Drivers of birdsong evolution on micro- and macroevolutionary scales

By

Katherine Tuuri Snyder

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in

Biological Sciences

May 12th, 2023

Nashville, Tennessee

Approved:

Kenneth Catania, Ph.D.

John Capra, Ph.D.

Maria Da Silva Pinto Jorge, Ph.D.

Emily Hodges, Ph.D.

Nicole Creanza, Ph.D.

Acknowledgments

I have so many people who have helped me along the way to my PhD, I cannot possibly name everyone. To all of my friends and family, thank you for your support.

I would like to thank my committee:
Ken Catania - For numerous chats about animal behavior, the opportunity to lecture on songbird neuroethology several times in what was one of my favorite courses, and for serving as Chair of my committee
Malu Jorge - For being a major factor in developing my skills in R, and for the privilege of getting to spend time and build connections in the Earth and Environmental Sciences department, learn about peccaries, and stay connected to my ecology roots
Tony Capra and Emily Hodges - For your input, thoughtful questions, and guidance throughout my candidacy

My PhD was a wonderful experience in large part thanks to members of the Creanza Lab, past and present: Fellow grad students Dr. Cristina Robinson, Dr. Abigail Searfoss, Kerri-Ann Anderson, Parker Rundstrom, Yakov Pichkar, Darra Boyer, and Ximena Leon. Postdoc Dr. Emily Hudson. My mentees, Maria Sellers and Aleyna Loughran-Pierce. Collaborators Alex Surowiec and Kara Boyer, and all of the other amazing Creanza Lab undergrads. Bird care technician Lena Friedman.

Above all, thank you to my advisor, Dr. Nicole Creanza. It was an honor and a privilege to be your first graduate student, and I can't imagine a better or more supportive mentor. I am so grateful for all of the opportunities I've had to build skills, gain experience, and become a better scientist above and beyond anything I could've hoped for in graduate school because of your dedication, enthusiasm, and willingness to fully engage with my big ideas and help me turn them into reality. Even if it means we end up with an absurd number of projects at any given time, it has all been such a joy and so rewarding.

I also want to thank:

My research advisors from my undergraduate alma mater: Joan Strassmann, Tom E.X. Miller, and Evan Siemann

Other Vanderbilt faculty who provided important mentorship and guidance: Patrick Abbot, Nicole Joseph, Christine Konradi, Doug McMahon, Kathy Friedman

Members of the Inclusivity in the Biosciences Association - your solidarity and collective desire to strive for cultural change wherever it was needed was a highlight of my graduate experience

Biological Sciences grad students: Stephanie Birnbaum, Michelle Moon, Juan Ortiz, Dom Vita, Danielle Golovin, Stephen Ferguson, Katrina Ngo, Jacob Steenwyk, Bryan Gitschlag, Luísa Jabbur, Annie Hatmaker, Reese Martin, and others. My original BioSci direct-admit cohort: Cait Kirby, Dylan Shropshire, Betty Xie, and Zi Ye

Earth and Environmental Science department grads and postdocs: Thank you for keeping me as an honorary member of EES, the camping trips, and the beeroclock shenanigans. Kristy Ferraro, Eva Biedron, George Duffy, Michaela Peterson, Genna Chiaro, Brad Pitcher, Chris Tasich, Jennifer Bradham, and so many more.

Other grad friends: Abigail Trozenski, Lauren Walker, Andrew Legan, Nate Faber, Leah Roberts, Alex Nackenoff; Pandemic Legacy champions: Sierra Palumbos, Mike Tackenberg, David Sprinzen

My roommates throughout my time in Nashville: Walker Pickens, Joshua Dent, Ally Brown, and Emily Ronna. Thank you for the music, the support, the friendship, and the care for Moki when I would work long evenings in lab.

Queer Birds - Kensuke Nakamura, Amery Wyn, and Kenzie Bottoms

Dear Rice friends - Tatiana Fofanova, Rachel Estes, Annie Blay, Daniel Zales, David Tam

Dear Nashville friends - Jennifer Rand, Sierra Gelbard, Lyle Reed, May Wild, Hope von Gunten, Brette Garner, Kate Mittendorf, Natalie Paul, Cáti Muñiz, Taylor Carroll, Kelly McKernan

The friends who collectively trained me in effective expository and persuasive writing for almost 15 years: Kim, Deanna, Sarah, Amy, Kisha, Lilly, Sara, Drew, Liz, Sonic, Kirsten, Nick, Tracy, Bridget, Mel, Wes, et al.

My animals for the emotional support and antics: my dog Moki, horse Chancee (and Avalon Farm for taking such good care of her), finches Hermione, Pippin, Tic-Tac, and Beaker

My family for their endless support - Mom, Dad, Grandma, Natalie, Emily and Andrew, Anders and Ellie, Ward, Nancy and Dean, Tom and Donna, Barb and Gar, Bonnie, Alison, Adam, and everyone else who has helped me become who I am today.

Funding sources: NSF, Vanderbilt Evolutionary Studies Institute, Vanderbilt Graduate School, Society for the Study of Evolution

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION:

BIRDSONG IN AN EVOLUTIONARY CONTEXT

The life of a songbird in nature is fraught with adversity. In its first few weeks of life, a chick is entirely dependent upon adults for protection and sustenance. Parents, meanwhile, must be utterly dedicated to the task of foraging and provisioning their offspring for several weeks until their offspring fledge and become independent (Low et al. 2008), buffering them from any ill effects of environmental scarcity or hostile conditions. Even when young birds become nutritionally independent, they must continue to develop their ability to forage in order to consistently sustain high energetic demands while avoiding predation themselves.

Despite the crucial need for nutrition, energy efficiency, and concealment from predators, juvenile songbirds spend significant time and energy in their first year on an endeavor that does nothing to promote their own survival: learning to sing. This expenditure of time and energy on this ornamental trait is instead an investment in their future reproductive success (Nowicki and Searcy 2004). In many songbird species, song is primarily produced by males, although female song has recently been found to be more widespread than previously assumed (Odom et al. 2014; Odom and Benedict 2018; Riebel et al. 2019). Male song has two primary functions, both important in sexual selection: territory defense, or male-male aggression, and mate attraction (Catchpole and Slater 2003). When male birds sing, they advertise their presence to nearby females, who in turn can identify them as conspecifics (Emlen 1972) as well as assess male song quality to inform their mate choice decisions (Byers and Kroodsma 2009; Nottebohm 1972; Reid et al. 2004; Robinson and Creanza 2019). Song can also reinforce pair bonding, especially in species that engage in duetting behaviors, where females and males sing complementary parts of a song (Hall 2004).

Because song plays a role in mate choice and pair maintenance, disruptions in the song-learning process can be detrimental to a learner's reproductive fitness (Kroodsma and Byers 1991; S. Nowicki, Searcy, and Peters 2002; Snyder and Creanza 2019). Despite the risk that poor song learning poses to an individual's reproductive success, vocal learning has persisted for ~40 million years in the Oscine suborder of the passerine birds (Order: Passeriformes). This contrasts with the sister suborder of the oscines, the suboscines, which generally sing unlearned songs ((Kroodsma and Konishi 1991; Touchton, Seddon, and Tobias 2014), but see (Kroodsma et al. 2013)). For effective communication to take place, a sender must transmit a signal that is correctly recognized by a receiver; over evolutionary time, deviations in these signals are generally selected against (i.e. stabilizing selection) since these deviations are likely to impede recognition (Ryan and Rand 1993). The core function of song in species recognition might therefore seem incompatible with learning, a comparatively risky and unstable form of inheritance relative to genes. Despite this risk, vocal learning has been resistant to loss over evolutionary time; within the Oscine suborder, the ability to learn vocalizations has apparently never been lost in any of the 4500+ extant species. The result is an extraordinary diversity of species-specific songs that evolved and diverged in the context of complex ecological, physiological, and social factors.

The drivers of birdsong evolution on micro- and macroevolutionary scales are thus likely to be multifaceted, necessitating the integration of diverse concepts, data types, and approaches to examine both direct observations of individual species and the extraordinary diversity of vocalizations across the songbird phylogeny. In this thesis, I will present studies that encompass both within- and between-species song variation in the context of different selection pressures and life history strategies. To approach the complex questions of how and why songs evolve, I draw from several different areas of study that are integral to understanding songbird evolution, including the composition and complexity of birdsong, the song-learning process, neurodevelopmental physiology and developmental stress, environmental effects on phenotype development, and social mating systems. To orient the reader, I give an introduction to these conceptual frameworks in the subsections that follow.

Structure and function of birdsong

Songs are composed of syllables, which are typically defined as periods of continuous production of sound, separated by a brief period of silence. Syllables can be composed of multiple notes, which are defined as continuous sounds on both a time and frequency scale. The syrinx, the muscular song production organ in the songbird respiratory tract, is split into two branches, each side of which can be controlled independently along with the corresponding lung (Schmidt 2008). Due to this unique structure, a bird can produce two distinct sounds simultaneously, thus enabling complex multi-note syllables (Suthers 1997). A song 'bout' can be defined as a period of singing that is separated from other such bouts by a period of silence that is generally much longer than inter-syllable silences. When these bouts are composed of multiple repetitions of the same syllable sequence, these repeated patterns of syllables are termed phrases or motifs (Brenowitz, Margoliash, and Nordeen 1997). (**Figure 1.1**) Songs can vary intraspecifically and interspecifically on each of these scales. The syllables present in song, the internal composition of song, and the order of the syllables and the pattern of motifs are all typically learned (Tchernichovski et al. 2004). Any of these components can contribute to the complexity or elaboration of a song, but, generally, song complexity is measured as the number of unique syllables or songs that a bird uses, also known as its syllable or song repertoire (Slater 2000). Alternatively, complexity can be considered the internal diversity of a bird's songs by using the mean number of unique syllables per song as a metric (Read and Weary 1992). Individual syllables can also vary in complexity based on the number of notes they contain. Historically, tabulating syllable repertoire and song repertoire has been done subjectively, with a human visual assessment of syllable similarity and uniqueness, since there is a certain level of variation within syllable types that are essentially the same (e.g. (Hughes et al. 1998)).

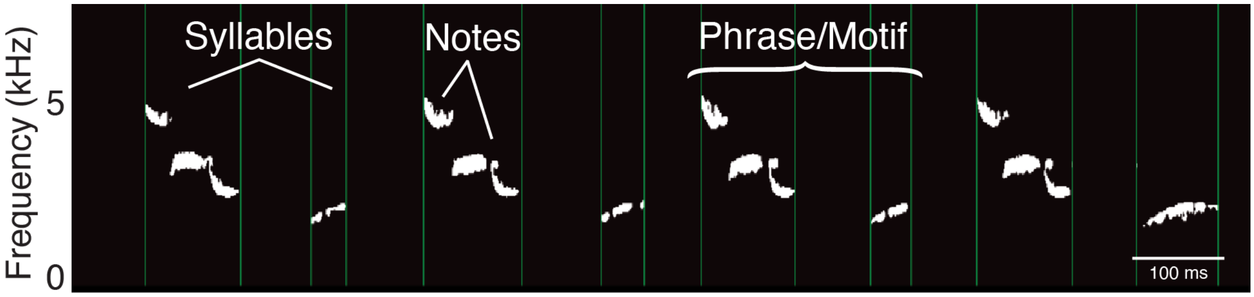


Figure 1.1. Spectrogram annotated with elements of a song. Syllables are separated from one another by periods of silence, but within a syllable there can be multiple notes, each a distinct pulse of sound at different frequencies. Repeated sets of syllables are called phrases or motifs.

The song features that are associated with higher attractiveness may differ between species. For example, in swamp sparrows, studies suggest that females prefer males with a higher rate of syllable production (Ballentine, Hyman, and Nowicki 2004). In zebra finches, females prefer songs containing syllables with clearly defined harmonics, as opposed to noisier syllables (Tomaszycki and Adkins-Regan 2005). In several other species, females tend to choose males with more complex or elaborate songs, potentially because these songs are more difficult to learn (Byers and Kroodsma 2009; Robinson and Creanza 2019; Soma and Garamszegi 2011)

In song sparrows, females have been shown to prefer males that learned their songs more accurately (Lachlan et al. 2014; Stephen Nowicki, Searcy, and Peters 2002). Learning accuracy is typically assessed by directly measuring song similarity between a tutor's and pupil's songs and includes the percentage of syllables that are copied as well as how closely each copied syllable matches the original; computational tools for quantifying pupil-tutor song similarity are broadly used in songbird research (Tchernichovski et al. 2000). Female preference is measured by several different methods that often differ between species; these measures are also challenging to compare between lab and field studies, complicating the interpretation of female preference tests (Byers and Kroodsma 2009; Robinson and Creanza 2019). These different methods include female choice tests in laboratory settings, with or without estrogen administration, which can be conducted by tallying copulation-solicitation displays to different song playbacks and by allowing females to choose which of two song playbacks to trigger or approach (Baker et al. 1986; Cardoso and Mota 2004; Catchpole, Dittami, and Leisler 1984), as well as female preference tests in the field, which are generally observational studies in which a proxy of female choice, such as pairing date or first egg date, is associated with properties of the male's song (Krebs, Ashcroft, and Webber 1978; Mountjoy and Lemon 1996; Searcy 1984)

The song-learning process

The understanding of birdsong as a learned behavior in Western science is relatively recent, and its discovery was rooted in a phenomenon that emerges from learning on a geographically structured population scale: dialects (Marler 1952). In this case, chaffinch subpopulations were observed to sing slightly different songs in somewhat spatially separated areas. This is reminiscent of language dialects in humans, and the observation inspired early birdsong researchers to test whether songs were also learned, and thus whether cultural transmission explained the song dialect patterns that were observed. Because song is learned, it may be influenced by both genetic and cultural evolutionary pressures. It might therefore exhibit more complex evolutionary dynamics than genetically inherited traits since its features can be transmitted obliquely, from any adult, instead of only vertically, from parent to offspring (Williams 2021).

While plasticity is inherent to learned behaviors, there do exist biological constraints to song development. First, there are morphological constraints: the types, frequencies, and rates of notes produced are dependent on the structural features of the beak, the respiratory system, and the syrinx (Podos and Sung 2020). For example, in Darwin's finches, beak size and shape is under selection related to food acquisition, but both the size and shape of the beak influence the frequency bandwidth and trill rate that a bird is capable of producing (Podos 2001). In addition to these morphological constraints, many species have an innate auditory song template that guides juveniles' naïve preference for and attentiveness to conspecific song (Marler et al. 1988; Marler and Peters 1977; Nelson and Marler 1993; Thorpe n.d.). In the cacophony of the natural world, the features of a song that cue a juvenile to learn a conspecific song vary between species. For example, some species, such as song sparrows, are cued by a certain syntactical structure, while others, such as the closely related swamp sparrow, are attuned to songs that contain syllables with certain phonetic characteristics (Marler and Peters 1977, 1988). There are also species that respond to a specific sound at the beginning of their species' song, such as the distinct whistles at the start of white-crowned sparrow and golden-crowned sparrow song (Hudson, Creanza, and Shizuka 2020; Soha and Marler 2000). Some species, such as those in the family Mimidae, which includes mockingbirds and other mimics, seem to have a comparatively permissive innate song template, as evidenced by their tendency to learn a wide range of heterospecific songs, calls of other animals, and anthropogenic sounds as well as to improvise and invent syllables (Goller and Shizuka 2018; Hatch 1967; Kroodsma et al. 1997; Love, Hoepfner, and Goller 2019). In species with more prescriptive innate templates, the template seems to enable socially isolated juveniles to develop crude songs that still exhibit some rudimentary elements of typical conspecific song (Baptista and Petrinovich 1984; Kroodsma 1977; Marler and Sherman 1985) and deafened juveniles to make sounds that, while abnormal, still retain the most basic features of song, such as containing discrete syllables (Love et al. 2019; Marler and Sherman 1983). Isolated birds inherently improvise or invent the elements in their songs, but improvisation of syllables, song segments, and song syntaxes has also been observed in birds of several species, including the sedge wren and the American robin, who were exposed to conspecific song in laboratory settings(Johnson 2006; Kroodsma et al. 1999). It is not well understood why a bird developing in a typical social environment might improvise elements of its song, but one hypothesis is that birds improvise or invent additional song elements when their tutor's song has low complexity, thereby increasing the complexity of its own song (Tchernichovski, Eisenberg-Edidin, and Jarvis 2021).

Song development takes place over several stages that occur during a "song-learning window," which may also be referred to as a "critical period" or a "sensitive period" (Marler and Peters 2010). [N.B. There is variation in the terminology in the literature; while the latter terms exist in other fields such as human language acquisition, where they have distinct meanings, they are often used interchangeably in birdsong literature.] This process and the timing and duration of these stages differs between species and varies based on the duration of the song-learning window. In many species, song learning begins with a sensory period, during which juveniles can begin to memorize songs they are exposed to without actively practicing vocalizations. Upon hearing and identifying a conspecific song, a juvenile or young adult forms a memory of the song, termed the “acquired auditory template” (Adret 2004). The sensory period is followed by the sensorimotor phase, defined as the period in which a juvenile practices singing, initially producing non-stereotyped “subsong” that is modified through a process of practicing a vocalization, assessing whether it is closer to matching their memorized song template, and self-editing, to converge on a song structure. Vocalizations produced in the later part of the sensorimotor phase are considered “plastic” song, where the overall structure of a song is stable, but changes can still occur while the bird continues to practice and converge on a stereotyped song. Song learning concludes with song crystallization, after which point a bird cannot alter its song. The timing of these phases varies by species; in some species, such as the zebra finch, the sensory and sensorimotor periods overlap, while in others they are separated by a significant period of time. This latter pattern typically occurs in seasonally breeding birds, such as the white-crowned sparrow, which acquires its song template in its hatch year, but does not begin practicing its song until the following breeding season (Brainard and Doupe 2002). Additionally, the age at which crystallization occurs varies widely between species (Kipper and Kiefer 2010; Rundstrom and Creanza 2021). Zebra finches undergo song crystallization at approximately 90 days post-hatch, and are unable to modify their songs after this point (Tchernichovski et al. 2001). Other species maintain plastic song until one or two years post-hatch. These species that at some point cease changing their song and maintain a stable song for the rest of their lifetime have been historically referred to as “closed-ended learners”. A few species do not appear to ever fully crystallize their songs. These are “open-ended learners.” (Brenowitz 2004)

Song neurobiology

The neural basis of song learning and production relies on the interdependent interactions between several nuclei of the songbird brain, collectively termed the “song system.” (Nottebohm 2005) The song system comprises two major, intertwined pathways: the song motor pathway (SMP) controls the production of sounds and songs, and the anterior forebrain pathway (AFP) mediates song learning. The robust nucleus of the arcopallium (RA) and Passeriformes-specific brain region HVC (used as a proper name) are integral in both pathways. The SMP originates from projection neurons in the HVC that propagate the signal directly into the RA, which then activates the motor neurons necessary for song production. The AFP is responsible for auditory processing during song production. This pathway sends real-time feedback to the RA, thus modulating song production. (**Figure 1.2**) Song processing for communication and learning occurs via the ascending auditory pathway, simplified in **Figure 1.2** with blue lines. Auditory signals from the cochlear nucleus (CN) are directed via two pathways, both ultimately ending in the HVC. While listening to song, the recurrent motor pathway engages several nuclei that are integral to song production (Mooney 2009). (**Figure 1.2**) The organization and interconnection of these pathways broadly resembles that of the human language learning and speech production pathways (Reiner et al. 2004). This parallels a key behavioral similarity between songbird and human vocal learning; in both systems, practicing the production of sounds and actively comparing those produced sounds to the memorized auditory template is essential for effective learning.

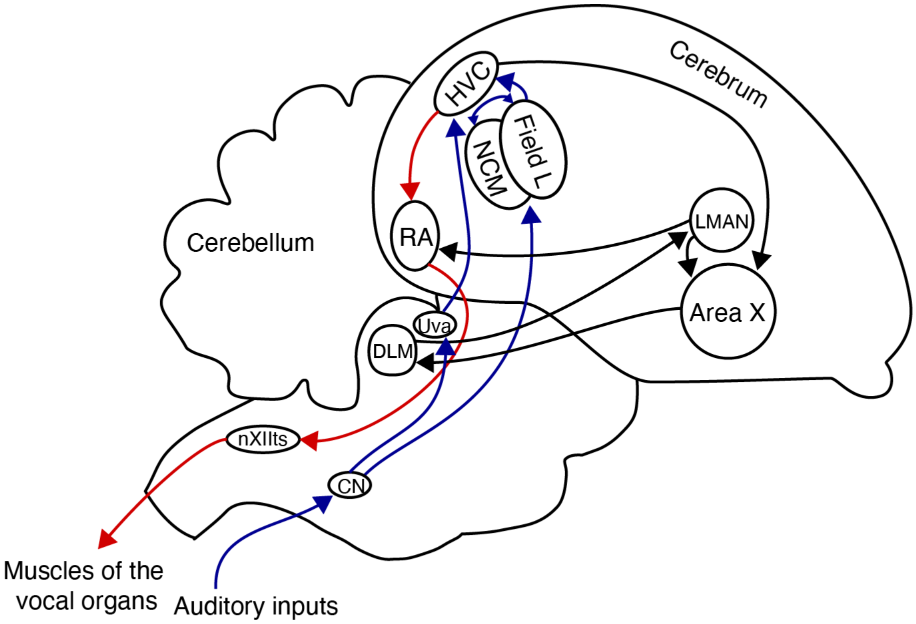


Figure 1.2: The song system in the songbird brain. Motor and sensory neural activity are integrated during song learning and production. The song motor pathway, which controls song production, is shown in red; the anterior forebrain pathway, which mediates song learning, is shown in black; and a simplified auditory pathway, which connects sounds heard through the ear to the song pathways via HVC, is shown in blue.

Paralleling the diversity of bird song, songbird neural development and anatomy can vary widely between species, individuals, and even seasons to reflect life history, individual variation, or the importance of song production at a given time. The volume of HVC covaries with song complexity interspecifically; species with simple songs have smaller HVC volume relative to total brain volume than species with more complex songs (Devoogd et al. 1993). HVC also varies intraspecifically; non-song-producing females typically have smaller song nuclei than males of the same species, and song lengths and complexities covary with HVC and RA volume in males within species. (Gahr, 2004; Garamszegi & Eens, 2004; Wade & Arnold, 2004) The volume of and neuron density within key song nuclei also vary within an individual on a seasonal basis; testosterone induces neurogenesis in the HVC leading up to the breeding season in seasonal breeders, such as canaries, corresponding to changes in various song metrics (Brenowitz and Larson 2015; Nottebohm and Nottebohm 1978). This occurs presumably in all songbird species that breed seasonally (Brenowitz 2004; Brenowitz and Larson 2015). In the zebra finch, an opportunistic breeder and year-round singer, neurogenesis slows after song learning has ended, but continues into adulthood, ultimately doubling the number of neurons in the HVC without increasing the volume (Kirn 2010; Walton, Pariser, and Nottebohm 2012; Wilbrecht and Kirn 2004).

Sexual Selection

Sexual selection is the part of natural selection that concerns only reproductive fitness, or the ability to propagate genes into future generations. There are several mechanisms by which this occurs, but the two main types are intrasexual selection, or competition between members of the same sex, and intersexual selection, or mate choice. In songbirds, males are typically the sex that competes and is chosen by females. Song, a trait used in both male-male competition and mate choice, functions in some ways as an ornament, a biologically frivolous trait that exists primarily to mediate conspecific social interactions by signaling individual quality to potential competitors and mates. Many ornamental traits are thought to serve as honest signals of fitness expressly because they are difficult to develop and maintain, so can only be exhibited by individuals fit enough to thrive despite them (Zahavi 1975). Individuals are motivated to choose a fit mate for two main reasons: first, to gain direct benefits, which is resource or action that gives a fitness benefit to the choosing individual and the pair’s potential brood, such as a high quality nest or territory or active and reliable nutrient provisioning during incubation and brood rearing; second, gain indirect benefits by choosing a mate that will pass on traits that increase the fitness of offspring. The two main hypotheses for how indirect benefits are secured by choosing a mate based on the elaborateness of an ornament are “sexy sons” and “good genes.” The “sexy sons” hypothesis, named such because males are typically the sex that develops ornaments, supposes that females that prefer males with an elaborate ornament ensure that male offspring inherit genes that encode that augmented ornament, thereby increasing their future reproductive fitness, while female offspring inherit a preference for the elaborate trait. This preference-elaboration cycle can result in a positive feedback loop termed “runaway sexual selection.” In the most extreme cases, it is predicted that the ornament would become so detrimental to overall fitness that the species is driven to extinction (Kokko and Brooks 2003). The “good genes” hypothesis posits that thriving in spite of having a cumbersome or expensive ornament indicates that a potential mate must have other traits that increase fitness that would be beneficial for offspring to inherit.

Developmental stress hypothesis

Females in numerous species have been shown to assess male songs and choose mates based on the quality of those songs (Catchpole 1987; Hasselquist, Bensch, and von Schantz 1996; Lachlan and Slater 1999; Reid et al. 2005). While there is no evidence that high-quality song provides any direct fitness benefit outside of territory defense, song quality has indeed been shown in various species to be an honest and reliable signal of male fitness as measured by body condition, immune system robustness, life span, and percent of offspring and further descendants surviving to be adults in the population (Pfaff et al. 2007; Reid et al. 2005; Spencer and

MacDougall-Shackleton 2011; Woodgate et al. 2012). This presents a puzzle: whereas honest signals typically evolve due to costs associated with making the signal (i.e. the risk of announcing presence to a predator, metabolic cost of producing signals, etc.), small changes in song structure and complexity amount to a negligible difference in direct costs of song production in males, so cost does not explain the evolution of female preference towards the highest quality versions of their species' stereotyped song.

The proposed solution to this puzzle of what makes song a costly signal likely hinges on the song-learning process itself (Searcy and Nowicki 2008). A developing bird must invest a lot of energetic resources in its neural development to robustly lay the groundwork for high-quality song learning. Indeed, juvenile exposure to any of nutritional, hormonal, thermal, social, or parasitic stress has been shown to negatively correlate with song quality in adulthood, likely because it cannot fully invest in the neural underpinnings of song learning (Anne Kristel Yu 2015; Banerjee et al. 2012; MacDougall-Shackleton and Spencer 2012; Nowicki et al. 2001). Consistent with this hypothesis, developmental stress is negatively correlated with the volume of HVC and Area X and neuronal recruitment to HVC (Gil et al. 2006). Further, several studies have shown that females tend to prefer the songs of males with a less stressful developmental history (Searcy et al. 2010; Spencer et al. 2005). Thus, it has been suggested that the honesty of song is maintained by the costs of learning song, especially elaborate song, and that poorer quality songs accurately signal a more difficult upbringing.

The evolutionary benefit of females accurately detecting a male's history of developmental stress remains understudied, but there are two main hypotheses: 1) a stressful upbringing could be indicative of deficiencies in foraging or other abilities, which could translate into a reduction in provisioning or other direct benefits from a potential mate, or a reduction in indirect benefits if those deficiencies have genetic underpinnings and are inherited by offspring; 2) developmental stress directly impacts health and fitness, so a developmental-stress-affected potential mate may be more susceptible to infection or early mortality, thus compromising their reliability as a breeding partner (Spencer and MacDougall-Shackleton 2011). There is a complex relationship between stress during development and later phenotypes, and birds who experience acute stressors at certain developmental stages can show either benefits or costs later in life, depending on the stressor (Wada and Coutts 2021). For example, acute temperature stress during development can lead to increased thermotolerance, enhancing fitness if the adult is then exposed to extreme temperatures (Hoffman, Finger, and Wada 2018; Nord and Giroud 2020). However, extended thermal stress during development increases mortality and decreases hatching success (Noiva, Menezes, and Peleteiro 2014; Patael et al. 2019; Wada et al. 2015). Similarly, while the stress response can enhance immune response to a pathogen in the short term, prolonged stress also significantly impairs immunity and leaves an organism with increased susceptibility to pathogens (Dhabhar 2009). Chronic stress is associated with shortened telomeres, which can lead to cellular dysfunction (Epel et al. 2004). While brains show remarkable plasticity and ability to tolerate injury, prolonged elevation of glucocorticoids, such as occurs during chronic stress, causes reduced neurogenesis and survival of new neurons in the short term in several model mammalian species (Joëls et al. 2007; Wong and Herbert 2006). Stress during development has augmented effects that can last well beyond the duration of the stressor, with impaired neuronal recruitment, particularly in the song system of the songbird brain (Gil et al. 2006). Finally, both developmental and chronic stress can act to downregulate hormones involved in several stages of reproduction, which can directly contribute to reduced fertility (Ahmed, Musa, and Sifaldin 2016; Rivier and Rivest 1991; Ubuka, Bentley, and Tsutsui 2013). In sum, there are complex and time-dependent effects of stress, but chronic exposure to stressors during development tends to have negative effects on physiology and fitness as well as on song, and a female's ability to detect stress-related song differences could help her choose a mate with a high potential reproductive output.

Stress and the environment

Understanding how environmental stress during development can affect a bird's song, and in turn its future reproductive success, is crucial in light of the shifting landscape of environmental conditions due to climate change. Environments shape organismal development in complex ways, and the amount of intraspecific variation that exists due to genetic variation as opposed to gene-by-environment interactions is difficult to determine. Environment effects are most significant during an individual's development, so in species where development happens over a relatively short time, even transient environmental fluctuations that happen to coincide with development can have large effects on an organism's phenotype that can persist into adulthood. Many of these fluctuations may be adaptive, such as epigenetic temperature acclimatization (Hu and Barrett 2017). Phenotypic plasticity can thus be evolutionarily beneficial, since it allows a species to hedge its bets, thriving in an environment that fluctuates over potentially multi-generational time-periods without losing the genetic architecture that encodes the potential alternate phenotype that would have a higher fitness in a different climate.

Songbird mating systems

The evolution of different mating strategies is the result of sexual selection in the context of complex social systems. The vast majority of songbird species (80-90%), typically form socially monogamous pair bonds with offspring often receiving biparental care. The rest exhibit social polygyny, polygynandry, promiscuity, or, rarely, polyandry (Cockburn, 2006). Polygyny, a mating system in which one male can be mated to several females simultaneously, has evolved numerous times in birds (Lack 1968; Rhoads 1890). Since polygynous male birds have the potential to raise clutches of offspring with multiple females, it is expected that there will be substantially higher variance in the fitness of individual males in polygynous species compared to monogamous species in which males have only one mate at a time (Emlen and Oring 1977). In other words, with a roughly equal sex ratio of males to females, a polygynous male could multiply his potential number of offspring by the number of mates he can attract. Thus, polygyny could cause differences in reproductive success between males on the order of several hundred percent. This is substantial, considering that fitness differences of only a few percent have the potential to shift the evolutionary dynamics in a population (e.g. (Lenski and Travisano 1994; Levy et al. 2015)). A polygynous mating strategy also leaves more males unpaired than monogamy; these males could have zero reproductive success unless they were successful in extra-pair fertilizations (Westneat, D. F., Sherman, P. W., & Morton, M. L. 1990). Since the stakes are higher to successfully attract one or more mates, the predicted high variance in reproductive success that is predicted for polygynous species could amplify the role of sexual selection compared to monogamous species. Females in polygynous species might also receive fewer fitness benefits from their mates—for example, less help with parental care—which could influence their mate choice decisions toward showy signals.

Darwin made this connection in *The Descent of Man and Selection in Relation to Sex*, stating that polygyny "leads to the same results as would follow from an actual inequality in the number of the sexes; for if each male secures two or more females, many males will not be able to pair; and the latter assuredly will be the weaker or less attractive individuals.…That some relation exists between [polygyny] and the development of secondary sexual characters, appears nearly certain; and this supports the view that a numerical preponderance of males would be eminently favourable to the action of sexual selection" ((Darwin 1873), pp. 265-266).

Extra-pair paternity (EPP) is common in Oscine species, and similar to polygyny, may also drive a variation in reproductive success in males in a population (Griffith, Owens, and Thuman 2002; Westneat and Stewart 2003). In both monogamous and polygynous species where extra-pair paternity is common, females will often form a socially monogamous bond with a single male partner, but seek extra-pair mates (Petrie and Kempenaers 1998). Seeking extra-pair copulations could act as a bet-hedging strategy that allows females to combine their genotypes with those of several males, which, in theory, could increase a female's indirect reproductive fitness by increasing the genetic diversity or fitness of her offspring (Cordero 1998; Yasui and Yoshimura 2018). Since sexual selection occurs in the choosing of these extra-pair mates, as has been shown in several Oscine species (Kempenaers et al. 1992; Wells, Safran, and Dale 2016), males with less attractive secondary characteristics may be doubly penalized; once in obtaining a social mate, and again in attaining copulations with already-mated females, but it remains unclear whether females choose social mates and extra-pair mates by the same criteria.

This thesis is broken into two sections based on the methods and types of data used in each, as well as the evolutionary scale they each cover. The first section consists of two projects investigating birdsong on the macroevolutionary scale. For these manuscripts, I collected species-level measurements of three song features that measure in some way song elaboration or complexity (syllable repertoire, syllables per song, and song repertoire) and two that are metrics associated with song performance (song duration and inter-song interval). The performance-associated metrics were also combined to make two additional metrics, song rate and continuity. I also collected species-level data on life history traits related to mating systems in order to assign classifications of socially monogamy or polygyny to 764 species and high or low rates of extra-pair paternity to 141 species (Chapter 2), as well as data on song learning window using adult song stability as a proxy metric for 67 species (Chapter 3). I obtained phylogenetic trees from BirdTree.org, which I used to perform phylogenetic comparative analyses using these databases. These computational methods combined with trees that were scaled by evolutionary time allowed me to probe the relationships between mating systems and song evolution, and learning and song evolution, and ask questions not only regarding how these traits are correlated in extant species, but also how evolutionary changes in one trait can drive evolution of the other.

In the second section, I investigate birdsong on a microevolutionary scale by focusing on populations of two common species, the song sparrow and dark-eyed junco, in a specific geographic region over a period of 10-15 years. During this period, from June through August 2016, western New York experienced an unprecedented drought that resulted in low vegetation growth for weeks to months. The area with the highest drought severity, which includes Ithaca, NY, received only 37% of its normal precipitation in June and July, 2016, after already experiencing a deficit of rainfall from March through May (64% of the norm). (NE Regional Climate Center). There was a sharp decrease in plant growth and water availability, which likely led to food scarcity and dehydration throughout the population. The area surrounding Ithaca, New York, with its proximity to Cornell University and the Cornell Ornithological Library, has a high density of bird song recordings obtained by both academics and community bird enthusiasts over the past few decades, with a sharp increase in recordings in the past five to ten years.

Since song is culturally transmitted, the fidelity with which it is learned by each generation is a key factor in its evolution. I posit that the 2016 drought served as a population-wide environmental stressor and increased the average intensity of stress would increase across the entire population, which could have theoretically caused an entire cohort to learn more poorly than normal. I hypothesize that this perturbation in song can be detected in subsequent years in the populations by analyzing the community-science recordings that are publicly available on the repositories Macaulay Library and xeno-canto. I test this hypothesis by analyzing average song features in each recording in both dark-eyed juncos and song sparrows (Chapter 4), as well as by tracking song sparrow population song composition by considering individual syllables as units of inheritance (Chapter 5).

References

Adret, Patrice. 2004. “In Search of the Song Template.” *Annals of the New York Academy of Sciences* 1016(1):303–24.

Ahmed, Abdelkareem A., Hassan H. Musa, and Amal Z. Sifaldin. 2016. “Prenatal Corticosterone Exposure Programs Growth, Behavior, Reproductive Function and Genes in the Chicken.” *Asian Pacific Journal of Reproduction* 5(4):271–78.

Anne Kristel Yu. 2015. “Early Life Cold Exposure Decreases Global Methylation Levels in Juveniles and Attenuates the Corticosterone Response after Restraint in Adult Zebra Finches (Taeniopygia Guttata).” Georgia State University.

Baker, Myron Charles, Tore K. Bjerke, Helene Lampe, and Yngve Espmark. 1986. “Sexual Response of Female Great Tits to Variation in Size of Males’ Song Repertoires.” *The American Naturalist* 128(4):491–98.

Ballentine, Barbara, Jeremy Hyman, and Stephen Nowicki. 2004. “Vocal Performance Influences Female Response to Male Bird Song: An Experimental Test.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 15(1):163–68.

Banerjee, Sunayana B., Adam S. Arterbery, Daniel J. Fergus, and Elizabeth Adkins-Regan. 2012. “Deprivation of Maternal Care Has Long-Lasting Consequences for the Hypothalamic-Pituitary-Adrenal Axis of Zebra Finches.” *Proceedings. Biological Sciences / The Royal Society* 279(1729):759–66.

Baptista, Luis F., and Lewis Petrinovich. 1984. “Social Interaction, Sensitive Phases and the Song Template Hypothesis in the White-Crowned Sparrow.” *Animal Behaviour* 32(1):172–81.

Brainard, Michael S., and Allison J. Doupe. 2002. “What Songbirds Teach Us about Learning.” *Nature* 417(6886):351–58.

Brenowitz, E. A., D. Margoliash, and K. W. Nordeen. 1997. “An Introduction to Birdsong and the Avian Song System.” *Journal of Neurobiology* 33(5):495–500.

Brenowitz, Eliot A. 2004. “Plasticity of the Adult Avian Song Control System.” *Annals of the New York Academy of Sciences* 1016:560–85.

Brenowitz, Eliot A., and Tracy A. Larson. 2015. “Neurogenesis in the Adult Avian Song-Control System.” *Cold Spring Harbor Perspectives in Biology* 7(6). doi: 10.1101/cshperspect.a019000.

Byers, Bruce E., and Donald E. Kroodsma. 2009. “Female Mate Choice and Songbird Song Repertoires.” *Animal Behaviour* 77(1):13–22.

Cardoso, G. C., and P. G. Mota. 2004. “Female Serin (Serinus Serinus) Responses to Song Do Not Favour the Predominant Song Syntax.” *Ethology Ecology & Evolution* 16(4):329–38.

Catchpole, C. K. 1987. “Bird Song, Sexual Selection and Female Choice.” *Trends in Ecology & Evolution* 2(4):94–97.

Catchpole, C. K., J. Dittami, and B. Leisler. 1984. “Differential Responses to Male Song Repertoires in Female Songbirds Implanted with Oestradiol.” *Nature* 312(5994):563–64.

Catchpole, Clive K., and Peter J. B. Slater. 2003. *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Cordero, Pedro J. 1998. “Extra-Pair Paternity in Birds: `good-Genes’ and Something Else.” *Trends in Ecology & Evolution* 13(7):280.

Darwin, Charles. 1873. *The Descent of Man, and Selection in Relation to Sex / by Charles Darwin*.

Devoogd, T. J., J. R. Krebs, S. D. Healy, and A. Purvis. 1993. “Relations between Song Repertoire Size and the Volume of Brain Nuclei Related to Song: Comparative Evolutionary Analyses amongst Oscine Birds.” *Proceedings. Biological Sciences / The Royal Society* 254(1340):75–82.

Dhabhar, Firdaus S. 2009. “Enhancing versus Suppressive Effects of Stress on Immune Function: Implications for Immunoprotection and Immunopathology.” *Neuroimmunomodulation* 16(5):300–317.

Emlen, S., and L. Oring. 1977. “Ecology, Sexual Selection, and the Evolution of Mating Systems.” *Science* 197(4300):215–23.

Emlen, Stephen T. 1972. “An Experimental Analysis of the Parameters of Bird Song Eliciting Species Recognition.” *Behaviour* 41(1-2):130–71.

Epel, Elissa S., Elizabeth H. Blackburn, Jue Lin, Firdaus S. Dhabhar, Nancy E. Adler, Jason D. Morrow, and Richard M. Cawthon. 2004. “Accelerated Telomere Shortening in Response to Life Stress.” *Proceedings of the National Academy of Sciences of the United States of America* 101(49):17312–15.

Gil, Diego, Marc Naguib, Katharina Riebel, Alison Rutstein, and Manfred Gahr. 2006. “Early Condition, Song Learning, and the Volume of Song Brain Nuclei in the Zebra Finch (Taeniopygia Guttata).” *Journal of Neurobiology* 66(14):1602–12.

Goller, Maria, and Daizaburo Shizuka. 2018. “Evolutionary Origins of Vocal Mimicry in Songbirds.” *Evolution Letters* 2(4):417–26.

Griffith, Simon C., Ian P. F. Owens, and Katherine A. Thuman. 2002. “Extra Pair Paternity in Birds: A Review of Interspecific Variation and Adaptive Function.” *Molecular Ecology* 11(11):2195–2212.

Hall, Michelle L. 2004. “A Review of Hypotheses for the Functions of Avian Duetting.” *Behavioral Ecology and Sociobiology* 55(5):415–30.

Hasselquist, Dennis, Staffan Bensch, and Torbjörn von Schantz. 1996. “Correlation between Male Song Repertoire, Extra-Pair Paternity and Offspring Survival in the Great Reed Warbler.” *Nature* 381(6579):229–32.

Hatch, Jeremy John. 1967. “DIVERSITY OF THE SONG OF MOCKINGBIRDS (MIMUS POLYGLOTTOS) REARED IN DIFFERENT AUDITORY ENVIRONMENTS.”

Hoffman, Alexander J., John W. Finger Jr, and Haruka Wada. 2018. “Early Stress Priming and the Effects on Fitness-Related Traits Following an Adult Stress Exposure.” *Journal of Experimental Zoology. Part A, Ecological and Integrative Physiology* 329(6-7):323–30.

Hudson, Emily J., Nicole Creanza, and Daizaburo Shizuka. 2020. “The Role of Nestling Acoustic Experience in Song Discrimination in a Sparrow.” *Frontiers in Ecology and Evolution* 8. doi: 10.3389/fevo.2020.00099.

Hughes, Melissa, Stephen Nowicki, William A. Searcy, and Susan Peters. 1998. “Song-Type Sharing in Song Sparrows: Implications for Repertoire Function and Song Learning.” *Behavioral Ecology and Sociobiology* 42(6):437–46.

Hu, J., and R. D. H. Barrett. 2017. “Epigenetics in Natural Animal Populations.” *Journal of Evolutionary Biology* 30(9):1612–32.

Joëls, Marian, Henk Karst, Harmen J. Krugers, and Paul J. Lucassen. 2007. “Chronic Stress: Implications for Neuronal Morphology, Function and Neurogenesis.” *Frontiers in Neuroendocrinology* 28(2):72–96.

Johnson, Steven L. 2006. “Do American Robins Acquire Songs By Both Imitating And Inventing?” *The Wilson Journal of Ornithology* 118(3):341–52.

Kempenaers, Bart, Geert R. Verheyen, Marleen Van den Broeck, Terry Burke, Christine Van Broeckhoven, and André Dhondt. 1992. “Extra-Pair Paternity Results from Female Preference for High-Quality Males in the Blue Tit.” *Nature* 357(6378):494–96.

Kipper, Silke, and Sarah Kiefer. 2010. “Chapter 3 - Age-Related Changes in Birds’ Singing Styles: On Fresh Tunes and Fading Voices?” Pp. 77–118 in *Advances in the Study of Behavior*. Vol. 41, edited by H. J. Brockmann, T. J. Roper, M. Naguib, K. E. Wynne-Edwards, J. C. Mitani, and L. W. Simmons. Academic Press.

Kirn, John R. 2010. “The Relationship of Neurogenesis and Growth of Brain Regions to Song Learning.” *Brain and Language* 115(1):29–44.

Kokko, Hanna, and Robert Brooks. 2003. “Sexy to Die for? Sexual Selection and the Risk of Extinction.” *Annales Zoologici Fennici* 40(2):207–19.

Krebs, John, Ruth Ashcroft, and Mike Webber. 1978. “Song Repertoires and Territory Defence in the Great Tit.” *Nature* 271(5645):539–42.

Kroodsma, D. E. 1977. “A Re-Evaluation of Song Development in the Song Sparrow.” *Animal*

Behaviour 25(2):390–99.

Kroodsma, D. E., P. W. Houlihan, P. A. Fallon, and J. A. Wells. 1997. “Song Development by Grey Catbirds.” *Animal Behaviour* 54(2):457–64.

Kroodsma, Donald E., and Bruce E. Byers. 1991. “The Function (s) of Bird Song.” *American Zoologist* 31(2):318–28.

Kroodsma, Donald E., and Masakazu Konishi. 1991. “A Suboscine Bird (eastern Phoebe, Sayornis Phoebe) Develops Normal Song without Auditory Feedback.” *Animal Behaviour* 42(3):477–87.

Kroodsma, Donald E., Wan-Chun Liu, Elijah Goodwin, and Paul A. Bedell. 1999. “The Ecology of Song Improvisation as Illustrated by North American Sedge Wrens.” *The Auk* 116(2):373–86.

Kroodsma, Donald, Debra Hamilton, Julio E. Sánchez, Bruce E. Byers, Hernán Fandiño-Mariño, David W. Stemple, Jill M. Trainer, and George V. N. Powell. 2013. “Behavioral Evidence for Song Learning In the Suboscine Bellbirds (Procnias Spp.; Cotingidae).” *The Wilson Journal of Ornithology* 125(1):1–14.

Lachlan, R. F., R. C. Anderson, S. Peters, W. A. Searcy, and S. Nowicki. 2014. “Typical Versions of Learned Swamp Sparrow Song Types Are More Effective Signals than Are Less Typical Versions.” *Proceedings. Biological Sciences / The Royal Society* 281(1785):20140252.

Lachlan, Robert F., and Peter J. B. Slater. 1999. “The Maintenance of Vocal Learning by Gene–culture Interaction: The Cultural Trap Hypothesis.” *Proceedings of the Royal Society of London B: Biological Sciences* 266(1420):701–6.

Lack, David Lambert. 1968. *Ecological Adaptations for Breeding in Birds*.

Lenski, R. E., and M. Travisano. 1994. “Dynamics of Adaptation and Diversification: A 10,000-Generation Experiment with Bacterial Populations.” *Proceedings of the National Academy of Sciences of the United States of America* 91(15):6808–14.

Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. 2015. “Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking.” *Nature* 519(7542):181–86.

Love, Jay, Amanda Hoepfner, and Franz Goller. 2019. “Song Feature Specific Analysis of Isolate Song Reveals Interspecific Variation in Learned Components.” *Developmental Neurobiology* 79(4):350–69.

Low, Matthew, Sönke Eggers, Debora Arlt, and Tomas Pärt. 2008. “Daily Patterns of Nest Visits Are Correlated with Ambient Temperature in the Northern Wheatear.” *Journal of Ornithology / DO-G* 149(4):515–19.

MacDougall-Shackleton, Scott A., and Karen A. Spencer. 2012. “Developmental Stress and

Birdsong: Current Evidence and Future Directions.” *Journal of Ornithology / DO-G* 153(1):105–17.

Marler, P. 1952. “Variation in the Song of the Chaffinch Fringilla Coelebs.” *The Ibis* 94(3):458–72.

Marler, Peter, and Susan Peters. 1988. “The Role of Song Phonology and Syntax in Vocal Learning Preferences in the Song Sparrow, Melospiza Melodia.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 77(2):125–49.

Marler, Peter, and Susan Peters. 2010. “A Sensitive Period for Song Acquisition in the Song Sparrow, Melospiza Melodia: A Case of Age-Limited Learning.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 76(2):89–100.

Marler, Peter, and Virginia Sherman. 1985. “Innate Differences in Singing Behaviour of Sparrows Reared in Isolation from Adult Conspecific Song.” *Animal Behaviour* 33(1):57–71.

Marler, P., and S. Peters. 1977. “Selective Vocal Learning in a Sparrow.” *Science* 198(4316):519–21.

Marler, P., S. Peters, G. F. Ball, A. M. Dufty Jr, and J. C. Wingfield. 1988. “The Role of Sex Steroids in the Acquisition and Production of Birdsong.” *Nature* 336(6201):770–72.

Marler, P., and V. Sherman. 1983. “Song Structure without Auditory Feedback: Emendations of the Auditory Template Hypothesis.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 3(3):517–31.

McEwen, B. S., and E. Stellar. 1993. “Stress and the Individual. Mechanisms Leading to Disease.” *Archives of Internal Medicine* 153(18):2093–2101.

Mooney, Richard. 2009. “Neural Mechanisms for Learned Birdsong.” *Learning & Memory* 16(11):655–69.

Mountjoy, D. J., and Robert E. Lemon. 1996. “Female Choice for Complex Song in the European Starling: A Field Experiment.” *Behavioral Ecology and Sociobiology* 38(1):65–71.

Nelson, Douglas A., and Peter Marler. 1993. “Innate Recognition of Song in White-Crowned Sparrows: A Role in Selective Vocal Learning?” *Animal Behaviour* 46(4):806–8.

Noiva, Rute M., António C. Menezes, and Maria C. Peleteiro. 2014. “Influence of Temperature and Humidity Manipulation on Chicken Embryonic Development.” *BMC Veterinary Research* 10(1).

Nord, Andreas, and Sylvain Giroud. 2020. “Lifelong Effects of Thermal Challenges During Development in Birds and Mammals.” *Frontiers in Physiology* 11.

Nottebohm, Fernando. 1972. “The Origins of Vocal Learning.” *The American Naturalist* 106(947):116–40.

Nottebohm, Fernando. 2005. “The Neural Basis of Birdsong.” *PLoS Biology* 3(5):e164.

Nottebohm, Fernando, and Marta E. Nottebohm. 1978. “Relationship between Song Repertoire and Age in the Canary, Serinus Canarius.” *Zeitschrift Für Tierpsychologie* 46(3):298–305.

Nowicki, S., W. A. Searcy, and S. Peters. 2002. “Brain Development, Song Learning and Mate Choice in Birds: A Review and Experimental Test of the ‘Nutritional Stress Hypothesis.’” *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology* 188(11-12):1003–14.

Nowicki, Stephen, and William A. Searcy. 2004. “Song Function and the Evolution of Female Preferences: Why Birds Sing, Why Brains Matter.” *Annals of the New York Academy of Sciences* 1016:704–23.

Nowicki, Stephen, William A. Searcy, Melissa Hughes, and Jeffrey Podos. 2001. “The Evolution of Bird Song: Male and Female Response to Song Innovation in Swamp Sparrows.” *Animal Behaviour* 62(6):1189–95.

Nowicki, Stephen, William A. Searcy, and Susan Peters. 2002. “Quality of Song Learning Affects Female Response to Male Bird Song.” *Proceedings. Biological Sciences / The Royal Society* 269(1503):1949–54.

Odom, Karan J., and Lauryn Benedict. 2018. “A Call to Document Female Bird Songs: Applications for Diverse Fields.” *The Auk* 135(2):314–25.

Odom, Karan J., Michelle L. Hall, Katharina Riebel, Kevin E. Omland, and Naomi E. Langmore. 2014. “Female Song Is Widespread and Ancestral in Songbirds.” *Nature Communications* 5:3379.

Patael, Tomer, Yogev Piestun, Arad Soffer, Sharon Mordechay, Shlomo Yahav, Sandra G. Velleman, and Orna Halevy. 2019. “Early Posthatch Thermal Stress Causes Long-Term Adverse Effects on Pectoralis Muscle Development in Broilers.” *Poultry Science* 98(8):3268–77.

Petrie, Marion, and Bart Kempenaers. 1998. “Extra-Pair Paternity in Birds: Explaining Variation between Species and Populations.” *Trends in Ecology & Evolution* 13(2):52–58.

Pfaff, Jeremy A., Liana Zanette, Scott A. MacDougall-Shackleton, and Elizabeth A. MacDougall-Shackleton. 2007. “Song Repertoire Size Varies with HVC Volume and Is Indicative of Male Quality in Song Sparrows (Melospiza Melodia).” *Proceedings. Biological Sciences / The Royal Society* 274(1621):2035–40.

Podos, J. 2001. “Correlated Evolution of Morphology and Vocal Signal Structure in Darwin’s Finches.” *Nature* 409(6817):185–88.

Podos, Jeffrey, and Ha-Cheol Sung. 2020. “Vocal Performance in Songbirds: From Mechanisms to Evolution.” Pp. 245–68 in *The Neuroethology of Birdsong*, edited by J. T. Sakata, S. C. Woolley, R. R. Fay, and A. N. Popper. Cham: Springer International Publishing.

Read, A. F., and D. M. Weary. 1992. “The Evolution of Bird Song: Comparative Analyses.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*.

Reid, Jane M., Peter Arcese, Alice L. E. V. Cassidy, Sara M. Hiebert, James N. M. Smith, Philip K. Stoddard, Amy B. Marr, and Lukas F. Keller. 2004. “Song Repertoire Size Predicts Initial Mating Success in Male Song Sparrows, Melospiza Melodia.” *Animal Behaviour* 68(5):1055–63.

Reid, Jane M., Peter Arcese, Alice L. E. V. Cassidy, Sara M. Hiebert, James N. M. Smith, Philip K. Stoddard, Amy B. Marr, and Lukas F. Keller. 2005. “Fitness Correlates of Song Repertoire Size in Free-Living Song Sparrows (Melospiza Melodia).” *The American Naturalist* 165(3):299–310.

Reiner, Anton, David J. Perkel, Laura L. Bruce, Ann B. Butler, András Csillag, Wayne Kuenzel, Loreta Medina, George Paxinos, Toru Shimizu, Georg Striedter, Martin Wild, Gregory F. Ball, Sarah Durand, Onur Güntürkün, Diane W. Lee, Claudio V. Mello, Alice Powers, Stephanie A. White, Gerald Hough, Lubica Kubikova, Tom V. Smulders, Kazuhiro Wada, Jennifer Dugas-Ford, Scott Husband, Keiko Yamamoto, Jing Yu, Connie Siang, Erich D. Jarvis, Onur Gütürkün, and Avian Brain Nomenclature Forum. 2004. “Revised Nomenclature for Avian Telencephalon and Some Related Brainstem Nuclei.” *The Journal of Comparative Neurology* 473(3):377–414.

Rhoads, Samuel N. 1890. “Probable Causes of Polygamy Among Birds.” *The American Naturalist* 24(287):1024–36.

Riebel, Katharina, Karan J. Odom, Naomi E. Langmore, and Michelle L. Hall. 2019. “New Insights from Female Bird Song: Towards an Integrated Approach to Studying Male and Female Communication Roles.” *Biology Letters* 15(4):20190059.

Rivier, C., and S. Rivest. 1991. “Effect of Stress on the Activity of the Hypothalamic-Pituitary-Gonadal Axis: Peripheral and Central Mechanisms.” *Biology of Reproduction* 45(4):523–32.

Robinson, Cristina M., and Nicole Creanza. 2019. “Species-Level Repertoire Size Predicts a Correlation between Individual Song Elaboration and Reproductive Success.” *Ecology and Evolution* 9(14):8362–77.

Rundstrom, Parker, and Nicole Creanza. 2021. “Song Learning and Plasticity in Songbirds.” *Current Opinion in Neurobiology* 67:228–39.

Ryan, Michael J., and A. Stanley Rand. 1993. “Species Recognition And Sexual Selection As A Unitary Problem In Animal Communication.” *Evolution; International Journal of Organic Evolution* 47(2):647–57.

Schmidt, Marc F. 2008. “Using Both Sides of Your Brain: The Case for Rapid Interhemispheric Switching.” *PLoS Biology* 6(10):e269.

Searcy, William A. 1984. “Song Repertoire Size and Female Preferences in Song Sparrows.” *Behavioral Ecology and Sociobiology* 14(4):281–86.

Searcy, William A., and Stephen Nowicki. 2008. “Bird Song and the Problem of Honest Communication: How Is the Honesty of Animal Signals Maintained When Exaggeration and Bluff Are so Tempting?” *American Scientist* 96(2):114–21.

Searcy, William A., Susan Peters, Silke Kipper, and Stephen Nowicki. 2010. “Female Response to Song Reflects Male Developmental History in Swamp Sparrows.” *Behavioral Ecology and Sociobiology* 64(8):1343–49.

Slater, Peter J. B. 2000. “Birdsong Repertoires: Their Origins and Use.” *The Origins of Music* 49–63.

Snyder, Kate T., and Nicole Creanza. 2019. “Food-Seeking Behavior Has Complex Evolutionary Pressures in Songbirds: Linking Parental Foraging to Offspring Sexual Selection.” *The Behavioral and Brain Sciences* 42:e52.

Soha, J. A., and P. Marler. 2000. “A Species-Specific Acoustic Cue for Selective Song Learning in the White-Crowned Sparrow.” *Animal Behaviour* 60(3):297–306.

Soma, Masayo, and László Zsolt Garamszegi. 2011. “Rethinking Birdsong Evolution: Meta-Analysis of the Relationship between Song Complexity and Reproductive Success.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 22(2):363–71.

Spencer, Karen A., and Scott A. MacDougall-Shackleton. 2011. “Indicators of Development as Sexually Selected Traits: The Developmental Stress Hypothesis in Context.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 22(1):1–9.

Spencer, K. A., J. H. Wimpenny, K. L. Buchanan, P. G. Lovell, A. R. Goldsmith, and C. K. Catchpole. 2005. “Developmental Stress Affects the Attractiveness of Male Song and Female Choice in the Zebra Finch (Taeniopygia Guttata).” *Behavioral Ecology and Sociobiology* 58(4):423–28.

Suthers, R. A. 1997. “Peripheral Control and Lateralization of Birdsong.” *Journal of Neurobiology* 33(5):632–52.

Tchernichovski, Ofer, Sophie Eisenberg-Edidin, and Erich D. Jarvis. 2021. “Balanced Imitation Sustains Song Culture in Zebra Finches.” *Nature Communications* 12(1):2562.

Tchernichovski, Ofer, Fernando Nottebohm, Ching Elizabeth Ho, Bijan Pesaran, and Partha Pratim Mitra. 2000. “A Procedure for an Automated Measurement of Song Similarity.” *Animal Behaviour* 59(6):1167–76.

Tchernichovski, O., T. J. Lints, S. Deregnaucourt, A. Cimenser, and P. P. Mitra. 2004. “Studying the Song Development Process: Rationale and Methods.” *Annals of the New York Academy of Sciences* 1016:348–63.

Tchernichovski, O., P. P. Mitra, T. Lints, and F. Nottebohm. 2001. “Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song.” *Science* 291(5513):2564–69.

Thorpe, W. H. n.d. “The Learning of Song Patterns by Birds, with Especial Reference to the Song of the Chaffinch Fringilla Coelebs.” *The Ibis*. doi: 10.1111/j.1474-919X.1958.tb07960.x.

Tomaszycki, Michelle L., and Elizabeth Adkins-Regan. 2005. “Experimental Alteration of Male Song Quality and Output Affects Female Mate Choice and Pair Bond Formation in Zebra Finches.” *Animal Behaviour* 70(4):785–94.

Touchton, Janeene M., Nathalie Seddon, and Joseph A. Tobias. 2014. “Captive Rearing Experiments Confirm Song Development without Learning in a Tracheophone Suboscine Bird.” *PloS One* 9(4):e95746.

Ubuka, Takayoshi, George E. Bentley, and Kazuyoshi Tsutsui. 2013. “Neuroendocrine Regulation of Gonadotropin Secretion in Seasonally Breeding Birds.” *Frontiers in Neuroscience* 7:38.

Wada, Haruka, and Victoria Coutts. 2021. “Detrimental or Beneficial? Untangling the Literature on Developmental Stress Studies in Birds.” *The Journal of Experimental Biology* 224(19). doi: 10.1242/jeb.227363.

Wada, Haruka, Buddhamas Kriengwatana, Natalie Allen, Kimberly L. Schmidt, Kiran K. Soma, and Scott A. MacDougall-Shackleton. 2015. “Transient and Permanent Effects of Suboptimal Incubation Temperatures on Growth, Metabolic Rate, Immune Function and Adrenocortical Responses in Zebra Finches.” *The Journal of Experimental Biology* 218(Pt 18):2847–55.

Walton, Clare, Eben Pariser, and Fernando Nottebohm. 2012. “The Zebra Finch Paradox: Song Is Little Changed, but Number of Neurons Doubles.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32(3):761–74.

Wells, Sarah J., Rebecca J. Safran, and James Dale. 2016. “Piecing Together Female Extra-Pair Mate Choice: Females Really Do Prefer More Ornamented Males.” *Molecular Ecology* 25(15):3521–24.

Westneat, David F., and Ian R. K. Stewart. 2003. “Extra-Pair Paternity in Birds: Causes, Correlates, and Conflict.” *Annual Review of Ecology, Evolution, and Systematics* 34(1):365–96.

Westneat, D. F., Sherman, P. W., & Morton, M. L. 1990. “The Ecology and Evolution of Extra-Pair Copulations in Birds.” *Current Ornithology* 7:331–69.

Wilbrecht, Linda, and John R. Kirn. 2004. “Neuron Addition and Loss in the Song System: Regulation and Function.” *Annals of the New York Academy of Sciences* 1016:659–83.

Williams, Heather. 2021. “Mechanisms of Cultural Evolution in the Songs of Wild Bird Populations.” *Frontiers in Psychology* 12:643343.

Wingfield, John C., Donna L. Maney, Creagh W. Breuner, Jerry D. Jacobs, Sharon Lynn, Marilyn Ramenofsky, and Ralph D. Richardson. 1998. “Ecological Bases of Hormone—behavior Interactions: The ‘emergency Life History Stage.’” *American Zoologist* 38(1):191–206.

Wong, E. Y. H., and J. Herbert. 2006. “Raised Circulating Corticosterone Inhibits Neuronal Differentiation of Progenitor Cells in the Adult Hippocampus.” *Neuroscience* 137(1):83–92.

Woodgate, Joseph L., Mylene M. Mariette, Andrew T. D. Bennett, Simon C. Griffith, and Katherine L. Buchanan. 2012. “Male Song Structure Predicts Reproductive Success in a Wild Zebra Finch Population.” *Animal Behaviour* 83(3):773–81.

Yasui, Yukio, and Jin Yoshimura. 2018. “Bet-Hedging against Male-Caused Reproductive Failures May Explain Ubiquitous Cuckoldry in Female Birds.” *Journal of Theoretical Biology* 437:214–21.

Zahavi, Amotz. 1975. “Mate selection—A Selection for a Handicap.” *Journal of Theoretical Biology* 53(1):205–14.

CHAPTER 2.

POLYGYNY IS LINKED TO ACCELERATED BIRDSONG EVOLUTION AND DESTABILIZED EXTREMES IN SONG COMPLEXITY AND PERFORMANCE METRICS

Kate T. Snyder[1] and Nicole Creanza[1*]

[1] Vanderbilt University Department of Biological Sciences, 465 21st Ave. South, Nashville, TN 37240
[*] Address for correspondence: nicole.creanza@vanderbilt.edu

Adapted from Snyder, K.T., Creanza, N. Polygyny is linked to accelerated birdsong evolution but not to larger song repertoires. *Nat Commun* 10, 884 (2019). https://doi.org/10.1038/s41467-019-08621-3

Online Supplementary Materials can be found: https://doi.org/10.1038/s41467-019-08621-3

Introduction

Polygyny, a social mating system in which one male can be mated to several females simultaneously, has evolved numerous times in birds (Lack 1968; Rhoads 1890). Since polygynous male birds have the potential to raise clutches of offspring with multiple females, it is expected that there will be substantially higher variance in the fitness of individual males in polygynous species compared to monogamous species in which males have only one social mate at a time (Emlen and Oring 1977; Payne 1984). In other words, a polygynous male could multiply his potential number of offspring by the number of mates he can attract, leading to differences in reproductive success that could be much larger than the fitness differences of only a few percent that have been shown to shift evolutionary dynamics in a population (e.g. (Lenski and Travisano 1994; Levy et al. 2015)). A polygynous mating strategy also leaves more males unpaired than monogamy; these males could have zero reproductive success unless they were successful in extra-pair fertilizations (Westneat, D. F., Sherman, P. W., & Morton, M. L. 1990). Since the stakes are higher to successfully attract one or more mates, the high variance in reproductive success predicted for polygynous species could amplify the role of sexual selection compared to monogamous species.

Extra-pair paternity (EPP) is common in Oscine species and, like polygyny, may affect the variance of reproductive success in males in a population (Griffith, Owens, and Thuman 2002; Westneat and Stewart 2003). In both monogamous and polygynous species where EPP is common, females will often form a social bond with a single male partner but also copulate with other males (Petrie and Kempenaers 1998). Seeking extra-pair fertilizations could act as a bet-hedging strategy that, in theory, could increase a female's indirect reproductive fitness, for example by increasing the genetic diversity or the fitness of her offspring (Cordero 1998; Yasui and Yoshimura 2018). Since sexual selection can occur when choosing these extra-pair mates (Kempenaers et al. 1992; Wells, Safran, and Dale 2016), males with less attractive secondary characteristics may be doubly penalized; once in obtaining a social mate, and again in attaining copulations with already-mated females. Thus, EPP could potentially increase variance in reproductive fitness (e.g. (Byers et al. 2004), but see (Lebigre et al. 2012)). On the other hand, EPP could putatively decrease the variance in fitness in a population, since socially paired males might invest resources in non-genetically-related offspring due to EPP, and even unpaired males would have the opportunity to sire offspring via EPP (e.g. (Vedder et al. 2011)). In this context, it is unclear what kind of association we would expect to find between EPP and learned song. On one hand, females could potentially diversify their mates by exhibiting a different set of song preferences when they seek out extra-pair copulations than when they choose social mates, thereby placing context-dependent evolutionary pressure on song. Alternatively, high EPP could cause more frequent opportunities for females to act upon the same preferences displayed for social mates, which would hypothetically drive song complexity higher.

For songbirds (class: Aves, order: Passeriformes, suborder: Oscine), song is an important and nearly ubiquitous behavior that has functions in mate choice, intrasexual competition, and mediating other social interactions (Baptista and Morton 1988; Lachlan and Slater 2003; Liu and Kroodsma 2006). Birdsong could be a particularly salient target for sexual selection since, as a learned behavior, it has the potential for rapid change over generations and large variation within a population. Indeed, songs of male songbirds have been demonstrated to influence mate choice in several species; female songbirds across multiple passerine families have been shown to choose mates with larger syllable repertoires, more syllables per song, longer songs, and larger song repertoires (Byers and Kroodsma 2009; Nowicki and Searcy 2004; Soma and Garamszegi 2011). For example, in the great reed warbler, a polygynous species, females preferred larger repertoires both when selecting a social mate and when seeking extra-pair copulation and fertilization (Hasselquist, Bensch, and von Schantz 1996). However, counter-examples have also been reported; for instance, female collared flycatchers appeared to prefer males that had smaller repertoires (Hegyi et al. 2010).

These findings suggest a possible link between the mating system of a species and its song evolution. In particular, a hypothesis has been repeatedly proposed: the intensified sexual selection in polygynous mating systems should drive the evolution of increased song complexity (Catchpole 1980; Irwin 1990). Several papers have investigated this hypothesis that mating system and song evolution are not independent, with mixed results (Searcy and Yasukawa 1995)

(**Table 2.1**). For example, a study of nine species of North American wrens found that the polygynous species had larger song repertoires (Kroodsma 1977). In a meta-analysis of 142 species across several families that did not account for shared ancestry beyond superfamily classification, Read and Weary (Read and Weary 1992) found that polygynous species have higher numbers of syllables per song. In contrast, a study of 21 New World blackbird species found no correlation between song and syllable repertoire sizes and mating system (Irwin 1990). Further, two studies showed that the two polygynous species of *Acrocephalus* warblers had simpler repertoires than the four monogamous species included in the studies (Catchpole 1980; Leisler and Catchpole 1992). Taken together, these results hint at a possible relationship between song and mating system, but it does not appear to consistently follow the prediction that increased sexual selection via non-monogamous mating systems favors more elaborate song repertoires.

EPP has been previously studied in the context of song characteristics, but mostly within the scope of a single species. Some studied species show a positive trend: rates of EPP were higher in males with larger repertoires (great reed warblers (Hasselquist et al. 1996)) or increased song diversity (reed bunting (Suter et al. 2009)). In contrast, there was no observed correlation between song characteristics and EPP in the song sparrow (Hill et al. 2010) and a negative correlation, with greater EPP associated with smaller repertoires, in the sedge warbler (Marshall, Buchanan, and Catchpole 2007). Further, some studies that tested for a correlation between repertoire size and rates of EPP found other factors to be predictive instead, such as singing earlier (blue tit (Poesel et al. 2006)), more consistently (chestnut-sided warbler (Byers 2006)), or at higher amplitude (dusky warbler (Forstmeier et al. 2002)). On a larger scale, EPP was not correlated with song complexity in a study of 65 species in which the authors controlled for shared ancestry but did not include syllable repertoire data (Garamszegi and Møller 2004). More recently, a study across 78 species indicated that within-song complexity (e.g. syllables per song, song duration) was correlated with rates of EPP occurrence, but between-song complexity (e.g. song repertoire, syllable repertoire) was not (Hill et al. 2017) (**Table 2.1**).

Citation	# species	Phylogenetic control	Song parameters	Mate parameters	Test	Results
(Kroodsma 1977)	9	One family: Troglodytidae	Syllable repertoire, Syllables/song, Song-type repertoire, Duration, Continuity	Monogamy / Polygyny	Qualitative observation	Polygynous species had longer and more complex songs, spent more bout time singing, and switched songs more rapidly
(Catchpole 1980)	6	One family: Acrocephalidae	Duration, Complexity	Monogamy / Polygyny	Qualitative observation	2 polygynous species shorter, simpler, less variable songs
(Catchpole and McGregor 1985)	5	One family: Emberizidae	Song repertoire, Variability within a population	Monogamy / Polygyny	Qualitative observation	One polygynous species: smaller song repertoire, less variation within populations
(Irwin 1990)	17	One family: Icteridae Analyzes more closely related species first (5 groups: cowbirds, grackles, ageline blackbirds, meadowlarks, orioles/caciques)	Syllable repertoire, Song repertoire, Versatility	Monogamy / Polygyny	Rank order	Agelaius blackbirds and cowbirds: versatility associated with monogamy. Orioles/caciques: syll rep possibly associated with polygyny. Grackles: versatility associated with polygyny
(Shutler and Weatherhead 1990)	56	One family: Parulinae Some analyses within genera	Syllables/song, Song repertoire, Duration, Song rate, Time singing, Frequency	Monogamy / Polygyny	Mann-Whitney	Monogamous species had higher syllable repertoires
(Read and Weary 1992)	142	Test within superfamilies: Tyrannoidea, Corvoidea, Fringilloidea, Sylvioidea, Turdoidea	Syllables/song, Song repertoire, Interval, Duration, Song rate, Continuity, Versatility	Monogamy / Polygyny	Binomial Rank order	Polygyny associated with lower song rates across all species, Sylls/song positively associated with polygyny across all species
(Garamszegi and Møller 2004)	65	Phylogenetic control - generalized least squares models via software Continuous (Pagel, 1997, 1999)	Syllables/song, Song repertoire, Interval, Duration, Song rate, Continuity, Versatility	EPP (Continuous)	Generalized least squares models for continuous variables	No correlation between song characteristics and EPP
(Soma and Garamszegi 2011)	26, 24	None (for these data)	"Complexity" term encompassing syllable repertoire, song repertoire, and song versatility	EPP (3 groups); Monogamy, Fac. Polygyny, Polygyny	Meta-regression analysis	No significant correlation between song complexity and EPP or mating system
(Hill et al. 2017)	78	Phylogenetic control - PGLS analysis	Syllable repertoire, Syllables/ song, Song repertoire, Duration, Versatility, Syll. transitions/song, Within-song complexity	EPP (continuous); Monogamy / "Polygamy" / Cooperative	Linear regression	Syllables per song (unique), syllable transitions per song, overall within-song complexity positively correlated with EPP

Current study	890	Phylogenetic control	Syllable repertoire (N=120), Syllables/song (N=178), Song repertoire (N=225), Interval (N=131), Duration (N=241), Song rate (N=126), Continuity (N=126)*	EPP (Low/High) (N=142); Monogamy / Polygyny (N=764)**	PhylANOVA, Brownie, BayesTraits, PGLS, GLMM (see methods)	Syllable repertoire and song duration evolve faster in polygynous species; Syllable repertoire is smaller in species with high EPP

Table 2.1. Previous comparisons of song and mating system evolution. *Definitions of song terms tested in this study: Syllable repertoire: average total number of unique syllables an individual produces. Syllables per song (Syllables/song): average number of unique syllables used per song. Song repertoire: average total number of unique songs an individual produces. Intersong Interval (Interval): average length of time separating songs within a period of consistent singing behavior, measured in seconds. Song duration (Duration): average length of a song in seconds. (6) Song rate: the number of full song cycles produced per minute, calculated as 60/(Duration + Interval). (7) Song continuity, the proportion of total song performance time spent producing song, calculated as Duration/(Duration + Interval). Some previous studies used different terms to refer to the same behavioral trait; see Methods. **Definitions of mating system terms used in this study: Monogamy/Polygyny: social monogamy vs social polygyny, based on qualitative or quantitative descriptions. If quantitative, populations with <5% males with multiple social mates classified as monogamous. EPP: Extra-pair paternity, primarily quantitative based on genetic parentage testing of chicks in a population. Species with <10% offspring in a population sired by male who is not the social mate of the female considered to have a low rate of EPP.

Here, we seek to understand the relationship between mating strategies and song across the evolutionary history of songbirds through computational and phylogenetically informed analyses. First, we broaden and deepen the scope of this type of analysis by compiling data from many sources on mating system, extra-pair paternity, and numerous song characteristics from published literature and curated field guides (see **Table 2.1** for definitions of these traits). Gathering data on multiple song metrics enables us to study features of song that have been traditionally categorized as measurements of song complexity, such as total individual syllable repertoire, average syllables per song, and individual song repertoire, as well as features that are often used to measure song performance, such as average song duration, intersong interval, and computed metrics of song rate and continuity. In contrast to most prior studies, we analyze song evolution in the context of both mating system (monogamy/polygyny) and rate of EPP. We perform these analyses while controlling for shared ancestral history using large-scale avian phylogenies (Jetz et al. 2012) of the Oscine suborder.

Our findings suggest a more complex evolutionary relationship between polygyny and EPP and sexual selection on song characteristics than previously has been hypothesized. We test numerous song parameters in this context; in particular, we highlight results from our analyses of syllable repertoire size, a measure of song complexity that enables comparison across many songbird families. Contrary to the hypothesis, we find that polygynous birds do not have systematically larger syllable repertoires than monogamous birds. However, we do find a significant difference in the rate of song evolution between polygynous and monogamous birds: syllable repertoire size seems to have evolved significantly faster in polygynous lineages, but this rapid evolution does not push the syllable repertoire size consistently higher. Instead, we find that the combination of polygyny and very small repertoires or very large repertoires are both evolutionarily unstable; polygyny seems to drive the evolution of more moderate-sized repertoires. Our analyses of extra-pair paternity and song characteristics also yield results that run counter to expectation: we find that syllable repertoire sizes are significantly larger in species with low EPP, and the combination of low EPP and small syllable repertoires is a rare and seemingly unstable state in evolutionary history.

Results

Assembled Database

Here, we compiled a database of mating systems and song characteristics across the songbird lineage from published literature and curated field guides (data in **Online Supplementary Data 1**, variable description in **Table 2.1**, full data curation protocol in Methods); this database includes 764 species with a mating system classification of either social monogamy or polygyny and 141 species that could be identified as having high or low incidence of extra-pair paternity. In addition, we catalogued song characteristic data (defined in **Table 2.2**) for 352 oscine species: 122 species with syllable repertoire size data, 171 species with syllables per song data, 217 species with song repertoire size data, 127 species with song interval data, 228 species with song duration data, and 122 species with both song interval and duration data, allowing calculation of song rate and continuity. In total, the dataset catalogues two mating strategy traits and seven song traits; 890 species across 79 families have data for one or more of these traits (**Table 2.1**). Thirty-nine of these species are suboscines, an outgroup to the oscine songbirds in which learned song has not been observed; thus, they were not used in the analyses testing correlation between mating system and song evolution in this study. We used these species to root our phylogenetic trees and included any with relevant mating system/EPP data in the calculation of the rates of transition between monogamy and polygyny and between low EPP and high EPP, which were used in later analyses. When we found multiple measures for the same species, we included all of them in our database. If an analysis required a single species measure for a song feature, we performed the analysis on the median value for each species and repeated the analysis with minimum and maximum values. If the analysis could accommodate multiple measures from a species, we included all measures.

Correlations between song characteristics and mating system

To test whether mating system and extra-pair paternity were associated with differences in song characteristics while controlling for shared ancestry, we used a published phylogeny containing 9993 avian species (Jetz et al. 2012). Since nearly all extant passerine species are included in this tree, we were able to perform species-level phylogenetically controlled analyses on all of our mating system, EPP, and song data. We measured whether closely related species tended to have similar song characteristics using two different methods, Pagel's λ (Pagel 1999) and Blomberg's κ (Blomberg, Garland, and Ives 2003); all seven song characteristics showed significant phylogenetic signal with both Pagel's λ and Blomberg's κ (**Online Supplementary Table 1**).

To compare the distribution of song phenotypes between species with different mating systems while accounting for the statistical non-independence of closely related taxa, we performed a phylogenetic ANOVA (PhylANOVA, R package: phytools (Revell 2011), based on the algorithm in (Garland et al. 1993)). We found that no song characteristic was significantly different between monogamous and polygynous species after multiple-hypothesis correction for seven tests with the Holm-Bonferroni method (**Figure 2.1, Table 2.3**). We then reconstructed the ancestral states of both mating system and song characteristic and mapped them onto the phylogenies (**Figure 2.2, Online Supplementary Figures 1–6**). No tests of correlation between mating system and song traits were significant, though we note that our results for syllables per song and song rate trend in the same direction as the results of Read and Weary (Read and Weary 1992), who found that that polygynous species tended to have a lower song rate and higher mean syllables per song, which they term "syllable repertoire". (**Figure 2.1**).

When testing for correlations between EPP and song characteristics, we find that syllable repertoire is significantly higher in species with low EPP, even when controlling for multiple hypothesis testing (**Figures 2.1 and 2.3, Table 2.3, Online Supplementary Figures 7-12**). These results did not qualitatively differ with a jackknife resampling test, removing each family in turn (**Online Supplementary Data 2**). No other song characteristics significantly differed between high and low EPP species after correcting for multiple hypothesis testing; we found a trend toward higher syllables per song in species with low rates of EPP that did not pass the threshold for significance with the Holm-Bonferroni correction (**Online Supplementary Table 2**).

Rates of song evolution under different mating systems

Next, we assessed whether song characteristics evolved at different rates during polygynous versus monogamous periods in evolutionary history using the Brownie algorithm (O'Meara et al. 2006). This algorithm tests whether a continuous trait evolves at different rates during the different ancestral states of a discrete trait, for example, whether syllable repertoire size evolved faster in polygynous or monogamous lineages. First, we estimated the ancestral states of a discrete mating system trait—here, polygyny vs. monogamy and low vs. high EPP—and rates of transition across a phylogeny using all oscine and suboscine species for which we had trait data. Using the resulting estimated rates of transition between these states, we generated

1000 unique stochastic character maps to simulate the evolutionary history of these ancestral states on the phylogeny (e.g. **Online Supplementary Figure 13**) (Bollback 2006; Revell 2011). We then used the Brownie algorithm to test whether mating system affects the evolutionary rate of song characteristics by comparing the average log-likelihood of the one-rate model (i.e. null model: song trait evolution occurs at same rate regardless of mating system) to that of the two-rate model (i.e. the rate of evolution of a song trait differs in monogamous vs polygynous states: **Online Supplementary Figures 14–15**). We did not include runs that failed to converge within 75,000 iterations in our average log-likelihood.

With this procedure, we find that the rate of syllable repertoire size evolution is higher in polygynous lineages than monogamous lineages, a relationship that is robust to jackknife analysis across families and persists when using either the minimum or maximum syllable repertoire values found in the literature (**Figure 2.4A**, **Table 2.3, Online Supplementary Figure 16, Supplementary Data 3**). Thus, polygyny appears to lead to faster evolution of syllable repertoire size, but not to systematically larger repertoires. This rapid but non-directional evolution could indicate that stochastic fluctuations of syllable repertoire occurred frequently in evolutionary history, or that the direction of evolution varies for different ranges of syllable repertoires, such as polygyny driving syllable repertoire toward either extreme or moderate values.

With the same type of analysis, we also observed that song duration evolved significantly faster in monogamous lineages, again without producing an appreciable difference in duration between monogamous and polygynous species **(Figure 2.4B)**. This result was also mostly robust to a jackknife analysis: with each family removed in turn, we observed a similar rate distribution, but with the removal of two families, the difference in rates was no longer significant (Brownie

Jackknife likelihood-ratio test: Icteridae *p*=0.066, Fringillidae *p*=0.053; **Online Supplementary Figure 17, Supplementary Data 3**).

We also find that the rate of evolution of syllables per song is significantly higher in monogamous lineages than polygynous lineages. However, this result was not robust to our jackknife analysis as it appears to be driven solely by the Mimidae family: when the four mimid species in this dataset, all monogamous, are removed, we find the opposite result that the rate of evolution for syllables per song is higher in the polygynous state. EPP did not significantly affect the rate of evolution of any song trait, including syllable repertoire (**Figure 2.4C**).

Correlated evolution between song and mating variables

Do changes in syllable repertoire size occur independently of mating behaviors, or are song changes more (or less) likely in non-monogamous lineages? To detect non-linear evolutionary trends that might produce the complex patterns of evolutionary rate and direction observed in **Figures 2.1–2.4**, we devised a method to test for correlated evolution between mating strategy and song characteristics across the full range of continuous song trait values by separating these continuous values into a series of discrete categories. The results of this analysis to detect correlated evolution (BayesTraits (Pagel 1994; Pagel and Meade 2006)) add a layer of complexity that helps explain the first two results. When we break down our syllable repertoire size values into discrete transitions between smaller and larger repertoires, we find evidence for correlated evolution between mating system and syllable repertoire size (**Figure 2.5**): the combination of polygyny and very small syllable repertoires is evolutionarily unstable, with an **ure 2.5A**). In addition, polygyny with very large syllable repertoires is also an unstable combination, with species tending toward smaller syllable repertoires (**Figure 2.5C**).

Interestingly, we found a qualitatively similar pattern of correlated evolution for each of our studied song characteristics. When combined with polygyny, both very high and very low values of song characteristics were relatively unstable, with elevated rates of evolutionary transitions toward more moderate values of syllables per song, song repertoire size, intersong interval, song duration, song rate, and song continuity (**Online Supplementary Figures 18–23**).

We also tested for correlated evolution of song characteristics and EPP. Our results complement our finding that syllable repertoire sizes are larger in species with low rates of EPP: we find patterns of correlated evolution that suggest that the combination of low rates of EPP and small syllable repertoires is unstable, with elevated transition rates leaving this state (**Figure 2.6**). Similarly, there are rapid evolutionary transitions away from the combination of large repertoires and high rates of EPP. These observations agree with the findings that species with low rates of EPP have larger syllable repertoires on average. When we tested for correlated evolution between EPP and other song characteristics, we found a similar pattern with syllables per song; rates of evolutionary transition were fastest when leaving the combinations of low EPP with low syllables per song and high EPP with high syllables per song (**Online Supplementary Figure 24**). For the remaining song characteristics, we observed either no significant evidence for correlated evolution with EPP (song repertoire size, song rate) or qualitatively different patterns from that of EPP and syllable repertoire size (intersong interval, song duration, song continuity) (**Online Supplementary Figures 25-29**).

In order to account for variation in tree branch lengths and topology, we repeated PhylANOVA, Brownie, and BayesTraits analyses using a consensus tree computed from trees only containing species with genetic data and using multiple individual trees. These results were consistent with those performed using our original consensus tree (**Online Supplementary Figures 32-35,** and **Supplementary Table 3**).

Combined effects of mating system and EPP on song evolution

Finally, we sought to test whether there are interacting effects of mating system and EPP on the evolution of song characteristics. We performed phylANOVAs to test whether there was a significant difference in song characteristics between the different combinations of mating system and EPP: Monogamy+LowEPP (N=7), Monogamy+HighEPP (N=31), Polygyny+LowEPP (N=6), Polygyny+HighEPP (N=8). Syllable repertoire size was significantly associated with species' combined mating system/EPP classification (PhylANOVA $p = 0.0265$). This significant result seems to be attributable to the significantly larger syllable repertoires in the Polygyny+LowEPP group (N=6) compared to the Monogamy+HighEPP group (N = 31) (PhylANOVA $p = 0.045$); no other pairwise comparisons were significant (**Online Supplementary Figure 36** and **Online Supplementary Table 4**). Next, we used PGLS to test whether song characteristics show a relationship to mating system, EPP, and the interaction between the two. The PGLS recapitulated the result showing High EPP to be correlated with lower syllable repertoire, but did not show evidence that the interaction between EPP and mating system has influenced syllable repertoire (**Online Supplementary Table 5**). Finally, we performed a Generalized Linear Mixed Model (GLMM), which allowed us to both test the interacting effects of mating system and EPP and control for multiple song measures from a single species when we had more than one published estimate in our database. This GLMM analysis also reaffirmed our result that EPP significantly predicts syllable repertoire (GLMM $p < 0.001$), but no other song variable was associated with EPP. Mating system and the interaction between mating system and EPP were not significant predictors of any song variable (**Online Supplementary Figures 37 and 38** and **Supplementary Table 5**). Finally, the GLMM analysis corroborated the results of our tests for phylogenetic signal: building a linear mixed model with only phylogenetic relationships (without any mating behavior data) led to reasonable predictions of most song characteristics (**Online Supplementary Figure 39**).

Discussion

Here, we assemble a large-scale database of mating system classifications and song characteristics across songbird species, which promises to be a useful tool for future evolutionary studies. By extracting information from published sources, including academic journals and curated field guides, we produce a synthesized database including mating system classification (polygynous vs. monogamous) for 764 species, EPP classification (high vs. low rates) for 142 species, and at least one song characteristic (out of syllable repertoire size, song repertoire size, syllables per song, song duration, song rate, song interval, and song continuity) for 360 species (352 oscine, 8 suboscine) (**Online Supplementary Data 1**). We synthesize this database with an avian phylogeny and show that song characteristics have strong phylogenetic signal, underscoring the importance of controlling for shared ancestry.

With this database, we perform phylogenetically controlled analyses to assess whether mating strategies alter the evolutionary dynamics of learned song. Since polygyny in a species likely increases the variance in reproductive success and potentially augments sexual selection pressures (Payne 1984), there is a long-debated hypothesis that polygynous species should evolve more complex or elaborate songs (Catchpole 1980; Catchpole and McGregor 1985; Irwin 1990; Kroodsma 1977; Lande 1981; Read and Weary 1992; Shutler and Weatherhead 1990). We show that, contrary to this prediction, polygynous species overall do not have systematically larger syllable or song repertoires than monogamous species. However, we find that the rate of evolution of syllable repertoire size is significantly higher in polygynous lineages, but this rapid evolution does not consistently push the syllable repertoire toward increased complexity. Instead, we find that the combinations of polygyny with very small repertoires and polygyny with very large repertoires are both unstable, and polygyny appears to drive an accelerated rate of song evolution toward moderate values of syllable repertoires (**Figure 2.5**).

When we analyze song evolution in the context of extra-pair paternity, we find different patterns from what has been previously predicted: syllable repertoires are significantly larger in species with low rates of EPP, and there is correlated evolution that favors transitions to states combining low rates of EPP with large syllable repertoires and high rates of EPP with small syllable repertoires. These results are in contrast to the recent analysis of 78 species by Hill *et al.* (Hill et al. 2017), which presented a phylogenetically controlled regression analysis of EPP percentage and their own measured data from three individual song recordings for each species. With these quite different methods and data types, they reported a positive association between EPP and within-song complexity, but not for song or syllable repertoire size, warranting follow-up studies with more comparable datasets.

Our results generate new questions that necessitate further investigation. First, why do polygynous species appear to evolve moderate-sized song characteristics? One hypothesis could be that the amplified sexual selection of polygyny favors repertoire sizes, and perhaps song characteristics more generally, that strike a balance between being challenging to learn for the male and requiring a low time investment to assess by the female. Second, why might polygyny lead to accelerated rates of evolution of syllable repertoire size and song duration? This result appears to be linked to our observation that very large and very small repertoires, as well as very long and very short songs, evolve to more moderate values in polygynous lineages. The ecological mechanism underlying this rapid evolution remains unclear, but, intuitively, we might think of this as the effect of combined forces of two repellant states at either extreme end of the spectrum for a given song characteristic at which selective pressure increases. Third, does EPP increase or decrease sexual selection pressure, and is this effect different in polygynous versus monogamous species? Our results indicate that, contrary to predictions, EPP seems to decrease the intensity of selection for large syllable repertoire, potentially to the point of selecting against larger syllable repertoires. This result also stands in contrast to our findings regarding mating system-dependent rates of evolution of syllable repertoire. Taken together, this suggests that EPP and polygyny do not have uniform effects on sexual selection of song. Thus, their interactions may produce more complex patterns of selection, leading us to hypothesize that the effects of EPP on variance in male reproductive success is intrinsically dependent upon the dominant social mating system of the species. Alternatively, females in monogamous vs polygynous species may have different preferences in song or differentially value other features relative to song in social mates versus extra-pair mates. None of our analyses that combined mating system and EPP yielded significant results for any song characteristic, but this may be due to the relatively low numbers of species that had a combination of mating strategies other than socially monogamous with high EPP. This question should be revisited as mating strategy and song data becomes available for more species.

It is difficult to directly compare song characteristics across bird species that structure their songs very differently. In order to mitigate these difficulties, we highlight results on syllable repertoire size as a metric that can be interpreted across the songbird lineage. Other metrics, such as syllables per song and song repertoire size, can be influenced both by how a species structures its song and by how human observers define a 'unit' of song, making them less ideal for a large cross-species analysis. These metrics may be particularly unreliable for species without stereotyped song bouts, such as Mimid species. Indeed, excluding the Mimidae family from some analyses led to statistically different results (Mating System + Syllsong: Brownie) compared to when Mimids are included (**Online Supplementary Data 2-3**). In addition, for tractability we encoded both mating system (polygyny and monogamy) and rate of extra-pair paternity (high EPP and low EPP) as binary traits. However, the mating systems of passerines are more nuanced than is reflected in this study; for example, rates of polygyny can vary widely among polygynous species. Additionally, the rarer mating systems of polyandry, polygynandry, and promiscuity are not reflected in this study.

Further, although song complexity and repertoire size can frequently predict mate choice in laboratory experiments, mate choice in the wild is complicated by numerous other factors. Territory quality, particularly in the sense of food availability, might be particularly salient to potential mates of polygynous males, since the territory might need to support multiple clutches with less contribution to nest provisioning from the polygynous male. Morphological traits such as plumage coloration and sexual size dimorphism are also subject to sexual selection and may play a greater role in mate choice in some species. Song evolution is also made more complicated by other functions of song that may also be selected upon. Some of these functions, such as intrasexual aggression and territory defense, may put similar evolutionary pressures on song as mate choice, but other social functions may not. Finally, the evolution of female song has been historically understudied, but could be correlated with male song evolution while being subject to different selective pressures.

Here, we use phylogenetically controlled computational analyses to demonstrate that mating system and extra-pair paternity can affect the evolution of song in multiple ways—by influencing the direction of evolution, the rate of evolution, or the likelihood of transitioning to a particular state. In contrast to the long-discussed prediction that polygyny should lead to the evolution of more elaborate songs, these observations suggest more complex, nonlinear dynamics, indicating that the evolutionary associations between non-monogamous mating and sexually selected characters should be analyzed in broader contexts and using methods that incorporate the rate and direction of evolution. In addition, extra-pair paternity had a significant relationship to song evolution in the opposite direction than has been typically predicted for non-monogamous mating, suggesting that different types of non-monogamous mating behaviors can have dramatically different effects on sexual selection. Building upon these complex dynamics, we propose that EPP and mating system may not influence sexual selection on song independently from one another; further investigation into their interacting effects is needed.

