POPULATION GENETIC STRUCTURE OF EARLY-STAGE PARAPATRIC ECOLOGICAL SPECIATION IN THE ATLANTIC SONG SPARROW

by

Jonathan D. Clark

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Environmental Science & Policy

Committee:

_____ Dr. Haw Chuan Lim, Thesis Chair

_____ Dr. David A. Luther, Thesis Co-Chair

_____ Dr. Jesus E. Maldonado, Committee Member

_____ Dr. A. Alonso Aguirre, Department Chairperson

_____ Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science

_____ Dr. Ali Andalibi, Interim Dean, College of Science

Date: _____ Summer Semester 2019
George Mason University
Fairfax, VA

Population Genetic Structure of Early-Stage Parapatric Ecological Speciation in the Atlantic Song Sparrow

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Jonathan D. Clark
Bachelor of Science
George Mason University, 2016

Directors: Haw Chuan Lim & David Luther, Professors
George Mason University

Summer Semester 2019
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my late grandmother, Eula ‘Fran’ Cox Thomas. Thank you for always encouraging me and for pushing me to pursue what I love.

ACKNOWLEDGEMENTS

I would like to thank my incredible committee for granting me the opportunity to pursue this project, in particular Dr. Lim for taking me on as a student despite my initial lack of experience with molecular genetics. I would also like to thank my family, whose love and support made the pursuit of this degree possible, especially: my father, Jonathan Clark, Sr.; mother, Christine Clark; grandfather, David Thomas, Sr.; and sister Sarah Clark. Additionally, this project would not have been possible without the sampling endeavors of others, especially the work of Raymond Danner and Russell Greenberg of the Smithsonian Migratory Bird Center. I would also like to thank the dedicated collections staff in the Division of Birds at the Smithsonian National Museum of Natural History for providing additional samples, which were especially valuable for this project. I owe much of my success to Marcus Chua and Masi Sikaroodi who helped shepherd me through laboratory work. I would finally like to acknowledge the generous financial support provided by the Smithsonian-Mason School of Conservation, the Catherine & Richard Becker Fellowship, and the Virginia Society of Ornithology, which funded my studies and research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

POPULATION GENETIC STRUCTURE OF EARLY-STAGE PARAPATRIC ECOLOGICAL SPECIATION IN THE ATLANTIC SONG SPARROW

Jonathan D. Clark, M.S.

George Mason University, 2019

Thesis Directors: Drs. Haw Chuan Lim & David Luther

The most basic model of speciation requires two main components: divergent natural selection and isolation. But how does natural selection facilitate the rise of new species without isolation? If a species occupies different habitats across its range, then parapatric (i.e. adjacent) populations can be exposed to divergent selection, possibly leading to speciation.

The song sparrow (Passerellidae: *Melospiza melodia*) is a common songbird with a variety of subspecies found across North America. One subspecies, the Atlantic song sparrow (*M. m. atlantica*), is a habitat specialist found in the dunes and saltmarshes of the east coast. We investigated the genetic differences of this subspecies from parapatric populations of the eastern song sparrow (*M. m. melodia*), a widespread generalist. Ecologically-driven parapatric divergence is a fundamental mechanism of speciation, but previous studies have had difficulty characterizing parapatric divergence at the genomic level due to limitations in resolution. We used a contemporary genomic method, RADseq,

in conjunction with an assay of a mitochondrial gene to assess how the genomic differentiation of these divergently-adapted, parapatric subspecies have been shaped by ecological selection.

We found that a putatively neutral genetic marker did not exhibit divergence between the subspecies, which suggests that they may interbreed frequently in their contact zone and/or have not been reproductively isolated for long enough for divergence to occur, as would be expected in the early stages of parapatric divergence. Analysis of RAD-markers revealed a clinal relationship in the proportion of genetic ancestry assignment with what appears to be extensive intergradation in transitional habitats, which may be due to hybrid superiority or an influx of genes from both parental types in these habitats. These patterns are consistent with our current understanding of parapatric ecological divergence and provide a framework for further investigations into the genomics of ecological speciation.

CHAPTER 1 | INTRODUCTION

Ecological Speciation and Genomics

The study of ecological genetics was first described as an approached that bridged field and genetic laboratory methods to study how populations are adapted to their environments (Ford, 1964). Since the advent of next-generation sequencing (NGS), the resolution of genetic studies has expanded to the genomic level, thus the modern study of ecological genomics investigates how ecological factors shape the genomes of species. Ecological genomics involves linking biological data across scales of the genome, the phenome, and the environment to make inferences about how populations adapt in response to external selection pressures.

Historically, speciation events were considered to adhere to one of two geographic modes, i.e. allopatric or sympatric speciation (Mayr, 1942). Allopatric speciation occurs as the result of two populations diverging in geographic isolation; sympatric speciation is the result of barriers to gene flow arising between two distinct groups of a population in the same place. In a modern context, speciation events are viewed on a continuum of allopatry to sympatry and defined by the mechanism of divergence. Ecological speciation is a specific process in which the divergence of lineages is driven by the emergence of barriers to gene flow as a result of disruptive ecological selection pressures (Rundle & Nosil, 2005).

Ecological speciation can occur in any geographic context, including allopatry, sympatry, or parapatry.

As outlined by Rundle & Nosil (2005), ecological speciation without isolation, i.e. sympatric or parapatric, requires three major components: 1) a divergent ecological selection pressure, 2) a mechanism of reproductive isolation between populations, and 3) a genetic mechanism linking the two. Ecological selection is a selection pressure that emerges as a result of how an organism interacts with its environment. Reproductive isolation can emerge from divergence in mate choice behaviors, postzygotic incompatibility, decreased hybrid fitness, selection against migrants between habitats, or simply differences in habitat choice. Rundle & Nosil (2005) suggest that ecological divergence and reproductive isolation can be genetically associated by mechanisms of linkage disequilibrium or pleiotropy.

The genomic patterns of ecological divergence are particularly interesting for evolutionary biologists, as ecological divergence is likely a major driving mechanism of both sympatric and parapatric speciation events (Nosil, 2008). However, studying the genomics of ecological divergence in parapatric or sympatric populations can be difficult, as it generally requires high-resolution methods, such as NGS, that can identify small genomic regions recalcitrant to gene flow between populations. Populations undergoing parapatric ecological divergence should generally interbreed in their contact zone, resulting in introgression, as gene flow of neutral loci limits divergence to loci associated with phenotypes under selection. However, population genetic studies have historically focused

on variation in few putatively neutral loci, such as mitochondrial genes or microsatellite DNA markers (microsatellites).

When evaluating the degree of drift caused by prolonged periods of isolation and subsequent secondary contact, neutral loci are a valid means of identifying population divergence (Kimura, 1968, 1979). For example, if divergence occurred during prolonged allopatry, then clear subdivision would likely exist between the mitochondrial genomes of populations due to drift. However, populations which regularly interbreed or have undergone rapid ecological divergence should mainly show differentiation in loci associated with phenotypes under selection. Thus, if populations diverged without isolation or diverged recently, then their mitochondrial genomes should not show strong divergence. Instead, if populations diverged in contact with gene flow, genetic differences would be primarily observed in small linked sets of nuclear loci under disruptive selection, known as genomic islands of differentiation (Marques et al., 2016; Rundle & Nosil, 2005).

Characterizing the genomic signatures of parapatric ecological speciation has only become feasible since the advent of NGS. The genomic differences of divergent saltmarsh, brackish, and freshwater ecotypes in common scurvygrass (Brassicaceae: *Cochlearia officinalis*) were characterized using NGS; freshwater ecotypes were found to have arisen independently several times from the ancestral saltmarsh ecotype (Brandrud, Paun, Lorenzo, Nordal, & Brysting, 2017). In the threespine stickleback (Gasterosteidae: *Gasterosteus aculeatus*), NGS was used to investigate ecologically divergent parapatric lake and stream populations; they were found to have several small islands of differentiation resistant to gene flow (Marques et al., 2016). Studies like these have only

begun to characterize the genomic patterns of early-stage parapatric ecological divergence in non-model species.

Reproductive isolation as a result of parapatric divergence may be common in bird species, but there are few well described examples. Parapatric divergence of song in populations of the greenish warbler (Phylloscopidae: *Phylloscopus trochiloides*) around the Himalayan Plateau has been proposed as a cause of reproductive isolation between distinct subspecies, but there are few well-characterized examples in avian taxa of reproductive isolation resulting from ecological selection without geographic isolation (Irwin, Bensch, & Price, 2001). The clapper rail (Rallidae: *Rallus crepitans*) and king rail (*R. elegans*) are believed to be in the late stages of speciation as the result of ecological divergence in parapatry (Maley, 2012). Evidence for the parapatric ecological divergence of king and clapper rails includes limited hybridization between the species in a transitional zone of brackish marsh where their respective saltmarsh and freshwater marsh habitats abut in Louisiana. As rails have high dispersal ability, ecological barriers, rather than geographic barriers, to gene flow likely led to their divergence. The bay-capped wren-spinetail (Furnariidae: *Spartonoica maluroides*) is another avian species that likely diverged with gene flow in parapatry (Cardoni, Greenberg, Maldonado, & Isacch, 2013). Two distinct morphotypes utilize different adjacent habitats; a dark morph breeds in coastal marshes and a pale morph breeds in upland fields. Despite being morphologically and ecologically distinct, these ecotypes cannot be distinguished by analyses of microsatellites or mitochondrial genes.

During parapatric divergence, populations would be expected to exhibit some clinal variation in phenotypic and genotypic traits. By evaluating the transition of phenotypic or genotypic traits between populations across a contact zone, i.e. geographic cline analysis, inferences can be made about how selection pressures affect gene flow (Brumfield, Jernigan, McDonald, & Braun, 2001). A study of two disjunct yellow-rumped warbler subspecies used geographic cline analysis to determine that an intermediate population was a remnant of intergrades from historical contact (Milá, Toews, Smith, & Wayne, 2011). Another study used geographic cline analysis in a hybrid zone of two toad species (Bufonidae: *Bufo bufo* and *B. spinosus*) to demonstrate the existence of barriers to gene flow (Arntzen et al., 2016). A study of three manakin species in the Amazon (Pipridae: *Lepidothrix* spp.) evaluated morphology and population structure across a contact zone to demonstrate a speciation event resulted from historic hybridization (Barrera-Guzmán, Aleixo, Shawkey, & Weir, 2017).

In summary, inferences can be made about the evolutionary relationship of divergent populations by evaluating shifts in allele frequencies and environmentally-relevant phenotypes across a contact zone. If divergent populations are suspected to have never been disjunct, then geographic cline analysis provides a framework to test for genomic signatures of ecological divergence that would precede a parapatric ecological speciation event. Therefore, evaluating clinal relationships of genomics and phenotypes in parapatric populations under divergent environmental conditions allows for *in situ* studies of ecological speciation processes.

Selection Pressures on the Avian Bill

The avian bill is a trait classically studied for its role in evolutionary processes, due to its great diversity of forms and functions (Cooney et al., 2017). The avian bill is particularly useful for studying natural selection, because of its role as a tool. The avian bill is well known as a tool for foraging, but also plays a role in other functions, such as thermoregulation and defense (Rico-Guevara, Rubega, Hurme, & Dudley, 2019; Tattersall, Arnaout, & Symonds, 2016). The myriad biological functions of the avian bill have likely played an important role in driving many avian speciation events.

Because the bill is associated with so many aspects of avian biology, it is often subject to intense natural and sexual selection pressures concurrently. The avian bill can play a major role in intrasexual competition and mate attraction, because it is partially responsible for the type of and quality of songs that a bird can produce. Mate attraction in songbirds is often tied to vocal performance, which is constrained by the ability to achieve a large frequency bandwidth, i.e. the difference in minimum and maximum pitch frequencies, and a high trill rate, i.e. the number of repeated notes per second, in song production (Podos, 1997). The size and shape of the bill is directly related to the ability of an individual to rapidly trill and maximize trill bandwidth in song production, thus physically affecting vocal performance (Podos, Southall, & Rossi-Santos, 2004).

New environmental conditions that rapidly change selection pressures on a population can result in sudden shifts in traits that are under strong influence from natural selection. For example, the Everglades snail kite (Accipitridae: *Rostrhamus socialbilis plumbeus*), an endangered wetland-dependent raptor, was observed rapidly undergoing a

population-level shift in bill size as a response to directional selection (Cattau, Jr, Kimball, Miller, & Kitchens, 2018). Everglades snail kites use their hooked bill to extract snails from their shells, feeding almost exclusively on a single species, the Florida apple snail (Ampullariidae: *Pomacea paludosa*). Following the invasion of a much larger exotic congener prey species, the island apple snail (*P. maculata*), snail kites rapidly adapted to the new food source. Within eight years of the island apple snail invasion, bill length as well as body mass of fledgling snail kites increased by approximately one standard deviation, showing that directional selection can influence the bill morphology of a long-lived avian species at the population level within less than a generation.

Further evidence of the ability of directional selection to suddenly alter the bill size of avian populations can be found in Darwin's finches from the Galapagos Islands of Ecuador. The bill size and shape of the medium ground-finch (Thraupidae: *Geospiza fortis*) and the congener cactus finch (*G. scandens*) were found to rapidly and repeatedly change as a response to stochastic environmental changes on the island of Daphne Major over a 30-year period (Grant & Grant, 2002). For example, following a period of intense drought, the bill size of the medium ground-finch population increased substantially, as individuals with larger bills were able to consume large seeds that became the dominant food source as a result of the drought (Boag & Grant, 1981). Simple mechanisms have been proposed to explain how variation in the bill shape of Darwin's finches can rapidly respond to selection. For example, a study demonstrated how varied levels of bone morphogenetic protein 4 (BMP4) expression correlated to bill structure between species of Darwin's finches, and the alteration of BMP4 levels in chicken embryos reproduced this effect

(Abzhanov, Protas, Grant, Grant, & Tabin, 2004). Further, a whole-genome resequencing investigation of Darwin's finches found that much of the evolution of bill size and shape in this clade may be associated with variation in a single gene, the ALX homeobox 1 (ALX1) gene (Lamichhaney et al., 2015).

The ability of the avian bill to rapidly change in response to environmental circumstances as well as its role in both ecological and reproductive fitness make it an excellent candidate to behave as a so-called magic trait. A magic trait refers to a characteristic that when subjected to disruptive ecological selection concurrently forms reproductive barriers to gene flow (Servedio, Doorn, Kopp, Frame, & Nosil, 2011). For example, in the medium ground-finch where sympatric morphs with small and large bills are under disruptive ecological selection, differences in song between morphotypes appears to have formed assortative mating barriers (Podos, 2010).

