Appendix: Supplemental Figures and Tables

Cultural shifts after punctuated environmental stress: a study of song distributions in Dark-eyed Junco and Song Sparrow populations

Supplemental Tables

	PC1	PC2	PC3
Bout duration (ms)	-0.1665459	-0.0036175	**-0.527575**
Rate of syllable production (Number of syllables divided by bout duration)	**-0.4292783**	0.15686655	0.08475945
Average syllable duration (ms)	**0.43914745**	-0.1548555	-0.1179026
Standard deviation of syllable duration (ms)	0.21535134	0.28518245	-0.0033744
Average syllable upper frequency (Hz)	0.1801962	-0.0907469	**-0.5171299**
Average syllable lower frequency (Hz)	-0.1943956	0.21603638	-0.0846287
Overall syllable frequency range (Hz)	0.26030926	-0.0903027	**-0.3856307**
Number of notes	-0.1073708	**0.4212219**	**-0.3498447**
Number of notes per syllable	0.24612039	**0.3398911**	-0.2007135
Number of syllables	**-0.4480448**	0.12904599	-0.207895
Number of unique syllables	0.06202007	**0.43995637**	0.10039593
Degree of repetition (Number of syllables divided by number unique)	**-0.3512576**	-0.2797122	-0.2353986
Mean syllable stereotypy	-0.1032396	**-0.4749758**	-0.0766961
Proportion of Variance	0.3065	0.2408	0.1317
Eigenvalues	3.984	3.130	1.713

Table S1A: Dark-eyed Junco principal component loadings by song feature. PC1 and PC3 were significantly different in the control region from before to after 2016, while none were significant in the drought region. The proportion of variance explained by each principal component is given in the last row. All loadings that have an absolute value greater than 0.30 are indicated in bold font. See **Figure S4**.

	PC1	PC2	PC3
Bout duration (ms)	0.00532734	**-0.4228521**	0.16240812
Number of syllables	**-0.3458911**	**-0.3243144**	0.20378128
Rate of syllable production (Number of syllables divided by bout duration)	**-0.4226098**	-0.0564736	0.11013442
Average syllable duration (ms)	**0.42324314**	-0.0529731	-0.0739714
Standard deviation of syllable duration (ms)	**0.36019293**	-0.1267016	0.24656715
Number of unique syllables	-0.18839	**-0.4027593**	-0.3541724
Degree of repetition (Number of syllables divided by number unique)	-0.1685814	0.11461891	**0.61940589**
Average syllable upper frequency (Hz)	0.11807169	-0.2409636	0.29407423
Average syllable lower frequency (Hz)	0.04844985	0.15346658	**0.37252899**
Overall syllable frequency range (Hz)	-0.0826014	**-0.3750778**	-0.0079921
Number of notes	0.09335848	**-0.4648783**	0.00834772
Number of notes per syllable	**0.36555865**	-0.1786854	-0.1467235
Largest syllable duration (ms)	**0.33641169**	-0.1659606	**0.31019402**
Smallest syllable duration (ms)	0.22430933	0.14398953	-0.0381991
Proportion of Variance	0.3275	0.2108	0.1097
Eigenvalues	4.585	2.951	1.535

Table S1B: Song Sparrow principal component loadings by song feature. PC1 and PC3 were significantly different in the Drought region from before to after 2016, while none were significant in the control region. The proportion of variance explained by each principal component is given in the last row. All loadings that have an absolute value greater than 0.30 are indicated in bold font. See **Figure S4**.