Methods

Database assembly

To assemble the most comprehensive database of mating systems for passerine birds, we compiled information from multiple sources. These included curated field guides, previous studies that investigated mating system evolution, song evolution, or both, and other studies that we found via targeted searches using Google Scholar and Web of Science, as detailed below.

Mating system and extra-pair paternity data collection

Birds of North America Online (BNA (Ferer n.d.)) is an online encyclopedia curated by the Cornell Lab of Ornithology containing profiles of several hundred avian species that reside in or migrate through North America, including Hawaii. Profiles include life histories compiled from peer-reviewed literature, field guides, and personal observations of profile authors, as well as multimedia from Macaulay Library (Cornell University Cornell Lab of Ornithology 2009) and eBird (Sullivan et al. 2014). Profiles for 329 Oscine and 34 Suboscine species were available, with varying degrees of documentation. The available search function in BNA only queried species names, so, in order to generate an initial database, we navigated to each species page and used the search page function to find the terms “monogamy”/“monogamous” and “polygyny”/“polygynous” using the character strings “monogam” and “polygyn”, respectively. This search yielded preliminary mating system data for 291 species. These included passages that did not qualify to be included in the final dataset in isolation due to ambiguous wording (i.e. “probably monogamous,” “usually monogamous,” etc.). If a citation for a digitally available peer-reviewed study was provided, we attempted to find the cited source and other primary literature by searching “(species name)” + “monogam*/polygyn*” (based on each species’ BNA entry) in Google Scholar and Web of Science, though we did not remove otherwise definitively classified mating system data for which we could not locate the original source. We searched more broadly for sources that compiled data on species mating systems using Google Scholar and Web of Science with the search terms “Passeriformes”/“songbird” + “mating system”/“social monogam*”/“polygyn*”/”extra-pair”.

We supplemented data on mating systems and EPP found in traditional publications with species descriptions found in Handbook of Birds of the World Online (HBW (del Hoyo et al.

n.d.)). While not peer-reviewed, HBW is rigorously edited on a yearly basis with a demonstrably high standard for data, and several other studies (Price and Griffith 2017; Quintero and Jetz 2018; Sayol et al. 2016) have used it as a primary source. We searched for pages using the following search terms with associated yields: (1) polygyn* AND song NOT "presumably polygynous" NOT "probably polygynous" = 135 species; (2) monogam* AND song NOT "presumably monogamous" NOT "probably monogamous" = 826 species. "We included "song" as a search term to attempt to limit results to passerine species and manually eliminated search results from non-passerine species. We read the "Breeding" and "Voice" section of each species that was returned in the highlighted searches above to gather any available data on mating system and song characteristics, respectively.

Any species with unambiguous mating system information was included in the final dataset. We also included species described with the terms "primarily monogamous/polygynous". We included species with notes including "apparently" or "appears monogamous/polygynous" only if there was another corroborating source. We used species described as "mostly," "normally," "typically," and "generally monogamous/polygynous" only if quantitative measurements were also given. Species with "Occasional," "Opportunistic," or "Facultative" polygyny were considered monogamous unless other published studies provided disputing evidence.

Song data collection

We performed searches in Google Scholar and Web of Science using the queries generated by the following search terms: [species name] + "syllable repertoire", "Passeriformes" + "syllable repertoire", and "songbird" + "syllable repertoire". We then repeated these search queries using “song repertoire” instead of “syllable repertoire”. For all searches, [species name] was replaced with the the taxonomic and common name of each species for which Birds of North America yielded mating system data. This was supplemented with values obtained via HBW.

Mating system data classification

There were often multiple sources for a species’ mating system. In the event that sources disagreed, we used the following procedure to decide which mating system to code in our database. 1) If any source gave a percentage of the male population that were polygynous, defined as males with >1 social mates (Hasselquist and Sherman 2001), we used that value with a threshold of 5% of sampled males exhibiting polygyny to determine the species mating system; i.e. if <5% of males exhibited polygyny, we classified the mating system as social monogamy; if ≥5% of males exhibited polygyny, we classified the mating system as polygyny. We do not suggest that this threshold value is inherently biologically relevant, but it is consistent with previous studies (Hasselquist and Sherman 2001; Lack 1968; Møller 1986; Soma and Garamszegi 2011; Verner and Willson 1969), allowing us to evaluate our results in the context of earlier research. 2) If multiple sources provided percentages of polygyny in a population that yielded conflicting mating system classification, the median value of those percentages was used. 3) If multiple sources provided qualitative data on mating system that were in disagreement, HBW was used to determine the final mating system classification. If HBW did not mention mating system in its entry for a particular species when there was such a discrepancy, the species was omitted from the dataset.

Extra-pair paternity data classification

To determine whether a species exhibited high or low incidence of extra-pair mating, we prioritized sources with genetic parental determination data. A review of extra-pair paternity studies estimated the cross-species average to be ~11% of offspring per nest to be attributable to extra-pair mates (Griffith et al. 2002). In line with this estimate and with previous studies (Soma and Garamszegi 2011), we used a 10% threshold for either extra-pair young or nests containing at least one extra-pair chick to estimate the frequency of extra-pair paternity in that species (<10% = low EPP; ≥10% = high EPP). If a source provided data on both the percentage of offspring in a population that were genetically unrelated to their social father and the percentage of nests that contained at least 1 extra-pair young, we used the former metric, which is more commonly used, to determine species EPP. Either value might be partially determined by cryptic female mate selection, which is a potentially important mechanism by which females may influence which males achieve genetic paternity after extra-pair copulation. Rarely, studies have reported extra-pair mating behavior in terms of observed copulations. We did not include these values for categorization of EPP in our database due to the rarity of the data and since all species for which this was reported also had a reported %EPP. If we found multiple studies with unique %EPP values for a species, we based the classification on the median of all reported %EPP values.

Controlling for different metrics of extra-pair paternity

When studies report a rate of extra-pair paternity, they typically use one of two metrics. The first, referred to as “extra-pair young,” (EPY) “extra-pair fertilizations,” (EPF) or just EPP (as used in this manuscript), refers to the percent of the total young in a population that are not genetic offspring of their social father. The second is usually referred to as “extra-pair broods”

(EPB), and represents the percent of broods in a population that have at least one egg/chick that is not the genetic offspring of their social father. We prioritized EPP/EPF/EPY values and used EPB values only if the former were unavailable. To test whether the different metrics of EPP might influence our conclusions, we performed another test using EPP classifications that were inclusive of EPB values. In this method, if a study reported both EPY and EPB for a single population, we used the mean of the values as that study's EPP value, instead of just the EPY/EPF value. We then used the median of all studies, including studies that only reported EPB, to obtain the EPP value we used in the 10% threshold classification of the species. Using this method did not change the result of phylANOVA for EPP and syllable repertoire, the only test involving EPP that yielded a significant result, nor of any other test.

Song data classification

Nomenclature for song characteristics is variable across sources. For example, in some sources, "syllable repertoire" means the average total number of syllables an individual uses in song; however, several others, including Read and Weary (Read and Weary 1992), have considered this term to mean "average number of unique syllables per song." We checked the primary source methods whenever possible to ensure we classified each song parameter correctly. We categorized several measures that have generally been used as proxies of song complexity, elaborateness, or variety, as follows: (1) Syllable repertoire size (Syll Rep), defined as the average total number of unique syllables an individual uses across songs. Some studies also called this 'song repertoire,' so we read the relevant methods sections for clarification. (2) Syllables per song (Syll/song), the average number of unique syllables used per song (also occasionally called syllable repertoire, e.g. (Read and Weary 1992)) (3) Song repertoire size

(Song Rep), the average total number of unique songs an individual produces. We also categorized song metrics that have traditionally been used as proxies of song performance, as follows: (4) Intersong Interval (Interval), the average length of time separating songs within a period of consistent singing behavior, measured in unit: seconds. (5) Song duration (Duration), the average length of a song, defined as the number of seconds of consistent singing between periods of silence; sources may have differed in this definition. (6) Song rate (Rate), the number of full song cycles produced per minute, calculated as 60/(Duration + Interval). (7) Song continuity, the proportion of total song performance time spent producing song, calculated as Duration/(Duration + Interval).

While we included other song characteristics when available, we only actively searched for species' syllable repertoires and song repertoires. Several sources and previous studies provided these other metrics in addition to repertoire or mating system data. Any value we encountered during this search we recorded in **Supplementary Data 1**, but the database should not be considered exhaustive. For values of syllable repertoire obtained from Moore *et al.* (Moore et al. 2011), we used only the value from the single study they utilized, since they were systematic in their selection of data. Occasionally we could infer the syllable repertoire of a species if we had data for 'syllables per song' and the song repertoire was equal to 1.

If any song data was given for any species as a range, we used the median value. If multiple sources were found for a particular species, we used the median value across sources and noted in the dataset the minimum and maximum values observed for that species. We $\log_{10}$-normalized the data for all song characteristics. Some species, in particular those belonging to the Mimid family, do not have discrete, stereotyped songs, instead improvising their songs for an extended, often uninterrupted period of time. Occasionally these species had measured values for syllable repertoire, syllables per song, and/or song repertoire in the literature, but more often they were qualitatively described as “large.” Previous studies on song evolution included these species in quantitative analyses by assigning an arbitrarily high value to these characteristics ranging from 100 to 1000, depending on the study (Garamszegi and Møller 2004; Macdougall-Shackleton 1997; Read and Weary 1992; Soma and Garamszegi 2011). For species that were assigned ‘large’ or arbitrary repertoires in the literature and had no other quantitative assessment of repertoire available, we assigned syllable repertoire, song repertoire, syllables per song, and song duration to have a minimum value of 100, a maximum value of 1000, and a median value of 500.

Assembling phylogenies

We assembled two phylogenies for use in our analyses: one including only avian species which had genetic sequence data integrated into the tree and one which also integrated species without genetic sequence data. For the latter, we obtained 1000 trees, each containing 9993 species from Birdtree.org (Jetz et al. 2012, 2014, n.d.) by randomly choosing one of ten sets of 1000 trees available in the BirdTree data downloads. These trees (HackettStage2 7001–8000) were built by generating relaxed-clock molecular trees for each of 158 clades, then arranging these clades on a avian-wide tree using the backbone determined in Hackett *et al.* (Hackett et al. 2008) (full methods in Jetz *et al.*(Jetz et al. 2012)). We used consense in Phylip (Feisenstein 1989) to generate a consensus tree that specified how many out of the 1000 input trees supported each node (full consensus tree in **Supplementary Data 4–5**). We used the function consensus.edges (R package: phytools, method: mean.edge) to compute a consensus tree with branch lengths from the 1000-tree sample. For any nodes that resulted in multifurcation, we used function multi2di (R package: ape) to randomly reassign multifurcations as a series of dichotomous nodes, resulting in a bifurcating tree necessary for trait analyses. The edges generated through this method are defined by default as having a length of zero, which the statistical tests below cannot process. We set these edge lengths to an arbitrarily very low value of 10^{-19}. (We found no qualitative differences in our results with different values of this branch length; see Supplementary Information.) For more efficient computation, we used drop.tip (package: ape) to eliminate species for which we had neither song characteristics nor mating system data. For subsequent analyses comparing mating system and song characteristics, we dropped tips of species with no data for a given comparison (see code in **Supplementary Software**).(Hackett et al. 2008) We repeated this procedure using a set of 1000 trees with only the 6670 species for which Jetz *et al. (Jetz et al. 2012)* had a genetic sequence (HackettStage1 7001-8000). To ensure that results were robust to differences in branch lengths and tree topography, we repeated all analyses using this Gene Tree, as well as on each of the first 100 trees in HackettStage2 7001-8000 for PhylANOVA and Brownie, and the first 20 trees for BayesTraits.

Phylogenetic comparative analyses

We used a set of phylogenetic comparative analyses to determine the relationship between the evolution of mating systems and each song characteristic in turn. First, we performed an ancestral character estimation (Paradis, Claude, and Strimmer 2004) using the ace function (R package: ape) for each trait separately. We calculated the ancestral state at each node for each trait, both continuous (song variables) and discrete (mating system classification, extra-pair paternity). We used this to visualize the likelihood that each trait was present at each ancestral node of the tree.

To assess the phylogenetic signal of each song characteristic, we calculated both Pagel's λ (Pagel 1999)and Blomberg's κ (Blomberg et al. 2003) using the function phylosig in phytools (Revell 2011). For Pagel's λ, we assessed significance with the included permutation test, resampling 100,000 times. For each song characteristic, we analyzed all of the species for which we had data for that characteristic, even if we did not have corresponding mating system data for that species.

To assess whether bird species had significantly different song characteristics dependent on their mating strategies, we used the function phylANOVA (R package: phytools (Revell 2011)) to determine whether song traits were significantly different between polygynous and monogamous species or between species with low EPP and high EPP, controlling for phylogeny. We also calculated the residuals of the ANOVA (**Online Supplementary Figures 40–41**). Next, we used the Brownie algorithm (O'Meara et al. 2006) to assess whether the rate of evolution of song characteristics was significantly different between polygynous and monogamous branches of the phylogeny. First, we map the evolution of mating system onto the phylogeny and test whether evolutionary transitions between monogamy and polygyny occur at equal rates (Equal Rates (ER) model) or whether the transition from monogamy to polygyny occurs at a significantly different rate than the transition from polygyny to monogamy (All Rates Different (ARD) model). To do this, we calculated the rates of evolution of mating system across the full tree and dataset using ace (R package ape) using both an ER and ARD model to estimate transition rates. We performed an anova (R package: stats) on the two models to determine whether the ARD transition rate model was a significantly better fit. We similarly tested the transitions between high and low rates of EPP. The ARD model was a significantly better fit for polygyny (ANOVA $p = 6.05\times10^{-19}$; log-likelihood ER = −302.59, ARD = −263.06), and EPP (ANOVA $p = 0.0140$; log-likelihood ER = −84.49, ARD = −81.47) so we recorded the transition rates from the ARD model (monogamous to polygynous: 0.0962, polygynous to monogamous: 0.0114; low EPP to high EPP: 0.0407, high EPP to low EPP: 0.0193) and used them to generate the stochastic character maps (simmaps (Bollback 2006)) for all subsequent Brownie tests. For each song trait, we generated 1500 simmaps from the subsetted tree and dataset for that mating system/song trait combination using the rates derived from the full tree.

We then ran brownie.lite (R package: phytools) once per simmap. For each run, this function calculates the log-likelihood of an ER model (i.e. the song trait evolves at equal rates regardless of the mating system state the ancestral species occupies) and ARD model (i.e. the song trait evolves at a rate dependent on an ancestral species' mating system). To determine whether the ARD model is a significantly better fit than the ER model, we calculated the average log-likelihood of the one-rate model and the average log-likelihood of the two-rate model over our 1500 runs and performed a likelihood ratio test, assessing significance with the function pchisq in R. For some runs, the Brownian motion run does not converge by the time it reaches the maximum number of iterations (we used maxit = 75000, increased from a default of 2000). If the model did not converge or did not finish in the maximum runtime (16 seconds), we discard these runs from our final analysis and do not include them in the average log-likelihood.

Currently, there is no published method to test for correlated evolution between a discrete trait and a continuous trait, to our knowledge. We developed a method to utilize the maximum likelihood test of dependent trait evolution of two discrete traits available in BayesTraits to test for correlated evolution between mating strategy and song across the full range of continuous song trait values. This algorithm calculates whether the rate of evolution of one trait is dependent on the state of another trait. For example, we can use this algorithm to ask whether an evolutionary change in mating system alters the likelihood of an evolutionary change in a song characteristic. BayesTraits compares current species data with simple and complex continuous Markov models of evolution of discrete traits on a given phylogeny to determine whether the complex model (i.e. dependent trait evolution) describes the data sufficiently better than the simple model (i.e. independent trait evolution) to justify accepting the complex model. BayesTraits reports marginal likelihoods for the complex and simple models (function Discrete in package btw). We used function LRtest (package: lmtest), which returns the likelihood ratio statistic (LRstat) and *p*-value, to perform the likelihood ratio test to determine whether to accept the complex model of dependent evolutionary rates over the simpler model of independent evolutionary rates. Since this model requires both traits to be binary, we classified continuous song characteristics as binary groups (low or high) based on a delineating threshold. Instead of choosing the threshold arbitrarily, we used R package btw (BayesTraits Wrapper) to perform the tests using each unique value of the song characteristic data as the threshold, repeating the test 100 times at each threshold. We plotted the *p*-values over the threshold used for all tests. (**Figure 2.5D**, **2.6D**)

Each repetition of the test analyzed whether the evolution of mating system and song were correlated by performing a log-likelihood comparison between a continuous-time Markov model where the traits evolve independently and a model of correlated evolution, in which the evolution of one trait is dependent on the state of the other trait. BayesTraits Wrapper function btw returns the computed rates of transition between the four total states and a *p*-value indicating whether we can reject the null hypothesis that evolution of the two traits occurred independently.

The computed transition rates vary dramatically across the range of song trait values, depending on the threshold delineating the low and high value categories. When the threshold is set as a lower value, the test effectively evaluates the rate of song trait values switching from very low to moderate and vice versa. When it is set as a higher value, the calculated rates are the rates of switching between very high song trait values and medium-low values. To account for this nuance, we segmented the vector of unique song trait values into bins and calculated the mean and 95% confidence interval of each state transition rate from all of the runs across all of the thresholds in each bin. We present the results with the thresholds binned into three groups in the main text, but also include the analyses with the thresholds divided into two, four, and five groups in **Online Supplementary Figures 42–47**.

Accounting for variation in measurements

To account for variation in song characteristic measurements across studies, we repeated PhylANOVA and Brownie calculations with the minimum and maximum values of each species' song characteristics obtained from the literature. For these runs, any species for which we had only one data source for the song trait in question kept that trait value. We replaced the median values of any species with multiple sources with the minimum and maximum trait values, respectively, for a total of three runs per song characteristic. We calculated the values for song rate and continuity in our database based on the median song duration and intersong interval values in the literature. For both song rate and continuity, the minimum and maximum values were calculated using the maximum duration and interval values and minimum duration and interval values, respectively.

Testing for an interaction between mating system and EPP

We performed a phylogenetic generalized least squares (PGLS) analysis using the function gls (R package: nlme (Pinheiro et al. 2014)), computing phylogenetic correlation using the function corBrownian (R package ape). This technique enables us to analyze whether there is a relationship between song characteristics and multiple variables at once: mating system, EPP, and the interaction between mating system and EPP.

We also performed a GLMM for each song characteristic using function MCMCglmm (R package: MCMCglmm (Hadfield 2010)) which allowed us to include all values found in the literature for each species. For this GLMM, we subsetted the data to include only those species for which we had both mating system and EPP classifications, which also allowed us to evaluate the interacting effects of mating system and EPP. We verified the computed model using posterior predictive checks using functions predict and simulate in the package MCMCglmm.

Controlling for differences between families

To ensure that our Passeriformes-wide findings were robust and not driven by any particular family, we performed a jackknife analysis in which we reran all analyses on a series of subsetted datasets with each family removed in turn. We determined the family of each species based on its classification in the 2017 version of the eBird Clements Integrated Checklist (Clements 2007).

Testing the effect of multifurcations on our results

The avian phylogeny constructed by Jetz et al. contains numerous multifurcations, which we observed both in the sample of 1000 trees that we extracted from their tree distribution and in the consensus tree we constructed from those 1000 trees. Since a dichotomous, bifurcating tree is required for phylogenetic comparative analyses, we used the function multi2di (package: ape) to restructure multifurcated nodes on the consensus tree into randomly organized dichotomous nodes with arbitrarily small branch lengths between them. Since the bifurcations are assigned randomly, we repeated the multi2di procedure 10 times with an arbitrary branch length of 10^{-19} and 10 times with an arbitrary branch length of 10^{-4}. Then, we repeated our phylANOVA and Brownie analyses with each of these trees. None of these computed differences led to qualitative differences in the results of any test.

Data Availability

The authors declare that all data supporting the findings of this study are available within the paper and its supplementary information files. Data are also stored on GitHub.com/CreanzaLab/MatingBehaviorsAndSongEvolution.

Code Availability

The authors declare that all code used in the analysis of data within this paper are available in the associated supplementary information files (**Supplementary Software**). Code is also available at GitHub.com/CreanzaLab/MatingBehaviorsAndSongEvolution.

Song Trait	**Definition**
Syllable repertoire	Mean total number of unique syllables an individual uses across songs
Syllables per song	Mean number of unique syllables used per song
Song repertoire	Mean total number of unique songs an individual produces
Intersong interval	Mean length of time separating songs within a period of consistent singing behavior (unit: seconds)
Song duration	Mean length of a song, measured as the length of time of consistent singing or discrete songs between periods of silence; sources may have differed in definition based on the song structure of a studied species (unit: seconds)
Song rate	Number of full song cycles produced per minute
Song continuity	Proportion of total song performance time spent producing song

Table 2.2. Definitions for song characteristics used in this paper. Here, we note the definition that we used throughout our analyses. Some previous studies have characterized birdsong in different terms; for example, what we term a ‘syllable’ is also called a strophe, note, element, etc. Further, Read and Weary (1992) define a ‘syllable repertoire’ as the number of syllables in a single song, whereas we classified that data as ‘syllables per song’ in our database. When we gathered song characteristic data from cited sources, we classified these data according to the definitions given in that source, regardless of the terms used.

	Syllable repertoire	**Syllables per song**	**Song repertoire**	**Intersong interval**	**Song duration**	**Song rate**	**Song continuity**
Mating system # of species	96	133	145	95	177	91	91
PhylANOVA *p*-value	0.25	0.0597	0.063	0.579	0.899	0.0698	0.221
Min; Max	0.342; 0.236	0.028; 0.084	0.048; 0.078	0.76; 0.496	0.952; 0.874	0.696; 0.58	0.35; 0.1632
Jackknife resampling	None significant	7/32 families $p<0.05$	5/32 families $p<0.05$	None significant	None significant	2/24 families $p=0.04$	None significant
Brownie *p*-value	0.0056	0.4248	0.1478	0.4881	0.0075	0.0266	0.3985
Rate higher in:	Polygyny	N/A	N/A	N/A	Monogamy	Polygyny	N/A
Min; Max	0.014; 0.008	0.042; 0.159	0.088; 0.127	0.007; 0.646	0.024; 0.007	0.373; 0.535	0.465; 0.384
Jackknife resampling	25/25 families $p<0.05$	2/32 families $p<0.05$	None significant	0/25 families $p<0.05$	43/45 families p<0.05, Icteridae $p=0.066$, Fringillidae $p=0.053$	19/24 families $p<0.05$	None significant
EPP # of species	57	67	72	45	64	45	45
PhylANOVA *p*-value	0.001	0.02	0.566	0.045	0.329	0.714	0.052
Min; Max	0.004; 0.004	0.048; 0.036	0.334; 0.656	0.098; 0.052	0.22; 0.398	0.772; 0.518	0.064; 0.053
Jackknife resampling	24/24 families $p\leq0.015$	24/25 families $p<0.05$	None significant	6/17 families $p<0.05$	None significant	None significant	6/17 families $p<0.05$
Brownie *p*-value	0.1156	0.2764	0.3792	0.3650	0.1590	0.5785	0.3898
Rate higher in:	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Min; Max	0.121; 0.166	0.477; 0.187	0.532; 0.220	0.026; 0.408	0.142; 0.075	0.018; 0.056	0.291; 0.421
Jackknife resampling	None significant	None significant	None significant	None significant	1/25 families $p<0.05$	None significant	None significant

Table 2.3. Results of phylogenetically controlled analyses of mating behaviors and song characteristics. We tested whether song characteristics were significantly different between polygynous and monogamous species and between high and low EPP species (phylANOVA). We also tested whether song characteristics evolved faster in polygynous versus monogamous lineages or in high versus low EPP lineages. For each analysis, we assessed the robustness of our

findings by testing whether the minimum and maximum values of song characteristics from the literature yielded the same results (Min; Max). In addition, we removed each avian family from the analysis in turn and repeated the analyses (Jackknife resampling), summarized here but reported in full in **Online Supplementary Data 2-3**. If the removal of any family led to significant results at the 0.05 level, we note the number of families that met this threshold out of the total number tested.

Figures

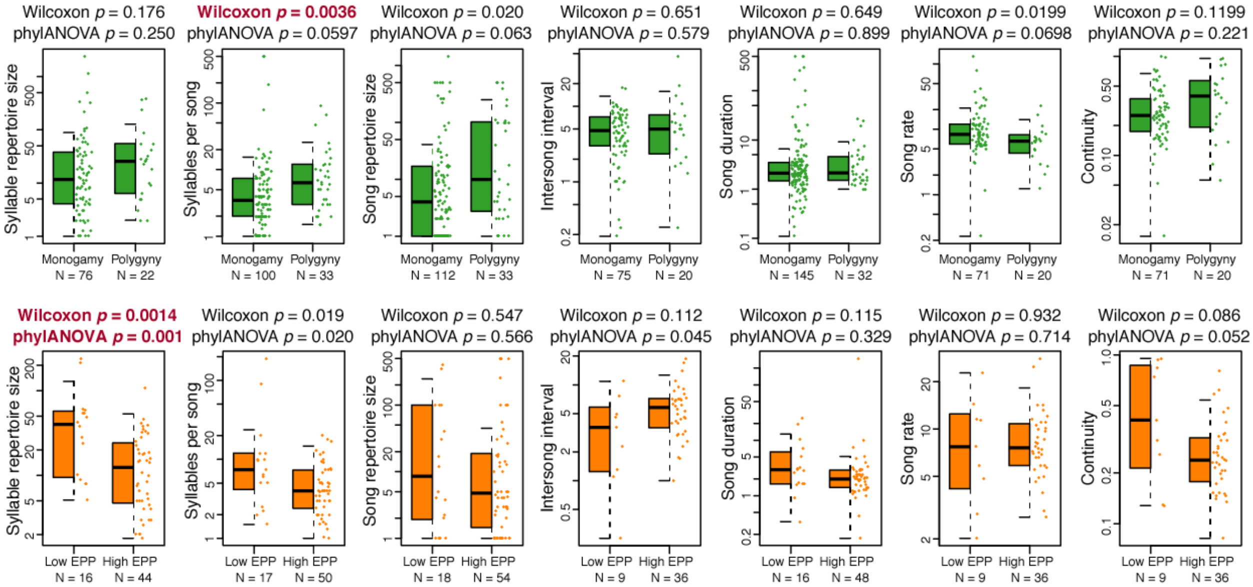


Figure 2.1: Differences in song characteristics for different rates of polygyny and extra-pair paternity. Each plot shows the distribution of song characteristics for species in our database. The top row (green) compares monogamous and polygynous species, with polygyny defined as >5% of males taking multiple mates when quantitative data is available, and the bottom row (orange) compares species with low and high rates of EPP, with a high rate of EPP defined as >10% of offspring in a population being the product of extra-pair fertilization (see methods for full classification criteria). Box plots indicate the median (black bar) interquartile range (IQR, box) and Q1−1.5*IQR, Q3+1.5*IQR (whiskers) of each distribution, and scatter plots of the data are shown to the right of each box plot. We compared each pair of distributions with phylogenetic ANOVA (phylANOVA) tests, which control for shared ancestry. *p*-values for these tests are shown above each box plot, with statistically significant results shown in red.

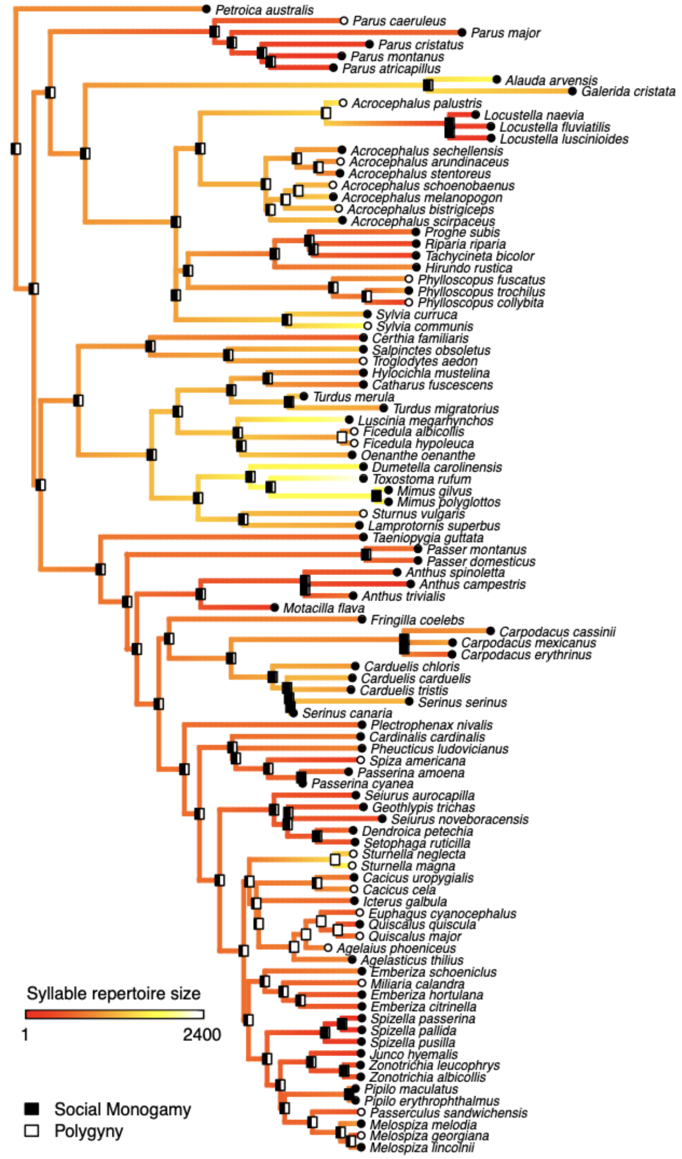


Figure 2.2: Ancestral character estimation of polygyny and syllable repertoire. At the tips of the tree, monogamy is indicated by black circles and polygyny is indicated by white circles. At the nodes of the tree, bars indicate the results of an ancestral character estimation algorithm, with the black fraction of the bar indicating the percent likelihood that the ancestor at that node was monogamous. The colors along the branches of the tree indicate the estimated ancestral syllable repertoire size. The syllable repertoire sizes ranged from 1 to 2400 in these species and were

$\log_{10}$ transformed for analysis. Asterisks indicate nodes that had less than 70% support across 1000 tree replicates; no node had less than 50% support on this tree. Monogamous and polygynous species did not have significantly different syllable repertoire sizes (PhylANOVA $p = 0.250$). Images representing taxa were used or modified from PhyloPic (http://phylopic.org). Several images are used under Creative Commons licenses, with changes made where indicated: Sturnidae (credit to Maxime Dahirel, http://creativecommons.org/licenses/by/3.0/), Estrildidae (credit to Jim Bendon for photography and T. Michael Keesey for vectorization, https://creativecommons.org/licenses/by-sa/3.0/), Fringillidae (credit to Francesco Veronesi (vectorized by T. Michael Keesey), https://creativecommons.org/licenses/by-nc-sa/3.0/), Emberizidae (credit to L. Shyamal, https://creativecommons.org/licenses/by-sa/3.0/; this image was also adapted for Parulidae), and Mimidae and Motacillidae (credit to Michelle Site, https://creativecommons.org/licenses/by-nc/3.0/, both adapted from the original image).

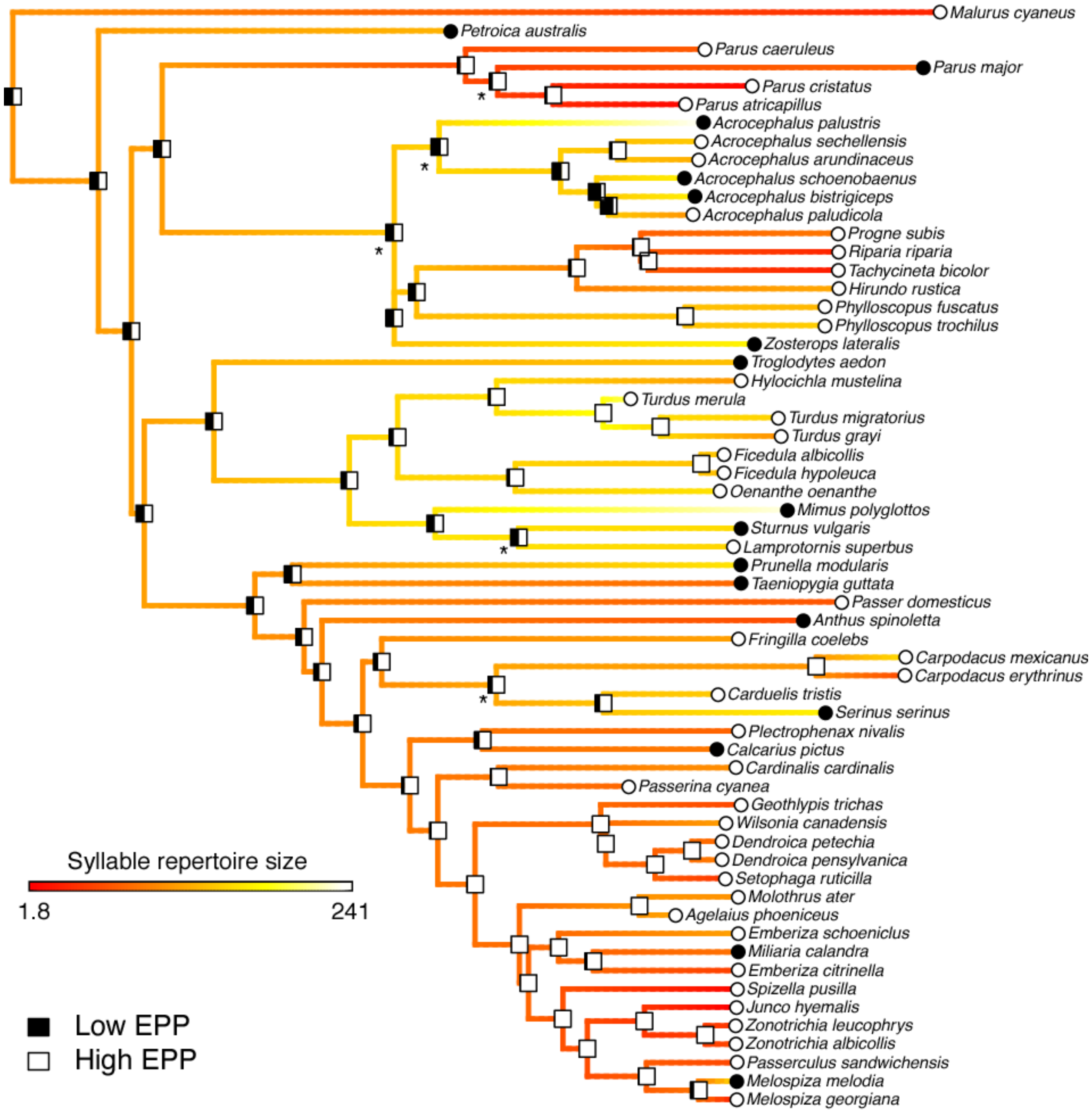


Figure 2.3: Ancestral character estimation of extra-pair paternity and syllable repertoire. At the tips of the tree, low (<10%) EPP is indicated by black circles and high EPP is indicated by white circles. At the nodes of the tree, bars indicate the results of an ancestral character estimation algorithm, with the black fraction of the bar indicating the percent likelihood that the ancestor at that node had low EPP. As in **Figure 2.2,** the colors from red to white along the

branches of the tree indicate the estimated ancestral syllable repertoire size. Asterisks indicate nodes that had less than 70% support across 1000 tree replicates; no node had less than 50% support on this tree. The syllable repertoire sizes ranged from 1.8 to 241 in these species and were $\log_{10}$ transformed for analysis. We found that species with high EPP had significantly smaller syllable repertoires than species with low EPP when controlling for phylogeny (PhylANOVA $p = 0.001$). Images representing taxa were used or modified from PhyloPic (http://phylopic.org). Several images are used under Creative Commons licenses, with changes made where indicated: Sturnidae (credit to Maxime Dahirel, http://creativecommons.org/licenses/by/3.0/), Estrildidae (credit to Jim Bendon for photography and T. Michael Keesey for vectorization, https://creativecommons.org/licenses/by-sa/3.0/), Fringillidae (credit to Francesco Veronesi (vectorized by T. Michael Keesey), https://creativecommons.org/licenses/by-nc-sa/3.0/), Emberizidae (credit to L. Shyamal, https://creativecommons.org/licenses/by-sa/3.0/; this image was also adapted for Parulidae), and Mimidae and Motacillidae (credit to Michelle Site, https://creativecommons.org/licenses/by-nc/3.0/, both adapted from the original image).

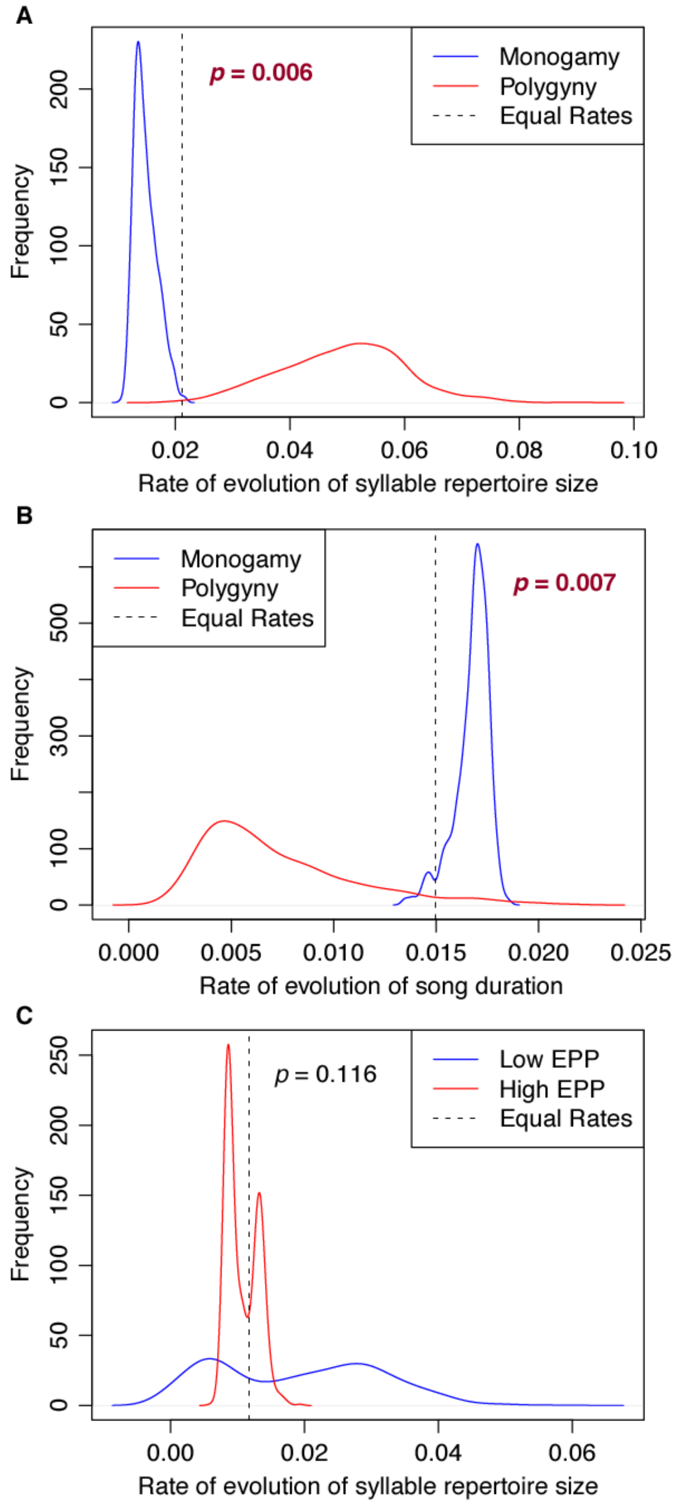


Figure 2.4: Analysis of the effect of mating system and EPP on the rate of syllable repertoire size evolution. (A) Mating system and syllable repertoire: We generated 1000

stochastic character maps—simulations of the evolutionary history of monogamy and polygyny mapped onto the phylogeny—and then we tested whether syllable repertoire size evolved at different rates in monogamous versus polygynous branches of the tree. From all runs that converged out of 1000 total runs of the Brownie algorithm, we plot the distribution of the rate of evolution of syllable repertoire size in monogamous lineages (blue) and the rate of evolution of syllable repertoire size in polygynous lineages (red). Distributions are kernel density plots generated using the R function density with a Gaussian smoothing kernel. The dashed line indicates the rate of evolution estimated when syllable repertoire size is assumed to evolve at the same rate in monogamous and polygynous periods of evolutionary history. We find that syllable repertoire size evolves significantly faster in polygynous branches (Brownie likelihood-ratio test $p = 0.006$). **(B)** Mating system and song duration: The rate of evolution of song duration also differed, evolving significantly faster in monogamous lineages. **(C)** EPP and syllable repertoire: We performed a similar analysis with high and low rates of EPP mapped onto the phylogeny, and tested whether syllable repertoire size evolved at different rates in periods of high (red) versus low (blue) rates of EPP. We do not reject the null hypothesis that syllable repertoire size evolved at the same rate in high-EPP and low-EPP branches of the tree.

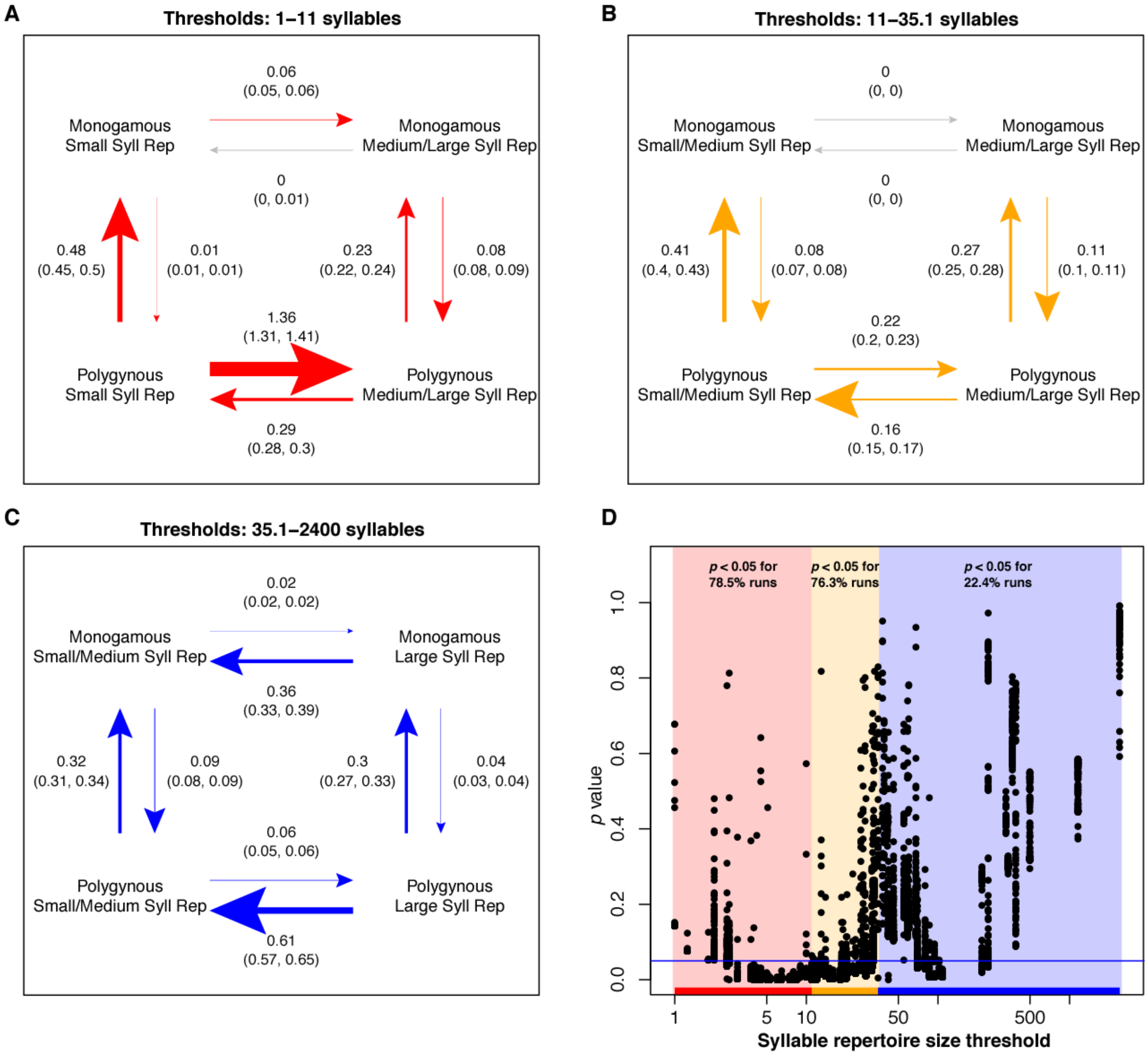


Figure 2.5: Detecting correlated evolution of mating systems and syllable repertoire size. We tested the correlated evolution of syllable repertoire size and polygyny using BayesTraits, with syllable repertoire size made binary based on a threshold delineating smaller versus larger syllable repertoires. Each value of syllable repertoire wais used as the threshold for 100 runs of BayesTraits per threshold. For each plot, there are eight possible transitions between the four trait pairs, shown by arrows. **(A–C)** We generated transition plots by calculating the mean rate and 95% confidence interval (in parentheses) for each of the eight transitions. **(A)** When the

threshold between smaller and larger syllable repertoires is in the lowest third of observed values (24 unique thresholds ≥1 and <11 syllables, red arrows), polygyny is unstable with low syllable repertoires. **(B)** When the threshold between smaller and larger syllable repertoires is in the middle third of observed values (24 unique thresholds ≥11 and <35.1 syllables, yellow arrows), the transitions between smaller and larger syllable repertoires do not appear to be elevated in either monogamy or polygyny. **(C)** When the threshold between smaller and larger syllable repertoires is in the highest third of observed values (24 unique thresholds ≥35.1 and <2400 syllables, blue arrows), the combination of large syllable repertoires and polygyny is unstable, with the highest transition rate pointing to a decrease in repertoire size in the presence of polygyny. These results were robust to jackknife resampling across families (**Online Supplementary Fig. 30**). **(D)** For each run of BayesTraits, we performed a likelihood-ratio test to assess whether the model of correlated evolution between mating system and syllable repertoire size was a significantly better fit to the data than the independent evolution model; *p*-values are plotted against the syllable repertoire size values defining the threshold.

