, J. R. Walters and R. C. Fleischer. 2013. Sexual dimorphism in a feeding apparatus is driven by mate choice and not niche

partitioning. Behavioural Ecology. 00:1-12.

Olson N. D., S. P. Lund, R. E. Colman, J. T. Foster, J. W. Sahl, J. M. Schupp, P. Keim,J. B. Morrow, M. L. Salit and J. M. Zook. 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Frontiers in Genetics. 6:235.

Peele, A. M., E. H. Burtt Jr, M. R. Schroeder and R. S. Greenberg. 2009. Dark color of the Coastal Plain Swamp Sparrow (*Melospiza georgiana nigrescens*) may be an evolutionary response to occurrence and abundance of salt-tolerant feather-degrading bacilli in its plumage. The Auk. 126:531-535.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7:e37135.

Pethick, J. S. 1984. An introduction to coastal geomorphology. Dept. of Geography, Univ. of Hull, U.K.

Podos, J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature. 409(6817):185-188.

Podos, J., S. K. Huber and B. Taft. 2004. Bird song: the interface of evolution and mechanism. Annual Review of Ecology Evolution and Systematics. 35:55-87.

Poelstra, J. W., N. Vijay, M. P. Hoeppner and J. B. Wolf. 2015. Transcriptomics of colour patterning and coloration shifts in crows. Molecular Ecology. 24(18):4617-4628.

Price, A. L., A. Tandon, N. Patterson, K. C. Barnes, N. Rafaels, I. Ruczinski, T. H. Beaty, R. Mathias, D. Reich and S. Myers. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genetics. 5(6):e1000519.

Pritchard, J. K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945-959.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution. 47(6):1637-1653.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecology Letters. 8(3):336-352.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology and Evolution. 16(7):372-380.

Smith, J. W, and C. W. Benkman. 2007. A coevolutionary arms race causes ecological speciation in crossbills. The American Naturalist. 169(4):455-65.

Warren, R. S., and W. A. Niering. 1993. Vegetation change on a Northeast tidal marsh: interaction of sea-level rise and marsh accretion. Ecology. 74(1):96-103.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38(6):1358-1370.

Wegmann, D., D. E. Kessner, K. R. Veeramah, R. A. Mathias, D. L. Nicolae, L. R. Yanek, Y. V. Sun, D. G. Torgerson, N. Rafaels, T. Mosley and L. C. Becker. 2011. Recombination rates in admixed individuals identified by ancestry-based inference. Nature Genetics. 43(9):847-853.

Wiley, R. H., and D. G. Richards. 1978. Physical constraints on acoustic communication in atmosphere: implications for evolution of animal vocalizations. Behavioural Ecology and Sociobiology. 3:69-94.

Wright, S. 1965. The interpretation of population structure by F-statistic with special regard to systems of mating. Evolution. 19(3): 395-420.

CHAPTER 2

WHOLE GENOME SCAN REVEALS THE MULTIGENIC BASIS OF RECENT TIDAL MARSH ADAPTATION IN A SPARROW

Abstract

Natural selection acts on functional molecular variation to create local adaptation, the "good fit" we observe between an organism's phenotype and its environment. Genomic comparisons of lineages in the earliest stages of adaptive divergence have high power to reveal genes under natural selection because molecular signatures of selection on functional loci are maximally detectable when overall genomic divergence is low. We conducted a scan for local adaptation genes in the North American swamp sparrow (*Melospiza georgiana*), a species that includes geographically connected populations that are differentially adapted to freshwater vs. brackish tidal marshes. The brackish tidal marsh form has rapidly evolved tolerance for salinity, a deeper bill, and darker plumage since colonizing coastal habitats within the last 15,000 years. Despite their phenotypic differences, background genomic divergence between these populations is very low, rendering signatures of natural selection associated with this recent coastal adaptation highly detectable. We recovered a multigenic snapshot of ecological selection via a whole genome scan that revealed robust signatures of selection at 31 genes with functional connections to bill shape, plumage melanism and salt tolerance. As in Darwin's finches, BMP signaling appears responsible for changes in bill depth, a putative magic trait for ecological speciation. A signal of selection at BNC2, a melanocyte transcription factor responsible for human skin color saturation, implicates a shared genetic mechanism for sparrow plumage color and human skin tone. Genes for salinity tolerance

constituted the majority of adaptive candidates identified in this genome scan (23/31) and included vasoconstriction hormones that can flexibly modify osmotic balance in tune with the tidal cycle by influencing both drinking behavior and kidney physiology. Other salt tolerance genes had potential pleiotropic effects on bill depth and melanism (6/31), offering a mechanistic explanation for why these traits have evolved together in coastal swamp sparrows, and in other organisms that have converged on the same "salt marsh syndrome". As a set, these candidates capture the suite of physiological changes that coastal swamp sparrows have evolved in response to selection pressures exerted by a novel and challenging habitat.

Introduction

Natural selection is one of the most fundamental evolutionary forces, solely responsible for generating present day adaptive diversity by sorting functional molecular variation within lineages to drive local adaptation. When selection is strong, and when adaptive phenotypic divergence influences reproductive isolation, locally adapted lineages may advance along the speciation continuum toward ecological speciation (Mayr 1963; Harrison 1991; Schluter 2000, 2001; Coyne and Orr 2004; Nosil et al. 2005; Rundle and Nosil 2005; Nosil 2012; Harrison 2012). This process can therefore be better understood by characterizing the molecular mechanisms underlying local adaptation. The functional effects of genes responsible for adaptive phenotypes offer one such window onto the machinery of evolution.

Recent studies have searched for genes under selection by conducting whole genome comparisons of closely related taxa (eg. Rubin et al. 2010; Cao et al. 2011; Jones et al. 2012; Malinsky et al. 2015; Burri et al. 2015; Lopes et al. 2016; Toews et al. 2016). Modern genome scans typically reveal heterogeneous patterns due to a combination of factors that include selection, gene flow, incomplete lineage sorting,

and recombination rate variation (Payseur and Rieseberg 2014). Peaks that stand out from the genomic background are often considered "islands of divergence" (Turner et al. 2005; Seehausen et al. 2014). Because the most commonly used divergence metric (F_{ST}) is a relative measure, sensitive to both allele frequency differences and within-lineage variation at a particular site, islands of divergence may represent impermeable regions in a semipermeable genome that have been shielded from gene flow by selection (Harrison 1986, 1990; Wu 2001; Turner et al. 2005; Kane et al. 2009) or regions that have been purged of molecular variation by selective sweeps (Cruickshank and Hahn 2014; Delmore et al. 2015). Both must be distinguished from background selection in low recombination regions, a confounding process also capable of generating F_{ST} peaks (Charlesworth et al. 1993; Charlesworth 1998; Noor and Bennet 2010; Nachman and Payseur 2012; Cruickshank and Hahn 2014; Burri et al. 2015; Van Doren et al. 2017).

Populations in the earliest stages of ecological divergence across an environmental gradient provide an opportunity to characterize local adaptation genes (Harrison and Larson 2014; Seehausen et al. 2014). Recent divergence times and ongoing gene flow make adaptive regions of the genome maximally detectable against a highly similar genomic background (Seehausen et al. 2014; Payseur and Rieseberg 2016). The North American swamp sparrow (*Melospiza georgiana*) fits these criteria well, making it a tractable system in which to identify molecular targets of selection during adaptive divergence. Multiple lines of evidence suggest that swamp sparrows have recently expanded their range to colonize brackish coastal tidal marshes, as much of the northeastern and mid-atlantic coast was under an ice sheet at the Last Glacial Maximum. Mid-atlantic coasts were also hydrologically unstable during much of the Pleistocene due to glacial outflow from rivers and streams (Malamud-Roam et al. 2006), and tidal marshes require sediment accretion from gradually rising sea levels in

order to establish (Pethick 1984; Warren and Niering 1993). Coastal swamp sparrows have therefore likely colonized this novel habitat within the last 15,000 years (Greenberg and Droege 1990). In the present, habitat isolation is the only potential barrier to gene flow between coastal and inland swamp sparrows, and they continue to interbreed at the ecotone between inland freshwater and coastal brackish habitats (Greenberg et al. 2016; Chapter 1).

Local adaptation to tidal marshes. Despite a very recent divergence time and active contemporary gene flow with inland populations, coastal swamp sparrows are morphologically, ecologically and behaviorally distinct from inland populations (Greenberg and Droege 1990; Chapter 1). At least 10 well-characterized traits distinguish coastal swamp sparrows that breed in tidal marshes (*M. g. nigrescens*) from the more common inland freshwater subspecies (*M. g. georgiana*), including a tolerance for salinity, a deeper bill, and more melanic plumage. (Figure 2.1; Greenberg and Droege 1990; Grenier and Greenberg 2005; Olsen et al. 2013). Common garden experiments have revealed that differences in bill depth and plumage melanism between inland and coastal swamp sparrows persist when they are reared together in the laboratory, indicating that these components of phenotypic divergence have a heritable genetic basis (Ballentine and Greenberg 2010), and the adaptive value of these traits have been explicitly tested and demonstrated in natural populations (summarized in Chapter 1). Physiological and molecular mechanisms responsible for salt tolerance, bill depth and melanism have been characterized in many other species, providing a robust functional genomic literature from which to extract candidate genes, pathways and processes that may underlie these traits in swamp sparrows.

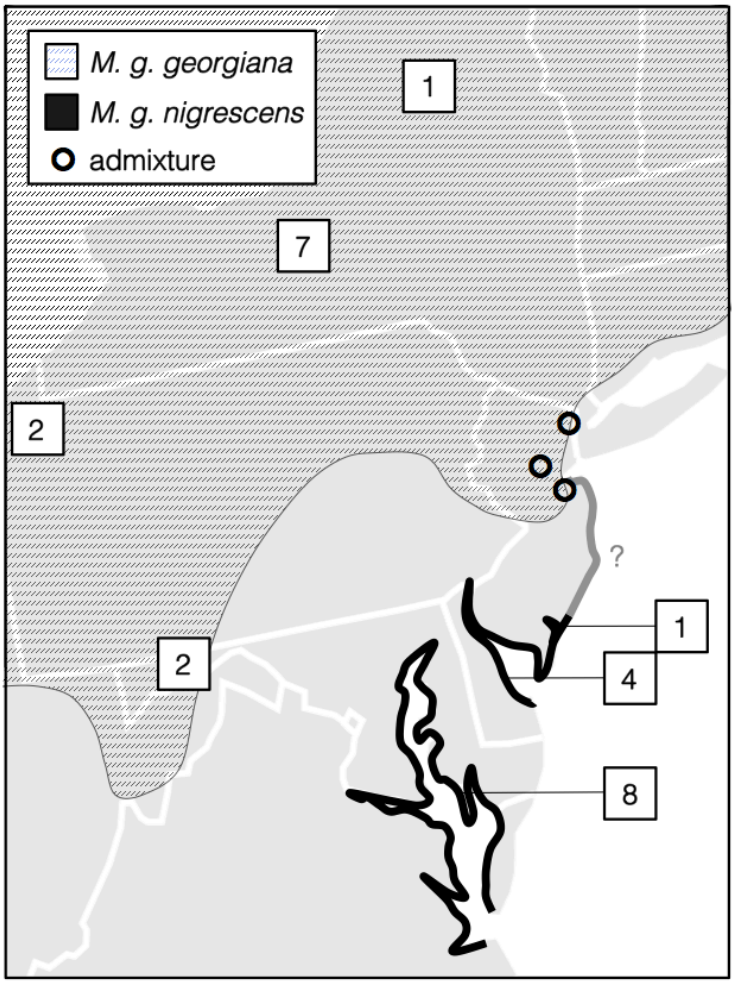


Figure 2.1 Breeding range of inland and coastal swamp sparrow subspecies in the mid-atlantic US, and their contact zone at the ecotone between freshwater and brackish tidal marshes in northern New Jersey (NJ). Tidal marsh habitat along the NJ shore contains previously unsampled (and therefore as-yet undiagnosed) populations (grey line). Numbers indicate samples from each site.

Salinity tolerance. Tidal salinity poses a fluctuating osmoregulatory challenge that all organisms breeding in tidal marshes must contend with physiologically, and it is one of the reasons that tidal marsh habitats are predominantly inhabited by specialist taxa (Correll et al. 2016). As tidal marsh specialists, coastal swamp sparrows are somehow able to tolerate increased salt intake from brackish drinking water and from foraging for invertebrate prey in brackish habitats (Greenberg et al. 2006). A coastal individual's exposure to dietary salinity, and the degree of osmoregulatory stress they must tolerate, is a product of where they drink water and where their preferred invertebrate prey live (Figure 2.2; Goldstein 2006).

Pelagic seabirds have specialized glands to concentrate and excrete salt (Heatwole and Taylor 1987), as does the saltmarsh-breeding clapper rail (Olson 1997), but passerine birds like sparrows have no such glands. Instead, passerines in saline habitats must tolerate salt through behavioral modifications that limit salt ingestion and physiological modifications that eliminate excess dietary salt via the kidney and intestine (Goldstein 2006). In a controlled comparison in a laboratory setting, sparrow species that breed in saltmarshes decreased drinking rate and maintained body mass better at higher salinity compared to inland sparrows (Bartholemew and Cade 1963). This decrease in drinking rate demonstrates that saltmarsh-adapted sparrows possess behavior-modifying mechanisms that aid osmoregulation, and their ability to maintain body mass demonstrates that they have physiological mechanisms to prevent water loss. For example, anatomical comparisons revealed that saltmarsh-adapted sparrows have structural adaptations to reduce salt reabsorption by the intestine (a smoother colon with a reduced brush border; Goldstein et al. 1990), and enlarged kidneys with an enhanced system for urine concentration (more medullary cones and loops of Henle in the nephron; Goldstein 2006).

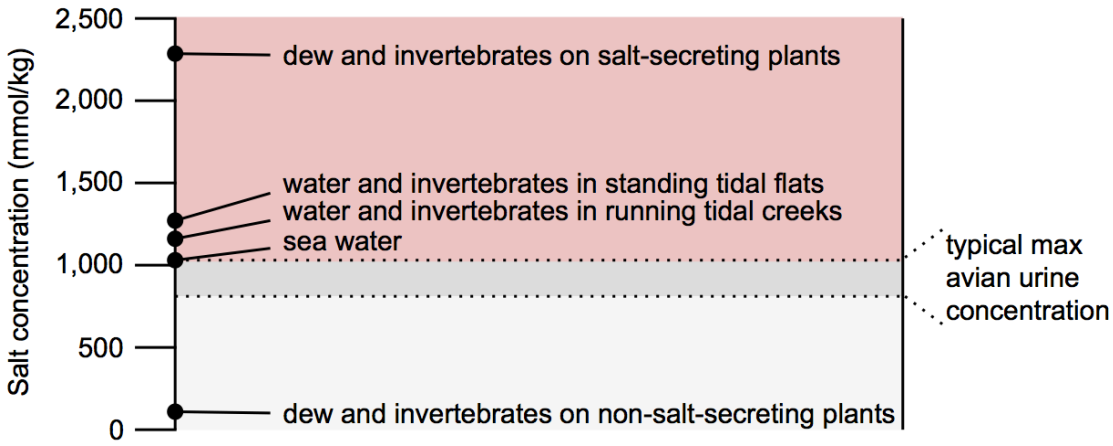


Figure 2.2 Comparison of salt concentration in different food and water sources in a tidal marsh with the maximum salt-concentrating ability of the avian kidney (copied from Goldstein et al. 1990, and adapted to include information from Goldstein 2006). Tidal marsh sparrows must evolve specialized behavioral and/or physiological mechanisms of salt tolerance, since most available food or water sources contain salt concentrations that exceed the maximum concentration of typical avian urine.

The genetic changes responsible for these specific behavioral and physiological adaptations in saltmarsh-adapted sparrows are not yet known. To date, molecular mechanisms of salt tolerance have largely been characterized for species that cannot behaviorally escape salinity, such as plants (eg. Shi et al. 2000, Xue et al. 2004; Kane et al. 2007), fish (eg. Hyndman and Evans 2007, 2009; Rengmark et al. 2007; Purcell et al. 2008) and aquatic invertebrates (eg. Patrick et al. 2000). In saline environments these organisms may increase the transport and excretion of Na^+ and Cl^- to maintain optimal osmotic balance. In plants and fish, this is often achieved by increasing the expression, distribution or affinity of Na^+ and Cl^- ion channels (Munns 2005; Hauser and Horie 2010). Much of this regulation occurs in the distal convoluted tubule (DCT) of the kidney, where the final excreted concentration of ions is determined via transporter-mediated reabsorption (de Baaij et al. 2015). In plants, ion channels expressed in the membranes of intracellular vesicles sequester and compartmentalize excess salt under salt stress conditions (Batelli et al. 2007). Pathways that regulate vesicle trafficking and turnover can confer salt tolerance by accelerating degradation or secretion of products sequestered in vacuoles, endosomes or lysosomes (Mazel et al. 2004). Organisms must also reduce intestinal permeability and transport of other ions like Mg^{2+} to maintain homeostasis without wasting water during excretion (Grosell et al. 2011; Esbaugh et al. 2014; Romani et al. 2007). Finally, in environments like tidal marshes where salinity stress changes throughout the tide cycle, rates of ion transport may be temporarily modified via vasoconstriction or vasodilation of the renal arteries or glomerular capillaries (Hyndman 2015).

Deep bills as radiators. For coastal birds, deeper bills offer a larger surface area for convective radiative heat loss, reducing thermal stress in the sun-exposed, water-limited conditions that characterize tidal marshes (Greenberg et al. 2012). Natural

selection for deeper bills in coastal swamp sparrows has, in turn, driven divergence in song (Ballentine 2006): coastal males are biomechanically constrained from singing the broad-bandwidth trills that inland females prefer, but coastal females prefer the alternative songs of coastal males, leading to non-random mating (Ballentine et al. 2013a, 2013b). This mirrors the evolution of incipient reproductive isolation in Darwin's finches and crossbills, where ecological selection favored larger bills in some populations, which in turn exerted constraints on song and promoted non-random mating (Podos 2001; Podos et al. 2004; Huber and Podos 2006; Huber et al. 2007; Smith and Benkman 2007). Swamp sparrow bill depth may therefore act as a "magic trait": a mechanism by which natural selection itself drives ecological speciation between inland and coastal lineages over time (Gavrilets 2004).

Genetic determinants of bill depth variation in birds are already well understood because this trait plays a role in several evolutionarily and agriculturally important model systems. Bone morphogenetic proteins (BMPs) are members of the Transforming Growth Factor β (TGFβ) superfamily, and BMP signaling is responsible for variation in bill depth in Darwin's finches, chickens, ducks and quail via their action during craniofacial development (Abzhanov et al. 2004; Wu et al. 2004, 2006; Brugmann et al. 2010). Over-expression of the growth factor BMP4 in the prenasal cartilage of developing chicken embryos caused those individuals to grow larger, deeper bills (Abzhanov et al. 2004; Wu et al. 2004). Other genes acting during development can lead to craniofacial deformities in humans, and several of these have also been implicated in bill shape evolution in birds (eg. Lamichhaney et al. 2015).

Increased melanism for feather preservation and territoriality. Coastal swamp sparrows may have evolved more melanic plumage than their inland relatives because feather-degrading bacteria are more abundant in in humid coastal habitats like tidal

marshes, and feathers with a higher concentration of melanins resist bacterial degradation better (Peele et al. 2009). Increased melanism also enhances the size and conspicuousness of their black forehead patch, a plumage badge that indicates dominance and aids males in territorial competition. Male-male competition is more intense for coastal swamp sparrows in comparison to inland swamp sparrows because breeding densities are very high in tidal marshes (Olsen et al. 2008a). Although it has been observed that anoxic iron sulfides cause tidal marsh sediments to be darker on average than inland marshes, inspiring speculation that melanism may also enhance crypsis (Greenberg and Droege 1990), this hypothesis has yet to be tested in swamp sparrows.

Melanic plumage is generated via the action of genes in the melanogenesis pathway (Mundy 2005), and via the function of specialized cells in the skin called melanocytes. Melanocytes contain melanosomes, specialized vesicles that collect the products of melanogenesis, melanins, and deploy them during feather development (Yu et al. 2004). Melanocyte function has been heavily studied in the context of feather patterning (eg. Lin et al. 2013), and several different members of the melanogenesis pathway have been implicated as functional candidates for melanic plumage or pelage in several different species. For example, mutations in the melanocortin-1 receptor (MC1R; Theron et al. 2001; Uy et al. 2016; reviewed in Mundy 2005) or it's antagonist, agouti signalling protein (ASIP; Manceau et al. 2011; Poelstra et al. 2014; Toews et al. 2016; Campagna et al. 2016), appear responsible for the switch between phaeomelanin (brown) and eumelanin (black) production in a range of different birds and mammals (Hubbard et al. 2010).

Melanic plumage and large bills, a saltmarsh "syndrome". Coastal subspecies of the swamp sparrow, saltmarsh sparrow, seaside sparrow, song sparrow and marsh wren are all more melanic than their inland relatives (Grinnell 1913; Von Bloeker 1932; Phillips 1986; Greenberg and Droege 1990; Luttrell et al. 2014). Saltmarsh melanism has also been described for black rails, shrews, voles, harvest mice, and the gulf saltmarsh snake (Von Bloeker 1932; Neill 1958; Pettus 1963; Conant and Lazell 1973; Myers 1988; Gaul 1996). In addition to this pervasive pattern of melanism in saltmarsh taxa, all 10 North American sparrow species or subspecies that breed in salt marshes have evolved larger bills than their closest inland relative (Grenier and Greenberg 2005). For example, in savannah sparrows (*Passerculus sandwichensis*), the tidal marsh subspecies *rostratus* has a much deeper bill than inland savannah sparrows (Grenier and Greenberg 2005) and in song sparrows (*Melospiza melodia*) the tidal marsh subspecies *maxillaris* has both a larger bill and more melanic plumage than inland song sparrows (Marshall 1948; Grenier and Greenberg 2005; Luttrell et al. 2014). Increased summer temperatures in tidal marshes relative to inland marshes offers an adaptationist explanation for this repeated evolution of large-billed forms, because increased bill volume improves radiative heat loss (Greenberg et al. 2012). Extensive phenotypic convergence would seem to constitute strong evidence for the adaptive value of deep bills and melanic plumage for a wide range of organisms living in saltmarsh habitats. As a recently diverged, tractable system that offers high detectability for adaptive variation in the genome, the swamp sparrow provides a means to understand a very taxonomically broad pattern of adaptive diversity.

We conducted a scan to identify peaks of divergence in the genomes of inland and coastal swamp sparrows, and to determine whether peaks coincide with the locations of candidate genes. We defined a candidate gene as any gene with one or more

functional connections to previously known physiological, cellular or molecular mechanisms of salinity tolerance, bill morphogenesis or melanogenesis (Table 2.1). Where possible, we differentiate between peaks in several different locations relative to transcription start sites of candidate genes, representing functionally distinct categories of molecular evolution: 1) in coding regions, 2) immediately upstream at the putative promoter or primary enhancer, 3) in introns of the target gene or neighboring genes, or 4) in adjacent non-coding regions. The last two categories represent possible sites of secondary regulatory elements (eg. "shadow enhancers"; Hong et al. 2008), which may influence the expression of target genes many kilobases away (eg. Bishop et al. 2000; Lettice et al. 2002).

Table 2.1 Candidate mechanisms for three main categories of coastal adaptation.

Adaptation		Mechanism
Salinity tolerance		Kidney function, ion transporters, vesicle trafficking
Deep bill		TGFβ / BMP signaling, craniofacial development
Melanic plumage		Melanogenesis pathway, melanocytes, melanosomes

Methods

Reference genome construction. We sequenced the genome of a male swamp sparrow in breeding condition, sampled 14 June 2014 from an active territory at a coastal breeding site in Fishing Bay Wildlife Management Area (MD) on the eastern shore of the Chesapeake Bay (38°24'36"N 76°00'09"W). This site represents the most southern coastal breeding population sampled to date, located at the greatest distance from the southern range edge of the inland subspecies. The skin specimen is vouchered at the Cornell University Museum of Vertebrates (Accession no. N). We extracted 5.5 ug of DNA from this sample using the DNeasy kit (Qiagen) and submitted it to the Genomics and Epigenomics Core Facility of the Weill Cornell

Medicine Core Laboratories Center (CLC) for whole genome library preparation and sequencing. Weill CLC prepared two mate-pair libraries with 3kb and 8kb inserts using Nextera Mate Pair library preparation kits, and one 180bp fragment library using a Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA) (Table A2.1). We confirmed target fragment size and concentration for each library with a Bioanalyzer (Agilent, Santa Clara, CA, USA), and paired end sequencing (2x100bp) was performed on the Illumina HiSeq2500 platform. Both large-insert mate pair libraries were sequenced on a single lane, and the fragment library was sequenced across two lanes.

We filtered reads failing the Illumina chastity filter and assembled remaining reads (3kb 507,715,832 reads; 8kb 466,743,564 reads; fragment 542,406,170 reads) into scaffolds using the ALLPATHS-LG pipeline, which additionally incorporates read quality and base pair uncertainty into the assembly algorithm (Gnerre et al. 2011; Ribeiro et al. 2012). We also annotated the genome using MAKER 2.32 (Cantarel et al. 2008), creating gene models from Ensembl protein and cDNA databases for the zebra finch (*Taeniopygia guttata*) and gene predictions from the program SNAP (Korf 2004).

Genome resequencing and genotyping. We compiled 24 blood or tissue samples from breeding males at allopatric coastal and inland sites (N=12 per subspecies; Figure 12; Table A2.1). Previous work has demonstrated an absence of genetic substructure within either subspecies, and no effect of isolation by distance at this geographic scale (Chapter 1). We documented phenotypes by taking a complete set of high resolution photographs for all birds banded in 2012-2014, using a Canon DSLR camera against a standardized color reference target (Xrite Digital SG Colorchecker), and collected standard morphological data for all banded birds. Blood was preserved in Queen's

lysis buffer, tissue was preserved in ethanol, and both were stored at -20°C. We extracted DNA from all samples using the DNeasy kit (Qiagen), quantified concentrations on a Qubit (Invitrogen, Carlsbad, CA, USA), and balanced concentrations to 2.3 ng/ul (119 ng total in 52.5 ul) across samples using either dilution or vacuum filtration.

We prepared a whole genome library in 350bp sonicated fragments for each individual using the Illumina TruSeq Nano library preparation kit, and performed paired-end sequencing of all 24 libraries on one lane of an Illumina NextSeq500 platform (2x150bp). One coastal sample (SWSP12, Egg Harbor NJ) did not result in high quality reads and was dropped prior to read processing and genotyping. We used the program AdapterRemoval (Lindgreen 2012) to trim Ns from reads, truncate read pairs that contain adapter sequence, and collapse any overlapping read pairs into a single read. We required that every read have a Phred quality score and a minimum length of at least 20.

Bowtie2, a gapped-read aligner (Langmead and Salzberg 2012), was used to map paired reads and unpaired collapsed reads from each individual to the reference assembly using the "very sensitive local" flag, which applies the following parameters to the alignment algorithm: a maximum effort of 20 seed extension attempts and 3 re-seed attempts, 0 mismatches allowed, a seed substring length of 20, and the interval between seed substrings defined by a square root function with a constant of 1 and coefficient of 0.5. We used Samtools (Li et al. 2009) to convert individual alignment files from sequence alignment/map format (SAM) to the binary version (BAM), and to index the fasta file for the reference genome. We applied tools from the program Picard (Wysoker et al. 2012) to add sequencing group information to individual BAM files prior to genotyping, and coordinate contig names across the dataset. We flagged individual fragments that had been sequenced more than one time as duplicates, so

that they did not mislead downstream genotyping, and indexed the resulting flagged alignments. We also created a dictionary of contig names and sizes for the reference genome.

Individual genotypes and genotype likelihoods were assigned using the UnifiedGenotyper from Genome Analysis Toolkit (GATK; DePristo et al. 2011). When identifying variant sites and assigning genotypes we required a minimum Phred-scaled base quality score of 17, and a minimum confidence threshold of 10 (34,358,667 variant sites from the alignments met these criteria). We then used VCFtools (Danecek et al. 2011) to apply more stringent filtering, requiring that each SNP be genotyped for all samples, have a minimum mapping quality score of 20, and a minimum mean depth of 5 reads across all individuals (13,819,090 SNPs).

Quantifying diversity, differentiation, and divergence. For each SNP we estimated Weir and Cockerham F_{ST} and nucleotide diversity (π) using VCFtools. We calculated 50 kb sliding window estimates of F_{ST} and π using a window step size of 10 kb, and plotted them using the Manhattan plot R package qqman (Turner 2014) to scan for regions of the genome that differed from the mean genomic background value in these stats. In comparison to blocked sliding windows (in which each regional estimate is independent), sliding windows with overlapping steps provide smoothed estimates that reduce noise in the dataset while accurately capturing any signal of gradual change across regions of interest. We defined differentiated regions as 50 kb windows for which mean F_{ST} was more than five standard deviations above the genome-wide mean, considering only the largest and best quality scaffolds in our assembly (scaffolds 0 to 150). We also scanned for long runs of homozygosity (LROH) in each differentiated region with VCFtools to test for the presence of a recent selective sweep, and calculated Tajima's D for 1 kb windows across the scaffold to infer other signatures of

selection or demography from allele frequency spectra in each window. Positive values of Tajima's D can result from excess polymorphism at intermediate frequencies, a pattern that can be generated at individual loci via balancing selection or genetic structure within the sampled population. To distinguish among these alternatives we tested for population structure by estimating F_{ST} between coastal sites using the SNPs that defined each peak with positive Tajima's D. When sites under peaks were invariant within coastal swamp sparrows, we expanded the region used to estimate F_{ST} by 1Mb (500kb in either direction). Negative values of Tajima's D indicate an abundance of rare alleles, a pattern that can be generated at individual loci via mutational accumulation of variation following a selective sweep, or genome-wide via population expansion after a bottleneck (Tajima 1989; Fu and Li 1993; Jensen et al. 2005).