Feature	log Transformed	GLS Model ANOVA p-value	ANOVA EraRegion_Pval	Holm-Bonferroni p-value threshold for ANOVA	ANOVA significant after Holm-Bonferroni	Before vs. After Drought adjusted p-value	Before vs. After Control adjusted p-value
Bout duration (ms)	TRUE	0.00000	**0.00002**	0.00385	**TRUE**	**0.02972**	**0.00250**
Rate of syllable production (Number of syllables divided by bout duration)	TRUE	0.21972	0.08965	0.00625	FALSE	0.49660	0.30911
Average syllable duration (ms)	TRUE	0.55248	0.17608	0.00714	FALSE	0.98802	0.21707
Standard deviation of syllable duration (ms)	TRUE	0.38112	0.22953	0.00833	FALSE	0.85697	0.21497
Average syllable upper frequency (Hz)	TRUE	0.28981	0.83666	0.05000	FALSE	0.98253	0.93992
Average syllable lower frequency (Hz)	FALSE	0.48985	0.80492	0.02500	FALSE	0.95494	0.98707
Overall syllable frequency range (Hz)	FALSE	0.20999	0.23523	0.01000	FALSE	0.82179	0.50029
Number of notes	TRUE	0.93393	**0.00026**	0.00417	**TRUE**	**0.00903**	**0.01367**
Number of notes per syllable	TRUE	0.02066	0.38254	0.01250	FALSE	0.07828	0.93970
Number of syllables	TRUE	0.13100	**0.00074**	0.00455	**TRUE**	0.15941	**0.00671**
Number of unique syllables	TRUE	0.00007	0.59496	0.01667	FALSE	0.99997	0.81726
Degree of repetition (Number of syllables divided by number unique)	TRUE	0.00006	**0.00169**	0.00500	**TRUE**	0.18574	**0.03243**
Mean syllable stereotypy	TRUE	0.35367	0.08171	0.00556	FALSE	0.62669	*0.03667*

Table S2. Statistical analysis of Dark-eyed Junco songs. We performed Shapiro-Wilk tests on the subset of data in the control region and the subset of data in the drought region. When either of these were significant, indicating non-normality, we log-transformed the feature values ('TRUE' in column 'log transformed'). We used nlme::gls() to test whether there were significant differences in any features between the four groups. We first performed this gls() as a simple model with no weights and as a complex model, with weights set by varIdent(form = ~1 | Era*Region). We performed an ANOVA to test whether the complex model was a significantly better fit than the simple model and used the simple model if this was not significant, and the complex model if it was (p-values in column "GLS model ANOVA p-value"). We performed an ANOVA on the selected model for each song feature and corrected for multiple hypothesis testing with a Holm-Bonferroni correction. Four song features varied significantly between groups, bout duration, the total number of syllables, degree of syllable repetition, and the total number of notes per bout. (bolded) One other feature, syllable stereotypy, trended towards significance (italics). For the four song features with significant ANOVA results, we performed a post-hoc test to test whether they were significantly different before vs. after the drought in the drought and control regions using emmeans::emmeans(), adjusting p-values using emmeans::test() p-adjustment method "mvt". All four song features differed in the control region, and bout duration and total number of notes also differed in the drought region. Syllable stereotypy approached significance in the control region only.

Feature	log Transformed	GLS Model ANOVA p-value	ANOVA EraRegion_Pval	Holm-Bonferroni p-value threshold for ANOVA	ANOVA significant after Holm-Bonferroni	Before vs. After Drought adjusted p-value	Before vs. After Control adjusted p-value
Bout duration (ms)	TRUE	0.29550	0.17625	0.00833	FALSE	0.38504	0.66716
Number of syllables	FALSE	0.65918	**0.00035**	0.00385	**TRUE**	**0.00181**	0.06993
Rate of syllable production (Number of syllables divided by bout duration)	TRUE	0.03806	**0.00160**	0.00417	**TRUE**	**0.00605**	0.11626
Average syllable duration (ms)	TRUE	0.00579	*0.00773*	0.00455	FALSE	*0.03925*	0.16804
Standard deviation of syllable duration (ms)	TRUE	0.00010	0.11816	0.00556	FALSE	0.30681	0.46208
Number of unique syllables	TRUE	0.35866	0.66125	0.01250	FALSE	0.89607	0.66437
Degree of repetition (Number of syllables per number unique)	TRUE	0.00582	**0.00006**	0.00357	**TRUE**	**0.00009**	0.11308
Average syllable upper frequency (Hz)	TRUE	0.00135	0.70977	0.01667	FALSE	0.96779	0.94271
Average syllable lower frequency (Hz)	TRUE	0.05068	0.79582	0.02500	FALSE	0.40549	0.45888
Overall syllable frequency range (Hz)	TRUE	0.00011	0.35661	0.01000	FALSE	0.48824	0.87767
Number of notes	TRUE	0.13728	0.87595	0.05000	FALSE	0.70086	0.93917
Number of notes per syllable	TRUE	0.00032	*0.02713*	0.00500	FALSE	*0.00151*	0.53597
Largest syllable duration (ms)	TRUE	0.00145	0.14997	0.00714	FALSE	0.25391	0.59648
Smallest syllable duration (ms)	TRUE	0.44932	0.12925	0.00625	FALSE	0.81418	0.05474