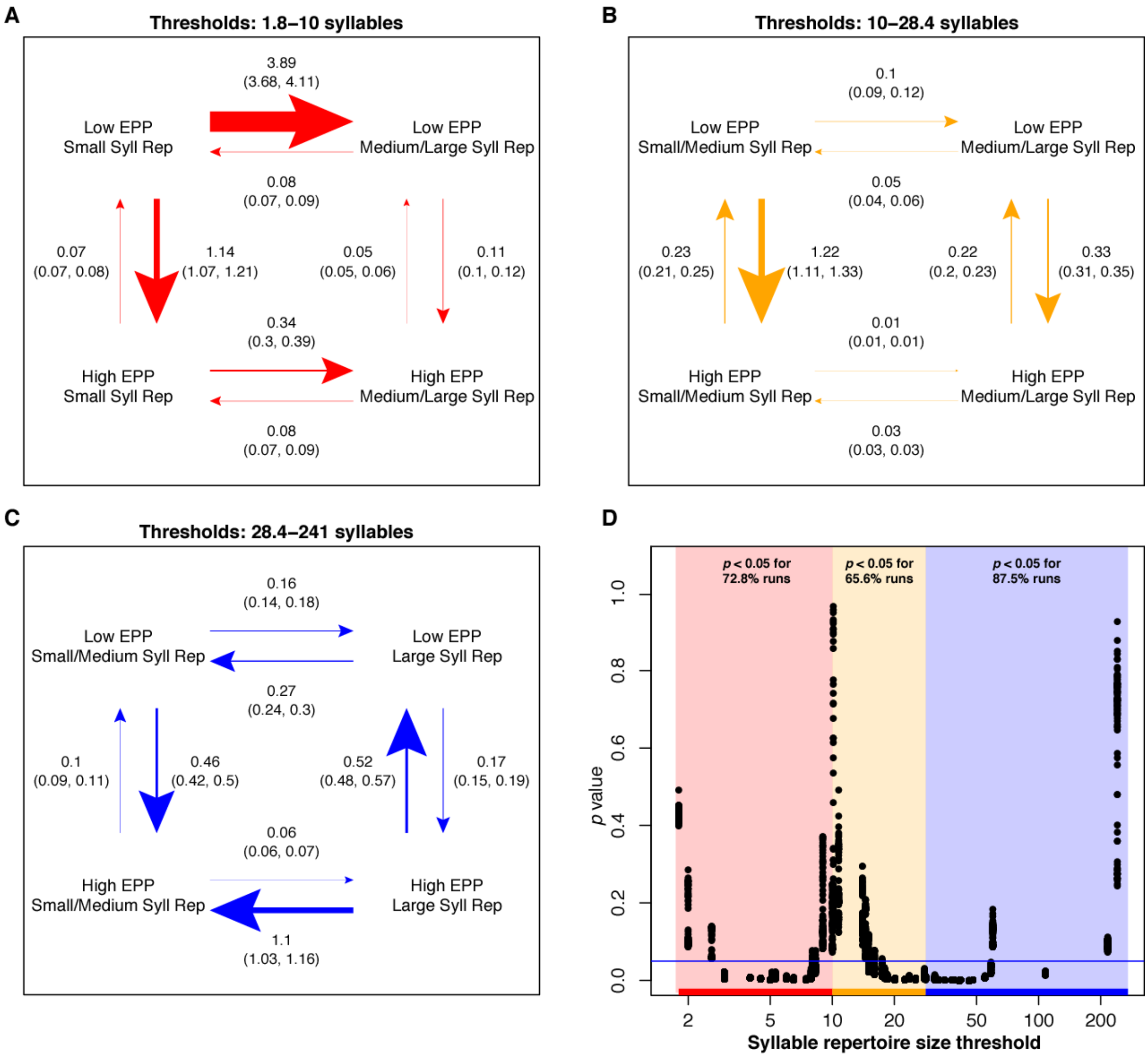


Figure 2.6: **Detecting correlated evolution of extra-pair paternity and syllable repertoire size.** We tested for correlated evolution of syllable repertoire size and EPP using the same procedure as in **Figure 2.5**. **(A)** When the threshold between smaller and larger syllable repertoires is in the lowest third of observed values (16 unique thresholds between 1.8 and 10 syllables, red arrows), low rates of EPP with small syllable repertoires are unstable, and we observe elevated transition rates either toward larger repertoires or toward higher rates of EPP. **(B)** When the threshold between smaller and larger syllable repertoires is in the middle third of

observed values (16 unique thresholds ≥10 and <28.4 syllables, yellow arrows), low rates of EPP are again unstable with small syllable repertoires, and evolutionary transitions toward higher rates of EPP are elevated. **(C)** When the threshold between smaller and larger syllable repertoires is in the highest third of observed values (16 unique thresholds ≥28.4 and <241 syllables, blue arrows), the combination of large syllable repertoires and high rates of EPP is unstable, and we observe elevated transition rates either toward smaller repertoires or toward lower rates of EPP. We averaged rate values from all runs, regardless of significance. These results were robust to jackknife resampling across families. In the middle segment, only removing Zosteropidae qualitatively altered the dominant rates of transition such that there was accelerated evolution from Low to High EPP regardless of syllable repertoire (**Online Supplementary Fig. 31**). **(D)** For each run of BayesTraits, we performed a likelihood-ratio test to assess whether the model of correlated evolution between EPP and syllable repertoire size was a significantly better fit to the data than the independent evolution model; *p*-values are plotted against the syllable repertoire size values defining the threshold.

References

Baptista, Luis F., and Martin L. Morton. 1988. “Song Learning in Montane White-Crowned Sparrows: From Whom and When.” *Animal Behaviour* 36(6):1753–64.

Blomberg, Simon P., Theodore Garland Jr, and Anthony R. Ives. 2003. “Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile.” *Evolution; International Journal of Organic Evolution* 57(4):717–45.

Bollback, Jonathan P. 2006. “SIMMAP: Stochastic Character Mapping of Discrete Traits on Phylogenies.” *BMC Bioinformatics* 7:88.

Byers, Bruce E. 2006. “Extrapair Paternity in Chestnut-Sided Warblers Is Correlated with Consistent Vocal Performance.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 18(1):130–36.

Byers, Bruce E., and Donald E. Kroodsma. 2009. “Female Mate Choice and Songbird Song Repertoires.” *Animal Behaviour* 77(1):13–22.

Byers, Bruce E., Herman L. Mays, Ian R. K. Stewart, David F. Westneat, and M. T. Murphy. 2004. “EXTRAPAIR PATERNITY INCREASES VARIABILITY IN MALE REPRODUCTIVE SUCCESS IN THE CHESTNUT-SIDED WARBLER (DENDROICA PENSYLVANICA), A SOCIALLY MONOGAMOUS SONGBIRD.” *The Auk* 121(3):788–95.

Catchpole, Clive K. 1980. “Sexual Selection and the Evolution of Complex Songs Among European Warblers of the Genus Acrocephalus.” *Behaviour* 74(1):149–65.

Catchpole, Clive K., and Peter K. McGregor. 1985. “Sexual Selection, Song Complexity and Plumage Dimorphism in European Buntings of the Genus Emberiza.” *Animal Behaviour* 33(4):1378–80.

Clements, James F. 2007. *The Clements Checklist of Birds of the World*. Comstock Publishing Associates.

Cordero, Pedro J. 1998. “Extra-Pair Paternity in Birds: ‘Good-Genes’ and Something Else.” *Trends in Ecology & Evolution* 13(7):280.

Cornell University Cornell Lab of Ornithology. 2009. *Macaulay Library: Archive of Animal Sounds and Video*.

Emlen, S., and L. Oring. 1977. “Ecology, Sexual Selection, and the Evolution of Mating Systems.” *Science* 197(4300):215–23.

Feisenstein, Joseph. 1989. *PHYLIP: Phylogeny Inference Package Version 3.2 Manual*.

Ferer, Elise. n.d. “Birds of North America Online.” *CC Advisor*.

Forstmeier, Wolfgang, Bart Kempenaers, Axel Meyer, and Bernd Leisler. 2002. “A Novel Song Parameter Correlates with Extra-Pair Paternity and Reflects Male Longevity.” *Proceedings. Biological Sciences / The Royal Society* 269(1499):1479–85.

Garamszegi, L. Z., and A. P. Møller. 2004. “Extrapair Paternity and the Evolution of Bird Song.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 15(3):508–19.

Garland, Theodore, Allan W. Dickerman, Christine M. Janis, and Jason A. Jones. 1993. “Phylogenetic Analysis of Covariance by Computer Simulation.” *Systematic Biology* 42(3):265.

Griffith, Simon C., Ian P. F. Owens, and Katherine A. Thuman. 2002. “Extra Pair Paternity in Birds: A Review of Interspecific Variation and Adaptive Function.” *Molecular Ecology* 11(11):2195–2212.

Hackett, Shannon J., Rebecca T. Kimball, Sushma Reddy, Rauri C. K. Bowie, Edward L. Braun, Michael J. Braun, Jena L. Chojnowski, W. Andrew Cox, Kin-Lan Han, John Harshman, Christopher J. Huddleston, Ben D. Marks, Kathleen J. Miglia, William S. Moore, Frederick H. Sheldon, David W. Steadman, Christopher C. Witt, and Tamaki Yuri. 2008. “A Phylogenomic Study of Birds Reveals Their Evolutionary History.” *Science* 320(5884):1763–68.

Hadfield, Jarrod D. 2010. “MCMC Methods for Multi-Response Generalized Linear Mixed Models: TheMCMCglmmRPackage.” *Journal of Statistical Software* 33(2). doi: 10.18637/jss.v033.i02.

Hasselquist, Dennis, Staffan Bensch, and Torbjörn von Schantz. 1996. “Correlation between Male Song Repertoire, Extra-Pair Paternity and Offspring Survival in the Great Reed Warbler.” *Nature* 381(6579):229–32.

Hasselquist, Dennis, and Paul W. Sherman. 2001. “Social Mating Systems and Extrapair Fertilizations in Passerine Birds.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 12(4):457–66.

Hegyi, Gergely, Eszter Szöllosi, Susanne Jenni-Eiermann, János Török, Marcel Eens, and László Zsolt Garamszegi. 2010. “Nutritional Correlates and Mate Acquisition Role of Multiple Sexual Traits in Male Collared Flycatchers.” *Die Naturwissenschaften* 97(6):567–76.

Hill, Christopher E., Çağlar Akçay, S. Elizabeth Campbell, and Michael D. Beecher. 2010. “Extrapair Paternity, Song, and Genetic Quality in Song Sparrows.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 22(1):73–81.

Hill, Samuel D., Christophe Amiot, Michael G. Anderson, and Weihong Ji. 2017. “It’s Complicated: The Association between Songbird Extrapair Paternity and within-Song Complexity.” *Animal Behaviour* 130:187–97.

del Hoyo, J., A. Elliott, J. Sargatal, D. A. Christie, and E. (eds). de Juana. n.d. “Handbook of the

Birds of the World Alive.” *Lynx Edicions, Barcelona*. Retrieved 2017 (http://www.hbw.com/).

Irwin, Rebecca E. 1990. “Directional Sexual Selection Cannot Explain Variation in Song Repertoire Size in the New World Blackbirds (Icterinae).” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 85(3):212–24.

Jetz, Walter, Gavin H. Thomas, Jeffrey B. Joy, David W. Redding, Klaas Hartmann, and Arne O. Mooers. 2014. “Global Distribution and Conservation of Evolutionary Distinctness in Birds.” *Current Biology: CB* 24(9):919–30.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. “The Global Diversity of Birds in Space and Time.” *Nature* 491(7424):444–48.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, D. Redding, and A. O. Mooers. n.d. “A Global Phylogeny of Birds.” Retrieved September 2017 (birdtree.org).

Kempenaers, Bart, Geert R. Verheyen, Marleen Van den Broeck, Terry Burke, Christine Van Broeckhoven, and André Dhondt. 1992. “Extra-Pair Paternity Results from Female Preference for High-Quality Males in the Blue Tit.” *Nature* 357(6378):494–96.

Kroodsma, Donald E. 1977. “Correlates of Song Organization Among North American Wrens.” *The American Naturalist* 111(981):995–1008.

Lachlan, R. F., and P. J. B. Slater. 2003. “Song Learning by Chaffinches: How Accurate, and from Where?” *Animal Behaviour* 65(5):957–69.

Lack, David Lambert. 1968. *Ecological Adaptations for Breeding in Birds*.

Lande, R. 1981. “Models of Speciation by Sexual Selection on Polygenic Traits.” *Proceedings of the National Academy of Sciences of the United States of America* 78(6):3721–25.

Lebigre, Christophe, Peter Arcese, Rebecca J. Sardell, Lukas F. Keller, and Jane M. Reid. 2012. “Extra-Pair Paternity and the Variance in Male Fitness in Song Sparrows (Melospiza Melodia).” *Evolution; International Journal of Organic Evolution* 66(10):3111–29.

Leisler, B., and C. K. Catchpole. 1992. “The Evolution of Polygamy in European Reed Warblers of the Genus Acrocephalus: A Comparative Approach.” *Ethology Ecology & Evolution* 4(3):225–43.

Lenski, R. E., and M. Travisano. 1994. “Dynamics of Adaptation and Diversification: A 10,000-Generation Experiment with Bacterial Populations.” *Proceedings of the National Academy of Sciences of the United States of America* 91(15):6808–14.

Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. 2015. “Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking.” *Nature* 519(7542):181–86.

Liu, Wan-Chun, and Donald E. Kroodsma. 2006. “SONG LEARNING BY CHIPPING SPARROWS: WHEN, WHERE, AND FROM WHOM.” *The Condor* 108(3):509.

Macdougall-Shackleton, Scott A. 1997. “Sexual Selection and the Evolution of Song Repertoires.” Pp. 81–124 in *Current Ornithology*.

Marshall, Rupert C., Katherine L. Buchanan, and Clive K. Catchpole. 2007. “Song and Female Choice for Extrapair Copulations in the Sedge Warbler, Acrocephalus Schoenobaenus.” *Animal Behaviour* 73(4):629–35.

Møller, Anders Pape. 1986. “Mating Systems among European Passerines: A Review.” *The Ibis* 128(2):234–50.

Moore, Jordan M., Tamás Székely, József Büki, and Timothy J. Devoogd. 2011. “Motor Pathway Convergence Predicts Syllable Repertoire Size in Oscine Birds.” *Proceedings of the National Academy of Sciences of the United States of America* 108(39):16440–45.

Nowicki, Stephen, and William A. Searcy. 2004. “Song Function and the Evolution of Female Preferences: Why Birds Sing, Why Brains Matter.” *Annals of the New York Academy of Sciences* 1016:704–23.

O’Meara, Brian C., Cécile Ané, Michael J. Sanderson, and Peter C. Wainwright. 2006. “Testing for Different Rates of Continuous Trait Evolution Using Likelihood.” *Evolution; International Journal of Organic Evolution* 60(5):922–33.

Pagel, M. 1994. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” *Proceedings of the Royal Society B: Biological Sciences* 255(1342):37–45.

Pagel, Mark. 1999. “Inferring the Historical Patterns of Biological Evolution.” *Nature* 401(6756):877–84.

Pagel, Mark, and Andrew Meade. 2006. “Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible‐Jump Markov Chain Monte Carlo.” *The American Naturalist* 167(6):808–25.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20(2):289–90.

Payne, Robert B. 1984. “Sexual Selection, Lek and Arena Behavior, and Sexual Size Dimorphism in Birds.” *Ornithological Monographs* (33):iii – 52.

Petrie, Marion, and Bart Kempenaers. 1998. “Extra-Pair Paternity in Birds: Explaining Variation between Species and Populations.” *Trends in Ecology & Evolution* 13(2):52–58.

Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkar. 2014. “Linear and Nonlinear Mixed Effects Models.” *R Package Version 3.*

Poesel, Angelika, Hansjoerg P. Kunc, Katharina Foerster, Arild Johnsen, and Bart Kempenaers. 2006. "Early Birds Are Sexy: Male Age, Dawn Song and Extrapair Paternity in Blue Tits, Cyanistes (formerly Parus) Caeruleus." *Animal Behaviour* 72(3):531–38.

Price, J. Jordan, and Simon C. Griffith. 2017. "Open Cup Nests Evolved from Roofed Nests in the Early Passerines." *Proceedings. Biological Sciences / The Royal Society* 284(1848). doi: 10.1098/rspb.2016.2708.

Quintero, Ignacio, and Walter Jetz. 2018. "Global Elevational Diversity and Diversification of Birds." *Nature* 555(7695):246–50.

Read, A. F., and D. M. Weary. 1992. "The Evolution of Bird Song: Comparative Analyses." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 338(1284):165–87.

Revell, Liam J. 2011. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)." *Methods in Ecology and Evolution / British Ecological Society* 3(2):217–23.

Rhoads, Samuel N. 1890. "Probable Causes of Polygamy Among Birds." *The American Naturalist* 24(287):1024–36.

Sayol, Ferran, Joan Maspons, Oriol Lapiedra, Andrew N. Iwaniuk, Tamás Székely, and Daniel Sol. 2016. "Environmental Variation and the Evolution of Large Brains in Birds." *Nature Communications* 7:13971.

Searcy, William A., and Ken Yasukawa. 1995. *Polygyny and Sexual Selection in Red-Winged Blackbirds*.

Shutler, Dave, and Patrick J. Weatherhead. 1990. "TARGETS OF SEXUAL SELECTION: SONG AND PLUMAGE OF WOOD WARBLERS." *Evolution; International Journal of Organic Evolution* 44(8):1967–77.

Soma, Masayo, and László Zsolt Garamszegi. 2011. "Rethinking Birdsong Evolution: Meta-Analysis of the Relationship between Song Complexity and Reproductive Success." *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 22(2):363–71.

Sullivan, Brian L., Jocelyn L. Aycrigg, Jessie H. Barry, Rick E. Bonney, Nicholas Bruns, Caren B. Cooper, Theo Damoulas, André A. Dhondt, Tom Dietterich, Andrew Farnsworth, Daniel Fink, John W. Fitzpatrick, Thomas Fredericks, Jeff Gerbracht, Carla Gomes, Wesley M. Hochachka, Marshall J. Iliff, Carl Lagoze, Frank A. La Sorte, Matthew Merrifield, Will Morris, Tina B. Phillips, Mark Reynolds, Amanda D. Rodewald, Kenneth V. Rosenberg, Nancy M. Trautmann, Andrea Wiggins, David W. Winkler, Weng-Keen Wong, Christopher L. Wood, Jun Yu, and Steve Kelling. 2014. "The eBird Enterprise: An Integrated Approach to Development and Application of Citizen Science." *Biological Conservation* 169:31–40.

Suter, Stefan M., David Ermacora, Nadia Rieille, and Dietrich R. Meyer. 2009. "A Distinct Reed

Bunting Dawn Song and Its Relation to Extrapair Paternity." *Animal Behaviour* 77(2):473–80.

Vedder, Oscar, Jan Komdeur, Marco van der Velde, Elske Schut, and Michael J. L. Magrath. 2011. "Polygyny and Extra-Pair Paternity Enhance the Opportunity for Sexual Selection in Blue Tits." *Behavioral Ecology and Sociobiology* 65(4):741–52.

Verner, Jared, and Mary F. Willson. 1969. "Mating Systems, Sexual Dimorphism, and the Role of Male North American Passerine Birds in the Nesting Cycle." *Ornithological Monographs* (9):1–76.

Wells, Sarah J., Rebecca J. Safran, and James Dale. 2016. "Piecing Together Female Extra-Pair Mate Choice: Females Really Do Prefer More Ornamented Males." *Molecular Ecology* 25(15):3521–24.

Westneat, David F., and Ian R. K. Stewart. 2003. "Extra-Pair Paternity in Birds: Causes, Correlates, and Conflict." *Annual Review of Ecology, Evolution, and Systematics* 34(1):365–96.

Westneat, D. F., Sherman, P. W., & Morton, M. L. 1990. "The Ecology and Evolution of Extra-Pair Copulations in Birds." *Current Ornithology* 7:331–69.

Yasui, Yukio, and Jin Yoshimura. 2018. "Bet-Hedging against Male-Caused Reproductive Failures May Explain Ubiquitous Cuckoldry in Female Birds." *Journal of Theoretical Biology* 437:214–21.

CHAPTER 3.

CORRELATED EVOLUTION BETWEEN REPERTOIRE SIZE AND SONG PLASTICITY PREDICTS THAT SEXUAL SELECTION ON SONG PROMOTES OPEN-ENDED LEARNING

Cristina Robinson[1,*], Kate T. Snyder[1,*], Nicole Creanza[1,✉]
[1] Department of Biological Sciences, Vanderbilt University, Nashville, TN USA 37204
* Contributed equally
✉ nicole.creanza@vanderbilt.edu

Adapted from Cristina M Robinson, Kate T Snyder, Nicole Creanza (2019) Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning eLife 8:e44454 https://doi.org/10.7554/eLife.44454

Supplementary Materials (Labelled "figure supplement", "S", or "source data") can be found: https://doi.org/10.7554/eLife.44454

Introduction

Song is a learned behavior with a complex evolutionary history in the oscine songbirds. Birds' songs have multiple functions, including species recognition, territory defense, and mate attraction (Catchpole and Slater 2003). All species studied in this expansive clade have certain life stages during which they are more likely to learn and acquire songs, termed sensitive periods (Brenowitz and Beecher 2005; Marler 1990; Marler and Peters 1987; Rauschecker and Marler 1987). In some species, learning is restricted to a short sensitive period early in life, also called a 'critical period' (for example, ~day 25–90 in zebra finches), after which no new song elements are acquired (Böhner 1990; Immelman 1969; Nottebohm 1984). Other species appear to delay song crystallization until some time in adulthood (Dowsett-Lemaire 1979; Kipper and Kiefer 2010; Martens and Kessler 2000); for example, chipping sparrows have been observed to have a second sensitive period immediately after their first migration, following which the song crystallizes (Liu and Kroodsma 2006; Liu and Nottebohm 2007). Still other species can continue to acquire new syllables or songs throughout their lives (Adret-Hausberger, Güttinger, and Merkel 2010; Espmark and Lampe 1993; Gil, Cobb, and Slater 2001; Hausberger, Jenkins, and Keene 1991; Mountjoy and Lemon 1995; Price and Yuan 2011). Typically, this spectrum of variation in the timing of the sensitive period is simplified into a dichotomy of 'open-ended learning' and 'closed-ended learning.' While these temporal differences in the song-learning window have been studied for decades, it is unknown how they interact with the evolution of song. Previous hypotheses have suggested that seasonal factors, such as environmental variation and breeding season length, play a role in shaping adult song learning (Nottebohm, Nottebohm, and Crane 1986; Smith et al. 1997; Tramontin et al. 2001). However, evidence from a small-scale comparative analysis suggests that a longer learning window in a species may be associated with larger average syllable repertoire sizes (Creanza, Fogarty, and Feldman 2016).

Birdsong is composed of both culturally and genetically inherited features, any of which may be subject to evolutionary pressures. Two key modes of selection on song might act in conjunction: female choice can favor certain song characteristics, such as superior repertoire size, learning quality, or song performance (Beecher and Brenowitz 2005; Gil and Gahr 2002; Searcy and Andersson 1986; Searcy and Marler 1984), while the inherent metabolic cost of neuroplasticity should theoretically favor a shorter song-learning window and thus reduce the opportunity for a bird to alter its song in adulthood (Garamszegi and Eens 2004; Nottebohm et al. 1986; Tramontin and Brenowitz 2000). Therefore, while learning in adulthood or elongated sensitive periods have not been shown to be directly under positive selection (sexual or natural) or to play an explicit role in female preferences, sexual selection acting on certain song features could indirectly favor longer or shorter song-learning windows. However, this theorized connection between sexual selection and adult learning hinges on establishing the evolutionary relationship between song as the target of sexual selection and the neurobehavioral phenotype of song learning, which has not yet been done.

Furthermore, sexual selection is hypothesized to be amplified in polygynous social mating systems and in species with high rates of extra-pair paternity (EPP). A recent large-scale study found that polygyny drives faster, but non-directional, evolution of syllable repertoire size, and that syllable repertoire size is negatively correlated with the rate of EPP (Snyder and Creanza 2019). Because of the evolutionary links between these mating strategies and song features, higher rates of EPP and polygyny could potentially be associated with longer learning windows. We therefore investigate the evolutionary relationship between the critical learning period and mating strategies.

Here, we take a comparative, computational approach to the evolutionary history of open- and closed-ended song learning. We mined the literature for longitudinal field and laboratory observations to classify species as exhibiting 'adult song stability' or 'adult song plasticity'. This classification is a quantifiable proxy for closed- and open-ended learning, as the true length of the song-learning window is difficult to assess directly in nature; ultimately, we obtained data for the classification of 67 species. For these species, we also compiled a database of seven species-level song characteristics that can represent either song complexity (syllable repertoire, syllables per song, and song repertoire) or song performance (song duration, inter-song interval, song rate, and song continuity). We then performed phylogenetically controlled analyses to evaluate the evolution of song and mating strategies alongside the relative plasticity of song over time. We find that adult song plasticity has evolved numerous times in bird species. Further, we find evidence of correlated evolution between the adult song stability and plasticity and social mating system, with shifts in social mating system occurring more rapidly in lineages with adult song plasticity. In addition, we find a significant evolutionary pattern: species with plastic songs generally have larger repertoires than species with stable songs. Specifically, the evolution of larger syllable and song repertoires appears to drive an evolutionary transition toward open-ended learning.

Results

In this study, we analyzed the length of the song-learning window using three classification schemes. First, we classified 67 species as having either adult song plasticity (i.e. those that change their syllable repertoire after their first breeding season ends) or adult song stability. Second, when possible, we reclassified the species into three categories: 1) those that stop changing their song before their first breeding season, 2) those that modify their songs during their first breeding season but not after, and 3) those that modify their songs after their first breeding season has ended. This three-state classification was possible for 59 out of the 67 species. Third, for the same 59 species, we used reported estimates of the ages at which song stabilizes to create a continuous measure of the song-learning window. There were exceedingly few studies that examined song changes after the second breeding season, so our continuous metric ranges from zero to two years. We then examined the evolutionary patterns of adult song stability and plasticity, as well as their interactions with species-level song characteristics and mating behaviors.

How has adult song plasticity evolved across clades?

We were first interested in examining the rate of evolution of adult song stability versus adult song plasticity and when and where evolutionary transitions in these traits occurred on a phylogenetic tree (Jetz et al. 2012) using ancestral state reconstruction. As with any reconstruction of evolutionary history, these simulations cannot exactly predict the ancestral states but aim to approximate them. Furthermore, we note that only a subset of oscine families could be included in our analysis, and most of the early branching lineages that would be required to assess the ancestral state for all oscine species could not be included in our dataset. Ultimately, we could not make a conclusion about ancestral state for the last common ancestor for the species included in this study, but our results hint that there might have been several early transitions in this trait, leading to clades that predominantly have adult song stability or plasticity, coupled with a number of more recent transitions (see pie charts in **Figure 3.1A** for the predicted likelihood of each state at each node). We found that a model allowing the transition rate from song stability to plasticity to be different from the transition rate from plasticity to stability (all-rates-different model [ARD]) did not fit the data significantly better than a simpler model allowing for only one rate of transition back and forth between song stability and plasticity (equal rates model [ER]) ($\text{LogLikelihood}_{\text{ER}}$=-38.22, $\text{LogLikelihood}_{\text{ARD}}$=-38.21, p=0.87). At least 14 transitions were required to explain the current binary song-stability states of our subset of bird species. Explaining the distribution of song plasticity in our subset of species most parsimoniously requires at least 9 transitions to adult song plasticity if the last common ancestor was song-stable and 7 transitions to song stability if the common ancestor was song-plastic. (**Figure 3.1 - figure supplement 1**).

Which song traits differ between species with stable versus plastic songs?

We next tested whether song characteristics were affected by the length of the song-learning window on an evolutionary scale. Intuitively, it makes sense that a species that has a longer time-window to learn might be able to accumulate a larger repertoire. Indeed, this relationship is consistent with the pattern of song stability and repertoire size in several clades, such as the *Phylloscopus* species (**Figure 3.2**). However, many individual species do not follow this prediction: for example, *Acrocephalus palustris* appears to learn a large repertoire in a single year (Dowsett-Lemaire 1979), and *Philesturnus rufusater* modifies its song for multiple years but maintains a small repertoire (Jenkins 1978). Further, numerous species with adult song plasticity do not significantly increase their repertoire sizes over time (Eriksen, Slagsvold, and Lampe 2011; Galeotti et al. 2001; Garamszegi et al. 2005; Nicholson et al. 2007). Thus, an evolutionary link between adult song plasticity and larger repertoire sizes cannot be assumed. Using a phylogenetically controlled ANOVA (Garland et al. 1993; Revell 2011), we found that species with adult song plasticity did possess significantly larger syllable repertoires than species with adult song stability (**Figures 3.1A, 3.3A, Table 3.1**). This concurs with a previous analysis using a smaller dataset (Creanza et al. 2016). Similarly, we found that song-plastic species had significantly larger song repertoires than song-stable species (**Figure 3.3B, Figure 3.1 - figure supplement 2, and Table 3.1**).

There were no significant differences between song-plastic and song-stable species for the other song characteristics that we tested: syllables per song, inter-song interval, song duration, song rate (calculated as 60/(interval + duration)), or song continuity (calculated as duration/(duration + interval)) (**Table 3.1, Figure 3.1 - figure supplements 3-7**). When we used the classification scheme with three states, we could only test for differences in syllable repertoire, song repertoire, and syllables per song between groups, as there were very few early song-stable species for which we had data on the other song traits. We found no significant differences between early song-stable and delayed song-stable species for any tested traits, but both of these groups had significantly smaller syllable and song repertoires compared to song-plastic species (**Figures 1B, 3C, Tables 3.2 and 3.3**). When performing a phylogenetic generalized least squares (PGLS) analysis using continuous estimates of the duration of song plasticity, we found similar results; both syllable repertoire and song repertoire were correlated with duration of song plasticity, such that repertoire size increased with the song-plasticity duration (**Figure 3.3D, Table 3.4**).

Do song characteristics evolve at different rates in song-stable or song-plastic species?

Our result that species with adult song plasticity had significantly larger syllable and song repertoires raised the question of whether song stability versus plasticity also affected the rate of evolution for any of the song characteristics. To examine this possibility, we used the Brownie algorithm (O'Meara et al. 2006), which tests whether a model with two rates of evolution for each song characteristic—one rate for ancestral periods of song stability and another rate for song plasticity—fits the data significantly better than a model that allows for only one rate of evolution of each song characteristic regardless of the ancestral states of song stability. Each calculation of the two-rate model is based on one stochastic projection of the ancestral traits across the phylogenetic tree, so we generated 1300 different stochastic simulation maps to use with Brownie. We plotted the distribution of potential rates (**Figures 3.4-5**) and compared the average log likelihood of the two-rate models to the log likelihood of the one-rate model (**Table 3.5**).

We found that allowing for two different rates of song trait evolution depending on song stability or plasticity did not lead to a significantly better fit model than using one Brownian-motion rate for either syllable repertoire size or song repertoire size, even though syllable repertoires and song repertoires were both significantly larger in species with adult song plasticity (**Figure 3.4A and B and Table 3.5).** In contrast, the two-rate model led to a significantly better fit for syllables per song, song rate, inter-song interval, and song duration (**Figure 3.4C, 5 and Table 3.5**), indicating that evolution of these song characteristics was faster in song-plastic lineages (**Figure 3.4C and 5, red traces**).

We repeated this analysis with the three-state categorization for syllable repertoire, song repertoire, and syllables per song; for other song characteristics, we did not have enough species in the early song stability group. We found that the three-rate model was significantly better than the one-rate model for syllables per song and song repertoire, but not for syllable repertoire (**Figure 3.4D-F and Table 3.6**). However, the three-rate model was only significantly better than the two-rate model for song repertoire (**Table 3.7**). Thus, the two-rate model sufficiently approximates the evolution of syllables per song. We noticed that for both song repertoire and syllable repertoire, the rate of evolution in delayed song-stable lineages (**purple traces in Figure 3.4 D, E**) was very similar to the rate in song-plastic lineages (**corresponding red traces**). We tested one more set of models where we combined delayed song-stable species with song-plastic species to create a “longer learning” group, while early song-stable species were assigned to a “shorter learning” group. For this comparison of shorter versus longer learning, the two-rate model was significantly better than the one-rate model for song repertoire and trending in that direction for syllable repertoire (**Figure 3.4 G, H and Table 3.8**). The three-rate model was not significantly better than the longer/shorter-learning two-rate model for either syllable or song repertoire (**Table 3.9**). Taken together with our phylANOVA results, this pattern suggests that species with early song stability evolve their song repertoires and potentially their syllable repertoires at a slower rate than delayed song-stable and song-plastic species; however, only song-plastic species directionally evolve towards larger song and syllable repertoires.

Is song stability or plasticity influenced by the evolution of song traits and mating strategies, or vice versa?

While the Brownie algorithm tested whether adult song plasticity affected the rate of evolution for song characteristics, it did not address whether the opposite might be true. We used BayesTraits (Pagel 1994; Pagel and Meade 2006) to test whether the rate and order of evolutionary transitions in one trait is dependent on the state of another trait. Because the song features were continuous variables, we binarized them by setting a series of threshold values to delineate “low” and “high” categories, using each observed song feature value as a threshold in turn. We then tested whether there was correlated evolution between the binary classifications of adult song plasticity versus stability and each of the seven song characteristics.

In the lowest third of syllable repertoire thresholds, adult song plasticity with small syllable repertoires was an evolutionarily unstable state, with rapid transitions primarily towards a song-stable state and secondarily towards larger repertoires (82% of runs significant in this range, **Figure 3.6**). In the middle third of syllable repertoire thresholds, song-stability with smaller syllable repertoires is an evolutionarily stable attractor state, with high rates of transition observed from large to small syllable repertoires in song-stable species and from plasticity to stability with a small syllable repertoire. These rate differences are highly significant (100% of runs significant in this range). In the highest third of syllable repertoire thresholds, adult song stability with a large syllable repertoire is an evolutionarily unstable state, transitioning primarily towards adult song plasticity (86% of runs significant in this range, **Figure 3.6**). We found similar trends when using two, four, and five bins for the song characteristic threshold values, with subtle differences. When using four or five bins, we still observe that song stability with larger syllable repertoires is an unstable combination. However, for the highest bin of threshold values, the transition rates are faster when changing to song plasticity, whereas for the second-highest bin, we observe faster transition rates toward repertoire size increases (**Figure 3.6 - figure supplement 1**). To rule out the possibility that syllable repertoire size evolution is faster in species with larger repertoire sizes regardless of learning program, we tested the rates of evolution of syllable repertoire size across monophyletic species pairs in our dataset. We found that lineages with larger syllable repertoire sizes do not systematically undergo faster or slower syllable repertoire size evolution (**Figure 3.1 - figure supplement 8**).

At low song repertoire threshold values, song plasticity with a small repertoire is an evolutionarily unstable state; there is rapid transition away from this combination, predominantly trending towards song stability but also transitioning secondarily to a larger song repertoire, with very high significance (100% of runs significant in this range). At moderate song repertoire thresholds, the highest rate is observed for species with small repertoires transitioning from song plasticity to song stability, also with very high significance (100% of runs significant in this range). At high song repertoire thresholds, the primary shift is from song stability to song plasticity in species with large song repertoires (89% of runs significant in this range) (**Figure 3.7**). When analyzing the results using five bins, transitions in the upper range of song repertoire values becomes more nuanced; in the highest bin, song stability with a larger song repertoire is very unstable, but is relatively stable in the second-highest bin. In the lower four bins, the dominant transition is from plastic to stable song with smaller song repertoires (**Figure 3.6 - figure supplement 1**). The results for syllables per song show some general trends that are complicated by the strong effect of the mimid species (see results of jackknife analysis below). We did not find evidence for correlated evolution between adult song stability or plasticity state and any of the other song characteristics (**Figure 3.6 - figure supplements 2-6**).

In addition, it has been proposed that polygyny and extra-pair paternity (EPP) may increase sexual selection pressures on sexually selected traits, including song (Emlen and Oring 1977; Payne 1984), and increased selection pressure due to polygyny was theorized to accelerate the evolution of song learning in a mathematical model (Aoki 1989). We tested for correlated evolution between adult song plasticity versus stability and both social mating system (polygyny vs. monogamy) and extra-pair paternity (low vs. high EPP), with the caveat that many species in our dataset lacked mating behavior classifications (57 species with social mating system data, 41 with EPP data). We did not find evidence for correlated evolution between song stability and EPP. There was, however, evidence for correlated evolution between polygynous/monogamous mating systems and song plasticity (100% of runs significant), with elevated rates of transition between polygyny and monogamy in the song-plastic state (**Figure 3.6 - figure supplement 7**).

Do our results depend on the particular values or families included in analyses?

In many cases, there were multiple studies that gave different estimates for a given song trait in one species, so we used the median values across studies for our main analysis. To test whether our results depended on the particular values we used, we repeated the PhylANOVA and Brownie analysis using either the maximum or minimum values reported in the literature. This did not significantly alter our PhylANOVA results for any song feature when species were split into song-stable and song-plastic groups (**Table S1**), though when the early song-stable, delayed-song stable, and song-plastic dataset was used, using the maximum values for song repertoire led to non-significant results (**Table S2-3**). When species were split into those with adult song plasticity and adult song stability, the Brownie analysis suggested that syllables per song was evolving significantly faster in song-plastic lineages when we used the median value, but using the minimum values led to non-significant results (**Table S4**). When species were split into three states (early song stability, delayed song stability, and song plasticity), we found that the three-rate model was not better than the one-rate model when the minimum values for song repertoire were used, while the three-rate models for syllable repertoire became significant when either the maximum or the minimum values were used (**Table S5**).

In our dataset, 24 songbird families were represented by 1 to 11 species each. This meant that families had unequal influence on the outcomes of the analyses. We performed a jackknife analysis to examine whether our results were affected by excluding individual families represented by 4 or more species in the full dataset. Exclusion of individual families did not significantly alter any of the phylANOVA results (**Tables S6-11**). In general, exclusion of individual families did not affect the Brownie results (**Tables S12-17**) or the BayesTraits results

(**Figure 3.6 - figure supplements 8-9, Figure 3.7 - figure supplement 1**). However, there were several notable exceptions, detailed below.

In several cases, removing a single family altered the significance of our findings. Removal of the Muscicapidae (3 species) from the Brownie analysis of inter-song interval created a two-rate model that did not fit the data significantly better than the one-rate model (**Table S13)**. For song duration, removal of Fringillidae (3 species) led to a two-rate model that was not significantly better fit than the one-rate model (**Table S14**). Removal of Mimidae (4 species) from the Brownie analysis of syllables per song drastically changed the results, such that the two-rate model was no longer significantly better fit than the one-rate model **(Table S15**). All of the included Mimidae species are song-plastic, so these data suggest that mimids may be driving the fit of the two-rate model for syllables per song. In addition, removal of Mimidae from the BayesTraits analysis of song plasticity and syllables per song altered the observed trends. At low threshold values of syllables per song, song-plasticity with low syllables per song is an unstable state, with high rates of transition towards either higher syllables per song or towards song stability (77% of runs significant in this range); however, when mimid species are removed, gaining more syllables per song in the song-plastic state becomes far more likely. At moderate values of syllables per song, there are elevated rates of transition both towards more syllables per song and towards fewer syllables per song in song-plastic species (100% of runs significant in this range); there is a comparable trend when mimid species are removed. At high threshold values, more syllables per song with song-stability is an unstable state (63% of runs significant in this range, **Figure 3.6 - figure supplement 2**); however, with mimids removed, more syllables per song with song plasticity becomes the most unstable state (**Figure 3.6 - figure supplement 9**).

Due to the strong effect that the inclusion of Mimidae had on the Brownie analysis of syllables per song, we performed a second jackknife analysis, in which each of the four mimid species was removed in turn. Exclusion of *Toxostoma rufum* or *Dumetella carolinensis* had little effect on the results (**Table S18**), leading to significant support for the two-rate model. However, exclusion of *Mimus polyglottos* (*p*=0.074) or of *Mimus gilvus* (*p*=0.509) led to a two-rate model that did not fit the data significantly better than the one-rate model (**Table S18**). Therefore, we concluded that the two *Mimus* species drove the estimated evolutionary rate of syllables per song in song-plastic species to be much greater than in song-stable species, and that faster evolution for syllables per song may not necessarily be a universal trend for song-plastic species.

Members of Mimidae are renowned for their vocal mimicry, frequently exhibiting improvisation and invention of syllables beyond simple imitation, and thus they lack the generally stereotyped song structure shown in other oscine families. Furthermore, mimids often have periods of continuous singing with minimal repetition of elements and irregular syllable spacing. Thus, quantification of song duration or number of syllables per song for mimids could be highly susceptible to listener perception (Wildenthal 1965). Therefore, though we acknowledge that mimids are an important case study in extended learning durations, our results for the evolutionary rate of syllables per song might be more meaningful across all bird species when mimids are excluded, in which case we find the rate of evolution of syllables per song is be independent of adult song plasticity.

Discussion

Previously, it was unknown whether the song-learning window evolved in concert with song features associated with sexual selection, as predicted by a computational model of song learning (Creanza et al. 2016). This is a critical missing piece of the puzzle of song learning evolution, as previous evidence has suggested that sexual selection only acts upon the features of song and not the length of the song-learning window or maintenance of the song-learning pathways in the brain. Here, we performed phylogenetically controlled analyses to assess the interactions between the length of the song-learning window—using adult song stability versus plasticity as a proxy—and the evolution of song characteristics. Interestingly, we noted that several evolutionary events relatively early in passerine evolution accounted for much of the diversity in the song-plasticity period in our sample of species. Our results predict that a bird's ability to modify its song as an adult not only affects the characteristics of its own song but also may interact with the evolution of larger repertoire sizes, since sexual selection for large repertoires could drive evolution of increased song plasticity.

We found two key trends in the trait correlation (phylANOVA) and evolutionary rate (Brownie) analyses. First, song plasticity affected the *direction* of evolution in traits that can be considered metrics of song complexity (syllable and song repertoire size, **Figure 3.1 and 3.3, Table 3.1**), leading to larger repertoires in species with adult song plasticity. Further, species with early song stability evolved their repertoires at a slower rate than species with longer learning (delayed song stability and song plasticity, **Table 3.5**), but song-plastic species did not evolve their repertoires at a faster rate than species with early or delayed song stability combined (**Table 3.6-9**). Thus, while repertoires only evolve directionally in song-plastic species (**Table 3.2-3**), our results suggest that extended learning through the first breeding season allows for faster, but not directional, evolution of repertoire size. It is possible, then, that delayed song learning allows individuals to modify their songs after migration and thus adapt their song to their new surroundings once they establish a territory, without necessarily corresponding with sexual selection for larger repertoires. This ability for an individual to adapt to a new local song might be beneficial, particularly when species have local dialect structure; however, this would not lead to directional evolution for any particular song feature. Second, song plasticity increased the *rate* of evolution primarily in traits that can be considered metrics of song performance (song duration, intersong interval, and song rate **Figure 3.5, Table 3.2**). While these performance-related song traits evolved faster in song-plastic lineages, this rapid evolution did not lead to significant differences in those song characteristics compared to song-stable lineages. A possible explanation may be that increases in repertoire size necessitate changes to song structure, but multiple structural aspects of song can be altered to accommodate these changes. Thus, there is no overall pattern of directional evolution in these other song characteristics. Alternatively, bird species may be required to adapt to increasing repertoire sizes while maintaining species-specific constraints imposed by innate aspects of song structure or female preferences for different performance characteristics. In the latter case, if information about innate characteristics and female preferences are known, it may be possible to predict how song traits will change in response to increasing repertoire sizes and greater adult song plasticity.

With our analyses of correlated evolution, we aimed to detect whether the state of the repertoire size or adult song stability versus plasticity consistently changes first in evolutionary history, facilitating a change in the other trait. Overall, our results suggest that there is not a consistent order of evolutionary transitions (**Figure 3.6-7**). For example, song stability with very large syllable or song repertoires, and song plasticity with very small syllable or song repertoires both formed evolutionarily unstable states, with high evolutionary rates of transition in both repertoire size and song-learning window. However, we do note that the fastest rates of transition in our analyses were those switching between song-stability and song-plasticity to leave those unstable states. This trend suggests that the magnitude of a species' repertoire may be more likely to drive the evolution of learning window than vice versa (**Figure 3.6-7**). This is consistent with the idea that selection acting upon song features could indirectly place selective pressures on the learning window. We propose several hypotheses that could explain these evolutionary dynamics: 1) it may be disproportionately costly to maintain song plasticity when syllable or song repertoire sizes are very small, perhaps because the benefit of extra time to learn does not outweigh the metabolic cost of maintaining plasticity, or 2) species with small syllable or song repertoires may have highly stereotyped songs which are selected for based on accuracy of learning and consistency of song production, favoring males that only learn from their fathers or early-life neighbors. Alternatively, in species where females prefer larger repertoires, 3) it may simply require more time to learn a large song or syllable repertoire than is available with a short learning window, or 4) learning large repertoires may require too much energy devoted to song practice during the crucial period of development before and during fledging, favoring birds that can learn for longer periods. Further research into the physiological or reproductive costs of song plasticity is needed.