Functional annotation. We mapped the swamp sparrow sequence under each F_{ST} peak to the zebra finch (*Taeniopygia guttata*) genome assembly (Taeniopygia_guttata-3.2.4, reference Annotation Release 103) using BLASTn (NCBI). We required that sequences mapped well (>70% query coverage, >80% sequence identity, E value 0.0) to a single location in the assembly in order for the genes under that peak to be considered candidates. Assuming conservation of synteny with the zebra finch, we queried NCBI for the location of the centromere of each chromosome harboring a divergence peak in swamp sparrows to assess whether peaks coincided (<1MB) with these regions of low recombination. We referred to gene entries from Entrez, UniProtKB/Swiss-Prot, GeneCards and Tocris databases to assign functional descriptions to each candidate gene located under or adjacent to the peak.

Results

Coastal swamp sparrow reference genome. Our ALLPATHS-LG reference assembly contained 4,778 scaffolds, with an N50 scaffold size of approximately 10Mb and an estimated genome size of 1.2Gb (Table A2.2; Table A2.3).

Regions of high divergence. Genome-wide divergence between inland and coastal swamp sparrows was very low (mean F_{ST}=0.015; Figure 2.3). Average genome-wide estimates of Tajima's D were negative in both inland (D = -0.73) and coastal birds (D = -0.42), representing a genome-wide signal of population expansion after a recent bottleneck in both populations. Restricting our scan to the largest scaffolds in our assembly (scaffolds 0-150), we identified 41 regions where the 50kb sliding window estimate of F_{ST} exceeded five standard deviations above the mean (0.08; Figure 3). The average size of a divergence peak was 171kb (± 294kb), the largest being 1.5Mb (scaffold 148). These windows contained highly differentiated SNPs (max F_{ST}=0.8), and deviations from the genomic baseline in π and Tajima's D. BLAST mapping of the sequence underlying each peak revealed that, out of the 138 genes located under peaks, 15 (11%) had functional effects that made them candidates for coastal adaptation. An additional 16 candidates were located adjacent to peaks. In total, 31 genes with functional connections to salt tolerance, bill depth or melanogenesis were located immediately under or adjacent to 22 of the 41 peaks characterized genome-wide (Figures 4-7; Table 2.2; Table A2.4; Figures A2.2-2.4). The remaining 19 peaks (Table A2.4) either did not contain candidate genes for coastal adaptation (N=16), or did not BLAST well to an annotated location in the zebra finch assembly (N=3).

Most candidate genes were located in regions with very low or zero π relative to the genomic background (25/31; Figure 2.4-2.5; Table 2.2; Figures A2.1-A2.4). A different but overlapping subset of candidates (13/31) coincided with runs of

homozygosity in coastal swamp sparrows. All 31 F_{ST} peaks exhibited concomitant runs of homozygosity, reduced nucleotide diversity, or both. Positive estimates of Tajima's D were explained by allele frequency divergence between coastal breeding sites for three of nine loci, but the remainder exhibited no structure within coastal swamp sparrows ($F_{ST} < 0.1$; Table A2.5). Most peaks containing candidate genes were not located in putative centromeric regions with low recombination rates (25/31).

Divergent peaks in regions of low recombination. Peaks containing 6/31 candidate genes for coastal adaptation coincided with centromeric regions in Zebra Finch (Table A2.6), and in 4/6 of these regions we detected a pattern in which sub-regions with higher F_{ST} and low-to-zero π abruptly alternated with subregions of higher π and low F_{ST} (Scaffold 10, Figure 2.4b; Scaffold 44, Figure A2.2f; Scaffold 137, Figure A2.4b; Scaffold 148, Figure 2.5b). This pattern was present at only one non-centromeric candidate gene peak (Scaffold 127, Figure A2.3f).

Figure 2.3 Genome-wide Manhattan plot between inland and coastal swamp sparrows, based on 50 kb sliding window estimates of F_{ST} across the largest scaffolds in our genome assembly (0-150). Genome-wide average F_{ST} is very low (F_{ST}=0.015; white line). Colored flags represent the location and functional annotation of candidate genes for coastal adaptation located at F_{ST} peaks more than five standard deviations (0.013) above the mean (outlier threshold=0.08; dashed line) (Table 2.4). Refer to Table 2.1 for color key.

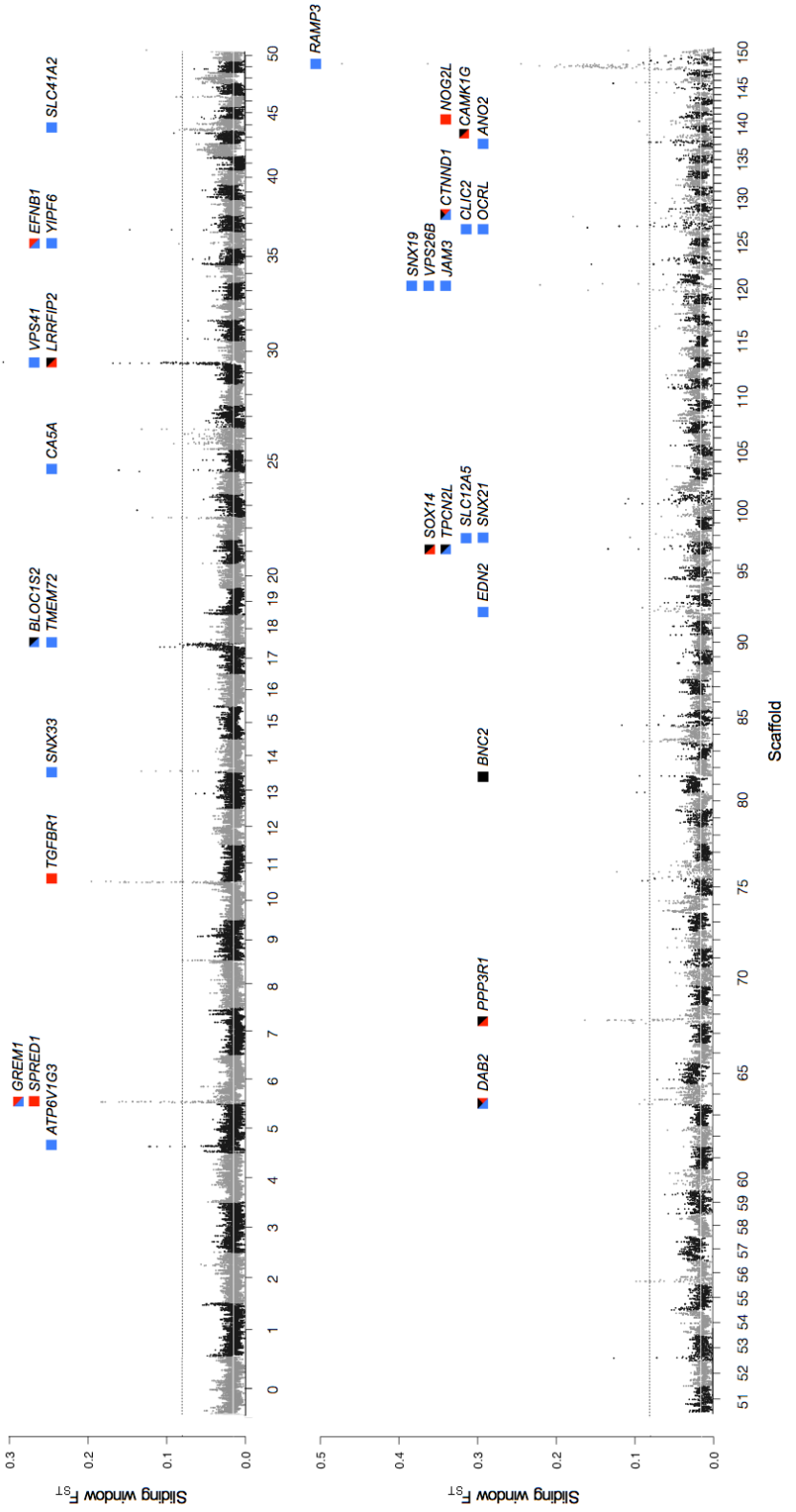
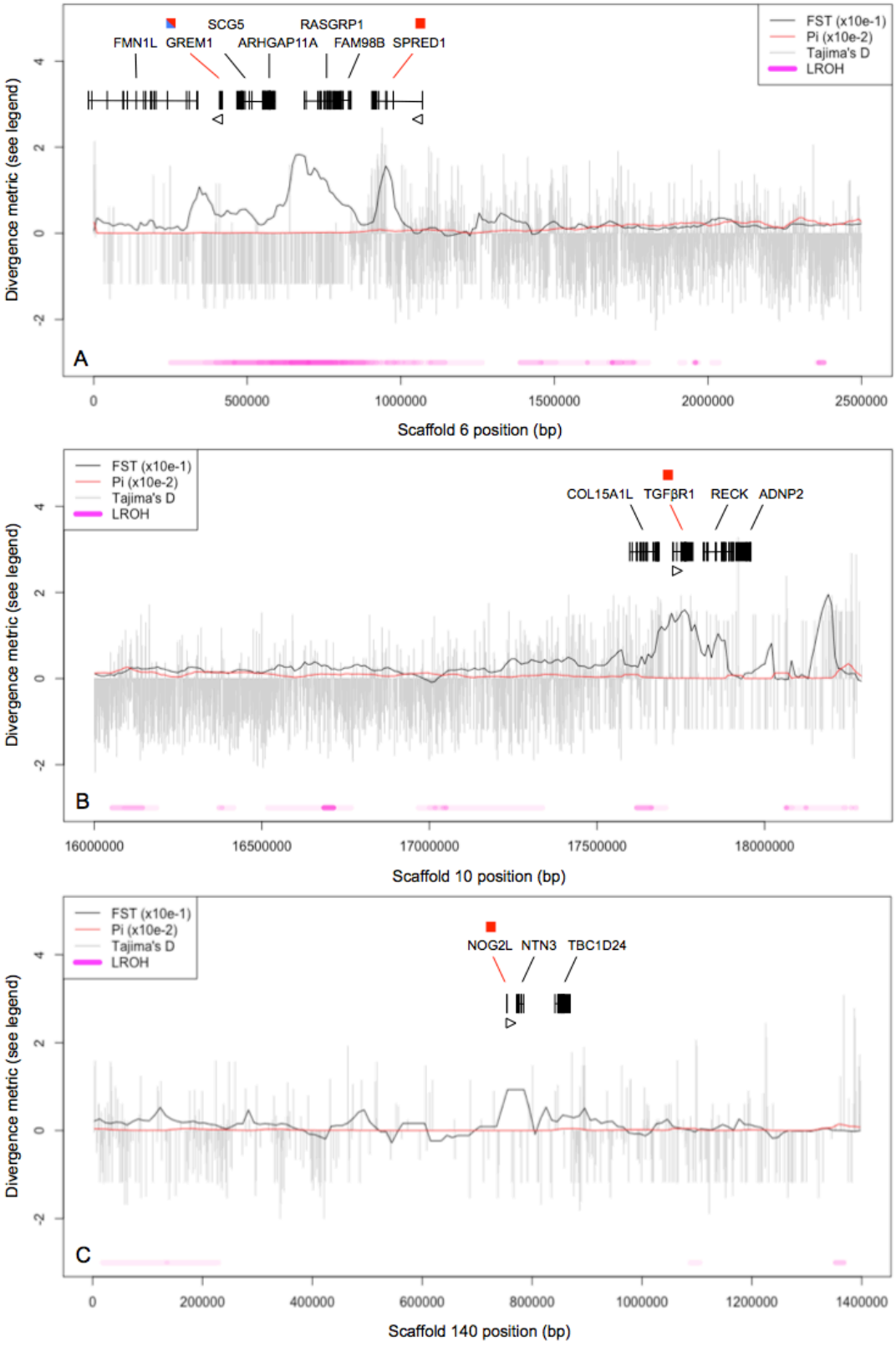

GREM1
SPRED1
ATP6V1G3
TGFBR1
SNX33
BLOC1S2
TMEM72
CA5A
VPS41
LRRFIP2
EFNB1
YIPF6
SLC41A2
RAMP3
SNX19
VPS26B
JAM3
CTNND1
NOG2L
CLIC2
CAMK1G
OCRL
ANO2
SOX14
TPCN2L
SLC12A5
SNX21
EDN2
BNC2
DAB2
PPP3R1
Sliding window F_{ST}
Scaffold

Figure 2.4 F_{ST} peaks on scaffolds 6 (A), 10 (B) and 140 (C) coincide with regions of low or zero nucleotide diversity (π) in coastal and inland swamp sparrows, consistent with recent selective sweeps, and contain three members of the BMP signaling cascade. On scaffold 6 (A), elevated F_{ST} also coincides with long runs of homozygosity (LROH) and an excess of rare alleles (Tajima's $D < 0$) in coastal swamp sparrows, further diagnostic of a sweep. Sites lacking Tajima's D estimates are invariant in coastal swamp sparrows. For LROH, color intensity corresponds to the count of coastal individuals with runs of homozygosity at that position.


FST (x10e-1)
Pi (x10e-2)
Tajima's D
LROH
FMN1L
GREM1
SCG5
ARHGAP11A
RASGRP1
FAM98B
SPRED1
Divergence metric (see legend)
A
0
500000
1000000
1500000
2000000
2500000
Scaffold 6 position (bp)
COL15A1L
TGFβR1
RECK
ADNP2
B
16000000
16500000
17000000
17500000
18000000
Scaffold 10 position (bp)
NOG2L
NTN3
TBC1D24
C
200000
400000
600000
800000
1000000
1200000
1400000
Scaffold 140 position (bp)

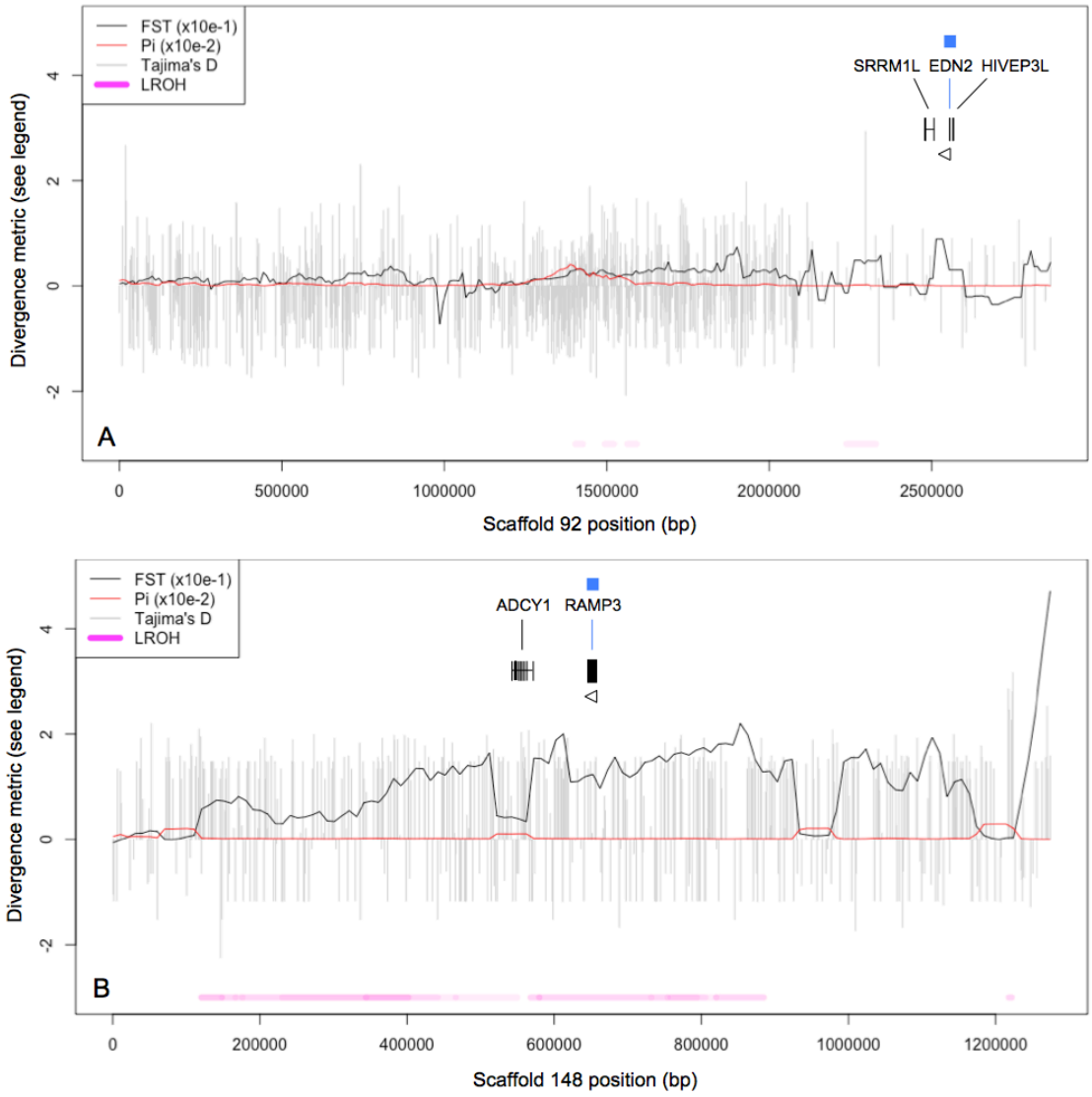


Figure 2.5 F_{ST} peaks on scaffolds 92 (A) and 148 (B) map to regions containing genes that encode kidney vasoconstriction and dilation hormones. Both show reduced nucleotide diversity (π). For LROH, color intensity corresponds to the count of coastal individuals with runs of homozygosity at that position.

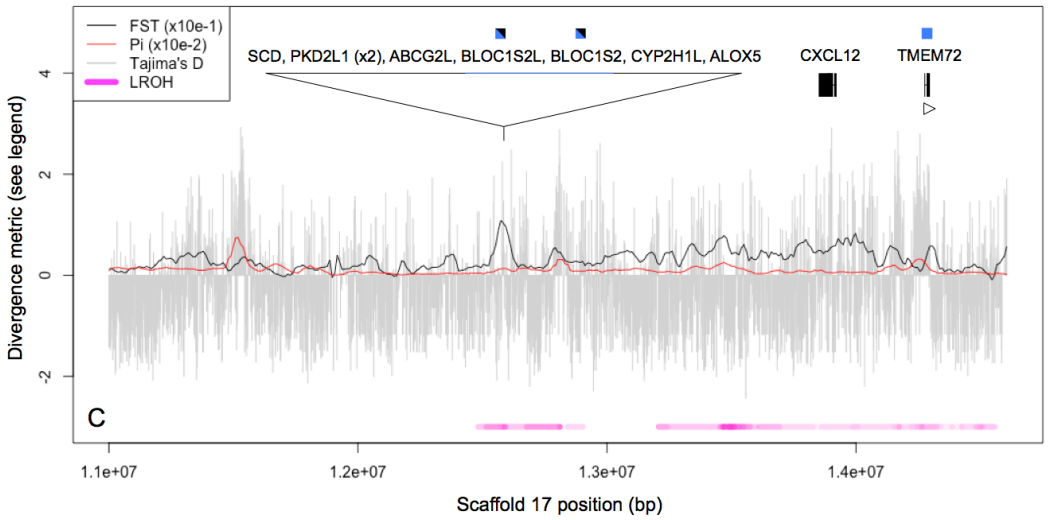


Figure 2.6 F_{ST} peaks on scaffolds 17 map to regions containing salinity tolerance genes that also show runs of homozygosity (LROH) and divergence in Tajima's D relative to the genomic background. For LROH, color intensity corresponds to the count of coastal individuals with runs of homozygosity at that position. The candidate BLOC1S2 may exert pleiotropic effects on both salinity tolerance and melanogenesis.

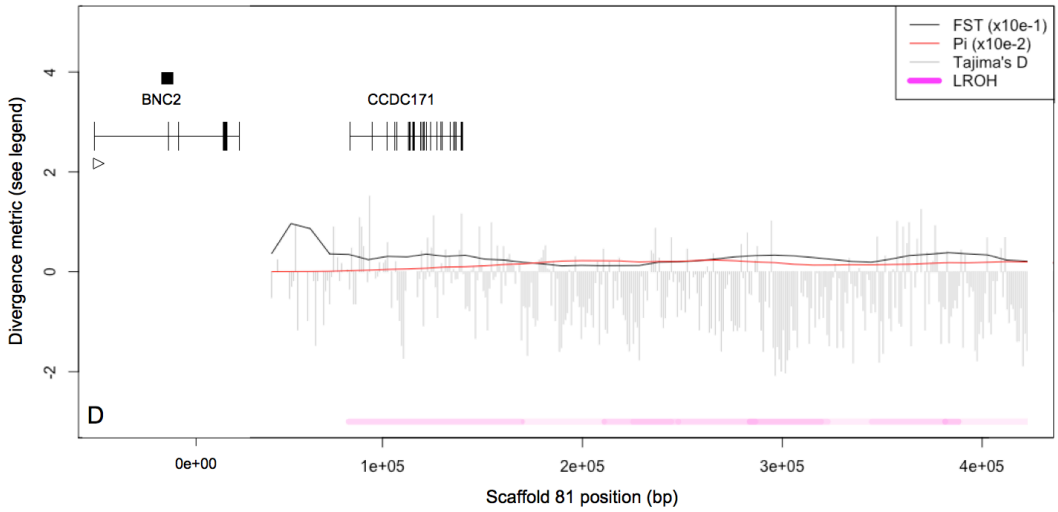


Figure 2.7 The F_{ST} peak on scaffold 81 maps to a region containing a candidate gene for melanic plumage that influences skin color saturation in human populations. Runs of homozygosity (LROH) are present across this small scaffold in several coastal individuals. For LROH, color intensity corresponds to the count of coastal individuals with runs of homozygosity at that position.

Table 2.2 Candidate genes for coastal adaptation located at F_{ST} peaks in our genome scan (N=31). Regions containing these genes exhibited runs of homozygosity (H), reduced nucleotide diversity (π), or both. Negative Tajima's D is consistent with a sweep, and positive D indicates balancing selection or structure among coastal sites at that locus. Signatures of selection at coding regions (EXON), proximal regulatory elements like primary enhancers (1°RE) or downstream/distal elements like secondary enhancers (2°RE) are given for peaks with sufficient resolution.

Gene	Function	H	π	D	Peak location
SLC41A2	Ion channel: plasma membrane Mg^{2+} transporter	Y	↓	+	
SLC12A5	Ion channel: neuron-specific Cl^- transporter		↓		
CLIC2	Ion channel: membrane-inserting Cl^- transporter	Y	↓	+	
ANO2	Ion channel: Ca^{2+} activated Cl^- channel	Y	↓	+	1°RE
TMEM72	Putative ion channel: expressed in kidney where Na^+ reabsorbed	Y		±	2°RE
TPCN2L	Ion channel: expressed in kidney, affects human hair color		↓		2°RE
ATP6V1G3	Vesicle trafficking: catalytic subunit of a vesicular ATPase H^+ pump	Y		±	2°RE
VPS41	Vesicle trafficking: regulator of vesicle transport to lysosomes	Y		-	EXON
VPS26B	Vesicle trafficking: regulator of vesicle cargo recycling		↓		1°RE
SNX19	Vesicle trafficking: regulator of vesicle trafficking and endocytosis		↓		2°RE
SNX21	Vesicle trafficking: regulator of vesicle trafficking and endocytosis		↓		
SNX33	Vesicle trafficking: regulator of vesicle trafficking and endocytosis		↓		
YIPF6	Vesicle trafficking: associated with vesicle budding and transport	~	↓		
OCRL	Vesicle trafficking: regulator of vesicle, lysosome, endosome traffic		↓		
BLOC1S2	Vesicle trafficking: produces lysosomes and melanosomes	Y	~	±	
DAB2	Vesicle trafficking: regulator of vesicle traffic, CFTR, TGFβ/WNT signal	Y		-	
EDN2	Kidney nephron vasoconstriction: the hormone endothelin 2		↓		2°RE
RAMP3	Kidney nephron vasodilation: a regulator of the hormone adrenomedullin	Y	↓	+	1°RE / 2°RE
CA5A	Excretion: modifies water balance during excretion, affects ureagenesis		↓		
JAM3	Solute permeability: adhesion molecule, prevents intracellular solute flow		↓		1°RE
CTNND1	Solute permeability: assoc. with cadherin, affects adhesion, WNT signal		↓	-	
GREM1	TGFβ signaling: BMP4 binding antagonist, affects bill/kidney development	Y	↓	-	2°RE
NOG2L	TGFβ signaling: BMP4 binding antagonist		↓		2°RE
TGFBR1	TGFβ signaling: receptor for TGFβ ligands, phosphorylates SMAD2	Y	↓	+	EXON
SPRED1	FGF signaling: antagonist of fibroblast growth factors, affects face shape	Y	↓	±	EXON
LRRFIP2	WNT signaling: may help activate WNT/β-catenin signaling	Y		-	
PPP3R1	MAPK signaling: phosphatase that binds Ca^{2+} and calmodulin		↓	-	2°RE
CAMK1G	MAPK signaling: Ca^{2+} and calmodulin dependent kinase		↓		1°RE
EFNB1	Craniofacial development: receptor ligand associated with deformities	~	↓		
SOX14	Craniofacial development: transcription factor associated with deformities		↓		2°RE
BNC2	Melanocyte zinc finger protein affecting human skin color saturation	~	↓		2°RE

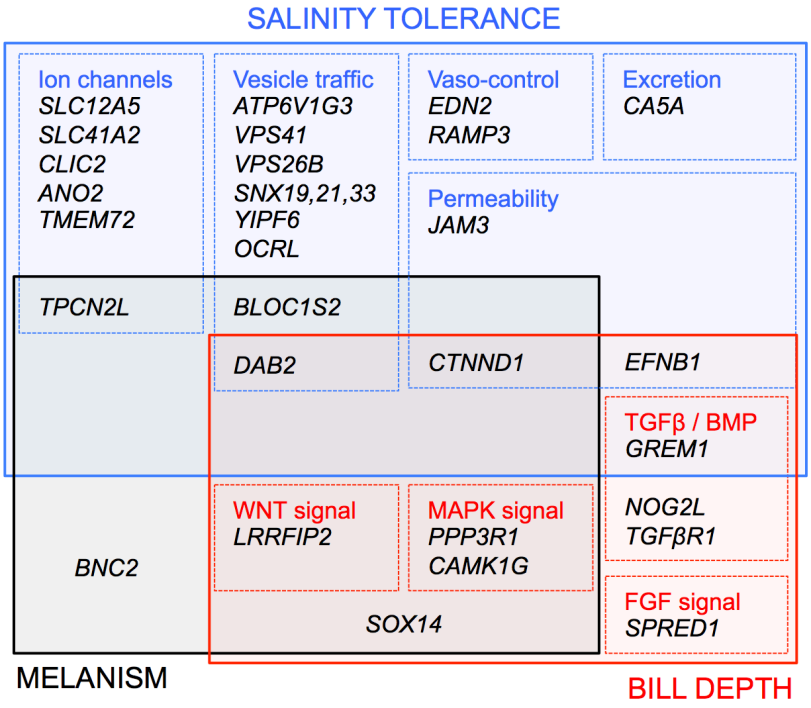


Figure 2.8 Functional categories encompassing candidate genes for tidal marsh adaptation in swamp sparrows (N=31). In several cases, candidates from one functional category may exert pleiotropic phenotypic effects on other adaptive traits.

Discussion

Inferring demography and selection from genome-wide patterns. Genome-wide F_{ST} and nucleotide diversity in inland and coastal swamp sparrows was very low, consistent with a very recent split, high levels of incomplete lineage sorting, ongoing hybridization, or all of the above. In coastal birds, on average, Tajima's D was negative in regions outside of divergence peaks, representing an excess of rare alleles across the genome. This pattern is consistent with a bottleneck having occurred during coastal colonization followed by recent population expansion (Tajima 1989). Although negative genome-wide Tajima's D provides further support for the demographic context proposed for the evolutionary origin of coastal swamp sparrows (very recent colonization and local adaptation), it might compromise our power to detect regions of negative D due to locus-specific selective sweeps. Runs of homozygosity and nucleotide diversity therefore provided additional lines of evidence with which to identify the signature of locus-specific sweeps. All 31 candidate loci located at F_{ST} peaks contained runs of homozygosity, reduced nucleotide diversity, or both.

Several candidate genes that exhibited evidence of a selective sweep via runs of homozygosity and reduced diversity were characterized by positive instead of negative values of Tajima's D within coastal birds (9/31), a departure from the overall negative genomic background that indicates either balancing selection or structure at that locus within coastal populations (Table 2.2). Since there is no genome-wide signal of structure due to drift or isolation-by-distance across breeding sites within the coastal population (Chapter 1), scenarios that could explain this pattern at certain candidate genes include 1) the presence of several different haplotypes in the coastal population that each harbor adaptive variants, or 2) selection for different haplotypes at different coastal breeding sites. We found evidence of allele frequency divergence

between coastal breeding sites at three of these loci (ATP6V1G3, SPRED1 and CLIC2), but the remainder (including TGFβR1, BLOC1S2, TMEM72 and RAMP3) exhibited minimal allele frequency divergence within coastal birds. The combined signatures of inland/coastal F_{ST} divergence, runs of homozygosity, low diversity and positive Tajima's D can be explained if adaptive variation at these particular genes was present on more than one haplotype background at the time of coastal colonization, and if each standing variant has since been driven to higher frequency by selection in coastal populations. The biogeographic context of swamp sparrow divergence over longer timescales could have provided a source of that adaptive standing variation on several different genetic backgrounds: coastal tidal marsh habitat has likely established and disappeared in sync with several glacial cycles over time (Malamud-Roam et al. 2006), providing an opportunity for periods of local adaptation followed by reabsorption into the pool of variation present in inland populations during time spent in shared glacial refugia. Genes bearing positive values of Tajima's D may therefore represent those that have been under selection during previous interglacial episodes of coastal adaptation by swamp sparrows, and persisted as standing variation in the population that recently re-colonized the coast.