Another case of the bill behaving as a magic trait can be found in populations of the red crossbill (Fringillidae: *Loxia curvirostra*), which have highly-specialized bills suited to feed on seeds inside of pinecones. There is variation of bill forms, as different populations of red crossbills specialize on different pine species (Benkman, 2003; Parchman, Benkman, & Britch, 2006). Each pine species has a slightly different pinecone morphology, making it difficult for individuals to extract the seeds of a particular species without a having a bill specialized for that single species. Female red crossbills tend to prefer males who are efficient foragers, further increasing reproductive fitness of individuals with bills specialized to local pine species (Snowberg & Benkman, 2009). The combination of ecological and reproductive pressure for regionally adapted bills has

resulted in distinct populations within this complex, which do not seem to regularly interbreed, possibly due to the observed variation in vocalization associated with bill form (Parchman et al., 2006). Therefore, the bill may be a major driver of ecological speciation processes in avian systems, because of its ability to rapidly respond to selection and its potential to behave as a magic trait.

The Song Sparrow as a Model for the Study of Evolution

The song sparrow (Passerellidae: *Melospiza melodia*) is a ubiquitous North American passerine with a diverse array of subspecies spanning nearly every terrestrial habitat on the continent. A taxonomic review of the song sparrow complex identified 25 morphologically diagnosable subspecies, making it one of the most polytypic avian species in the northern hemisphere (Patten & Pruett, 2009). Within the song sparrow complex, there are several subspecies adapted to a range of climatically challenging environments, such as southwestern deserts, coastal saltmarshes, and the Aleutian Islands.

Because of the breadth of geographic distribution and habitat use in the song sparrow complex, it illustrates many classical ecogeographic relationships, including Allen's, Bergman's, and Gloger's rules (Patten & Pruett, 2009). In avian species, Gloger's rule is a consequence of the properties of dark pigments, such as eumelanin, which help fortify feathers against degradation from microbes that thrive in wet environments (Burtt & Ichida, 2004). Gloger's rule can be demonstrated in the song sparrow complex by comparison of the dark plumaged subspecies that reside in moist regions of the northwest, such as *M. m. morphna*, to their paler counterparts in the arid southwest, such as *M. m. fallax* (Burtt & Ichida, 2004; Patten & Pruett, 2009). The song sparrow complex also

provides support for the well-known Bergman's rule, which states that members of a homeothermic clade that reside in cold environments tend to have larger bodies (Salewski & Watt, 2017). The smallest subspecies, *M.m fallax*, resides in the scorching heat of the Sonoran Desert, while the largest, towhee-sized subspecies, *M. m. maxima*, inhabits the frigid Aleutian Islands (Patten & Pruett, 2009). Members of the complex under specific ecological conditions have also been shown to adhere to Allen's rule, which states that members of a clade that reside in cold environments will have proportionally smaller appendages to prevent the loss of body heat (Danner & Greenberg, 2015; Greenberg & Danner, 2012). Extreme phenotypic diversity associated with variation in habitat use makes the song sparrow complex an ideal vertebrate model for *in situ* studies of ecological and evolutionary processes.

In addition to providing examples of vertebrate ecogeographic relationships, the song sparrow complex also provides an ideal case for the study of population divergence processes. For example, several units of the complex have been suggested to be a rare example of an avian ring species, demonstrating the process of divergence by geographic distance (Patten & Pruett, 2009; Price, 2008). The majority of subspecific diversity in the complex is restricted to insular and peninsular regions of the west, and several subspecies are endemic to highland islands in Mexico (Patten & Pruett, 2009). Many of these groups may, therefore, be the result of standard allopatric divergence mechanisms, i.e. the result of drift or differential selection in historic isolation.

There are also several widespread generalist subspecies, such as *M. m. melodia* in the east, and *M. m. merrilli* and *M. m. montana* in the west, with extensive but narrow

contact zones that abut conspecific specialist populations (Aldrich, 1984; Patten & Pruett, 2009). The most widespread of these generalist subspecies is *M. m. melodia*, which occurs from the Rocky Mountains east to the Atlantic Ocean and from the Appalachian Mountains north to the Hudson Bay (Patten & Pruett, 2009). Contact zones between geographically limited and widespread subspecies provide cases to evaluate how discrete populations are maintained despite gene flow.

If populations are in secondary contact following a period of divergence in allopatry, the maintenance of subspecific groups could indicate selection against intergrades or the existence of reproductive barriers (Price, 2008). Postzygotic barriers caused by ecological selection against subspecific intergrades would prevent panmixia despite interbreeding, as is the case in the yellow-rumped warbler complex (Parulidae: *Setophaga coronata*; Hubbard, 1969; Toews, Brelsford, Grossen, Milá, & Irwin, 2016). Alternatively, prezygotic reproductive barriers may exist between populations due to divergence in mating systems during a period of allopatry, as in altitudinally differentiated populations of grey-breasted wood-wren (Troglodytidae: *Henicorhina leucophrys*; Caro, Caycedo-Rosales, Bowie, Slabbekoorn, & Cadena, 2013). In both cases, maintenance of distinct groups in secondary contact requires a mechanism restricting gene flow, and populations should still exhibit clear genetic divergence unless they are undergoing post-contact fusion.

Conversely, a lack of divergence in neutral loci between song sparrow populations would suggest that some specialist groups are the result of ongoing parapatric divergence due to disruptive ecological selection pressures without historical inhibition of contact or

reproductive barriers (Aldrich, 1984). While mechanisms of parapatric and ecological speciation have been widely theorized, little evidence of parapatric ecological divergence has been demonstrated in avian taxa, as it is difficult to characterize (Price, 2008; Rundle & Nosil, 2005). The clear association of many subspecies' ranges with Köppen's climate zones is potential evidence that divergence in the song sparrow complex is driven primarily by ecological selection (Patten & Pruett, 2009). For example, the most widespread subspecies, *M. m. melodia*, is found throughout most of the humid continental zone, while the neighboring specialist subspecies, *M. m. atlantica*, is restricted to the northern coastal portion of the humid subtropical zone (Patten & Pruett, 2009).

The most authoritative reviews of song sparrow subspecies to date relied on evaluations of plumage and morphology due to the limited resolution of genetic methods at the time (Aldrich, 1984; Patten & Pruett, 2009). Now that genomic technologies have advanced, investigations of song sparrow subspecies using extensive genomic data are possible. Therefore, extensive genomic reviews of the remarkably diverse song sparrow complex are necessary to corroborate morphologically described populations. Such genomic studies of the song sparrow complex will certainly advance the understanding of the genomic mechanisms underlying speciation processes.

The Atlantic Song Sparrow

One subspecies of song sparrow, the Atlantic song sparrow (*M. m. atlantica*; hereafter *atlantica*), is an extreme habitat specialist that only occurs in the sand dunes and saltmarshes of the Atlantic coast from North Carolina to New York (Patten & Pruett, 2009; Todd, 1924). Despite inhabiting the region of North America first colonized by the British,

atlantica remarkably went unrecognized until the 20th century, possibly due to its general resemblance to the widespread and geographically adjacent nominate subspecies, the eastern song sparrow (*M. m. melodia*; hereafter *melodia*; Todd, 1924). These subspecies were initially distinguished by differences in their plumage; *atlantica* was first described as "much grayer above, with the blackish streaking more distinct, and the reddish brown feather-edging reduced to a minimum" (Todd, 1924).

Additionally, *atlantica* is distinguished from *melodia* by its larger bill (Patten & Pruett, 2009; Todd, 1924). Several studies have demonstrated the selective advantage for *atlantica* that comes with having a larger bill (Danner & Greenberg, 2015; Danner, Gulson-Castillo, et al., 2016; Greenberg, Cadena, Danner, & Tattersall, 2012; Luther & Danner, 2016). A relatively larger bill is thought to be a selective advantage for *atlantica*, because coastal sand dunes are hot environments with limited freshwater, and bills can be used to radiate body heat without losing water (Greenberg, Danner, Olsen, & Luther, 2012; Luther & Greenberg, 2014; Tattersall, Andrade, & Abe, 2009; Tattersall et al., 2016).

It is reasonable to assume that small passerine birds would not need to use the bill as a radiator of heat, because they naturally have a high surface area to body volume ratio (Greenberg, Cadena, et al., 2012). However, a high-temperature environment poses a serious physiological challenge to small birds, as they rapidly lose water by panting to thermoregulate (Tattersall et al., 2016). The challenge of thermoregulation is especially severe for species which inhabit an environment that is both hot and freshwater-limited. Therefore, the bill often plays a key role as a thermoregulatory organ for birds in hot, freshwater-limited environments (Danner, Gulson-Castillo, et al., 2016).

A study using infrared imaging compared the response of *atlantica* and *melodia* individuals to heat stress in a controlled setting and found that *atlantica* individuals were able to dissipate approximately 30% more heat through the bill than *melodia* individuals (Greenberg, Cadena, et al., 2012). Of the individuals examined, the average surface area of the bill in *atlantica* was more than 16% greater than that of *melodia*. Overall, song sparrows were found to rely on the bill to dissipate up to 10% of radiated heat despite the bill making up less than 2.5% of their total body surface area. Specimens of *atlantica* were also found to diffuse heat from the base of their bill much more efficiently than *melodia*. Further work using the same methods found that bills are also used as dry radiators of heat in large-billed species of Darwin's finches (Tattersall, Chaves, & Danner, 2017).

A study using high precision computed tomography compared the interior structure of the bill between *melodia* and *atlantica* individuals and found that *atlantica* had longer nasal cavities and greater surface area of the nasal conchae (Danner, Gulson-Castillo, et al., 2016). The nasal conchae, also known as nasal respiratory turbinates, are structures that prevent the evaporative loss of water from respiration by condensing exhaled moisture (Geist, 2000; Tieleman, Williams, Michaeli, & Pinshow, 1999). Therefore, increased surface area of the nasal conchae should be an adaptive advantage in freshwater-limited coastal habitats. The nasal conchae are also highly vascularized, potentially aiding heat dissemination through the bill.

The homeostatic advantage of *atlantica* morphology versus *melodia* morphology in hot environments is likely marginal from a perspective of mortality (Danner, Gulson-Castillo, et al., 2016). However, *atlantica* individuals with larger bills are more active in

the heat of day than *atlantica* with smaller bills (Luther & Danner, 2016). A study found that *atlantica* males with the largest bills sing at rates nearly twice that of males with the smallest bills regardless of the ambient temperature (Luther & Danner, 2016). But, a limit exists in the ability of song sparrows to radiate body heat from the bill as ambient temperatures converge with body temperature, meaning that selection for larger bills in *atlantica* has an upper bound (Greenberg, Cadena, et al., 2012).

One adaptive disadvantage to a large bill is increased heat loss during the winter (Ryeland, Weston, & Symonds, 2017). Both *atlantica* and *melodia* were unable to restrict blood flow to the bill when subjected to cold temperatures, possibly because of large muscles at the base of the bill that require a constant blood supply (Greenberg, Cadena, et al., 2012). While there appears to be selection against larger bills for *melodia* individuals associated with winter temperatures, selection against larger bills associated with winter temperatures does not seem to affect *atlantica* individuals (Danner & Greenberg, 2015). Therefore, the reproductive benefits of large bills during the breeding season for the *atlantica* population may outweigh negative selection pressure during the winter. The wintering grounds of *atlantica* may have a milder climate than that of *melodia*, but little is known about the migratory behavior and wintering ecology of *atlantica* (Danner, Olsen, & Luther, 2016; Greenberg, Cadena, et al., 2012).

Disruptive selection pressures that explain differences in bill size between *atlantica* and *melodia* are well described, but no studies have specifically investigated the genetics of these parapatric subspecies to evaluate how they are related. An understanding of the genetic variation underlying the morphological differences between *atlantica* and *melodia*

is critical to determining their evolutionary relationship. A clear first step is to compare the population genetic structure of these groups to identify if they are clearly differentiated and if they are reproductively isolated. Investigating population division in these subspecies will pave the way for future in-depth analysis of the *atlantica* genome to determine the genes underlying their specialized phenotypes.

Based on the general morphological similarities and the modern distributions of *atlantica* and *melodia*, it is reasonable to deduce that *atlantica* likely diverged from the generalist *melodia* in parapatry. During glacial recession the generalist *melodia* may have expanded its range and colonized coastal habitats, facilitating divergent adaptations that gave rise to the *atlantica* subspecies. By describing geographic variation in population genetics across the contact zone of *atlantica* and *melodia*, the population genetic structure of early-stage parapatric ecological divergence can be evaluated in these conspecific populations.

Studying the Population Genetics of the Atlantic Song Sparrow

Because *atlantica* and *melodia* fit the general assumptions of a biological cline, i.e. varied morphotypes on opposite sides of a habitat gradient, a one-dimensional sampling transect that represents individuals from across the habitat gradient would provide an ideal means of investigating the relationship of these populations. Prior fieldwork by researchers at George Mason University and the Smithsonian Migratory Bird Center included the collection of blood samples from >200 song sparrows representing both the *atlantica* and *melodia* subspecies, as well as possible introgressed intergrades, along a coastal-to-inland transect.

The easterly portion of the sampling transect started at the Atlantic shore of Delaware and Maryland on the Delmarva Peninsula and moved inland terminating at sites in western Maryland and southwestern Virginia (**Fig. 1**). Additionally, the Smithsonian National Museum of Natural History (USNM) provided several DNA samples from *atlantica* individuals collected in saltmarshes of the upper Chesapeake Bay (**Table 1**). The availability of this broad set of DNA samples makes a thorough examination of the population genetic structure and evolutionary history of *atlantica* and *melodia* possible.