Table S3. Statistical analysis of Song Sparrow songs. Methods are the same as in **Table S2**. Three song features varied significantly between groups, number of syllables, rate of syllable production (number of syllables divided by bout duration), and degree of syllable repetition (number of syllables divided by number of unique syllables) (bolded), and two others (average syllable duration and mean number of notes per syllable) showed trends towards significance (italics). For the three song features with significant ANOVA results, we performed a post-hoc test to test whether they were significantly different before vs. after the drought in the drought and control regions using emmeans::emmeans(), adjusting p-values using emmeans::test() p-adjustment method “mvt”. All three song features differed in the drought region but not in the control region.

	Dark-eyed Junco - Drought	Dark-eyed Junco - Control	Song Sparrow - Drought	Song Sparrow - Control
Mean distance between all points (km)	24.47	103.6	39.90	63.51
Max distance between any points (km)	257.3	296.2	278.9	218.0
Mean distance to nearest neighbor (km)	3.024	4.660	1.144	5.063
Area of convex hull polygon (sq km)	18360	31320	14920	24920

Table S4: Spatial metrics of each region. Latitude and longitude of each recording transformed using UTM zone 18 projection and sp:spTransform(). Convex hull polygon calculated using rgeos:gConvexHull() and rgeos:gArea().

Supplemental Figures

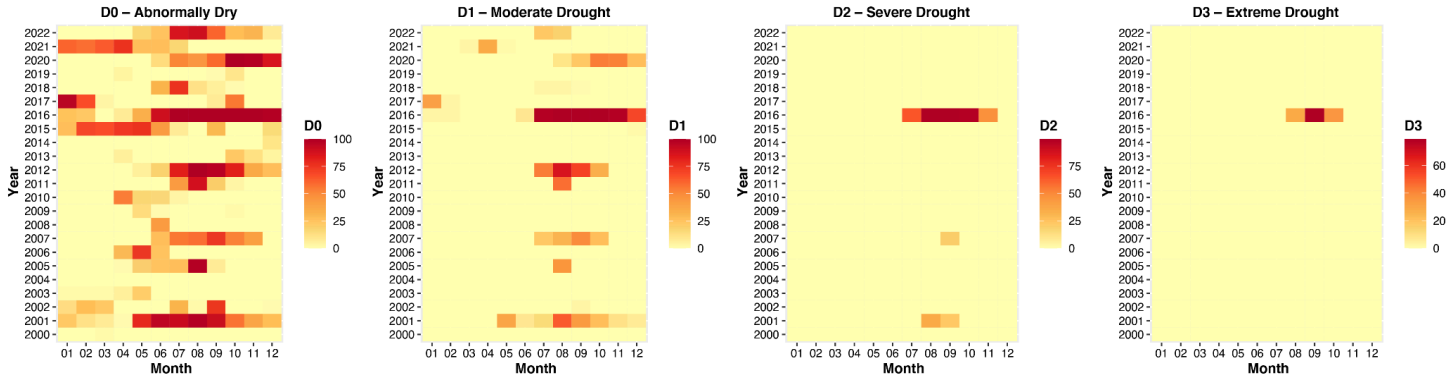

Figure S1: Mean percent of land area in New York counties in the drought region classified as drought level D0-D3 or higher by month and year from 2000-2022. Counties included are Cayuga, Erie, Genesee, Livingston, Ontario, Schuyler, Seneca, Tompkins, Wyoming, and Yates. The year 2016 was a clear outlier in drought severity and area, as it was the only year that most of the land area across these counties was in drought level D2 ("severe") and the only year that any of the listed counties had land classified as drought level D3 ("extreme") or higher. The drought levels referenced here are assigned by National Oceanic and Atmospheric Administration's National Integrated Drought Information System (NIDIS) program and take into account variables including temperature, precipitation, soil moisture, water levels in streams and lakes, and others, with drought level D3 indicating extreme drought, corresponding to major crop and pasture losses and widespread water shortages or restrictions. Data obtained from the Drought Monitor (Svoboda et al. 2002), accessed via https://droughtmonitor.unl.edu/NADM/Home.aspx. The colorbar indicates the average percentage of each county that met or exceeded the indicated level of drought.