Accounting for alternative processes when interpreting F_{ST} peaks. In this comparison of locally adapted swamp sparrow genomes, we interpret peaks of divergence in the relative measure F_{ST} that coincide with other divergence in other metrics (ROH, reduced π, Tajima's D) as targets of natural selection. However, background purifying selection against deleterious mutations in genes located in centromeres, inversions or other low recombination regions is also capable of purging variation and generating F_{ST} peaks across the genome that do not represent regions important for local adaptation (Charlesworth et al. 1993; Charlesworth 1998; Noor and Bennet 2010;

Nachman and Payseur 2012; Cruickshank and Hahn 2014). We acknowledge this alternative, but since background selection is often considered an effect of linked selection with weakly deleterious mutations (Nordberg et al. 1996; Charlesworth 1998), and divergence between inland and coastal lineages is so recent, it is unlikely that there has been sufficient time for this gradual process (Wolf and Ellegren 2016). Most divergence peaks harboring candidate genes in our study did not coincide with known centromeric regions in the zebra finch (25/31), but a subset did (TGFβR1, SLC41A2, ANO2, ATP6V1G3, RAMP3, NOG2L). However, even those that occurred in low recombination regions contained long runs of homozygosity localized to the peak (5/6), a fingerprint of selective sweeps that is not readily mimicked by background selection. Peaks at these genes were also small relative to the putative region of reduced recombination around centromeres. These lines of evidence suggest that divergent natural selection, not background purifying selection, is largely responsible for peaks of divergence between inland and coastal swamp sparrow genomes, and genes located at these peaks are therefore robust functional candidates for coastal adaptation. In fact, reduced recombination may have facilitated divergence at the subset of candidate genes located at centromeres (Butlin 2005; Hoffman and Rieseberg 2008; Ellegren 2012), particularly since inland and coastal swamp sparrows experience active contemporary gene flow at the ecotone between freshwater and brackish marshes (Turner et al. 2005; Noor and Bennett 2009).

BMP signaling and bill depth, a putative magic trait. Three central participants in the TGFβ / BMP signaling cascade emerged as candidates in our genome scan: TGFβ receptor 1 (TGFβR1)and two binding antagonists of BMP4 (NOG2L, GREM1). TGFβR1 propagates the BMP signaling cascade by phosphorylating SMAD2 (Von Bubnoff and Cho 2001). Noggin (NOG) is an antagonist inhibitor that binds BMP4

(McMahon et al. 1998; Groppe et al. 2002) and diffuses readily through membranes to create morphogenic gradients in BMP4 signal during development (Jones and Smith 1998). Over-expression of NOG causes birds to grow smaller, thinner bills (Abzhanov et al. 2004). The candidate gene under a peak in our genome scan is a Noggin-2 like sequence (NOG2L). Gremlin 1 (GREM1) is another binding antagonist of BMP4, and mutations in GREM1 are strongly associated with human craniofacial defects like a cleft lip or palate (Mostowska et al. 2015; Ludwig et al. 2016). All three are strong candidate genes for bill depth in swamp sparrows since studies in several other birds, including Darwin's finches, have already established a strong functional connection between this pathway and increases in the depth axis of bills during bone morphogenesis (Abzhanov et al. 2004; Wu et al. 2004, 2006; Brugmann et al. 2010).

The location of divergence peaks at the two antagonists NOG2L and GREM1 suggests selection on downstream regulatory elements that are distal to the promoter, like shadow enhancers. This type of distal secondary enhancer is fairly common (eg. Markstein et al. 2002; Zeitlinger et al. 2007) and often shares redundancy with the primary enhancer. That redundancy makes shadow enhancers more evolutionarily labile than primary enhancers, free to evolve new binding sites or other features that change expression of the target gene (Hong et al. 2008). The location of the divergence peak at TGFβR1 suggests selection on a coding region, influencing the structure and function of the receptor. Divergence at secondary regulatory elements of BMP antagonists, and at coding regions of the TGFβ receptor, may therefore represent the molecular mechanisms by which natural selection is driving adaptive bill divergence in coastal swamp sparrows. Since selection for deeper coastal bills drives song divergence and positive assortative mating between coastal and inland populations, these molecular changes represent a mechanism by which natural

selection could drive ecological speciation between inland and coastal swamp sparrows over evolutionary time.

A shared mechanism for sparrow plumage and human skin color. The candidate gene BNC2 from scaffold 81 is a transcription factor that acts specifically within melanocytes, and BNC2 is one of the main genetic markers associated with skin color saturation in human populations (Jacobs et al. 2013). In human melanocytes, the allelic variant present at an intergenic SNP proximal to the BNC2 enhancer region determines chromatin accessibility at the enhancer, leading to variation in BNC2 expression in the melanocyte and resulting variation in skin saturation (Visser et al. 2014).

This gene is made even more compelling as a candidate for plumage melanism in coastal swamp sparrows by the observation that coastal birds exhibit increased melanism in their soft parts as well as their plumage, including the skin of their legs and their lower bill (Figure A2.5). Coastal swamp sparrows experiencing ecological selection for increased melanism may therefore have converged on the same molecular mechanism as human populations. In humans, functional SNPs that associate with skin color fall in an upstream sequence motif that structurally influences chromatin accessibility at the enhancer of BNC2 (Visser et al. 2014). In swamp sparrows we detect a peak of divergence immediately downstream of the gene body, so different molecular features governing BNC2 expression are likely under selection.

Multiple physiological mechanisms for salt tolerance. Genes with a potential to confer physiological tolerance to salinity constituted the majority of candidates from our genome scan (74%), and these encoded ion transporters, hormones governing kidney vaso-control, regulators of vesicle trafficking, an enzyme affecting water balance

during excretion, and adhesion proteins that reduce intestinal permeability to ions. Salt tolerance genes involved in vesicle trafficking were particularly well represented in the list of candidates, and these eight genes are associated with a range of different trafficking networks, vesicle types, transport destinations and vesicle life cycle timepoints.

Two different divergence peaks implicate endothelin and adrenomedullin hormones as being important for coastal adaptation in swamp sparrows due to their effects on kidney nephron vaso-control. Fish express endothelin receptors in their gills and kidneys that modify blood pressure to alter sodium excretion and confer homeostasis (Hyndman 2015). Since changes to blood flow are temporary and reversible, endothelins are particularly important for maintaining osmotic balance in estuarine fish like killifish, where salinity changes dramatically with the tide cycle (Hyndman and Evans 2007, 2009). In rats fed a high salt diet, adrenomedullin was upregulated in the kidney (Cao et al. 2003), conferred protection against kidney damage (Nishikimi et al. 2002) and inhibited a rat's appetite for salt (Samson and Murphy 1997), indicating that this hormone can regulate osmotic balance by modifying both excretory physiology and behavior. This behavioral effect of adrenomedullin matches well with known mechanisms of salt tolerance in salt marsh specialist sparrows, since they actively decrease drinking rate as salinity increases (Bartholemew and Cade 1963). The peak encompassing RAMP3, a required activator of the adrenomedullin receptor (McLatchie et al. 1998), is the largest in our comparison between inland and coastal swamp sparrows, indicating that natural selection acting on adrenomedullin binding has been particularly strong during local adaptation to tidal marshes. Vasoconstriction and vasodilation hormones may therefore provide a powerful and temporally flexible functional mechanism by which coastal swamp sparrows physiologically respond to changing tidal salinity.

Pleiotropic effects of salt tolerance genes. Candidate gene BLOC1S2 encodes a subunit of the BLOC1 protein that, in mice, is required to produce normal lysosomes in the kidney and normal melanosomes in the skin. BLOC1 knockout mice exhibited impaired kidney function and a pale coat color (Theriault and Hurley 1970; Nguyen et al. 2002; Dell-Angelica 2004). Candidate gene TPCN2 encodes an ion channel expressed in the kidney, and harbors SNPs that determine blond vs. brown hair in humans (Sulem et al. 2008; Zong et al. 2009). The potential for these two candidate genes to exert pleiotropic effects on both salt tolerance and melanism may explain how melanic plumage has evolved multiple times in bird populations adapting to brackish tidal marshes or saltmarshes. As is the case in swamp sparrows, melanic plumage or pelage may confer an adaptive benefit for saltmarsh mice, shrews, voles, rails and snakes, but pleiotropy with salt tolerance mechanisms may strengthen and simplify natural selection for both traits. Since habitat isolation is often the only barrier between saltmarsh taxa and their inland relatives, this could help adaptive divergence proceed, or be maintained, even in the face of ongoing gene flow (Kondrashov and Mina 1986; Via 2001). Present day inland swamp sparrows exhibit variation in plumage melanism (Greenberg and Droege 1990). If functional variants of these genes were segregating as standing variation in the source population that first colonized tidal marsh habitats, darker birds could have been "pre-adapted" to deal better with salinity and may have experienced differential survival. In other species where saltmarsh melanism has evolved but there is no strong evidence for the adaptive value of that melanism, it could be that melanism is in fact a molecular "spandrel": a selectively neutral, pleiotropic consequence of selection for salt tolerance (Barrett and Hoekstra 2011; Gould and Lewontin 1979). Future studies characterizing tissue-specific gene expression throughout development are required to test whether

BLOC1S2 or TPCN2L do exert pleiotropic effects on swamp sparrow physiology as they do on human and mouse physiology.

A reduction in BMP4 activity due to antagonism by the candidate bill depth gene GREM1 is also required for normal kidney development (Michos et al. 2004, 2007). Like the pleiotropic effect of salt tolerance candidates on melanogenesis, the dual role of the BMP4 antagonist GREM1 in bill and kidney function may help explain an additional axis of parallel evolution among salt marsh birds: larger bill size. Under this scenario the pleiotropic effects of GREM1 on kidney development could have strengthened selection, contributing to patterns of parallel evolution across species like swamp, song and savannah sparrows that have colonized tidal marsh habitat and presumably experienced similar combinations of selective pressures. Alternatively, instead of both traits being adaptive, the convergent evolution of larger bills across tidal marsh sparrows could be a non-adaptive byproduct of selection on GREM1 for only one of its roles, either adaptation to heat stress (bill size) or salt stress (kidney development).

Peaks without candidates. We did not detect candidate genes for salt tolerance, bill depth or melanism at almost half of the divergence peaks we characterized genome-wide (19/41). Since at least eight other traits distinguish inland and coastal swamp sparrows, including degrees of sexual dimorphism, reproductive biology, molt timing, migration, and the corticosterone stress response (summarized in Chapter 1), these peaks could harbor functional genes for other coastally adaptive traits. Alternatively, they could harbor functional genes for traits under selection in inland freshwater swamp sparrows.

Conclusions. Swamp sparrows have recently colonized coastal tidal marshes and have rapidly evolved a suite of local adaptations. We detected signatures of selection on 20 different salt tolerance genes representing at least five different physiological mechanisms; a diverse panel of candidates that speaks to the multigenic nature of adaptation to a systemic environmental stressor like salinity. We also detected selection at genes with important evolutionary roles in other systems, including BMP signaling factors that influence bill depth in Darwin's finches and the melanocyte transcription factor BNC2 that determines human skin tone. The targets of natural selection for swamp sparrow bill depth and plumage melanism are therefore examples of taxonomically broad functional convergence, and are evidence of the repeatability of evolution. The results of our genome scan also emphasize the perhaps-underappreciated role of pleiotropy in local adaptation. Six of the candidate genes for salt tolerance also have well-documented effects on bill shape and plumage melanism. In cases like the swamp sparrow, in which the adaptive value of deep bills or melanism has been tested and demonstrated (Olsen et al. 2008a; Peele et al. 2009; Greenberg et al. 2012), pleiotropy with salt tolerance mechanisms may have facilitated divergence by offering natural selection a single target capable of influencing several adaptive phenotypes. Alternatively, pleiotropy could cause divergence in a suite of traits as a by-product of selection on a single trait, and this needs to be incorporated as a component of null models in future studies of adaptive phenotypic divergence.

REFERENCES

Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant and C. J. Tabin. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. Science. 305:1462–1465.

Batelli, G., P. E. Verslues, F. Agius, Q. Qiu, H. Fujii, S. Pan, K. S. Schumaker, S. Grillo, J. K. Zhu. 2007. SOS2 promotes salt tolerance in part by interacting with the vacuolar H+-ATPase and upregulating its transport activity. Molecular and Cellular Biology. 27(22):7781-90.

Barrett, R. D, and H. E. Hoekstra. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature Reviews Genetics. 12(11):767-80.

Bartholomew, G. A., and T. J. Cade. 1963. The water economy of land birds. The Auk. 80(4):504-39.

Bishop, C. E., D. J. Whitworth, Y. Qin, A. I. Agoulnik, I. U. Agoulnik, W. R. Harrison, R. R. Behringer, and P. A. Overbeek. 2000. A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. Nature Genetics. 26(4):490-4.

Brugmann, S. A., K. E. Powder, N. M. Young, L. H. Goodnough, S. M. Hahn, A. W. James, J. A. Helms and M. Lovett. 2010. Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. Human Molecular Genetics. 19:920-930.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, and S. Hogner. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Research. 25(11):1656-65.

Butterworth, M. B. 2010. Regulation of the epithelial sodium channel (ENaC) by membrane trafficking. Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease. 1802(12):1166-77.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. I. Tubaro, and I. J. Lovette. Repeated divergent selection on pigmentation genes in a rapid finch radiation driven by sexual selection. bioRxiv. 2016 Jan 1:075713.

Cantarel, B. L., I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado and M. Yandell. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research. 18:188-196.

Cao, Y. N., K. Kitamura, J. Kato, K. Kuwasako, K. Ito, H. Onitsuka, Y. Nagoshi, T. Uemura, T. Kita, and T. Eto. 2003. Chronic salt loading upregulates expression of adrenomedullin and its receptors in adrenal glands and kidneys of the rat. Hypertension. 42(3):369-72.

Cao, J., K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, and X. Wang. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature Genetics. 43(10):956-63.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution. 15(5):538-43.

Chou, S. Y., K. S. Hsu, W. Otsu, Y. C. Hsu, Y. C. Luo, C. Yeh, S. S. Shehab, J. Chen, V. Shieh, G. A. He, and M. B. Marean. 2016. CLIC4 regulates apical exocytosis and renal tube luminogenesis through retromer-and actin-mediated endocytic trafficking. Nature Communications. 20:7.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

De Baaij, J. H., F. J. Arjona, M. Van Den Brand, M. Lavrijsen, A. L. Lameris, R. J. Bindels, and J. G. Hoenderop. 2016. Identification of SLC41A3 as a novel player in magnesium homeostasis. Scientific Reports. 6:28565.

De Baaij, J. H., J. G. Hoenderop, and R. J. Bindels. 2015. Magnesium in man: implications for health and disease. Physiological Reviews. 95(1):1-46.

De Baaij, J. H., M. J. Koerkamp, M. Lavrijsen, F. van Zeeland, H. Meijer, F. C. Holstege, R. J. Bindels, and J. G. Hoenderop. 2013. Elucidation of the distal convoluted tubule transcriptome identifies new candidate genes involved in renal Mg2+ handling. American Journal of Physiology-Renal Physiology. 305(11):F1563-73.

Dell'Angelica, E. C. 2004. The building BLOC(k)s of lysosomes and related organelles. Current Opinion in Cell Biology. 16(4):458-64.

Delmore, K. E., S. Hübner, N. C. Kane, R. Schuster, R. L. Andrew, F. Câmara, R. Guigó, and D. E. Irwin. 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Molecular Ecology. 24(8):1873-88.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, and A. McKenna. 2011. A

framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 43(5):491-498.

Esbaugh, A. J., T. Kristensen, H. Takle, and M. Grosell. 2014. The effects of sustained aerobic swimming on osmoregulatory pathways in Atlantic salmon (*Salmo salar*) smolts. Journal of Fish Biology. 85(5):1355-68.

Gnerre, S., I. MacCallum, D. Przybylski, F. Ribeiro, J. Burton, B. Walker, T. Sharpe, G. Hall, T. Shea, S. Sykes, A. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander and D. B. Jaffe. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the Royal Society B. 108(4):1513-1518.

Goldstein, D. L. 2006. Osmoregulatory biology of saltmarsh passerines. Studies in Avian Biology. 32:110.

Goldstein, D. L., and E. J. Braun. 1989 Structure and concentrating ability in the avian kidney. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology. 256(2):R501-9.

Goytain, A., and G. A. Quamme. 2005. Functional characterization of the human solute carrier, SLC41A2. Biochemical and biophysical research communications. 330(3):701-5.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London. Series B, Biological Sciences 205(1161): 581-98.

Greenberg, R., and S. Droege. 1990. Adaptation to tidal marshes in breeding populations of the Swamp Sparrow. The Condor. 92:393-404.

Greenberg, R., K. M. Cammen, A. G. Wilson, B. J. Olsen, B. Ballentine, N. C. McInerney, R. C. Fleischer. 2016. Geographic population structure and subspecific boundaries in a tidal marsh sparrow. Conservation Genetics. 17(3):603-13.

Grinnell, J. 1909. Three new Song Sparrows from California. University of California Publications in Zoology 5: 265-269.

Grosell, M., A. P. Farrell, and C. J. Brauner. 2011. *The Role of the Gastrointestinal Tract in Salt and Water Balance*. San Diego, CA: Elsevier, Inc

Hartzell, H., K. Yu, Q. Xiao, L. T. Chien, and Z. Qu. 2009. Anoctamin/TMEM16 family members are Ca^{2+}-activated Cl^- channels. The Journal of Physiology. 587(10):2127-39.

Habuka, M., L. Fagerberg, B. M. Hallström, C. Kampf, K. Edlund, Å. Sivertsson, T. Yamamoto, F. Pontén, M. Uhlén, and J. Odeberg. 2014. The kidney transcriptome and proteome defined by transcriptomics and antibody-based profiling. PLoS One. 9(12):e116125.

Hauser, F., and T. Horie. 2010. A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K+/Na+ ratio in leaves during salinity stress. Plant, Cell & Environment. 33(4):552-65.

Hebert, S. C., D. B. Mount, and G. Gamba. 2004. Molecular physiology of cation-coupled Cl– cotransport: the SLC12 family. Pflügers Archive. 447(5):580-93.

Hirata, Y., H. Hayakawa, Y. Suzuki, E. Suzuki, H. Ikenouchi, O. Kohmoto, K. Kimura, K. Kitamura, T. Eto, K. Kangawa, and H. Matsuo. 1995. Mechanisms of adrenomedullin-induced vasodilation in the rat kidney. Hypertension. 25(4):790-5.

Hyndman, K. A. 2015. The evolution and comparative physiology of endothelin regulation of sodium transport. In *Sodium and Water Homeostasis* (pp. 119-140). Springer: New York.

Hyndman, K. A., and D. H. Evans. 2007. Endothelin and endothelin converting enzyme-1 in the fish gill: evolutionary and physiological perspectives. Journal of Experimental Biology. 210(24):4286–4297.

Hyndman, K. A., and D. H. Evans DH. 2009. Effects of environmental salinity on gill endothelin receptor expression in the killifish, *Fundulus heteroclitus*. Comparative Biochemical Physiology A: Molecular Integrative Physiology. 152(1):58–65.

Jacobs, L. C., M. A. Hamer, D. A. Gunn, J. Deelen, J. S. Lall, D. Van Heemst, H. W. Uh, A. Hofman, A. G. Uitterlinden, C. E. Griffiths, and M. Beekman. 2015. A genome-wide association study identifies the skin color genes IRF4, MC1R, ASIP, and BNC2 influencing facial pigmented spots. Journal of Investigative Dermatology. 135(7):1735-42.

Jacobs, L. C., A. Wollstein, O. Lao, A. Hofman, C. C. Klaver, A. G. Uitterlinden, T. Nijsten, M. Kayser, and F. Liu. 2013. Comprehensive candidate gene study

highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. Human Genetics. 132(2):147-58.

Jensen, J. D., Y. Kim, V. B. DuMont, C. F. Aquadro, and C. D. Bustamante. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics. 170(3):1401-10.

Ji, W., J. N. Foo, B. J. O'Roak, H. Zhao, M. G. Larson, D. B. Simon, C. Newton-Cheh, M. W. State, D. Levy and R. P. Lifton. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nature Genetics. 40(5):592-9.

Jones, F. C., M .G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, and E. Birney. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484(7392):55-61.

Jones, C. M., and J. C. Smith. 1998. Establishment of a BMP-4 morphogen gradient by long-range inhibition. Developmental Biology. 194(1):12-7.

Kane, N. C., and L. H. Rieseberg. 2007. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics. 175(4):1823-34.

Kodzhahinchev, V., D. Kovacevic, and C. Bucking. 2017. Identification of the putative goldfish (Carassius auratus) magnesium transporter SLC41a1 and functional regulation in the gill, kidney, and intestine in response to dietary and environmental manipulations. Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology. 206:69-81.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics. 5(1):1.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C. J. Rubin, C. Wang, N. Zamani, and B. R. Grant. 2015. Evolution of Darwin/'s finches and their beaks revealed by genome sequencing. Nature. 518(7539):371-5.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human Molecular Genetics. 12(14):1725-35.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genomes Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078-2079.

Li, X., N. M. Young, S. Tropp, D. Hu, Y. Xu, B. Hallgrímsson, and R. S. Marcucio. 2013. Quantification of shape and cell polarity reveals a novel mechanism underlying malformations resulting from related FGF mutations during facial morphogenesis. Human Molecular Genetics. 22(25):5160-72.

Lin, S. J., J. Foley, T. X. Jiang, C. Y. Yeh, P. Wu, A. Foley, C. M. Yen, Y. C. Huang, H. C. Cheng, C. F. Chen, and B. Reeder. 2013. Topology of feather melanocyte progenitor niche allows complex pigment patterns to emerge. Science. 340(6139):1442-5.

Lindgreen, S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Resources Notes. 5:337.

Lopes, R. J., J. D. Johnson, M. B. Toomey, M. S. Ferreira, P. M. Araujo, J. Melo-Ferreira, L. Andersson, G. E. Hill, J. C. Corbo, and M. Carneiro. 2016. Genetic basis for red coloration in birds. Current Biology. 26(11):1427-34.

Ludwig, K. U., S. T. Ahmed, A. C. Böhmer, N. B. Sangani, S. Varghese, J. Klamt, H. Schuenke, P. Gültepe, A. Hofmann, M. Rubini, and K. A. Aldhorae. 2016. Meta-analysis reveals genome-wide significance at 15q13 for nonsyndromic clefting of both the lip and the palate, and functional analyses implicate GREM1 as a plausible causative gene. PLoS Genetics. 12(3):e1005914.

Luttrell, S. A., S. T. Gonzalez, B. Lohr, and R. Greenberg. 2014. Digital photography quantifies plumage variation and salt marsh melanism among Song Sparrow (Melospiza melodia) subspecies of the San Francisco Bay. The Auk. 132(1):277-87.

Malamud-Roam, K. P., F. P. Malamud-Roam, E. B. Watson, J. P. Collins and B. L. Marshall, J. T. 1948. Ecological races of Song Sparrows in the San Francisco Bay Region. The Condor 50: 233-256.

Manceau, M., V. S. Domingues, R. Mallarino, and H. E. Hoekstra. 2011. The developmental role of Agouti in color pattern evolution. Science. 331(6020):1062-5.

Mazel, A., Y. Leshem, B. S. Tiwari, and A. Levine. 2004. Induction of salt and osmotic stress tolerance by overexpression of an intracellular vesicle trafficking protein AtRab7 (AtRabG3e). Plant Physiology. 134(1):118-28.

McLatchie, L. M., N. J. Fraser, M. J. Main, A. Wise, J. Brown, N. Thompson, R. Solari, M. G. Lee, and S. M. Foord. 1998. RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. Nature. 393(6683):333-9.

McMahon, J. A., S. Takada, L. B. Zimmerman, C. M. Fan, R. M. Harland, and A. P. McMahon. 1998. Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. Genes & Development. 12(10):1438-52.

Michos, O., A. Gonçalves, J. Lopez-Rios, E. Tiecke, F. Naillat, K. Beier, A. Galli, S. Vainio, and R. Zeller. 2007. Reduction of BMP4 activity by gremlin 1 enables ureteric bud outgrowth and GDNF/WNT11 feedback signalling during kidney branching morphogenesis. Development. 134(13):2397-405.

Michos, O., L. Panman, K. Vintersten, K. Beier, R. Zeller, and A. Zuniga. 2004. Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. Development. 131(14):3401-10.

Mostowska, A., K K. Hozyasz, P. Wójcicki, K. Żukowski, A. Dąbrowska, A. Lasota, M. Zadurska, A. Radomska, I. Dunin‐Wilczyńska, and P. P. Jagodziński. 2015. Association between polymorphisms at the GREM1 locus and the risk of nonsyndromic cleft lip with or without cleft palate in the Polish population. Birth Defects Research Part A: Clinical and Molecular Teratology. 103(10):847-56.

Mundy, N. I. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proceedings of the Royal Society of London B: Biological Sciences. 272(1573):1633-40.

Munns, R. 2005. Genes and salt tolerance: bringing them together. New Phytologist. 167(3):645-63.

Nguyen, T., M. Kermani, J. Fluhr, M. L. Wei, E. K. Novak, L. L. Peters, and R. T. Swank. 2002. Melanosome morphologies in murine models of Hermansky–Pudlak syndrome reflect blocks in organelle development. Journal of Investigative Dermatology. 119(5):1156-64.

Nishikimi, T., Y. Mori, N. Kobayashi, K. Tadokoro, X. Wang, K. Akimoto, F. Yoshihara, K. Kangawa, and H. Matsuoka. 2002. Renoprotective effect of chronic adrenomedullin infusion in Dahl salt-sensitive rats. Hypertension. 39(6):1077-82.

Patrick, M. L., and T. J. Bradley. 2000. The physiology of salinity tolerance in larvae of two species of *Culex* mosquitoes: the role of compatible solutes. Journal of Experimental Biology. 203(4):821-30.

Peele, A. M., E. H. Burtt Jr, M. R. Schroeder and R. S. Greenberg. 2009. Dark color of the Coastal Plain Swamp Sparrow (*Melospiza georgiana nigrescens*) may be an evolutionary response to occurrence and abundance of salt-tolerant feather-degrading bacilli in its plumage. The Auk. 126:531-535.

Peng, X., X. Ding, T. Chang, Z. Wang, R. Liu, X. Zeng, Y. Cai, and Y. Zhu. 2014. Overexpression of a vesicle trafficking gene, OsRab7, enhances salt tolerance in rice. The Scientific World Journal. 483526.

Purcell, K. M., A. T. Hitch, P. L. Klerks, and P. L. Leberg. 2008. Adaptation as a potential response to sea-level rise: a genetic basis for salinity tolerance in populations of a coastal marsh fish. Evolutionary Applications. 1(1):155-60.

Rengmark, A. H., A. Slettan, W. J. Lee, Ø. Lie, and F. Lingaas. 2007. Identification and mapping of genes associated with salt tolerance in tilapia. Journal of Fish Biology. 71:409-22.

Ribeiro, F., D. Przybylski, S. Yin, T. Sharpe, S. Gnerre, A. Abouelleil, A. M. Berlin, A. Montmayeur, T. P. Shea, B. J. Walker, S. K. Young, C. Russ, I. MacCallum, C. Nusbaum and D. B. Jaffe. 2012. Finished bacterial genomes from shotgun sequence data. Genome Research. 22**:**2270-2277.

Romani, A. 2007. Regulation of magnesium homeostasis and transport in mammalian cells. Archives of Biochemistry and Biophysics. 458(1):90-102.

Rong, Y., M. Liu, L. Ma, W. Du, H. Zhang, Y. Tian, Z. Cao, Y. Li, H. Ren, C. Zhang, and L. Li. 2012. Clathrin and phosphatidylinositol-4, 5-bisphosphate regulate autophagic lysosome reformation. Nature Cell Biology. 14(9):924-34.

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, and F. Hallböök. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 464(7288):587-91.

Samson, W. K., and T. C. Murphy. 1997. Adrenomedullin inhibits salt appetite 1. Endocrinology. 138(2):613-6.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G. P. Saetre, C. Bank, Å. Brännström, and A. Brelsford. 2014. Genomics and the origin of species. Nature Reviews Genetics. 15(3):176-92.

Shah, G. N., T. S. Rubbelke, J. Hendin, H. Nguyen, A. Waheed, J. D. Shoemaker, and W. S. Sly. 2013. Targeted mutagenesis of mitochondrial carbonic anhydrases VA and VB implicates both enzymes in ammonia detoxification and glucose metabolism. Proceedings of the National Academy of Sciences. 110(18):7423-8.

Shi, H., M. Ishitani, C. Kim, and J. K. Zhu. 2000. The *Arabidopsis thaliana* salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. Proceedings of the National Academy of Sciences. 97(12):6896-901.

Smith, J. W., and C. W. Benkman. 2007. A coevolutionary arms race causes ecological speciation in crossbills. The American Naturalist. 169(4):455-65.

Song, L., A. Mercado, N. Vázquez, Q. Xie, R. Desai, A. L. George, G. Gamba, and D. B. Mount. 2002. Molecular, functional, and genomic characterization of human KCC2, the neuronal K–Cl cotransporter. Molecular Brain Research. 103(1):91-105.

Sulem, P., D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S. A. Gudjonsson, A. Palsson, G. Thorleifsson, and S. Pálsson. 2008. Two newly identified genetic determinants of pigmentation in Europeans. Nature Genetics. 40(7):835-7.

Supuran, C. T. 2008. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. Nature Reviews Drug Discovery. 7(2):168-81.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585-95.

Theriault, L. L., and L. S. Hurley. 1970. Ultrastructure of developing melanosomes in C57 black and pallid mice. Developmental Biology. 23(2):261-75.

Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs, and N. I. Mundy. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. Current Biology. 11(8):550-7.

Toews, D. P., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Current Biology. 26(17):2313-8.

Turner, S.D. 2014. Qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. BiorXiv DOI:10.1101/005165.

Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Molecular Ecology. 26:3982-3997.

Visser, M., R. J. Palstra, and M. Kayser. 2014. Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. Human Molecular Genetics. 23:5750-5762.

