The role of learned song in the evolution and speciation of Eastern and Spotted towhees

Ximena León Du'Mottuchi[1,2] & Nicole Creanza[1,2,*]

[1]Department of Biological Sciences, Vanderbilt University

[2]Evolutionary Studies Initiative, Vanderbilt University

[*]Correspondence: nicole.creanza@vanderbilt.edu

ABSTRACT

Oscine songbirds can learn complex vocalizations that function in mate attraction and territory defense. Sexual selection pressures on such learned behavioral traits have been proposed to accelerate speciation. The Eastern and Spotted towhees are sister species that diverged relatively recently (0.28 Ma) but now have a partially overlapping range; widespread community-science recordings of these species, including songs from within their zone of overlap, enable us to investigate whether song differentiation could have facilitated their reproductive isolation. Here, we quantify 16 song features to analyze the geographic variation in Spotted and Eastern towhee songs and test for species-level differences. We then use a random-forest model to measure how accurately the two species' songs can be distinguished from one another, as well as whether songs within their region of overlap are more easily distinguishable than those from non-overlapping regions. While no single song feature can reliably distinguish the two species, their songs can be distinguished with 89.5% accuracy using a random-forest model trained on 16 features. For each song feature, there is variation in the strength of the correlation across the longitudinal gradient and considerable differences within and between species. Finally, we analyze the limited publicly available genetic data from each species and find support for the hypothesis that they are reproductively isolated. Together, our results suggest that, in combination, small variations in song features may contribute to these sister species' ability to recognize their species-specific songs.

INTRODUCTION

In the Oscine songbirds, birdsong is a set of learned vocalizations used primarily for attracting mates and defending territories, and the characteristics of these songs vary widely across different species (Catchpole and Slater 1995). Song also varies within a species, with some differences in birdsong characteristics occurring at large spatial scales (Searfoss, Liu, and Creanza 2020). This geographic variation is thought to be influenced by a combination of genetic and environmental factors, such as adaptation to local habitats and ecological conditions (Podos and Warren 2007). Furthermore, under isolation-by-distance due to spatially limited dispersal, songs can accumulate differences gradually over time and become increasingly different with geographic distance (Rivera-Gutierrez et al. 2010). In vocal learners, such as songbirds, geographic variation can also accumulate through cultural transmission of learned vocalizations because of many factors, including, but not limited to, physical isolation and divergence in sexual selection (Podos and Warren 2007).

Divergence in traits that are under sexual selection is important for premating reproductive isolation (Price 2007), since mate preference based on these traits can limit interpopulation mating (Ptacek 2000; MacDougall-Shackleton and MacDougall-Shackleton 2001; Slabbekoorn and Smith 2002; Lachlan and Servedio 2004; Podos 2007). For example, in a study of chorus frogs (*Pseudacris n. nigrita* and *Pseudacris triseriata feriarum*), the (unlearned) calls of the two species show significant differences in pulse rate and pulse number in sympatry, suggesting that these characteristics may have diverged via reproductive character displacement (Fouquette 1975). A similar pattern has been observed for a learned behavior: Ratcliffe and Grant (1985) found that although songs of *Geospiza fortis* and *Geospiza fulginosa* were virtually indistinguishable, individuals were still able to discriminate between species songs in a playback experiment, suggesting that they are likely using subtle song differences to inform their mate choice. Since birds can use their songs for species recognition, these song differences can inhibit birds from recognizing birds from different locations with less familiar song characteristics, preventing individuals from mating (Podos 2007; Price 2007). In theory, a rapid accumulation of song changes can lead to reproductive isolation, accelerating speciation (Lachlan and Servedio 2004).

Hybridization can occur when two genetically distinct populations come into contact with one another and mate (Barton & Hewitt 1985, 1989). Searching for a mate can be costly to

females if conspecific males are difficult to find in a given area. Therefore, females of the rarer species could have an increased tendency to hybridize with mates of the more common species, particularly if the signal of the heterospecific male is similar to that of the conspecific male (Price 2007). Stability of the hybrid zone can be maintained via spatial segregation of the two populations: as the hybrid zone is traversed, individuals of one species will become less common, leading them to mate more frequently with the other species and produce hybrids with reduced fitness (Price 2007).

In some hybrid zones, the hybrids between two species have high fitness and are not selected against, so the hybrid zone can be quite large (Grant and Grant 1992; Price 2007). However, cross-mating does not necessarily lead to the populations merging into one. For example, selection that favors the prevalent phenotype of a given area constrains the width of the hybrid zone because individuals that disperse in the range of the other species exhibit inferior fitness (Moore and Price 1993; Price 2007). Further, separation of species can be maintained by reinforcement if the strength of selection is greater for the characteristics that fall in the extremes rather than those of the hybrids (Price 2007), such that premating isolation is strengthened. In other words, if sexual selection favors more extreme characteristics, then hybrids will have lower mating success For example, a study of two species of chorus frogs showed evidence that hybrid fitness was significantly reduced, suggesting that reinforcement, via sexual selection against hybrid males, is the primary mechanism behind the reproductive character displacement in these species (Lemmon and Lemmon 2010).

The Spotted towhee (*Pipilo maculatus*) and the Eastern towhee (*Pipilo erythrophthalmus*) are sister species of the Passerellidae family of oscine songbirds. It is estimated that the Eastern and Spotted towhees diverged 280,000 years ago (Hirase et al. 2016; DaCosta et al. 2009). Evidence from the geological history of the Great Plains, in combination with patterns of the current geographic distribution of the towhees and their character gradients, suggest that present hybridization of these sister species in the Great Plains is a result of secondary contact that occurred after the Pleistocene glaciation fragmented their breeding range and accelerated speciation (Sibley and West 1959; Hirase et al. 2016; Price 2007). They were previously considered one species, the Rufous-sided towhee, and the idea that the Spotted and Eastern towhees should be considered separate species has been contentious for many years. These sister species were separated into their respective groups primarily due to distinctions in their

geographic distribution, morphology, and genetic data (Ball and Avise 1992). Secondarily, early studies noted some differences between songs in the eastern and western populations, but also indicated that towhees in the Great Plains sang 'intermediate' songs (Sibley and West 1959).