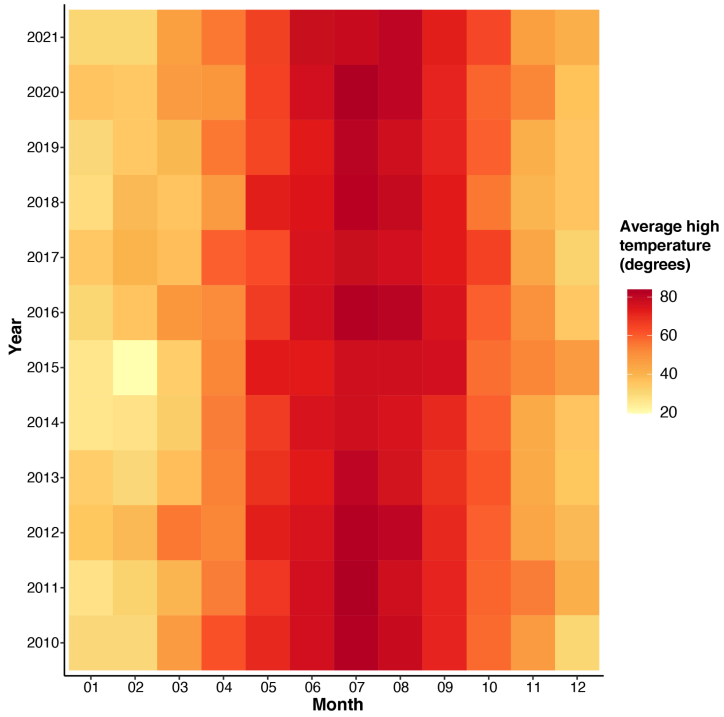

Figure S2: Mean of mean maximum daily recorded temperature per month across stations with complete records in Tompkins County, New York. There is no year that is a clear outlier in the period between 2010 and 2021. Data from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information, accessed via https://www.ncei.noaa.gov/cdo-web/datatools.