Von Bubnoff, A., and K. W. Cho. 2001. Intracellular BMP signaling regulation in vertebrates: pathway or network? Developmental Biology. 239(1):1-4.

Wilke T.A., S. Gubbels, J Schwartz, and J.M. Richman. 1997. Expression of fibroblast growth factor receptors (FGFR1, FGFR2, FGFR3) in the developing head and face. Developmental Dynamics 210:41-52.

Wine, J. 2003. Rules of conduct for the cystic fibrosis anion channel. Nature Medicine. 9(7):827.

Wu, P., T. X. Jiang, J. Y. Shen, R. B. Widelitz and C. M. Chuong. 2006. Morphoregulation of avian beaks: comparative mapping of growth zone activities and morphological evolution. Developmental Dynamics. 235:1400–1412.

Wu, P., T. X. Jiang, S. Suksaweang, R. B. Widelitz and C. M. Chuong. 2004. Molecular shaping of the beak. Science. 305:1465– 1466.

Wysoker, A., K. Tibbetts and T. Fennell. 2012. Picard Tools.

Xue, Z. Y., D. Y. Zhi, G. P. Xue, H. Zhang, Y. X. Zhao, and G. M. Xia. 2004. Enhanced salt tolerance of transgenic wheat (*Tritivum aestivum* L.) expressing a vacuolar Na+/H+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na+. Plant Science. 167(4):849-59.

Yu, M., Z. Yue, P. Wu, D. Y. Wu, J. A. Mayer, M. Medina, R. B. Widelitz, T. X. Jiang, and C. M. Chuong. 2004. The developmental biology of feather follicles. The International Journal of Developmental Biology. 48:181.

Zong, X., M. Schieder, H. Cuny, S. Fenske, C. Gruner, K. Rötzer, O. Griesbeck, H. Harz, M. Biel, and C. Wahl-Schott. 2009. The two-pore channel TPCN2 mediates NAADP-dependent Ca2+-release from lysosomal stores. Pflügers Archive - European Journal of Physiology. 458(5):891-9.

CHAPTER 3

SWEEPS FOR CHEAP: A NOVEL METHOD FOR DETECTING SELECTIVE SWEEPS FROM STANDING VARIATION

Abstract

Whole genome sequence data are being collected and analyzed for an ever-increasing number of taxa with the goal of identifying molecular signatures of selection. Very recently diverged sister taxa contain the most detectable signal of selection because demography, drift and neutral mutation have not yet obscured the picture. Because whole genome data still represent a substantial investment in terms of library preparation, sequencing and computing, we present a novel method that leverages the clarity of signal present in recently diverged lineages to conduct a quick and easy scan for one category of selective sweep using reduced representation data. We predicted that F_{ST} analog statistics that incorporate levels of locus-specific variation (F'_{ST}) and the extent of sequence divergence between haplotypes (Φ_{ST}) could be used to identify soft sweeps from standing variation in comparisons of very shallowly diverged sister taxa. Here we develop the theory underlying this prediction and conduct an empirical test of it using ddRADseq in the swamp sparrow (*Melospiza georgiana*), a system in which the genomic location of recent selective sweeps is already known from whole genome scans. We predicted that ddRAD loci near genes that have experienced a selective sweep in one lineage would exhibit relatively large allele frequency differences between the source and sweep populations when scaled by the level of variation present in each population (high F'_{ST}), but the mutational distance between haplotypes at that locus in source vs. sweep populations (Φ_{ST}) would be no greater than that present in shared ancestral variation. One of the four ddRAD loci that met

these criteria mapped near a gene known to have experienced a recent sweep (NOG2L), providing validation of the method. With some caveats, we demonstrate that comparison of F'_{ST} and Φ_{ST} is a simple but informative tool for identifying genes that have experienced soft selective sweeps, and one that could be widely applied to suspected cases of recent adaptation from standing variation.

Introduction

Detecting fingerprints of selection in the genomes of natural populations is a daunting task, despite a revolution in sequencing technology that has made it possible to collect genome-scale data for non-model organisms. When two taxa are hypothesized to be under divergent selection, genomic comparisons may be conducted to identify F_{ST} outliers, regions with increased genetic distance (D_{XY}), departures in the allele frequency spectrum (Tajima's D), runs of homozygosity (ROH, EHH; Sabeti et al. 2002; Vitti et al. 2013), or divergence in composite statistics that integrate across several of these molecular signatures (CLR; Nielsen et al. 2005) (reviewed in Nadeau and Jiggins 2010; Wolf and Ellegren 2016). However powerful these statistical approaches may be, the strength of inference possible from genomic comparisons is still fundamentally constrained by the evolutionary history of the species being studied. If species are too distantly related, background levels of divergence may be too high to identify particular loci under selection (Seehausen et al. 2014; Wolf and Ellegren 2016), or it may become difficult to distinguish genomic "islands of divergence" due to divergent selection (Turner et al. 2005) from background purifying selection against deleterious mutations in regions of low recombination (Charlesworth 1998; Noor & Bennett 2009; Nachman and Payseur 2012; Cruickshank and Hahn 2014; Burri et al. 2015; Van Doren et al. 2017). Many natural populations also have uncertain and potentially complex demographic histories that can create spurious

signals of divergence in the genome, particularly when gene flow is limited or absent between the lineages being compared (Cruickshank and Hahn 2014; Wolf and Ellegren 2016).

Recently diverged taxa offer tractable systems in which to search for un-confounded signatures of selection, because the demographic context of divergence is often simple, and because there has been insufficient time for other molecular processes, such as the neutral mutational accumulation of genome-wide divergence, to obscure the signal of recent selective sweeps (Seehausen et al. 2014; Wolf and Ellegren 2016). If reproductive barriers are also incomplete, gene flow and recombination "shuffle the deck" and create new linkage relationships, allowing regions of the genome to act as independent units (Barton and Hewitt 1985; Hewitt 1988; Harrison 1990; Wu 2001; Kane et al. 2009). This homogenization of the genomic background maximizes detectability of genes and gene regions that have experienced selective sweeps due to their role in reproductive isolation and local adaptation (Machado and Hey 2003; Dopman et al. 2005; Payseur and Nachman 2005; Noor and Feder 2006; Buerkle and Lexer 2008; Gompert and Buerkle 2010; Payseur 2010). Within a population responding to selection, sweeps may occur on new mutations (classic "hard" sweeps; Maynard Smith and Haigh 1974) or on standing variation ("soft" sweeps; Messer and Petrov 2013). Soft sweeps are one example of a molecular signature of selection that presents a substantial detection challenge (Messer and Petrov 2013; Berg and Coop 2015), and are generally only detected in recently diverged taxa (eg. Colosimo et al. 2005; Studer et al. 2011) since the demographic context of divergence must be simple enough to have a reasonable hypothesis about the population that served as the source of that variation, and because neutral rates of mutation accumulation over time must not have obscured haplotype similarity

between the lineage experiencing the sweep and the putative source population (Colosimo et al. 2005; Barrett and Schluter 2008).

Since sweeps are highly detectable in recently diverged taxa, these systems can be expected to yield robust, high-resolution results from whole genome comparisons (eg. Chapter 2; Rubin et al. 2010; Cao et al. 2011; Jones et al. 2012; Malinsky et al. 2015; Burri et al. 2015; Lopes et al. 2016; Toews et al. 2016). The flip side of that coin is that they are also tractable systems in which to take computational shortcuts to the same answer. We present a new concept in the latter category; a quick and easy method for detecting sweeps from standing variation using restriction site associated DNA sequencing (RADseq). The protocol for library preparation, cost of sequencing, and computing requirements for data analysis with RADseq are readily accessible for most modern molecular labs, which is still not the case for *de novo* whole genome assembly and re-sequencing. Although the spare distribution of RAD markers across the genome make them inferior to whole genome sequencing methods for holistic selection detection (Lowry et al. 2017), reduced representation methods still have high power to detect sites under selection for systems in which linkage disequilibrium is high (McKinney et al. 2017), including those that have experienced recent divergence and strong selection.

The impacts of a sweep on F_{ST} analog statistics. Wright's Fixation Index (F_{ST}), a relative measure of molecular divergence, estimates allele frequency differentiation between two populations at a particular site or SNP (Wright 1965; Weir and Cockerham 1984). Conventional F_{ST} outlier approaches to candidate gene discovery are imperfectly suited to detecting sweeps from standing variation, however. We present an example involving a comparison between two recently diverged lineages, each containing the same subset of variation present in the ancestral population at the

time of divergence (Figure 3.1). F_{ST} outlier tests perform well with reciprocally fixed or nearly fixed haplotype frequencies (Weir and Cockerham 1984; Wolf and Ellegren 2016), but they would not detect a strong signal of selection when a haplotype variant is fixed or nearly fixed in one lineage that has experienced a selective sweep, if that haplotype was present at moderately high frequency in the ancestral population, and is therefore still present in the population with which it is being compared. This lower likelihood of detection is made even more unlikely if the genome is sampled sparsely via GBS, RAD or ddRAD sequencing, because these methods are only expected to detect signals of F_{ST} divergence when regions surrounding a selected variant also show divergence due to physical linkage. Linkage diminishes with distance, however, so an already-subtle signal would not be detectable over a large region.

Two F_{ST} analogs, F'_{ST} and Φ_{ST}, capture additional features of molecular divergence and are therefore better suited to detecting the unique signature of a recent sweep from standing variation. F'_{ST} is scaled relative to the level of variation present within each population (Meirmans 2006; Meirmans and Hedrick 2010; Bird et al. 2011), and is therefore a useful single statistic that captures the combined signals of allele frequency divergence (high F_{ST}) and low nucleotide diversity (low π), frequently applied as a criteria for detection of a selective sweep in genome scans. Φ_{ST} incorporates the mutational distance between haplotypes (Excoffier et al. 1992). As an indicator of the extent of neutral mutation that has accumulated between two lineages, this metric has largely been used as a proxy for time-in-isolation in population genetic studies (eg. Friesen et al. 2005; Joly et al. 2016).

Since large differences in the variation present at a locus between two populations would drive F'_{ST} higher, and overall similarity between haplotypes in those populations would keep Φ_{ST} low (Meirmans 2006; Excoffier et al. 1992; Meirmans and Hedrick 2010; Bird et al. 2011), we predicted that regions of the

genome that have experienced a recent sweep from standing variation would exhibit relatively large allele frequency differences between the source and sweep populations when scaled by the level of variation present in each population (high F'_{ST}), but the mutational distance between haplotypes at that locus in source vs. sweep populations would be small (low Φ_{ST}) (Figure 3.1).

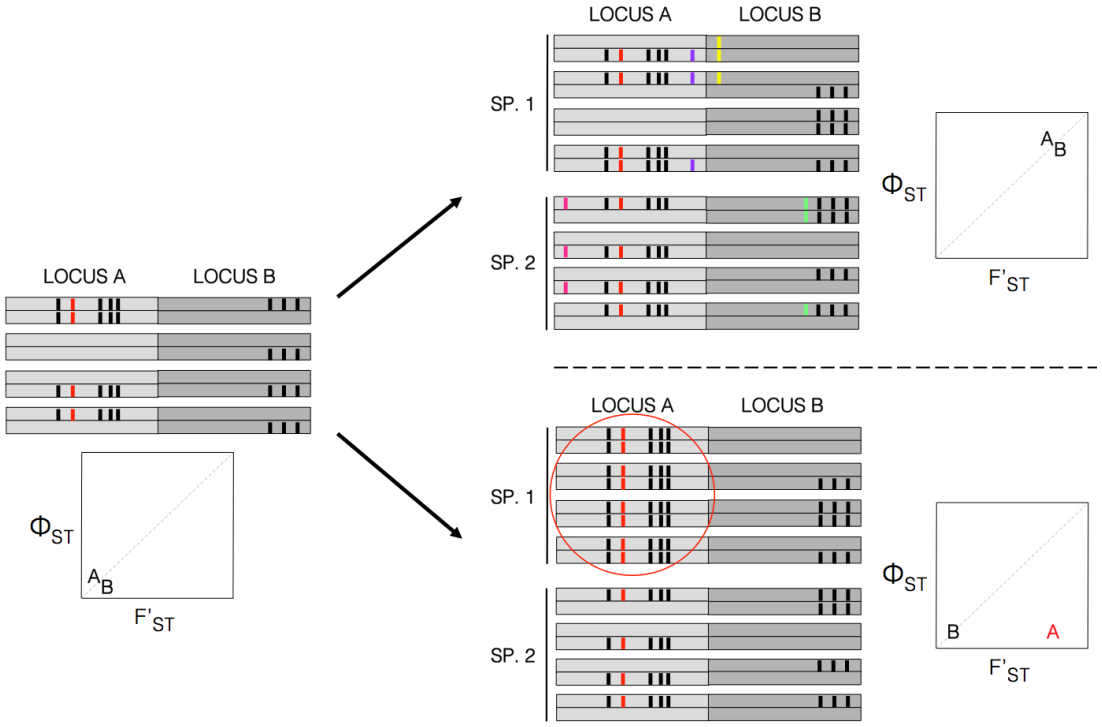


Figure 3.1 The F_{ST} analog Φ_{ST} has historically been used in population genetics, to estimate the degree of sequence divergence between two lineages (sp. 1 and 2) via the neutral accumulation of new mutations after they become physically and/or reproductively isolated (top right panel). Here we propose to use Φ_{ST} to detect loci exhibiting allele frequency differences (high F_{ST} or F'_{ST}) that do not show substantial sequence divergence between two lineages (bottom right panel) due to recent selective sweeps in one lineage (sp. 1) occurring on standing haplotype variation from a source population (left panel) that persists at moderate frequency in the lineage with which it is being compared (sp. 2). We illustrate this rationale using two different alleles on each of two loci (A and B), with each allele at 50% frequency (4/8) in a source population of four diploid individuals. The red allele on locus A is presumed to be selectively neutral, experiencing positive selection only in sp. 1 in the bottom right example.

Sweeps on standing variation in recently diverged swamp sparrows. The North American swamp sparrow (*Melospiza georgiana*) includes two subspecies that breed in inland freshwater (*M. g. georgiana*) or coastal brackish tidal marshes (*M. g. nigrescens*). Inland and coastal swamp sparrows are morphologically, ecologically and behaviourally distinct (Greenberg and Droege 1990) and common garden experiments have confirmed a genetic basis for their phenotypic differences (Ballentine and Greenberg 2010), despite a very recent divergence time. The northeastern and mid-atlantic coasts were under an ice sheet at the Last Glacial Maximum (LGM), and tidal marshes require stabilization of glacial outflow from coastal rivers and streams combined with sediment accretion from gradually rising sea levels in order to establish (Pethick 1984; Warren and Niering 1993), so these habitats, and the specialist swamp sparrows that now inhabit them, are less than 15,000 years young (Malamud-Roam et al. 2006). During the LGM, swamp sparrows likely occupied a freshwater refugial range, and population genomic data suggest that a founding group of birds harbouring a subset of the variation present in inland populations colonized and adapted to tidal marshes after the habitat became available (Chapter 1). Morphological data also strongly suggest that coastal adaptation occurred via selection on standing variation present in inland populations, because trait values for adaptive phenotypes in coastal birds are an exaggerated subset of the variation present in contemporary inland populations (Chapter 1; Greenberg and Droege 1990).

We have since constructed a reference genome for swamp sparrows and conducted a genome-wide scan that identified 31 genes with functional connections to coastal adaptation that bore signatures of selection in coastal populations. In addition to being F_{ST} outliers, these genes exhibited low π and/or runs of homozygosity, diagnostic of a recent selective sweep (Chapter 2). These data provide an opportunity to explicitly test whether F'_{ST} / Φ_{ST} comparisons using ddRAD data can reliably detect

regions known to have experienced selective sweeps. As a reduced representation method, ddRAD would not be expected to detect the full panel of known sweeps across the genome, but does this method provide a shortcut to identifying a subset of genes known to be under selection?

Methods

We previously generated a double-digest RADseq dataset (Peterson et al. 2012) for a population genetic study of swamp sparrows (N=92). In brief, this dataset included 4,256 loci from across the genome, each 137bp long, assembled using the STACKS pipeline (Catchen et al. 2011; refer to Chapter 1 for detailed methods). To test the suitability of F'_{ST} / Φ_{ST} comparisons for detecting recent sweeps from standing variation, we extracted estimates of these metrics for each locus (as well as F_{ST}, π and heterozygosity) from the output of STACKS.

We defined the predicted signal of a sweep from standing variation at (or near enough to experience linkage with) a given ddRAD locus as moderate divergence in F'_{ST} ($F'_{ST} \geq 0.5$), without an increase in Φ_{ST} above the maximum level estimated from loci that exhibited no divergence in F'_{ST} ($F'_{ST} = 0$). The rational for this threshold ($\Phi_{ST} \leq 0.04$ when $F'_{ST} = 0$ in swamp sparrows, N=798 loci) is based on the interpretation that shared ancestral sequence variation within two recently diverged lineages will result in non-zero estimates of Φ_{ST} for some loci even if those loci exhibit no allele frequency divergence as measured by F'_{ST}. A sweep occurring on a genetic background of shared variation would therefore be expected to yield Φ_{ST} values no greater than this threshold immediately after the sweep occurred.

Each ddRAD locus with F'_{ST} and Φ_{ST} estimates within these predicted bounds of a very recent sweep from standing variation was mapped to annotated scaffolds of the Medium Ground-finch assembly (*Geospiza fortis*, "GeoFor_1.0"; Zhang et al.

2012) using the *blastn* suite through NCBI and requiring an E-value of $1x10^{-5}$ to assign significant matches. We investigated a 250kb region around each significant *blast* hit to determine whether any of the genes that were previously identified as targets of selective sweeps in our whole genome scan (Chapter 2) were located near enough to the mapping site of the ddRAD locus for F'_{ST} / Φ_{ST} estimates to be capturing signatures of selection on that gene or regulatory elements adjacent to the gene via linkage. This is a conservative window since half of the conserved non-coding elements that regulate genes in the human genome are located more than 250 kb away (Vavouri et al. 2006).

At a broader scale, we calculated mean heterozygosities and patterns of allele frequency change for quantiles of the distribution of F'_{ST} and Φ_{ST} to test whether general trends in the data fit our predictions about the suitability of these metrics to capture the signal of a sweep from standing variation. We also tested the degree to which F'_{ST} and Φ_{ST} were correlated across our dataset by conducting a Pearson product-moment correlation.

Results

General trends in the distributions of F'_{ST} and Φ_{ST}. Both F_{ST} analogs that we estimated (F'_{ST} and Φ_{ST}) were significantly correlated across 4,256 RAD loci (ϱ=0.65, $p < 2.2x10^{-16}$). However, the mutational distance between inland and coastal haplotypes (Φ_{ST}) varied substantially across loci for any particular value of F'_{ST} (Figure 3.2). Estimates of Φ_{ST} ranged from $7x10^{-4}$ to 0.77 among highly differentiated loci in the top 5% of F'_{ST}. Estimates of heterozygosity for these same loci are consistent with our prediction that high F'_{ST} / low Φ_{ST} loci could be generated via a recent sweep on standing variation in coastal swamp sparrows. Inland and coastal swamp sparrows had similar levels of heterozygosity genome-wide (mean=0.06 for both, p=0.96), however

in coastal birds, loci in the top 5% of F'_{ST} with the smallest mutational distance from inland haplotypes (1st quartile of Φ_{ST}) had lower heterozygosity than those with larger mutational distances from inland haplotypes (4th quartile of Φ_{ST}) ($p<0.05$) (Figure 3.3). The same comparison was not significant in inland swamp sparrows ($p=0.16$). SNPs on loci with low vs. high Φ_{ST} also showed different patterns of allele frequency change between allopatric inland and coastal sites. For loci in the top 5% of F'_{ST}, alleles on loci within the 1st quartile of Φ_{ST} values existed at significantly higher frequencies in inland populations than those on loci within the 4th quartile of Φ_{ST} values (Figure 3.4).

Loci within the predicted bounds for recent sweeps on standing variation. Four ddRAD loci contained the combined signal of high F'_{ST} and low Φ_{ST} that we hypothesized would be generated by a recent selective sweep from standing variation. All four also showed reduced heterozygosity and nucleotide diversity (π) in coastal populations compared to inland populations (Table 3.1). The locus with the lowest estimate of Φ_{ST} (locus 4612) among all high F'_{ST} loci ($F'_{ST} > 0.5$; N=101) mapped to a location 183kb away from the gene Noggin-2-Like (NOG2L), one of 31 genes that bore a signature of a selective sweep in our previous genome scan. The remainder blasted to annotated locations on the *G. fortis* assembly that did not overlap with regions previously known to have swept in coastal swamp sparrows (Table A3.1).

Table 3.1 Four ddRAD loci fell within the bounds predicted for a recent sweep on standing variation ($F'_{ST} > 0.5$, $\Phi_{ST} \leq 0.04$). All four also exhibited reduced coastal heterozygosity and nucleotide diversity relative to inland birds.

Locus	F'_{ST}	Φ_{ST}	ΔHet	$\Delta\pi$
4612	0.60	0.02	-0.006	-0.045
1965	0.67	0.04	-0.0048	-0.054
12128	0.51	0.04	-0.022	-0.084
3803	0.51	0.04	-0.001	-0.089

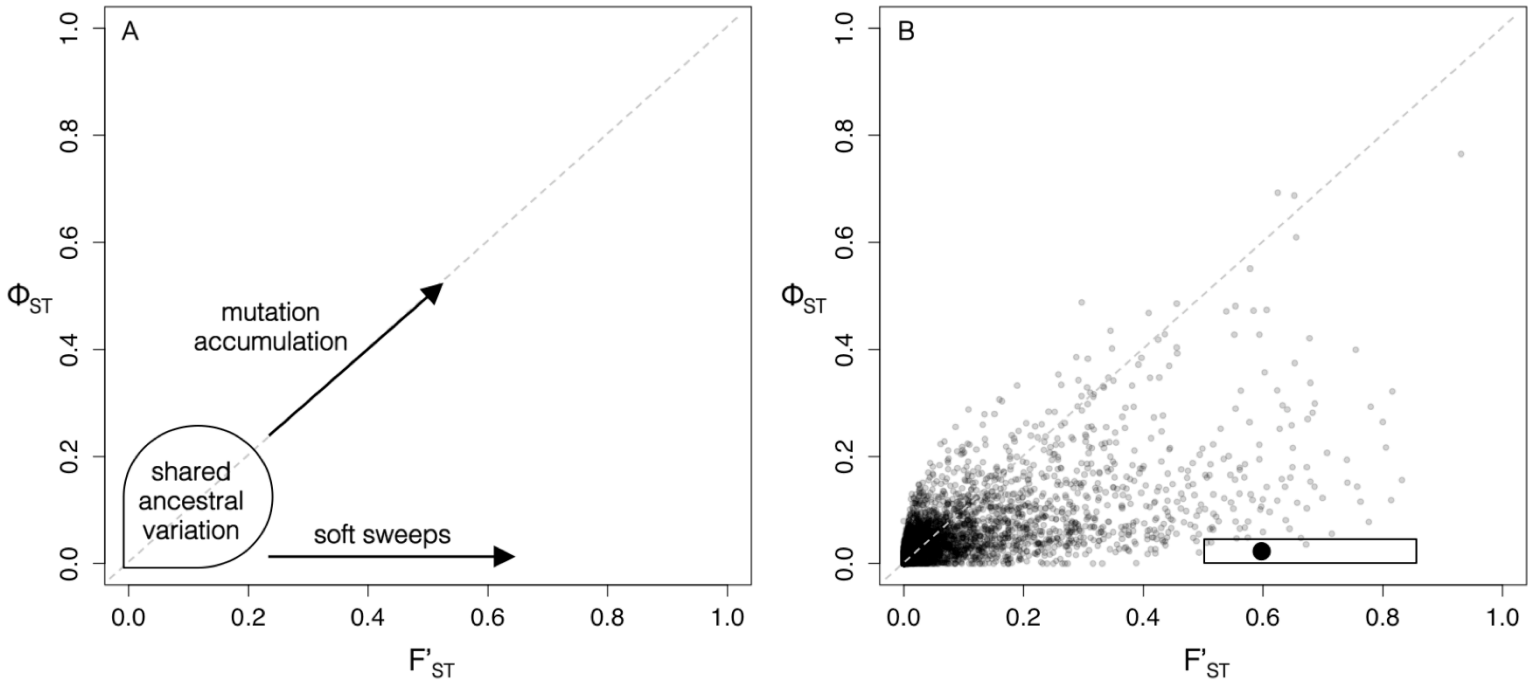


Figure 3.2 (A) We predicted that ddRAD loci linked to regions that have experienced recent sweeps on standing variation in one population would be characterized by low Φ_{ST} and high F'_{ST} estimates. (B) Comparison of F'_{ST} and Φ_{ST} for 4,256 loci between allopatric inland and coastal swamp sparrows revealed substantial Φ_{ST} variation among high F'_{ST} loci. Four loci exhibited divergence in F'_{ST} (≥ 0.5) without an increase in Φ_{ST} above the level present in shared ancestral variation ($\Phi_{ST} \leq 0.04$; black box), and one of these mapped near NOG2L (black circle), a gene with strong functional connections to coastal adaptation that bore the signature of a recent selective sweep in a previous whole genome scan (Table 3.1; Chapter 2).

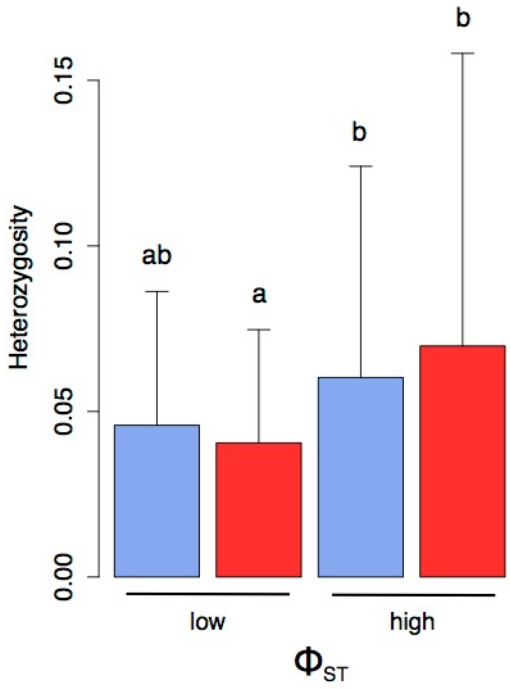


Figure 3.3 Estimates of heterozygosity are consistent with the prediction that high F'_{ST} / low Φ_{ST} loci capture the signal of a recent sweep. In coastal swamp sparrows (red bars), loci exhibiting the smallest mutational distance from inland haplotypes (1st quartile of Φ_{ST}, "low") had lower heterozygosity than those with larger mutational distances from inland haplotypes (4th quartile of Φ_{ST}, "high") ($p<0.05$). The same comparison was not significantly different in inland swamp sparrows, a large and broadly distributed population not known to have experienced recent sweeps (blue bars; $p=0.16$).

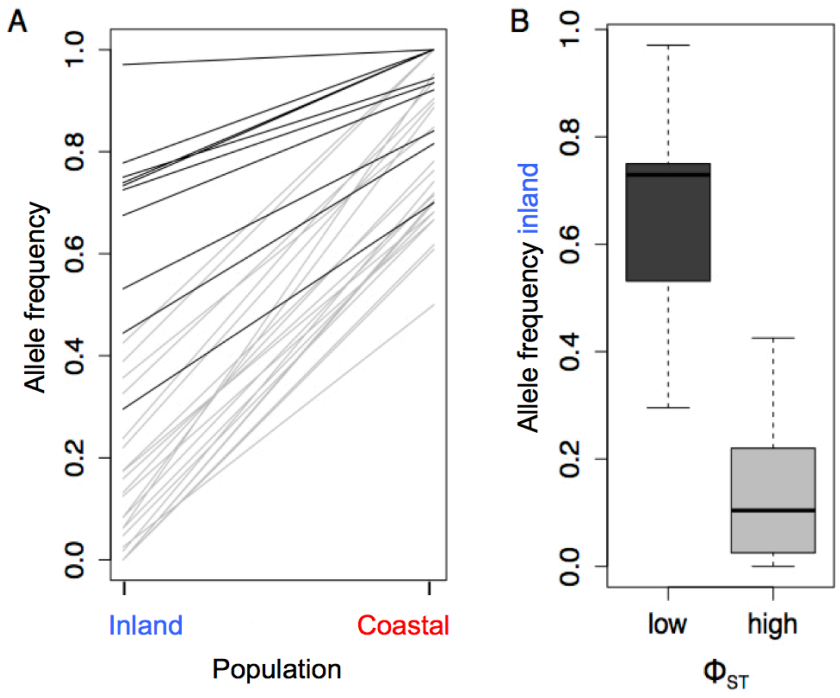


Figure 3.4 Allele frequencies in inland vs. coastal swamp sparrows provide additional evidence that high F'_{ST} / low Φ_{ST} loci are generated via sweeps from standing variation. Among highly divergent loci (those in the top 5% of F'_{ST}), alleles on loci with low sequence divergence from inland haplotypes (1st quartile of Φ_{ST}; black) are more common in that presumptive source population than loci with high sequence divergence from inland haplotypes (4th quartile of Φ_{ST}; grey) ($p=1.76 \times 10^{-5}$). Allele frequencies were calculated from the SNP with highest individual F_{ST} for each locus.