The breeding range of the Spotted towhee lies on the west side of the United States extending into Mexico and southern Canada, while the breeding range of the Eastern towhee lies on the east side of the United States extending into southern Canada. Hybridization of the Spotted and Eastern towhees have been reported in the Great Plains area, particularly in Nebraska, with birds of this region exhibiting intermediate songs. Sibley and West (1959) used a spotting index to score birds in the Great Plains based on spotting of the wing coverts (males) and color of the head and back (females). Although hybridization occurs, they are still able to maintain themselves as separate species. While they share many similarities in their vocalizations, subtle but consistent differences in their songs could allow individuals to distinguish between the different species and identify potential conspecific mates, allowing separation of species to be maintained, a form of behavioral premating isolation.

The typical song of both the Spotted towhee and Eastern towhee is a loud and clear "drink-your-tea", composed of short introductory notes and a fast trill. Previous studies suggest that the Eastern towhees tend to have songs that are more variable and complex, with a greater number of syllables, while the Spotted towhees tend to have fewer syllables or no introductory syllables with a faster trill (Borror 1959; Kroodsma 1971). In this study, we aim to distinguish any geographic patterns in the song characteristics of the Spotted towhee and Eastern towhee, which have not been quantified between the two species. By analyzing the songs of these two species individually and as a two-population cline, we aim to understand if there are song patterns that are continuous across the geographic ranges of the Eastern and Spotted towhees or if there are discontinuities in the geographic distribution that clearly distinguish the songs of the two species from one another. Additionally, we use machine learning to investigate whether the two species differ their song characteristics such that birds could potentially use multiple song features to distinguish between conspecific and heterospecific individuals.

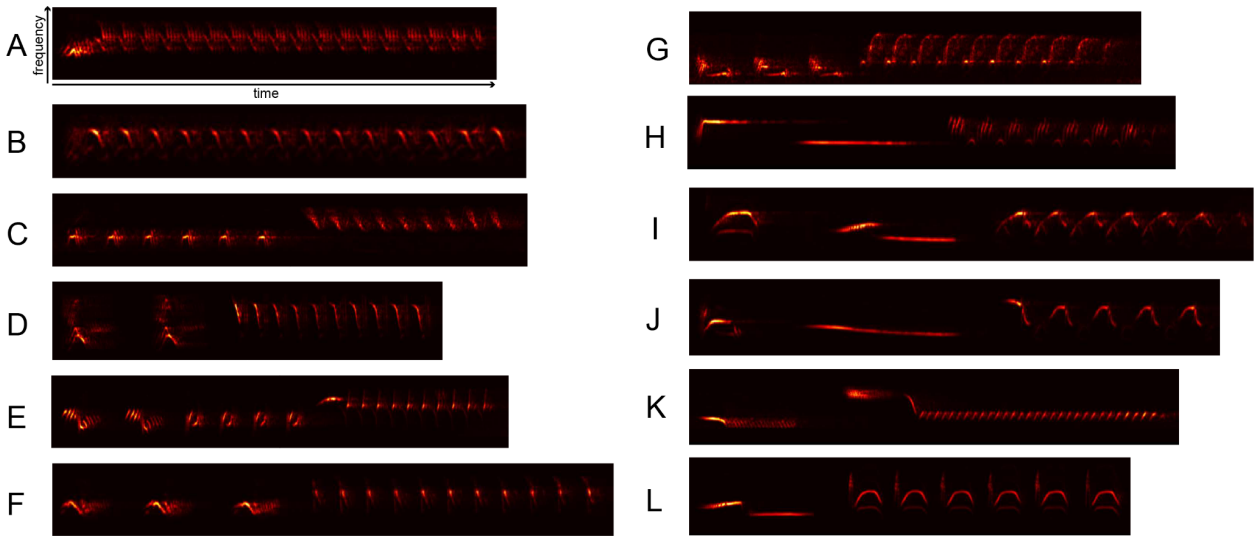

Figure 1. Spectrograms of (A-F) Spotted towhee songs and (G-L) Eastern towhee songs. The repeated syllable at the end of each song is often called the 'trill'.

METHODS

Generating a database of song bouts

Recordings of Spotted and Eastern towhees were downloaded from the Macaulay Library (Cornell Lab of Ornithology 2009) by requesting access to the relevant song files, as well as from Xeno-canto (Planque and Vellinga 2005) using the WarbleR package (Araya-Salas and Smith-Vidaurre 2017) in R. We filtered the recordings to eliminate those without song by specifying "type:song". For further analysis, we constructed a spreadsheet of metadata for each recording, which included species, date, location, latitude, longitude, and recordist. To minimize the chance of including repeated song recordings of the same bird, we discarded any duplicate entries with the same recordist, date, and location. For recordings that named a location but not geographic coordinates, we estimated the latitude and longitude by identifying the location on a map. Each recording file was opened in Audacity version 3.1.3 (https://www.audacityteam.org/) to manually extract between one or more bouts depending on the number of unique song types from each recording. Files from Xeno-canto are generally stored in MP3 format and files from Macaulay Library in WAV format; we exported each bout as a WAV file with a sampling rate of 44100 Hz to standardize the recordings for analysis. We considered the breeding season to be between April and August based on the "Breeding" section of the Birds of the World database

(Bartos and Greenlaw 2020; Greenlaw 2020), and we conducted analyses on all recordings (main text) and on breeding season recordings (**Table S1**, **Supplementary Materials**).

To assign the zone of overlap between the ranges of the two species, we plotted the ratio of the number of recordings of each species over the total number of samples at each degree longitude during the breeding season. In other words, we calculated the fraction of Spotted towhees over the total number of Spotted and Eastern towhee samples combined, and then calculated the same for the Eastern towhees. Based on our findings, we chose to focus on the area between 102°W and 91°W; which contained the majority of the species' range overlap (**Fig. S1**). In addition, we downloaded from eBird all sightings of putative Spotted towhee x Eastern towhee hybrids during the breeding season and plotted their locations; we observed that hybridization events primarily occur in this region (**Fig. S2**).

Song analysis (syllable features and syllable segmentation)

To begin analyzing the features of the syllables within these bouts, we used the Chipper software (Searfoss, Pino, and Creanza 2020), which was designed to facilitate the syllable segmentation and analysis of field recordings with different levels of background noise. This software parses each song into syllables by identifying periods of sound separated by silences, and then allows the user to enable high-pass and low-pass filters and to modify the segmentation when syllables have been incorrectly parsed, such as when background noise occurs during the inter-syllable silences (**Fig. 2**). To prevent the lower amplitude syllables from being segmented incorrectly, we normalized the amplitude across each song. Additionally, we adjusted the signal-to-noise threshold, minimum syllable duration, and minimum silence duration to most accurately define the syllables in the song. If any syllables appear to be parsed incorrectly after this procedure, Chipper allows the user to manually modify the syllable segmentation. If, when visualized in Chipper, we determined that the song overlapped with other birds singing or had excessive background noise, we removed the recording from the analysis. We provide the analyzed song bouts and analysis parameters used in Chipper to extract song features, our code for statistical analysis and visualization of the song feature data, and metadata about each song recording at https://github.com/CreanzaLab/TowheeAnalysis.