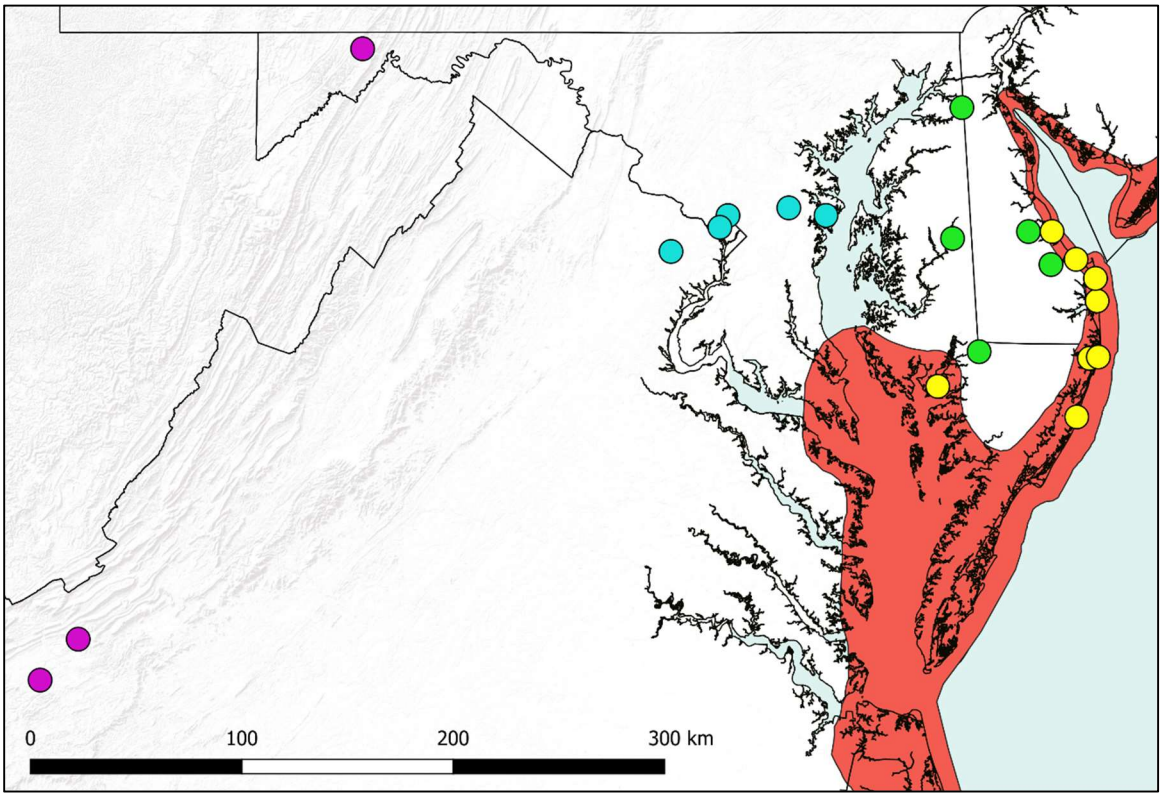

Figure 1 | Map of sampling sites. Blood samples were collected from song sparrows throughout Delaware, Maryland, and Virginia. Yellow points are sites in *M. m. atlantica* habitat (coastal); green points are upland sites on the Delmarva Peninsula (peninsular); blue points are sites on the mainland proximal to the coast (proximal inland); purple points are sites on the mainland distal to the coast (distal inland). The red area is the range of *M. m. atlantica* in the region; *M. m. melodia* occurs throughout the rest of the region shown (*sensu* Patten & Pruett, 2009).

Table 1 | List of samples from USNM

USNM Skin ID	USNM Tissue ID	USNM Biorepository ID	DNA ID	dsDNA Conc. (ng/µL)
634160	B18526	AA0AP22	Yellow_23	6.61
634171	B18624	AA0AQ32	Yellow_19	5.82
634177	B18630	AA0AQ38	Yellow_5	3.48
636202	B18806	AA0AS25	Yellow_10	1.91
638668	B22915	AA1BD37	Yellow_31	10.2
638682	B22939	AA1BD61	Yellow_27	3.43
641291	B26727	AA0AU80	Yellow_15	2.07

CHAPTER 2 | MITOCHONDRIAL POPULATION STRUCTURE ANALYSIS

Prior Population Genetic Studies of the Song Sparrow

The neutral theory of molecular evolution posits that most DNA mutations that occur within a population will have no effect on the fitness of an organism (Kimura, 1979). These neutral mutations can become fixed in isolated populations, making them useful for identifying population structure when populations diverged after prolonged allopatry. However, if populations diverged recently, rapidly, or without isolation, then neutral genetic markers may not correlate with apparent geographic or phenotypic differences.

Prior studies of the song sparrow complex have relied on few genetic markers putatively considered to be neutral, such as mitochondrial genes and microsatellites. Remarkably, both mitochondrial restriction fragment length polymorphism (RFLP) analysis and sequencing of the mitochondrial control region (CR) failed to discern population structure between song sparrow populations at the continental scale (Fry & Zink, 1998; Zink & Dittmann, 1993). A lack of mitochondrial divergence between populations on opposite sides of the continent suggests that the phenotypic diversity of the complex arose recently and that populations may have ecologically adapted to the environments of various regions without complete barriers to gene flow (Aldrich, 1984; Fry & Zink, 1998; Nabholz, Lanfear, & Fuchs, 2016).

At finer spatial scales, neutral genetic markers have had limited success at differentiating song sparrow populations. In the parapatric desert subspecies *M. m. fallax*

and *M. m. heermanni*, analysis of microsatellites was able to differentiate populations, suggesting that barriers to gene flow exist between these populations (Patten, Rotenberry, & Zuk, 2004). For the several subspecies found throughout the San Francisco Bay region and the Pacific Northwest, microsatellites demonstrated limited population differentiation (Chan & Arcese, 2002; Wilson, Arcese, Chan, & Patten, 2011). For song sparrow populations on the Channel Islands, sequencing of the CR demonstrated limited structure, but microsatellites showed inter-island population structure and divergence from mainland populations (Wilson, Chan, Taylor, & Arcese, 2015).

The general inability of neutral genetic markers to differentiate phenotypically distinct and geographically distant populations of song sparrow suggests that neutral loci are exchanged between populations or have not yet become fixed in isolated populations. In cases with divergence despite extensive gene flow, several authors have suggested that limited genomic islands of differentiation under local ecological selection can be the principal cause of variation (Emelianov, Marec, & Mallet, 2004; Marques et al., 2016; Turner, Hahn, & Nuzhdin, 2005; Via & West, 2008). Generally, to detect these genomic islands of differentiation requires scanning large portions of the genome.

In avian species that diverged by drift or selection in isolation, such as the grey-cheeked fulvetta (Pellorneidae: *Alcippe morrisonia*), the second subunit of the mitochondrial nicotinamide adenine dinucleotide dehydrogenase (ND2) gene, has been used to successfully discern population structure (Zou, Lim, Marks, Moyle, & Sheldon, 2007). For lineages that have adapted to utilize different habitats in parapatry, such as the bay-capped wren-spinetail, analysis of the ND2 gene has failed to discern population structure (Cardoni et al., 2013). However, analysis of differentiation of the ND2 gene in

the ecologically speciated clapper and king rails, did discern structure (Maley, 2012). These studies suggest that the ND2 gene is an appropriate marker for evaluating if parapatric populations recently diverged with gene flow.

Hypotheses

We hypothesize that (**i**) if there is ongoing gene flow or these populations diverged recently, then a mitochondrial gene, ND2, would not exhibit clear structure between *atlantica* and *melodia*. Alternatively, if there are barriers to gene flow between the *atlantica* and *melodia* subspecies, then the ND2 gene should clearly differentiate these populations. To evaluate differentiation in ND2 sequences, we used: an analysis of molecular variance (AMOVA), a minimum spanning network (MSN), and a Mantel test.

Laboratory Material & Methods

DNA Samples

After specimen collection, DNA was extracted from blood and stored frozen at the Smithsonian Migratory Bird Center in 2013. As these samples were several years old, their quality was assessed prior to laboratory work. In 2018, gel electrophoresis was used to confirm the presence of intact genomic DNA in each sample, then DNA concentrations were measured using an Invitrogen™ Qubit 4 Fluorometer and a dsDNA HS Assay Kit (Thermo Fisher Scientific). While many of the samples had lower than optimal DNA concentrations, all but three samples registered a detectable amount of DNA. Samples with a DNA concentration <5ng/μL were concentrated by evaporation in a vacuum centrifuge.

ND2 Amplification

A subset of 48 individuals representing both subspecies from across the sampling transect was selected for mitochondrial DNA analysis. A roughly 800bp partial sequence

of the mitochondrial gene, ND2, was sequenced to evaluate if neutral genetic markers could be used to differentiate *atlantica* from *melodia* populations. Amplification of ND2 was performed using the primers L5215 and H6113 and following a protocol adapted from prior studies on avian species (**Table 2**; Gawin et al., 2014; Zou et al., 2007).

Thirty cycles of polymerase chain reaction (PCR) were performed in a 25µl reaction using 1.25µL of each primer at a concentration of 10µM, 0.5 units of Invitrogen™ Platinum II *Taq* Hot-Start DNA polymerase (Thermo Fisher Scientific), and approximately 20ng of template DNA. The thermal cycling regimen began with an initial denaturing at 94ºC for 2.5 minutes, then an annealing period at 57ºC for 30 seconds, followed by an extension period at 72ºC for 1 minute. This was repeated 29 times with denaturation shortened to 30 seconds for all subsequent cycles. At the end of the final cycle, there was an additional 5 minutes of extension time at 72ºC.

Table 2 | Primers used for ND2 amplification

Primer	Sequence
L5215 (Forward)	TATCGGGCCCATACCCCGAAAAT
H6113 (Reverse)	CAGTATGCAAGTCGGAGGTAGAAG

Amplicon Purification

After PCR, amplicon purification was performed using 37.5µL of 1% solids Sera-Mag™ SpeedBeads magnetic particles (MilliporeSigma). The magnetic particles were added directly to each of the 25µL amplicon solutions. Each solution was incubated at room temperature for 5 minutes to allow the amplicons to bind to the magnetic particles, then each solution was exposed to a magnetic stand for 2 minutes to bind the magnetic

particles to the side of the tube. The fluid was then aspirated, leaving behind the magnetic particles and the bound amplicons. The magnetic particles were rinsed twice with 150µL of 70% ethanol while on the magnetic stand. After aspiration of the second ethanol rinse, the magnetic particles were left to dry for approximately 10 minutes. After drying, the magnetic particles were suspended in 12µL ultrapure water and incubated for 2 minutes off the magnetic stand to resuspend the amplicons. Finally, each solution was exposed to a magnetic stand for 2 minutes to bind the magnetic particles, and the 12µL of solution containing the amplicons was aspirated and transferred to a new tube. The presence of amplicons of the appropriate approximate size in the final solutions was confirmed visually using gel electrophoresis.

Thermal Cycle Sequencing

The ND2 amplicons were cycle sequenced using a BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). The same primers used for amplification were used for cycle sequencing (L5215 and H6113). Forward and reverse primers were diluted to 3.2 pM, and separate reactions were performed with 1µL of the dilute forward and reverse primer for each sample. First, 1µL of BigDye™ Terminator Cycle Sequencing Ready Reaction Mix and 3µL of BigDye™ Terminator 5X Sequencing Buffer were added to 3µL aliquots of cleaned ND2 amplicons. Next, 2µL of ultrapure water was added for a total volume of 10µL.

After adding reagents, the samples were immediately thermal cycle sequenced. First, the samples were denatured in the thermal cycler at 96ºC for 1 minute. Then 38 cycles of amplification were performed as follows: 96ºC for 30 seconds, 50ºC for 15 seconds, 60ºC for 4 minutes. Finally, the samples were brought to 12ºC and removed for

purification. Following thermal cycle sequencing, the reactions were purified using Sephadex™ G-50 Superfine (GE Healthcare). The purified samples were then sequenced on an Applied Biosystems™ PRISM 3130xl Genetic Analyzer (Thermo Fisher Scientific).

Analyses

Sequence Assembly

Output files for the forward and reverse reads were assembled in the software *Geneious* using the *Align/Assemble* function on default settings. The assembled contigs were then manually error corrected and trimmed to a length of 818bp. The generated consensus sequences were aligned using the *Align/Assenble* function on the default settings and compared to a partial mitochondrial sequence from a specimen of *M. m. montana* collected in Nevada County, California (GenBank Accession FJ236290; Museum of Vertebrate Zoology at Berkeley #173560). The *Transcribe* function of *Geneious* was then used to confirm the absence of erroneous stop codons in the gene sequences. An alignment of the final consensus sequences was exported from *Geneious* in *NEXUS* file format for analysis.

Analysis of Molecular Variance

Divergence of ND2 haplotypes between populations was tested using an analysis of molecular variance (AMOVA; Excoffier, Smouse, & Quattro, 1992). AMOVA is a common method used to test if the genetic distances between haplotypes of individuals are related to population groupings (see Cardoni et al., 2013). The outputs of an AMOVA are a population differentiation statistic (ϕ_{ST}), which is a value from 0-1 indicating the strength of the differentiation, and a *p*-value indicating significance of the result. If populations have

been reproductively isolated for an extensive time, then ϕ_{ST} should be large (e.g. $\geq$0.25) and the *p*-value should be small (i.e. $\leq$0.05).

We assigned individuals to groups based on samplings sites. Individuals of the *atlantica* subspecies from the coastal sites were grouped (coastal); individuals of the *melodia* subspecies collected at sites on the mainland were grouped (inland); and individuals of the *melodia* subspecies collected upland on the Delmarva Peninsula were grouped (peninsular), as they may have been in contact with *atlantica* populations. A single-level AMOVA was executed in the software *Population Analysis with Reticulated Trees* (*PopART*; Leigh & Bryant, 2015). The ND2 sequence of each individual was associated with its respective regional identity (i.e. coastal, peninsular, or inland) using the *Traits* options in *PopART*.

Minimum Spanning Network

After conducting an AMOVA on the haplotypes, *PopART* was used to generate a minimum spanning network (MSN) with the default tolerance value (ε=0; Bandelt, Forster, & Röhl, 1999). An MSN provides a visualization of the diversity and relative abundance of haplotype sequences within a population and division between populations. Unique haplotype sequences are plotted as circles and connected by lines with marks to indicate the number of nucleotide substitutions between them. If the populations have been reproductively isolated for an extended period, then an MSN should clearly show differentiation in ND2 haplotypes between coastal and inland populations.

Mantel Test

To evaluate if geography accounts for any observed genetic difference in ND2, a Mantel test was performed in *R* using the package *ade4* (Dray & Dufour, 2007; *R* Core

Team, 2013). A Mantel test can be used to determine if there is spatial autocorrelation of pairwise genetic differences making it possible to identify isolation by distance (Arguedas & Parker, 2000; Mantel, 1967). The output of a Mantel test is a regression value (r) indicating the strength and direction of the correlation and a p-value indicating the significance of that relationship.

If subspecific groups are found to be differentiated by an AMOVA, isolation by distance may be assessed as a possible confounding cause of this differentiation (Meirmans, 2012). Although, if subspecific groups are not differentiated by an AMOVA there is still the possibility of significant isolation by distance. If there is a significant effect of isolation by distance, then r should be positive and large (e.g. $\geq$0.50) and the p-value should be small (i.e. $\leq$0.05).

Results

Sequence Assembly

Following sequence assembly in *Geneious*, one individual (1811-53143) was excluded from analyses due to poor sequence quality, bringing the analyzed sample size to 47 individuals. Eleven unique ND2 haplotypes were identified in the sequenced individuals. There were 14 segregating sites identified in the sequences, of which 11 were parsimony-informative sites. Of the haplotypes identified, one accounted for more than half of the sequences and was common in both subspecies and across the sampling transect (Haplotype I in **Fig. 2**).