Discussion

F'_{ST} and Φ_{ST} comparisons with ddRAD data do detect selective sweeps on standing variation. We predicted that loci that have experienced a recent sweep on standing variation in coastal swamp sparrows would exhibit high F'_{ST} and low Φ_{ST} between inland and coastal populations, because a sweep on standing variation would drive a change in allele frequencies without an associated change in the degree of sequence divergence between inland and coastal haplotypes. We identified four high F'_{ST} loci with estimates of Φ_{ST} within the same range as loci that exhibit no haplotype differentiation between inland and coastal swamp sparrows. This interval represents a conservative approximation of the genomic background Φ_{ST} between the populations, or the amount of mutational distance present among shared ancestral haplotypes. Reduced heterozygosity and nucleotide diversity in coastal birds at these loci were also consistent with recent sweeps. The locus with the lowest overall estimate of Φ_{ST} among all high F'_{ST} loci mapped near the gene NOG2L, a gene known to have experienced a recent selective sweep from our previous genome scan. Assuming that each 137bp ddRAD locus represents a random sample from across the genome, the odds that one of four loci would randomly map within 250kb of the 0.3% portion of the genome bearing signatures of selection (3.8Mb in a 1.2Gb genome) are roughly 1 in 20. NOG2L plays a critical role in adaptation to heat stress in coastal birds, and is likely under strong selection (Chapter 2). Coastal swamp sparrows experience greater heat stress in comparison to inland swamp sparrows because coastal habitats offer little cover, and higher salinity drinking water means that they cannot engage in cooling methods that waste water, such as panting. Bill size is under strong selection because vascular structures within the bill allow the bill to act as a radiator, dissipating heat without wasting water. Larger bills make better radiators (Greenberg et al. 2012). NOG2L encodes an antagonist of bone morphogenetic protein 4 (BMP4; McMahon et

al. 1998; Groppe et al. 2002), and is likely responsible for establishing morphogenic gradients during development that influence adult bill size (Jones and Smith 1998).

Patterns of allele frequency change and heterozygosity in swamp sparrows provide additional evidence that F'_{ST} / Φ_{ST} comparisons do detect soft sweeps from standing variation in ddRAD data. Among high F_{ST} loci, those at either extreme of Φ_{ST} (1^{st} vs. 4^{th} quartile) exhibited different patterns of allele frequency change between inland and coastal swamp sparrows: alleles on loci with low Φ_{ST} existed at significantly higher frequencies in inland populations than those on loci with high Φ_{ST}. This is consistent with our prediction that variants on low Φ_{ST} ddRAD loci exist as standing variation in inland swamp sparrows, and have risen in frequency in coastal swamp sparrows in response to sweeping selection on nearby genes. Patterns of variation in heterozygosity across loci were also consistent with our prediction that haplotypes that have swept in coastal populations originated from inland standing variation: in coastal swamp sparrows, loci that more closely resembled inland haplotypes had lower heterozygosity than those that were more sequence divergent from inland haplotypes.

Why didn't the other three high F'_{ST} / low Φ_{ST} ddRAD loci map to known sweep locations from the genome scan? First, these loci may represent sweeps from standing variation that haven't swept to high enough frequency to be detected as outliers via the coarse sliding window resolution and stringent F_{ST} threshold we used in the genome scan. Second, they may be products of background purifying selection, although of the four only locus 4612 (near to NOG2L) is known to be located in a centromeric region (Table A3.1).

Accounting for background selection when applying this method. Background selection in genome regions with low recombination (eg. inversions or centromeric regions) could drive allele frequency differences higher while keeping the mutational distance between haplotypes low. This represents a confounding variable that must be accounted for in applying F'_{ST} / Φ_{ST} comparisons for the detection of sweeps from standing variation in other species or biological systems. In swamp sparrows, divergence between inland and coastal lineages is so recent that it is unlikely that there has been sufficient time for this process to generate the molecular signatures observed (Wolf and Ellegren 2016). Although some genes with signatures of sweeping selection and functional connections to adaptive phenotypes do broadly coincide with centromeric regions in the zebra finch (including NOG2L), the region bearing a signature of a sweep is small relative to the putative region of reduced combination around centromeres. Instead of reduced recombination creating spurious signals of selection, the sum of evidence from genome-wide data suggests that reduced recombination at these genes has helped maintain adaptive divergence between inland and coastal swamp sparrows in the face of active contemporary gene flow (Turner et al. 2005; Noor and Bennett 2009; Chapters 1 and 2). Given the results in swamp sparrows, we suspect that regions of reduced recombination may likewise not pose a substantial obstacle to accurate inference if F'_{ST} / Φ_{ST} comparisons are applied to other recently diverged taxa that experience similarly strong divergent selection.

Detection of a subset of known sweep regions. Comparison of F'_{ST} and Φ_{ST} with ddRAD data detected only one of 31 known genes exhibiting evidence of a selective sweep across the swamp sparrow. Although this single positive detection is evidence of the suitability of the method, why didn't it detect a larger proportion of known sweeps? First, the bounds we established to define shared ancestral variation were

very stringent, particularly in terms of Φ_{ST}. In the strictest sense, a sweep occurring on a genetic background of shared variation would be expected to yield Φ_{ST} values no greater than that present in shared ancestral variation immediately after the sweep occurred. However, some time has passed since coastal colonization and many loci that swept during coastal colonization and adaptation may have since accrued some additional sequence divergence in allopatry. For example, ddRAD locus 6803 was characterized by high F'_{ST} and low Φ_{ST} (F'_{ST}=0.57, Φ_{ST}=0.07), but fell just outside the Φ_{ST} threshold representing strictly shared variation. This locus mapped directly to a 70kb wide F_{ST} peak with reduced nucleotide diversity from our previous genome scan (Chapter 2). This peak was located near the gene carbonic anhydrase 5A (CA5A) which acts during ureagenesis to control excretory water balance, and represents a mechanism of coastal adaptation via salinity tolerance (Supuran 2008; Shah et al. 2013).

Additionally, reduced representation methods, and particularly ddRAD sequencing, represent very sparse sampling of the genome. A method such as the one proposed, that use reduced representation data for an easy and affordable peek at patterns of genomic variation, cannot be expected to offer comprehensive detection of all sweeps across the genome. Furthermore, not all of the sweeps detected in our previous whole genome scan are likely to be sweeps of the type best detected by F'_{ST} and Φ_{ST} (i.e. sweeps on moderately common standing variants that also persist at moderately high frequency in inland populations). These may represent sweeps from rare standing variants, or hard sweeps on novel mutations that arose in coastal populations, for which we would not predict low Φ_{ST}.

Evidence of a diversity of molecular processes during divergence. Many loci in our ddRAD dataset exhibited high estimates of both F'_{ST} and Φ_{ST}, demonstrating that, for

many regions of the swamp sparrow genome, allele frequency divergence (scaled by the variation present at each locus) has been accompanied by sequence divergence. This pattern, in contrast to the pattern expected for sweeps from standing variation, would be consistent with selection on novel mutations that have arisen in either inland or coastal birds. The variation we detect in F'_{ST} and Φ_{ST} reveal that diverse molecular, demographic or selective processes have acted across swamp sparrow genomes during divergence.

Φ_{ST} as a useful additional metric for whole genome scans. We cast F'_{ST} / Φ_{ST} comparisons as a useful method for In whole genome scans for signatures of selection in recently diverged taxa, divergence in Φ_{ST} (or lack thereof) could provide a statistical method for differentiating between sweeps that have occurred from standing variation in a source population, and sweeps that have occurred on novel variants. This represents a novel insight not captured by the standard statistics applied to genome scan data, since the molecular signals of different categories of sweeps are conflated when patterns of genomic divergence are estimated using F_{ST}, π, linkage disequilibrium or measures of haplotype homozygosity (Figure 3.5).

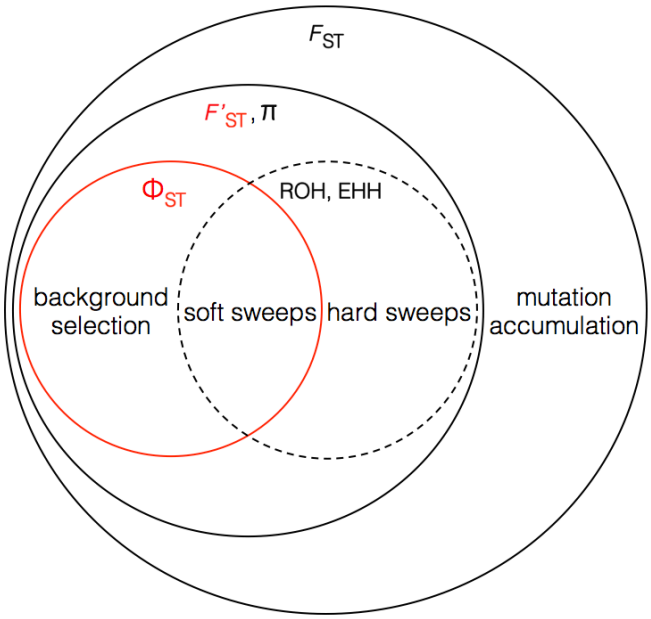


Figure 3.5 Different evolutionary processes are conflated in terms of the molecular statistics most commonly applied to detect selection from genomic data. F'_{ST} captures a combined signal of allele frequency differentiation (eg. F_{ST}) and nucleotide diversity (eg. π), providing a useful shortcut for differentiating processes that reduce sequence variation (eg. selective sweeps) from those that increase it (eg. the neutral accumulation of mutations in allopatry). Φ_{ST} can differentiate between sweeps on standing variation (soft sweeps) vs. novel mutations (hard sweeps) in cases of recent divergence when the source population of standing variation is known, and when the standing variant persists at moderate frequency in the source population (Figure 3.1). The dashed circle represents statistics that can identify sweeps by quantifying haplotype homozygosity (runs of homozygosity, ROH; extended haplotype homozygosity, EHH), but these cannot be estimated from short read length data like RADseq.

Conclusions. Our previous studies of inland and coastal swamp sparrows elucidated an important role for selective sweeps from standing variation during coastal adaptation, since adaptation has been rapid, and the distribution of phenotypic trait values and genome-wide variation among coastal birds is an adaptive subset of the variation present among inland birds (Greenberg and Droege 1990; Chapters 1 and 2). Genetic variation conferring an adaptive benefit in coastal environments during interglacials may have been maintained at moderate frequencies in inland refugial populations during glaciation. After glacial retreat, sweeps within coastal colonists adapting to newly available tidal marsh habitat generated divergence at genes encoding adaptive traits by reducing nucleotide diversity within the founding population of coastal birds. Although we originally described this signal via whole genome re-sequencing of inland and coastal populations (Chapter 2), we demonstrate here that we were able to detect this same molecular signal at one of the known locations of a selective sweep using ddRAD data and shortcut statistics.

Repeated formation and dissolution of hybrid zones in sync with major climatic oscillations are likely to have been common processes among species in temperate regions (Hewitt 2011; Payseur and Rieseberg 2016). If this process has been accompanied by periods of local adaptation followed by genomic homogenization, it could have contributed to the maintenance of adaptively relevant variation within lineages. Selective sweeps from standing variation may therefore be a relatively common mechanism of adaptation for temperate species that have undergone postglacial divergence. Indeed, mounting evidence suggests that soft sweeps may be the most common mode of adaptation overall (Messer and Petrov 2013). Comparison of Φ_{ST} and F'_{ST} may therefore represent a widely applicable tool for cheaply and easily scanning for candidate functional loci in species suspected of adaptive divergence from standing genetic variation.

REFERENCES

Ballentine, B., and R. Greenberg. 2010. Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. PLoS One. 5:e10229.

Barrett, R. D., and D. Schluter D. 2008. Adaptation from standing genetic variation. Trends in Ecology and Evolution. 23(1):38-44.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annual Review of Ecology, Evolution and Systematics. 16:113-148.

Berg, J. J., and G. Coop. 2015. A coalescent model for a sweep of a unique standing variant. Genetics. 201(2):707-25.

Bird, C. E., S. A. Karl, P. E. Smouse and R. J. Toonen. 2011. Detecting and measuring genetic differentiation. Pp. 31-55 *in* C. Heid, S. Koenemann and C. D. Schubart eds. Phylogeography and population genetics in Crustacea. CRC Press, Boca Raton, FL.

Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping. Trends in Ecology and Evolution. 23:686-694.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi and S. Hogner. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Research. 25(11):1656-1665.

Cao, J., K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, and X. Wang. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature Genetics. 43(10):956-63.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. Genes Genomes and Genetics. 1:171-182.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution. 15(5):538-43.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science. 307(5717):1928-33.

Cruikshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology. 23:3133-3157.

Dopman, E. B., L. Pérez, S. M. Bogdanowicz and R. G. Harrison. 2005. Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proceedings of the National Academy of Sciences USA. 102(41):14706-14711.

Excoffier, L., P. E. Smouse and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131(2):479-491.

Friesen, V. L., T. P. Birt, J. F. Piatt, R. T, Golightly, S. H. Newman, P. N. Hebert, B. C. Congdon, and G. Gissing. 2005. Population genetic structure and conservation of marbled murrelets (*Brachyramphus marmoratus*). Conservation Genetics. 6(4):607-14.

Gompert, Z., and C. A. Buerkle. 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Molecular Ecology Resources. 10:378-384.

Greenberg, R., and S. Droege. 1990. Adaptation to tidal marshes in breeding populations of the Swamp Sparrow. The Condor. 92:393-404.

Greenberg, R., R. Danner, B. Olsen and D. Luther. 2012. High summer temperature explains bill size variation in salt marsh sparrows. Ecography. 35:146-152.

Groppe J., J. Greenwald, E. Wiater, J. Rodriguez-Leon, A. N. Economides, W. Kwiatkowski, M. Affolter, W. W. Vale, J. C. Izpisua Belmonte and S. Choe. 2002. Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. Nature. 420(6916):636–42.

Harrison, R. G.. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma and J. Antonovics, eds. *Oxford surveys in evolutionary biology*. Oxford Univ. Press, New York, NY.

Hewitt, G. M. 1988. Hybrid zones – natural laboratories for evolutionary studies. Trends in Ecology and Evolution. 3:158-167.

Hewitt, G. M. 2011. Quaternary phylogeography: the roots of hybrid zones. Genetica. 139(5), 617-638.

Jones, C. M., and J. C. Smith. 1998. Establishment of a BMP-4 morphogen gradient by long-range inhibition. Developmental Biology. 194(1):12-7.

Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484, 55–61 (2012).

Kane, N. C., M. G. King, M. S. Barker, A. Raduski, S. Karrenberg, Y. Yatabe, S. J. Knapp and L. H. Rieseberg. 2009. Genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. Evolution. 63:2061-2075.

Lopes, R. J., J. D. Johnson, M. B. Toomey, M. S. Ferreira, P. M. Araujo, J. Melo-Ferreira, L. Andersson, G. E. Hill, J. C. Corbo, and M. Carneiro. 2016. Genetic basis for red coloration in birds. Current Biology. 26(11):1427-34.

Lowry, D. B., S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, and A. Storfer. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources. 17(2):142-52.

Machado, C. A., and J. Hey. 2003. The causes of phylogenetic conflict in a classic Drosophila species group. Proceedings of the Royal Society of London B. 270(1520):1193-1202.

Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science. 350(6267):1493-8.

McMahon, J. A., S. Takada, L. B. Zimmerman, C. M. Fan, R. M. Harland, and A. P. McMahon. 1998. Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. Genes and Development. 12(10):1438-52.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution. 60(11):2399-2402.

Meirmans, P. G., and P. W. Hedrick. 2010. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources. 11(1):5-18.

Messer, P. W., and D. A. Petrov. 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology and Evolution. 28(11):659-69.

McKinney, G. J., W. A. Larson, L. W. Seeb, and J. E. Seeb. 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. 2017. Molecular Ecology Resources. 17(3):356-61.

Nachman, M. W., and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B. 367:409-421.

Nadeau, N. J., and C. D. Jiggins. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends in Genetics. 26(11):484-492.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Reviews in Genetics. 39:197–218.

Noor, M. A., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nature Rev Genet. 7(11):851-861.

Noor, M. A., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity. 103(6):439-444.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources. 10:806-820.

Payseur, B. A., and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation. Molecular Ecology. 25:2337-2360.

Payseur, B. A., and M. W. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. Biological Journal of the Linnean Society. 84(3):523-534.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7:e37135.

Pethick, J. S. 1984. *An introduction to coastal geomorphology*. Deptartment of Geography, University of Hull, U.K.

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, and F. Hallböök. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 464(7288):587-91.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald and H.

C. Ackerman. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419(6909): 832-837.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G. P. Saetre, C. Bank, Å. Brännström and A. Brelsford. 2014. Genomics and the origin of species. Nature Reviews Genetics. 15(3):176-192.

Shah, G. N., T. S. Rubbelke, J. Hendin, H. Nguyen, A. Waheed, J. D. Shoemaker, and W. S. Sly. 2013. Targeted mutagenesis of mitochondrial carbonic anhydrases VA and VB implicates both enzymes in ammonia detoxification and glucose metabolism. Proceedings of the National Academy of Sciences. 110(18):7423-8.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genetical Research. 23(01):23-35.

Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nature Genetics. 43(11):1160-3.

Supuran, C. T. 2008. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. Nature Reviews Drug Discovery. 7(2):168-81.

Toews, D. P., L. Campagna, S. A. Taylor, C. N. Balakrishnan, D. T. Baldassarre, P. E. Deane-Coe, M. G. Harvey, D. M. Hooper, D. E. Irwin, C. D. Judy and N. A. Mason. 2015. Genomic approaches to understanding population divergence and speciation in birds. The Auk. 133(1):13-30.

Turner, T. L., M. W. Hahn and S. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biology. 3(9):e285.

Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Molecular Ecology. 26:3982-3997.

Vavouri, T., G. K. McEwen, A. Woolfe, W. R. Gilks and G. Elgar. 2006. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. Trends in Genetics. 22(1):5-10.

Vitti, J. J., S. R. Grossman and P. C. Sabeti. 2013. Detecting natural selection in genomic data. Annual Reviews in Genetics. 47:97-120.

Warren, R. S., and W. A. Niering. 1993. Vegetation change on a Northeast tidal marsh: interaction of sea-level rise and marsh accretion. Ecology. 74(1):96-103.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38(6):1358-1370.

Wolf, J. B., and H. Ellegren. 2016. Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics. 18:87-100.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19(3):395-420.

Wu, C. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology. 14:851-865.

Yeaman, S., and M. C. Whitlock. 2011. The genetic architecture of adaptation under migration–selection balance. Evolution. 65(7):1897-1911.

Zhang, G., P. Parker, B. Li, H. Li and J. Wang. 2012. The genome of Darwin's Finch (*Geospiza fortis*). Gigascience. 1:13.

CHAPTER 4

SALT MARSH MELANISM: CONVERGENT ADAPTATION OR MOLECULAR SPANDREL? DISENTANGLING THE EFFECTS OF SELECTION AND PLEIOTROPY ON MELANIC COASTAL SPARROW PHENOTYPES

Abstract

Examples of taxonomically broad phenotypic convergence are of interest to evolutionary biologists because they are often considered evidence that certain phenotypes have fundamental adaptive value in certain habitats. Salt marsh melanism is one such example, in which more than 10 different species of salt marsh birds, snakes, and mammals are darker than their closest inland relatives. This convergence has inspired a range of adaptive hypotheses about the benefit of melanism in salt marshes. We recently conducted a genomic selection scan in a species that exhibits salt marsh melanism, the swamp sparrow, and discovered robust signatures of selection on genes involved in both salt tolerance and melanogenesis in a population that has recently colonized coastal brackish habitats. Salt tolerance genes involved in intracellular vesicle trafficking were particularly well represented, indicating that natural selection has acted on the cellular life cycle of endosomes, lysosomes and vacuoles to excrete or sequester excess salt in the kidney. However vesicle trafficking pathways also influence melanosomes, raising the possibility that salt marsh melanism is a molecular "spandrel": an adaptively neutral pleiotropic consequence of selection for salt tolerance. To test this possibility, we deconstructed the melanic phenotype of coastal swamp sparrows into individual component traits, and used a naturally occurring zone of genetic and phenotypic admixture to quantify genotype-phenotype

associations at regions of the genome containing either single-effect melanogenesis genes or pleiotropic vesicle trafficking agents. We demonstrate that the overall melanic phenotype of coastal swamp sparrows is a product of both selection and spandrel effects. Enlargement of black plumage patches has been driven by selection on a melanin-specific transcription factor (BNC2), whereas increased melanism of the legs and feet is a systemic pleiotropic consequence of selection on a vesicle trafficking gene involved in salt tolerance (BLOC1S2). Therefore, in addition to adaptive hypotheses, adaptively neutral pleiotropy becomes a highly parsimonious explanation for the convergent evolution of melanism across the salt marsh species assemblage.

Introduction

Phenotypic convergence by taxonomically diverse species living in the same habitat is considered strong evidence of the adaptive value of that phenotype in that habitat. For example, the assemblage of lizards, mice, toads, crickets, grasshoppers, beetles and ants breeding in the white gypsum sand formation at White Sands, New Mexico exhibit blanched dorsal coloration compared to populations from neighboring darker-substrate habitats (Dice 1929, 1930; Smith 1943; Strohecker 1947; Stroud 1949, 1950). Now-classic studies on White Sands lizards demonstrated that strong selection for crypsis plays a critical role in maintaining this adaptive coloration, particularly in the face of ongoing gene flow from the surrounding habitat (Rosenblum 2005; Rosenblum and Harmon 2010).

Just as the White Sands species assemblage has converged on a less melanic phenotype, salt marsh species assemblages have converged on a more melanic phenotype. Coastal salt marsh populations of mice, shrews, voles, snakes, rails, wrens and sparrows are all more melanic than their closest inland relatives (Figure 4.1; Grinnell 1913; Von Bloeker 1932; Neill 1958; Pettus 1963; Conant and Lazell 1973;

Phillips 1986; Greenberg and Droege 1990; Gaul 1996; Luttrell et al. 2014). Salt marsh melanism and White Sands leucism are both the product of recent postglacial evolution in the absence of geographical barriers to gene flow (Malamud-Roam et al. 2006; Rosenblum 2005; Rosenblum and Harmon 2010), representing "yin" (dark) and "yang" (light) cases of rapid convergence. Melanic phenotypes are therefore assumed to be highly adaptive for salt marsh organisms (Greenberg et al. 1998), just as leucistic phenotypes are for White Sands organisms, but there is as yet no strong consensus on what the adaptive value of salt marsh melanism might be.

Brackish tidal estuaries are often characterized by grey-to-black sediments due to the presence of iron oxides, prompting hypotheses that salt marsh melanism enhances crypsis via background matching (VonBloeker 1932; Greenberg and Droege 1990). However differential predation rates for melanic and non-melanic forms have not been demonstrated in salt marshes. It is also unclear whether a diverse assemblage that includes small mammals, snakes and birds would be expected to experience similarly strong selection for background matching due to mortality risk from diurnal visual predators that occur in salt marshes. An alternative explanation pertaining to birds is that feather-degrading bacteria are more abundant in humid coastal habitats, and melanic feathers are better at resisting bacterial degradation (Peele et al. 2009). Territorial competition is also more intense for some birds in salt marshes, because the structure of the habitat drives higher breeding densities. Black plumage badges signal dominance, giving melanic males a competitive advantage during territory acquisition (Olsen et al. 2010). Although these last two factors may provide an adaptive explanation for melanism in salt marsh birds, they are insufficient to explain skin or pelage melanism exhibited by salt marsh snakes and mammals. The evolutionary processes responsible for convergence across such taxonomic breadth therefore remain somewhat mysterious.

Figure 4.1 Many salt marsh organisms have converged on a melanic phenotype, including (A) the gulf salt marsh snake (*Nerodia clarkii*), (B) the salt marsh harvest mouse (*Reithrodontomys raviventris*), (C) the clapper rail (*Rallus longirostris*), and (D) the extinct dusky seaside sparrow (*Ammodramus maritimus nigrescens*). All images from Wikipedia Commons: gulf salt marsh snake by Glenn Bartolotti, clapper rail by Magnus Manske.

The swamp sparrow is a North American species that has experienced postglacial divergence into phenotypically and genetically distinct inland freshwater and coastal brackish populations (Greenberg and Droege 1990; Malamud-Roam et al. 2006; Greenberg et al. 2016; Chapter 2), and the coastal form exhibits classic salt marsh melanism: the body feathers of coastal swamp sparrows are greyer and duskier than those of inland swamp sparrows (Greenberg and Droege 1990; Luttrell et al. 2014). Through measurements of museum specimens we have also recently documented that coastal swamp sparrows appear to exhibit increased melanism of their soft parts, including their bottom bill and their legs and feet (Chapter 1). Additionally, the plumage of coastal swamp sparrows is characterized by heavier black streaking and larger black patches (Greenberg and Droege 1990), including a conspicuous black forehead patch that has received robust empirical support as an adaptive dominance badge for male-male competition (Olsen et al. 2010). Inland birds do not experience strong selection for the dominance badge, and instead experience strong female preference for rusty crowns that act as a signal of good parental care. This sets up a tradeoff for the allocation of eumelanin (black) or phaeomelanin (rust) to crown plumage, and means that the size of the black forehead patch trait is under divergent directional selection in inland freshwater and coastal brackish populations (Olsen et al. 2010).

Each component of coastal swamp sparrow melanism mirrors the axes of convergence among other melanic salt marsh organisms. Overall dusky grey coloration in body plumage or pelage is the most broadly convergent trait, exhibited by song sparrows (Grinnell 1909; Marshall 1948; Nolan 1968), seaside and saltmarsh sparrows (Greenberg and Droege 1990), marsh wrens (Phillips 1986), and the clapper rail (Maley 2012), as well as all of the salt marsh mice, voles and shrews that exhibit melanism (Grinnell 1913; Von Bloeker 1932; Thaeler 1961; Wood et al. 1982). We

have observed that this is accompanied by increased melanism of the skin or other soft parts in many of these cases, including seaside and saltmarsh sparrows, but this pattern has yet to be quantified across taxa. In comparison, enlarged black patches are a less common form of salt marsh melanism, but are present in the Alameda song sparrow (Nolan 1968) and the dusky seaside sparrow (Baker 1973).

We recently conducted a genomic comparison of inland and coastal swamp sparrows and identified a set of 28 genes that had functional connections to physiological mechanisms of salinity tolerance or melanogenesis, and which all bore clear molecular signatures of recent selective sweeps. Vesicle trafficking genes comprised the largest functional category related to salinity tolerance, and included 10 different genes with the potential to influence absorption, excretion or storage of excess dietary salt by affecting rates of trafficking or turnover of intracellular vesicles like endosomes, lysosomes or vacuoles (Chapter 1). The apparent importance of vesicle trafficking for adaptation to salinity in swamp sparrows raises an interesting functional connection between salt tolerance and melanism, because melanosomes are also part of the vesicle network and are dependent on some of the same vesicle trafficking genes during development and maturation (Raposo and Marks 2007). For example, coastal swamp sparrows have experienced a selective sweep at the gene encoding a subunit of the Biogenesis of Lysosome-related Organelles Complex 1 (BLOC1; Chapter 2). BLOC1 mediates the development and trafficking of both kidney lysosomes and skin melanosomes (Theriault and Hurley 1970; Nguyen et al. 2001; Dell-Angelica 2004). Since vesicle trafficking can pleiotropically influence both kidney and melanosome function, we hypothesized that the convergent evolution of melanism in salt marsh taxa may, in fact, be a molecular "spandrel" (Gould and Lewontin 1979; Barrett and Hoekstra 2011): an adaptively neutral consequence of selection for salinity tolerance via vesicle trafficking mechanisms.

We test this hypothesis by conducting a genome-wide association study (GWAS) in the swamp sparrow, exploiting a naturally occurring zone of extensive admixture between inland and coastal swamp sparrows at the ecotone between freshwater and brackish habitats in New Jersey (Greenberg et al. 2016; Chapter 1). Hybrid zones provide powerful natural laboratories in which to conduct association tests because multi-generational gene flow and recombination have disrupted linkage (Barton and Hewitt 1985; Hewitt 1988; Harrison 1990; Wu 2001; Kane et al. 2009), allowing detection of functional variants at a fine resolution (Harrison and Larson 2014). If melanism is a pleiotropic by-product of selection for salinity tolerance, we predict significant associations between genotypes at pleiotropic vesicle trafficking genes and phenotype scores for both melanic and salt tolerance traits. However, significant associations between melanic phenotype scores and genotypes at single-effect melanogenesis genes would be consistent with the alternative hypothesis that melanism is indeed adaptive for coastal swamp sparrows.

Methods

Sample collection and phenotypic data. We had previously collected blood samples and phenotypic data from 24 male swamp sparrows breeding in allopatric inland and coastal marshes (Chapter 1), and to these we added blood samples and phenotypic data from 24 breeding males at inland freshwater and coastal brackish marshes throughout the putative zone of admixture in New Jersey (Table 4.1; Figure 4.2). Breeding habitat salinity for each individual was scored as a binary phenotype, as either freshwater or brackish. We quantified the degree of melanism exhibited by each individual by photographing the crown, plumage, legs and bottom bill using a Canon DSLR camera against a standardized color reference target (Xrite Digital SG Colorchecker). We defined the degree of melanism of the crown as the proportion of the total crown that was black as opposed to rust, and we measured this using either digital calipers or a wing rule (N=36/47). Two different people collected these crown measurements (authors R.G and P.D.), but we standardized our approach prior to data collection using explicit anatomical landmarks. We quantified the degree of melanism of the legs, bill and plumage through comparison with Munsell soil color charts (N=16). Variation in leg color was assigned according to variation in value within hue 7.5, regardless of chroma. Variation in bill color was assigned according to variation in chroma within hue 10 and value 6. Because the body plumage of swamp sparrows is highly heterogeneous, the overall degree of melanism of the plumage was inferred from the tail feathers, and was assigned according to variation in chroma within hue 10 and value 4. We tested whether these four components of the melanic phenotype were dependent or independent of one another using a spearman rank correlation test. We also collected a set of identical feathers from each individual (three crown feathers, the 3rd and 4th greater coverts, three right side tail coverts, and the 5th secondary from the left wing) for Munsell color verification under even more

controlled light conditions in the lab (N=16). Blood samples were preserved in Queen's lysis buffer and stored at -20°C.

Table 4.1 Locations and dates for swamp sparrows sampled in this study (N=48).