Beginning with Darwin (Darwin 1872), numerous researchers have proposed that polygynous mating systems could lead to amplified sexual selection (Emlen and Oring 1977; Payne 1984). The elevated rates of transition that we observed between social mating systems in song-plastic lineages suggest an interesting hypothesis for further investigation: perhaps having a plastic song-learning program facilitates evolutionary transitions in mating systems.

Our results provide key evidence that sexual selection upon song characteristics might indirectly act upon the song-learning window. We do not fully understand the mechanisms underlying the maintenance or reopening of the song learning window in adulthood, but genetic, environmental, hormonal, and social factors are likely contributors (Eales 1985, 1987; Kroodsma and Pickert 1980; Morrison and Nottebohm 1993; Nottebohm 1969). For example, when zebra finches were reared in isolation, their sensitive periods were lengthened. These isolated birds maintained both gene expression profiles associated with song learning in the song system and high levels of neuronal addition to the HVC (a key region in the song system of the songbird brain) for longer than birds reared with an adult male tutor (Kelly et al. 2018; Wilbrecht et al. 2006), linking the neural underpinnings of song learning to the length of the song-learning window (Brenowitz and Beecher 2005; Nordeen 1997; Nordeen and Nordeen 1990; Nottebohm 1992). Furthermore, a positive association between HVC volume and song repertoire size has been demonstrated both intraspecifically (in song sparrows, (Pfaff et al. 2007)) and interspecifically (Devoogd et al. 1993). In light of our findings that adult song plasticity correlates with an increase in song repertoire size, there is a logical prediction that extended song learning may be associated with increased HVC volume across species. This is an important avenue for future research.

Although our data set includes species from 24 different songbird families, many families are not represented due to a lack of data about song stability. It will be important to expand this dataset in future studies. It would also be interesting to explore the evolutionary interactions between adult song plasticity and mimicry of heterospecific sounds, which has been observed in Mimidae and numerous other clades across the songbird lineage (Goller and Shizuka 2018). With our current dataset, we could not adequately explore the effects of mimicry on the evolution of song learning outside of the mimids, but the repeated evolution of mimicry makes it a particularly interesting topic for follow-up studies on the length of the song-learning window. In addition, different song metrics that are tailored to mimicry would be important in studying the evolution of vocal mimics and the dynamics of their unique vocal patterns. Furthermore, there is increasing interest in the importance of female song in species, which is more common than previously thought (Langmore 1998; Odom et al. 2014). Our dataset includes numerous species wherein females are known to sing at least occasionally (Langmore 1998), but the length of the song-learning window in females has not been assessed in any of these species. There is, however, some evidence that female birds can modify their song preferences in adulthood (Nagle and Kreutzer 1997). Thus, it remains an open question whether song plasticity affects the evolution of female song in the same way it affects male song, and whether species with adult song plasticity in males also have adult song plasticity in females.

Our findings shed new light on the broader subject of song evolution, specifically the evolution of the process of song learning. We hypothesize that sexual selection on certain aspects of song could in turn alter the selection pressures on the length of the song-learning window. Here, we perform phylogenetically controlled analyses across 67 songbird species to assess the evolutionary interactions between song characteristics and song plasticity in adulthood. With these analyses, we show the first evidence for this evolutionary relationship. Adult song stability versus plasticity is evolutionarily dependent upon the properties of the song itself: large syllable and song repertoires appear to drive the evolution of adult song plasticity and thus open-ended song learning. This provides context for the remarkable interspecific variation in song-learning windows across the songbird lineage and suggests an evolutionary mechanism by which sexual selection might have influenced the evolution of songbird brains.

Methods

Data collection

It is difficult to precisely measure the length of the song-learning window in both field and lab studies. In field studies, if a bird is recorded singing a new song element in its second year, researchers often cannot rule out the possibility that it learned that element during its first year but did not incorporate it into its song until later (Marler and Peters 1981; Vargas-Castro, Sánchez, and Barrantes 2015). Likewise, if a bird learns a new element in its second year but elects not to produce it, a recordist would be unlikely to capture it. In contrast, raising birds in the laboratory with known song exposure enables researchers to assess when a bird learned a particular song element (Chaiken, Böhner, and Marler 1994; Marler and Peters 1988; Nelson 1998; Nottebohm 1969) but also raises questions about whether the lab-reared birds are exhibiting their normal behaviors (Baptista and Petrinovich 1984; Kroodsma and Pickert 1984). In addition, these lab-rearing procedures have only been performed in a handful of species. In this paper, we examined studies that include longitudinal measures of adult song stability versus plasticity (e.g. (Nordby et al. 2002)) to determine whether song is modified across time as a proxy for the length of the song learning window.

We performed a literature search to gather information on the stability or plasticity of male song in oscine species. Studies with information about learning style were found via Google Scholar using the following search terms: [species name] or [common name] in combination with “open-ended,” “close-ended,” “closed-ended,” “age-limited,” “crystal*,” “adult learning,” and “song changes.” We used three strategies to assign the song stability classification for a species. We first defined a species as having adult song stability (‘song-stable species’) if the literature indicated that males did not modify their songs after the first breeding season. Species in which males modified their syllable repertoires after their first breeding season were classified as having adult song plasticity ('song-plastic species'). This strategy was meant to approximate the dichotomy of open- and closed-ended learners often used in the field. We made two exceptions to this for the species *Cacicus cela* and *Phoenicurus ochruros*, in which males do not gain their mature plumage until their second breeding season and may therefore be delayed in reaching sexual maturity relative to other bird species (Draganoiu et al. 2014; Trainer and Parsons 2002). Because these birds cease modifying their repertoires before reaching their second breeding season with mature plumage, they were considered song-stable (**Dataset S1**). Additionally, past research defined *Melospiza lincolnii* as an "open-ended improviser." It is unclear whether improvisation throughout the lifespan is equivalent to learning throughout the lifespan, however it does fit our definition for adult song plasticity. For the main analysis, we considered this species to have adult song plasticity, but we repeated the main analysis with this species reclassified as having adult song stability. This reclassification produced a negligible effect on our results (**Tables S19-20**). For our second strategy, we separated song-stable species into two sub-groups: early song-stable (species that cease modifying their songs before the first breeding season) and delayed song-stable (species that modify their songs during their first breeding season but not after). There was not enough information to make this determination for some species, so our dataset was reduced to 59 species. We used these same 59 species to generate a continuous measure of song stability for our third strategy. Because no information was available about the prevalence of song changes beyond the second breeding season for most of the species in our dataset, this measure only ranges from 0 to 2 years, and all song-plastic species were assigned a value of 2. Furthermore, it was not clear exactly when most of the delayed song-stable species stopped learning, so they were given a value of 1.33, at which point the first breeding season should have ended. The two species mentioned above that display delayed song and plumage maturation were assigned a value of 2.

To gather data on the song characteristics, we performed a literature search via Google Scholar and Web of Science using the search terms “Passeriformes” or [species name] in combination with “syllable repertoire” or “song repertoire.” This yielded a mix of primary sources and studies that had previously aggregated repertoire size data. We also gathered data from the curated field guides Birds of North America (Rodewald 2015) and Handbook of Birds of the World (del Hoyo et al. n.d.). We did not perform explicit searches for any of the other included song characteristics, but we collected this data whenever we encountered it. Song characteristic nomenclature is variable across studies, so, when possible, we read the methods of the primary sources to ensure the authors used the same definitions for song characteristics that we did.

We utilized the following definitions for song characteristics:

Syllable repertoire: the mean number of unique syllables sung per individual

Syllables per song: the mean number of unique syllables per song

Song repertoire: the mean number of unique songs produced per individual

Song duration: the mean length of a song, delineated by pauses or beginnings of a repeated motif (seconds)

Intersong interval: the mean pause length between songs (seconds)

Song rate: the mean number of songs produced per minute (calculated value; Rate = 60/(Duration + Interval))

Continuity: the fraction of singing time spent actively producing song (calculated value; Continuity = Duration/(Duration + Interval))

For three species with song-learning window data, syllable repertoire size estimates were not available in the literature, so we estimated these repertoire sizes from published sources or song recordings. For *Philesturnus rufusater* (Table S21), Jenkins (Jenkins 1978) coded songs into different types and gave the song repertoire types of each male studied in Table 2 of that paper.Information from one male was missing from this table. We deduced the repertoire of the missing male by first looking at Table 1 from that paper (Jenkins's Table 1), which showed that 16 males had a repertoire size of one song. Only 15 of the males in Jenkins's Table 2 had a repertoire size of one song. Thus, the missing male had a repertoire size of one song. We then compared the bands of males present in Jenkins's Table 2 to the territory map in Figure 7 of that paper, and A_RW was the only male missing from Jenkins's Table 2. A_RW was located in the DC region of Jenkins's Figure 7, so we assigned that as his repertoire. Jenkins notes that neighboring males share song types, so the only other song A_RW could have known instead of DC was ZZ, which has the same number of unique syllables as DC. For *Geospiza fortis* (Table S22) and *G. scandens* (Table S23), we used recordings from the Macaulay Library at the Cornell Lab of Ornithology (Anon 2009) to estimate the syllable repertoire size and syllables per song. For *G. fortis*, we also used the sonogram examples present in (Grant, Rosemary Grant, and Grant 1996).

When the song repertoire for a species equalled one, we assumed that its syllable repertoire was equal to its number of syllables per song. In many cases, there were multiple studies that gave different estimates for a given song variable in one species. To handle these discrepancies, we created three datasets (**Dataset S2**). For one, the main dataset, we used the median value across studies. For the second dataset, we always used the minimum value reported in the literature, and for the third, we used the maximum value reported in the literature. We log-normalized all song trait data. The three datasets revealed similar results; analyses using the maximum and minimum literature values are presented in **Tables S1-S5**.

We also catalogued data on mating behavior at the species level. In particular, we assembled binary classifications of social mating system (monogamy vs polygyny) and extra-pair paternity (low EPP vs high EPP). We considered a species to be monogamous or polygynous when a source unambiguously categorized that species' social mating system, i.e. we did not assign a social mating system to species labeled "probably," "usually," "mostly," "normally," "typically," and "generally monogamous/polygynous," etc. unless quantitative measurements were also provided. When quantitative data were available, species were defined as polygynous when at least 5% of males have more than one social mate, as in (Snyder and Creanza 2019). A review of extra-pair paternity studies estimated an average of ~11% of offspring per nest were attributable to extra-pair mates across species (Griffith, Owens, and Thuman 2002). In accordance with this estimate and with previous studies that used a binary classification of EPP (Snyder and Creanza 2019; Soma and Garamszegi 2011), we used a 10% threshold for either extra-pair young or nests containing at least one extra-pair chick to estimate the frequency of extra-pair paternity in that species (<10% = low EPP; ≥10% = high EPP, **Dataset S2**).

Assessing the evolutionary history of adult song plasticity

To predict the rate of transition between adult song stability and adult song plasticity, we used the ace function from the R package Phytools and a phylogeny from Jetz et al. (Jetz et al. 2012). We note that this phylogeny has broad coverage of oscine songbirds but is based on limited genetic data, so the topology could change as more avian genomes are sequenced (Jarvis et al. 2014; Lamichhaney et al. 2015) With this tree, we tested whether an all rates different (ARD) model fit the data significantly better than the equal rates (ER) model using an ANOVA. We then used the better-fit equal rates model to generate 10000 trees with make.simmap (Phytools). This function uses the rate from ace and a phylogenetic tree with annotated tips to create stochastic simulation maps for the potential evolutionary transitions between the song-stable and song-plastic states. We found the predicted ancestral state for each of these 10000 simulations and used countSimmap (Phytools) to count the total number of transitions that occurred in each map. The minimum number of predicted evolutionary transitions across these 10000 simulations was considered to be the most parsimonious; we also compared this to a manual count of evolutionary transitions starting from either ancestral state.

Detecting differences in song characteristic evolution in song-stable vs. song-plastic species

To test whether there were significant differences between song-stable species and song-plastic species for the song traits, we performed a phylogenetically controlled ANOVA (phylANOVA, Phytools) for each song characteristic. We repeated this analysis with the subset of species we classified into early song-stable, delayed song-stable, and song-plastic. Because there was a limited number of early song-stable species in this dataset, we only performed this re-analysis for song traits that had data for at least 9 early song-stable species (syllable repertoire size: 9 species with early song stability, song repertoire size: 9 species, and syllables per song: 10 species). In this paper, we visualize the predicted ancestral traits on the phylogenetic tree with color and pie graphs, however, the raw values are available in **Figure 3.1 - source data 1**. To test for correlations between song characteristics and the continuous values for the duration of song plasticity, we performed a phylogenetic generalized least squares (PGLS) analysis. We used the function gls (R package: nlme), with the "correlation" parameter lambda computed using the function corPagel (R package: ape).

To test whether adult song stability or plasticity affected the rate of evolution for the song characteristics, we used the function brownie.lite (R package: phytools). This function first calculates a one-rate model of evolution for a song characteristic using a phylogenetic tree and the current states of the tips for that song trait. This one-rate model assumes that change in the value of the song characteristic is random across evolutionary time and can be approximated by Brownian motion. Next, a model is generated wherein two different rates are calculated; this two-rate model assumes that the evolution of the song characteristic has one rate in the song-stable state and a different rate in the song-plastic state. This model requires estimations for the ancestral states of song stability for each branch of the phylogeny. To create these estimates, we used the function ace to calculate the rate of transition between the song-stable and song-plastic states for the full dataset. We then used these transition rates to generate 1300 different stochastic simulation maps (make.simmap) for the subset of species that had data for each song trait. For the Brownie analysis, we tested whether the two-rate model fit the data significantly better than the one-rate model by performing a chi-square test on the mean log likelihoods of the two models. We repeated this analysis for the set of species we classified as early song-stable, delayed song-stable, and song-plastic for traits for which we had data on at least 9 early song-stable species. We compared the three-rate model to the one-rate model. We also reran the two-rate model in this reduced data set by combining the early and delayed song-stable groups and testing whether the three-rate model was better than the two-rate model. Because the delayed song-stable trace peaked at a similar position to the song-plastic trace for syllable and song repertoire size (**Figure 3.4D-E**), we also compared the three-rate model to another version of the two-rate model, in which one group was early song-stable (shorter learning), and the other was delayed song-stable and song-plastic combined (longer learning).

We used BayesTraits to test for correlated evolution between song stability and song characteristics, or, in other words, whether the rate and direction of evolutionary transitions of one trait are dependent on the state of another trait, and whether an order of transition events can be inferred. Specifically, we tested the hypothesis that an evolutionary change in song stability increases the likelihood of an evolutionary change in certain song variables or mating behaviors, or vice versa. BayesTraits compares two models of discrete trait evolution for a pair of binary traits and a given phylogenetic tree: 1) an independent model (i.e. the evolution of one trait does not depend on the other trait) and 2) a dependent model (i.e. the evolutionary transitions of each trait depend on the state of the other trait, suggesting correlated evolution). Using the maximum likelihood method, BayesTraits reports marginal likelihoods for the computed dependent and independent models (function Discrete in package btw [BayesTraitsWrapper]), allowing us to determine whether the dependent model describes the data significantly better than the independent model. We used function LRtest (package: lmtest) to perform the likelihood ratio test. Since this model requires both input traits to be binary, we classified the continuous song characteristics as binary groups (“low” or “high”) based on a delineating threshold. Instead of choosing the threshold arbitrarily, we used each unique value of the song characteristic data as the threshold and repeated the analysis 100 times at each threshold. This method of using a spectrum of thresholds to delineate the “low” and “high” value categories resulted in transition rates that varied dramatically depending on where the threshold was placed. In essence, when the threshold is set as a value in the bottom third of the unique trait values present in the data, the analysis evaluated the rate transition from low to moderate and larger values for a song trait and vice versa. When the threshold is set as a value in the upper range of the unique trait values present in the data, the analysis calculates the rates of transition from higher song trait values to medium and lower values. To account for this nuance, we binned the threshold data into two to five bins, with three bins as the default: low (bottom third of unique trait values), medium (middle third) and high (top third). We then calculated the mean of each state transition rate in each bin. In addition to the song characteristics, we also analyzed song stability versus social mating system (i.e. social monogamy or polygyny) and rate of EPP. These analyses were performed for 1000 runs each.

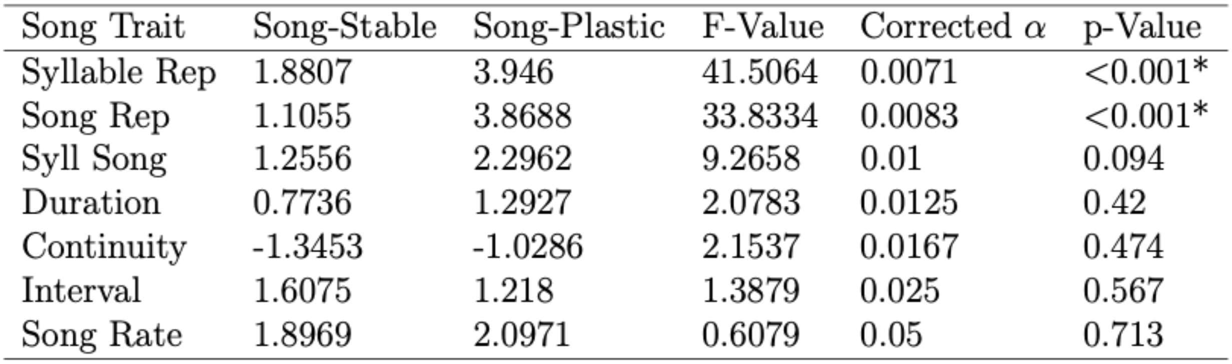
Jackknife Analysis

Some families of birds were well represented in our sample, while others were only represented by one or two species. To test whether any well-represented family significantly skewed our results, we removed each family that was represented by four or more species in the full dataset in turn, and repeated the phylANOVA, brownie.lite, and BayesTraits analyses. Jackknife analyses were only performed when significant results were obtained in the main analysis. Thus, all song traits except continuity were tested in the phylANOVA and brownie.lite jackknife analysis, while only syllable repertoire, song repertoire, and syllables per song were tested in the BayesTraits jackknife analysis. Each Brownie analysis was run on 1300 unique stochastic character maps, and each BayesTraits analysis was repeated 20 times. We determined the family of each species based on its classification in the 2017 version of the eBird Clements Integrated Checklist (Clements 2007). The family Locustella was combined with Acrocephalidae, as Acrocephalidae was paraphyletic when Locustella was considered to be a separate family. The Mimidae family alone had a large effect on the syllables per song metric, so we performed another jackknife analysis with phylANOVA and brownie.lite by removing each mimid species in turn.

Correction for Multiple Testing

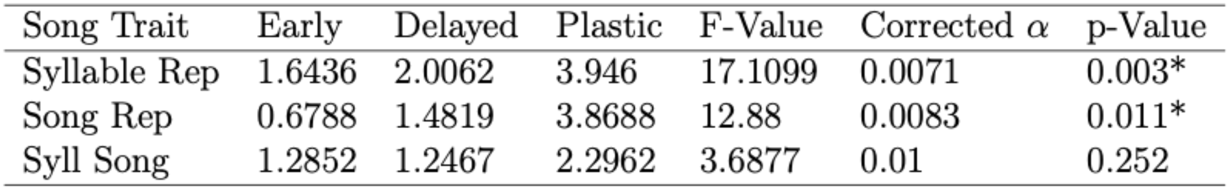
We used a Holm-Bonferroni correction to control for testing multiple hypotheses with the same data (Holm 1979). This correction is appropriate for data wherein the outcome of one test is likely to be related to the outcome of another test, as was likely to be the case for song characteristics.

Tables



Song Trait	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Syllable Rep	1.8807	3.946	41.5064	0.0071	<0.001*
Song Rep	1.1055	3.8688	33.8334	0.0083	<0.001*
Syll Song	1.2556	2.2962	9.2658	0.01	0.094
Duration	0.7736	1.2927	2.0783	0.0125	0.42
Continuity	-1.3453	-1.0286	2.1537	0.0167	0.474
Interval	1.6075	1.218	1.3879	0.025	0.567
Song Rate	1.8969	2.0971	0.6079	0.05	0.713

Table 3.1: PhylANOVA results for all song traits when birds are divided into species with adult song stability or adult song plasticity. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. * denotes traits with significantly different groups.



Song Trait	Early	Delayed	Plastic	F-Value	Corrected α	p-Value
Syllable Rep	1.6436	2.0062	3.946	17.1099	0.0071	0.003*
Song Rep	0.6788	1.4819	3.8688	12.88	0.0083	0.011*
Syll Song	1.2852	1.2467	2.2962	3.6877	0.01	0.252

Table 3.2: PhylANOVA results for all song traits when birds are divided into early song stability, delayed song stability, and song plasticity. Song traits are sorted from most to least significant. Early, Delayed, and Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. * denotes traits with significantly different groups.

Song Trait	State 1	State 2	T-Value	p-Value
Syllable Rep	Plastic	Delayed	4.8995	0.012*
Syllable Rep	Early	Plastic	4.6091	0.003*
Syllable Rep	Early	Delayed	0.6872	0.659
Song Rep	Plastic	Delayed	4.0268	0.044*
Song Rep	Early	Plastic	4.3074	0.015*
Song Rep	Early	Delayed	1.0444	0.55

Table 3.3: Post-hoc pairwise phylANOVA tests for significant song traits when birds are divided into early song stability, delayed song stability, and song plasticity. * denotes traits with significantly different groups.

Song Trait	Slope	Std Error	λ	T-Value	Corrected α	p-Value
Syllable Rep	0.9067	0.2449	0.8913	3.7021	0.0071	<0.001*
Song Rep	1.1013	0.3123	0.8316	3.5263	0.0083	<0.001*
Syll Song	0.3701	0.2224	0.4699	1.6642	0.01	0.1029
Interval	0.4221	0.2646	0.8823	1.5953	0.0125	0.1215
Continuity	-0.2135	0.1439	0.8832	-1.4838	0.0167	0.1486
Duration	0.3702	0.2569	1.0163	1.441	0.025	0.1578
Song Rate	-0.2113	0.25	0.7307	-0.8453	0.05	0.4048

Table 3.4: Results of PGLS analysis between song characteristics and continuous song stability. Test performed on the natural log scale values of song characteristics. λ is the value by which off-diagonal elements in the Brownian motion model are multiplied to make the correlation structure. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Song traits are sorted from most to least significant. * denotes significant slopes.

Song Trait	One Rate	Two Rates	p-Value
Syll Song	-110.6482	-100.7673	<0.001*
Song Rate	-43.4397	-38.4938	0.002*
Interval	-45.2842	-40.5004	0.002*
Duration	-71.2042	-66.3122	0.002*
Continuity	-25.6471	-24.7285	0.175
Syllable Rep	-120.2983	-120.0695	0.499
Song Rep	-113.5829	-113.3706	0.515

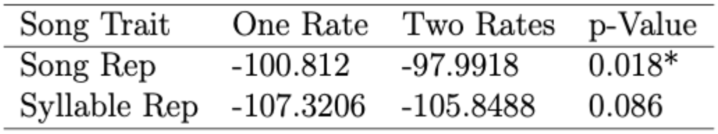
Table 3.5: Brownie results for song traits when birds are divided into species with adult song stability or adult song plasticity. For **Tables 3.5-9**, rate columns show mean log likelihood. Song traits are sorted from most to least significant. * denotes traits where the more complex model fit the data significantly better than the simpler model.

Song Trait	One Rate	Three Rates	p-Value
Syll Song	-97.8349	-86.3206	<0.001*
Song Rep	-100.812	-97.7647	0.014*
Syllable Rep	-107.3206	-105.5895	0.063

Table 3.6: Brownie results for song traits when birds are divided into early song stability, delayed song stability, and song plasticity.

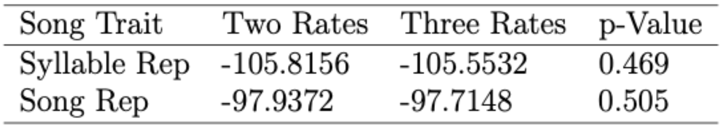
Song Trait	Two Rates	Three Rates	p-Value
Song Rep	-100.691	-97.7148	0.015*
Syllable Rep	-107.1332	-105.5532	0.075
Syll Song	-86.3125	-86.3447	1

Table 3.7: Brownie results for song traits when birds are divided into either song stability (early and delayed) and song plasticity (Two Rates) or early song stability, delayed song stability, and song plasticity (Three Rates).



Song Trait	One Rate	Two Rates	p-Value
Song Rep	-100.812	-97.9918	0.018*
Syllable Rep	-107.3206	-105.8488	0.086

Table 3.8: Brownie results for song traits when birds are divided into shorter learning (early song stability) and longer learning (delayed song stability and song plasticity).



Song Trait	Two Rates	Three Rates	p-Value
Syllable Rep	-105.8156	-105.5532	0.469
Song Rep	-97.9372	-97.7148	0.505

Table 3.9: Brownie results for song traits when birds are divided into either shorter learning (early song stability) and longer learning (delayed song stability and song plasticity) (Two Rates) or early song stability, delayed song stability, and song plasticity (Three Rates).

Figures

Figure 3.1: Syllable repertoire size is larger in species with adult song plasticity even when controlling for phylogeny. These phylogenies show the calculated evolution of natural-log transformed syllable repertoire size and either (A) stable and plastic song stability states or (B) early song-stable, delayed song-stable, and song-plastic states. Dots at the tips of branches represent the current song-stability state. Pie charts represent the likelihood that the common

ancestor at that node was in each song-stability state. Dark purple colors represent small syllable repertoires while white represents large repertoires. For the sake of visualization, the color range was truncated based on the distribution of the data, such that the lowest value was the 25th percentile minus the range of the 25th to 50th percentile and the highest value was the 75th percentile plus the range of the 50th to 75th percentile. See **Table 3.1** for PhylANOVA results.

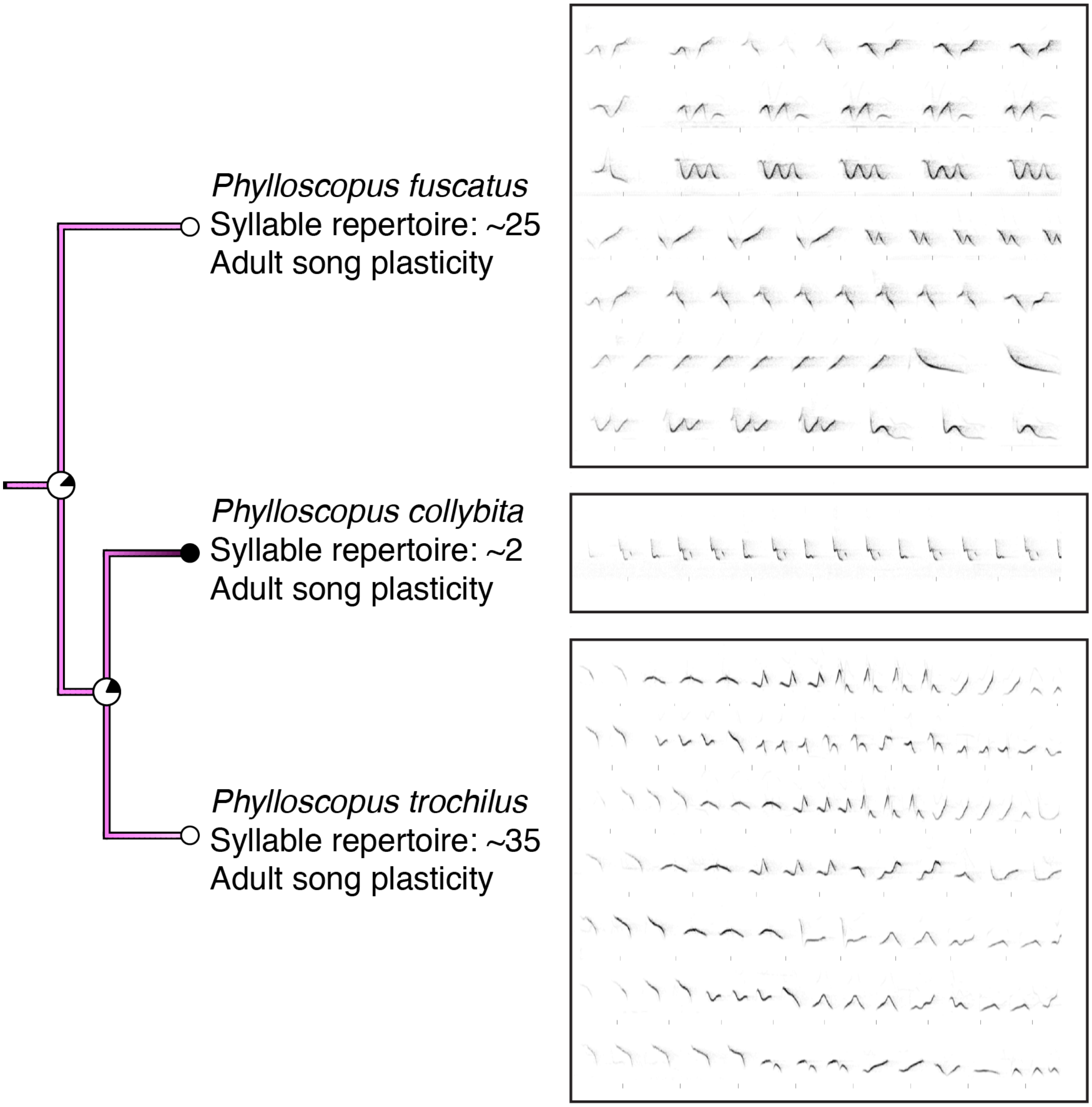


Figure 3.2: Comparison of song stability state and song examples in *Phylloscopus* species. *P. collybita*, a species with adult song stability, has a smaller syllable repertoire size than *P. trochilus* and *P. fuscastus*, two species with adult song plasticity. Colors of branches and nodes correspond with **Figure 3.3**. Sonograms were generated from recordings obtained from xeno-canto.org: XC340281 recorded by Tom Wulf (*P. fuscatus*), XC414221 recorded by Frank

Lambert (*P. collybita*), and XC402265 recorded by Hans Matheve (*P. trochilus*). Sonograms are used only to demonstrate comparative repertoire size from one individual for each species and were stretched horizontally to fit the allotted space.

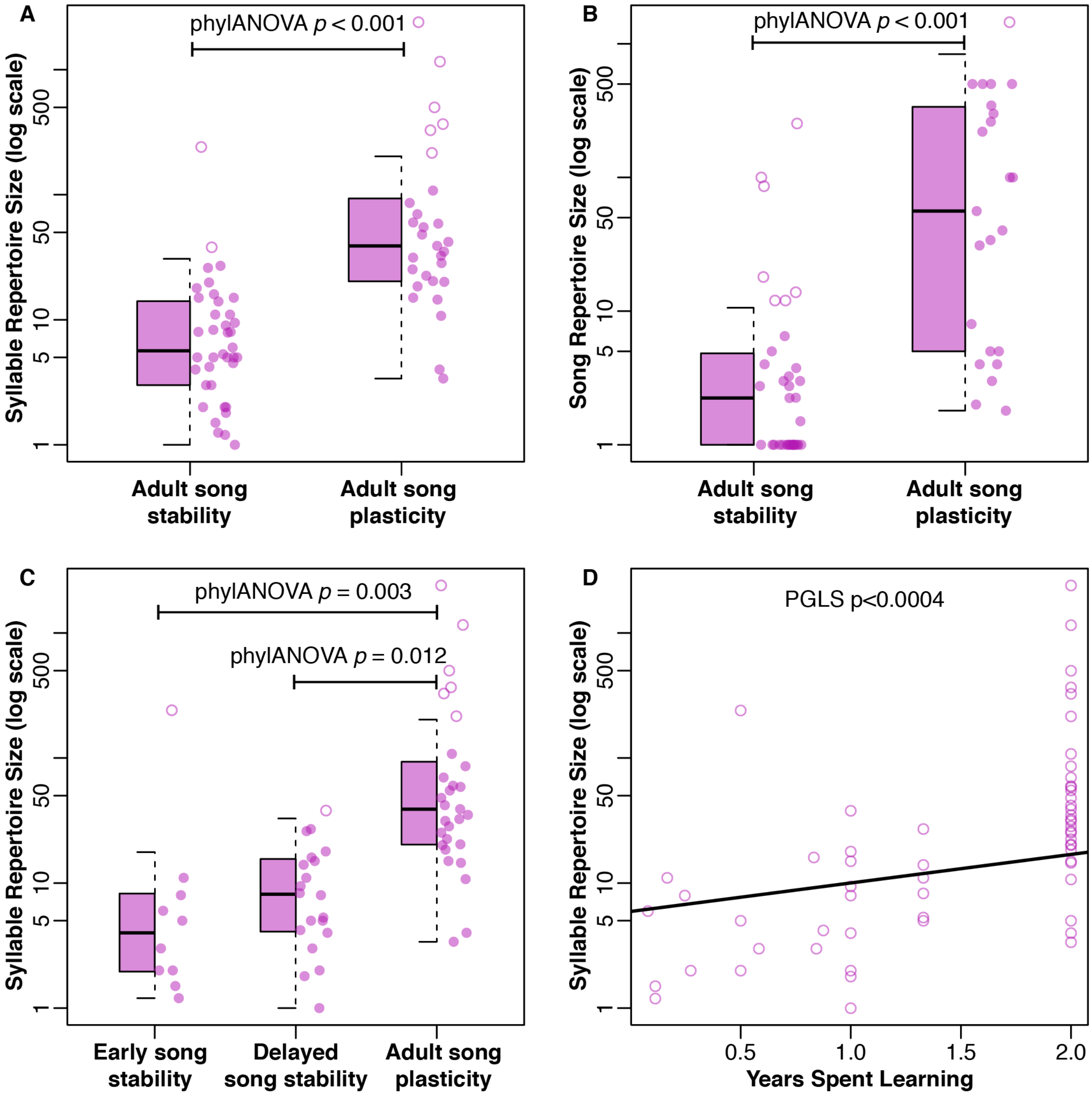


Figure 3.3: Distribution of repertoire sizes in species with different song stability states. (A) shows the distribution of syllable repertoires and (B) shows the distribution of song repertoires when species are broken into two groups based on song stability. (C) shows the distribution of syllable repertoires when species are broken into three groups based on song stability. Boxes indicate the 25th, 50th, and 75th percentile. The lower whisker is either the minimum value or the 25th percentile minus 1.5 times the interquartile range, whichever was larger. The upper

whisker is either the maximum value or the 75th percentile plus 1.5 times the interquartile range, whichever was smaller. Dots are the raw values as a scatter plot. Solid dots are within the range of the box and whiskers, while open dots are outliers. (A-B) Species with adult song plasticity had significantly larger syllable and song repertoires. See **Table 3.1** for full PhylANOVA results. (C) Species with adult song plasticity had significantly larger syllable repertoires than early song stable and delayed song stable species, but there was no significant difference between early and delayed song stable species (p=0.659). See **Tables 3.2 and 3.3** for full PhylANOVA results. (D) shows the continuous relationship between syllable repertoire size and song stability when song stability is truncated at 2 years due to lack of data in subsequent years.

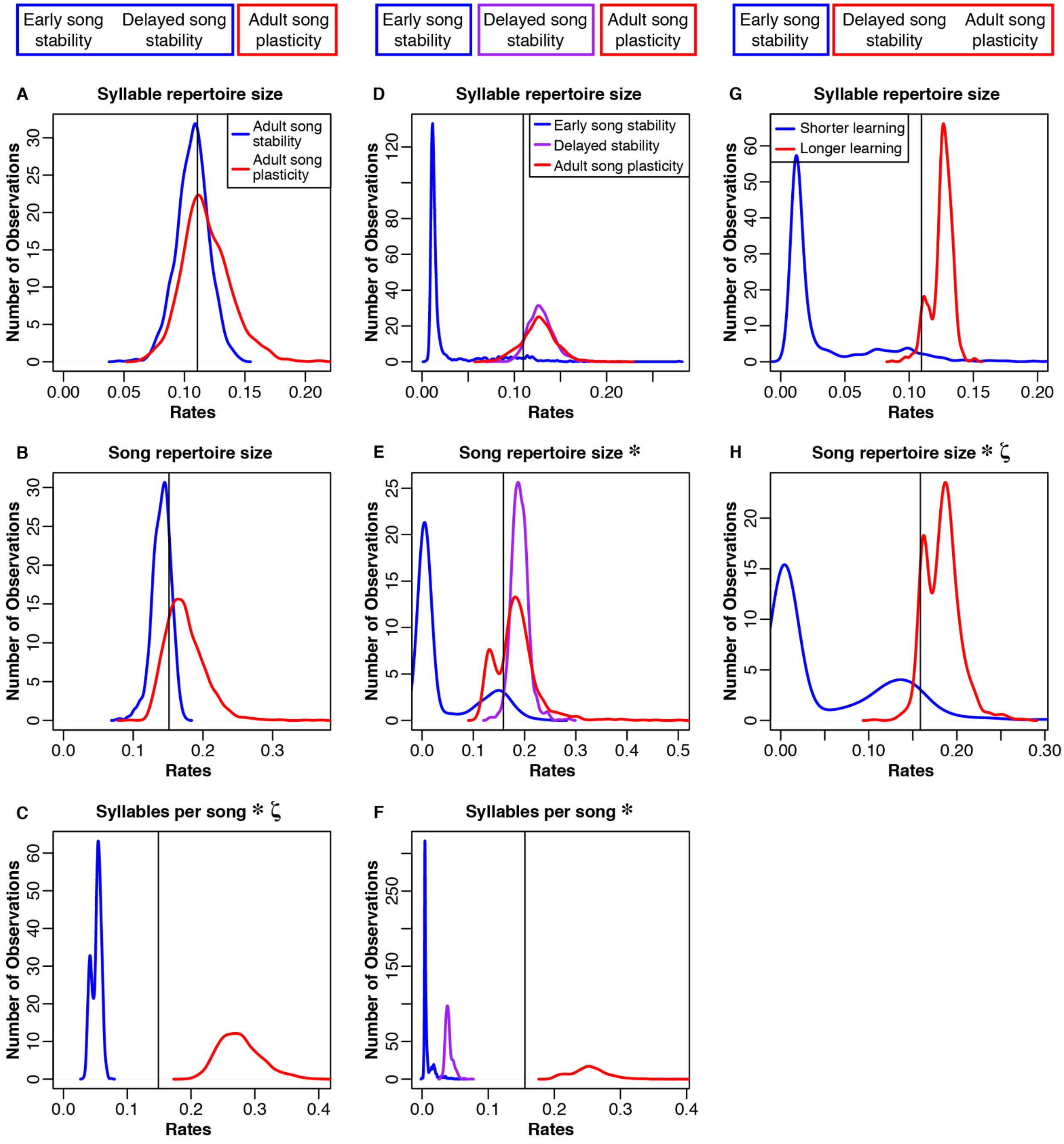


Figure 3.4: Distributions of rates for natural-log transformed song traits related to complexity. The boxes at the top illustrate how we grouped the species for each model. Column 1 (A-C): Blue traces are song-stable, while red traces are song-plastic. Column 2 (D-F): Blue traces are early song-stable, purple traces are delayed song stable, and red traces are song-plastic. Column 3 (G-H): Blue traces are early song-stable, while red traces are delayed song-stable and

song-plastic combined. The black line shows the rate value for the one-rate model in all columns. Asterisks indicate that the rate of evolution of that song characteristic significantly differed between groups. Lowercase zeta (ζ) indicates the multi-rate model that best fit the data while using the fewest number of rates. In the case of syllable repertoire, the multi-rate models were not significantly better than the one-rate model. See **Tables 3.5-9** for chi-square test results.

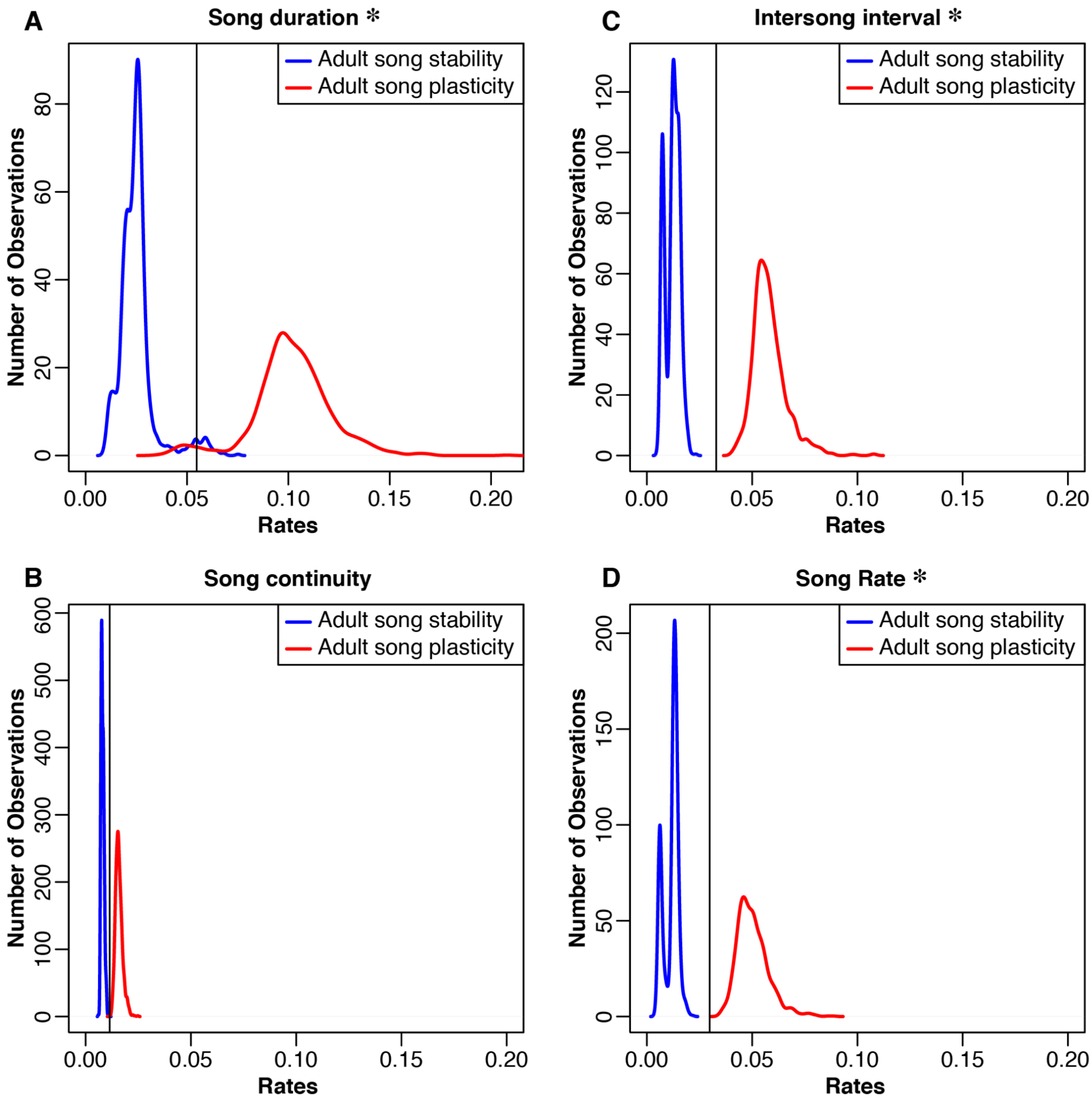


Figure 3.5: Distributions of rates for natural-log transformed song traits related to performance. Blue traces are song-stable, while red traces are song-plastic. The black line shows the rate for the one-rate model. Asterisks indicate that the rate of evolution of that song characteristic significantly differed between song-stable and song-plastic lineages. See **Table 3.5** for chi-square test results.