Analysis of Molecular Variance

The single-level AMOVA showed that there is no significant difference in ND2 sequences between the populations as defined (ϕ_{ST}=0.034, p=0.169). The near-zero ϕ_{ST}

indicated that differentiation was nearly non-existent. The large *p*-value (>0.05) indicated that the observed difference in the populations was not statistically significant.

Minimum Spanning Network

The MSN did not show any clear population structure between subspecies, consistent with the result of the AMOVA (**Fig. 2**). There appeared to be two clustered haplotype groups which were distinct, but they were not associated with subspecies groups. Of the 47 individuals sequenced, 81% were within ≤2 mutations of Haplotype I. The remaining 19% were ≥6 mutations from Haplotype I. Notably, the reference individual from California (not included in the analyses) was an exact match to Haplotype I.

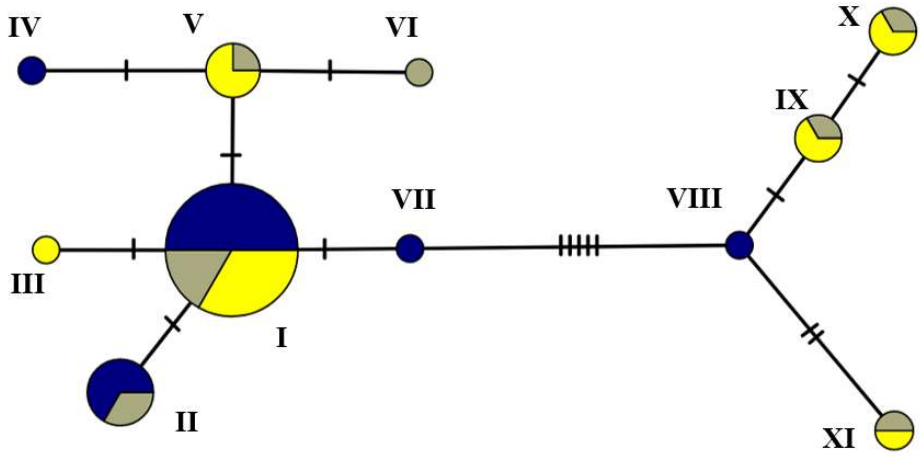

Figure 2 | Minimum spanning network of ND2 haplotypes. Yellow represents *atlantica* (coastal), blue represents mainland *melodia* (proximal inland and distal inland), and grey represents individuals from upland on the Delmarva Peninsula (peninsular). Hatch marks represent a mutation between the haplotypes and the size of the haplotypes corresponds to the proportion of individuals with that sequence.

Mantel Test

The Mantel test confirmed that there is no significant spatial autocorrelation of pairwise genetic distances (r=-0.045, p=0.655). The negative, near-zero r indicated that

there was no correlation between spatial distance and genetic distance of ND2 between individuals. The large *p*-value (>0.05) confirmed that any observed correlation was statistically insignificant.

Discussion

Consistent with our alternative hypothesis (**i**) we found no population structure in a mitochondrial gene, ND2, between populations of *atlantica* and *melodia*. Although mitochondrial genes are not strictly neutral, they can putatively be considered neutral markers (see Ballard & Whitlock, 2004; Hill, 2017). This suggests that these subspecies are not reproductively isolated, and/or they have diverged recently without time for different mutations in the ND2 gene to accumulate. Divergence of populations solely due to varied environmental pressures may mean assortative mating is weak or nonexistent, thus subspecies may interbreed in the broad contact zones where their habitats abut. However, for the maintenance of frequently interbreeding subspecies, selection against migrants and intergrade offspring on either side of the contact zone must inhibit panmixia (Price, 2008; Rundle & Nosil, 2005). In the case of ecological speciation, panmixia is inhibited by intense natural selection against intergrades and/or migrants despite any interbreeding (Rundle & Nosil, 2005).

Little is known about the population genetic structure of the song sparrow where subspecies abut. In the desert southwest, phenotypic intergradation between adjacent populations is common, despite the maintenance of distinct subspecies (Patten & Pruett, 2009; Patten et al., 2004). One of the most focused efforts to study inter-subspecies genetic contact in the song sparrow examined two desert subspecies, *M. m. fallax* and *M. m. heermanni*; across the contact zone, these two subspecies are both phenotypically and

vocally distinct (Patten et al., 2004). Males and females were found to be more responsive to songs of the their own subspecies, and microsatellite variation correlated with plumage differentiation between subspecies (Patten et al., 2004). In this case, neutral genetic differentiation is associated with behavioral reproductive barriers to gene flow caused by subspecific differences in mate recognition by plumage and song. Therefore, inhabitation of panmixia in this contact zone is likely a result of weak behavioral reproductive isolation between adjacent populations that has incrementally increased with isolation by distance, as in ring speciation (Irwin et al., 2001; Patten & Pruett, 2009).

There are anecdotal differences in vocalization of *atlantica* and *melodia*, meaning that there may be some weak assortative mating. If assortative mating does occur in these populations, the reproductive barrier has arisen too recently for genetic drift to accumulate sufficient differences in neutral genes, such as ND2. Alternatively, the lack of population differentiation in the ND2 gene indicates a lack of assortative mating barriers, as neutral genes continue to move between populations uninhibited by selection.

Our results are consistent with the findings of Fry & Zink (1998) who found no population structure in another highly variable mitochondrial gene, the CR, between populations of the song sparrow complex at the continental scale. As the reference individual from California had a sequence identical to the majority of the individuals sequenced, it is likely that there is also a total lack of population structure in the ND2 gene of song sparrows at the continental scale. As suggested by Zink & Dittmann (1993) and Fry & Zink (1998), the radiation of song sparrow diversity may have occurred recently during glacial recession at the end of the Pleistocene. With a warming climate, new niches in shrubland and successional habitats likely became available, allowing the ancestral song

sparrow population to radiate into a diversity of locally adapted subspecies (Pielou, 1991). Where these recently diverged populations are in contact, assortative mating is likely to be weak or non-existent, allowing for the exchange of neutral genes, such as the mitochondrial genome, between distinct populations despite maintenance of distinct phenotypes. Ecologically and phenotypically distinct populations with mostly unrestricted gene flow are sometimes referred to as ecotypes instead of true subspecies, but this distinction can be ambiguous.

The lack of mitochondrial divergence between *atlantica* and *melodia* suggests that these populations do or recently did interbreed and supports the assumption that they diverged with gene flow in parapatry. As mentioned, divergence with gene flow in parapatry due to disruptive ecological selection would prevent differentiation of neutral loci such as ND2. A mitochondrial investigation of a polytypic conger, the swamp sparrow (Passerellidae: *Melospiza georgiana*), also found no divergence in mitochondrial genes despite geographic morphological variation (Greenberg, Cordero, Droege, & Fleischer, 1998). A lack of mitochondrial divergence could alternatively be a result of incomplete lineage sorting despite recent isolation, i.e. when populations became isolated they may each contained the same random set of mitochondrial haplotypes. More clarity on the genetic relationship of the *atlantica* and *melodia* subspecies will thus require the use of methods that capture high-resolution genomic data, allowing for the identification of loci under selection from the environment that may underly phenotypic differences between them and provide more detailed data on the extent of division between the populations. For example, it took whole-genome resequencing efforts to characterize genomic differences

between populations of the golden-winged warbler and blue-winged warbler, which hybridize frequently (Toews, Taylor, et al., 2016).

Historically, *atlantica* and *melodia*, while not fully isolated, were somewhat restricted from contact by habitat differences. Populations of *melodia* were likely unable to occupy the forested habitats adjacent to the coast, as the *melodia* is primarily adapted to the inland successional habitats now associated with human development (Aldrich & Coffin, 1979). Coastal regions of the United States have been disproportionately affected by rapid growth of the human population and the spread of urban sprawl, which has likely made habitats adjacent to the coast more hospitable to *melodia* (Aldrich & Coffin, 1979; Alig, Kline, & Lichtenstein, 2004). It is possible that the human-altered habitats adjacent to the coast may be facilitating a higher degree of contact between *melodia* and *atlantica* than existed historically, exacerbating the extent of gene flow between them.

We believe these data accurately represent this biological system, as they were consistent with findings from a previous mitochondrial studies in similar systems, e.g. the swamp sparrow and the bay-capped wren-spinetail, and a comprehensive mitochondrial study of the song sparrow across North America (Cardoni et al., 2013; Fry & Zink, 1998; Greenberg et al., 1998). Samples were taken along a transect from the beaches of the Atlantic Ocean to the Appalachian Mountains covering habitat gradients from coastal dunes and saltmarshes, to upland coastal habitats, to inland habitats nearly 400km from the coast. The range of habitats represented in the samples sequenced captures the environmental variation that exists across this whole gradient and includes the coastal-adjacent contact zone between these two subspecies.

We believe the sample size of 47 used in this analysis would have been sufficient to identify structure between these populations on the basis of mitochondrial DNA if a deep lineage split was present. A previous study of swamp sparrow mitochondrial variation concluded there was no population structure across a larger geographic area with only 29 samples (Greenberg et al., 1998). A study of the bay-capped wren-spinetail failed to elucidate mitochondrial population structure with 140 samples over a similar spatial scale (Cardoni et al., 2013). Ultimately, if these subspecies were clearly differentiated mitochondrially, then the populations should not have all been dominated by a single haplotype. Haplotype I accounted for 47.1%, 36.4%, and 63.2% of the sequences found in the coastal, peninsular, and inland populations respectively. Further, an inland individual from southwestern Virginia had an ND2 variant more similar to several coastal individuals (Haplotype VIII) than to any other inland individuals. We suspect adding additional individuals to this analysis would not change the overall pattern but rather proportionally add more individuals with Haplotype I and rarer haplotypes, possibly adding a few new rare haplotypes that exist randomly between localities.

While our purpose is not to evaluate the systematic designation of the Atlantic song sparrow as a subspecies, our results demonstrate the difficulty that may arise in classifying phenotypically distinct populations that have diverged ecologically in parapatry when using standard mitochondrial methods. Song sparrow subspecies have been considered by some to be more like plant ecotypes than traditional vertebrate subspecies (Aldrich, 1984). The apparent lack of geographic and reproductive isolation between *atlantica* and *melodia* supports this comparison. But a lack of reproductive barriers between phenotypically distinct organisms does not necessarily merit declassifying them as subspecies. It is our

view that if populations are maintained as a unique evolutionary lineage by divergent ecological selection despite a lack of isolation, then they are an intrinsic biological unit worthy of being systematically classified and protected, i.e. an evolutionarily significant unit (Moritz, 1994, 2002; Paetkau, 1999).

Though conservation biology often treats the population as the unit of conservation, there is an implicit need to maintain variation by preserving genetic diversity (Shafer et al., 2015). Subpopulations with locally-specialized phenotypes, and therefore some unique underlying genetic variants, should thus be given special attention in a conservation context. Populations like the Atlantic song sparrow highlight the complex nature of species and demonstrate how overgeneralized classification might confound conservation planning. In the case of the Atlantic song sparrow, phenotypic differences are likely contained in small portions of the genome under local ecological selection (Nosil, 2008). We assert that these locally adapted genomic islands qualify the Atlantic song sparrow as an evolutionarily significant unit, despite that the majority of their genome may not be reciprocally monophyletic from the eastern song sparrow.

CHAPTER 3 | NEXT-GENERATION POPULATION STRUCTURE ANALYSIS

Ecological Divergence in the Song Sparrow

Long before the advent of contemporary genomic technologies, a researcher at the Smithsonian National Museum of Natural History systematically described morphological variation in song sparrows across the continent and evaluated how this morphology was associate with regional environments (Aldrich, 1984). This endeavor evoked a longstanding debate in biology: does interpopulation variation primarily arise through ecological adaptation of organisms to their local environment or as a result of random genetic drift in isolation? From his investigation of the song sparrow complex, Aldrich (1984) concluded that "…the ecological forces selecting adaptive genetic differences have a greater effect on morphological change or microevolution than do geographical separation or isolation [in the song sparrow complex]… based primarily on the fact that significant morphological variation is noted between populations with no apparent impediment to the exchange of genes other than marked ecological differences in their habitats..." Aldrich went so far as to compare the observed variation in song sparrow populations to ecotypes in plants (see Brandrud et al., 2017). This conclusion was consonant with a prior review of the song sparrow complex that suggested most lineages arose as a result of natural selection without complete geographic isolation (Miller, 1956).

Based on Aldrich's conclusion, we would expect that the genetic differentiation of most parapatric song sparrow subspecies will follow a clinal pattern across a habitat gradient, as selection in different habitats has gradually pulled populations apart. Biological clines were first introduced as a more nuanced taxonomic approach than subspecies for describing interpopulation variation (Huxley, 1938). Put simply a biological cline is variation in a character, either genotypic or phenotypic, that occurs across a spatial dimension of a population's range. This spatial dimension is typically across a contact zone or a habitat gradient.

The limited body of work on mitochondrial genes in the song sparrow has so far supported Aldrich's conclusion, but these studies provide little clarity on how the genomes of song sparrow lineages ecologically diverging with gene flow are differentiated. We sought to corroborate Aldrich's conclusion by elucidating the pattern of population structure in these ecologically differentiated subspecies with contemporary genomic technologies.

Restriction-site Associated DNA Sequencing

Thus far, no contemporary genomic methods have been used to investigate population genetics of the Atlantic song sparrow. As neutral markers are unable to differentiate the subspecies *atlantica* and *melodia*, genome-wide scans are likely necessary to detect genomic islands of differentiation are necessary to investigate the evolutionary relationship of these subspecies. One contemporary method appropriate for genomic investigations of non-model organisms is restriction-site associated DNA sequencing (RADseq), which uses restriction enzymes to isolate loci randomly across the genome but

consistently between individuals (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Davey & Blaxter, 2010; Lemmon & Lemmon, 2013). Several studies have used RADseq to investigate population structure and evolutionary genomics in a variety of taxa (e.g. Brandrud, Paun, Lorenzo, Nordal, & Brysting, 2017; Combs et al., 2018; Hansson et al., 2018; Lim et al., 2017; Ng et al., 2017; Ruegg, Anderson, Boone, Pouls, & Smith, 2014). For example, in the threespine stickleback, RADseq was used to capture population divergence limited to genomic islands of differentiation (Marques et al., 2016).

RADseq methods involve the digestion of genomic DNA by restriction enzymes, which are biochemicals produced naturally in prokaryotes that cleave DNA at enzyme-specific sequences, known as restriction sites, to remove inserted viral DNA from the genome (Krüger & Bickle, 1983). By incubating the DNA of a target organism with certain restriction enzymes, the DNA is cleaved consistently across the genome. Fragments of a certain size range can be purified from the sample, given a molecular tag unique to the individual organism, and sequenced using high-throughput methods (Lemmon & Lemmon, 2013).