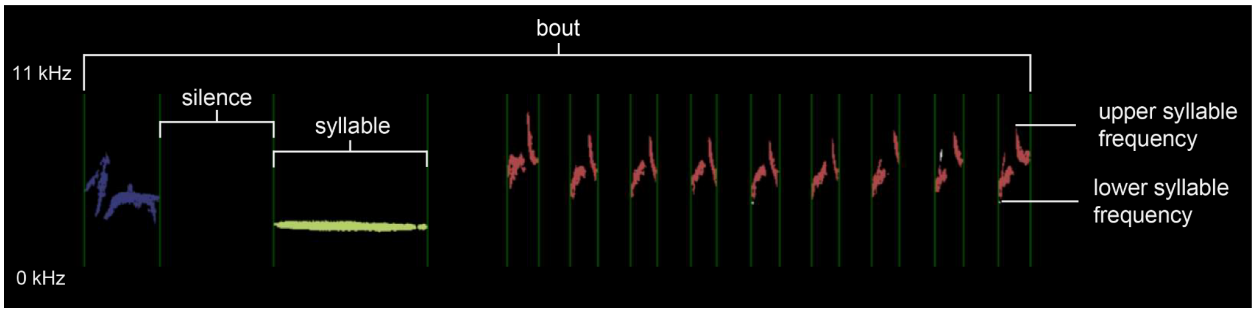

Figure 2. Spectrogram of Spotted towhee showing syllable, silence, bout, maximum syllable frequency and minimum syllable frequency. Color of syllable indicates unique syllable type.

After segmentation, we examined song bouts in Chipper to estimate a noise threshold, or the minimum number of matrix elements in the spectrogram that a note must contain in order to not be discarded as noise (Spotted towhee = 87, Eastern towhee = 70), and syllable similarity threshold, or the percent syllable overlap that determines whether two syllables are considered the same or different (Spotted towhee = 26.6, Eastern towhee = 25.5). We then ran the song analysis function in Chipper, which uses the spectrogram and the syllable segmentation data to provide automated measurements of numerous song features, which were used for statistical analysis. We used 16 song-feature outputs from Chipper: average syllable lower frequency (Hz), number of unique syllables, bout duration (ms), smallest syllable duration (ms), number of syllables per bout duration (1/ms), maximum syllable frequency (Hz), average syllable frequency range (Hz), minimum syllable frequency (Hz), largest syllable duration (ms), number of syllables per number of unique syllables, average syllable duration (ms), smallest syllable frequency range (Hz), number of syllables, overall syllable frequency range (Hz), largest syllable frequency range (Hz), and average syllable upper frequency (Hz).

Statistical analysis

To investigate the geographic variation in birdsong of the Spotted and Eastern towhees, we first determined that the song features were not normally distributed, so we used the natural log of each feature. To visualize the data, we plotted each log-transformed feature against the longitude and latitude coordinates of its recording.

To test whether a directional pattern exists between the longitude and latitude and the song features of the Spotted towhees and the Eastern towhees, we used Spearman rank correlation to measure the strength and direction of association across the 3515 song bouts. We

used the Bonferroni multiple hypothesis correction to adjust our alpha threshold for statistical significance for 16 song features and analysis of longitude and latitude ($\alpha_{adjusted}$=0.00156). Additionally, we used a Wilcoxon rank-sum test between Eastern and Spotted towhees to test for species-level differences in song features across a longitude gradient using Bonferroni multiple hypothesis correction for 16 song features ($\alpha_{adjusted}$=0.003125).

Principal components and Procrustes analyses of song feature data

We used a Principal Component Analysis (PCA) to visualize the variation in the song data, and we measured the proportion of variance explained by each principal component (PC). Our analysis contained a total of 1427 Spotted towhee samples and 2088 Eastern towhee samples and the same 16 song features. We followed up with a Procrustes analysis using PC1 and PC2 from the PCA of our song data. For this analysis, the matrix of PC scores was rotated and transformed onto the target matrix containing the longitude-latitude coordinates of the samples, such that the sum of squared-distances between the corresponding points of the transformed and target matrix was minimized and plotted on a map using the 'procrustes' function from the vegan package (Oksanen et al. 2023) in R.

Machine learning

We used a random forest model (RFM) to assess how accurately Spotted and Eastern towhee songs can be distinguished by their songs, using the ranger package (Wright and Ziegler 2017) in R. The model was trained on the 16 song features from the training set and multiple decision trees were averaged to make predictions of the species identity in the test set. We used our random forest algorithm to train two models: 1) a model trained on a geographically unbiased subset of the entire data set and 2) a model trained on a subset of samples obtained from the non-overlap zone. For the former, we obtained a random subsample (75% for training and 25% for testing) of the entire data set, and we then downsampled the training set to obtain a balanced subsample of Eastern towhees equal to the number of Spotted towhee samples ($N_{Spotted_towhee}$=1062 and $N_{Eastern_towhee}$=1062). We tested the model on 879 samples, none of which were resolved from the training set. For the latter, we separated the samples from the zone of overlap and from those in the non-overlap regions, and trained only on a subset of the non-overlap samples, retaining a sample size equal to that of the zone of overlap for testing. Our

training set was downsampled, such that we obtained a balanced subsample of Eastern towhees equal to that of Spotted towhees in our training set ($N_{Spotted_owhee}$=1062 and $B_{Eastern_towhee}$=1062). We tested this model on both a sample of towhees from the non-overlap zones (N=299) and an equally sized sample of towhees from the zone of overlap (N=299). For both models, we created a 10-fold cross-validation control and tuned the hyperparameters using a 10 x 16 grid using the caret package (Kuhn 2008) in R, which allows us to evaluate every possible combination of the number of predictors to randomly sample at each split and the minimum number of samples needed to keep splitting nodes using Gini impurity to split nodes. Additionally, we tested the models using both 100 and 1000 trees and obtained similar results (**Table S2**). Our code for all analyses is available at github.com/CreanzaLab/TowheeAnalysis.