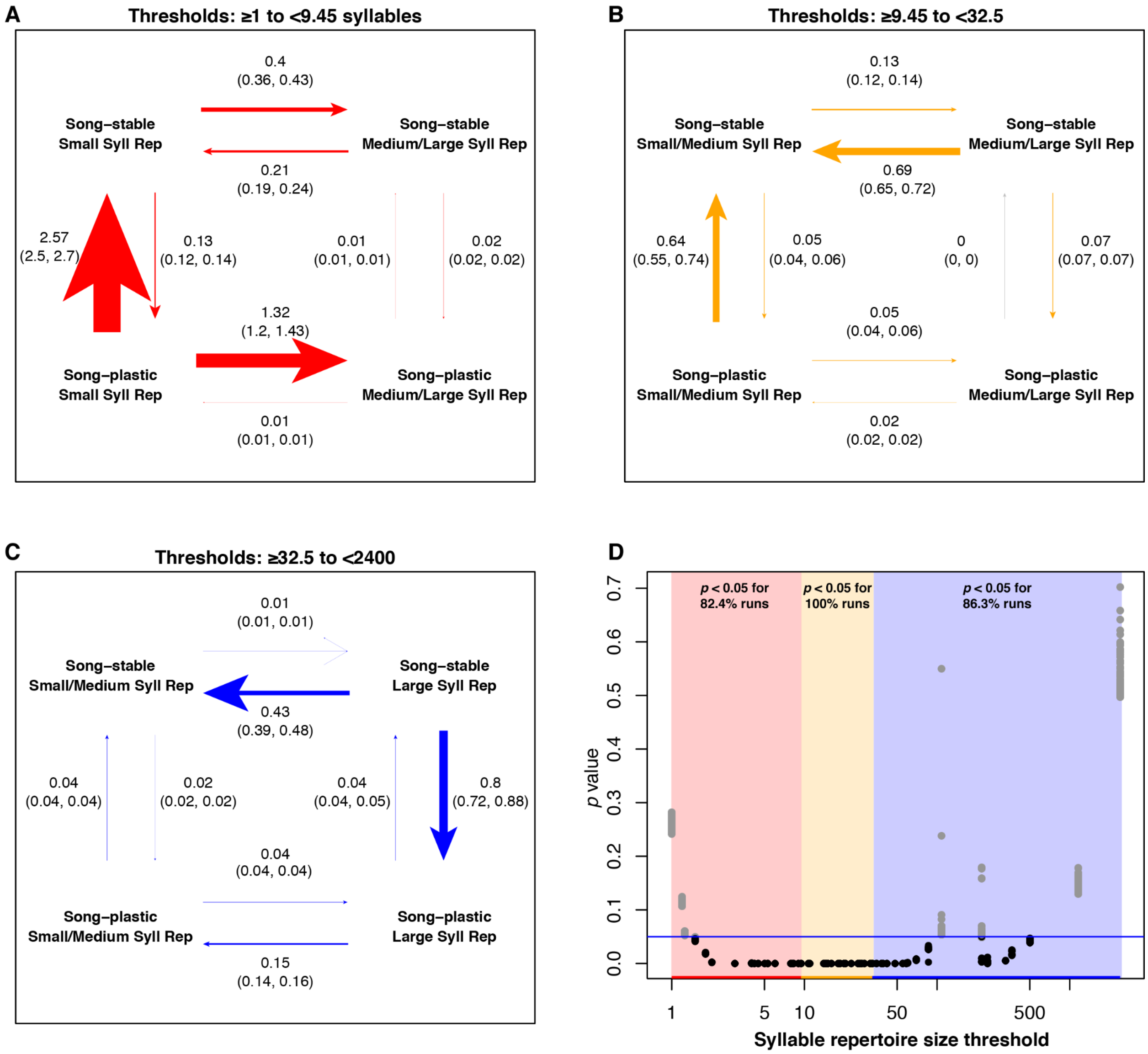


Figure 3.6: Analysis of correlated evolution between adult song stability/plasticity and syllable repertoire size. We repeated the BayesTraits test using each value of the continuous song trait as the threshold delineating the larger and smaller syllable repertoire groups. We performed a total of 100 runs per threshold. We pooled the results of all of the runs into three groups based on whether the threshold was in the lowest, middle, or highest third of the unique trait values. Within these groups, we computed the mean percentage of runs that were significant at $p < 0.05$ at each threshold. (A-C) Rate of transition plots when the lowest (red), middle

(yellow), and highest (blue) thirds of the unique syllable repertoire values in the dataset were used as the threshold. Rates are the average across all runs when the threshold denoting small/large repertoire sizes was defined as each value within each segment. Arrows are labeled with the mean rate and the 95% confidence interval. Arrow weights are scaled to the mean rate values. (D) P-values from the 100 runs per threshold, plotted against threshold. Colored bars denote low, middle, and high threshold segments. Blue line denotes $p = 0.05$.

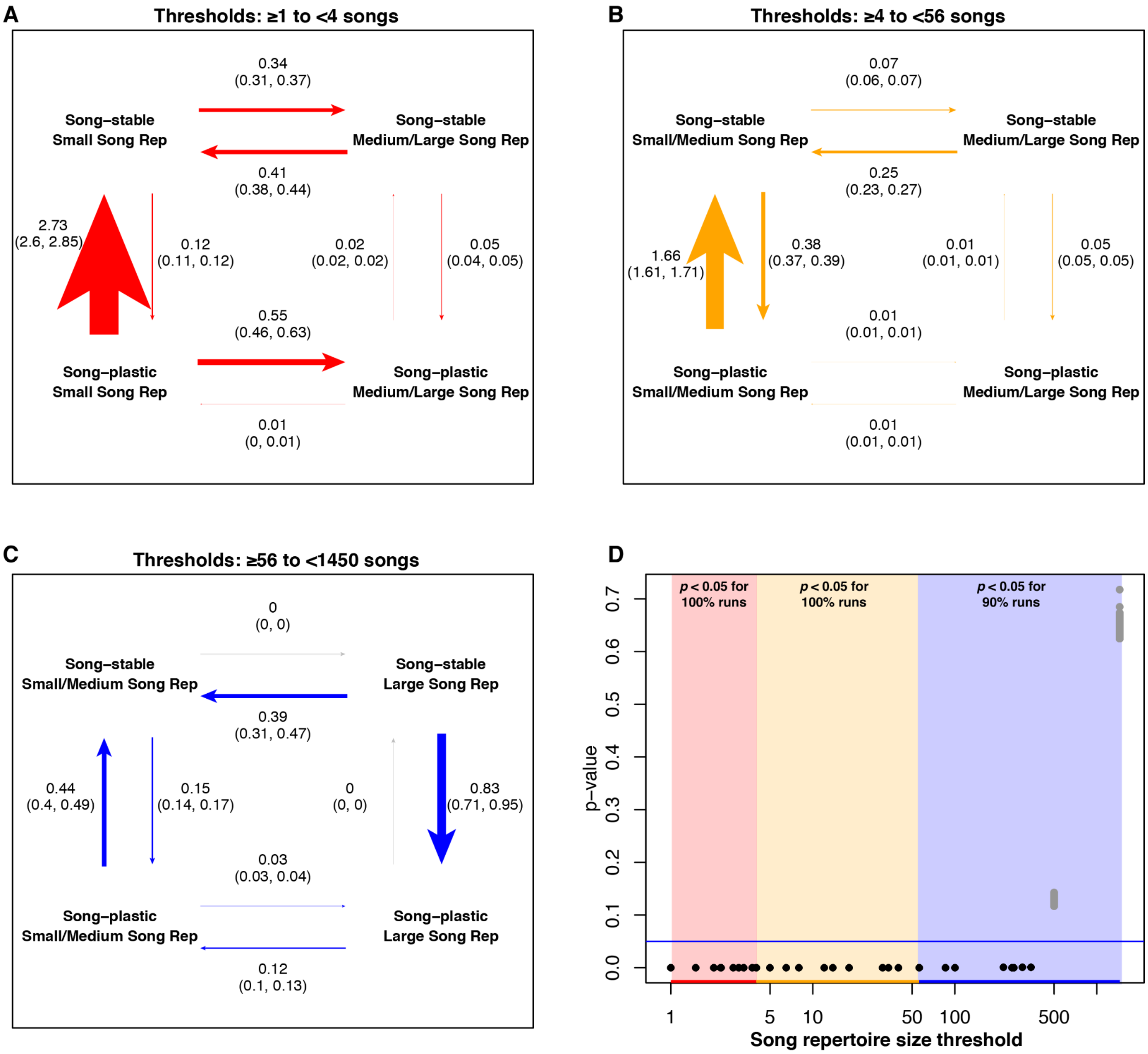


Figure 3.7: Analysis of correlated evolution between adult song stability/plasticity and song repertoire size. Labeling is the same as in **Figure 3.6**

References

Adret-Hausberger, Martine, Hans-Rudolf Güttinger, and Fritz W. Merkel. 2010. "Individual Life History and Song Repertoire Changes in a Colony of Starlings (Sturnus Vulgaris)." *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 84(4):265–80.

Anon. 2009. "Macaulay Library: Sound and Video Catalog." *Choice Reviews Online* 46(05):46–2661.

Aoki, Kenichi. 1989. "A Sexual-Selection Model for the Evolution of Imitative Learning of Song in Polygynous Birds." *The American Naturalist* 134(4):599–612.

Baptista, Luis F., and Lewis Petrinovich. 1984. "Social Interaction, Sensitive Phases and the Song Template Hypothesis in the White-Crowned Sparrow." *Animal Behaviour* 32(1):172–81.

Beecher, Michael D., and Eliot A. Brenowitz. 2005. "Functional Aspects of Song Learning in Songbirds." *Trends in Ecology & Evolution* 20(3):143–49.

Böhner, Jörg. 1990. "Early Acquisition of Song in the Zebra Finch, Taeniopygia Guttata." *Animal Behaviour* 39(2):369–74.

Brenowitz, Eliot A., and Michael D. Beecher. 2005. "Song Learning in Birds: Diversity and Plasticity, Opportunities and Challenges." *Trends in Neurosciences* 28(3):127–32.

Catchpole, Clive K., and Peter J. B. Slater. 2003. *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Chaiken, Marthaleah, Jörg Böhner, and Peter Marler. 1994. "Repertoire Turnover and the Timing of Song Acquisition in European Starlings." *Behaviour* 128(1):25–39.

Clements, James F. 2007. *The Clements Checklist of Birds of the World*. Comstock Publishing Associates.

Creanza, Nicole, Laurel Fogarty, and Marcus W. Feldman. 2016. "Cultural Niche Construction of Repertoire Size and Learning Strategies in Songbirds." *Evolutionary Ecology* 30(2):285–305.

Darwin, Charles. 1872. *The Descent of Man, and Selection in Relation to Sex*. London: J. Murray.

Devoogd, T. J., J. R. Krebs, S. D. Healy, and A. Purvis. 1993. "Relations between Song Repertoire Size and the Volume of Brain Nuclei Related to Song: Comparative Evolutionary Analyses amongst Oscine Birds." *Proceedings. Biological Sciences / The Royal Society* 254(1340):75–82.

Dowsett-Lemaire, Françoise. 1979. "The Imitative Range Of The Song Of The Marsh Warbler Acrocephalus Palustris, With Special Reference To Imitations Of African Birds." *The Ibis*

121(4):453–68.

Draganoiu, Tudor I., Aurélien Moreau, Lucie Ravaux, Wim Bonckaert, and Nicolas Mathevon. 2014. “Song Stability and Neighbour Recognition in a Migratory Songbird, the Black Redstart.” *Behaviour* 151(4):435–53.

Eales, Lucy A. 1985. “Song Learning in Zebra Finches: Some Effects of Song Model Availability on What Is Learnt and When.” *Animal Behaviour* 33(4):1293–1300.

Eales, Lucy A. 1987. “Song Learning in Female-Raised Zebra Finches: Another Look at the Sensitive Phase.” *Animal Behaviour* 35(5):1356–65.

Emlen, S., and L. Oring. 1977. “Ecology, Sexual Selection, and the Evolution of Mating Systems.” *Science* 197(4300):215–23.

Eriksen, Ane, Tore Slagsvold, and Helene M. Lampe. 2011. “Vocal Plasticity - Are Pied Flycatchers, Ficedula Hypoleuca, Open-Ended Learners?” *Ethology* 117(3):188–98.

Espmark, Yngve O., and Helene M. Lampe. 1993. “VARIATIONS IN THE SONG OF THE PIED FLYCATCHER WITHIN AND BETWEEN BREEDING SEASONS.” *Bioacoustics* 5(1-2):33–65.

Galeotti, Paolo, Nicola Saino, Elena Perani, Roberto Sacchi, and Anders Pape Møller. 2001. “Age‐related Song Variation in Male Barn Swallows.” *Italian Journal of Zoology* 68(4):305–10.

Garamszegi, László Z., Dieter Heylen, Anders P. Møller, Marcel Eens, and Florentino de Lope. 2005. “Age-Dependent Health Status and Song Characteristics in the Barn Swallow.” *Behavioral Ecology* 16(3):580–91.

Garamszegi, László Zsolt, and Marcel Eens. 2004. “Brain Space for a Learned Task: Strong Intraspecific Evidence for Neural Correlates of Singing Behavior in Songbirds.” *Brain Research. Brain Research Reviews* 44(2-3):187–93.

Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. “Phylogenetic Analysis of Covariance by Computer Simulation.” *Systematic Biology* 42(3):265–92.

Gil, Diego, James L. S. Cobb, and Peter J. B. Slater. 2001. “Song Characteristics Are Age Dependent in the Willow Warbler, Phylloscopus Trochilus.” *Animal Behaviour* 62(4):689–94.

Gil, Diego, and Manfred Gahr. 2002. “The Honesty of Bird Song: Multiple Constraints for Multiple Traits.” *Trends in Ecology & Evolution* 17(3):133–41.

Goller, Maria, and Daizaburo Shizuka. 2018. “Evolutionary Origins of Vocal Mimicry in Songbirds.” *Evolution Letters* 2(4):417–26.

Grant, B. Rosemary, B. Rosemary Grant, and Peter R. Grant. 1996. “Cultural Inheritance of Song and Its Role in the Evolution of Darwin's Finches." *Evolution* 50(6):2471.

Griffith, Simon C., Ian P. F. Owens, and Katherine A. Thuman. 2002. "Extra Pair Paternity in Birds: A Review of Interspecific Variation and Adaptive Function." *Molecular Ecology* 11(11):2195–2212.

Hausberger, M., P. F. Jenkins, and J. Keene. 1991. "Species-Specificity and Mimicry in Bird Song: Are They Paradoxes? A Reevaluation of Song Mimicry in the European Starling." *Behaviour* 117(1):53–81.

Holm, S. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics, Theory and Applications* 6(2):65–70.

del Hoyo, J., A. Elliott, J. Sargatal, D. A. Christie, and E. (eds). de Juana. n.d. "Handbook of the Birds of the World Alive." *Lynx Edicions, Barcelona*. Retrieved 2017 (http://www.hbw.com/).

Immelman, K. 1969. "Song Development in the Zebra Finch and Other Estrildid Finches." *Bird Vocalizations* 61–74.

Jarvis, Erich D., Siavash Mirarab, Andre J. Aberer, Bo Li, Peter Houde, Cai Li, Simon Y. W. Ho, Brant C. Faircloth, Benoit Nabholz, Jason T. Howard, Alexander Suh, Claudia C. Weber, Rute R. da Fonseca, Jianwen Li, Fang Zhang, Hui Li, Long Zhou, Nitish Narula, Liang Liu, Ganesh Ganapathy, Bastien Boussau, Md Shamsuzzoha Bayzid, Volodymyr Zavidovych, Sankar Subramanian, Toni Gabaldón, Salvador Capella-Gutiérrez, Jaime Huerta-Cepas, Bhanu Rekepalli, Kasper Munch, Mikkel Schierup, Bent Lindow, Wesley C. Warren, David Ray, Richard E. Green, Michael W. Bruford, Xiangjiang Zhan, Andrew Dixon, Shengbin Li, Ning Li, Yinhua Huang, Elizabeth P. Derryberry, Mads Frost Bertelsen, Frederick H. Sheldon, Robb T. Brumfield, Claudio V. Mello, Peter V. Lovell, Morgan Wirthlin, Maria Paula Cruz Schneider, Francisco Prosdocimi, José Alfredo Samaniego, Amhed Missael Vargas Velazquez, Alonzo Alfaro-Núñez, Paula F. Campos, Bent Petersen, Thomas Sicheritz-Ponten, An Pas, Tom Bailey, Paul Scofield, Michael Bunce, David M. Lambert, Qi Zhou, Polina Perelman, Amy C. Driskell, Beth Shapiro, Zijun Xiong, Yongli Zeng, Shiping Liu, Zhenyu Li, Binghang Liu, Kui Wu, Jin Xiao, Xiong Yinqi, Qiuemei Zheng, Yong Zhang, Huanming Yang, Jian Wang, Linnea Smeds, Frank E. Rheindt, Michael Braun, Jon Fjeldsa, Ludovic Orlando, F. Keith Barker, Knud Andreas Jønsson, Warren Johnson, Klaus-Peter Koepfli, Stephen O'Brien, David Haussler, Oliver A. Ryder, Carsten Rahbek, Eske Willerslev, Gary R. Graves, Travis C. Glenn, John McCormack, Dave Burt, Hans Ellegren, Per Alström, Scott V. Edwards, Alexandros Stamatakis, David P. Mindell, Joel Cracraft, Edward L. Braun, Tandy Warnow, Wang Jun, M. Thomas P. Gilbert, and Guojie Zhang. 2014. "Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds." *Science* 346(6215):1320–31.

Jenkins, P. F. 1978. "Cultural Transmission of Song Patterns and Dialect Development in a Free-Living Bird Population." *Animal Behaviour* 26:50–78.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. "The Global Diversity of Birds in Space and Time." *Nature* 491(7424):444–48.

Kelly, Theresa K., Somayeh Ahmadiantehrani, Adam Blattler, and Sarah E. London. 2018. “Epigenetic Regulation of Transcriptional Plasticity Associated with Developmental Song Learning.” *Proceedings. Biological Sciences / The Royal Society* 285(1878). doi: 10.1098/rspb.2018.0160.

Kipper, Silke, and Sarah Kiefer. 2010. “Age-Related Changes in Birds’ Singing Styles.” Pp. 77–118 in *Advances in the Study of Behavior*.

Kroodsma, Donald E., and Roberta Pickert. 1980. “Environmentally Dependent Sensitive Periods for Avian Vocal Learning.” *Nature* 288(5790):477–79.

Kroodsma, Donald E., and Roberta Pickert. 1984. “Sensitive Phases for Song Learning: Effects of Social Interaction and Individual Variation.” *Animal Behaviour* 32(2):389–94.

Lamichhaney, Sangeet, Jonas Berglund, Markus Sällman Almén, Khurram Maqbool, Manfred Grabherr, Alvaro Martinez-Barrio, Marta Promerová, Carl-Johan Rubin, Chao Wang, Neda Zamani, B. Rosemary Grant, Peter R. Grant, Matthew T. Webster, and Leif Andersson. 2015. “Evolution of Darwin’s Finches and Their Beaks Revealed by Genome Sequencing.” *Nature* 518(7539):371–75.

Langmore, N. E. 1998. “Functions of Duet and Solo Songs of Female Birds.” *Trends in Ecology & Evolution* 13(4):136–40.

Liu, Wan-Chun, and Donald E. Kroodsma. 2006. “Song Learning by Chipping Sparrows: When, Where, and from Whom.” *The Condor* 108(3):509.

Liu, Wan-Chun, and Fernando Nottebohm. 2007. “A Learning Program That Ensures Prompt and Versatile Vocal Imitation.” *Proceedings of the National Academy of Sciences of the United States of America* 104(51):20398–403.

Marler, P. 1990. “Song Learning: The Interface between Behaviour and Neuroethology.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 329(1253):109–14.

Marler, Peter, and Susan Peters. 1987. “A Sensitive Period for Song Acquisition in the Song Sparrow, Melospiza Melodia: A Case of Age-Limited Learning.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 76(2):89–100.

Marler, Peter, and Susan Peters. 1988. “Sensitive Periods for Song Acquisition from Tape Recordings and Live Tutors in the Swamp Sparrow, Melospiza Georgiana.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 77(1):76–84.

Marler, P., and S. Peters. 1981. “Sparrows Learn Adult Song and More from Memory.” *Science* 213(4509):780–82.

Martens, Jochen, and Philipp Kessler. 2000. “Territorial Song and Song Neighbourhoods in the Scarlet Rosefinch Carpodacus Erythrinus.” *Journal of Avian Biology* 31(3):399–411.

Morrison, R. G., and F. Nottebohm. 1993. “Role of a Telencephalic Nucleus in the Delayed Song Learning of Socially Isolated Zebra Finches.” *Journal of Neurobiology* 24(8):1045–64.

Mountjoy, James D., and Robert E. Lemon. 1995. “Extended Song Learning in Wild European Starlings.” *Animal Behaviour* 49(2):357–66.

Nagle, L., and M. L. Kreutzer. 1997. “Adult Female Domesticated Canaries Can Modify Their Song Preferences.” *Canadian Journal of Zoology* 75(8):1346–50.

Nelson, D. A. 1998. “External Validity and Experimental Design: The Sensitive Phase for Song Learning.” *Animal Behaviour* 56(2):487–91.

Nicholson, Joanne S., Katherine L. Buchanan, Rupert C. Marshall, and Clive K. Catchpole. 2007. “Song Sharing and Repertoire Size in the Sedge Warbler, Acrocephalus Schoenobaenus: Changes within and between Years.” *Animal Behaviour* 74(5):1585–92.

Nordby, J. Cully, J. Cully Nordby, S. Elizabeth Campbell, and Michael D. Beecher. 2002. “Adult Song Sparrows Do Not Alter Their Song Repertoires.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 108(1):39–50.

Nordeen, E. J., and K. W. Nordeen. 1990. “Neurogenesis and Sensitive Periods in Avian Song Learning.” *Trends in Neurosciences* 13(1):31–36.

Nordeen, K. W. 1997. “Neural Correlates of Sensitive Periods in Avian Song Learning.” *Annals of the New York Academy of Sciences* 807:386–400.

Nottebohm, Fernando. 1969. “The ‘Critical Period’ for Song Learning.” *The Ibis* 111(3):386–87.

Nottebohm, Fernando. 1984. “Birdsong as a Model in Which to Study Brain Processes Related to Learning.” *The Condor* 86(3):227.

Nottebohm, Fernando. 1992. “The Search for Neural Mechanisms That Define the Sensitive Period for Song Learning in Birds.” *Netherlands Journal of Zoology* 43(1):193–234.

Nottebohm, F., M. E. Nottebohm, and L. Crane. 1986. “Developmental and Seasonal Changes in Canary Song and Their Relation to Changes in the Anatomy of Song-Control Nuclei.” *Behavioral and Neural Biology* 46(3):445–71.

Odom, Karan J., Michelle L. Hall, Katharina Riebel, Kevin E. Omland, and Naomi E. Langmore. 2014. “Female Song Is Widespread and Ancestral in Songbirds.” *Nature Communications* 5:3379.

O’Meara, Brian C., Cécile Ané, Michael J. Sanderson, and Peter C. Wainwright. 2006. “Testing for Different Rates of Continuous Trait Evolution Using Likelihood.” *Evolution; International Journal of Organic Evolution* 60(5):922–33.

Pagel, Mark. 1994. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” *Proceedings of the Royal Society of London.*

Series B: Biological Sciences 255(1342):37–45.

Pagel, Mark, and Andrew Meade. 2006. “Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible‐Jump Markov Chain Monte Carlo.” *The American Naturalist* 167(6):808–25.

Payne, Robert B. 1984. “Sexual Selection, Lek and Arena Behavior, and Sexual Size Dimorphism in Birds.” *Ornithological Monographs* (33):iii – 52.

Pfaff, Jeremy A., Liana Zanette, Scott A. MacDougall-Shackleton, and Elizabeth A. MacDougall-Shackleton. 2007. “Song Repertoire Size Varies with HVC Volume and Is Indicative of Male Quality in Song Sparrows (Melospiza Melodia).” *Proceedings. Biological Sciences / The Royal Society* 274(1621):2035–40.

Price, J. Jordan, and David Yuan. 2011. “Song-Type Sharing and Matching in a Bird with Very Large Song Repertoires, the Tropical Mockingbird.” *Behaviour* 148(5):673–89.

Rauschecker, Josef P., and Peter Marler. 1987. *Imprinting and Cortical Plasticity: Comparative Aspects of Sensitive Periods*. Wiley-Interscience.

Revell, Liam J. 2011. “Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things).” *Methods in Ecology and Evolution / British Ecological Society* 3(2):217–23.

Rodewald, P., ed. 2015. *The Birds of North America: Https://birdsna.org*. Cornell Laboratory of Ornithology, Ithaca, NY.

Searcy, W., and M. Andersson. 1986. “Sexual Selection and the Evolution of Song.” *Annual Review of Ecology and Systematics* 17(1):507–33.

Searcy, William A., and Peter Marler. 1984. “Interspecific Differences in the Response of Female Birds to Song Repertoires.” *Zeitschrift Für Tierpsychologie* 66(2):128–42.

Smith, G. T., E. A. Brenowitz, M. D. Beecher, and J. C. Wingfield. 1997. “Seasonal Changes in Testosterone, Neural Attributes of Song Control Nuclei, and Song Structure in Wild Songbirds.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 17(15):6001–10.

Snyder, Kate T., and Nicole Creanza. 2019. “Polygyny Is Linked to Accelerated Birdsong Evolution but Not to Larger Song Repertoires.” *Nature Communications* 10(1):884.

Soma, Masayo, and László Zsolt Garamszegi. 2011. “Rethinking Birdsong Evolution: Meta-Analysis of the Relationship between Song Complexity and Reproductive Success.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 22(2):363–71.

Trainer, Jill M., and Ryan J. Parsons. 2002. “DELAYED VOCAL MATURATION IN POLYGYNOUS YELLOW-RUMPED CACIQUES.” *The Wilson Bulletin* 114(2):249–54.

Tramontin, Anthony D., and Eliot A. Brenowitz. 2000. “Seasonal Plasticity in the Adult Brain.” *Trends in Neurosciences* 23(6):251–58.

Tramontin, Anthony D., Nicole Perfito, John C. Wingfield, and Eliot A. Brenowitz. 2001. “Seasonal Growth of Song Control Nuclei Precedes Seasonal Reproductive Development in Wild Adult Song Sparrows.” *General and Comparative Endocrinology* 122(1):1–9.

Vargas-Castro, Luis E., Natalie V. Sánchez, and Gilbert Barrantes. 2015. “Song Plasticity over Time and Vocal Learning in Clay-Colored Thrushes.” *Animal Cognition* 18(5):1113–23.

Wilbrecht, Linda, Heather Williams, Nidhi Gangadhar, and Fernando Nottebohm. 2006. “High Levels of New Neuron Addition Persist When the Sensitive Period for Song Learning Is Experimentally Prolonged.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 26(36):9135–41.

Wildenthal, Joyce L. 1965. “Structure in Primary Song of the Mockingbird (Mimus Polyglottos).” *The Auk* 82(2):161–89.

CHAPTER 4.

CULTURAL SIGNATURES OF PUNCTUATED ENVIRONMENTAL CHANGE: MEASURING THE EFFECTS OF POPULATION-WIDE STRESS ON SONG DISTRIBUTION IN DARK-EYED JUNCOS AND SONG SPARROWS

Kate T. Snyder[1], Maria Sellers[1], Nicole Creanza[1]
[1] Department of Biological Sciences, Vanderbilt University, Nashville, TN USA 37204

A modified version of this chapter is available as a preprint on bioRxiv with doi: 10.1101/2022.12.03/519005

Introduction

The natural world is punctuated by catastrophic events that can dramatically change natural selection pressures, altering the course of a species' evolution (Price et al. 1984; Donihue et al. 2020). The documented examples of these short-term phenomena on a population scale are characterized by unusually high mortality rates after all individuals are subjected to extremely hostile conditions, where those individuals that exhibit certain adaptive traits disproportionately survive to pass on their genes to the next generation (B. R. Grant and Grant 1993; P. R. Grant and Grant 2011; Brown and Brown 1998). These studies have centered on traits that are assumed to be genetically determined and not plastic in adulthood; however, recent research has shown that gene-by-environment interactions can have significant outcomes on phenotypes, which could theoretically also drive a species' evolution, since natural selection primarily acts on expressed phenotypes. There are several potential mediators of phenotypic change based on interactions between organisms and their environments, including epigenetic modifications, microbiome composition, and learned or conditioned behaviors (Sepers et al. 2019; Henry et al.

2021; Caspi et al. 2022). These factors can all be affected by the physical and social environment, especially those experienced during development, in potentially long-lasting ways that can be adaptive, conditionally adaptive in a given environment, or maladaptive (Boyce, Sokolowski, and Robinson 2020; Brass et al. 2021; Gienapp et al. 2017; Regan et al. 2022).

Birdsong is a learned behavior that is extremely diverse across the 4000+ species that exhibit it in the songbirds (suborder: Oscine, Order: Passeriformes). Learning duration varies between species, with some species learning only in the first year of life, after which they do not modify their songs (termed "closed-ended learners"), and others learning for several years or throughout their lifetimes ("open-ended learners"). Song evolution is influenced by sexual selection, as it primarily functions in mate attraction, species identification, and territory defense. Songs are composed of syllables, defined as periods of continuous sound that are separated by a short period of silence. The diversity of song is typically characterized using a set of metrics that include measurements of pitch, frequency modulation, duration, spectral entropy, and syllable diversity. One key metric is "syllable repertoire," or the total number of unique syllables an individual produces. The average syllable repertoire within a species ranges from 1 (e.g. in the chipping sparrow (Marler and Isaac 1960)) to over 2000 (in the brown thrasher (Kroodsma and Parker 1977)), and is considered a metric of song complexity and learning capacity (Robinson, Snyder, and Creanza 2019; Creanza, Fogarty, and Feldman 2016). The song preferences of females vary widely between species, but in laboratory and field studies across several species, it has been found that females tend to prefer more complex songs, as measured by a male's syllable repertoire, number of unique syllables per song, or number of unique songs (Catchpole, Dittami, and Leisler 1984; Searcy 1984; Reid et al. 2004).

Birdsong is an energetically costly behavior, requiring significant investment in specialized neural development and time dedicated towards practicing song (Nowicki and Searcy 2004). Stress during development has been shown to impair song learning ability in several species, using song learning accuracy or song complexity as proxy metrics (Brumm, Zollinger, and Slater 2009; MacDougall-Shackleton and Spencer 2012). The developmental stress hypothesis posits that birdsong is an honest indicator of fitness because females can gain accurate information about a male's developmental environment by assessing song quality, enabling them to choose a mate that had a non-stressful upbringing (Nowicki and Searcy 2004). The conditions causing stress, such as food availability and parasite load, are usually assumed to vary across a population, with some birds experiencing little stress during development, and attractive song characteristics are favorably selected by females. However, certain extreme environmental events might effectively apply stress to an entire population simultaneously, which could result in an entire generation of birds with impaired song learning.

Genetically determined phenotypes in a population are constantly undergoing minor changes as individuals exhibit varied success from year to year, but major shifts in population phenotype happen far more rapidly under extreme conditions (Price et al. 1984). Could a similar phenomenon occur in a culturally transmitted phenotype such as song that does not give a selective advantage under the environmental change? Specifically, if a stressor was applied across an entire breeding songbird population simultaneously, such that virtually all juveniles experienced impaired learning, would there be a detectable perturbation in population song in the next breeding season? If these juveniles' songs became the models for future generations, could such a punctuated stressor ultimately induce a lasting change in the populations' songs? We predict that ecosystem-level stressors may cause population-level changes to bird song as entire cohorts that experience developmental stress subsequently learn their songs imperfectly. Further, if these individuals become the song tutors for future generations, these altered songs could theoretically spread in the population and become an established dialect.

While the biological effect of developmental stress upon the song quality of individuals has been well studied, the cultural effect of an environmental stressor applied to an entire natural population has not, to our knowledge, been investigated. We have identified an ecological opportunity through which to attempt to observe the effects of such a phenomenon. From June through November 2016, New York State, which typically has a humid climate, experienced a severe drought. The most severely affected region, central-western New York State, including Ithaca and the surrounding area, spent over two months under the classification of "Extreme Drought" [Information from the Drought Monitor (Svoboda et al. 2002), accessed via https://droughtmonitor.unl.edu/NADM/Home.aspx]. Air temperatures in 2016 in this region were also higher than normal, around the 90th percentile relative to the previous 60 years (Sweet et al. 2017). These combined factors had significant effects on vegetation, using agriculture losses as a proxy metric; in the most heavily impacted region, New York Drought Region VII, which encompasses Tompkins, Cayuga, Onondaga, Schuyler, Seneca, Yates, Ontario, Livingston, and Wyoming counties, farms reported average non-irrigated crop losses ranging from 45% loss of field crops to 69% loss of fruit crops (Sweet et al. 2017). Incidentally, this area is also a hotspot for songbird recordings: the area surrounding Ithaca, New York, which is home to Cornell University, the Cornell Lab of Ornithology, and the Macaulay Library of Natural Sounds, has a high density of publicly available bird song recordings obtained by both academics and community bird enthusiasts over the past few decades, with a sharp increase in recordings in the past five to ten years.

While there does not seem to be robust analysis or systematic documentation of effects on passerines during this drought, research on passerine species from other drought events can provide some predictions as to the potential effects. In many species, including song sparrows, reproduction and reproductive physiology is negatively affected by the absence of water (Wingfield et al. 2012; Prior, Heimovics, and Soma 2013; Kozlovsky et al. 2018). This extends to decreases in nesting behavior and egg-laying (Langin et al. 2009). Drought leads to physiological stress and mortality in individuals by reducing the availability of plant and arthropod biomass that serve as food sources for many bird species, as well as through dehydration (Langin et al. 2009; Boag and Grant 1984). These effects may be most pronounced in specialist species, but generalists are also affected negatively by lack of water availability and tree and plant mortality (Roberts et al. 2019).

Here, we present analyses of the songs of two focal species, the dark-eyed junco (*Junco hyemalis*) and the song sparrow (*Melospiza melodia*), recorded in central-western New York State between 2006 and 2019. We analyze these recordings obtained from publicly accessible repositories with a specific prediction that the population distributions of certain song features will shift from before the 2016 drought to after the drought. We hypothesize in particular that metrics of song complexity in song sparrows will shift towards simpler song, since song complexity has previously been shown to negatively correlate with metrics of stress in song sparrows (Schmidt et al. 2012). We predict that there will also be a larger variance in these feature distributions after the drought, since recordings after 2016 are likely to represent a mix of individuals that learned their songs prior to 2016, individuals that learned their songs during drought period, and individuals that may have learned their songs from the drought-year birds.

Methods

Site Selection

In the summer of 2016, a large area of New York State experienced a drought that was unprecedented in recorded history for the region. This provided a unique opportunity for a natural experiment to study the effects of environmental stress on birdsong on a population-wide scale. Tompkins County, home to the Cornell Lab of Ornithology, was classified as being under drought conditions from late June until well into November of that year, with over two months spent under the classification of "Extreme Drought." Using the ample community-science birdsong recording data available in this area, collected before, during and after the drought, we quantify features of the songs of individual species to assess whether the drought had a detectable effect on songs in the affected populations. We compare the songs from the drought area with songs in a control region of approximately similar land area. The control regions were selected by finding the area with the next highest density of recordings after western New York from the pool of available recordings on Macaulay Library and xeno-canto. We assessed the density of recordings for each target species and both time periods, 2006-June 2016 (before drought) and 2017-2019 (after drought) and selected the regions that were not impacted or were less severely impacted by the 2016 drought and that had the greatest number of available recordings before and after the drought period. For both species, we only considered regions in the eastern half of the continental U.S. since there are well-documented differences in song dialects and singing behavior between east and west coast populations of song sparrows (Hughes et al. 1998), and different dark-eyed junco subspecies are present in the West (Ferree 2013). Since no regions outside of western New York had at least 20 recordings for dark-eyed juncos from 2006-2016, we chose the dark-eyed junco control region based on the abundance of recordings from 2017-2019, and, of those options, chose the region that did not also experience severe drought during 2016. (**Figures S4.1A-B**) For song sparrows, only one potential control area had relatively dense recording availability both before and after 2016, and it had experienced moderate drought during 2016. (**Figures S4.1C-D**)

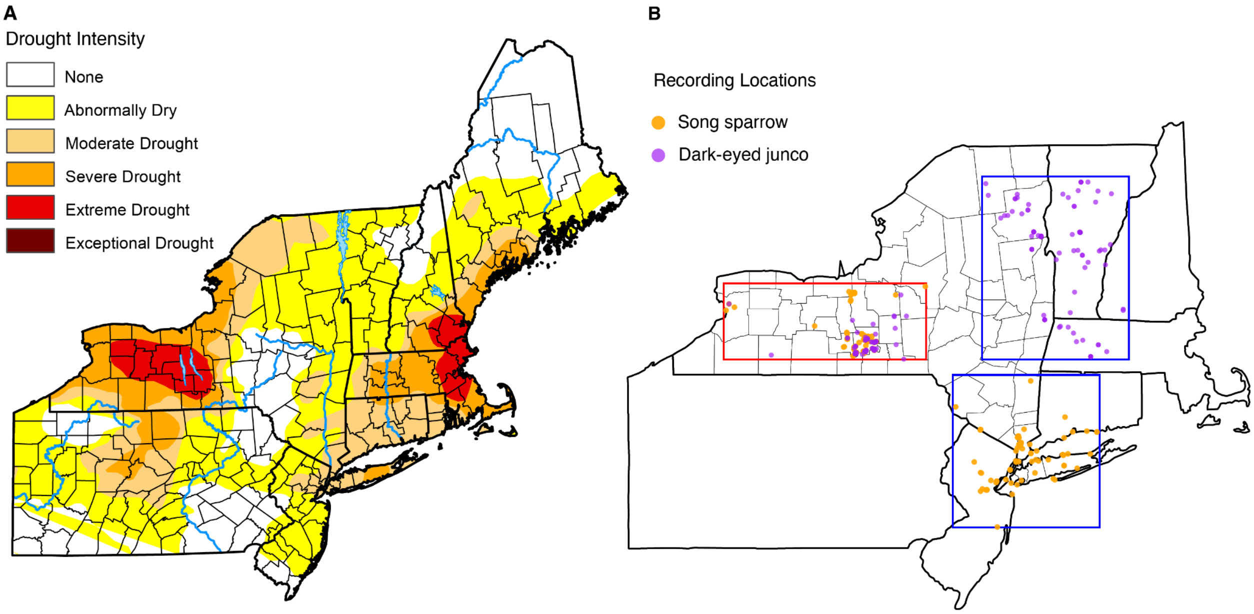


Figure 4.1. Maps of drought severity and recording locations. (A) Map from U.S. Drought Monitor (Svoboda et al. 2002) from September 2016 showing the severity and geographic extent of the drought in western New York. (B) Map of songs for song sparrows and dark-eyed juncos in our analyses. The red box encompasses the recordings of both song sparrows (orange) and dark-eyed juncos (purple) in the drought region. For a similarly sized control region that did not experience a drought, we found the densest sampling of song sparrow recordings in northern New Jersey and southern New York and Connecticut and of dark-eyed junco recordings in eastern New York, Vermont, New Hampshire, and western Massachusetts.

Species Selection Criteria

We selected song sparrows and dark-eyed juncos as target species based on several factors related to 1) feasibility of study, 2) likelihood of drought impact on individuals, and 3) likelihood that any drought-associated impact of song could be detected in years following the drought. Regarding the feasibility of study, we considered the number of song recordings available and song structure. In early 2017, we retrieved metadata for all audio recordings from Tompkins County, New York which were tagged as containing “song” in the Macaulay Library online repository. We excluded any species with less than 100 recordings. We also eliminated any species that sing continuously, or that do not have a clear separation between song bouts to facilitate consistent song analysis using our acoustic analysis program, Chipper (Searfoss, Pino, and Creanza 2020). Drought-associated impacts on song within a population would be diluted by high rates of immigration or emigration, so we sought species for which there is limited dispersal or at least some evidence of philopatry, so that adult birds recorded in a location are likely to have hatched near that location. We also focused on species that do not modify their song repertoires after their first breeding season (i.e. “closed-ended” learners), since we hypothesized that effects of the drought on song learning could be less pronounced in “open-ended” learners, who can continue to modify their song in subsequent breeding seasons (Brenowitz and Beecher 2005). Based on these criteria, we selected the dark-eyed junco (*Junco hyemalis*) and the song sparrow (*Melospiza melodia*), the two songbird species with the most recordings in the drought area, both of which are closed-ended learners and show evidence of philopatry (Weatherhead and Forbes 1994; Liebgold, Gerlach, and Ketterson 2013), suggesting that there would be continuity in population makeup in the area from before to after the drought.

Environmental limitations to food sources may have been a major contributor to developmental stress during the drought (Huberty and Denno 2004; White 1984). The 2016 drought has been shown to have had a large detrimental effect on agricultural vegetation and wild vegetation alike (Sweet et al. 2017). The diet of *Junco hyemalis* is composed primarily of plant matter, which comprises 76% of their diet, 62% of which is seeds, and they have been known to supplement their diet with insects on occasion (Del Hoyo et al. 1992). Although the diet of *Melospiza melodia* is also primarily composed of plant matter for most of the year, during the breeding season their diets rely more heavily on animal food sources, such as beetles, butterflies, and small gastropods (Del Hoyo et al. 1992).

The two focal species for our study have notable variation in song complexity. The dark-eyed junco has a relatively simple and consistent song, typically consisting of one syllable sung repetitively, whereas the song sparrow sings many songs composed of multiple syllables. Dark-eyed juncos have small repertoires of around 4 songs on average, and have been known to sing the same song up to 120 times in a row before switching to a new song (Cardoso et al. 2009; Williams and MacRoberts 1977; Newman, Yeh, and Price 2008). Song sparrows, on the other hand, typically cycle between 5-13 different song types, often singing several renditions of one type before switching to the next (Wood and Yezerinac 2006).

Obtaining Recordings

Two citizen-science recording libraries, Macaulay Library and Xeno-canto, provided most of the recordings for this study. We utilized the search engines of these digital databases with the search terms "*Junco hyemalis*" and *"Melospiza melodia*" with the filters "Location = New York, United States" and "Sounds = Song." From these digital databases, we were able to obtain a total of 203 recordings for dark-eyed juncos, and 564 recordings for song sparrows from within the two analysis regions from 2006-2019. After examining each recording, we discarded any that misidentified the species, only contained calls, were noted in the recording metadata to contain juvenile or plastic song, or had excessive background noise. In cases where multiple recordings existed from the same recordist on the same date and at the same time or location, we used only one recording in order to avoid double-counting individual birds who may have been recorded more than once, unless it was clearly noted in the remarks that the recordings were of different individuals.

We supplemented these recordings with ones recorded in the field by one author (KTS) in July 2017. During this trip, KTS visited several publicly accessible locations in Tompkins County, within the region that had experienced the heaviest drought during 2016. This added an additional 8 dark-eyed junco and 7 song sparrow recordings that were usable. Recordings from before 2006 through June 2016 were categorized as before the drought. Anything recorded from July 2016 through December 2016 was considered to have occurred during the drought, and thus removed from analysis. Any recordings from 2017 or later were defined as after the drought. Due to the sparseness of recordings in public repositories from earlier years, we did not include any recordings prior to 2006.

Processing Song Recordings

Many of the recordings collected contained multiple song bouts from the same bird. For the purposes of this study, we define a bout as a continuous period of syllable production within the recording that is visibly separate from other periods of song. Within these bouts, syllables are separated from one another by silence, and discrete pulses of signal that are not separated by silence are termed notes (**Figure 4.2**).

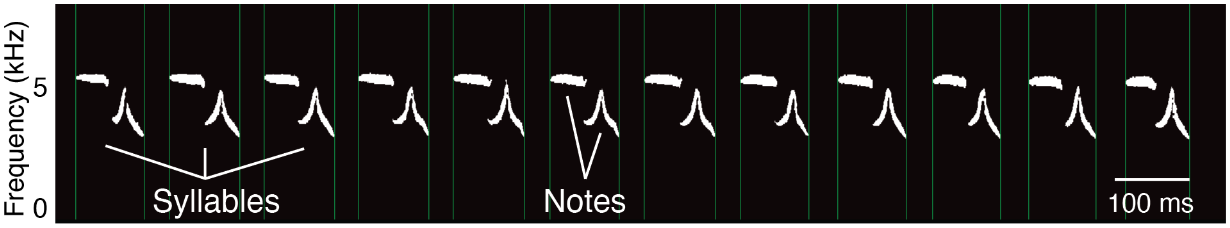


Figure 4.2: Example spectrogram of a single bout of dark-eyed junco song with parts of the song highlighted. Syllables are periods of signal separated by silences. Within these syllables, notes are the discrete pulses of signal that are not separated from one another by silences.

Individual bouts were extracted from recordings using Audacity version 2.4.2 and exported as separate WAV files with sampling rate set to 44100 Hz. Due to inherent differences between species song, different bout-selection protocols were implemented for juncos and song sparrows. Since dark-eyed juncos repeat the same song type many times in a row, most of the recordings in community-science databases are of a single song type (Cardoso et al. 2009; Williams and MacRoberts 1977). If there were multiple distinct song types in a single recording, we collected at least one bout of each type, but otherwise, we collected up to three bouts per recording, preferentially choosing ones that were least affected by background noise. Song sparrows, in contrast, cycle between song types more frequently (Wood and Yezerinac 2006). In order to obtain the greatest possible variation in songs, we parsed out every available bout from song sparrow recordings.

Syllable Segmentation in Chipper

Bouts were segmented into discrete syllables by a researcher blind to the region and year of each recording using the open source software Chipper (Searfoss, Pino, and Creanza 2020). Chipper uses a Fast Fourier Transform to generate a spectrogram from the WAV file of each song bout and then predicts syllable boundaries using fluctuations in the signal amplitude. The user can first implement high-pass and low-pass filters to reduce background noise and then modify the predicted syllable boundaries by adjusting parameters such as the signal-to-noise threshold, the minimum silence duration required to distinguish between syllables, and the minimum duration of syllables. Song bouts were discarded if they had a signal-to-noise ratio that made it difficult to reliably detect syllable boundaries. In addition, any files that had birds singing in the background that interfered with the song of the bird of interest in the recording were removed. Since dark-eyed juncos generally sing a repeated syllable, we could extract syllable properties even when only a subset of the song bout was usable; these files were excluded from analyses of two song features, number of syllables and bout duration.

The resulting spectrogram of each song is stored as a matrix in which (at a sampling rate of 44100 Hz) every row represents ~43 Hz and every column represents 0.317 milliseconds (Searfoss, Pino, and Creanza 2020). The value in each element of this matrix represents the signal intensity at that frequency and time; each element in the matrix becomes a pixel in the spectrogram image (e.g. **Figure 4.2**). After the user adjusts the signal-to-noise threshold, signal intensities above this value are retained in a binary spectrogram where ‘1’ or ‘0’ indicates that signal is present or absent at that frequency and time.

Chipper also executes an additional stage of noise reduction by measuring each note, defined as an area of continuous sound across the time and frequency axes, and discarding any that have an area in pixels less than a user-specified noise threshold. Chipper allows the user to empirically determine this average noise threshold (the minimum area in pixels that a note must be in order to not be discarded as noise) as well as an average syllable similarity threshold (the percent overlap that determines whether two syllables are considered the same or different). To assess these thresholds, we selected ~80-100 single-bout files for each species and manually set the noise threshold and syllable similarity threshold for each file. Across these test files, the average noise threshold for song sparrows was 59.1 pixels, and the average similarity threshold was 47.3% syllable overlap. The average noise threshold for dark-eyed juncos was 142 pixels, and the average similarity threshold was 29% syllable overlap. We used these average values as the thresholds in subsequent analyses. Occasionally, a syllable would only contain notes with areas below the noise threshold; we manually checked these recordings to decide whether the songs should be re-analyzed with different parameters or discarded.