Because of the ability to identify thousands of genetic markers dispersed throughout a genome with little or no background information on the target species, RADseq is better-suited than other existing methods for population-level investigations of non-model species. Former techniques for population genetic investigations required screening for specific markers, which involved arduous work prior to a study to discover the markers, develop assays, and validate the assays in subsamples of the population (Davey & Blaxter, 2010). Such methods include screenings for individual microsatellites, single nucleotide

polymorphisms (SNPs), and insertion-deletion polymorphisms (indels). While RADseq methods can identify polymorphisms, such as SNPs, they do not require such screenings of the genome on the front-end. The ability to attain massive amounts of sequence data on each organism also makes RADseq more informative than prior restriction-enzyme-based methods, such as restriction fragment length polymorphisms (RFLPs), which were based on fragment length profiles rather than SNPs within fragments. However, like RFLPs, the use of restriction enzymes reduces the complexity of the genome, making RADseq better-suited for large-scale, exploratory population genetic investigations than whole-genome re-sequencing methods.

A further advantage to NGS methods, such as RADseq, is the ability to multiplex, which is to attach unique molecular tags to samples. Having individual RADseq libraries uniquely tagged allows for the parallel sequencing of several samples at once, known as high-throughput sequencing. The ability to use high-throughput methods makes sequencing of RADseq libraries efficient and economical.

Investigating the song sparrow complex with contemporary genomic methods, such as RADseq, promises to clarify the evolutionary history of subspecies and to elucidate the genomic mechanisms of ecological divergence. Recently, novel improvements have been made to the existing RADseq technology, making it even more affordable, efficient, and reliable. One such improved RADseq method, 3RAD, produces a quadruple-indexed RADseq library for analysis (Glenn et al., 2017). 3RAD has several specific advantages, including a workflow with a reduced number of processing steps. In the 3RAD protocol, ligation of adapters, i.e. enzymatically attaching oligos at the 3’ and 5’ ends of the genomic

DNA that make the sequences readable by Illumina sequencing machines, occurs in the presence of the three restriction enzymes: *Eco*RI, *Xba*I, and *Nhe*I. By ligating adapters in the presence of restriction enzymes, the formation of chimeras, i.e. the binding of formerly non-adjacent DNA fragments, is inhibited. A key characteristic of 3RAD is the addition of the third restriction enzyme, *Nhe*I, which cleaves adapter-dimers, thereby decreasing the quantity of the expensive adapters lost to dimerization during ligation. Inhibiting these adapter dimers also prevents them from contaminating the final libraries and being sequenced. Additionally, modifications to the molecular tagging system allows for reliable pooling of a larger number of samples.

Hypotheses

We hypothesize that (**ii**) *atlantica* and *melodia* have limited genomic islands of differentiation in their nuclear genomes as a result of divergent ecological selection, and therefore RADseq methods will be able to differentiate the subspecies. To discern genomic differentiation in these subspecies, principal component analysis (PCA) and *STRUCTURE* analysis were performed on genotype data from RADseq.

Because *atlantica* and *melodia* are putatively in the early stages of parapatric ecological divergence and likely interbreed, we also hypothesize that (**iii**) there will be a clinal genomic relationship between the subspecies moving inland from *atlantica* habitat, with extensive intergradation in transitional habitats proximal to the coast. To test this, we evaluated the shift in cluster assignment from *STRUCTURE* analysis geographically by plotting it against distance from *atlantica* habitat.

Laboratory Materials & Methods

RADseq Library Preparation

The 218 available DNA samples collected by former research efforts by the Smithsonian Migratory Bird Center and George Mason University along a coastal-to-inland transect, including several samples from saltmarshes in Chesapeake possessed by the Smithsonian National Museum of Natural History (USNM), were prepared into RADseq libraries following the protocol of 3RAD (Glenn et al., 2017). Restriction enzymes used were obtained from New England Biosystems and had concentrations of 20U/μL. For each sample, 20ng of DNA (or 5μL if DNA concentrations were still too low after concentration via vacuum centrifugation) was added to 1.5μL of 10X CutSmart® Buffer (New England Biosystems), 0.5μL *Xba*I, 0.5μL *Eco*RI, 0.5μL *Nhe*I, and 1.0μL each of labeled iTru *Nhe*I and *Eco*RI adapters (Glenn et al., 2016). Thus, the total reaction volume was 15μL. The samples were digested at 37ºC for one hour.

Immediately after restriction digestion, 1.5μL of 10mM rATP, 0.5μL of 10X Ligase Buffer, 0.25μL of 400U/μL DNA ligase, and 2.75μL ultrapure water were added to each reaction, bringing the total volume to 20μL. Ligation of uniquely barcoded read 1 and read 2 sequencing primers (i.e. the iTru adapters) was then performed using the following thermal cycling regime: 22ºC for 20 minutes, 37ºC for 10 minutes, 22ºC for 20 minutes, 37ºC for 10 minutes, and 80ºC for 10 minutes. The samples were then cooled to 10ºC. Half of each reaction was then reserved for archiving.

After restriction digestion and iTru adapter ligation, 10μL from each reaction were pooled into groups of 8. These pooled libraries were then purified using 100μL of 1% solids

Sera-Mag™ SpeedBeads magnetic particles (MilliporeSigma) following manufacture guidelines (see Amplicon Purification in Chapter 2 for detailed procedure). Following purification, the total volume of each pool of 8 libraries was 10µL.

Once the pooled libraries were purified, a cycle of PCR was performed using a KAPA HiFi PCR Kit (Kapa Biosystems) to create a single copy of each fragment with the necessary Illumina sequencing primer and an 8N index unique to each molecule to allow for decloning of raw reads after amplification. To each 10µL pooled library, 10µL of 5X buffer, 1.5µL of 10µM dNTPs, 22.5µL ultrapure water, 1µL of 1U/µL DNA polymerase, and 5µL of 5µM iTru5 8N primer were added, totaling 50µL. The libraries were then thermal cycled once as follows: 98ºC for 60 seconds, 60ºC for 30 seconds, 72ºC for 6 minutes, and cooled to 15ºC.

Purification was performed again following the protocol described above, this time using 70µL of magnetic particles in the 50µL post-PCR solution, then resuspended in 5µL of ultrapure water. Then, 5µL 5X buffer, 0.75µL of 10µM dNTPs, 8.75µL ultrapure water, 0.5µL of 1U/µL DNA polymerase, 2.5µL of 5µM P5 primer, and 2.5µL of 5µM iTru7 primer were added to the 5µL pooled libraries bringing the total volume to 25µL. The samples were initially incubated at 98ºC for 2 minutes, then PCR was performed for 6 cycles as follows: 98ºC for 20 seconds, 60ºC for 15 seconds, 72ºC for 30 seconds. After the final cycle, samples were held at 72ºC for 5 minutes, then cooled to 15ºC.

The samples were purified again following the same procedure but with 37.5µL of magnetic particles and resuspended in 12µL ultrapure water. The DNA concentration of each pooled library was quantified using an Invitrogen™ Qubit 4 Fluorometer (Thermo

Fisher Scientific). The concentration of the samples were found to be too low following library preparation, so the samples were amplified with 5 more cycles of PCR as outlined above and purified again. The pooled libraries were then equimolar pooled into one single library for size selection.

Next-generation Sequencing

Size selection of 450-600bp adapter-ligated amplified restriction fragments was done using an e-Gel precast agarose electrophoresis system (Thermo Fisher Scientific). After size selection, the concentration of the final pooled library of all samples was quantified using an Invitrogen™ Qubit 4 Fluorometer (Thermo Fisher Scientific). The concentration of the size-selected final library was too low for sequencing, so it was amplified a final time with 5 cycles of PCR and then purified and quantified. The final pooled library of all samples was sequenced using a HiSeq 4000 (Illumina), which output 150bp paired-end reads.

Bioinformatic Pipeline

Filtering Reads

The raw reads were received demultiplexed by the i7 index shared by each sub-pool of 8 samples. The program *STACKS* was then used to demultiplex each group of 8 into individual samples by the two inline barcodes annealed in library preparation (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Using the function *process_radtags* in *STACKS*, reads with any uncalled base or those with a quality score of less than 10 in any 10% window of the read were discarded during demultiplexing. Following demultiplexing, the reads of each individual were decloned by the random i5

index to remove error introduced by any biased over-amplified PCR clones from the genotyping.

Parameter Selection

Optimal parameters for genotyping in *STACKS* were selected following methods outlined in previous works (Paris, Stevens, & Catchen, 2017; Rochette & Catchen, 2017). First, a subset of 45 individuals representing both subspecies from across the sampling transect and with varying average coverage per read were selected to create a catalog of loci (Rochette & Catchen, 2017). As no reference genome exists for the song sparrow, these individuals were run through the *de novo* map pipeline (*denovo_map.pl*) built into *STACKS* with varied constraint parameters to evaluate how these parameters affected loci and SNP outputs for the group.

After removing individuals with very few decloned reads (<10,000), analyses were conducted on loci from the catalog found in at least 80% of individuals (hereafter r80 loci), as recommended by the authors of the program (Rochette & Catchen, 2017). The output response of the r80 loci for individuals in the catalog was used as an indicator for effectiveness of different constraint parameters (Rochette & Catchen, 2017). There are two major constraint parameters in *STACKS*: the number of nucleotide differences allowed between alleles of an individual to call them a locus (*M*) and the number allowed between the loci of individuals to call loci the same (*n*). The most constrained value for *M* and *n* that gave a stable number of r80 loci and SNPs for the catalog individuals was chosen, which was 4 (**Fig. 3**). As the populations are in contact, there is no reason for *n* to be greater than *M*, as is the case for isolated populations. The number of unique reads required within

an individual to create an allele (*m*) was left at the default value of *m*=3, as this is generally reliable and unlikely to be erroneous, especially as the reads were decloned (Paris et al., 2017).

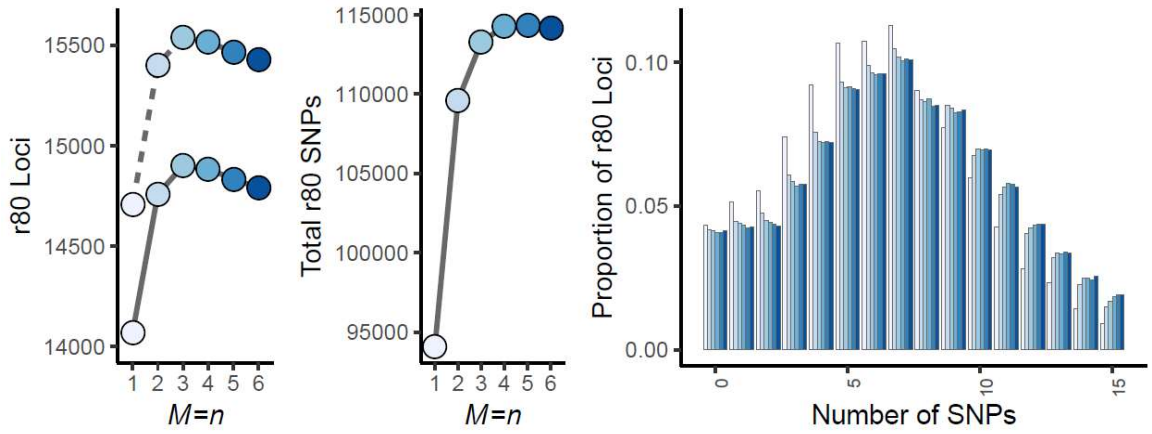

Figure 3 | Outputs of the catalog subset at varied parameters. Stabilization appeared to occur at an *M=n* value of 4 for the number of total (dashed) and polymorphic (solid) r80 loci, the total number of SNPs, and the proportion of loci by number of SNPs (bar colors correspond to *M=n* values of figures to the left).

Genotyping Individuals

After parameter selection, functions in *STACKS* were used to create genotype data for each individual. The function *ustacks* was first used to assemble the alleles of each individual from the decloned reads. Next, *sstacks* was used to match the alleles of each individuals against the catalog of loci from the chosen subset. Then *tsv2bam* and *gstacks* were used to compile this data into genotype information for each individual. Individuals with <10,000 genotyped loci were excluded at this point. Finally, *populations* was used to extract the genotype data of the r80 loci for all individuals as one file for analysis.

Final Genotype Outputs

The final genotype data output from *STACKS* was 11,698 loci with a total of 69,232 SNPs. This robust dataset provides an opportunity for detailed analysis of population structure between these likely interbreeding subspecies. Individuals missing more than 20% of the r80 loci were removed, bringing the total number of individuals for analysis from the initial 218 down to 163.

Analyses

Principal Component Analysis

Principal component analysis (PCA) is an exploratory analysis method that is frequently used to assess population structure from large genotype datasets. For example, a PCA was performed from whole-genome data to assess population structure in the yellow-rumped warbler complex (Toews, Brelsford, et al., 2016). Another study used a method similar to PCA, principal coordinate analysis, to evaluate a hybrid speciation event in South American manakin species (Barrera-Guzmán et al., 2017). We conducted a PCA in *R* using the package *Adegenet* to assess how the genetic data were structured (Jombart, 2008). If an individual was missing data for a particular locus, it was set to the average value to prevent it from affecting the analysis. When using PCA on a large genotype dataset, such as this, the number of principal components (PCs) will equal the number of individuals minus one.

STRUCTURE

The genotype outputs from *STACKS* were analyzed using the program *STRUCTURE*, which evaluates the ancestry of organisms using a Bayesian clustering

method based on Markov Chain Monte Carlo (MCMC) estimation (Porras-Hurtado et al., 2013; Pritchard, Stephens, & Donnelly, 2000). The program was run with the assumed number of ancestral populations (*K*) set to 2 and 3. As *STRUCTURE* requires unlinked alleles, the data were output from *STACKS* using only the first SNP of each locus, (Pritchard et al., 2000; Rochette & Catchen, 2017). The program was run with a burnin period of 5000 MCMC steps and 50,000 post-burnin MCMC steps.

The output of *STRUCTURE* analysis is a matrix of membership coefficients, which are scores indicating what portion of an individual's genome belongs to each infered ancestral genetic cluster. This process was conducted using the admixture model and with a constant allele frequencies parameter (λ) of 1.0, which accounts for the possibility that populations interbreed and considers allele frequencies to be independent between loci. Because these computations rely on a naïve Bayesian approach, which treats each locus within the organisms' genome as an independent variable, small sample size can severely affect the results (Porras-Hurtado et al., 2013). The data available, i.e. >11,000 loci from 163 individuals, were sufficient to quell these concerns.

Geographic Cline

Geographic clines of genotypic traits are commonly used to evaluate how divergent selection affects populations across a contact zone (Arntzen et al., 2016; Brumfield et al., 2001). In the past, these geographic cline analyses may have evaluated the shift in allele frequency of one or few, putatively neutral markers. With more robust genomic data, such as the results of *STRUCTURE* analysis from thousands of RAD-loci, geographic clines can be used to summarize more complex genomic relationships. We plotted assignment to the

apparent *atlantica* cluster from *STRUCTURE* analysis against distance from *atlantica* habitat to determine how genomic ancestry varies across the sampling transect.