Genomic analysis

We downloaded fasta files of mtDNA sequences of the cytochrome oxidase subunit I regions of the Spotted and Eastern towhee genomes available from the Barcode of Life Data Systems (BOLD) database. We obtained a total of 18 Spotted towhee and 5 Eastern towhee sequences. We aligned these sequences using MAFFT (Katoh, Asimenos, and Toh 2009), and we used a pairwise F_{ST} (fixation index) in Arlequin (Excoffier and Lischer 2010) to quantify the level of genetic differentiation between the two species by measuring the proportion of genetic variation that is due to differences between populations. It also measures the extent to which individuals within populations are similar to one another, where a larger F_{ST} describes a greater difference in allele frequencies within a population. We also did an Analysis of Molecular Variance (AMOVA) to quantify the proportion of genetic variation that is due to differences between and within populations.

PCA and Procrustes analysis of genomic data

Using the glPCA function from the adegenet package (Jombart 2008) in R, we ran a PCA on the aligned fasta files from our genomic analysis. We followed with a Procrustes analysis using PC1 and PC2, and we transformed the PC matrix onto the target matrix containing the longitude-latitude coordinates of the samples using the procrustes function from the vegan package (Oksanen et al. 2023) in R.

RESULTS

After gathering all song recordings of Eastern and Spotted towhees from the largest public repositories of natural sounds, Xeno-canto and Macaulay Library, and eliminating recordings that lacked song or had excessive background noise, we were able to analyze 2088 song bouts from 1563 individual Eastern towhees and 1427 song bouts from 874 individual Spotted towhees. For each song bout, we segmented the song into syllables, defined as periods of sound separated by periods of silence, and we used the song-analysis software Chipper (Searfoss, Pino, and Creanza 2020) to automatically extract 16 syllable and song features. In the Spotted towhee, several samples had very rapid trills that could not be further parsed into syllables (i.e. there were no periods of silence within the trill (Fig. 1)) and, thus, were considered to be a single syllable. Therefore, a subset of the Spotted towhee samples with one syllable in the bout were recorded from birds with a rapid trill that could not be further subdivided by silences.

Statistical analysis

Statistical analysis revealed several patterns in the song features of the Spotted and Eastern towhees. Spearman rank correlation showed statistical significance ($p < 0.00156$) in 14 syllable features across longitude and 11 features across latitude (**Table 1**). Stronger correlations ($\rho > 0.400$) existed in song features across the longitude gradient (e.g. bout duration, number of unique syllables, average syllable upper frequency, and smallest syllable frequency range). A Wilcoxon rank-sum test showed statistical significance ($p < 0.003125$) in all but 2 of the song features. We observed qualitatively similar results when we restricted the analysis to only songs that were recorded during the breeding season: the same set of song features show significant associations with latitude and longitude (**Table S1**).

PCA and Procrustes analysis of song data

A PCA of the song data revealed considerable overlap in song space using PC1 and PC2. Individual samples in both the range of overlap and non-overlap were scattered randomly in PC space; if the individuals in the zone of overlap had songs that showed greater species-level differentiation, we would expect that these points would have greater separation in PC space than those in the zone of non-overlap. PC1 contributed 35.4% of the variance, with the highest

loading being average syllable duration. PC2 contributed 24.7% with average syllable upper frequency most closely associated to it. A Procrustes analysis comparing the first two principal components of the song data to the latitude-longitude coordinates of the recording location indicated that, despite the extensive overlap between the two species, there is significant geographic signal in the song data (Procrustes rotation 0.26, $p < 0.001$).

Song Feature	Spearman						Wilcoxon	
	Latitude			Longitude				
	S	*p*-value	ρ	S	*p*-value	ρ	W	*p*-value
bout duration (ms)	7397584966	1.92E-01	-0.022	3748450720	4.40E-204	0.482	654894	1.18E-175
number of syllables	7827213145	1.35E-06	-0.081	7359761359	3.19E-01	-0.017	1626277	3.67E-06
number of syllables per bout duration (1/ms)	7622436939	1.64E-03	-0.053	8589473544	6.15E-29	-0.187	1924310	5.88E-49
largest syllable duration (ms)	6771885315	1.33E-04	0.064	5522445809	4.39E-46	0.237	896183	8.97E-90
smallest syllable duration (ms)	6731656553	3.30E-05	0.070	7423283837	1.29E-01	-0.026	1474316	6.01E-01
average syllable duration (ms)	6680317932	4.79E-06	0.077	6266818563	1.36E-15	0.134	1135866	4.61E-33
number of unique syllables	6992434693	4.42E-02	0.034	3453642343	6.34E-246	0.523	534212.5	4.72E-244
number of syllables per number of unique syllable	7832890160	1.07E-06	-0.082	9075790484	7.82E-53	-0.254	2089206	1.10E-91
average syllable upper frequency (Hz)	8248493705	9.26E-17	-0.140	10162112044	4.08E-138	-0.404	2438590	3.02E-226
average syllable lower frequency (Hz)	7277970268	7.44E-01	-0.006	8380047396	4.99E-21	-0.158	1864955	6.12E-37
maximum syllable frequency (Hz)	8436848622	4.89E-23	-0.166	7929152107	1.42E-08	-0.095	1880803	5.53E-40
minimum syllable frequency (Hz)	7183062081	6.52E-01	0.008	10128201861	1.09E-134	-0.399	2204676	2.00E-129
overall syllable frequency range (Hz)	8398422023	1.15E-21	-0.160	6640075404	9.34E-07	0.083	1548046	4.86E-02
largest syllable frequency range (Hz)	8230917361	3.14E-16	-0.137	8329654663	2.49E-19	-0.151	1923190	1.01E-48
smallest syllable frequency range (Hz)	7901898144	5.13E-08	-0.092	10290529942	1.32E-151	-0.422	2393045	2.25E-205
average syllable frequency range (Hz)	8075151267	6.12E-12	-0.116	9367136726	4.24E-71	-0.294	2187303	3.26E-123

Table 1. Spearman rank correlation ($\alpha_{adjusted}$=0.00156) and Wilcoxon rank-sum test ($\alpha_{adjusted}$=0.003125) of Spotted towhee and Eastern towhee song features (N=3515; $N_{Spotted_towhee}$=1427; $N_{Eastern_towhee}$=2088) .