Code	Sampling locations	N	Date
HC	Hamilton County, NY	1	2013
TC	Tompkins County, NY	7	2008-2014
ER	Erie National Wildlife Refuge, PA	2	2001
GC	Garrett County, MD	2	2001
ME	Meadowlands, NJ	4	2001
GS	Great Swamp National Wildlife Refuge, NJ	6	2001
CH	Cheesequake State Park, NJ	3	2001
AP	Assunpink Wildlife Management Area, NJ	2	2015
WH	Whitesbog Preservation Trust, NJ	2	2015
MA	Manahawkin Wildlife Management Area, NJ	1	2015
BR	Bass River, Edwin B. Forsythe National Wildlife Refuge, NJ	6	2016
EH	Egg Harbor, NJ	1	2001
PM	Port Mahone, DE	4	2001
FB	Fishing Bay Wildlife Management Area, MD	7	2014

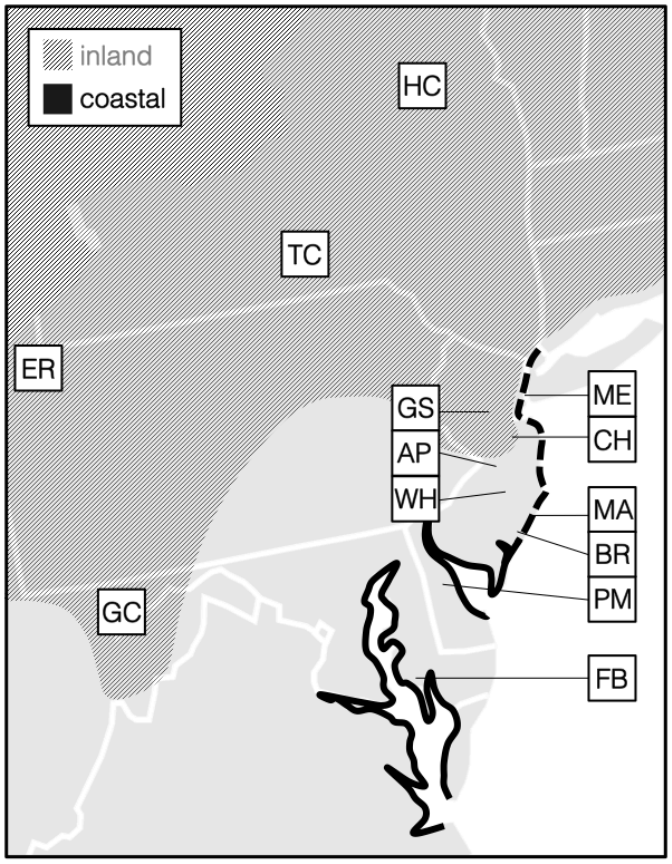


Figure 4.2 Breeding range of inland freshwater (grey) and coastal brackish (black) swamp sparrows (Greenberg and Droege 1990; Beadell et al. 2003). Sampling sites are shown with two-letter site codes (Table 4.1).

Genome resequencing and genotyping. We extracted DNA using the DNeasy kit (Qiagen) and quantified concentrations on a Qubit (Invitrogen, Carlsbad, CA, USA). We balanced concentrations to approximately 2 ng/ul across samples using either dilution or vacuum filtration, and used approximately 105 ng total to prepare individual whole genome libraries in 350bp sonicated fragments using the Illumina TruSeq Nano library preparation kit. We performed paired-end sequencing of all 48 libraries on two lanes of an Illumina NextSeq500 platform (2x150bp; allopatric and admixed samples were run on separate lanes). One sample from Egg Harbor NJ did not result in high quality reads and was dropped prior to read processing and genotyping. We used the program AdapterRemoval (Lindgreen 2012) to trim uncalled sites from reads, truncate read pairs that contain adapter sequence, and collapse any overlapping read pairs into a single read. We required that every read have a Phred quality score and minimum length of 20.

Bowtie2, a gapped-read aligner (Langmead and Salzberg 2012), was used to map paired reads and unpaired collapsed reads from each individual to our coastal swamp sparrow reference genome assembly (Chapter 2) using the "very sensitive local" flag, which applies the following parameters to the alignment algorithm: a maximum effort of 20 seed extension attempts and 3 re-seed attempts, 0 mismatches allowed, a seed substring length of 20, and the interval between seed substrings defined by a square root function with a constant of 1 and coefficient of 0.5. We used Samtools (Li et al. 2009) to convert individual alignment files from sequence alignment/map format (SAM) to the binary version (BAM). We applied tools from the program Picard (Wysoker et al. 2012) to add sequencing group information to individual BAM files prior to genotyping, and coordinate contig names across the dataset. We flagged individual fragments that had been sequenced more than one time as duplicates, so that they did not mislead downstream genotyping, and indexed the

resulting flagged alignments.

Individual genotypes were assigned using the UnifiedGenotyper from Genome Analysis Toolkit (GATK; DePristo et al. 2011). When identifying variant sites and assigning genotypes we required a minimum Phred-scaled base quality score of 17, and a minimum confidence threshold of 10. We then used VCFtools (Danecek et al. 2011) to reduce the dataset to only those scaffolds known to contain candidate functional genes for salt tolerance or melanism (19 scaffolds, 2.5-23.5Mb in size, total target size 138Mb; Chapter 2), and to apply more stringent genotype filtering, requiring that each SNP be biallelic, be genotyped for all samples (zero missing data), have a minimum mean depth of 3 reads, and have allele frequencies greater than 5% (to filter sequencing errors). In total, 526,447 SNPs met these criteria, allowing us to confidently assign genotypes at a density of roughly 1 SNP / 262bp.

Testing genotype-phenotype associations. We tested for genotype-phenotype associations across scaffolds containing candidate salt tolerance, melanogenesis or pleiotropic genes using the program PLINK (Purcell et al. 2007). Associations with salt tolerance were tested using a binary case-control framework, where breeding in a saline habitat was considered a "case" and breeding in a freshwater habitat was considered a "control". Significant associations between particular genotypes and breeding in saline habitats were determined via an asymptotic p-value. Associations with melanic traits (black crown patch size, tail plumage, leg color and bill color) were tested in a quantitative framework, in which significant associations were determined via the Wald test statistic. In each case we required a threshold of $p=1x10^{-5}$ to assign significant associations.

Results

Significant associations between salt tolerance genes and habitat salinity. Four of the 23 salt tolerance genes that bore signatures of selection in our genome scan (17%) exhibited significant associations with habitat salinity (BLOC1S2, ATP6V1G3, DAB2, RAMP3; Figure 4.3, Figures A4.1-4.3). Two of these (BLOC1S2 and DAB2) are involved in vesicle trafficking, and have functional implications for both salt tolerance and melanism (Theriault and Hurley 1970; Nguyen et al. 2001; Wine 2003; Dell-Angelica 2004).

Variation in components of the melanic phenotype. We documented variation in all four melanic traits in swamp sparrows (Figure 4.4), with black crown patch size and leg color exhibiting more variation than tail plumage or bill color. The size of the black crown patch ranged from 31% to 65% of the total crown (mean 52%), and leg color spanned five value categories of the Munsell color system (mean value 3.6). Only crown and leg traits exhibited sufficient variation to test for correlations between traits, and they varied independently from one another ($p=0.33$).

Significant associations with melanic traits. Three of the nine melanin-associated genes identified from our previous genome scan exhibited significant associations with black crown size, body color, bill color or leg color. Genotypes at a SNP in the intron of a gene adjacent to the melanocyte transcription factor Basonuclin 2 (BNC2) strongly predicted the size of the black crown patch of swamp sparrows (Figure 4.5A). All birds with an AA genotype at this site had large black crown patches (>40% of the crown), whereas all birds with one or more copies of the G allele had smaller black patches (<40% of the crown). Genotypes at the high-order signaling factor Delta Catenin (CTNND1) also exhibited significant associations with black crown size

(Figure A4.4). The pleiotropic vesicle trafficking gene BLOC1S2 exhibited significant associations with both breeding habitat salinity and a melanic trait, and was the only gene in our dataset for which this was the case. The genotype of birds at an intergenic site approximately 250kb downstream from BLOC1S2 strongly predicted the degree of melanism in the legs and feet of swamp sparrows (Figure 4.5B). Birds with an AA genotype exhibited the darkest leg coloration (corresponding to Munsell color values 2.5 or 3), and the single bird we genotyped as homozygous for the alternative allele (GG) exhibited the lightest leg coloration (Munsell color value 6). All birds with leg color corresponding to intermediate Munsell value 4 were heterozygotes with an AG genotype.

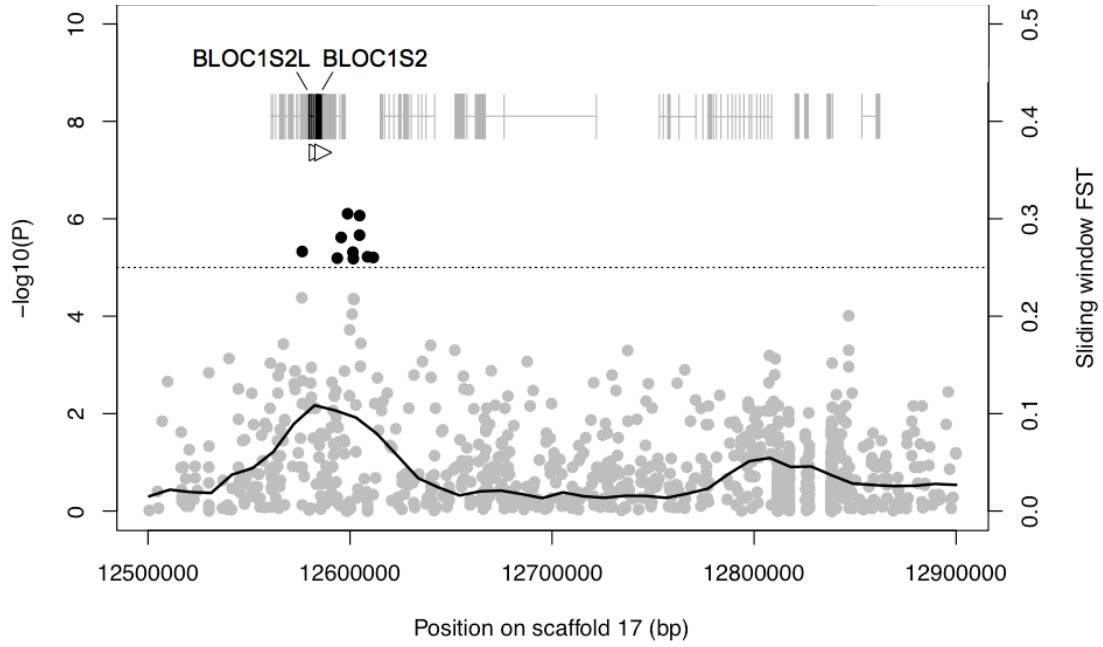


Figure 4.3 Association of genotypes near the vesicle trafficking gene BLOC1S2, with breeding habitat salinity. Dotted line denotes a significance threshold of $p = 1x10^{-5}$. Arrow indicates the location and direction of the transcription start site of each candidate functional gene, with exons represented by vertical bars. Adjacent genes without functional connections to salt tolerance or melanism are shown in grey (refer to Table A4.1 for gene identities). A gene sequence highly similar to BLOC1S2 (BLOC1S2L) is located immediately upstream of the gene (NCBI).

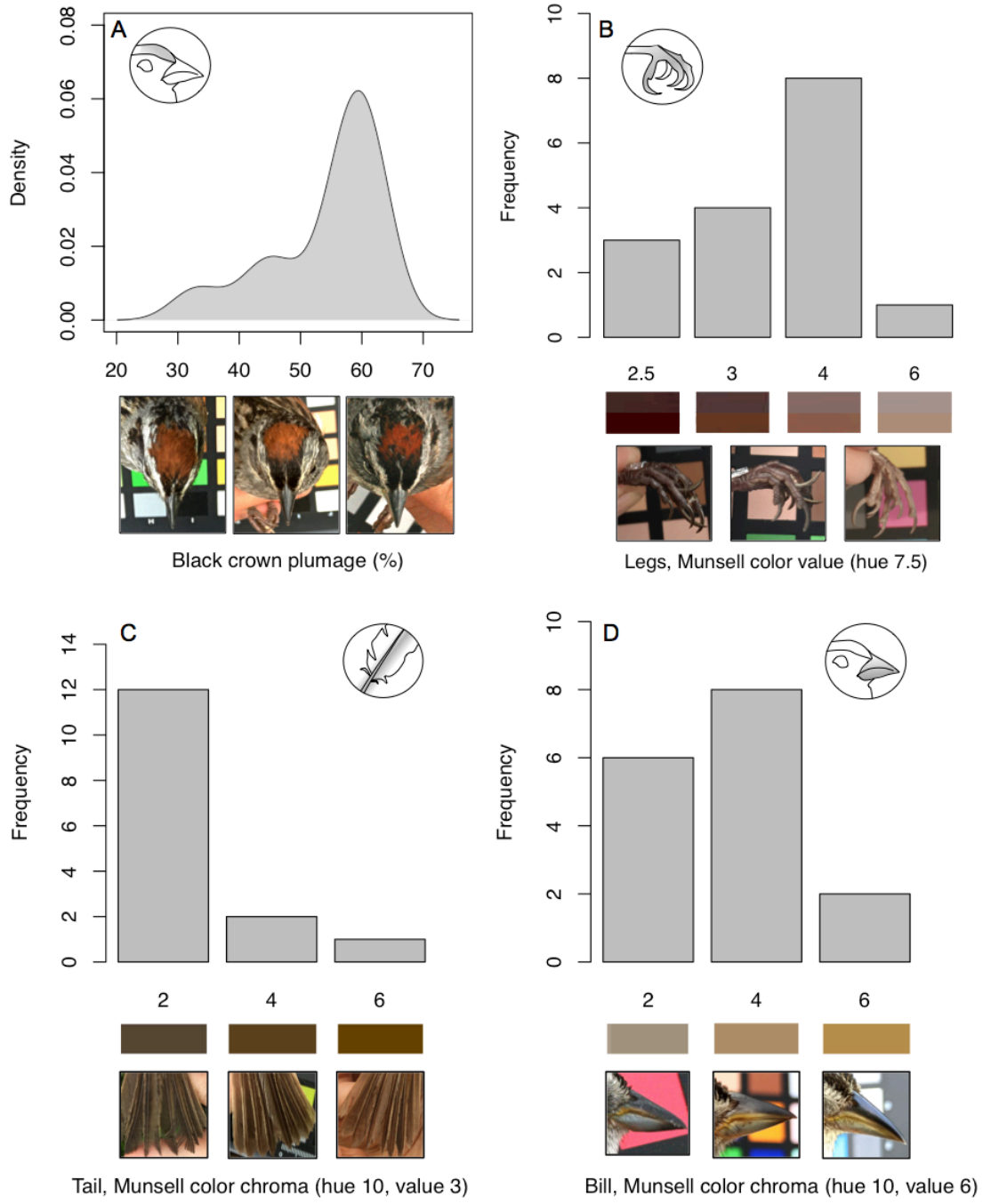


Figure 4.4 Variation in four melanic phenotypes in swamp sparrows: (A) black crown patch size, (B) color of the legs and feet, (C) chroma of the tail feathers, and (D) chroma of the bottom bill.

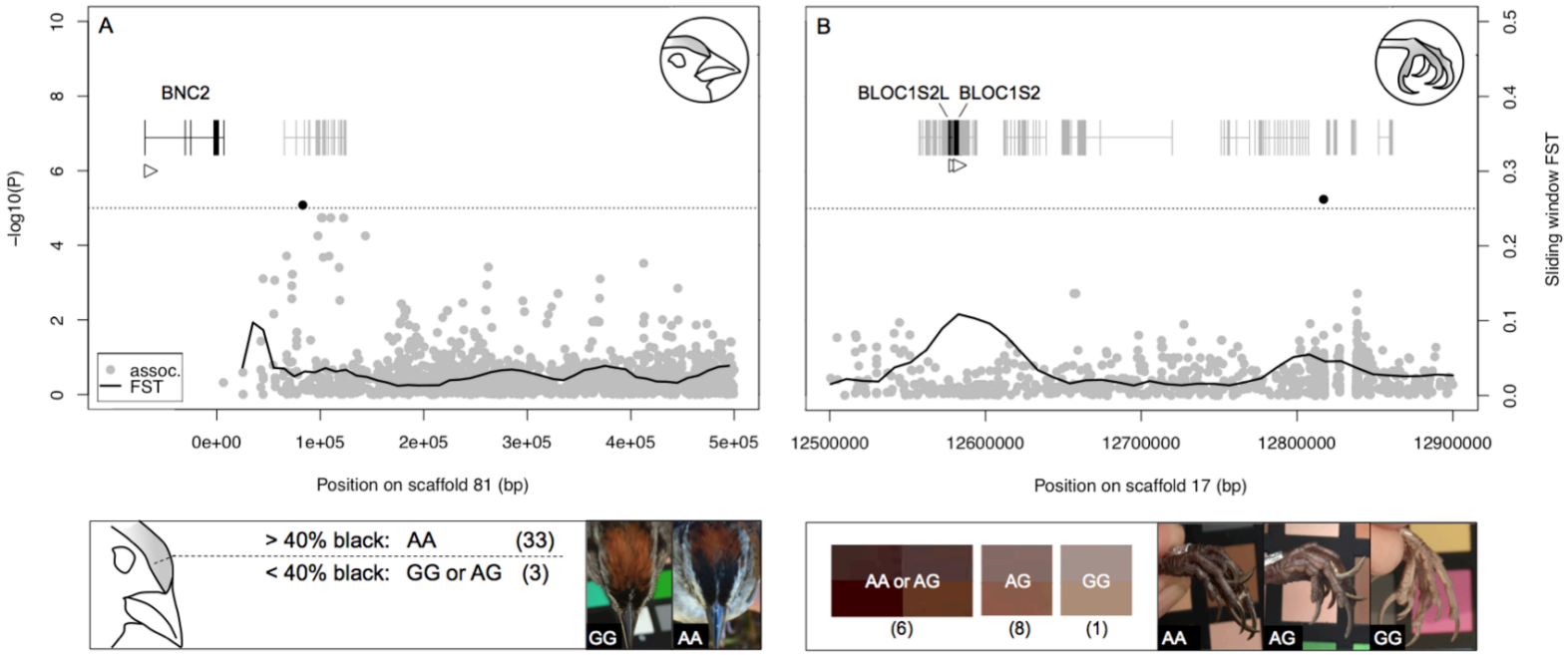


Figure 4.5 Genotype-phenotype associations at sites near: (A) the melanocyte transcription factor BNC2, with black crown patch size, and (B) the pleiotropic vesicle trafficking gene BLOC1S2, with leg color. Dotted lines in association plots denote a significance threshold of $p=1x10^{-5}$, and SNPs that fall above this threshold are indicated in black. Arrow indicates the location and direction of the transcription start site of each candidate functional gene, with exons represented by vertical bars. Adjacent genes without functional connections to salt tolerance or melanism are shown in grey (refer to Table A4.1 for gene identities). Numbers in brackets indicate individuals with each genotype-phenotype combination.

Discussion

Melanism as a product of positive selection. Genotypes in an intron of the gene adjacent to BNC2 strongly predicted the size of the black crown patch in swamp sparrows. BNC2 is a transcription factor that operates specifically within melanocytes, and it is one of the main predictors of melanic skin pigmentation in humans (Jacobs et al. 2013). A haplotype responsible for a reduction in melanic pigmentation was introduced into the human lineage via Neanderthal introgression during early northern range expansion (Vernot and Akey 2014). BNC2 is a particularly strong candidate for the functional gene underlying the enlarged black crown patch of coastal swamp sparrows because the size of this patch is sexually dimorphic (Olsen et al. 2010), and BNC2 is on the Z chromosome in the annotated Zebra Finch genome assembly (NCBI). The downstream SNP associated with the melanic crown phenotype of swamp sparrows may therefore represent (or be very closely linked to) a distal secondary enhancer of BNC2. Alternatively, plumage color in swamp sparrows may be regulated by the same mechanisms as melanic pigmentation in humans: via a distal sequence motif that influences expression of BNC2 by controlling chromatin accessibility at the primary enhancer (Visser et al. 2014).

Melanism as a by-product of salt tolerance. The genomic region containing the vesicle trafficking gene BLOC1S2 was associated with both habitat salinity and melanism of the legs and feet of swamp sparrows. Genotypes at eight SNPs immediately downstream of BLOC1S2, and one SNP immediately upstream of an adjacent BLOC1S2 duplication, strongly predicted whether sparrows bred in freshwater or saltwater habitats. These SNPs are presumably non-synonymous polymorphisms (or very closely linked to such polymorphisms) in proximal regulatory elements that confer different physiological tolerances for salinity due to the role of the BLOC1

complex in kidney lysosome trafficking (Nazarian et al. 2003; Gautam et al. 2004; Dell-Angelica 2004).

The genotype of an individual at an intergenic site further downstream from BLOC1S2 also strongly predicted the degree of melanism in the legs and feet of swamp sparrows. In addition to regulating vesicle traffic in the kidney, the BLOC1 complex plays a central role in vesicle trafficking during melanosome development. BLOC1-knockout mice exhibit a dramatic reduction in the number and size of mature melanosomes in the skin, leading to a systemic *pallid* blond phenotype that extends to the skin of the ears, feet and tail as well as the fur (Theriault and Hurley 1970; Nguyen et al. 2001; Dell-Angelica 2004). The SNP associated with leg color is a distance from the gene body of BLOC1S2, but there are no other genes with functional connections to melanism in the region, and *cis*-regulation on this scale is common in the well-studied human genome (Vavouri et al. 2006). This SNP may therefore represent (or be closely linked to) a functional change at a distal enhancer of BLOC1S2, or some other type of distal regulatory element.

Although both are located in the region of BLOC1S2, different SNPs associate with habitat salinity and melanism. We did not detect any individual SNP that significantly associated with both traits, which would be the expectation under pleiotropy in the strictest sense. Instead, sites proximal to the gene appear to determine salinity tolerance and a site farther downstream determines leg color. Selection on salt tolerance SNPs could nevertheless influence the SNP for leg color through linkage. If a selective sweep occurred on one or more of the functional SNPs for salt tolerance, the SNP for leg color could sweep with it because it is nearby. Even if that linkage association has been broken up in the present day by hybridization and recombination, past genetic hitchhiking could therefore have led to an increase in melanism that was an adaptively neutral by-product of selection for salt tolerance.

Limitations posed by the habitat salinity phenotype. Because molecular signatures of selection were originally detected through genomic comparisons of allopatric populations of inland freshwater and coastal brackish swamp sparrows, there is the risk of some circularity in testing for associations between genotypes under those F_{ST} peaks and "breeding habitat salinity" (scored as a binary trait) using a panel of birds that included those allopatric individuals. The association was no longer significant when allopatric individuals were excluded, but it is unclear whether this is the effect of a loss of detection power due to the reduced size of the association panel, particularly in terms of freshwater breeding individuals. Only 17% of F_{ST} peaks harboring candidate salt tolerance genes exhibited significant associations with habitat salinity, even when the association panel included allopatric as well as admixed individuals. In fact, only one of the eight largest F_{ST} peaks containing salt tolerance genes exhibited significant associations with habitat salinity (RAMP3), emphasizing that elevated F_{ST} captures a signal of genetic differentiation that is biologically distinct from a significant association between individual genotypes and phenotypes. The other three genes that did exhibit significant associations with breeding habitat salinity (ATP6V1G3, DAB2, BLOC1S2) were located under much smaller F_{ST} peaks.

Conclusions. Salt marsh melanism is a general pattern of phenotypic convergence, but even within a single species a melanic phenotype may be the product of change in several distinct components. More than one evolutionary process may therefore be responsible for convergence on an overall melanic form. In coastal swamp sparrows, increased melanism of the crown plumage has likely evolved via sexual selection, but increased melanism of the legs and feet is likely an adaptively neutral by-product of selection on vesicle trafficking genes that contribute to salt tolerance. Swamp sparrows therefore provide the first evidence that phenotypic convergence across the

salt marsh species assemblage may not be driven exclusively by selection directly on coloration, and suggests that explanatory hypotheses based on molecular pleiotropy are a parsimonious alternative to conventional adaptive hypotheses for some components of the overall melanic phenotype. In addition to providing a classic case study of convergence by natural selection, salt marsh melanism may now provide a novel window onto the under-appreciated role of pleiotropy in evolution.

REFERENCES

Barrett, R. D., and H. E. Hoekstra. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature Reviews Genetics. 12(11):767-80.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annual Review of Ecology and Systematics. 16:113-148.

Baker, J. L. 1973. Preliminary studies of the dusky seaside sparrow on the St. Johns National Wildlife Refuge. Proceedings of the Annual Conference of the Southeastern Association of Game Fishery Commissions. 27:207-214.

Beadell, J., R. Greenberg, S. Droege and J. A. Royle. 2003. Distribution, abundance, and habitat affinities of the Coastal Plain Swamp Sparrow. Wilson Bulletin. 115:38-44.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Dell'Angelica, E. C. 2004. The building BLOC(k)s of lysosomes and related organelles. Current Opinion in Cell Biology. 16(4):458-64.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. McKenna. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 43(5):491-498.

Dice, L. R. 1929. Description of two new pocket mice and a new woodrat from New Mexico. Occasional Papers of the Museum of Zoology, University of Michigan. 203:1-4.

Dice, L. R. 1930. Mammal distribution in the Alamogordo region, New Mexico. Occasional Papers of the Museum of Zoology, University of Michigan. 213:1-32.

Gaul, R. W. 1996. An investigation of the genetic and ecological status of the Carolina salt marsh snake, *Nerodia sipedon williamengelsi*. East Carolina University (Doctoral Dissertation).

Gautam, R., S. Chintala, W. Li, Q. Zhang, J. Tan, E. K. Novak, S. M. Di Pietro, E. C. Dell'Angelica, and R. T. Swank. 2004. The Hermansky-Pudlak syndrome 3 (cocoa) protein is a component of the biogenesis of lysosome-related organelles complex-2 (BLOC-2). Journal of Biological Chemistry. 279(13):12935-42.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London. Series B, Biological Sciences 205(1161): 581-98.

Greenberg, R., K. M. Cammen, A. G. Wilson, B. H. Olsen, B. Ballentine, N. C. McInerney, and R. C. Fleischer. 2016. Geographic population structure and subspecific boundaries in a tidal marsh sparrow. Conservation Genetics. 17(3):603-13.

Greenberg, R., P. J. Cordero, S. Droege, and R. C. Fleischer. 1998. Morphological adaptation with no mitochondrial DNA differentiation in the coastal plain swamp sparrow. The Auk. 1:706-712.

Greenberg, R., and S. Droege. 1990. Adaptation to tidal marshes in breeding populations of the Swamp Sparrow. The Condor. 92:393-404.

Grinnell, J. 1909. Three new Song Sparrows from California. University of California Publications in Zoology. 5:265-269.

Grinnell, J. 1913. The species of the mammalian genus *Sorex* of west-central California with a note on the vertebrate palustrine fauna of the region. University of California Publications in Zoology. 20:179-205.

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. Journal of Heredity. 105:795-809.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Pp. 69-128 *in* D. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology. Oxford Univ. Press, New York, NY.

Hewitt, G. M. 1988. Hybrid zones – natural laboratories for evolutionary studies. Trends in Ecology and Evolution. 3:158-167.

Kane, N. C., M. G. King, M. S. Barker, A. Raduski, S. Karrenberg, Y. Yatabe, S. J. Knapp, and L. H. Rieseberg. 2009. Genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. Evolution. 63:2061-2075.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359.

Lazell Jr, J. D., and J. A. Musick. 1973. The Kingsnake, *Lampropeltis getulus sticticeps*, and the ecology of the Outer Banks of North Carolina. Copeia. 28:497-503.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genomes Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078-2079.

Lindgreen, S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes. 5:337.

Luttrell, S. A., S. T. Gonzalez, B. Lohr, and R. Greenberg. 2014. Digital photography quantifies plumage variation and salt marsh melanism among Song Sparrow (*Melospiza melodia*) subspecies of the San Francisco Bay. The Auk. 132(1):277-87.

Malamud-Roam, K. P., F. P. Malamud-Roam, E. B. Watson, J. P. Collins and B. L. Marshall. 1948. Ecological races of Song Sparrows in the San Francisco Bay Region. The Condor. 50:233-256.

Malamud-Roam, K. P., F. P. Malamud-Roam, E. B. Watson, J. P. Collins and L. B. Ingram. 2006. The Quaternary geography and biogeography of tidal saltmarshes. Studies in Avian Biology. 32:11-31.

Maley, J. M. 2012. Ecological speciation of king rails *(Rallus elegans)* and clapper rails *(Rallus longirostris)* (Doctoral dissertation).

Nazarian, R., J. M. Falcón-Pérez, and E. C. Dell'Angelica. 2003. Biogenesis of lysosome-related organelles complex 3 (BLOC-3): a complex containing the Hermansky–Pudlak syndrome (HPS) proteins HPS1 and HPS4. Proceedings of the National Academy of Sciences. 100(15):8770-5.

Neill, W. T. 1958. The occurrence of amphibians and reptiles in saltwater areas, and a bibliography. Bulletin of Marine Science. 8(1):1-97.

Nguyen, T., M. Kermani, J. Fluhr, M. L. Wei, E. K. Novak, L. L. Peters, and R. T. Swank. 2002. Melanosome morphologies in murine models of Hermansky–Pudlak syndrome reflect blocks in organelle development. Journal of Investigative Dermatology. 119(5):1156-64.

Nolan, V. J. 1968. Atlantic Song Sparrow. Pp. 1512 *In* A. C. Bent and O. Austin, eds. Life Histories of North American Cardinals, Grosbeaks, Buntings, Towhees, Finches, Sparrows, and Allies. Dover Press, New York, New York.

Olsen, B., R. Greenberg, I. A. Liu, J. M. Fetch and J. R. Walters. 2010. Interactions between sexual and natural selection on the evolution of a plumage badge. Evolutionary Ecology. 24:731-748.

Peele, A. M., E. H. Burtt Jr, M. R. Schroeder and R. S. Greenberg. 2009. Dark color of the Coastal Plain Swamp Sparrow (*Melospiza georgiana nigrescens*) may be an evolutionary response to occurrence and abundance of salt-tolerant feather-degrading bacilli in its plumage. The Auk. 126:531-535.