Results

Principal Component Analysis

The PCA did exhibit differentiation of individuals based on sampling region. The first two of the 162 principal components (PC1 and PC2) respectively accounted for 1.31% and 1.02% of the observed variance in the data, totaling 2.33%. While this value is low, it may still be usefule for identifying a pattern of differentiation between populations.

A strong clinal relationship can be observed along the most informative principal component, PC1, with the axis value increasing with distance from coastal habitat (**Fig. 4**). There is a notable overlap of peninsular and coastal individuals, suggesting that there may be extensive gene flow from the *atlantica* population into upland peninsular habitats. Distal inland individuals from sites in extreme western Maryland and southwestern Virginia cluster strongly together on PC1 despite their substantial geographic distance from each other, which suggests that the pattern observed on PC1 is driven by habitat more strongly than isolation by distance. The main pattern captured by PC2 appears to be variation within the *atlantica* population. Additional principal components after PC2, e.g. PC3 and PC4, do not show any clear patterns related to geography or habitat and thus may be separating closely related individuals from the rest of the samples.

STRUCTURE

The results of Bayesian clustering MCMC exhibited patterns consistent with our hypotheses. The pattern of ancestry was most clearly captured with the inferred number

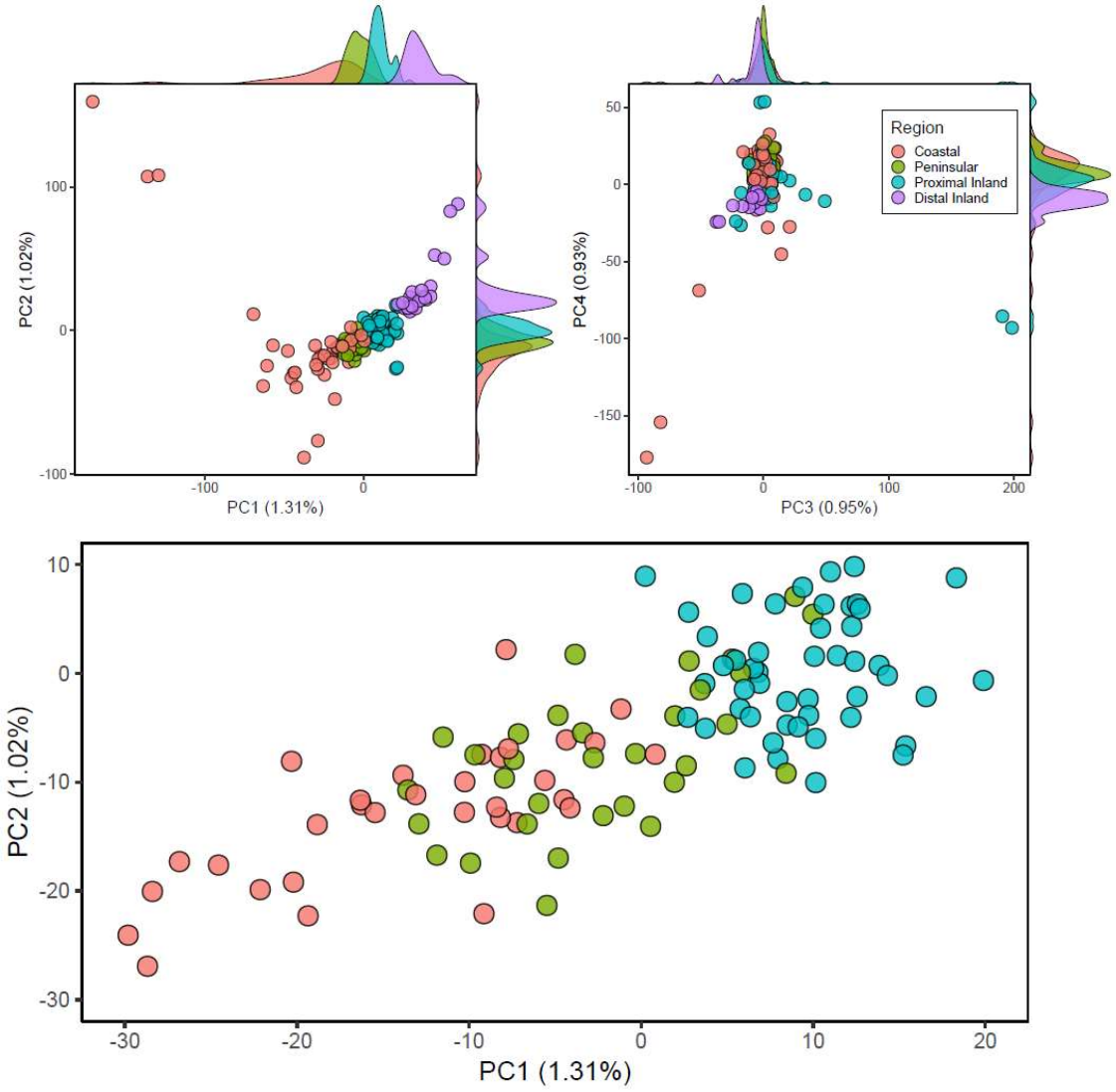

Figure 4 | Principal component analysis of SNPs from RADseq. There is a clear clinal differentiation by transect region along PC1 with variation within the *atlantica* populations illustrated on PC2. The addition of PC3 and PC4 does not appear to show any strong geographic pattern, but rather separates closely related individuals from the rest of the samples. Curves represent frequency of samples along the respective axis. A close up of the boundary between coastal and proximal inland individuals on PC1 and PC2 shows that these populations are discrete from one another and bridged by upland peninsular individuals, possible evidence of intergradation.

of ancestral populations (*K*) set to 2, as the estimated natural log of the probability of the data dropped sharply when *K* was changed to 3. Individuals from distal inland sites all had an assignment to the *atlantica* cluster of <11% with an average of 3.85%. The average assignment to the *atlantica* cluster for the proximal inland, peninsular, and coastal populations were 32.58%, 45.28%, 66.14%, respectively (**Fig. 5**). The standard deviations for assignment to the *atlantica* cluster for distal inland, proximal inland and peninsular populations were respectively 2.67%, 6.19%, 5.97%, while in the coastal populations the standard deviation was 17.07%. The relatively low average assignment in the coastal populations with a large standard deviation suggests that there may be extensive gene flow from *melodia* populations into populations designated as *atlantica*.

There is a strong correlation of PC1 with the estimated level of *atlantica* ancestry (**Fig. 6**), which corroborates the clinal differentiation in populations along the coastal-to-inland habitat gradient. Assignment to the *atlantica* cluster more clearly separates distal inland individuals from proximal inland individuals than PCA, suggesting the results may be more informative than PCA (**Fig. 5** and **Fig. 6**).

When *K* is set to 3 for *STRUCTURE* analysis, the individuals sampled on Assateague Island (the first 13 bars in **Fig. 5**) are the only ones substantially impacted. The three individuals which have a majority assignment to the Assateague cluster are also the three outliers on PC1 from PCA (**Fig. 6**). This may represent the isolation of *atlantica* individuals on Assateague Island from the mainland *melodia* population or be a result of some drift in isolation. However, as *K*=3 is not the most supported model, the Assateague cluster may simply be a result of close relation of the 3 unique individuals. The 13 birds

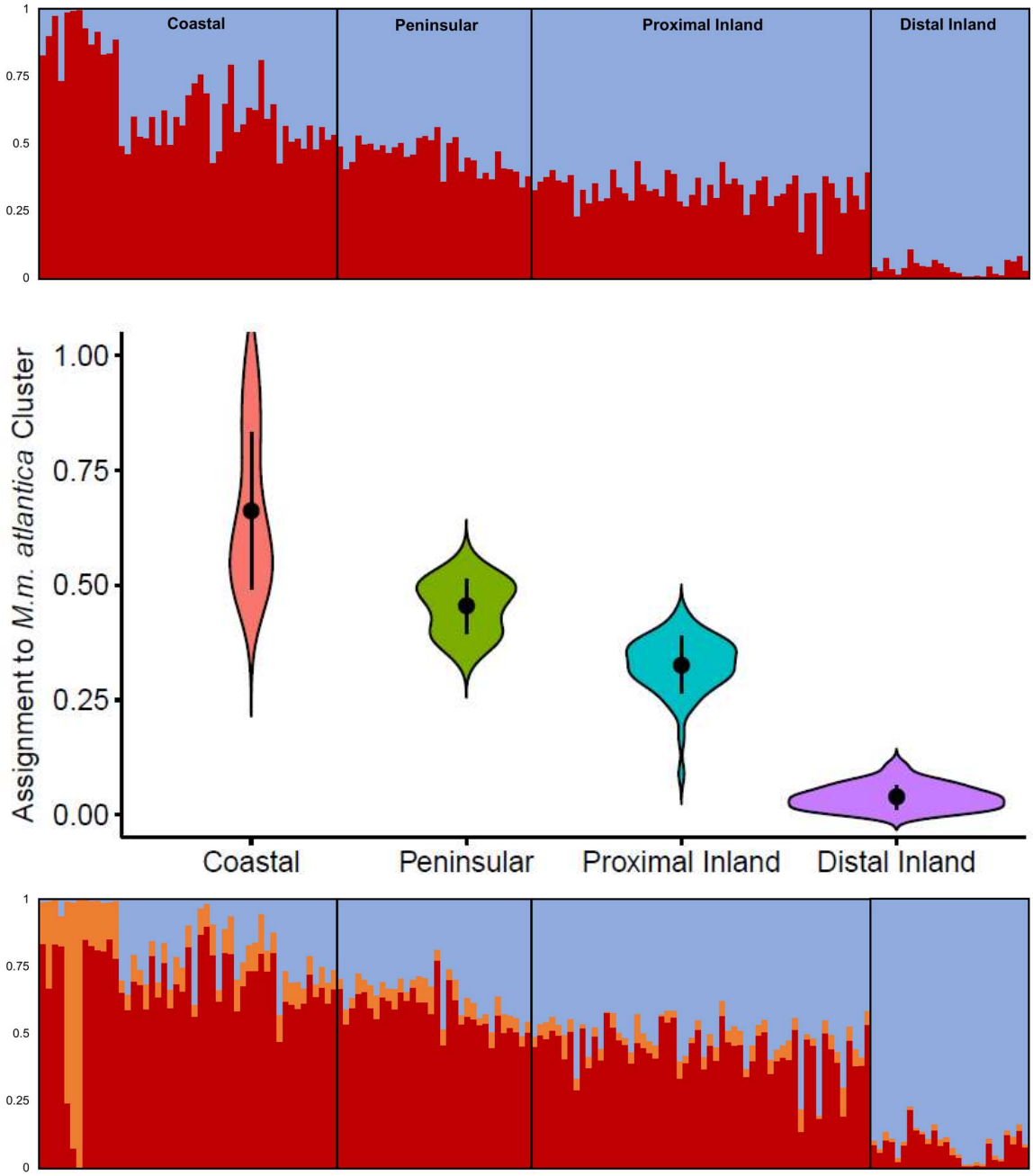

Figure 5 | Cluster assignments from *STRUCTURE* analysis. Barplots of individuals by sampling region with *K*=2 (top) and *K*=3 (bottom); red is the apparent *atlantica* cluster, blue is the apparent *melodia* cluster, and the orange where *K*=3 appears to be associated mainly with *atlantica* individuals sampled on Assateague Island. Violin plots (center) of ancestry assignment to the *atlantica* cluster by region with *K*=2; dots are the average value with lines for standard error.

from Assateague Island were also the most associated with the *atlantica* cluster with *K*=2, suggesting that this site typifies genetically pure *atlantica* individuals.

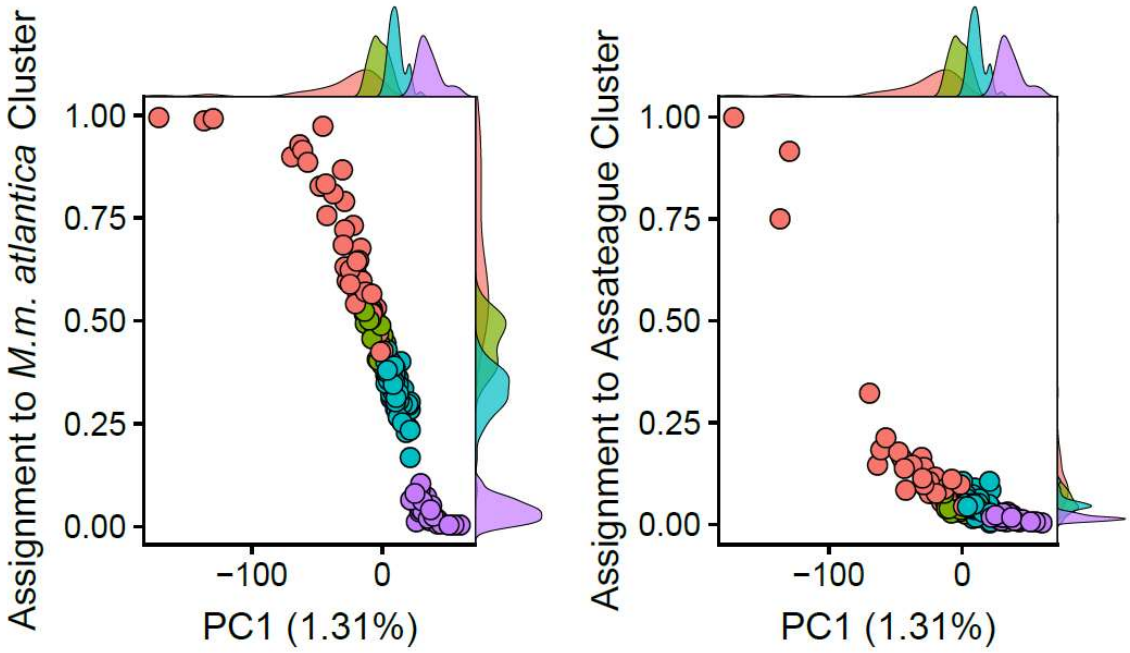

Figure 6 | Correlation of PCA with *STRUCTURE* analysis. PC1 strongly correlates with assignment to the *atlantica* cluster from *STRUCTURE* analysis (*K*=2). The three outliers on PC1 are associated with the Assateague cluster (*K*=3). Pink, green, blue, and purple correspond to coastal, peninsular, proximal inland, and distal inland, respectively as in **Fig. 5**.