Song analysis

After syllable segmentation, we used Chipper to measure a slightly different set of song features for dark-eyed juncos and song sparrows, choosing metrics that were more likely to be meaningful based on the structure and features of each species' song. For both species, we measured bout duration, total number of syllables, number of unique syllables, total number of notes, number of notes per syllable, mean syllable duration, standard deviation of syllable duration, rate of syllable production (calculated as the number of syllables divided by the duration of the song bout), degree of syllable repetition (calculated as the number of syllables divided by the number of unique syllables), overall frequency range, average maximum frequency of syllables, and average minimum frequency of syllables. For dark-eyed juncos, which generally sing multiple repetitions of the same syllable, we also measured the mean syllable stereotypy, defined as the mean syllable overlap value for syllables that were deemed to be repetitions of the same syllable. For song sparrows, which generally sing both short and long syllables, we also measured the duration of the shortest syllable and the duration of the longest syllable. Features that describe duration are measured in milliseconds, and features that describe frequency are measured in Hertz.

For each focal species and song feature, we used a set of Shapiro-Wilk tests to assess the distribution of our song-feature data before and after the drought in both the drought region and the control region. When at least one group in each comparison was not normally distributed (**Table S2**), we log-transformed the song-feature data. If the log-transformed values for that song feature were still not normally distributed, we analyzed the song features using non-parametric statistical tests, i.e. using a Wilcoxon rank-sum test instead of a *t*-test to compare data before versus after the drought in both the drought region and the control region, with a Holm-Bonferroni correction for multiple hypotheses. We also assessed the homogeneity of variance before versus after the drought in both regions using Brown-Forsythe tests. All statistical analyses were implemented in R with custom code, available at github.com/CreanzaLab/Birdsong_DroughtStress.

Results

Dark-eyed junco song analysis

We analyzed 403 dark-eyed junco bouts from 175 unique recordings in total (Table S1). We discarded 28 of our original 203 recordings; these were either recorded during the drought (5 recordings), recorded slightly outside the drought region, or had no analyzable song due to excessive noise or poor signal-to-noise ratio. In the drought region, we analyzed 46 recordings (131 bouts) from before the 2016 drought and 73 recordings (199 bouts) from after. From the control region, where the drought was less severe or absent, we analyzed 8 recordings (11 bouts) from before 2016 and 48 recordings (62 bouts) from after 2016.

Using Chipper, we segmented these dark-eyed junco song bouts into syllables and measured thirteen properties from each: bout duration, number of syllables, number of unique syllables, total number of notes, degree of syllable repetition (equal to the total number of syllables divided by the number of unique syllables per bout), mean number of notes per syllable, rate of syllable production, mean syllable duration, standard deviation of syllable duration, mean syllable stereotypy, frequency range, average maximum frequency of syllables, and average minimum frequency of syllables.

If a recording had multiple bouts, we took the mean value of each song feature across bouts so that each recording would be weighted equally in the comparison. For both the drought region and the control region, no song features were significantly different before vs. after the 2016 drought (**Figure 4.3, Table S4.2**). In addition, we did not reject the null hypothesis that each song feature had equal variance before versus after the drought (**Table S4.2**).

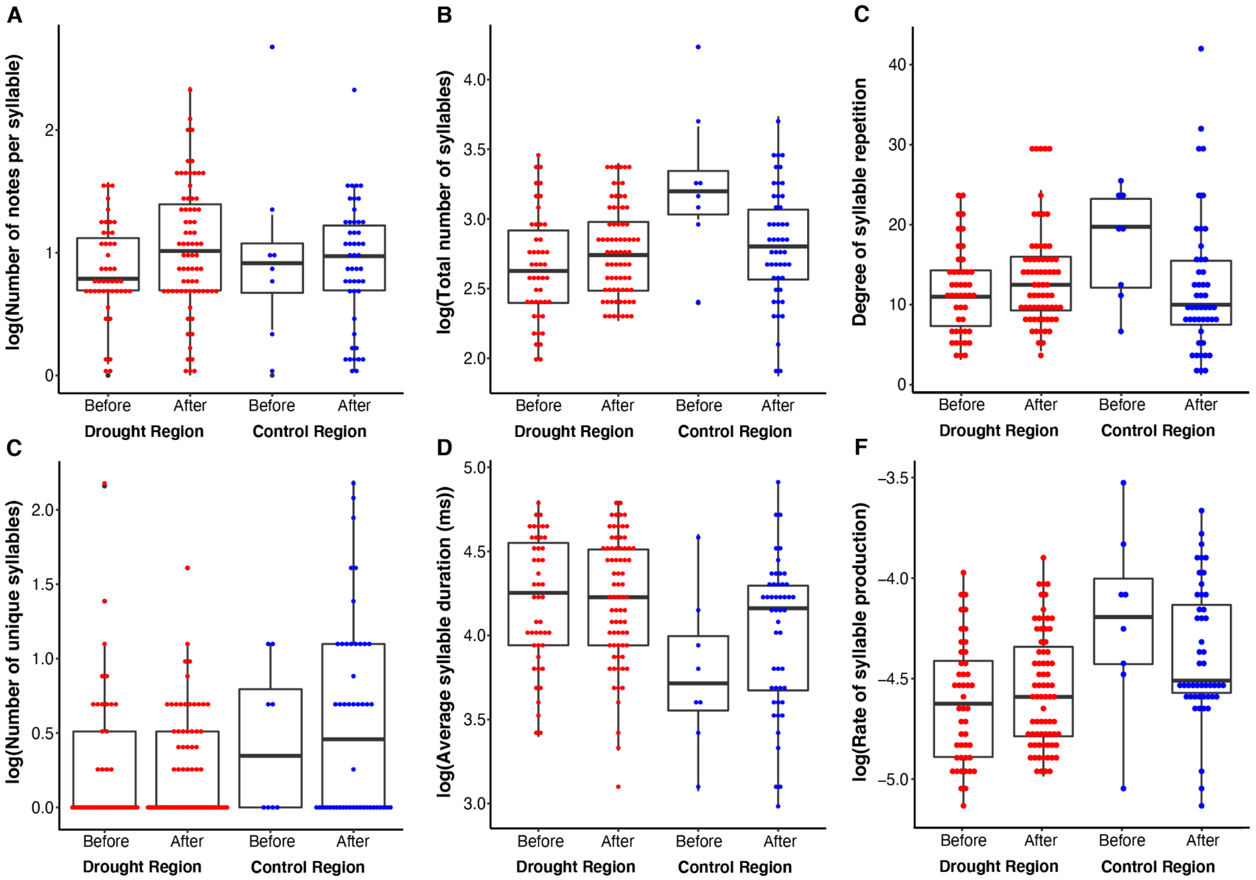


Figure 4.3. Song feature distributions within dark-eyed junco populations before and after the 2016 drought. Each boxplot shows dark-eyed junco mean song-feature values per recording: (A) number of notes per syllable, (B) number of syllables, (C) degree of repetition (number of syllables divided by the number of unique syllables), (D) number of unique syllables, (E) mean syllable duration, and (F) rate of syllable production.No dark-eyed junco song feature significantly differed in median or variance when we compared recordings before versus after the drought. Note that we log-transformed song features when they were not normally distributed (all panels except C). The midline of each boxplot represents the median, with the boxes representing the interquartile range (IQR). The whiskers are 1.5 * IQR. These specifications also apply to the box plots in Figure 4.6.

The dark-eyed junco sings a relatively simple song characterized by a single syllable repeated multiple times. We sought to visualize a 'typical' song from each of the four groups: before and after the drought from the drought region and before and after the drought from the control region. We identified the most representative bout for each group by first rescaling the song features with the largest differences from before to after the drought for each bout—the number of notes per bout, bout duration, and degree of repetition—so that their values would be comparable to one another (base R function 'scale'). We calculated the average value for each scaled feature among the bouts in each recording, and found the median value of each feature across recordings within each group. (**Figure 4.4**). We then identified and plotted the spectrograms of the bouts that were the smallest Euclidean distances from the point located at the median value of the three song features for each group (**Figure 4.5**).

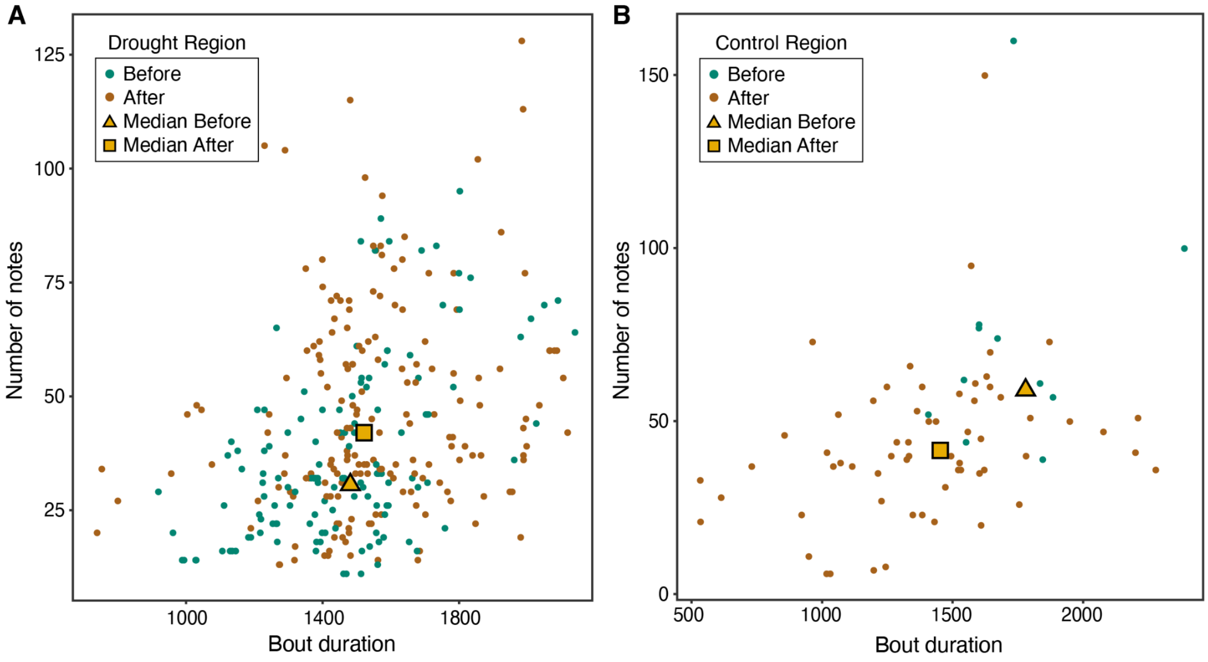


Figure 4.4. Visualizing dark-eyed junco song features. For every analyzed dark-eyed junco bout, we visualized the two song features with the largest differences before versus after the

drought, number of notes and bout duration, in both A) the drought region and B) the control region (**Table S4.2**), though these differences did not reach statistical significance with a Holm-Bonferroni correction. The yellow shapes indicate the median value of both of these song features.

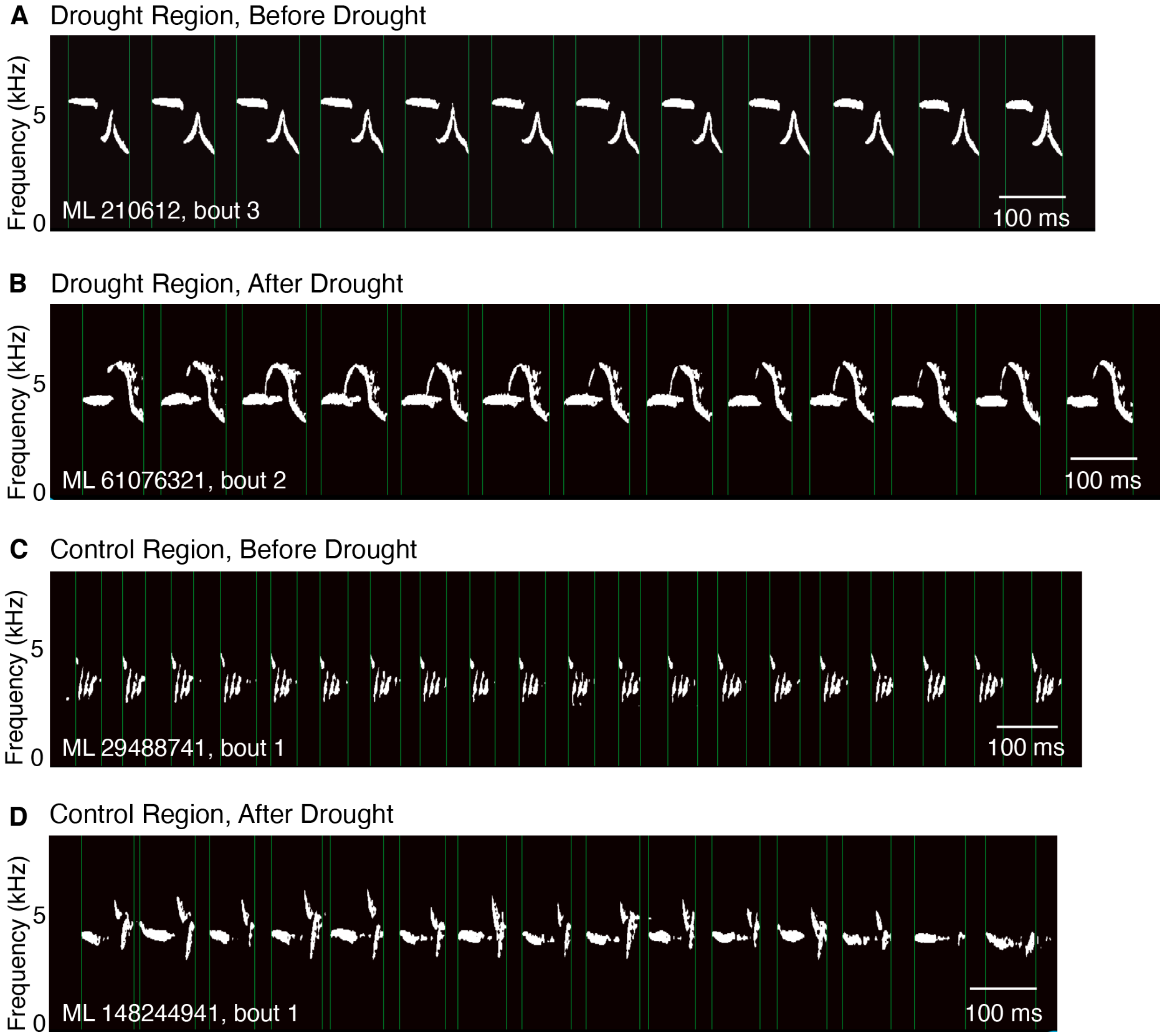


Figure 4.5. Spectrograms from representative dark-eyed junco songs. The songs that were closest to the median values of the three song features that showed the greatest difference before versus after the drought—the total number of notes, bout duration, and degree of repetition—are

shown for each region. The groups that these songs represent did not significantly differ from one another in any of the measured song features.

Song sparrow song analysis

We analyzed 1556 song sparrow bouts from 295 unique recordings in total (**Table S4.3**). We discarded 269 of our original 564 recordings; numerous recordings were submitted by the same recordist (or a group of recordists in the same location) on the same day, and we eliminated all but one from our sample unless we could confirm that the recordings were different birds based on the information provided by the recordists. In addition, we eliminated songs that were either recorded during the drought (7 recordings), recorded slightly outside the drought region, or had no analyzable song due to excessive background noise or poor signal-to-noise ratio. In the drought region, we analyzed 37 recordings (409 bouts) from before the 2016 drought and 159 recordings (745 bouts) from after. From the control region, where drought conditions reached only moderate severity, we analyzed 16 recordings (72 bouts) from before 2016 and 83 recordings (330 bouts) from after 2016.

For each of the fifteen song features we analyzed for song sparrows, we calculated the mean value across bouts in each recording and compared recordings before versus after the 2016 drought in both the drought and control regions. In the drought region only, three song features were significantly different after the 2016 drought: the number of syllables ($p < 2.6\times10^{-3}$), the number of notes per syllable ($p < 5.9\times10^{-4}$), and the degree of repetition ($p < 5.0\times10^{-5}$) (**Figure 4.6, Table S4.3**). However, as with the dark-eyed junco, we did not reject the null hypothesis that each song feature had equal variance before versus after the drought (**Table S4.3**).

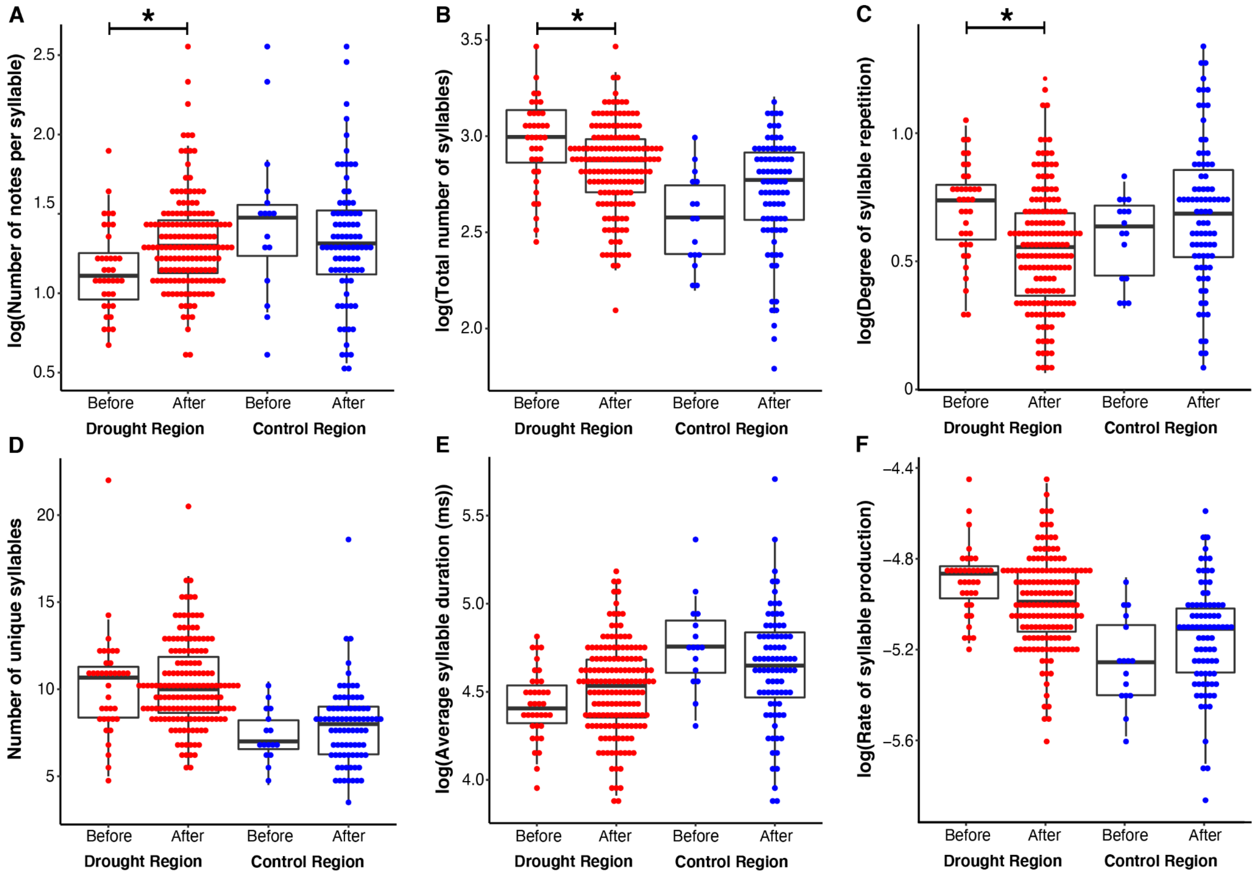


Figure 4.6: Song feature distributions within song sparrow populations before and after the 2016 drought. Each boxplot shows song sparrow mean song-feature values per recording: (A) number of notes per syllable, (B) number of syllables, (C) degree of repetition (number of syllables divided by the number of unique syllables), (D) number of unique syllables, (E) mean syllable duration, and (F) rate of syllable production. The number of notes per syllable, number of syllables, and degree of repetition (A-C) were significantly different from before to after the drought in the drought region only, with Bonferroni-Holm correction for multiple hypothesis testing. In the drought region, the difference in the rate of syllable production from before to after the drought had a Wilcoxon rank-sum test *p* = 0.0074, but was not significant after Bonferroni-Holm correction for multiple hypothesis testing. The number of unique syllables was

normally distributed (Shapiro-Wilk test $p > 0.05$) in both time periods and did not need to be log-transformed. (**Table S4.3**)

Song sparrows sing a relatively complex song, particularly for closed-ended learners (Marler and Peters 1987). Many of their syllables are composed of multiple notes, and they cycle through their syllable repertoire of approximately 35-38 syllables by singing different song types containing overlapping subsets of their syllable repertoire. To capture this variation, we analyzed as many bouts as possible from each individual recording. Bout-level data from the two song features that showed the strongest temporal difference in the drought region, the number of notes per syllable and the number of syllables per number of unique syllables, are shown in **Figure 4.7**. To visualize a 'typical' song bout from each region and each time period, we rescaled the three significantly different song features so that they would be comparable with one another, and we found the song bout that was closest to the median value of all three features (**Figure 4.8**).

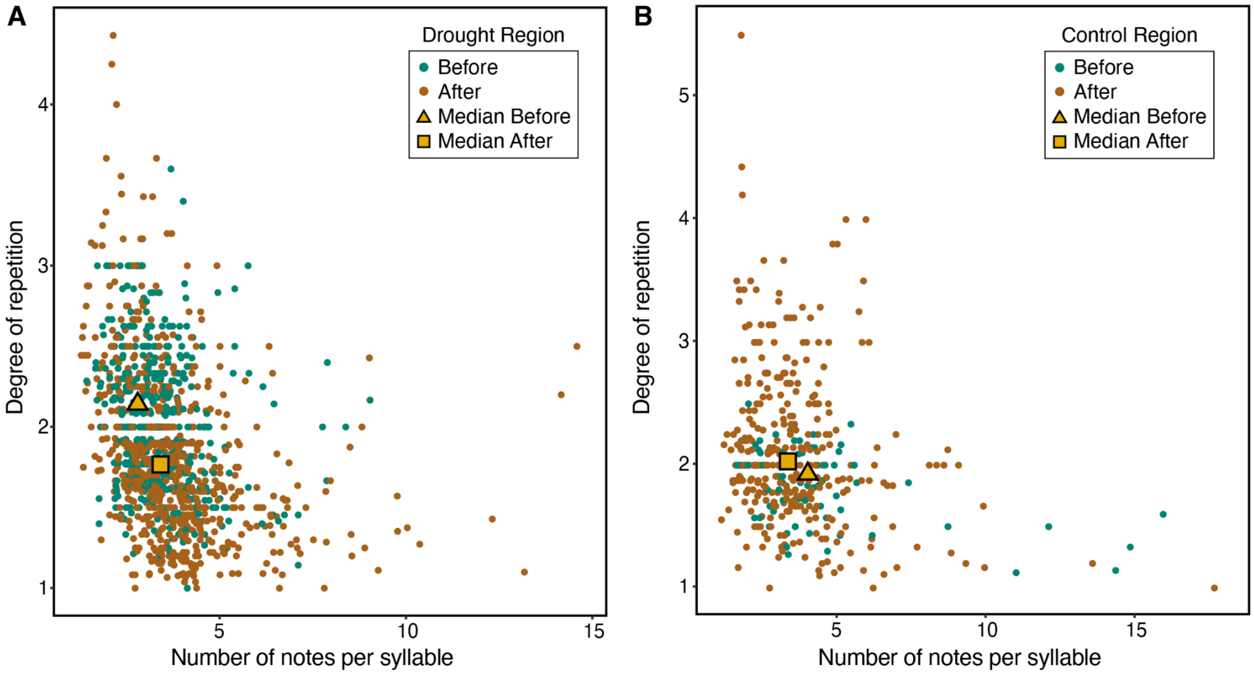


Figure 4.7. Visualizing song sparrow song features. For every analyzed song sparrow bout, we visualized the two song features with the largest differences before versus after the drought, number of notes per syllable and degree of repetition (**Table S4.3**), both of which were significantly different in the two time periods in the drought region (A) but not in the control region (B). After the drought, song sparrow songs on average had less syllable repetition and more notes per syllable. There were no significant differences in the control region.

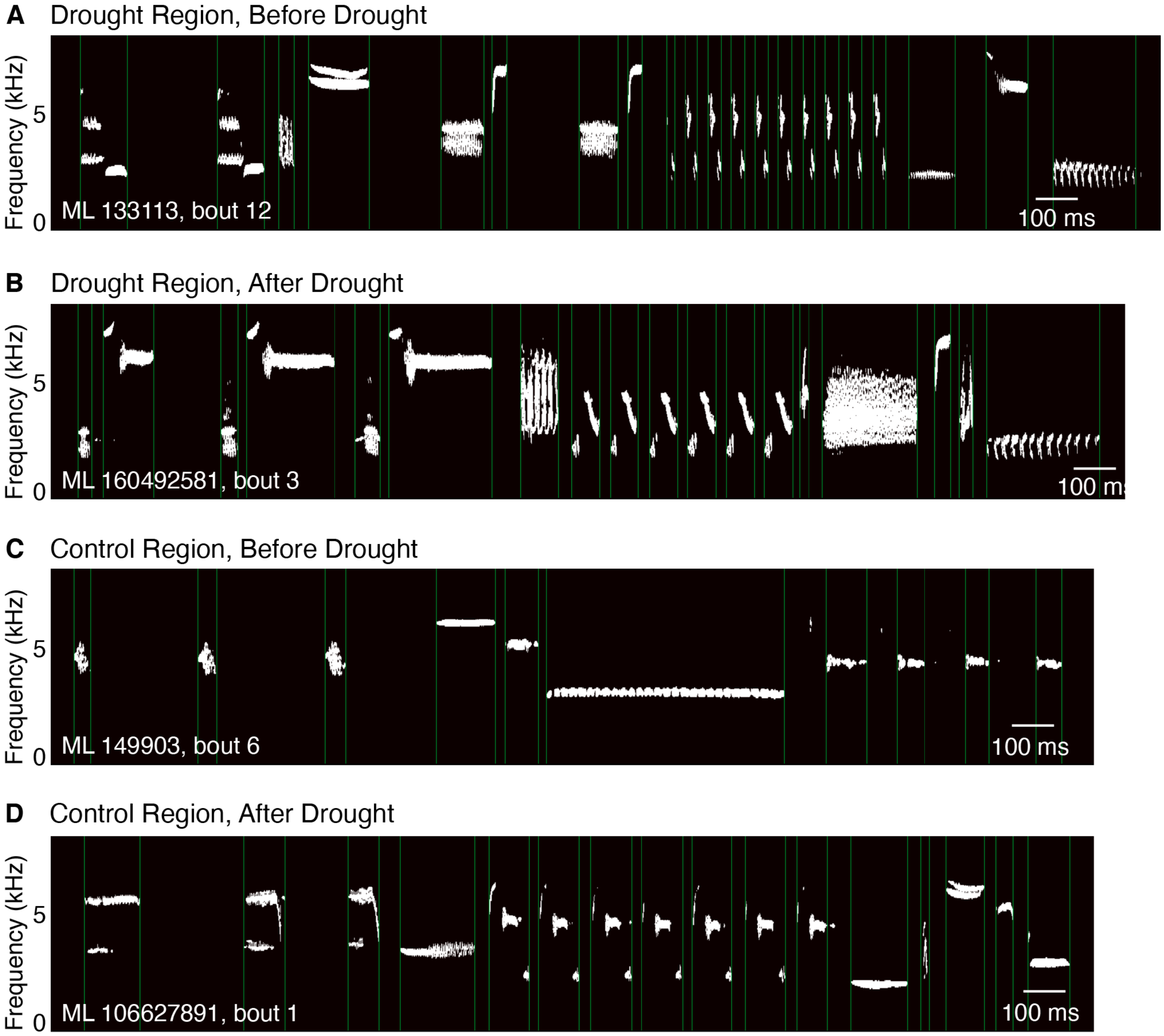


Figure 4.8. Spectrograms from representative song sparrow songs. These songs were closest to the median rescaled values of the three song features that showed significant differences

before versus after the drought in the drought region, as in **Figure 4.7**. Compared to before the drought, song sparrow songs after the drought had significantly more notes per syllable, fewer total syllables per bout, and a lower degree of repetition. Here, the 'median' song from before the drought has 21 syllables and 10 unique syllables, with a degree of syllable repetition of 2.1; the 'median' song after the drought has 18 syllables and 10 unique syllables, thus a degree of syllable repetition of 1.8.

Discussion

Here, we analyze community-science birdsong recordings from two common species of songbird, the dark-eyed junco and the song sparrow. In particular, we compare the time period before and after a severe drought in western New York, and compare these results to songs from control regions, in which drought was absent or only moderate, over the same time period (**Figure 4.1**). Because physiologically stressful conditions during development can disrupt song-learning on an individual level, we hypothesized that a population-wide stressor could have measurable effects on the song distribution of a population. For dark-eyed juncos, a species that generally sings a simple song of one repeated syllable, we observed no significant changes in any measured song feature in either the drought region or the control region. In the song sparrow, a species that has a relatively large syllable repertoire and produces complex multi-note syllables, we observed significant differences in song features related to both syllable complexity (notes per syllable) and song complexity (number of syllables, degree of repetition) in the drought region, but not the control region. Specifically, song sparrows after the drought sang songs with more notes per syllable, less repetition, and fewer syllables overall. It is notable that we see these significant effects in the species with more complex songs, since we hypothesized that species with complex songs would be more sensitive to the potential stress-associated learning disruption. However, the behavioral consequences of the changes we observed are not completely clear; more notes per syllable after the drought is suggestive that syllables in the later group might be more complex, whereas fewer overall syllables per bout could suggest reduced song complexity. The reduction in the degree of repetition after the drought (**Figure 4.6C**) is less straightforward, but when taken in concert with the trends towards longer average syllable durations and a lower rate of syllable production (**Figures 4.6E, F**), it may indicate that the shift occurs most prominently in the trill section that exists in most song sparrow songs, in which a single syllable is rapidly repeated on average 7-9 times (Borror 1965): a smaller number of syllable repetitions in the trill would decrease the overall degree of repetition and, if the trill was approximately the same duration, the rate of syllable production in the trill. We performed a follow-up analysis to check whether this interpretation is borne out in the data by obtaining for each bout the number of times the most-repeated syllable type was produced in that bout. We averaged these values across all bouts within each recording to obtain, for each recording, the "mean number of repetitions of the most-repeated syllable type per bout" and found that this was significantly lower after the drought in the drought region only **(Figure S4.2)**. This provides further evidence that trills may have meaningfully changed due to the drought, becoming either shorter in duration or slower in the rate of syllable production. Notably, a lower rate of syllable production could be indicative of poorer performance: though the characteristics that determine trill quality haven't been studied intraspecifically in song sparrows, trills, especially ones that cover large frequency ranges, are considered to be physically difficult to perform across species, and thus a higher rate of trill production is considered to be indicative of higher quality or greater skill (Podos 1997). There is evidence from several other species that trill rate is important in female choice, territory defense, and intrasexual conflict (Ballentine, Hyman, and Nowicki 2004; Cramer 2012; Caro et al. 2010; Searcy et al. 2010). Further, there is some evidence that rate of syllable production may be affected by developmental history in swamp sparrows, a close relative of song sparrows (Searcy et al. 2010).

Song sparrows have a long study history in the field of songbird neuroethology, providing additional context for these observations. There have been several lab-based song-preference studies in wild-caught song sparrows that have found that females exhibit stronger preferences for familiar song types than unfamiliar types (O'Loghlen and Beecher 1997; O'Lochlen and Beecher 1999; Hernandez et al. 2009) and for larger song repertoires (Searcy 1984). In field studies, males with larger song repertoires held territories longer and experienced greater reproductive success (Hiebert, Stoddard, and Arcese 1989), as well as had a higher probability of mating in their first breeding year (Reid et al. 2004), although other studies have not found such relationships between song repertoire and reproductive success in the field (Searcy 1984). Studies investigating stress and song in song sparrows have shown that treatment with corticosterone, the hormone that mediates the stress response in birds, or food restriction during development did not lead to a difference in trill rate or song-type stereotypy but did lead to smaller syllable and song repertoires (Schmidt et al. 2013). Further, adult assays of stress reactivity found a negative correlation between the magnitude of the acute stress response and song complexity as a combined measure of syllable and song repertoires (Schmidt et al. 2012). Together, these findings suggest that, in song sparrows, metrics of song complexity, such as song repertoire, may be both more biologically meaningful and more responsive to early-life stress than metrics of song performance, like the rate of syllable production.

The high baseline level of complexity of song sparrow songs may prevent us from drawing solid conclusions about biological causes and consequences of any shifts in song using only mean song feature values across recordings, since, for example, dramatic changes in syntax or syllable composition could occur but not be reflected in the average values we analyzed. However, the observed shifts in certain features may hint at more nuanced changes happening in song or syllable structure. Future analysis of individual syllable types could probe these facets of song in greater depth.

There are several ways in which this study system and site may not be ideal for assessing song changes on a population scale. As with all community-science-based analyses, we lack metadata that would be useful in interpreting our analyses, such as the age of the bird in the recording. One potential challenge of studying severe environmental stress is that members of a population might conserve resources by reducing their reproductive output, for example by abandoning nest-building activities or failing to provision chicks. In the context of this study, this would mean that there would be very few juveniles developing during the height of the drought, and thus a minimal effect on the represented songs in the following years. However, the timing of the drought might have mitigated this challenge; since the drought did not reach its highest level of severity until mid-summer, many species would likely have already bred and raised chicks to at least fledgling age, increasing the likelihood that enough juveniles were exposed to drought stress to significantly affect population song.

To detect shifts in song that could potentially be linked to impaired learning due to the drought, as opposed to occurring due to normal stochasticity in population song, we compared the drought region to a control region; however, the best-sampled control regions we could find still had much smaller numbers of recordings than the drought region, reducing our power to detect any potential shifts in the songs in these control regions. Further, the only reasonably well-sampled potential control regions also experienced abnormally dry and moderate drought conditions, respectively, during 2016, since this drought was a regional phenomenon that affected much of the northeast United States.

Song in a population could shift significantly after an environmental stressor for reasons other than impaired learning. First, if there was heightened mortality in the population, songs could undergo a bottleneck effect akin to the stochastic loss of genetic diversity after extreme population reduction. In such a case, we might expect to see a narrower distribution of one or more of the song features after the reduction in population diversity that occurs during this type of bottleneck event. We did not observe such a shift in variance in any song features. Alternatively, if song features were correlated with another trait that was under selection pressure during drought conditions, for example body size, there could be indirect effects on song independent of learning ability. Second, a drought could induce individuals to emigrate from the region in pursuit of more hospitable conditions, or to not return to it after migration. If these individuals did not return in the following year once drought conditions had abated, others may disperse into those territories, changing the region's song distribution by changing the members of the region. Third, adult songs, despite being crystallized and thus not expected to radically change, could theoretically be different during drought conditions, for example if the ambient temperature and humidity altered airflow through the syrinx or sound transmission through the air (Pandit, Bridge, and Ross 2022). In this case, juveniles perceiving those song differences could hypothetically learn the song as perceived through that "filter" and propagate it to future pupils.

Other systems may prove more robust for testing the hypotheses and approaches we propose here. An ideal study system would likely be a long-term field study on a isolated or island population where historical data exist on rates of mortality and recruitment per year, especially during years when a stressful environmental event occurs, ample historical recordings of songs from years prior to the environmental stressor, and environmental data including water availability and a survey of available nutrient sources. If the species predominantly exhibits high-fidelity vertical learning, or learning from parent to offspring with little improvisation on the part of the pupil, having data on parentage and identity-linked song recordings would allow for comparisons between songs of the tutor and pupil directly. By measuring the similarity between these songs, it would be possible to more directly attribute any changes in song to mistakes in learning and track the frequency of mistakes over stressful and non-stressful years.

Studying a learned behavior, such as song, in the context of population-wide evolutionary pressures is particularly interesting since the resulting cultural evolutionary dynamics could be significantly more rapid than genetic evolution. Even without selection for a novel song type, the combination of drift and oblique transmission—from an adult to an unrelated juvenile—could lead to fixation of novel song variants far more rapidly than possible with a genetic mutation. Whether or not transient learning deficits in a subset of a population could induce a persistent shift in song, and therefore influence the long-term trajectory of the population, likely would depend on several factors.

Here, we present a study of the songs of two well-recorded species in a region that experienced a severe drought compared to a region that did not, and we observe a suite of song changes before versus after the time of the drought in only the species with the more complex song. If the trends we observed in the three years since the drought were to persist and a population's song underwent a lasting change, there could be significant implications for conservation, particularly in the face of climate change that is likely to make severe weather events more common. If these changes were particularly salient in a given species' song—for example, if the song changed in features that are important for mate choice—a population's song could theoretically become unattractive to a sister population that maintained the typical ancestral song. This would suggest a potential side effect of anthropogenic climate change that has not been studied: that the sublethal experiences of individuals could impact the development of behaviors that are culturally transmitted to future generations, leading to an increase in reproductive isolation by sexual selection of this trait that is both plastic and sensitive to hostile conditions (Mendelson and Safran 2021).

References

Ballentine, Barbara, Jeremy Hyman, and Stephen Nowicki. 2004. “Vocal Performance Influences Female Response to Male Bird Song: An Experimental Test.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 15 (1): 163–68.

Boag, P. T., and P. R. Grant. 1984. “The Classical Case of Character Release: Darwin’s Finches (Geospiza) on Isla Daphne Major, Galápagos.” *Biological Journal of the Linnean Society. Linnean Society of London* 22 (3): 243–87.

Borror, Donald J. 1965. “Song Variation in Maine Song Sparrows.” *The Wilson Bulletin* 77 (1): 5–37.

Boyce, W. Thomas, Marla B. Sokolowski, and Gene E. Robinson. 2020. “Genes and Environments, Development and Time.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (38): 23235–41.

Brass, Dominic P., Christina A. Cobbold, David A. Ewing, Bethan V. Purse, Amanda Callaghan, and Steven M. White. 2021. “Phenotypic Plasticity as a Cause and Consequence of Population Dynamics.” *Ecology Letters* 24 (11): 2406–17.

Brenowitz, Eliot A., and Michael D. Beecher. 2005. “Song Learning in Birds: Diversity and Plasticity, Opportunities and Challenges.” *Trends in Neurosciences* 28 (3): 127–32.

Brown, Charles R., and Mary Bomberger Brown. 1998. “INTENSE NATURAL SELECTION ON BODY SIZE AND WING AND TAIL ASYMMETRY IN CLIFF SWALLOWS DURING SEVERE WEATHER.” *Evolution; International Journal of Organic Evolution* 52 (5): 1461–75.

Brumm, Henrik, Sue Anne Zollinger, and Peter J. B. Slater. 2009. “Developmental Stress Affects Song Learning but Not Song Complexity and Vocal Amplitude in Zebra Finches.” *Behavioral Ecology and Sociobiology* 63 (9): 1387–95.

Cardoso, Gonçalo C., Jonathan W. Atwell, Ellen D. Ketterson, and Trevor D. Price. 2009. “Song Types, Song Performance, and the Use of Repertoires in Dark-Eyed Juncos (Junco Hyemalis).” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 20 (4): 901–7.

Caro, Samuel P., Kendra B. Sewall, Katrina G. Salvante, and Keith W. Sockman. 2010. “Female Lincoln’s Sparrows Modulate Their Behavior in Response to Variation in Male Song Quality.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 21 (3): 562–69.

Caspi, Tal, Jacob R. Johnson, Max R. Lambert, Christopher J. Schell, and Andrew Sih. 2022. “Behavioral Plasticity Can Facilitate Evolution in Urban Environments.” *Trends in Ecology & Evolution* 37 (12): 1092–1103.

Catchpole, C. K., J. Dittami, and B. Leisler. 1984. “Differential Responses to Male Song Repertoires in Female Songbirds Implanted with Oestradiol.” *Nature* 312 (5994): 563–64.

Cramer, Emily R. A. 2012. “Vocal Deviation and Trill Consistency Do Not Affect Male Response to Playback in House Wrens.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 24 (2): 412–20.

Creanza, Nicole, Laurel Fogarty, and Marcus W. Feldman. 2016. “Cultural Niche Construction of Repertoire Size and Learning Strategies in Songbirds.” *Evolutionary Ecology* 30 (2): 285–305.

Del Hoyo, Josep, Josep Del Hoyo, Andrew Elliott, and Jordi Sargatal. 1992. *Handbook of the Birds of the World*. Vol. 1. Lynx edicions Barcelona.

Donihue, Colin M., Alex M. Kowaleski, Jonathan B. Losos, Adam C. Algar, Simon Baeckens, Robert W. Buchkowski, Anne-Claire Fabre, et al. 2020. “Hurricane Effects on Neotropical Lizards Span Geographic and Phylogenetic Scales.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (19): 10429–34.

Ferree, Elise D. 2013. “Geographic Variation in Morphology of Dark-Eyed Juncos and Implications for Population Divergence.” *The Wilson Journal of Ornithology* 125 (3): 454–70.

Gienapp, Phillip, Veronika N. Laine, A. C. Mateman, Kees van Oers, and Marcel E. Visser. 2017. “Environment-Dependent Genotype-Phenotype Associations in Avian Breeding Time.” *Frontiers in Genetics* 8 (August): 102.

Grant, Barbara Rosemary, and Peter Raymond Grant. 1993. “Evolution of Darwin’s Finches Caused by a Rare Climatic Event.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 251 (1331): 111–17.

Grant, Peter R., and B. Rosemary Grant. 2011. “Causes of Lifetime Fitness of Darwin’s Finches in a Fluctuating Environment.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (2): 674–79.

Henry, Lucas P., Marjolein Bruijning, Simon K. G. Forsberg, and Julien F. Ayroles. 2021. “The Microbiome Extends Host Evolutionary Potential.” *Nature Communications* 12 (1): 5141.

Hernandez, Alexandra M., Jeremy A. Pfaff, Elizabeth A. MacDougall-Shackleton, and Scott A. MacDougall-Shackleton. 2009. “The Development of Geographic Song Preferences in Female Song Sparrows Melospiza Melodia.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 115 (6): 513–21.

Hiebert, Sara M., Philip K. Stoddard, and Peter Arcese. 1989. “Repertoire Size, Territory Acquisition and Reproductive Success in the Song Sparrow.” *Animal Behaviour* 37 (February): 266–73.

Huberty, Andrea F., and Robert F. Denno. 2004. “PLANT WATER STRESS AND ITS CONSEQUENCES FOR HERBIVOROUS INSECTS: A NEW SYNTHESIS.” *Ecology*. https://doi.org/10.1890/03-0352.

Hughes, Melissa, Stephen Nowicki, William A. Searcy, and Susan Peters. 1998. “Song-Type Sharing in Song Sparrows: Implications for Repertoire Function and Song Learning.” *Behavioral Ecology and Sociobiology* 42 (6): 437–46.

Kozlovsky, Dovid Y., Carrie L. Branch, Angela M. Pitera, and Vladimir V. Pravosudov. 2018. “Fluctuations in Annual Climatic Extremes Are Associated with Reproductive Variation in Resident Mountain Chickadees.” *Royal Society Open Science* 5 (5): 171604.

Kroodsma, Donald E., and Linda D. Parker. 1977. “Vocal Virtuosity in the Brown Thrasher.” *The Auk* 94 (4): 783–85.

Langin, Kathryn M., T. Scott Sillett, Jongmin Yoon, Helen R. Sofaer, Scott A. Morrison, and Cameron K. Ghalambor. 2009. “Reproductive Consequences of an Extreme Drought for Orange-Crowned Warblers on Santa Catalina and Santa Cruz Islands.” In *Proceedings of the Seventh California Islands Symposium*. Institute for Wildlife Studies.

Liebgold, Eric B., Nicole M. Gerlach, and Ellen D. Ketterson. 2013. “Similarity in Temporal Variation in Sex-Biased Dispersal over Short and Long Distances in the Dark-Eyed Junco, Junco Hyemalis.” *Molecular Ecology* 22 (22): 5548–60.

MacDougall-Shackleton, Scott A., and Karen A. Spencer. 2012. “Developmental Stress and Birdsong: Current Evidence and Future Directions.” *Journal of Ornithology / DO-G* 153 (1): 105–17.

Marler, Peter, and Donald Isaac. 1960. “Physical Analysis of a Simple Bird Song as Exemplified by the Chipping Sparrow.” *The Condor* 62 (2): 124–35.

Marler, Peter, and Susan Peters. 1987. “A Sensitive Period for Song Acquisition in the Song Sparrow, Melospiza Melodia: A Case of Age-Limited Learning.” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 76 (2): 89–100.

Mendelson, Tamra C., and Rebecca J. Safran. 2021. “Speciation by Sexual Selection: 20 Years of Progress.” *Trends in Ecology & Evolution* 36 (12): 1153–63.

Newman, Melissa M., Pamela J. Yeh, and Trevor D. Price. 2008. “Song Variation in a Recently Founded Population of the Dark-Eyed Junco (junco Hyemalis).” *Ethology: Formerly Zeitschrift Fur Tierpsychologie* 114 (2): 164–73.

Nowicki, Stephen, and William A. Searcy. 2004. “Song Function and the Evolution of Female Preferences: Why Birds Sing, Why Brains Matter.” *Annals of the New York Academy of Sciences* 1016 (June): 704–23.

O’Lochlen, A. L., and M. D. Beecher. 1999. “Mate, Neighbour and Stranger Songs: A Female Song Sparrow Perspective.” *Animal Behaviour* 58 (1): 13–20.

O’Loghlen, A. L., and M. D. Beecher. 1997. “Sexual Preferences for Mate Song Types in Female Song Sparrows.” *Animal Behaviour* 53 (4): 835–41.

Pandit, Meelyn M., Eli S. Bridge, and Jeremy D. Ross. 2022. “Environmental Conditions Lead to Shifts in Individual Communication, Which Can Cause Cascading Effects on Soundscape

Composition." *Ecology and Evolution* 12 (10): e9359.