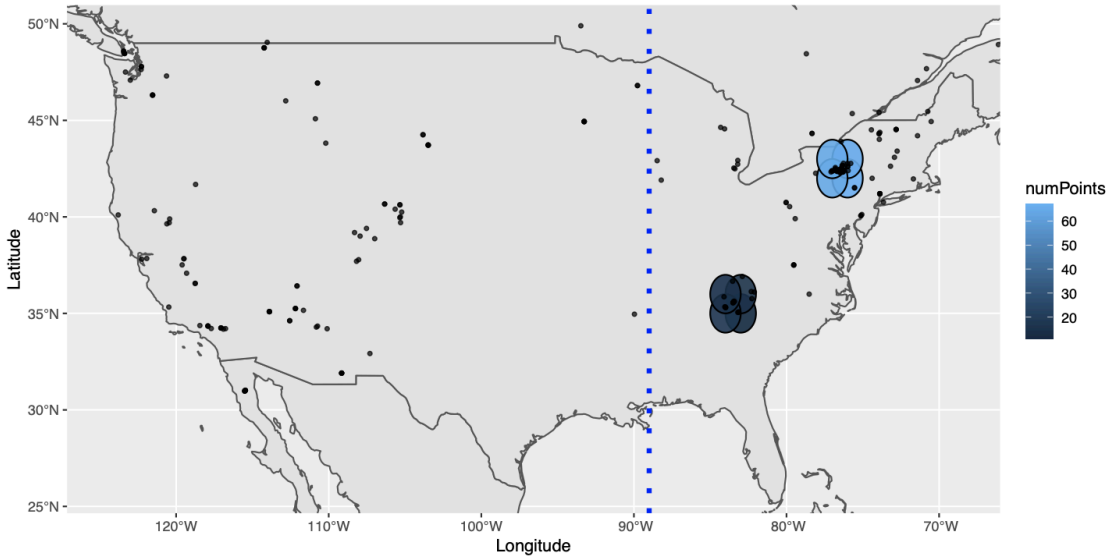

Figure S3A: Dark-eyed Junco recordings publicly available on Macaulay Library or Xeno-canto recorded between 2006 and 2016. Of all regions centered on each integer value of (Longitude, Latitude) ranging from 70°W through 89°W (the westernmost longitude we considered in deciding target regions, denoted by the dotted blue line) and 30°N through 47°N, the plotted, filled oval regions (area with a radius of 1°) are the only regions that contain at least 10 recordings between 2006 and 2016 (pre-Drought). The fill color is scaled to the number of recordings available within that area.

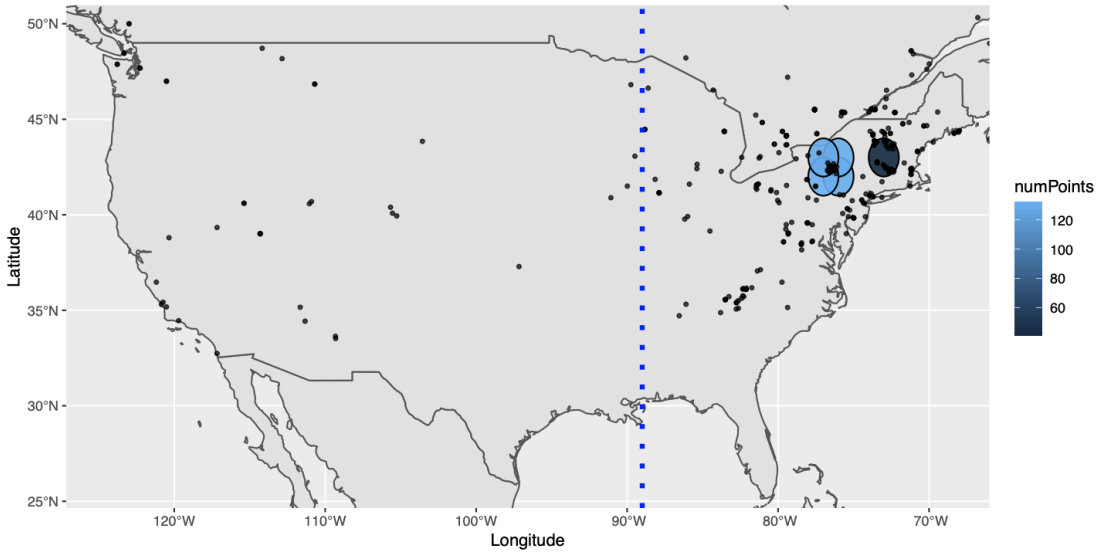

Figure S3B: Dark-eyed Junco recordings publicly available on Macaulay Library or Xeno-canto recorded between 2017 and 2019. Of all regions centered on each integer value of (Longitude, Latitude) ranging from 70°W through 89°W (the westernmost longitude we considered in deciding target regions, denoted by the dotted blue line) and 30°N through 47°N, the plotted, filled oval regions (area with a radius of 1°) are the only regions that contain at least 40 recordings between 2017 and 2019 (post-Drought).