Geographic Cline

Standard geographic cline analysis of genotypic traits evaluates the shifts in the allele frequencies of individual loci across a spatial axis. The output of *STRUCTURE* analysis allows us to capture genomic resolution of shifts in genotypes across a spatial dimension, as we can plot shifts in the ancestry assignment of individuals. When assignment to the *atlantica* cluster from *STRUCTURE* analysis was plotted against the

approximate distance inland from coastal saltmarsh and dune habitats, a clear geographic cline in genetics can be observed (**Fig. 7**).

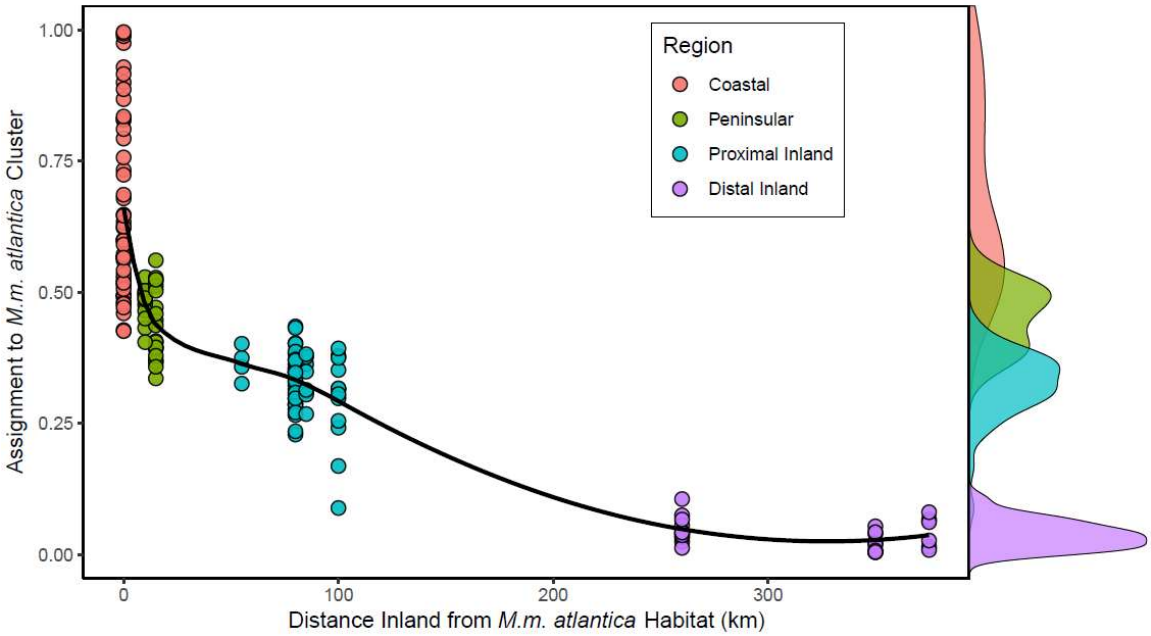

Figure 7 | Geographic cline of ancestry assignment. Assignment to the *atlantica* cluster from *STRUCTURE* analysis plotted by position along the coastal-to-inland transect. There is a clear clinal relationship in the genomes of the subspecies across the habitat gradient illustrated by a locally estimated scatterplot smoothing (LOESS) line.

The steep cline in genetic ancestry assignment supports the hypothesis that differentiation between the subspecies may be a consequence of natural selection despite gene flow. All peninsular individuals appear to be genetic intermediates of inland *melodia* and coastal *atlantica* populations, with extensive gene flow appearing to extending inland for at least 100km. Gene flow into coastal environments may be extreme, as 61% of sampled individuals from coastal habitats had an ancestry assignment to the *atlantica* cluster of <65%. The three individuals from saltmarshes in the Chesapeake Bay had a

greater than average assignment to the *atlantica* cluster, confirming that these habitats are used by *atlantica* (**Fig. 1**).

Discussion

As hypothesized, (**ii**) RADseq analyses of these DNA samples yielded discernable genetic population structure between the coastal *atlantica* populations and inland *melodia* populations. We confirmed that inland populations from distant sites are more similar genetically to each other than they are to those from sites proximal to coastal habitats. We found (**iii**) clinal variation in genotypes that corresponded to the occupation of different ecological habitats, consistent with what would arise during parapatric ecological divergence.

While a clinal pattern can also result from isolation by distance, the genetic similarity of distant inland populations suggests that this is not the case. A clinal pattern can also emerge by fusion in secondary contact, but based on the current geographic distribution of these subspecies, we suspect that they have never been strictly isolated and thus diverged with contact. The pattern observed in PCA was able to demonstrate a clear clinal relationship between sampling regions moving away from the coast and identified substantial variation within the *atlantica* population. This was confirmed by *STRUCTURE* analysis, which primarily separated individuals clinally by sampling region, and secondarily identified differences between *atlantica* populations.

Because so much of the coastal population consisted of genetic intermediates, selection against intergrades in littoral habitats could be weak or nonexistent, or an influx of individuals with primarily *melodia* ancestry has recently overwhelmed selection against

them. The elevated differentiation of *atlantica* individuals on the relatively unchanged Assateague Island from less isolated sites supports the idea that human modifications to the landscape may be driving fusion of these subspecies in populous areas. As human populations around the coast have grown, coast-adjacent habitats have become more suitable for *melodia* individuals, allowing for an unnatural amount of immigration from *melodia* populations into coastal habitats. Human development, therefore, may be exposing the *atlantica* population to a detrimental amount of gene flow, i.e. genetic swamping.

When genetic ancestry assignment was plotted against distance from *atlantica* habitat, a clear cline in ancestry assignment was observed, allowing for inferences to be drawn about how natural selection may affect gene flow between these subspecies (**Fig. 7**). Peninsular sites in the tension zone adjacent to the coast were strictly inhabited by genetic intermediates, which may indicate that there is hybrid superiority in transitional habitats, i.e. phenotypic intermediates of the subspecies are selected for in non-littoral habitats near the coast. Genetically intermediate individuals, which may also be morphologically intermediate, could strike the balance between the selective advantage of having a large bill in coast-adjacent habitats and having a smaller bill that does not lose excess heat during the winter (Danner & Greenberg, 2015). As the selective advantage for *atlantica* phenotypes decays moving inland, gene flow between the subspecies could be inhibited, resulting in the geographic cline observed from *STRUCTURE* analysis (**Fig. 7**).

The observed pattern of genetic ancestry assignment of *atlantica* and *melodia* populations across a habitat gradient is comparable to what has been described from RADseq investigations of parapatric ecotypes of common scurvygrass (Brandrud et al.,

2017). In common scurvy grass, populations from habitats with different levels of salinity are morphologically distinct due to divergent selection. Because mitochondrial DNA could not differentiate *atlantica* from *melodia* but RAD-loci could to an extent, we believe that genomic islands of differentiation may be responsible for many of the phenotypic differences between the subspecies. Genomic islands of differentiation mediated by ecological pressure have been described in populations of threespine stickleback that occupy different habitats (Marques et al., 2016). Habitat-mediated divergence of small portions of the genome is likely also the reason that thorough mitochondrial and microsatellite assays of the bay-capped wren-spinetail were unable to differentiate coastal and upland ecotypes (Cardoni et al., 2013). However, reproductive isolation must somehow arise as a byproduct of ecological divergence for genomic islands fixed between populations by natural selection to result in a speciation event (Rundle & Nosil, 2005).

Clapper and king rails are believed to be the end product of such an ecological speciation event, as some intrinsic selection pressures limit their hybridization (Maley, 2012). While extremely strong ecological pressures can inhibit panmixia and push populations apart into separate species, ideally assortative mating would concurrently evolve causing pre-zygotic reproductive isolation, as in populations of the red crossbill where females prefer males with beaks locally adapted for efficient foraging (Parchman et al., 2006; Snowberg & Benkman, 2009). Genomic islands of differentiation, like those that we suspect are responsible for morphological difference between *atlantica* and *melodia*, can be enough to create reproductive barriers between populations, as in divergent populations of malaria mosquitos (Turner et al., 2005).

If *atlantica* and *melodia* are maintained solely by natural selection without strong pre-zygotic or post-zygotic reproductive barriers, then *atlantica* may rapidly fuse back into *melodia* in response to environmental shifts or if *melodia* individuals migrate into *atlantica* habitat at a rate that overwhelms selection against them. Divergent morphs of the peppered moth (Geometridae: *Biston betularia*) are undergoing such fusion following reductions in the pollution levels that once gave dark morphs a selective advantage in urban environments (Saccheri, Rousset, Watts, Brakefield, & Cook, 2008). In the case of *atlantica* and *melodia*, if coast-adjacent habitats have become recently accessible to *melodia* due to anthropogenic development, then genetic swamping from nearby source populations of *melodia* may erode the morphological distinctions that define *atlantica*.

While it is not our goal to discern the precise boundaries between a subspecies and an ecotype in terrestrial vertebrates, our findings corroborated the comparison by Aldrich (1984) of song sparrow populations to plant ecotypes. That is, much of song sparrow diversity is likely the result of local ecological selection acting on a limited set of loci rather than the phenome of the whole organism. In this case, ecologically-selected loci should be divergent between populations, despite panmixia elsewhere in the genome. For such ecological divergence to result in speciation would require reproductive isolation to somehow arise as a byproduct of the divergent ecological selection

Our study highlights the need for further, comprehensive genomic studies of the song sparrow complex. To fully assess what makes *atlantica* and *melodia* populations different will require generating a reference genome, performing whole-genome resequencing, and evaluating genome-wide associations. The foundations built while

identifying genomic differences between *atlantica* and *melodia* can then be used to evaluate divergence within the entire song sparrow complex. Ultimately, these populations provide an ideal model for investigating the beginnings of speciation on the genomic level. Reassessing song sparrow subspecies with comprehensive genomic studies is likely to challenge our current understanding of systematics and speciation, but provide data, which will bring us into a more complete understanding of the genomic processes underlying evolution.

The pattern of genomic divergence illustrated by populations of *atlantica* and *melodia* in contact shows how speciation with gene flow can occur as a result of strong natural selection pressures. Characterizing evolutionarily significant units in cases where populations have diverged recently and without sufficient isolation, will likely requiring further studies at the scale of the whole-genome, such as this. Where whole-genome resequencing is not preferred or feasible, we have demonstrated that modern RADseq methods can provide useful data on populations with complex patterns of genetic divergence.

Speciation is a complex process, with no single path. The new age of accessible genomic technologies in which we live has created endless possibilities to investigate the underlying processes that lead to speciation. The more genomic data created on diverging populations, the more nuanced our understanding of evolution will become. Despite our best efforts, we may never be able to fully conceptualize all the complex ways that populations of organisms over time give rise to new species.

REFERENCES

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, *305*(5689). doi: 10.1126/science.1098095

Aldrich, J. W. (1984). Ecogeographical variation in size and proportions of song sparrows (*Melospiza melodia*). *Ornithological Monographs*, (35), 1–134. doi: 10.2307/40166779

Aldrich, J. W., & Coffin, R. W. (1979). Breeding bird populations from forest to suburbia after thirtyeven years. *American Birds*, *34*, 3–7.

Alig, R. J., Kline, J. D., & Lichtenstein, M. (2004). Urbanization on the US landscape: looking ahead in the 21st century. *Landscape and Urban Planning*, *69*(2), 219–234. doi: 10.1016/j.landurbplan.2003.07.004

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. doi: 10.1038/nrg.2015.28

Arguedas, N., & Parker, P. G. (2000). Seasonal Migration and Genetic Population Structure in House Wrens. *The Condor*, *102*(3), 517–528. doi: 10.2307/1369782

Arntzen, J. W., Trujillo, T., Butôt, R., Vrieling, K., Schaap, O., Gutiérrez-Rodríguez, J., & Martínez-Solano, I. (2016). Concordant morphological and molecular clines in a contact zone of the common and spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. *Frontiers in Zoology*, *13*. doi: 10.1186/s12983-016-0184-7

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744. doi: 10.1046/j.1365-294X.2003.02063.x

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Barrera-Guzmán, A. O., Aleixo, A., Shawkey, M. D., & Weir, J. T. (2017). Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proceedings of the National Academy of Sciences*, 201717319. doi: 10.1073/pnas.1717319115

Benkman, C. W. (2003). Divergent selection drives the adaptive radiation of crossbills. *Evolution*, *57*(5), 1176. doi: 10.1554/0014-3820(2003)057[1176:DSDTAR]2.0.CO;2

Boag, P., & Grant, P. (1981). Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galápagos. *Science*, *214*, 82–85.

Brandrud, M. K., Paun, O., Lorenzo, M. T., Nordal, I., & Brysting, A. K. (2017). RADseq provides evidence for parallel ecotypic divergence in the autotetraploid *Cochlearia officinalis* in Northern Norway. *Scientific Reports*, *7*(1), 5573. doi: 10.1038/s41598-017-05794-z

Brumfield, R. T., Jernigan, R. W., McDonald, D. B., & Braun, M. J. (2001). Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. *Evolution*, *55*(10), 2070–2087. Retrieved from http://www.jstor.org/stable/2680453

Burtt, E. H., & Ichida, J. M. (2004). Gloger's rule, feather-degrading bacteria, and color variation among song sparrows. *The Condor*, *106*(3), 681–690. doi: 10.1650/7383

Cardoni, D. A., Greenberg, R., Maldonado, J. E., & Isacch, J. P. (2013). Morphological adaptation to coastal marshes in spite of limited genetic structure in the Neotropical passerine *Spartonoica maluroides* (Aves: Furnariidae): South American Salt Marsh Bird Evolution. *Biological Journal of the Linnean Society*, *109*(1), 78–91. doi: 10.1111/bij.12019

Caro, L. M., Caycedo-Rosales, P. C., Bowie, R. C. K., Slabbekoorn, H., & Cadena, C. D. (2013). Ecological speciation along an elevational gradient in a tropical passerine bird? *Journal of Evolutionary Biology*, *26*(2), 357–374. doi: 10.1111/jeb.12055

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

Cattau, C. E., Jr, R. J. F., Kimball, R. T., Miller, C. W., & Kitchens, W. M. (2018). Rapid morphological change of a top predator with the invasion of a novel prey. *Nature Ecology & Evolution*, *2*(1), 108. doi: 10.1038/s41559-017-0378-1

Chan, Y., & Arcese, P. (2002). Subspecific differentiation and conservation of song sparrows (*Melospiza melodia*) in the San Francisco Bay region inferred by microsatellite loci analysis. *Auk*, *119*(3), 641–657. doi: http://dx.doi.org.mutex.gmu.edu/10.1043/0004-8038(2002)119(0641:SDACOS)2.0.CO;2

Combs, M., Byers, K. A., Ghersi, B. M., Blum, M. J., Caccone, A., Costa, F., … Munshi-South, J. (2018). Urban rat races: spatial population genomics of brown rats (*Rattus norvegicus*) compared across multiple cities. *Proc. R. Soc. B*, *285*(1880), 20180245. doi: 10.1098/rspb.2018.0245

Cooney, C. R., Bright, J. A., Capp, E. J. R., Chira, A. M., Hughes, E. C., Moody, C. J. A., … Thomas, G. H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, *542*(7641), 344–347. doi: 10.1038/nature21074

Danner, R. M., & Greenberg, R. (2015). A critical season approach to Allen's rule: bill size declines with winter temperature in a cold temperate environment. *Journal of Biogeography*, *42*(1), 114–120. doi: 10.1111/jbi.12389

Danner, R. M., Gulson-Castillo, E. R., James, H. F., Dzielski, S. A., Frank, D. C., Sibbald, E. T., & Winkler, D. W. (2016). Habitat-specific divergence of air conditioning structures in bird bills. *The Auk*, *134*(1), 65–75. doi: 10.1642/AUK-16-107.1

Danner, R. M., Olsen, B. J., & Luther, D. (2016). Migratory status, winter subspecies interactions, and habitat segregation of Atlantic song sparrows (*Melospiza melodia atlantica*). *The Wilson Journal of Ornithology; Fort Collins*, *128*(2), 434–437. Retrieved from https://search.proquest.com/docview/1806425321

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, *9*(5–6), 416–423. doi: 10.1093/bfgp/elq031

Dray, S., & Dufour, A.-B. (2007). The *ade4* package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(1). doi: 10.18637/jss.v022.i04

Emelianov, I., Marec, F., & Mallet, J. (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1534), 97–105. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1691566/

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491. Retrieved from https://www.genetics.org/content/131/2/479

Ford, E. B. (1964). *Ecological genetics*. London: Chapman and Hall.