Pettus, D. 1963. Salinity and subspeciation in *Natrix sipedon*. Copeia. 25:499-504.

Phillips, A. 1986. The known birds of North and Middle America. Part I: Denver Museum of Natural History. Denver, CO.

Purcell, K. M., A. T. Hitch, P. L. Klerks, and P. L. Leberg. 2008. Adaptation as a potential response to sea-level rise: a genetic basis for salinity tolerance in populations of a coastal marsh fish. Evolutionary Applications. 1(1):155-60.

Raposo, G., and M. S. Marks. 2007. Melanosomes—dark organelles enlighten endosomal membrane transport. Nature Reviews Molecular Cell Biology. 8(10):786-97.

Rosenblum, E. B., and L. J. Harmon. 2011. "Same same but different": replicated ecological speciation at White Sands. Evolution. 65(4):946-60.

Rosenblum, E. B. 2005. Convergent evolution and divergent selection: lizards at the White Sands ecotone. The American Naturalist. 167(1):1-5.

Smith, H. M. 1943. The White Sands earless lizard. Field Museum of Natural History; 1943.

Strohecker, H. F. 1947. Some southwestern Gryllacrididae (Orthoptera). Annals of the Entomological Society of America. 40(2):241-6.

Stroud, C. P. 1950. A survey of the insects of White Sands National Monument, Tularosa Basin, New Mexico. The American Midland Naturalist. 44(3):659-77.

Stroud, C. P. 1949. A white spade-foot toad from the New Mexico White Sands. Copeia. 1949(3):232-.

Thaeler, C. J. 1961. Variation in some salt marsh populations of *Microtus californicus*. University of California Publications in Zoology. 60: 67-94.

Theriault, L. L., and L. S. Hurley. 1970. Ultrastructure of developing melanosomes in C57 black and pallid mice. Developmental Biology. 23(2):261-75.

Von Bloeker Jr., J. C. 1932. Three new mammal species from salt marsh areas in Southern California. Proceedings of the Biological Society of Washington. 45:131-138.

Wine, J. 2003. Rules of conduct for the cystic fibrosis anion channel. Nature Medicine. 9(7):827.

Wood, C. A., W. Post, and C. W. Kilpatrick. 1982. *Microtus pennsylvanicus* (Rodentia: Muridae) in Florida: A Pleistocene relic in a coastal salt marsh. Bulletin of the Florida State Museum of Biological Sciences. 28:25-52.

Wu, C. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology. 14:851-865.

Wysoker, A., K. Tibbetts and T. Fennell. 2012. Picard Tools.

Yeaman, S., and M. C. Whitlock. 2011. The genetic architecture of adaptation under migration–selection balance. Evolution. 65(7):1897-1911.

APPENDIX

Additional Tables and Figures (Chapter 1)

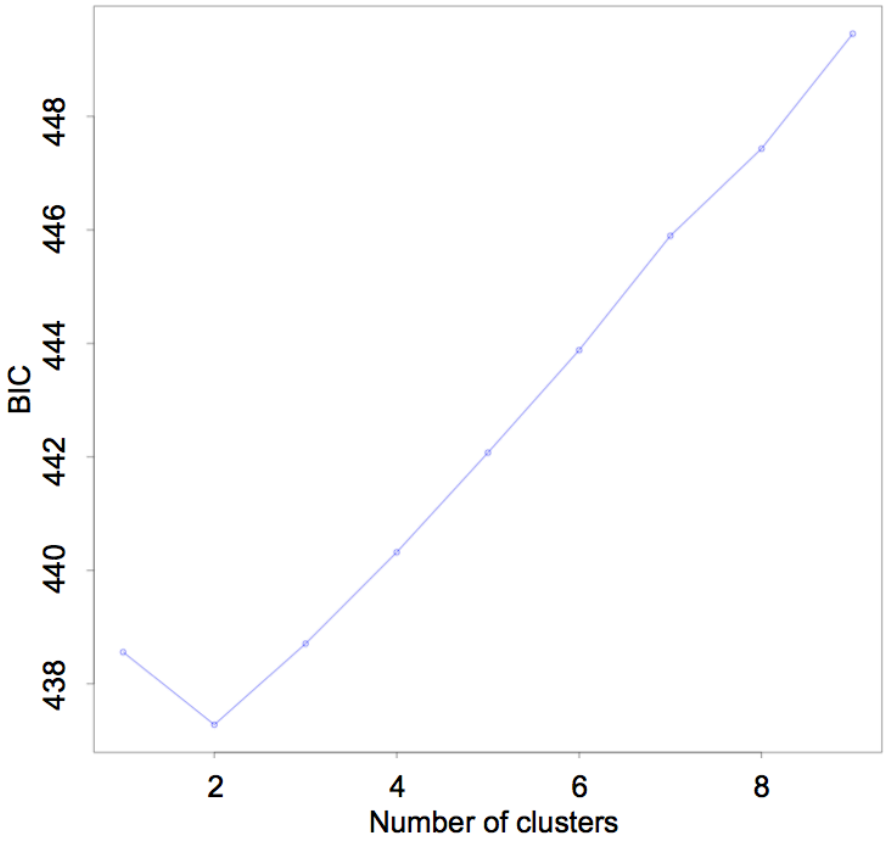


Figure A1.1 After applying k-means clustering to a principal components analysis of the variation present between inland and coastal swamp sparrows at 4,238 independent SNPs (DAPC, Jombart et al. 2010), the model that partitioned variation into two genetic clusters received the strongest support via the lowest Bayesian Information Criterion (BIC) scores.

Table A1.1 Individual ID, sampling date, and Unites States Geological Survey (USGS) band number or Cornell University Museum of Vertebrates (CUMV) accession number for each sample included in this study.

Sample locations		Sample ID	Band#	CUMV#	Sampling Date
HC	Hamilton County, NY	HCm1		BT4057	May 2013
		HCm2		BT4078	May 2013
		HCm3		BT4079	May 2013
TC	Tompkins County, NY	TCm1		BT3487	July 2012
		TCm2		BT3488	July 2012
		TCm3		BT3491	July 2012
		TCm4		BT4032	July 2012
		TCm5		BT4113	July 2013
		TCm6		BT4112	July 2013
		TCm7		BT1872	June 2008
		TCm8		BT2254	May 2008
		TCm9		BT3060	August 2009
		TCm10		BT4418	May 2014
		TCf1		BT2776	August 2009
		TCu1		NA	August 2012
GE	Genessee County, NY	GEu1		BT2710	May 2007
ER	Erie NWR, PA	ERm1	136127242		June 2001
		ERm2	136127241		June 2001
		ERm3	136127235		June 2001
		ERf1	136127237		June 2001
		ERf2	136127240		June 2001
		ERf3	136127238		June 2001
		ERf4	136127239		June 2001
		ERf5	136127243		June 2001
GC	Garrett County, MD	GCm1	136127223		June 2001
		GCm2	136127226		June 2001
		GCm3	136127221		June 2001
		GCm4	136127224		June 2001
		GCm5	136127229		June 2001
		GCf1	136127227		June 2001
		GCf2	136127220		June 2001
		GCf3	136127228		June 2001
HA	Hawley, PA	HAm1	120110739		July 2001
		HAm2	120110740		July 2001
		HAf1	120110738		July 2001
NE	Newton, NJ	NEm1	120110733		July 2001
		NEm2	120110734		July 2001
		NEm3	120110732		July 2001
		NEm4	120110728		July 2001
		NEf1	120110729		July 2001
		NEf2	120110731		July 2001
ME	Meadowlands, NJ	MEm1	001		June 2001
		MEm2	002		June 2001
		MEm3	136127269		June 2001
		MEm4	136127271		June 2001
		MEf1	003		June 2001
		MEf2	136127270		June 2001
GS	Great Swamp NWR, NJ	GSm1	136127261		June 2001
		GSm2	136127264		June 2001
		GSm3	136127263		June 2001
		GSm4	136127262		June 2001

Table A1.1 (Continued)

Sample locations	Sample ID	Band#	CUMV#	Sampling Date	
GS	Great Swamp NWR, NJ	GSm5	136127259		June 2001
		GSm6	136127260		June 2001
		GSm7	136127268		June 2001
		GSf1	008		June 2001
		GSf2	136127267		June 2001
		GSf3	136127266		June 2001
CH	Cheesequake State Park, NJ	CHm1	136127256		June 2001
		CHm2	136127254		June 2001
		CHm3	136127257		June 2001
		CHf1	136127258		June 2001
		CHf2	136127255		June 2001
EH	Egg Harbor, NJ	EHm1	120110746		July 2001
WB	Woodland Beach, DE	WBm1	136127272		June 2001
		WBf1	136127273		June 2001
BH	Bombay Hook, DE	BHm1	136127213		June 2001
		BHm2	136127206		May 2001
		BHm3	136127201		May 2001
		BHm4	136127203		May 2001
		BHm5	136127205		May 2001
		BHf1	136127208		May 2001
		BHf2	136127204		May 2001
		BHf3	136127202		May 2001
PM	Port Mahone, DE	PMm1	136127299		June 2001
		PMm2	136127300		June 2001
		PMm3	136127284		June 2001
		PMm4	136127210		May 2001
		PMm5	136127297		June 2001
		PMf1	136127285		June 2001
		PMf2	120110752		July 2001
		PMf3	136127296		June 2001
		PMf5	136127298		June 2001
FB	Fishing Bay WMA, MD	FBm1		BT4419	June 2014
		FBm2		BT4420	June 2014
		FBm3		BT4422	June 2014
		FBm4	169139703		June 2014
		FBm5	169139704		June 2014
		FBm6	169139705		June 2014
		FBm7	169139706		June 2014
		FBm8	169139708		June 2014
		FBf1		BT4421	June 2014
		FBf2	169139707		June 2014

Table A1.2 Illumina TruSeq index group and barcode combinations used for pooled sequencing. Individuals from different sites were distributed as evenly as possible across index or barcode sequences. Combinations marked NA were used for samples from another species (Song Sparrow, *Melospiza melodia*) that were sequenced and used for a different study.

		6-Base Barcodes						5-Base Barcodes					
		1	2	3	4	5	6	19	20	21	22	23	24
Index		ATCACG	CGATGT	TTAGGC	TGACCA	ACAGTG	GCCAAT	TTCTC	AGCCC	GTATT	CTGTA	AGCAT	ACTAT
1	CGTGAT	HCm1	TCm1	GEu1	ERm1	GCm1	HAm1	NEm1	MEm1	GSm1	CHm1	EHm1	WBm1
2	ACATCG	BHm1	PMm1	FBm1	NA	NA	HCm2	TCf1	ERf1	GCf1	HAf1	NEf1	MEm2
5	CACTGT	GSf1	CHf1	WBf1	BHf1	PMf1	FBm2	NA	HCm3	TCm2	ERm2	GCm2	HAm2
6	ATTGGC	NEm2	MEm3	GSm2	CHm2	BHm2	PMm2	FBf1	NA	TCu1	ERf2	GCf2	NEf2
7	GATCTG	MEf1	GSf2	CHf2	BHf2	PMf2	FBf2	TCm3	ERm3	NEm3	GCm3	MEm4	GSm3
8	TCAAGT	GSm4	FBm3	PMm3	TCm4	CHm3	TCm5	ERf3	GCm4	NEm4	MEf2	GSm5	GSf3
11	GTAGCC	BHm3	BHf3	TCm6	PMm4	FBm4	ERf5	PMf3	TCm7	FBm5	TCm8	BHm4	FBm6
12	TACAAG	FBm8	ERf4	GCm5	GSm6	TCm9	GCf3	BHm5	FBm7	PMm5	GSm7	PMf5	TCm10

Additional Tables and Figures (Chapter 2)

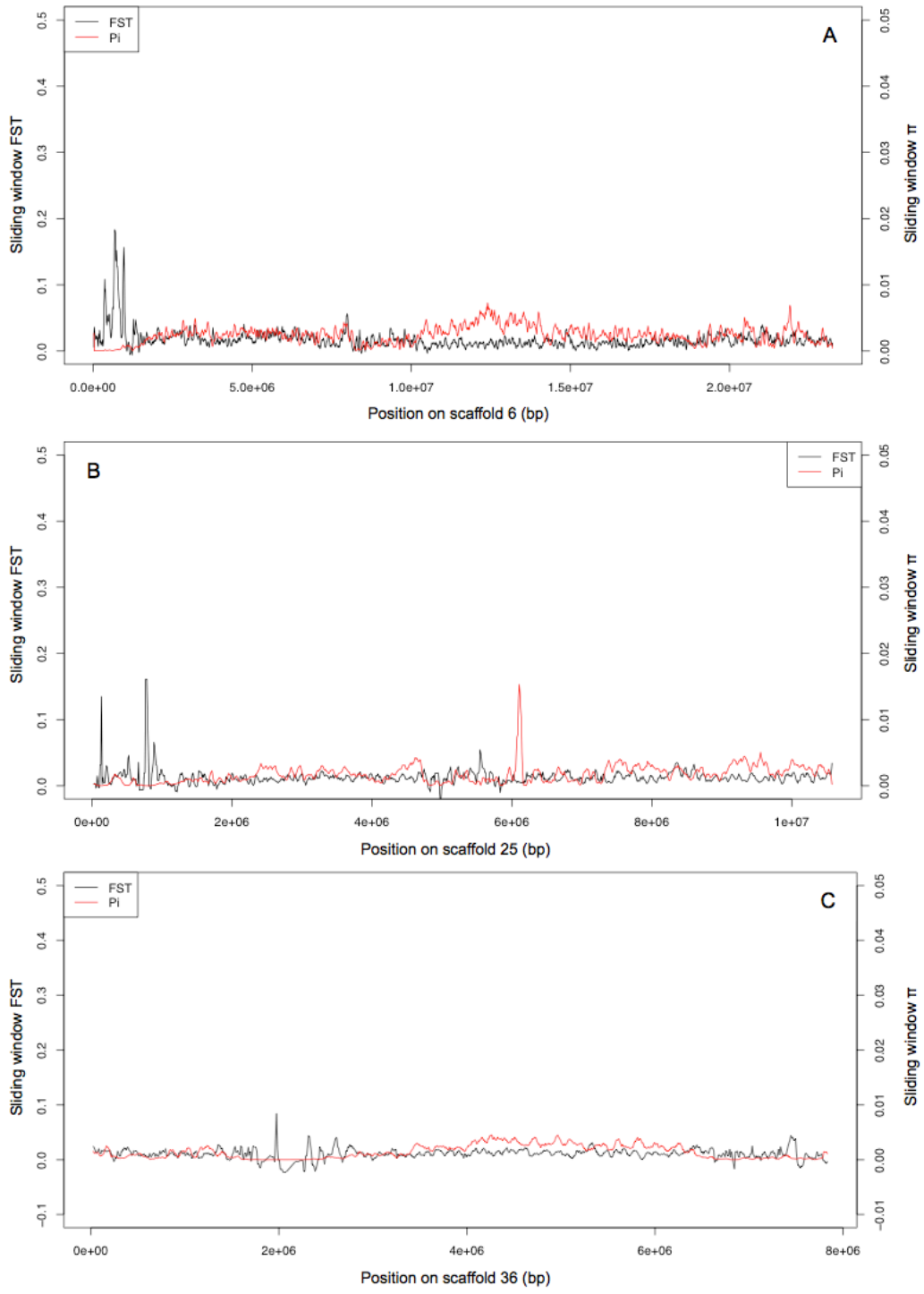


Figure A2.1 Three examples of scaffolds showing co-localization of elevated F_{ST} and reduced nucleotide diversity (π).

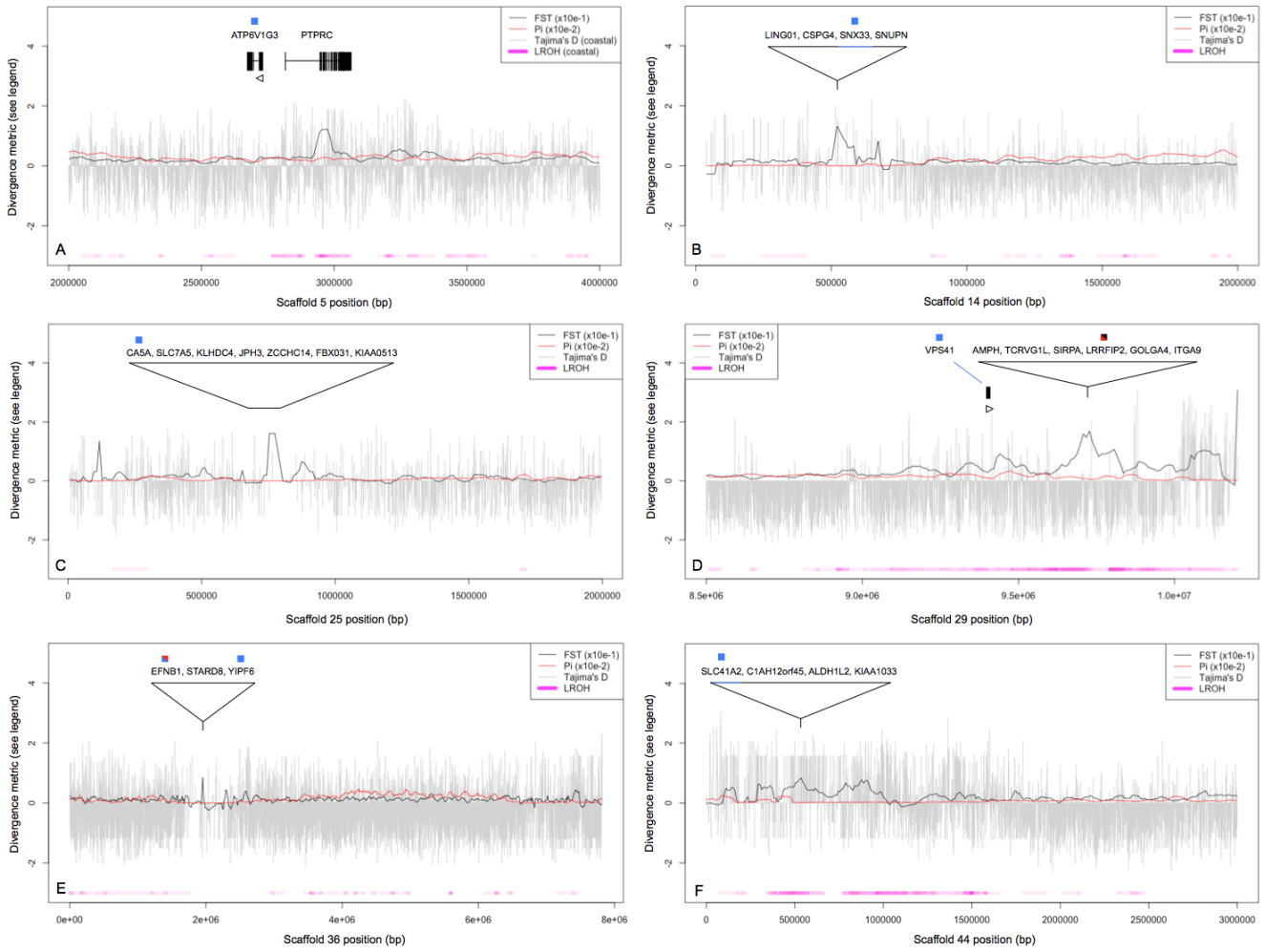


Figure A2.2 Patterns of variation in F_{ST}, π and Tajima's D in regions containing candidate genes for coastal adaptation (I).

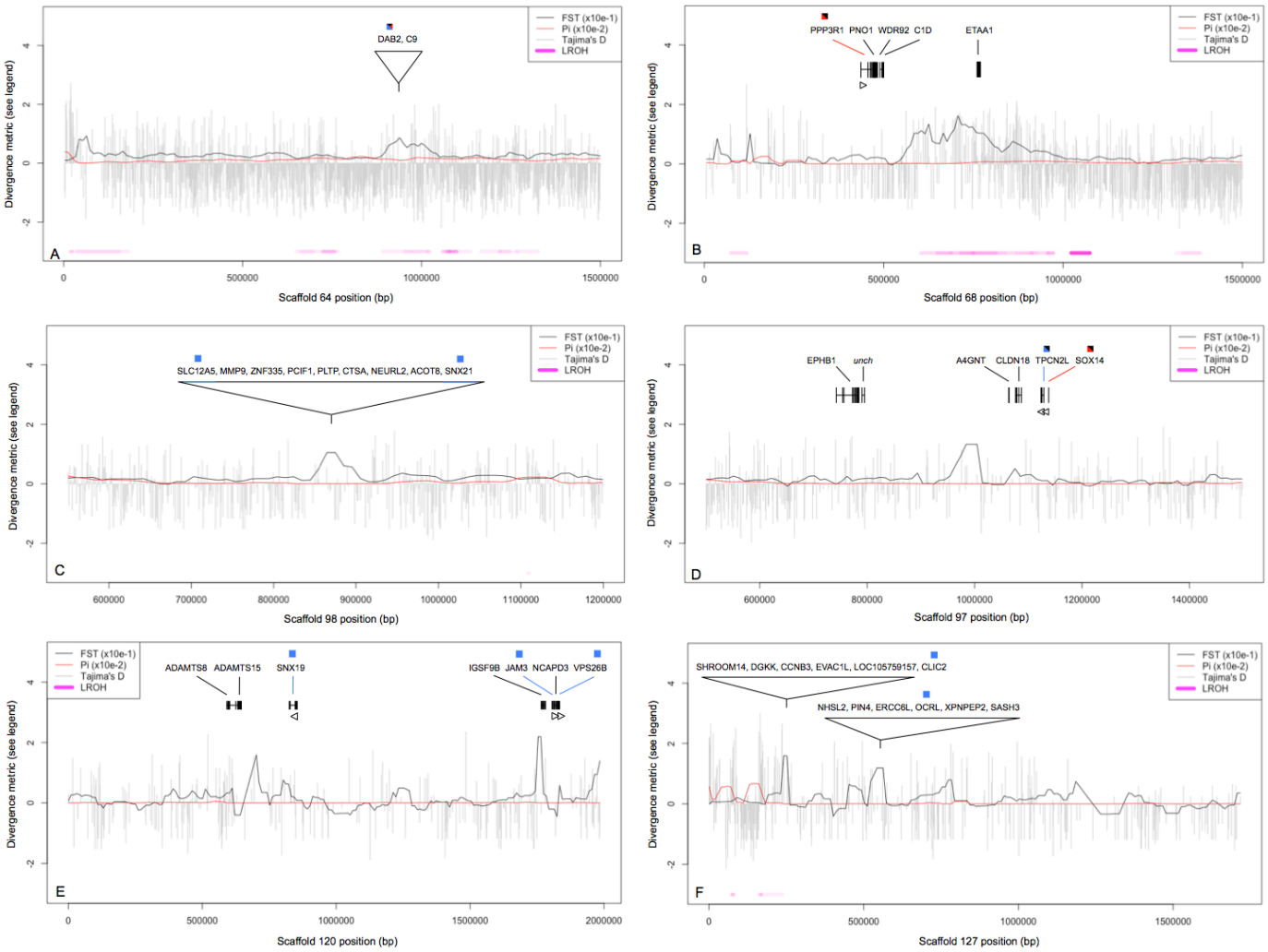


Figure A2.3 Patterns of variation in F_{ST}, π and Tajima's D in regions containing candidate genes for coastal adaptation (II).

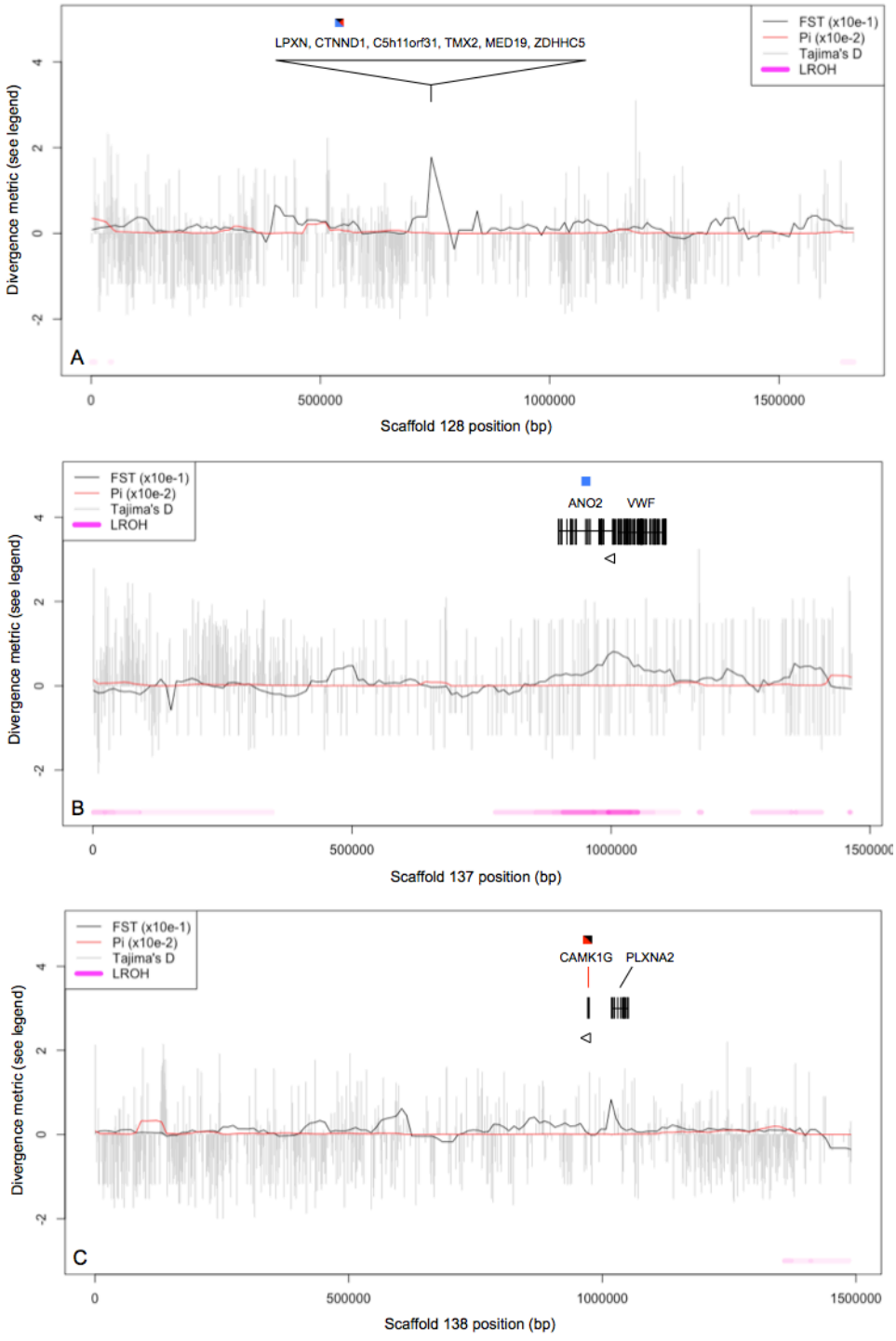


Figure A2.4 Patterns of variation in F_{ST}, π and Tajima's D in regions containing candidate genes for coastal adaptation (III).

Melospiza georgiana georgiana (inland)

Melospiza georgiana nigrescens (coastal)

Figure A2.5 In addition to enhanced plumage melanism, observations of swamp sparrows in the field and in museum collections suggest that coastal swamp sparrows also exhibit greater melanism in the soft tissue of their legs. Individuals shown are breeding males sampled from allopatric sites.

Table A2.1 Sampling information for breeding male swamp sparrows included in whole genome reference construction (*) or resequencing (N=25).

Melospiza georgiana nigrescens (coastal)

Sample ID	Band #/Accession ID	Location	Date
SWSPref *	BT4419	Fishing Bay, MD	06.2014
SWSP1	BT4420	Fishing Bay, MD	06.2014
SWSP2	BT4422	Fishing Bay, MD	06.2014
SWSP3	169139703	Fishing Bay, MD	06.2014
SWSP4	169139704	Fishing Bay, MD	06.2014
SWSP5	169139705	Fishing Bay, MD	06.2014
SWSP6	169139706	Fishing Bay, MD	06.2014
SWSP7	169139708	Fishing Bay, MD	06.2014
SWSP8	136127299	Port Mahone, DE	06.2001
SWSP9	136127300	Port Mahone, DE	06.2001
SWSP10	136127284	Port Mahone, DE	06.2001
SWSP11	136127210	Port Mahone, DE	05.2001

Melospiza georgiana georgiana (inland)

Sample ID	Band #/Accession ID	Location	Date
SWSP13	BT4113	Tompkins County, NY	07.2013
SWSP14	BT4112	Tompkins County, NY	07.2013
SWSP15	BT1872	Tompkins County, NY	06.2008
SWSP16	BT2254	Tompkins County, NY	05.2008
SWSP17	BT3060	Tompkins County, NY	08.2009
SWSP18	BT4418	Tompkins County, NY	05.2014
SWSP19	BT4417	Tompkins County, NY	05.2014
SWSP20	136127242	Erie County, PA	06.2001
SWSP21	136127241	Erie County, PA	06.2001
SWSP22	136127223	Garrett County, MD	06.2001
SWSP23	136127226	Garrett County, MD	06.2001
SWSP24	BT4079	Hamilton County, NY	05.2013

Table A2.2 Library statistics for the ALLPATHS-LG reference genome assembly for an allopatric coastal male swamp sparrow (Sample ID: SWSPref).

	Fragment	3kb	8kb
Number of original reads	542,406,170	507,715,832	466,743,564
Mean length (bp)	101	NA	NA
GC content (%)	44.4		
Mean original insert size (bp)	NA	3175	8065
Mean insert after error correction (bp)	NA	3130	8523
Reads used in assembly (%)	88.2	27.7	8.2
Number of pairs assembled	234,712,074	52,640,448	12,091,741
Sequence coverage	47.3X	13.9X	3.8X
Physical coverage	45.5X	165.8X	105.2X

Table A2.3 Summary stats from the ALLPATHS-LG reference genome assembly for an allopatric coastal male swamp sparrow.