Podos, Jeffrey. 1997. "A PERFORMANCE CONSTRAINT ON THE EVOLUTION OF TRILLED VOCALIZATIONS IN A SONGBIRD FAMILY (PASSERIFORMES: EMBERIZIDAE)." *Evolution; International Journal of Organic Evolution* 51 (2): 537–51.

Price, T. D., P. R. Grant, H. L. Gibbs, and P. T. Boag. 1984. "Recurrent Patterns of Natural Selection in a Population of Darwin's Finches." *Nature* 309 (5971): 787–89.

Prior, Nora H., Sarah A. Heimovics, and Kiran K. Soma. 2013. "Effects of Water Restriction on Reproductive Physiology and Affiliative Behavior in an Opportunistically-Breeding and Monogamous Songbird, the Zebra Finch." *Hormones and Behavior* 63 (3): 462–74.

Regan, Charlotte E., Kristina B. Beck, Keith McMahon, Sam Crofts, Josh A. Firth, and Ben C. Sheldon. 2022. "Social Phenotype-Dependent Selection of Social Environment in Wild Great and Blue Tits: An Experimental Study." *Proceedings. Biological Sciences / The Royal Society* 289 (1986): 20221602.

Reid, Jane M., Peter Arcese, Alice L. E. V. Cassidy, Sara M. Hiebert, James N. M. Smith, Philip K. Stoddard, Amy B. Marr, and Lukas F. Keller. 2004. "Song Repertoire Size Predicts Initial Mating Success in Male Song Sparrows, Melospiza Melodia." *Animal Behaviour* 68 (5): 1055–63.

Roberts, L. Jay, Ryan Burnett, James Tietz, and Sam Veloz. 2019. "Recent Drought and Tree Mortality Effects on the Avian Community in Southern Sierra Nevada: A Glimpse of the Future?" *Ecological Applications: A Publication of the Ecological Society of America* 29 (2): e01848.

Robinson, Cristina M., Kate T. Snyder, and Nicole Creanza. 2019. "Correlated Evolution between Repertoire Size and Song Plasticity Predicts That Sexual Selection on Song Promotes Open-Ended Learning." *eLife* 8 (September). https://doi.org/10.7554/eLife.44454.

Schmidt, Kim L., Ainsley A. Furlonger, Janet M. Lapierre, Elizabeth A. MacDougall-Shackleton, and Scott A. MacDougall-Shackleton. 2012. "Regulation of the HPA Axis Is Related to Song Complexity and Measures of Phenotypic Quality in Song Sparrows." *Hormones and Behavior* 61 (4): 652–59.

Schmidt, Kim L., S. Drew Moore, Elizabeth A. MacDougall-Shackleton, and Scott A. MacDougall-Shackleton. 2013. "Early-Life Stress Affects Song Complexity, Song Learning and Volume of the Brain Nucleus RA in Adult Male Song Sparrows." *Animal Behaviour* 86 (1): 25–35.

Searcy, William A. 1984. "Song Repertoire Size and Female Preferences in Song Sparrows." *Behavioral Ecology and Sociobiology* 14 (4): 281–86.

Searcy, William A., Susan Peters, Silke Kipper, and Stephen Nowicki. 2010. "Female Response to Song Reflects Male Developmental History in Swamp Sparrows." *Behavioral Ecology and Sociobiology* 64 (8): 1343–49.

Searfoss, Abigail M., James C. Pino, and Nicole Creanza. 2020. "Chipper: Open‐source

Software for Semi‑automated Segmentation and Analysis of Birdsong and Other Natural Sounds." *Methods in Ecology and Evolution / British Ecological Society* 11 (4): 524–31.

Sepers, Bernice, Krista van den Heuvel, Melanie Lindner, Heidi Viitaniemi, Arild Husby, and Kees van Oers. 2019. "Avian Ecological Epigenetics: Pitfalls and Promises." *Journal of Ornithology / DO-G* 160 (4): 1183–1203.

Svoboda, Mark, Doug LeComte, Mike Hayes, Richard Heim, Karin Gleason, Jim Angel, Brad Rippey, et al. 2002. "THE DROUGHT MONITOR." *Bulletin of the American Meteorological Society* 83 (8): 1181–90.

Sweet, Shannan K., David W. Wolfe, Arthur DeGaetano, and Rebecca Benner. 2017. "Anatomy of the 2016 Drought in the Northeastern United States: Implications for Agriculture and Water Resources in Humid Climates." *Agricultural and Forest Meteorology* 247 (December): 571–81.

Weatherhead, Patrick J., and Mark R. L. Forbes. 1994. "Natal Philopatry in Passerine Birds: Genetic or Ecological Influences?" *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 5 (4): 426–33.

White, T. C. R. 1984. "The Abundance of Invertebrate Herbivores in Relation to the Availability of Nitrogen in Stressed Food Plants." *Oecologia* 63 (1): 90–105.

Williams, Laidlaw, and Michael H. MacRoberts. 1977. "Individual Variation in Songs of Dark-Eyed Juncos." *The Condor* 79 (1): 106–12.

Wingfield, John C., Kimberley Sullivan, Jessica Jaxion-Harm, and Simone L. Meddle. 2012. "The Presence of Water Influences Reproductive Function in the Song Sparrow (Melospiza Melodia Morphna)." *General and Comparative Endocrinology* 178 (3): 485–93.

Wood, William E., and Stephen M. Yezerinac. 2006. "Song Sparrow (Melospiza Melodia) Song Varies with Urban Noise." *The Auk* 123 (3): 650–59.

CHAPTER 5

CONCLUSIONS: INTERACTIONS BETWEEN SONG, SEXUAL SELECTION, LEARNING, AND ENVIRONMENT AND RECOMMENDATIONS FOR THE FUTURE OF BIRDSONG RESEARCH

Birdsong is among the most complex and diverse auditory signals that exist in biology. The overarching goal of this thesis was to demonstrate how birdsong has reached this profound diversity via the consistent interaction of two key factors: sexual selection and learning. I have approached this task with diverse studies that probe the extremes of evolutionary history for this clade, from predicting the song properties of the most recent common ancestor of Oscine songbirds to exploring whether detectable changes in song patterns emerge within a single population as a result of individuals coping with an adverse environment through the stress response, an even more ancient physiological adaptation.

In Chapter 2, I showed, using non-monogamous mating systems as a proxy, that increased sexual selection does not drive species inexorably towards more elaborate songs via Fisherian runaway, but instead influences song evolution within a more complex reality of more subtle factors, such as differing rates of evolution, that together lead to instability in song elaboration over evolutionary time. In Chapter 3, I elucidated the identity and effect of one of these other factors, the length of the song-learning window. I demonstrated that extended song plasticity and song elaboration coevolve in a pattern of mutual reinforcement; either combination of one without the other is likely to have been evolutionarily unstable. However, the evolutionary processes and selection pressures that we suppose have acted upon these two features are distinct. Song elaboration is not genetically hard-wired, but is instead the product of cultural transmission and perhaps a genetic predisposition in the form of an innate song template, and the emergent behavior of a complex song is the feature that is actually exposed to sexual selection. On the other hand, the learning window is genetically encoded, with apparently little variation within species; however, there is no evidence that learning window itself is under sexual selection. An extended learning window might be directly selected against under natural selection if there are costs, for example metabolic costs, associated with it, but we can reasonably conclude that learning window evolution is otherwise driven by selection pressures on song. Sexual selection could thus indirectly drive learning window in either direction; positively, if song complexity is favored, or negatively, for example if songs lost recognizability due to excessive drift in features over an extended learning period. Indeed, our findings hinted at this pattern that sexual selection on song drives the extension of adult song plasticity.

While Chapters 2 and 3 were based on cross-species analyses to elucidate past interactions of these traits, in Chapter 4, I focused on the micro-evolutionary patterns of song change on the population scale. Any trend that occurs in evolution necessarily emerges after traveling through a long line of successive and perpetually changing, fusing, and splitting populations. This study was an endeavor to capture a snapshot of accelerated evolution in action in a novel way. This was based on a proposal that emergent behaviors may be linked over several scales: that hostile conditions change physiology, which responds adaptively to the proximate threat, but at a cost of potentially maladaptive side effects like impaired learning. This suggests another mechanism by which the environment may influence evolution: by influencing learning, and therefore, indirectly, learned behaviors. Interactions with biotic and abiotic factors in the environment form the basis of natural selection in its simplest form, survival against the elements. This proposed mechanism acts on evolution based on individuals' ability to thrive against the elements and therefore gain opportunities to reproduce, linking environment to sexual selection. Chapter 4 attempts to shed light on a single connection in this web of interacting phenomena, from environmental pressure to song, a behavior not directly linked to that pressure.

The Future of Large-scale Song Analysis

An essential part of this body of work is the use of birdsong recordings collected by community scientists and made freely and publicly available. These are most central to our work in Chapter 4, where community science song recordings are the primary data we used in the analyses. Additionally, I contributed recordings of my own for these analyses. While there are song recordings available from before 2017, the density of available recordings is generally quite low for any given species except near a few hotspots, so sampling to detect song differences over space is largely ineffective.

However, the number and geographic distribution of recordings has increased significantly in recent years, so future studies will be much more likely to be able to ask nuanced questions about population songs, dialects, song dispersal, and cultural transmission, as well as questions about changes over time on yearly or even monthly scales. For example, where certain migratory populations and species overwinter remains enigmatic for the vast majority of populations, despite this being extremely biologically relevant for these species. As community science sightings and recordings become more common in the Global South, two questions may become answerable: 1) Which species continue to sing in their overwintering grounds? 2) Where are the overwintering grounds of populations and species?

It is possible that there is more diversity in song learning between species than is currently known because there are several axes of song variation that are not well characterized in large numbers of species, for example, syntax, the fraction of syllables that are shared, and the fraction of whole songs that are shared locally. In song sparrows, it has been shown that in an Eastern U.S. population, songs are composed of syllables acquired from multiple tutors, while in a Western U.S. population, birds tend to learn each song entirely from one individual tutor (Hughes et al. 1998). What might drive these differences in learning tendencies? In the future, analysis of song structures in a geographic context could help determine whether various species typically inherit entire songs or develop songs composed of syllables from several tutors, or if patterns of acquisition vary within species. With enough species-level data, comparative methods may help uncover conditions that drive learning tendencies or identify trends of song feature evolution that emerge from different patterns of song acquisition.

While determining certain life history traits, such as social or sexual mating systems, and species-typical learning mechanisms on a species level still require intensive field studies on each species, ideally across multiple populations, determining song characteristics of species may be achievable through analysis of community science recordings. This will allow for greater sample sizes and therefore more robust studies using phylogenetic comparative methods in the future. Further, several methods of phylogenetically-controlled trait comparison can account for variation within species. Aggregating values across many regions will allow this functionality to be used to a greater extent. Certain metrics of song will be easier to calculate than others using community science recordings. For example, the number of unique syllables per song, song duration, and inter-song interval for an individual are all easy to calculate with even brief recordings. Syllable repertoire and song repertoire sizes would require longer recordings than are often contributed by community scientists, but recordings of single birds that are over 30 or 60 minutes in duration do exist, and these may be sufficient for many species with relatively small repertoires (Garamszegi et al. 2005; Peshek and Blumstein 2011). An important caveat is that this may bias sampling towards species with smaller repertoires, species that sing for longer periods of time in one location, and species that have smaller inter-song intervals. Further, in several species, it is common for individuals to alternate singing with a neighbor, which could be easily mistaken as a single bird in a recording; in these cases, it will be important that the number of individuals represented in the recording is noted in the recording metadata.

If the upward trend in song recording contribution by community scientists continues, at some point recordings in certain areas may be frequent enough that unique recordings cannot be reliably assumed to be different individual birds. For studies that treat individual recordings as unique birds in analyses, as I did in Chapter 4, this may result in oversampling prolific singers or birds that happen to occupy a territory that is accessible and regularly traversed by humans, such as a public park, hiking trail, or school campus. I encountered this when sampling from the Ithaca, NY region; while I had already been treating recordings that were collected by the same recordist on the same day as the same bird unless explicitly noted in the recording remarks, I found that I had missed cases where different recordists clearly recorded the same bird on the same day. In one case, four recordings by four different recordists were collected on the same day, coordinate location, and nearly the same time; unsurprisingly, inspection of the spectrograms revealed practically indistinguishable songs, and these were virtually guaranteed to be from the same bird. This seemed quite predictable once I saw the name of the location: Sapsucker Woods, the nature reserve adjacent to and affiliated with the Cornell Lab of Ornithology. After this finding, I re-checked the recordings I was using and discarded all but one of any group of recordings that were collected on the same date and location, unless the recordist explicitly noted that the individuals in the recordings were confirmed different individuals – an immensely beneficial piece of metadata that prevented the preemptive removal of valuable samples from the study. To ensure the utility and reliability of community science recordings in scientific inquiry in the future as repeated sampling inevitably becomes more common, establishing a norm or a designated metadata field for recordists to note whenever recordings are known to be of different individuals or a single individual would be immensely beneficial.

Alternatively, if song analysis could eventually lead to the identification of individuals based on song recordings alone, possibly through minute differences in timbre and harmonics that correspond with physiological differences rather than strictly cultural ones, the issue of unintentional overrepresentation of the same individual bird across multiple recordings could be ameliorated. If this individual-identification ability is developed, it would also open up significant possibilities for studying other elements of avian biology. For example, knowledge on dispersal and philopatry is extremely limited, since the main way that these phenomena are investigated is through capture-recapture techniques, and, to a much lesser extent, through geotagging and remote sensing. It may be possible to effectively scan large databases of passive audio recordings for vocal matches to specific individuals to determine whether the individual is present in an area in which it was previously detected, or even whether it is present in other regions. It is possible that this type of analysis could also utilize recordings of calls, as opposed to songs alone, to detect vocal signatures that characterize particular individuals.

Community science song recordings also have significant potential to enable follow-up studies to the work in Chapters 2 and 3. Certain hypotheses that follow naturally from the work in Chapters 2 and 3 were unable to be tested due to a lack of data for a sufficient number of species to perform a robust, phylogenetically controlled analysis. For example, the different effects on the evolution of song complexity by social non-monogamy vs. sexual non-monogamy (extra-pair paternity, EPP) would benefit from the follow-up of testing the interacting effects of these two traits. However, there were gaps in the data, and not all species with both a social monogamy vs social polygyny classification and a low-EPP vs high-EPP classification also had published values for song feature values. Species that have previously been understudied, especially species in the Global South, may be able to be documented by local community scientists, without researchers from elsewhere traveling to record song data themselves. More can and should be done to empower residents across geographies to take the lead in collecting and analyzing birdsong data from species that are locally significant.

Defining Meaningful Syllable Boundaries: A Neuroscientific Approach

A common thread across all of my chapters is the finding that the uniqueness of syllables is an important measure of song complexity and diversity. However, the methodology behind calculating syllable repertoire remains highly subjective. While compiling data for use in my first two chapters, and again surveying journal articles while designing the methods for my within-population analyses, I informally noted a consistent pattern; whenever unique syllables are determined, especially—but not exclusively—in field studies, the methods typically say little more than "syllables were enumerated by hand" or "unique syllables were determined visually" (e.g. (Vargas-Castro, Sánchez, and Barrantes 2012)). To make these approaches more robust, a common option has historically been to have multiple individuals perform the same syllable classification task. However, a common outcome appears to be that there are discrepancies in results when the tasks are done separately, and it is very difficult for individuals to reach the same conclusions without discussion and coordination. Some computational methods exist that do a decent job of grouping syllable types into clusters. Sound Analysis Pro does this for zebra finch song fairly well in the context of an ontogeny, where there is typically gradual change in a syllable type for several days or weeks (Tchernichovski et al. 2000). These software tools tend to work best in lab settings, in single species, within the same animal or even the same song. Challenges of computational syllable sorting and matching include variable noise that may mean quieter parts of syllables get tossed, harmonics (vibrations at multiples of the fundamental frequency), quality and signal-to-noise ratio differences, and time warping of syllables or trailing echoes, making overlaying syllables or matching based on feature values measured over small slices of time either not accurate or even more computationally intensive.

Very few studies that have multiple individuals perform this pattern-identification and discrimination task on the same recordings also publish the results of the participants separately, so it is not known what the potential margin of error is for these metrics in any given study. However, at least for syllable repertoire and the number of unique syllables per song, variation across studies and across individual birds within studies can be quite large. Little has been determined about how robust or persistent any differences between populations may be, or why those differences may exist. In Chapters 2 and 3, to account for these differences, we repeated all of our tests using the minimum and maximum values per population reported in prior studies for each species. The positive takeaway is that, despite species feature values frequently varying by up to 50% between studies, the patterns of how these features relate between species tend to be broadly maintained; all significant results we obtained regarding the correlated evolution of syllable repertoire and social mating system or learning window remained significant when repeating the test using the minimum and maximum values for each species instead.

The fact remains that studies of syllable diversity continue to be highly subjective and will continue to be as long as we leave syllable discrimination decisions up to individual observers, especially without any guidelines. It would be beneficial to reach a consensus with a set of criteria to consider when deciding whether two syllables or syllable groups should be considered the same. However, while this might go some way towards reducing the variability of findings across individual observers, this still does not address another fundamental issue with this methodology: an inherent bias towards human perception of patterns. For any consensus we might reach regarding how much variation is permissible within syllable types – or how much constitutes a different type altogether – how do we know that our classifications are biologically relevant to the birds themselves? What amount of variation in birdsong syllables is ethologically meaningful?

This would require an experiment to probe a bird's perception and an ability to measure a response that can indicate some kind of discrimination. The three main types of behavioral experiments that can be used to assess song discrimination are female preference tests, stimulating male aggression with simulated territory incursion, and tutoring; however, none of these are likely to provide interpretable results when discriminating between subtly different stimuli. Any behavioral experiment would first require sample songs. While the question is about the differences between individual syllable types, birds would not engage in song-triggered behaviors if provided with only one syllable, or one syllable repeated in a way inconsistent with the species' typical song structure. Instead, songs that are essentially identical except for one syllable would need to be used. The choice of test syllable will determine the types of variation that can be tested, but some examples: from a complex multi-note syllable, remove one of the notes; shift a syllable up or down in pitch; apply a time warp to the syllable to compress or stretch it, preferably without changing the timing or rhythm of the rest of the song, which could influence a bird's overall perception independent of the test syllable. All of these are examples of small variations that could be considered negligible by a human completing a syllable sorting task. In a female choice test, a female would be presented with two perches or keys at opposite ends of a cage that, when landed upon or pecked, would trigger the playback of one of two songs. The song which is triggered the most is interpreted to be the more attractive song (Fujii et al 2022). In a male aggression test, playback speakers are brought into the field, generally to the edge of a male's territory, and an observer measures how closely the male approaches a speaker as it plays a series of different conspecific songs, with pauses to prevent habituation; closer approaches to the speaker are used as a measure of the male's aggressiveness and thus the song's quality. However, songs used in these are typically much more different than one nuance of a syllable (e.g. (Derryberry 2007). Finally, tutoring a juvenile would involve tutoring with playback. One approach would be to tutor birds with the same song sets as in the female choice or male aggression tests. If the juveniles end up producing only one of the test syllables, it may mean they perceived the difference and chose to learn the syllable each preferred. But, unless there is a directional preference, it would be difficult to confidently make conclusions. A second approach would be to tutor many juveniles with the same song or live tutor, track the development of each syllable, and compare all of the individual syllables that came from one common syllable after crystallization. The variation in that syllable in the juveniles might indicate an amount of variation that can exist within one syllable type; however, that assumes that the juveniles attempted to learn accurately (i.e. did not improvise upon the syllable) and did not learn poorly - neither of which can be ruled out.

Researchers in the field may be able to take advantage of a unique feature of songbirds. Males that are anesthetized exhibit neural activity in the song system nuclei LMAN, Area X, mLD, and HVC when exposed to conspecific song, and these signals can be reliably measured with electrophysiology techniques. Fascinatingly, the amplitude of this signal increases significantly and activity spikes occur roughly in synchrony with each syllable of the playback when exposed to a highly specific stimulus: the bird's own song (Poirier et al. 2009; Solis and Doupe 1999). The response is evoked in juveniles when they hear a playback of their tutor's song, but in adulthood, the response to the tutor's song weakens while the response to hearing a playback of a bird's own song becomes robust. This highly specific gating mechanism could be exploited to essentially tap into a bird's perception to deduce whether it can tell the difference between its own unaltered song and its song but with one syllable artificially modified. Since a bird is not expected to become habituated to its own song, many repeated measurements of multiple test syllable modifications can be performed within a single session, which is not possible with any of the other behavioral discrimination tests.

Concluding Remarks

In sum, this body of work sheds light on key factors and forces in the evolution of birdsong. In addition, I propose a new avenue of study to examine the role of environmental and ecological factors on learning and learned behaviors, with broad potential impacts in the field of cultural evolution. I further provide a proof-of-concept study using community science data to investigate the effects of population-scale environmental stress and offer initial findings that will hopefully motivate further studies of this kind. I hope that I have laid the groundwork for future endeavors to succeed in showing a snapshot of evolution of a learned behavior by probing the effects of extreme climatic events, paralleling the earlier work of giants in the field of evolutionary biology.

References

Derryberry, Elizabeth P. 2007. “Evolution of Bird Song Affects Signal Efficacy: An Experimental Test Using Historical and Current Signals.” *Evolution; International Journal of Organic Evolution* 61(8):1938–45.

Garamszegi, László Z., Thorsten J. S. Balsby, Ben D. Bell, Marta Borowiec, Bruce E. Byers, Tudor Draganoiu, Marcel Eens, Wolfgang Forstmeier, Paolo Galeotti, Diego Gil, Leen Gorissen, Poul Hansen, Helene M. Lampe, Stefan Leitner, Jan Lontkowski, Laurent Nagle, Erwin Nemeth, Rianne Pinxten, Jean-Marc Rossi, Nicola Saino, Aurélie Tanvez, Russell Titus, János Török, Els Van Duyse, and Anders P. Møller. 2005. “Estimating the Complexity of Bird Song by Using Capture-Recapture Approaches from Community Ecology.” *Behavioral Ecology and Sociobiology* 57(4):305–17.

Hughes, Melissa, Stephen Nowicki, William A. Searcy, and Susan Peters. 1998. “Song-Type Sharing in Song Sparrows: Implications for Repertoire Function and Song Learning.” *Behavioral Ecology and Sociobiology* 42(6):437–46.

Peshek, Kathleen R., and Daniel T. Blumstein. 2011. “Can Rarefaction Be Used to Estimate Song Repertoire Size in Birds?” *Current Zoology* 57(3):300–306.

Poirier, Colline, Tiny Boumans, Marleen Verhoye, Jacques Balthazart, and Annemie Van der Linden. 2009. “Own-Song Recognition in the Songbird Auditory Pathway: Selectivity and Lateralization.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29(7):2252–58.

Solis, Michele M., and Allison J. Doupe. 1999. “Contributions of Tutor and Bird’s Own Song Experience to Neural Selectivity in the Songbird Anterior Forebrain.” *The Journal of Neuroscience* 19(11):4559–84.

Tchernichovski, O., F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra. 2000. “A Procedure for an Automated Measurement of Song Similarity.” *Animal Behaviour* 59(6):1167–76.

Vargas-Castro, Luis E., Natalie V. Sánchez, and Gilbert Barrantes. 2012. “Repertoire Size and Syllable Sharing in the Song of the Clay-Colored Thrush (Turdus Grayi).” *The Wilson Journal of Ornithology* 124(3):446–53.

APPENDIX

Supplementary Tables

Number of unique recordings- dark-eyed junco	Drought Region	Control Region
2006-June 2016 (Pre-drought)	131 bouts from 46 recordings (45 ML, 1 XC)	11 bouts from 8 recordings (8 ML)
July 2016-Dec 2016 (During)	14 bouts from 5 recordings (5 ML)	0
2017-2019 (Post-drought)	199 bouts from 73 recordings (64 ML, 1 XC, 8 self)	62 bouts from 48 recordings (47 ML, 1 XC)

Table S4.1A: Distribution of recordings of dark-eyed juncos in the drought and control regions.

Number of unique recordings- song sparrow	Drought Region	Control Region
2006-June 2016 (Pre-drought)	409 bouts from 37 recordings (34 ML, 3 XC)	72 bouts from 16 recordings (1 ML, 15 XC)
July 2016-Dec 2016 (During)	9 bouts from 4 recordings (4 ML)	9 bouts from 3 recordings (2 ML, 1 XC)
2017-2019 (Post-drought)	745 bouts from 159 recordings (148 ML, 4 XC, 7 self)	330 bouts from 83 recordings (74 ML, 9 XC)

Table S4.1B: Distribution of recordings of song sparrows in the drought and control regions.

Song Feature	Log transformed	Shapiro-Wilk Before, Drought Region	Shapiro-Wilk After, Drought Region	Shapiro-Wilk Before, Control Region	Shapiro-Wilk After, Control Region	p-value Before vs After, Drought Region	p-value Before vs After, Control Region	Holm-Bonferroni Threshold, Drought Region	Significant After Bonferroni, Drought Region	Holm-Bonferroni Threshold, Control Region	Significant After Bonferroni, Control Region	Brown-Forsythe test for unequal variance, Before vs After, Drought Region	Brown-Forsythe test for unequal variance, Before vs After, Control Region
average minimum frequency of syllables	No	0.975	0.724	0.287	0.240	0.788	0.849	0.0250	FALSE	0.0250	FALSE	0.653	0.192
overall frequency range	No	0.537	0.879	0.879	0.448	0.603	0.198	0.0100	FALSE	0.0063	FALSE	0.165	0.213
number of notes per syllable	Yes	0.050	0.479	0.178	0.045	0.040	0.898	0.0033	FALSE	0.0500	FALSE	0.032	0.527
number of unique syllables	Yes	5.25E-09	3.37E-09	0.016	2.40E-06	0.534	0.754	0.0071	FALSE	0.0250	FALSE	0.269	0.846
average maximum frequency of syllables	Yes	0.062	0.779	0.295	0.176	0.879	0.727	0.0250	FALSE	0.0167	FALSE	0.360	0.863
mean syllable duration	Yes	0.051	0.016	0.986	0.064	0.890	0.204	0.0500	FALSE	0.0083	FALSE	0.842	0.969
standard deviation of syllable duration	Yes	0.014	0.889	0.829	0.099	0.348	0.194	0.0063	FALSE	0.0071	FALSE	0.072	0.644
rate of syllable production	Yes	0.075	0.004	0.961	0.021	0.207	0.146	0.0056	FALSE	0.0063	FALSE	0.599	0.414
total number of syllables	Yes	0.380	0.010	0.542	0.938	0.069	0.060	0.0042	FALSE	0.0056	FALSE	0.259	0.846
mean syllable stereotypy	Yes	0.053	9.86E-05	0.972	0.027	0.702	0.034	0.0083	FALSE	0.0050	FALSE	0.252	0.379
degree of syllable repetition	No	0.552	0.332	0.160	0.123	0.098	0.015	0.0045	FALSE	0.0045	FALSE	0.417	0.255
total number of notes	Yes	0.793	0.754	0.625	0.006	0.005	0.005	0.0029	FALSE	0.0042	FALSE	0.966	0.639
bout duration	Yes	0.579	0.087	0.684	0.006	0.015	0.004	0.0031	FALSE	0.0038	FALSE	0.908	0.153

Table S4.2. Statistical analysis of dark-eyed junco songs. When all Shapiro-Wilk tests were not significant, indicating normality, we compared song features using a t-test. Otherwise, we log-transformed and used a Wilcoxon rank-sum test. We corrected for multiple hypothesis testing with a Holm-Bonferroni correction. No song features were significantly different before vs. after the drought in either the drought or control region.

Song Feature	Log transformed	Shapiro-Wilk Before, Drought Region	Shapiro-Wilk After, Drought Region	Shapiro-Wilk Before, Control Region	Shapiro-Wilk After, Control Region	p-value Before vs After, Drought Region	p-value Before vs After, Control Region	Holm-Bonferroni Threshold, Drought Region	Significant After Bonferroni, Drought Region	Holm-Bonferroni Threshold, Control Region	Significant After Bonferroni, Control Region	Brown-Forsythe test for unequal variance, Before vs After, Drought Region	Brown-Forsythe test for unequal variance, Before vs After, Control Region
degree of syllable repetition	Yes	0.452	0.273	0.093	0.447	5.00E-05	0.059	0.0036	TRUE	0.0036	FALSE	0.107	0.062
number of notes per syllable	Yes	0.491	3.13E-05	0.278	0.048	0.0006	0.267	0.0038	TRUE	0.0050	FALSE	0.753	0.481
total number of syllables	No	0.695	0.105	0.505	0.501	0.0026	0.020	0.0042	TRUE	0.0033	FALSE	0.551	0.359
rate of syllable production	Yes	0.043	0.571	0.860	0.282	0.0074	0.061	0.0045	FALSE	0.0038	FALSE	0.038	0.619
mean syllable duration	Yes	0.989	0.861	0.717	0.233	0.0199	0.089	0.0050	FALSE	0.0045	FALSE	0.086	0.284
average minimum frequency of syllables	Yes	0.510	0.017	0.238	0.534	0.1030	0.341	0.0056	FALSE	0.0071	FALSE	0.098	0.537
standard deviation of syllable duration	Yes	0.985	0.093	0.307	0.102	0.1788	0.270	0.0063	FALSE	0.0056	FALSE	0.048	0.859
bout duration	Yes	0.012	4.73E-04	0.240	0.279	0.2164	0.406	0.0071	FALSE	0.0125	FALSE	0.325	0.335
duration of the longest syllable	Yes	0.026	7.36E-05	0.306	0.013	0.2402	0.372	0.0083	FALSE	0.0083	FALSE	0.096	0.848
total number of notes	Yes	0.207	0.321	0.643	0.255	0.3965	0.808	0.0100	FALSE	0.0500	FALSE	0.182	0.263
overall frequency range	Yes	2.61E-04	1.49E-06	0.384	0.001	0.5581	0.328	0.0125	FALSE	0.0063	FALSE	0.954	0.184
duration of the shortest syllable	Yes	0.267	0.683	0.289	0.592	0.5730	0.087	0.0167	FALSE	0.0042	FALSE	0.409	0.492
average maximum frequency of syllables	Yes	0.188	0.005	0.340	0.068	0.8469	0.763	0.0250	FALSE	0.0250	FALSE	0.522	0.661
number of unique syllables	Yes	0.031	0.638	0.884	0.110	0.9474	0.381	0.0500	FALSE	0.0100	FALSE	0.634	0.361

Table S4.3. Statistical analysis of song sparrow songs. When all Shapiro-Wilk tests were not significant, indicating normality, we compared song features using a t-test. Otherwise, we log-transformed and used a Wilcoxon rank-sum test. We corrected for multiple hypothesis testing with a Holm-Bonferroni correction. Three song features were significantly different before vs. after the drought in the drought region, and no song features differed in the control region.

Supplementary Figures

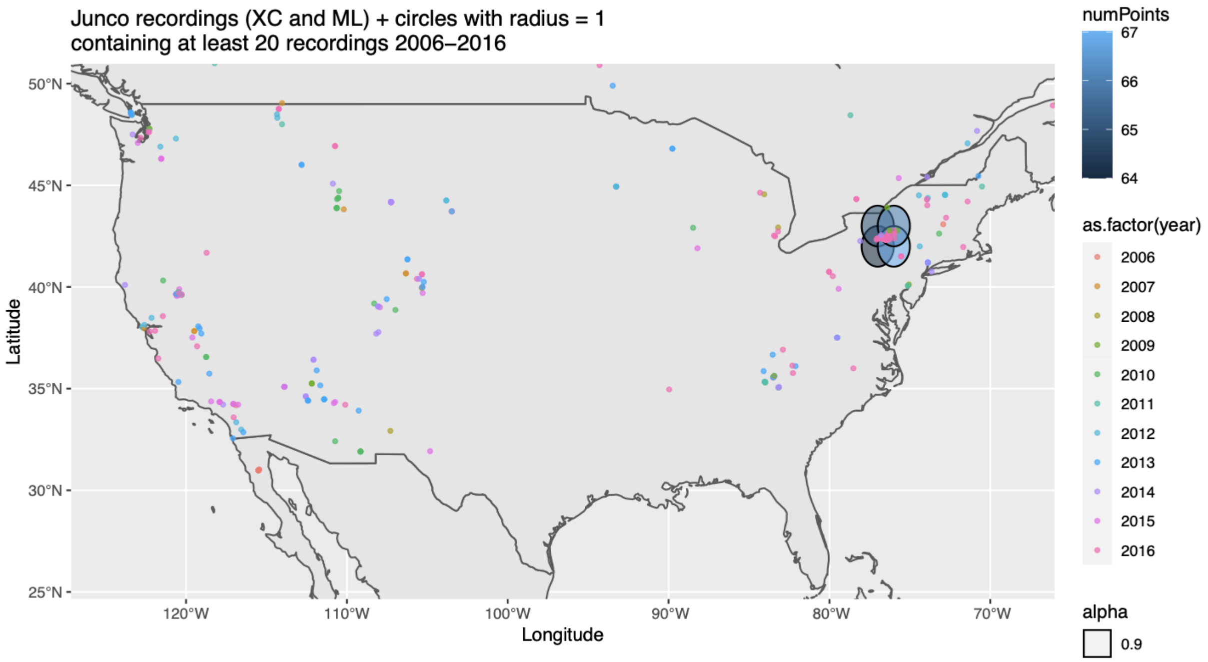
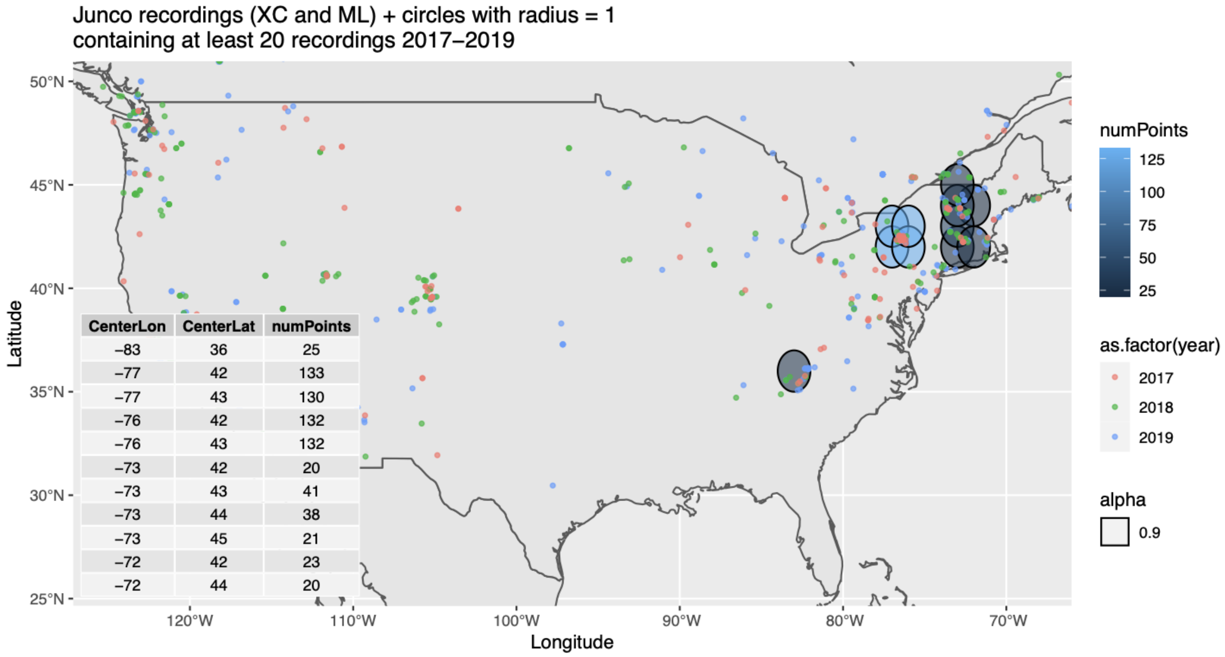


Figure S4.1A: Dark-eyed junco recordings publicly available on Macaulay Library or xeno-canto recorded between 2006 and 2016. Of all regions centered on each value of (Longitude, Latitude) ranging from 70W through 89W and 30N through 47N, the plotted, filled circle regions are the only regions that contain at least 20 recordings. The four circles shown here all include part of Tompkins County, NY.



CenterLon	CenterLat	numPoints
−83	36	25
−77	42	133
−77	43	130
−76	42	132
−76	43	132
−73	42	20
−73	43	41
−73	44	38
−73	45	21
−72	42	23
−72	44	20

Figure S4.1B: Dark-eyed junco recordings publicly available on Macaulay Library or xeno-canto recorded between 2017 and 2019. Of all regions centered on each value of (Longitude, Latitude) ranging from 70W through 89W and 30N through 47N, the plotted, filled circle regions are the only regions that contain at least 20 recordings. Inlaid table shows the exact counts for each circle and the (Longitude, Latitude) center of the plotted circle.

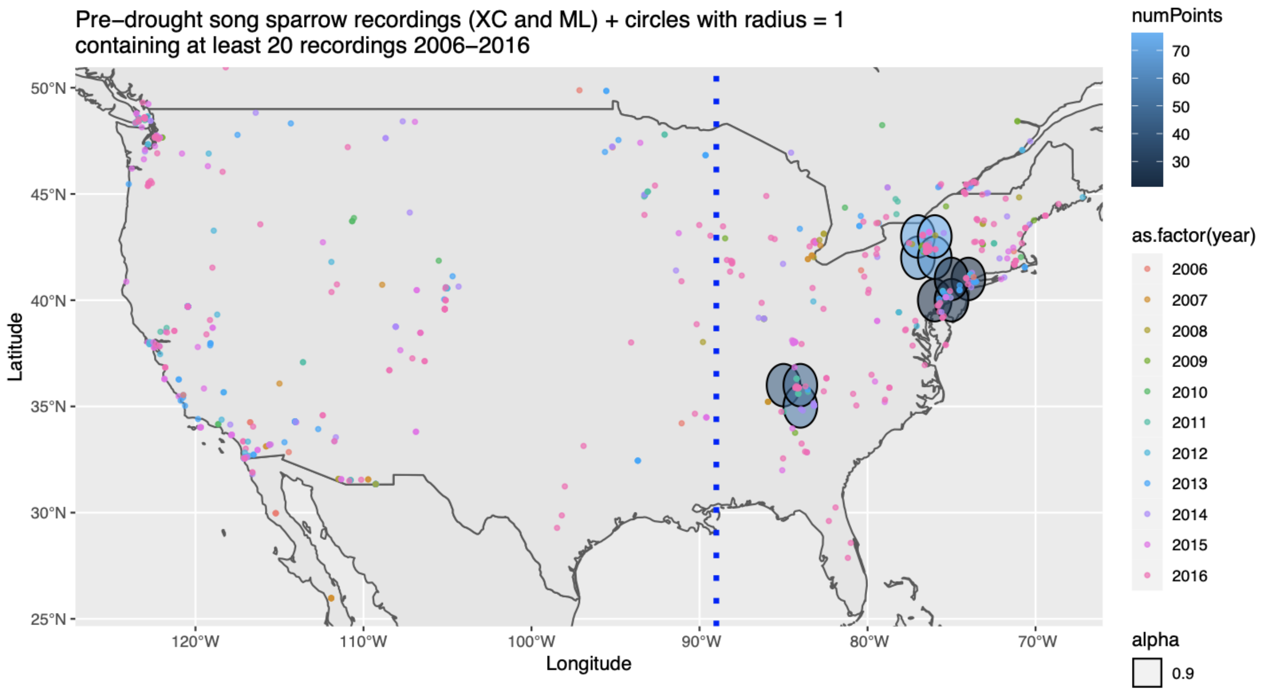


Figure S4.1C: Song sparrow recordings publicly available on Macaulay Library or xeno-canto recorded between 2006 and 2016. Of all regions centered on each value of (Longitude, Latitude) ranging from 70W through 89W and 30N through 47N, the plotted, filled circle regions are the only regions that contain at least 20 recordings.

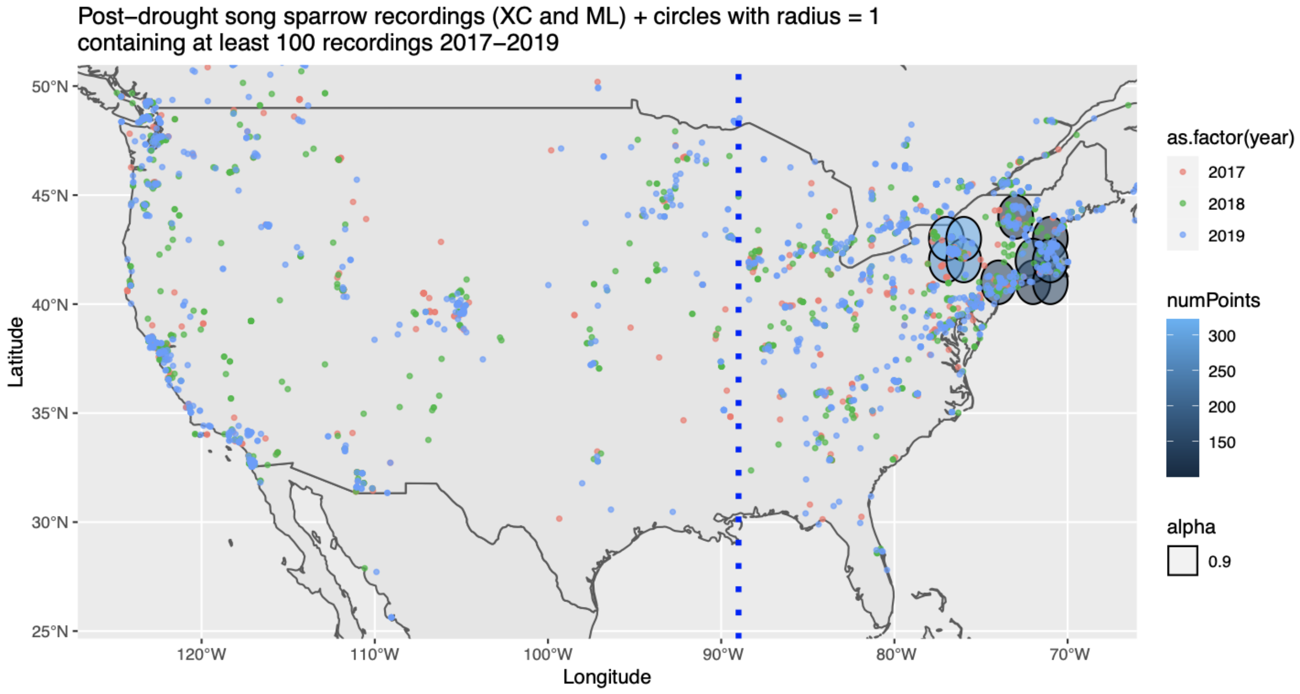


Figure S4.1D: Song sparrow recordings publicly available on Macaulay Library or xeno-canto recorded between 2017 and 2019. Of all regions centered on each value of (Longitude, Latitude) ranging from 70W through 89W and 30N through 47N, the plotted, filled circle regions are the regions that contain at least 100 recordings.

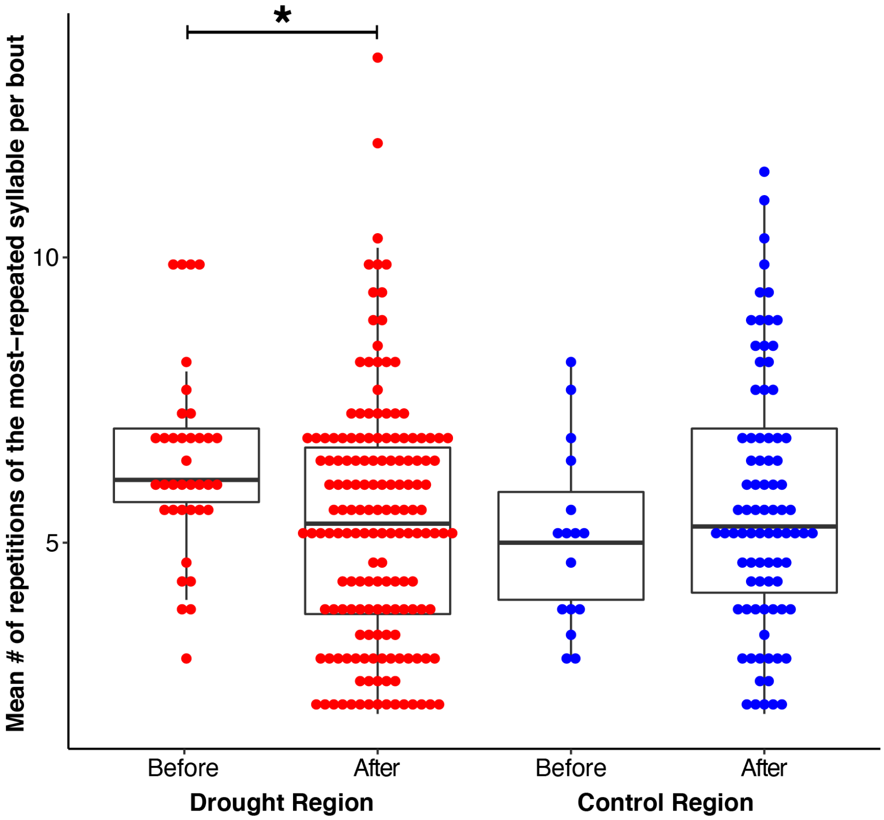


Figure S4.2: Distributions of average counts of the most-repeated syllable per bout per recording within song sparrow populations before and after the 2016 drought. For each bout, we found the most repeated syllable type based on the syllable type assignment from the Chipper analysis, counted the number of times the most repeated syllable was produced, and averaged those counts across all bouts sampled from a given recording. Overall, in recordings after the drought, the most-repeated syllable type per bout was repeated fewer times on average (t-test $p = 0.007$).