Fry, A. J., & Zink, R. M. (1998). Geographic analysis of nucleotide diversity and song sparrow (Aves: Emberizidae) population history. *Molecular Ecology*, *7*(10), 1303–1313. doi: 10.1046/j.1365-294x.1998.00462.x

Gawin, D. F., Rahman, M. A., Ramji, M. F. S., Smith, B. T., Lim, H. C., Moyle, R. G., & Sheldon, F. H. (2014). Patterns of avian diversification in Borneo: The case of the endemic mountain black-eye (*Chlorocharis emiliae*). *The Auk*, *131*(1), 86–99. doi: 10.1642/AUK-13-190.1

Geist, N. R. (2000). Nasal respiratory turbinate function in birds. *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches*, *73*(5), 581–589. doi: 10.1086/317750

Glenn, T. C., Bayona-Vasquez, N. J., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., … Faircloth, B. C. (2017). Adapterama III: quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per Sample (3RAD). *BioRxiv*, 205799. doi: 10.1101/205799

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Finger, J. W., Pierson, T. W., Bentley, K. E., … Faircloth, B. C. (2016). Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *BioRxiv*, 049114. doi: 10.1101/049114

Grant, P. R., & Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, *296*(5568), 707. doi: 10.1126/science.1070315

Greenberg, R., Cadena, V., Danner, R. M., & Tattersall, G. (2012). Heat loss may explain bill size differences between birds occupying different habitats. *PLOS ONE*, *7*(7), e40933. doi: 10.1371/journal.pone.0040933

Greenberg, R., Cordero, P. J., Droege, S., & Fleischer, R. C. (1998). Morphological adaptation with no mitochondrial DNA differentiation in the coastal plain swamp sparrow. *The Auk*, *115*(3), 706–712. doi: 10.2307/4089418

Greenberg, R., & Danner, R. M. (2012). The influence of the California marine layer on bill size in a generalist songbird. *Evolution*, *66*(12), 3825–3835. doi: 10.1111/j.1558-5646.2012.01726.x

Greenberg, R., Danner, R., Olsen, B., & Luther, D. (2012). High summer temperature explains bill size variation in salt marsh sparrows. *Ecography*, *35*(2), 146–152. doi: 10.1111/j.1600-0587.2011.07002.x

Hansson, B., Sigeman, H., Stervander, M., Tarka, M., Ponnikas, S., Strandh, M., … Hasselquist, D. (2018). Contrasting results from GWAS and QTL mapping on wing length in great reed warblers. *Molecular Ecology Resources*, *18*(4), 867–876. doi: 10.1111/1755-0998.12785

Hill, G. E. (2017). The mitonuclear compatibility species concept. *The Auk*, *134*(2), 393–409. doi: 10.1642/AUK-16-201.1

Hubbard, J. P. (1969). The relationships and evolution of the *Dendroica coronata* complex. *The Auk*, *86*, 393–432.

Huxley, J. (1938). Clines: An auxiliary taxonomic principle. *Nature*, *142*(3587), 219. doi: 10.1038/142219a0

Irwin, D. E., Bensch, S., & Price, T. D. (2001). Speciation in a ring. *Nature*, *409*(6818), 333. doi: 10.1038/35053059

Jombart, T. (2008). *adegenet*: a *R* package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624. doi: 10.1038/217624a0

Kimura, M. (1979). The neutral theory of molecular evolution. *Scientific American*, *241*(5), 98–129. Retrieved from http://www.jstor.org/stable/24965339

Krüger, D. H., & Bickle, T. A. (1983). Bacteriophage survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbioloical Reviews*, *47*(3), 16.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., … Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371–375. doi: 10.1038/nature14181

Leigh, J. W., & Bryant, D. (2015). *PopART*: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. doi: 10.1111/2041-210X.12410

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 99–121. doi: 10.1146/annurev-ecolsys-110512-135822

Lim, H. C., Gawin, D. F., Shakya, S. B., Harvey, M. G., Rahman, M. A., & Sheldon, F. H. (2017). Sundaland's east–west rain forest population structure: variable manifestations in four polytypic bird species examined using RAD-seq and plumage analyses. *Journal of Biogeography*, *44*(10), 2259–2271. doi: 10.1111/jbi.13031

Luther, D., & Danner, R. (2016). Males with larger bills sing at higher rates in a hot and dry environment. *The Auk*, *133*(4), 770–778. doi: 10.1642/AUK-16-6.1

Luther, D., & Greenberg, R. (2014). Habitat type and ambient temperature contribute to bill morphology. *Ecology and Evolution*, *4*(6), 699–705. doi: 10.1002/ece3.911

Maley, J. M. (2012). *Ecological speciation of king rails (Rallus elegans) and clapper rails (Rallus longirostris)* (Ph.D.). Louisiana State University.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220. Retrieved from http://cancerres.aacrjournals.org/content/27/2_Part_1/209

Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLOS Genetics*, *12*(2), e1005887. doi: 10.1371/journal.pgen.1005887

Mayr, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. doi: 10.1111/j.1365-294X.2012.05578.x

Milá, B., Toews, D. P. L., Smith, T. B., & Wayne, R. K. (2011). A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (Aves: *Dendroica coronata*). *Biological Journal of the Linnean Society*, *103*(3), 696–706. doi: 10.1111/j.1095-8312.2011.01661.x

Miller, A. H. (1956). Ecologic factors that accelerate formation of races and species of terrestrial vertebrates. *Evolution*, *10*(3), 262–277. doi: 10.1111/j.1558-5646.1956.tb02851.x

Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, *3*(4), 401–411. doi: 10.1111/j.1365-294X.1994.tb00080.x

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, *51*(2), 238–254.

Nabholz, B., Lanfear, R., & Fuchs, J. (2016). Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology*, *25*(18), 4438–4449. doi: 10.1111/mec.13780

Ng, N. S. R., Wilton, P. R., Prawiradilaga, D. M., Tay, Y. C., Indrawan, M., Garg, K. M., & Rheindt, F. E. (2017). The effects of Pleistocene climate change on biotic differentiation in a montane songbird clade from Wallacea. *Molecular Phylogenetics and Evolution*, *114*, 353–366. doi: 10.1016/j.ympev.2017.05.007

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2103–2106. doi: 10.1111/j.1365-294X.2008.03715.x

Paetkau, D. (1999). Using genetics to identify intraspecific conservation units: A critique of current methods. *Conservation Biology*, *13*(6), 1507–1509. doi: 10.1046/j.1523-1739.1999.98507.x

Parchman, T. L., Benkman, C. W., & Britch, S. C. (2006). Patterns of genetic variation in the adaptive radiation of New World crossbills (Aves: *Loxia*). *Molecular Ecology*, *15*(7), 1873–1887. doi: 10.1111/j.1365-294X.2006.02895.x

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373. doi: 10.1111/2041-210X.12775

Patten, M. A., & Pruett, C. L. (2009). The Song Sparrow, *Melospiza melodia*, as a ring species: patterns of geographic variation, a revision of subspecies, and implications for speciation. *Systematics and Biodiversity; Oxford*, *7*(1), 33–62. doi: http://dx.doi.org/10.1017/S1477200008002867

Patten, M. A., Rotenberry, J. T., & Zuk, M. (2004). Habitat selection, acoustic adaptation, and the evolution of reproductive isolation. *Evolution*, *58*(10), 2144–2155.

Pielou, E. C. (1991). *After the Ice Age: The Return of Life to Glaciated North America*. University of Chicago Press.

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*, *51*(2), 537–551. doi: 10.2307/2411126

Podos, J. (2010). Acoustic discrimination of sympatric morphs in Darwin's finches: A behavioural mechanism for assortative mating? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1543), 1031–1039. doi: 10.1098/rstb.2009.0289

Podos, J., Southall, J. A., & Rossi-Santos, M. R. (2004). Vocal mechanics in Darwin's finches: Correlation of beak gape and song frequency. *Journal of Experimental Biology*, *207*(4), 607–619. doi: 10.1242/jeb.00770

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of *STRUCTURE*: applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*. doi: 10.3389/fgene.2013.00098

Price, T. (2008). *Speciation in Birds* (1st ed.). Greenwood Village, Colorado, USA: Roberts and Company.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

R Core Team. (2013). R*: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/

Rico-Guevara, A., Rubega, M. A., Hurme, K. J., & Dudley, R. (2019). Shifting paradigms in the mechanics of nectar extraction and hummingbird bill morphology. *Integrative Organismal Biology*, *1*(1). doi: 10.1093/iob/oby006

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using *Stacks*. *Nature Protocols*, *12*(12), 2640–2659. doi: 10.1038/nprot.2017.123

Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, *23*(19), 4757–4769. doi: 10.1111/mec.12842

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. doi: 10.1111/j.1461-0248.2004.00715.x

Ryeland, J., Weston, M. A., & Symonds, M. R. E. (2017). Bill size mediates behavioural thermoregulation in birds. *Functional Ecology*, *31*(4), 885–893. doi: 10.1111/1365-2435.12814

Saccheri, I. J., Rousset, F., Watts, P. C., Brakefield, P. M., & Cook, L. M. (2008). Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16212–16217. doi: 10.1073/pnas.0803785105

Salewski, V., & Watt, C. (2017). Bergmann's rule: a biophysiological rule examined in birds. *Oikos*, *126*(2). doi: 10.1111/oik.03698

Servedio, M. R., Doorn, G. S. V., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: 'Magic' but not rare? *Trends in Ecology & Evolution*, *26*(8), 389–397. doi: 10.1016/j.tree.2011.04.005

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*(2), 78–87. doi: 10.1016/j.tree.2014.11.009

Snowberg, L. K., & Benkman, C. W. (2009). Mate choice based on a key ecological performance trait. *Journal of Evolutionary Biology*, *22*(4), 762–769. doi: 10.1111/j.1420-9101.2009.01699.x

Tattersall, G. J., Andrade, D. V., & Abe, A. S. (2009). Heat exchange from the toucan bill reveals a controllable vascular thermal radiator. *Science*, *325*(5939), 468–470. doi: 10.1126/science.1175553

Tattersall, G. J., Arnaout, B., & Symonds, M. R. E. (2016). The evolution of the avian bill as a thermoregulatory organ. *Biological Reviews*, *92*(3), 1630–1656. doi: 10.1111/brv.12299

Tattersall, G. J., Chaves, J. A., & Danner, R. M. (2017). Thermoregulatory windows in Darwin's finches. *Functional Ecology*, *32*(2), 358–368. doi: 10.1111/1365-2435.12990

Tieleman, B. I., Williams, J. B., Michaeli, G., & Pinshow, B. (1999). The role of the nasal passages in the water economy of crested larks and desert larks. *Physiological and Biochemical Zoology*, *72*(2), 219–226. doi: 10.1086/316658

Todd, W. E. C. (1924). A new song sparrow from Virginia. *The Auk*, *41*, 147–148.

Toews, D. P. L., Brelsford, A., Grossen, C., Milá, B., & Irwin, D. E. (2016). Genomic variation across the yellow-rumped warbler species complex. *The Auk*, *133*(4), 698–717. doi: 10.1642/AUK-16-61.1

Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, *26*(17), 2313–2318. doi: 10.1016/j.cub.2016.06.034

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285. doi: 10.1371/journal.pbio.0030285

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

Wilson, A. G., Arcese, P., Chan, Y. L., & Patten, M. A. (2011). Micro-spatial genetic structure in song sparrows (Melospiza melodia). *Conservation Genetics*, *12*(1), 213–222. doi: 10.1007/s10592-010-0134-4

Wilson, A. G., Chan, Y., Taylor, S. S., & Arcese, P. (2015). Genetic divergence of an avian endemic on the Californian Channel Islands. *PLOS ONE*, *10*(8), e0134471. doi: 10.1371/journal.pone.0134471

Zink, R. M., & Dittmann, D. L. (1993). Gene flow, refugia, and evolution of geographic variation in the song sparrow (*Melospiza melodia*). *Evolution*, *47*(3), 717–729. doi: 10.1111/j.1558-5646.1993.tb01228.x

Zou, F., Lim, H. C., Marks, B. D., Moyle, R. G., & Sheldon, F. H. (2007). Molecular phylogenetic analysis of the grey-cheeked fulvetta (*Alcippe morrisonia*) of China and Indochina: A case of remarkable genetic divergence in a "species." *Molecular Phylogenetics and Evolution*, *44*(1), 165–174. doi: 10.1016/j.ympev.2006.12.004

BIOGRAPHY

Jonathan D. Clark was born and raised in the Blue Ridge Mountains of southwestern Virginia, where he was privileged to be surrounded by a plethora of wildlife and scenic wilderness. He graduated from Glenvar High School in Roanoke County, Virginia in 2013, after which he began his undergraduate studies at George Mason University in Fairfax, Virginia. He received a Bachelor of Science from George Mason in 2016 and began the pursuit of a Master of Science at George Mason in 2017. Following the completion of his studies at George Mason, Jonathan will begin a Doctor of Philosophy program at the University of New Hampshire in Durham, New Hampshire, where he will continue to research the genomic ecology of coastal sparrow species.