Assembly metric	Reported value
Minimum contig size (bp)	1,000
Number of contigs	43,828
Number of scaffolds	4,778
Number of contigs per Mb	41
Number of scaffolds per Mb	4.47
Total contig length (Mb)	1.02
Total scaffold length with gaps (Mb)	1.07
N50 contig size (kb)	81.8
N50 scaffold size (kb)	10,057
N50 scaffold size, with gaps (kb)	8,220
Median scaffold gap size (bp)	451±42
Bases in captured gaps (%)	4.35
Bases in negative gaps (%)	0.05
Ambiguities per 10,000 bases	41.3
Repetitive sequences (%)	18
Estimated genome size from fragments (Gb)	1.21
Estimated genome coverage by fragments	34X

Table A2.4 BLAST results for 41 F_{ST} peaks (50kb window mean $F_{ST} > 0.8$) on the best quality scaffolds in our genome assembly (scaffolds 0 - 150). Sequences were queried against the zebra finch (ZF; Taeniopygia_guttata-3.2.4 reference Annotation Release 103). Square brackets contain genes located under each peak, and genes outside of brackets are adjacent to the peak. Arrows represent orientation of the mapped SWSP sequence relative to genes in the ZF annotation. Candidate genes for coastally adaptive traits indicated in bold.

SWSP scaffold	Window start position (mb)	Snps /50kb	Zf chr.	Total score	Query cover	Identity	Genes
5	2.92-3.01	633	8	17,479	97%	87%	**ATP6V1G3** [-> PTPRC]
6	0.33-1.01	63	5	$1.5x10^6$	96%	87%	[-> RYR3, FMN1L, FMNL, **GREM1**, SCG5, ARHGAP11A, RASGRP1, FAM98B, **SPRED1**]
10	17.67-18.24	28	2	$2.9x10^6$	93%	88%	[< ADNP2, RECK, **TGFBR1**, COL15A1L (x2), TRIP13, BRD9, TPPP, SLC9A3, LRRC14B]
14	0.49-0.70	7	10	$3.3x10^5$	92%	88%	[-> LING01, CSPG4, **SNX33**, SNUPN]
17	12.55-12.63	241	6	$1.7x10^5$	95%	90%	[-> SCD, PKD2L1 (x2), ABCG2L, **BLOC1S2L**, **BLOC1S2**, CYP2H1L, ALOX5]
	13.99-14.04	151		85,795	97%	88%	[-> CXCL12] **TMEM72**
22	10.36-10.44	16	4	$2.4x10^5$	94%	88%	SGCZ, DLC1L, DLC1, KIAA1456 [<- LONRF1, PAICS, PPAT] AASDH, KIAA1211, ERVK18
23	2.98-3.04	2	12	66,740	84%	90%	[-> DNAH1, BAP1, SEMA3G, ACY1, ABHD14AL, ABHD14B, PCBP4] PARP3, RBM5
25	0.11-0.81	1	11	$1.1x10^6$	87%	86%	[-> KIAA0513, FBX031, ZCCHC14, JPH3, KLHDC4, SLC7A5, **CA5A**, BANP, ZNF469]
26	2.78-2.83	18	17	79,497	69%	84%	COL5A1, OLFM1 [<-] PPP1R26
	9.69-9.77	5		88,589	71%	83%	[-> GLUL, ARRDC1]
29	9.43-9.49	485	2	$2.2x10^5$	92%	88%	[-> **VPS41**]
	9.68-9.86	180		$1.1x10^6$	90%	89%	[-> AMPH, TCRVG1L, SIRPA, **LRRFIP2**, GOLGA4, ITGA9]
	10.05-10.25	58		$1.5x10^6$	96%	87%	[-> ITGA9, NEUROD6]
36	1.95-2.00	1	4A	$1.1x10^5$	93%	91%	**EFNB1** [<-] STARD8, **YIPF6**
37	0.85-0.91	12	15	80,536	98%	89%	UBBL, SCARB1 [-> NCOR2, ZNF664-FAM101A]
43	4.88-4.93	3	3	89,709	78%	83%	[-> NKX2-2, NKX2-4]
44	0.53-0.58	48	1A	68,744	95%	87%	[<- KIAA1033, ALDH1L2] C1AH12orf45, **SLC41A2**
46	4.53-4.64	50	3	$3.0x10^5$	96%	88%	MIA3, TAF1A [<- DUSP10]
50	5.19-5.24	2	*scaffold edge, poor BLAST hits*				*NA*
53	0.50-0.55	3	7	24,081	43%	86%	[-> GALNT5, ERMN, CYTIP]
56	0.67-0.78	66	5	$4.3x10^5$	93%	84%	[-> METTL15] ERVK8L, ERVKL18L, COL3A1L
64	0.04-0.11	18	Z	$3.9x10^5$	78%	86%	[-> OXCT1, FBXO4]
	0.92-0.97	230		$1.5x10^5$	96%	86%	[-> **DAB2**, C9]
68	0.03-0.17	14	3	$3.2x10^5$	88%	88%	[<- BFSP1, PCSK2]
	0.57-0.82	80		$8.0x10^5$	90%	89%	[<- ETAA1] C1D, WDR92, PNO1, **PPP3R1**

Table A2.4 (Continued)

SWSP scaffold	Window start position (mb)	Snps /50kb	Zf chr.	Total score	Query cover	Identity	Genes
75	2.95-3.01	8	21	72,817	76%	87%	[-> MEGF6, ARHGEF16]
76	1.27-1.32	12	19	$1.7x10^5$	79%	85%	[-> WBSCR17]
81	0.01-0.07	10	Z	$1.1x10^5$	86%	88%	**BNC2** [->] CCDC171
	3.28-3.33	4		1,281	72%	84%	PTPRD [->] GP1L
84	0.23-0.28	124	1	71,596	92%	88%	HHLA2, IFT57 [-> CD47] BBX
85	0.17-0.23	10	19	80,943	76%	86%	LIG3 [->] TMEM132E
92	2.50-2.57	4	23	$3.9x10^5$	61%	85%	[-> **EDN2**]
97	0.95-1.03	7	9	$3.6x10^5$	92%	87%	**SOX14**, **TPCN2L**, CLDN18, A4GNT [<-]
98	0.84-0.90	12	20	77,870	89%	89%	**SLC12A5** [-> MMP9, ZNF335, PCIF1, PLTP, CTSA, NEURL2] ACOT8, **SNX21**
101	0.01-0.08	6	26	20,175	96%	89%	KCNA2, LOC105758877, KCNA3 [<-] LOC105758874, PNPLA1, ETV7, KCTD20
	0.98-1.03	1		66,327	67%	80%	[-> PTPRVL, LGR6] LYZL, KLHL12
120	0.70-0.75	1	24	52,921	44%	87%	**SNX19** [<-] ADAMTS15, ADAMTS8, ADAMTS8L
	1.75-1.81	2		$1.1x10^5$	81%	92%	**JAM3** [<- IGSF9B]
	1.96-2.03	3		44,670	79%	88%	[<- ACAD8, THYN1, **VPS26B**, NCAPD3]
123	0.06-0.11	1	*scaffold edge, poor BLAST hits*				*NA*
127	0.24-0.30	1	4A	70,340	77%	87%	[-> SHROOM14, DGKK, CCNB3, EVA1CL, LOC105759157] **CLIC2**
	0.53-0.61	4		$1.2x10^5$	83%	88%	[-> NHSL2, PIN4, ERCC6L, **OCRL**, XPNPEP2, SASH3]
128	0.74-0.79	8	5	61,679	89%	89%	[<- ZDHHC5, MED19, TMX2, C5h11orf31, **CTNND1**, LPXN]
137	1.00-1.05	28	1A	$2.4x10^5$	94%	88%	[<- VWF, **ANO2**]
138	1.01-1.06	8	26	71,336	76%	84%	**CAMK1G** [-> PLXNA2]
140	0.75-0.83	2	14	$1.0x10^5$	94%	89%	[<- NTN3] **NOG2L**
145	1.45-1.52	4	23	40,554	43%	83%	[->] HIVEP3L, **EDN2**
146	0.15-0.22	43	*scaffold edge, poor BLAST hits*				*NA*
148	0.17-1.32	36	2	$3.9x10^6$	88%	86%	[-> ICE1, HUS1, TNS3, **RAMP3**, ADCY1, IGFBP1, IGFBP3]
149	0.21-0.26	4	23	30,520	27%	79%	KPNA6, TXLNA [-> CCDC28B, EIF3I, MTMR9L] HDAC1, TMEM39B
150	1.16-1.21	1	28	$1.1x10^5$	66%	90%	[<- PIP5K1C, CACTIN, TBXA2R, GIPC3, HMG20B, MFSD12]

Table A2.5 Allele frequency divergence between coastal breeding locations may explain positive Tajima's D at three genes. Positive Tajima's D for the remainder (F_{ST} < 0.1 for 6/9 genes) may be due to selection acting on standing variation during coastal colonization (Ch. 2 Discussion). For genes in regions that are invariant within coastal swamp sparrows (*), the region used to estimate F_{ST} was expanded by 1Mb.

Candidate gene	F_{ST} between coastal sites
ATP6V1G3	0.30*
SPRED1	0.22
TGFβR1	0.08*
BLOC1S2 / TMEM72	0.08
SLC41A2	0.03
CLIC2	0.26
ANO2	-0.09
RAMP3	-0.05

Table A2.6 F_{ST} peaks implicating six of the 31 total candidate gene regions from our genome scan mapped to approximately the same location as centromeric regions of chromosomes 1A, 2, 8 and 14, assuming synteny with the zebra finch (Ellegren et al. 2012). "C" denotes the complementary strand.

Zf chr.	SWSP scaffold	Candidate gene	Gene location	Approximate centromere location	Overlap?
1A	44	SLC41A2	54,291,473-54,327,642	48-62Mb	Y
	137	ANO2	63,010,746-63,146,169		Y
2	10	TGFBR1	c75,639,906-75,667,037	58-98Mb	Y
	29	VPS41	34,833,041-34,946,622		N
		LRRFIP2	c35,152,325-35,207,644		N
	148	RAMP3	76,984,792-77,027,615		Y
3	68	PPP3R1	c3,515,047-3,553,175	28-39Mb	N
4A	36	EFNB1	5,792,636-5,812,169	Acro/telocentric,	N
		YIPF6	c6,362,174-6,365,560	end unknown	N
	127	CLIC2	15,652,595-15,661,001		N
		OCRL	15,930,829-15,962,670		N
5	6	GREM1	c30,061,804-30,062,492	Acrocentric	N
		SPRED1	c30,244,669-30,302,996		N
	128	CTNND1	7,360,802-7,372,751		N
6	17	BLOC1S2	16,616,252-16,619,081	Acrocentric	N
		TMEM72	18,255,900-18,261,601		N
8	5	ATP6V1G3	c4,976,229-4,987,371	Acrocentric, 0-5Mb	Y
9	97	SOX14	5,492,381-5,493,165	Acro/telocentric,	N
		TPCN2L	5,508,996-5,516,996	end unknown	N
10	14	SNX33	c1,712,115-1,715,870	Acro/telocentric, end unknown	N
11	25	CA5A	c10,784,894-10,795,589	Acro/telocentric, end unknown	N
14	140	NOG2L	c675,530-676,355	Acro/telocentric, end unknown	Y
20	98	SLC12A5	c7,493,161-7,505,535	Acro/telocentric,	N
		SNX21	c7,568,467-7,570,588	end unknown	N
23	92	EDN2	1,419,652-1,423,181	Acro/telocentric, end unknown	N
24	120	SNX19	6,826,587-6,851,839	Acro/telocentric,	N
		JAM3	c5,775,557-5,783,714	end unknown	N
		VPS26B	c5,717,164-5,725,457		N
26	138	CAMK1G	c3,340,304-3,358,538	Acro/telocentric, end unknown	N

Functional details on candidate genes (Chapter 2)

Salinity tolerance: ion transporters. Four genes under F_{ST} peaks between inland and coastal swamp sparrows encode ion channels that may confer tolerance to salinity via Na^+, Cl^- or Mg^{2+} transport mechanisms. These include two members of the Solute Carrier Family: SLC12A5 and SLC41A2. SLC12 genes encode a superfamily of cation-coupled Cl^- co-transporters that maintain ion concentrations in many cells types and tissues (Hebert et al. 2004). SLC12A5 is a K^+/Cl^- co-transporter specific to neurons (Song et al. 2002). Two closely related SLC12s, SLC12A1 and SLC12A3, contain rare mutations in humans that affect salt reabsorption in the kidney (Ji et al. 2008). The other SLC candidate gene, SLC41A2, encodes a plasma membrane Mg^{2+} transporter. SLC41A2 expression is responsive to extracellular Mg^{2+} concentrations in mice, making it likely that it plays a role in Mg^{2+} homeostasis (Goytain and Quamme 2005). Closely related transporters SLC41A1 and SLC41A3 control Mg^{2+} reabsorption in the kidney nephron DCT of mice and fish (de Baaij et al. 2013, 2016; Kodzhahinchev et al. 2017). Chloride Intracellular Channel 2 (CLIC2) is protein that can insert itself into membranes, creating a Cl^- channel (UniProtKB). Anoctamin 2 (ANO2) encodes a Ca^{2+} activated Cl^- channel (Hartzell et al. 2009). Transmembrane protein 72 (TMEM72) is a putative ion channel, expressed in the kidney nephron DCT where Na^+ is reabsorbed (Habuka et al. 2014).

Salinity tolerance: regulators of vesicle trafficking. Multiple candidate genes for salt tolerance were associated with vesicle trafficking (10/22). ATPase H^+ Transporting V1 Subunit G3 (ATP6V1G3) encodes the ATP catalytic subunit of a proton pump expressed in the membrane of vacuoles, endosomes and lysosomes (Entrez). ATPase proton pumps establish the initial electrochemical gradient that facilitates transport of other ions (like Na^+ or Cl^-) across the vesicle membrane (Batelli et al. 2007). Vacuolar

Protein Sorting genes (VPSs) mediate vesicle traffic to vacuoles, endosomes and lysosomes. VPS41 is a core component of the HOPS complex that regulates vesicle traffic to lysosomes. It is recruited to vesicle membranes by RAB7 (Entrez, UniProtKB), a gene that confers salt tolerance when over-expressed in *Arabidopsis thaliana* (Peng et al. 2014). VPS26B is a core component of the Cargo-Selective Complex (CSC), which regulates endosome traffic to the Trans-Golgi Network (TGN) for recycling. The CSC is recruited to vesicle membranes by RAB7A (UniProtKB). Rates of lysosome turnover and regeneration are controlled by the signaling molecule phosphatidylinositol 4,5-biphosphate ($PtdIns(4,5)P_2$) in association with the protein Clathrin (Rong et al. 2012). The gene OCRL encodes an inositol polyphosphate-5-phostphatase that converts $PtdIns(4,5)P_2$ into a different form, regulating the available pool of this signaling factor for lysosome regeneration (UniProtKB). Sorting Nexins like candidate genes SNX19, SNX21 and SNX33 are proteins that regulate several stages of vesicle trafficking and bind phosphoinositides (Entrez, UniProtKB). SNX33 promotes expression of Na^+ Channel Epithelial 1 Alpha Subunit (ENAC) at the cell surface, and this is the main ion channel driving Na^+ reabsorption in the DCT of the kidney (UniProtKB, Butterworth 2010). The gene Yip1 Domain Family Member 6 (YIPF6) is not well studied but is associated with vesicle transport and Clathrin-coated vesicle budding (Entrez). The Cl^- ion transporter CLIC2 (see "*ion transporters*", above) may also interact with vesicle transport. In mice, the closely related gene CLIC4 is expressed in kidney tubules and regulates early endosomal vesicle trafficking. CLIC4 knockout mice do not develop normal kidney tubules (Chou et al. 2016).

Salinity tolerance: excretion. Carbonic anhydrase 5A (CA5A) is a mitochondrially-localized enzyme involved in ammonia detoxification. This enzyme converts

bicarbonate (H_2CO_3) into carbon dioxide and water in kidney nephrons (Supuran 2008). CA5A knockout mice exhibited physiological effects related to dietary salt, requiring Na^+ and K^+ supplementation in order to yield the same number of offspring as wild type mice (Shah et al. 2013).

Salinity tolerance: adhesion proteins reducing permeability. Junctional Adhesion Molecule 3 (JAM3) creates tight junction seals that reduce epithelial and endothelial permeability, preventing solutes and water from passing between cells in tissues like the intestine (Entrez). Catenin Delta 1 (CTNND1) is a Cadherin-associated member of the Armadillo protein family that also creates cell-to-cell adhesion, as well as contributing to the action of the WNT signaling pathway (Entrez, UniProtKB).

Other candidate genes for bill depth. Fibroblast Growth Factor (FGF) signaling is another important pathway in early bill and facial development (Wilke et al. 1997), and Sprouty Related EVH1 Domain Containing 1 (SPRED1) is an antagonist of FGF signaling that was up-regulated in embryonic facial tissue after infection with the growth factor FGF8 (Li et al. 2013). Ephrin B1 (EFNB1) is a ligand of Eph-related receptor tyrosine kinases, and mutations in EFNB1 are associated with human craniofacial deformities that are sexually dimorphic and are characterized by a broad face and nose (RefSeq, Entrez, GeneCards).

Other pleiotropic candidate genes. Like OCRL, the candidate gene implicated in lysosomal regeneration (see "*regulators of vesicle trafficking*", above), the sorting protein encoded by the gene DAB2 also regulates lysosome traffic and turnover rates by binding to both Clathrin and PtdIns(4,5)P_2. DAB2 also functions during endocytosis of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), the

Cl^- channel responsible for salt reabsorption by intestinal cells, airways and sweat glands in humans (Wine 2003). Highly pleiotropic, DAB2 also inhibits WNT/β-catenin signaling, which could affect melanogenesis, and participates in TGFβ signaling by facilitating the phosphorylation of SMAD2 by TGFβR1 (UniProtKB). The protein encoded by the bill depth candidate gene EFNB1, implicated in facial and nasal deformities in humans, may also influence cell-to-cell adhesion (Entrez).

Higher order signaling factors. The remaining candidate genes act at high levels within core signaling cascades and therefore also have the potential to exert pleiotropic effects on phenotypes depending on downstream responses to that signal. These candidates include Ca^{2+}/Calmodulin Dependent Protein Kinase 1G (CAMK1G), a gene that resembles CAMK and may serve a similar function responding to Ca^{2+} or Calmodulin (CaM) and propagating the MAPK signaling cascade to influence gene expression (RefSeq). The candidate gene Protein Phosphatase 3 Regulatory Subunit B Alpha (PPP3R1) encodes another participant in MAPK signaling that binds Ca^{2+} or CaM (GeneCards). Leucine Rich Repeat Binding FLII Interacting Protein 2 (LRRFIP2) works with DVL3 to activate the WNT signaling pathway, upstream of β-catenin (UniProtKB). SRY-Box 14 (SOX14) is a transcription factor that functions in cell determination during development, and mutations in SOX14 are associated with Mobius syndrome, a craniofacial deformity (RefSeq, Entrez).

Additional Tables and Figures (Chapter 3)

Table A3.1 BLAST mapping locations to *Geospiza fortis* scaffolds (GeoFor_1.0) for four ddRAD loci with high F'_{ST} and low Φ_{ST} in the comparison between inland and coastal swamp sparrows (see methods). The locations of centromeric regions were estimated assuming synteny with the zebra finch (Ellegren et al. 2012). "C" denotes a location on the complementary strand.

Locus	Scaffold	Blast location	ZF chr	Centromere
4612	NW_005054383.1	830573-830715	14	Y
1965	NW_005054303.1	12906841-12906962	5	N
12128	NW_005054422.1	359277-359386	2	N
3803	NW_005054454.1	286461-286568	1B	? microchr

Additional Tables and Figures (Chapter 4)

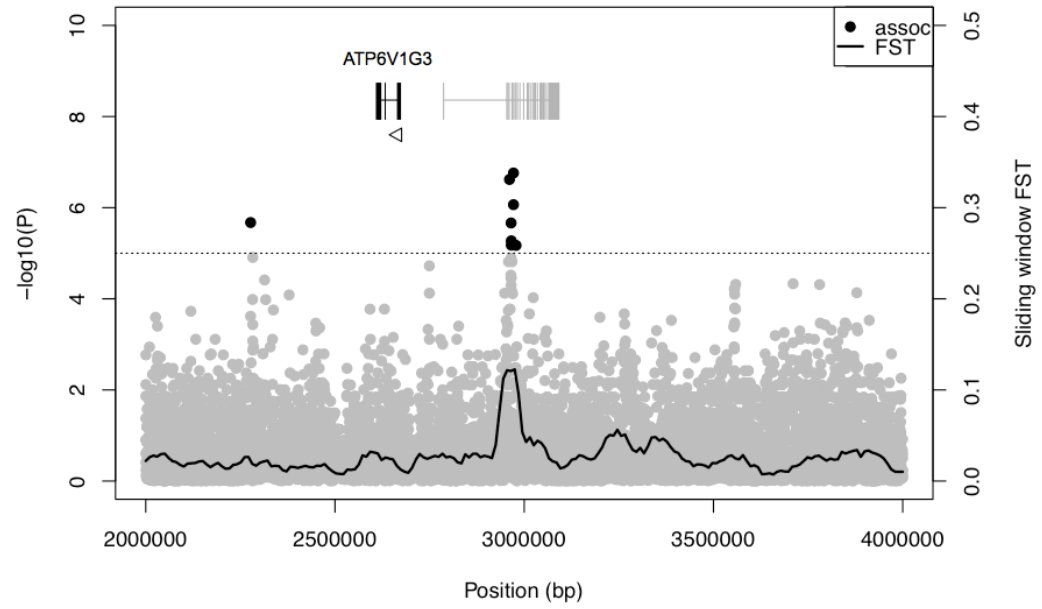


Figure A4.1 Association of genotypes at sites near the candidate salt tolerance gene ATP6V1G3 with salinity of the breeding habitat. Dotted line denotes a significance threshold of $p = 1x10^{-5}$.

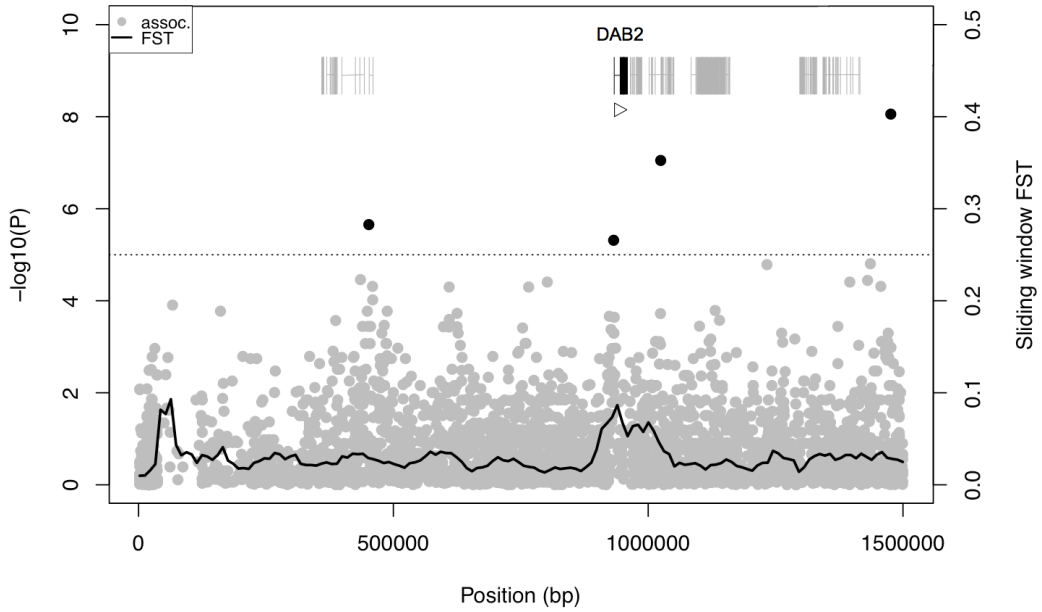


Figure A4.2 Association of genotypes at sites near the gene DAB2 with salinity of the breeding habitat. Dotted line denotes a significance threshold of $p = 1x10^{-5}$.

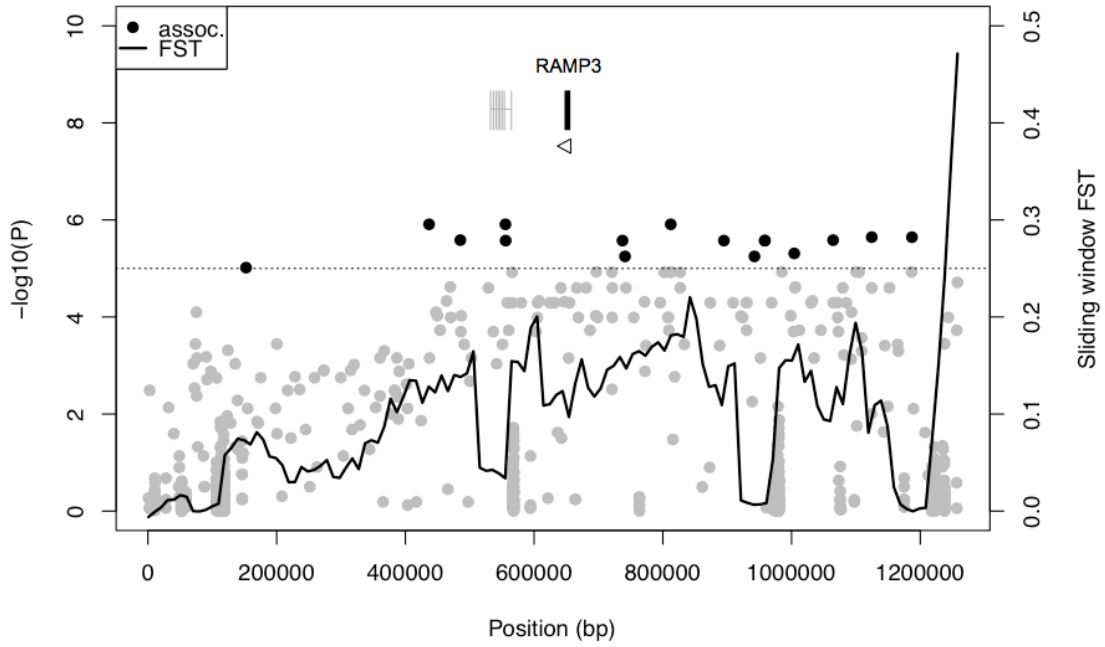


Figure A4.3 Association of genotypes at sites near the gene RAMP3 with salinity of the breeding habitat. Dotted line denotes a significance threshold of $p = 1x10^{-5}$.

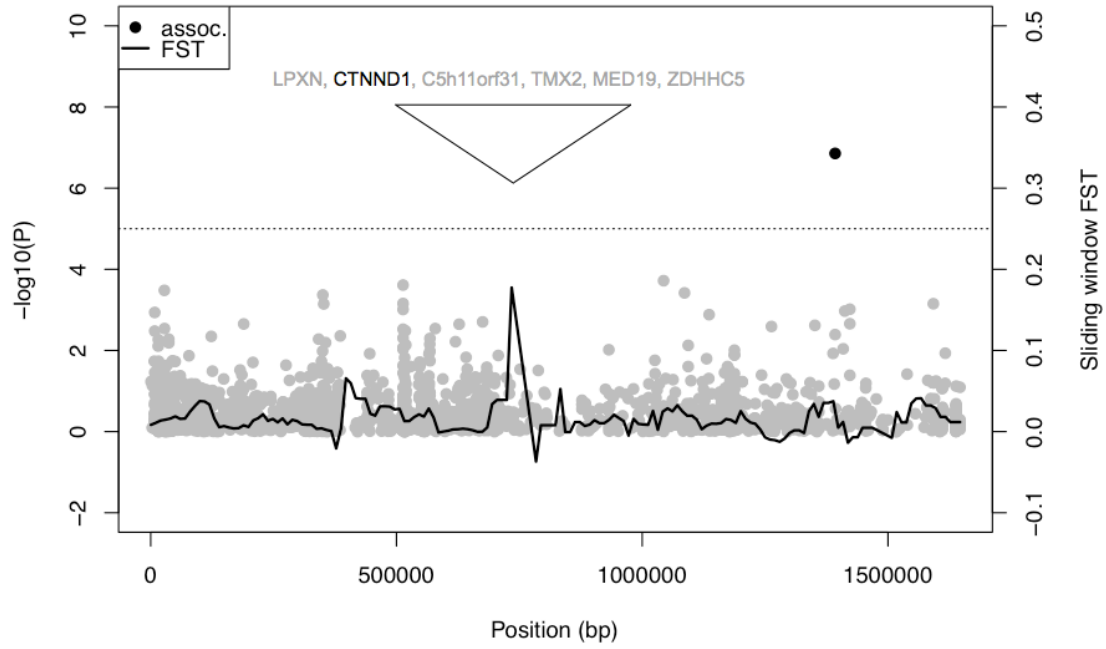


Figure A4.4 Association of genotypes at sites near the gene CTNND1 with black crown size. Dotted line denotes a significance threshold of $p = 1x10^{-5}$.

Table A4.1 Gene identities from plots of genotype-phenotype associations in swamp sparrows (Figures 4.3, 4.5, A4.1-4.4). Candidate genes for salt tolerance or melanic traits shown in bold (Chapter 2).

Scaffold	Genes
5	**ATP6V1G3**, PTPRC
17	ABCG2L, **BLOC1S2L**, **BLOC1S2**, CYP2H1L, ALOX5, MARCH8, ZFAND4, FAM21C, MSMBL(3)
64	TTC33, PTGER4, **DAB2**, C9, FYB, RICTOR, OSMRL, LIFR, EGFLAM
81	**BNC2**, CCDC171
128	LPXN, **CTNND1**, C5h11orf31, TMX2, MED19, ZDHHC5
148	ADCY1, **RAMP3**