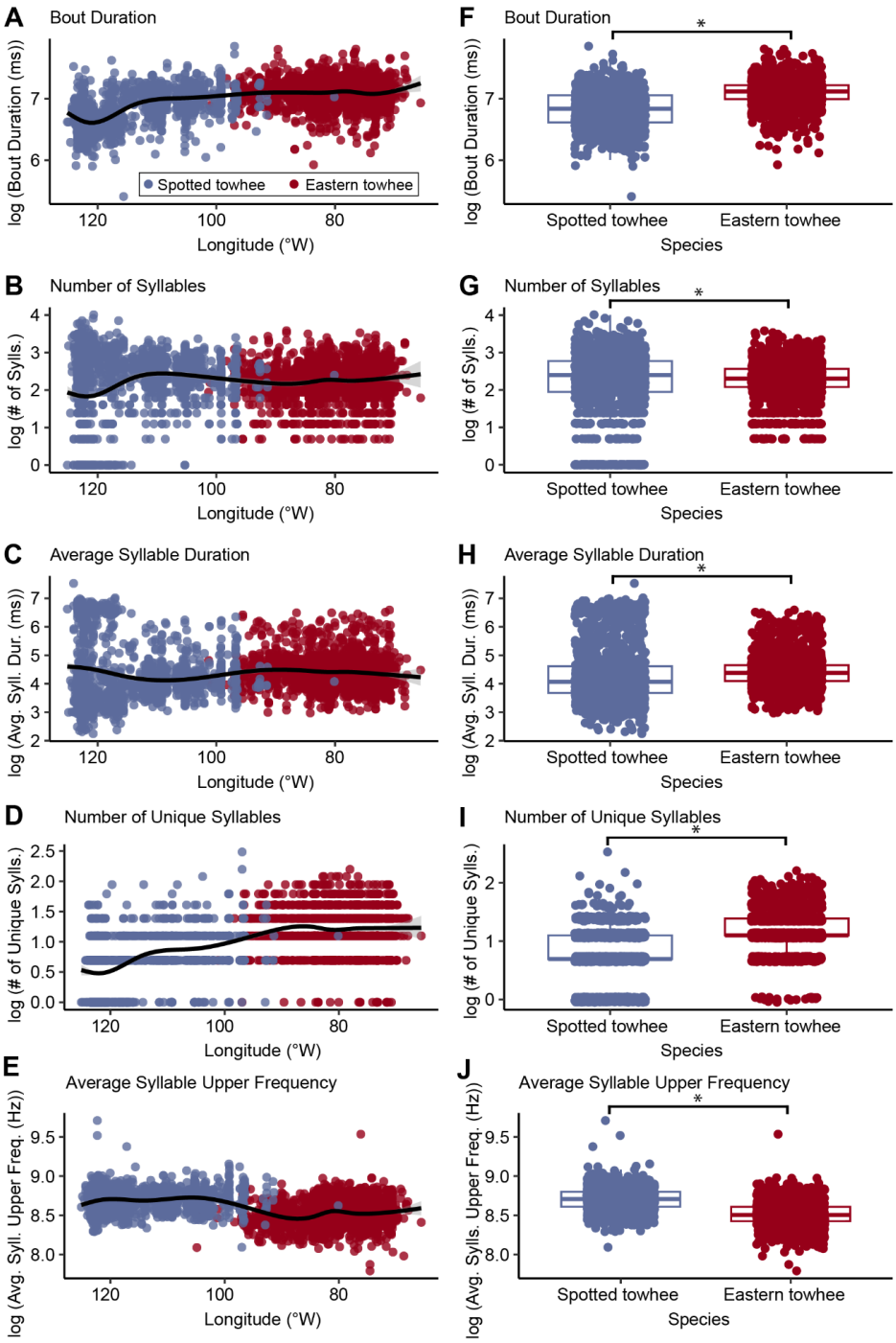

Figure 3. Spearman rank correlation (A-E) and Wilcoxon's rank sum test (F-J) of five song features in songs of Spotted towhees and Eastern towhees (N_{total_bouts}=3515; $N_{Spotted_towhee}$=1427; $N_{Eastern_towhee}$=2088). (Spearman: $\alpha_{adjusted}$=0.00156; Wilcoxon: $\alpha_{adjusted}$=0.003125). The black line on panels A-E represents a generalized additive model-fitting method. Asterisks (*) in panels F-J indicate statistical significance; Spearman's rank correlations were significant in panels A, C, D, and E; Number of Syllables was not significantly associated with longitude.

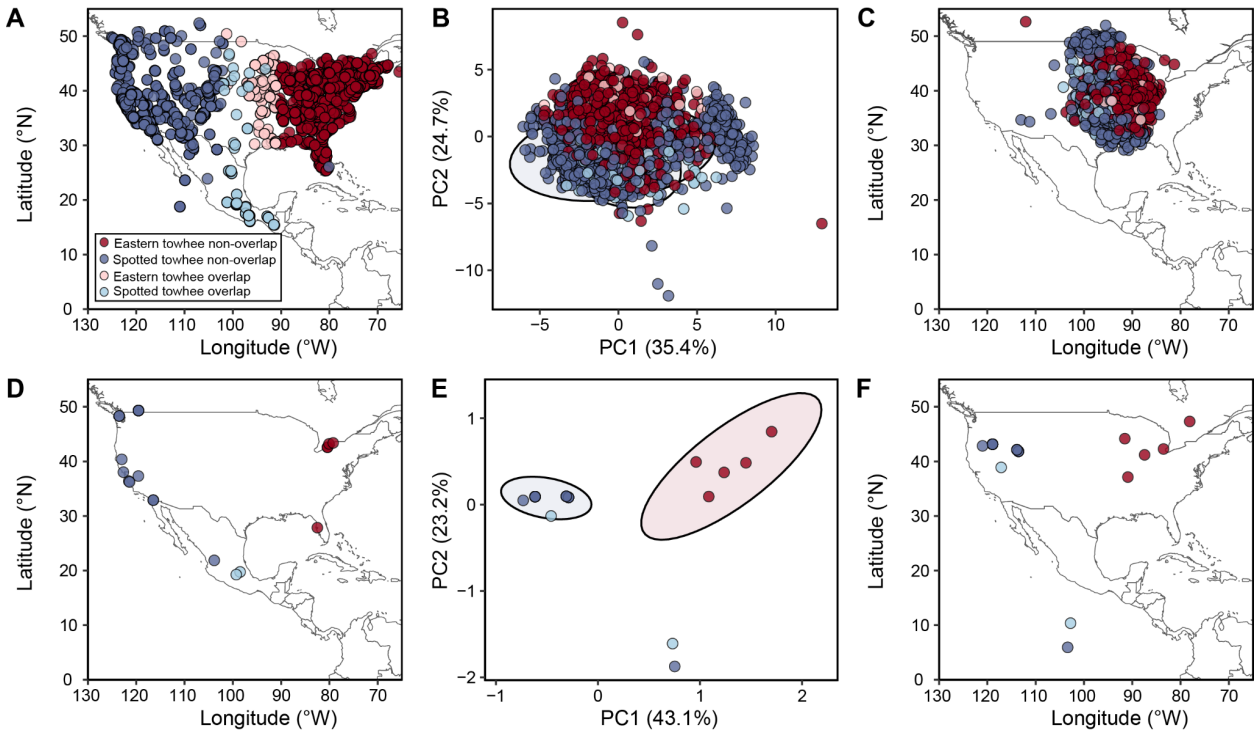

Figure 4. Spatial distribution of Spotted towhee and Eastern towhee song (A-C) and genomic (D-F) data. (A) Distribution of song recordings in North America ($N_{total_recordings}$=2788; $N_{Spotted_towhee}$=1069; $N_{Eastern_towhee}$=1719). (B) Principal component analysis (PCA) of song bouts using 16 song features (N_{total_bouts}=3515; $N_{Spotted_towhee}$=1427; $N_{Eastern_towhee}$=2088). (C) Procrustes analysis of song data using PC1 and PC2 from the PCA in panel B (N_{total_bouts}=3515; $N_{Spotted_towhee}$=1427; $N_{Eastern_towhee}$=2088). (D) Distribution of genetic sequences obtained from the Barcode of Life Data Systems database (N_{total}=23 ; $N_{Spotted_towhee}$=18; $N_{Eastern_towhee}$=5). (E) PCA using single nucleotide polymorphisms of aligned mtDNA sequences of the cytochrome oxidase subunit I regions of the genome (N_{total}=23 ; $N_{Spotted_towhee}$=18; $N_{Eastern_towhee}$=5). (F) Procrustes analysis of genomic data using PC1 and PC2 from the PCA analysis in panel E (N_{total}=23 ; $N_{Spotted_towhee}$=18; $N_{Eastern_towhee}$=5). Ellipses indicate 95% confidence intervals.

Machine learning

Our trained model using samples from the entire geographic range had an accuracy of 89.5% when tested on a subset of song samples. The most important feature in the decision tree was the average syllable upper frequency. Our model trained on the samples from the zone of only non-overlap was tested on a subset of samples from the zone of non-overlap and on samples from the zone of overlap which had an accuracy of 89.3% and 82.6%, respectively. The most important feature was the largest syllable duration.

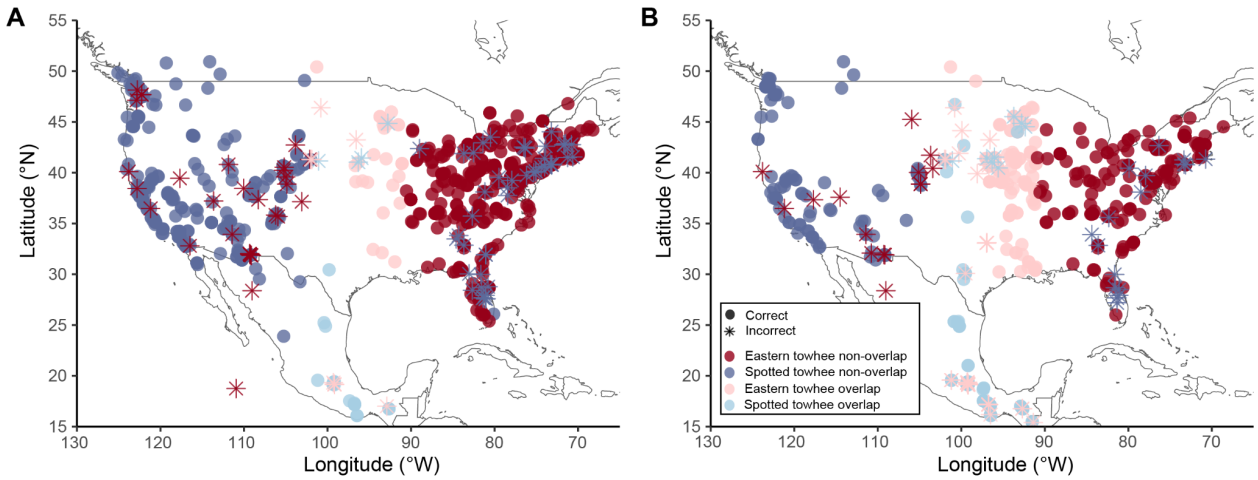

Figure 5. Geographic distribution of random forest model predictions of species identity based on a model trained on 16 song features from samples of Spotted towhees and Eastern towhees. (A) Predictions of subset of all song samples (N_{test}=879; accuracy = 89.5%) trained on song data from the entire geographic range of both species ($N_{Spotted_towhee}$=1062; $N_{Eastern_towhee}$=1062). (B) Model trained on a subset of samples obtained from the non-overlap zone ($N_{Spotted_towhee}$=1062; $N_{Eastern_towhee}$=1062). The model was tested on a random subsample of song bouts from both the zone of non-overlap ($N_{test_nonoverlap}$=299; accuracy = 89.3%) and the zone of overlap (i.e. 102°W - 91°W; $N_{test_overlap}$=299; accuracy = 82.6%).

PCA and Procrustes analysis of genetics data

A PCA using mtDNA sequences revealed 3 different clusters. Eastern towhees formed their own cluster, while the Spotted towhees were split into two clusters: 2 individuals from Mexico grouped together and a group which included all individuals from California, USA and Canada, and one individual from Mexico. Eastern and Spotted towhees cluster separately, suggesting that they are genetically different, at least using SNPs from COI mtDNA sequences. PC1 explained 43.1% of the variance in the genetic data, and PC2 explained 23.2%. A Procrustes analysis comparing the first two principal components of the genetic data to the latitude-longitude coordinates of the sampling location indicated that there is significant geographic signal in the genetic data (Procrustes rotation 0.85, $p < 0.001$).

Genomic analysis

Our genomic analysis revealed a pairwise F_{ST} value of 0.64, while an AMOVA showed that 63% of the variance was between species and 37% of the variance was within species. This agrees with the PCA analysis, where we see a separate cluster for the Eastern towhees (**Fig. 3E**).

DISCUSSION

Since Oscine songbirds learn their songs from conspecifics, they are an excellent model system for studying how culturally transmitted traits affect evolution and speciation. Specifically, the Spotted and Eastern towhees have broad ranges that span large distances. Since we have access to song recordings sampled throughout North America, we can understand variation that exists at a larger scale, allowing us to assess how traits can change with distance. This pair of sister species have overlapping ranges where it has been noted that individuals cross-mate (Fig. S2). Although hybridization occurs between these sister species, our analysis suggested levels of genetic differentiation similar to those of other avian species (Irwin et al. 2018). This suggests that the two species are genetically distinct, indicating reproductive isolation in these sister species. However, we had few samples, and the majority of them were from the outer edges of the ranges, which limits our ability to assess genetic differentiation in the overlap zone. Nonetheless, these populations are considered separate species due to genetic, morphological, and song differences and seem to maintain themselves as separate species. Therefore, individuals are likely using specific traits or a combination of traits to recognize conspecifics. A lot of variation in song features exists not only between sister species but also within species. In this study, we analyze potential song features that lead to the evolution and species of the Spotted and Eastern towhees, and we investigate whether song features could potentially allow individuals to reliably distinguish species-specific songs.

Our statistical analysis showed significant variation in song features across the longitude and latitude gradient overall, but showed stronger correlations and greater differences across longitude. Both number of syllables and bout duration decrease at around 120°W (**Fig. 3 A & B**). Additionally, the populations of Spotted towhees in the westernmost edge of the geographic range appear to have larger variation in some of the song features, including number of syllables and average syllable duration (**Fig. 3 B & C**); this pattern appears to correspond to the region where Spotted towhees have more rapid trills that occasionally were performed as a single syllable (with no separation between the repeated elements). Together, our results suggest that geographic variation alone is not responsible for the variation that exists in the songs of the Spotted and Eastern towhees. Further, in our Spearman rank correlation and Wilcoxon rank sum test, the average syllable upper frequency range seems to have a shift in the region where the two ranges meet (**Fig. 3 E & J**). This agrees with the highest loading in PC2, which captures a large

portion of the variation in the song data. [However, the highest loading for PC1 was average syllable duration.] The PCA revealed minimal separation in clustering using PC1 and PC2 (**Fig. 4B**), and a Procrustes analysis showed little geographic structure in the distribution of song data (**Fig. 4C**), suggesting that no one song feature can be used to reliably distinguish between species' songs. Even so, our random forest models were able to accurately predict songs of the Spotted and Eastern towhees with a minimum of 82.6% accuracy, supporting the notion that several song features, in combination with one another or with morphology, could allow individuals to reliably distinguish members of their own species. However, our random forest models trained on samples from the edges of the range were able to predict the species identification of songs from outside the zone of overlap better than those inside the zone of overlap (**Fig. 5B; Fig. S3**), supporting the notion that intermediate songs seem to exist (Sibley and West 1959). Our observation that songs of the Spotted and Eastern towhee are less distinguishable in the zone of overlap than at the edges of the ranges implies that this sister-species pair has not developed increased song differences in this zone of overlap to discourage hybridization. Nonetheless, the incorrect predictions are distributed relatively evenly across the geographic range.

Altogether, our analyses suggest that reinforcement of species boundaries is not readily detectable in towhee song, and other factors, such as habitat (e.g. vegetation density, altitude, climate), could also be contributing to the extensive amount of variation that we see in song features within and between species. For example, the acoustic adaptation hypothesis suggests that birdsong evolves under the constraints of the sound transmission properties of a given environment, and the songs are structured to increase the fidelity of song transmission in the native environment (Boughman 2002; Derryberry 2009; Morton 1975). As such, individuals with certain song traits may be more fit in specific environments if their song is able to transmit through the environment and be more easily detected. Additionally, as individuals transcend the zone of overlap into the other species' range, female preference for certain traits could lead to differential fitness in males. For example, if females prefer mates with traits that fall towards the extremes, those with intermediate songs would have decreased fitness. This may contribute to constraining the width of the hybrid zone and maintaining reproductive isolation. As such, future studies incorporating environmental factors at the population level would give us greater insight into factors affecting the evolution and speciation of avian species. Further, analyses that

account for geographic distance between songs could help assess whether songs accumulate changes by cultural drift, fitting an isolation-by-distance pattern. Additionally, differences in the trills of the Spotted and Eastern towhee may be features that are important for species recognition. A study by Richards (1981) suggests that the trills of the northeast populations of Eastern towhees function as a messaging component and may indicate species recognition. Therefore, future analyses could compare only the trills of Spotted and Eastern towhee songs to determine whether these trills have more species-specific differences than the songs as a whole. Finally, using playback experiments in the field to assess how individuals respond to songs of conspecifics versus songs of heterospecifics would directly test whether individuals can discriminate between songs and whether a preference exists for certain song features. In sum, we find that subtle song differences between Eastern and Spotted towhees can lead to relatively reliable species distinctions, but these distances are less apparent in the zone of species overlap, where they have been hypothesized to be most useful.

REFERENCES

Araya-Salas, Marcelo, and Grace Smith-Vidaurre. 2017. "WarbleR: An r Package to Streamline Analysis of Animal Acoustic Signals." *Methods in Ecology and Evolution* 8 (2): 184–91. https://doi.org/10.1111/2041-210X.12624.

Ball, R. Martin, and John C. Avise. 1992. "Mitochondrial DNA Phylogeographic Differentiation among Avian Populations and the Evolutionary Significance of Subspecies." *The Auk* 109 (3): 626–36.

Borror, Donald J. 1959. "Variation in the Songs of the Rufous-Sided Towhee." *The Wilson Bulletin* 71 (1): 54–72.

Boughman, Janette Wenrick. 2002. "How Sensory Drive Can Promote Speciation." *Trends in Ecology & Evolution* 17 (12): 571–77. https://doi.org/10.1016/S0169-5347(02)02595-8.

Catchpole, C. K., and Peter J. B. Slater. 1995. *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Cornell Lab of Ornithology. 2009. "Macaulay Library: Archive of Animal Sounds and Video." 2009. https://www.macaulaylibrary.org/.

DaCosta, Jeffrey M., Garth M. Spellman, Patricia Escalante, and John Klicka. 2009. "A Molecular Systematic Revision of Two Historically Problematic Songbird Clades: Aimophila and Pipilo." *Journal of Avian Biology* 40 (2): 206–16. https://doi.org/10.1111/j.1600-048X.2009.04514.x.

Derryberry, Elizabeth. 2009. "Ecology Shapes Birdsong Evolution: Variation in Morphology and Habitat Explains Variation in White-Crowned Sparrow Song." *The American Naturalist* 174 (February): 24–33. https://doi.org/10.1086/599298.

Excoffier, L, and H. E. L. Lischer. 2010. "Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows." *Molecular Ecology Resources*, no. 10: 564–67.

Fouquette, M. J., Jr. 1975. "Speciation in Chorus Frogs. I. Reproductive Character Displacement in the Pseudacris Nigrita Complex." *Systematic Biology* 24 (1): 16–23. https://doi.org/10.1093/sysbio/24.1.16.

Grant, Peter R., and B. Rosemary Grant. 1992. "Hybridization of Bird Species." *Science* 256 (5054): 193–98.

Hirase, Shotaro, Yusuke Yokoyama, Cin-Ty Lee, and Wataru Iwasaki. 2016. "The Pliocene-Pleistocene Transition Had Dual Effects on North American Migratory Bird Speciation." *Palaeogeography, Palaeoclimatology, Palaeoecology* 462 (November): 85–91. https://doi.org/10.1016/j.palaeo.2016.09.006.

Irwin, Darren E., Borja Milá, David P. L. Toews, Alan Brelsford, Haley L. Kenyon, Alison N. Porter, Christine Grossen, Kira E. Delmore, Miguel Alcaide, and Jessica H. Irwin. 2018. "A Comparison of Genomic Islands of Differentiation across Three Young Avian Species Pairs." *Molecular Ecology* 27 (23): 4839–55. https://doi.org/10.1111/mec.14858.

Jombart, Thibaut. 2008. "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24: 1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Katoh, Kazutaka, George Asimenos, and Hiroyuki Toh. 2009. "Multiple Alignment of DNA Sequences with MAFFT." In *Bioinformatics for DNA Sequence Analysis*, edited by David Posada, 39–64. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59745-251-9_3.

Kroodsma, Donald E. 1971. "Song Variations and Singing Behavior in the Rufous-Sided Towhee, Pipilo Erythrophthalmus Oregonus." *The Condor* 73 (3): 303–8.

https://doi.org/10.2307/1365756.
Kuhn, Max. 2008. “Building Predictive Models in R Using the Caret Package.” *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.
Lachlan, R. F., and M. R. Servedio. 2004. “Song Learning Accelerates Allopatric Speciation.” *Evolution* 58 (9): 2049–63. https://doi.org/10.1111/j.0014-3820.2004.tb00489.x.
MacDougall-Shackleton, Elizabeth A., and Scott A. MacDougall-Shackleton. 2001. “Cultural and Genetic Evolution in Mountain White-Crowned Sparrows: Song Dialects Are Associated with Population Structure.” *Evolution* 55 (12): 2568–75. https://doi.org/10.1111/j.0014-3820.2001.tb00769.x.
Moore, William S., and Jeff T. Price. 1993. “Nature of Selection in the Northern Flicker Hybrid Zone and Its Implications for Speciation Theory.” In *Hybrid Zones and the Evolutionary Process*. Oxford University Press.
Morton, Eugene S. 1975. “Ecological Sources of Selection on Avian Sounds.” *The American Naturalist* 109 (965): 17–34. https://doi.org/10.1086/282971.
Oksanen, J, G. Simpson, F. Blanchet, and Kindt et al. 2023. “Vegan: Community Ecology Package.” *R Package Version 2.6-5*. https://github.com/vegandevs/vegan.
Planque, R, and W. P. Vellinga. 2005. “Xeno-Canto: Sharing Bird Sounds from around the World.” www.xeno-canto.org.
Podos, Jeffrey. 2007. “Discrimination of Geographical Song Variants by Darwin’s Finches.” *Animal Behaviour* 73 (5): 833–44. https://doi.org/10.1016/j.anbehav.2006.11.001.
Podos, Jeffrey, and Paige S. Warren. 2007. “The Evolution of Geographic Variation in Birdsong.” In *Advances in the Study of Behavior*, 37:403–58. Academic Press. https://doi.org/10.1016/S0065-3454(07)37009-5.
Price, Trevor. 2007. *Speciation in Birds*. 1st ed. W. H. Freeman.
Ptacek, Margaret B. 2000. “The Role of Mating Preferences in Shaping Interspecific Divergence in Mating Signals in Vertebrates.” *Behavioural Processes* 51 (1): 111–34. https://doi.org/10.1016/S0376-6357(00)00123-6.
Richards, Douglas G. 1981. “Alerting and Message Components in Songs of Rufous-Sided Towhees.” *Behaviour* 76 (3/4): 223–49.
Rivera-Gutierrez, Hector F., Erik Matthysen, Frank Adriaensen, and Hans Slabbekoorn. 2010. “Repertoire Sharing and Song Similarity between Great Tit Males Decline with Distance between Forest Fragments.” *Ethology* 116 (10): 951–60. https://doi.org/10.1111/j.1439-0310.2010.01813.x.
Searfoss, Abigail M., Wan-chun Liu, and Nicole Creanza. 2020. “Geographically Well-Distributed Citizen Science Data Reveals Range-Wide Variation in the Chipping Sparrow’s Simple Song.” *Animal Behaviour* 161 (March): 63–76. https://doi.org/10.1016/j.anbehav.2019.12.012.
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 11 (4): 524–31. https://doi.org/10.1111/2041-210X.13368.
Sibley, Charles G., and David A. West. 1959. “Hybridization in the Rufous-Sided Towhees of the Great Plains.” *The Auk* 76 (3): 326–38. https://doi.org/10.2307/4081810.
Slabbekoorn, Hans, and Thomas B. Smith. 2002. “Bird Song, Ecology and Speciation.” *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357 (1420): 493–503. https://doi.org/10.1098/rstb.2001.1056.

Wright, M. N., and A. Ziegler. 2017. “Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.” *Journal of Statistical Software* 77 (1): 1–17.