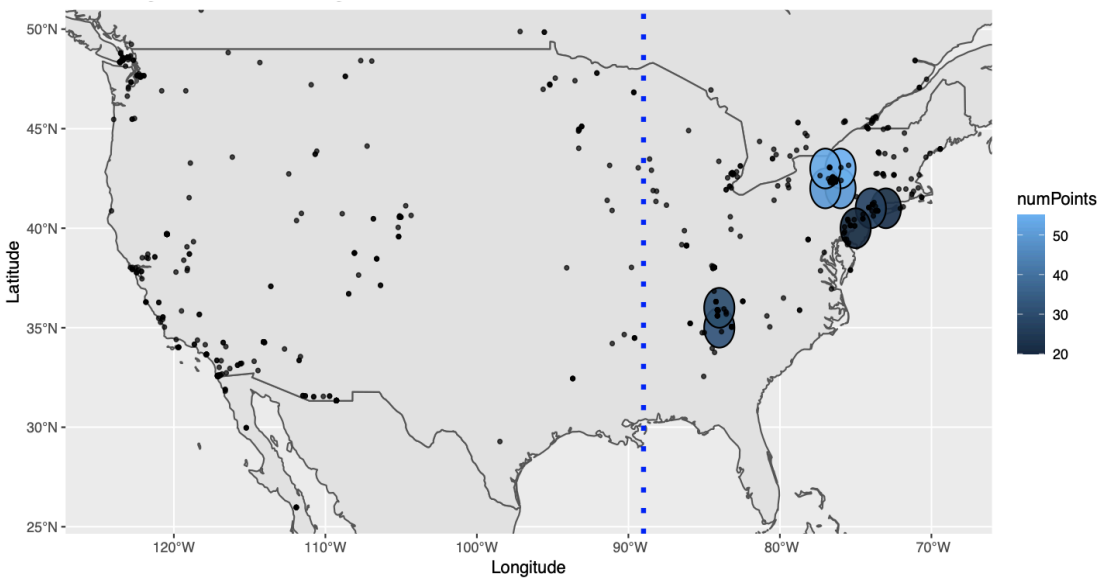

Figure S3C: Song Sparrow recordings publicly available on Macaulay Library or Xeno-canto recorded between 2006 and 2016. Of all regions centered on each integer value of (Longitude, Latitude) ranging from 70°W through 89°W (the westernmost longitude we considered in deciding target regions, denoted by the dotted blue line) and 30°N through 47°N, the plotted, filled oval regions (area with a radius of 1°) are the only regions that contain at least 20 recordings between 2006 and 2016 (pre-Drought).

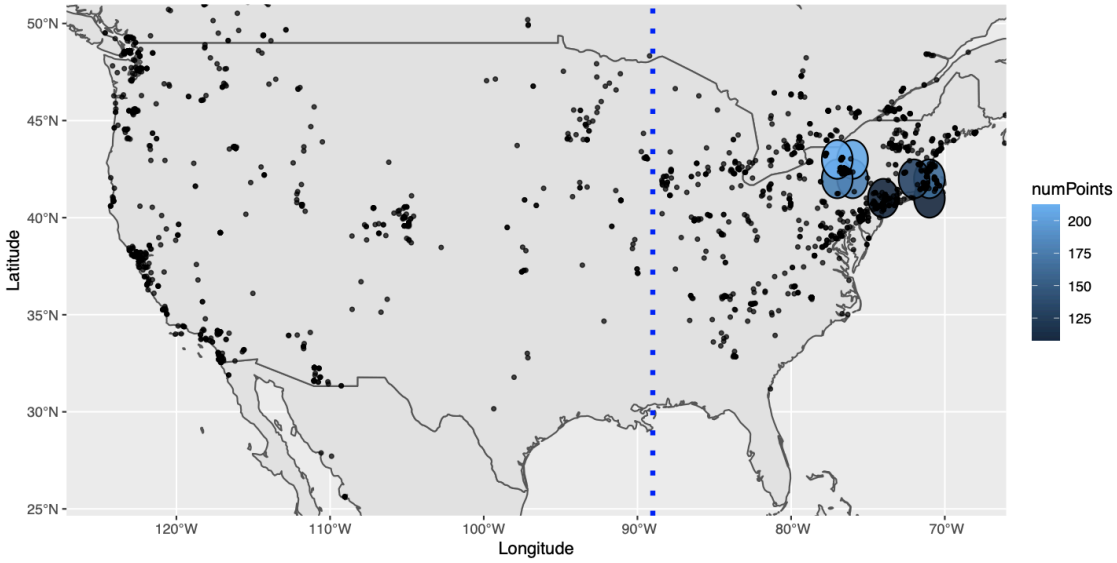

Figure S3D: Song Sparrow recordings publicly available on Macaulay Library or Xeno-canto recorded between 2017 and 2019. Of all regions centered on each integer value of (Longitude, Latitude) ranging from 70°W through 89°W (the westernmost longitude we considered in deciding target regions, denoted by the dotted blue line) and 30°N through 47°N, the plotted, filled oval regions (area with a radius of 1°) are the regions that contain at least 100 recordings between 2017 and 2019 post-Drought).

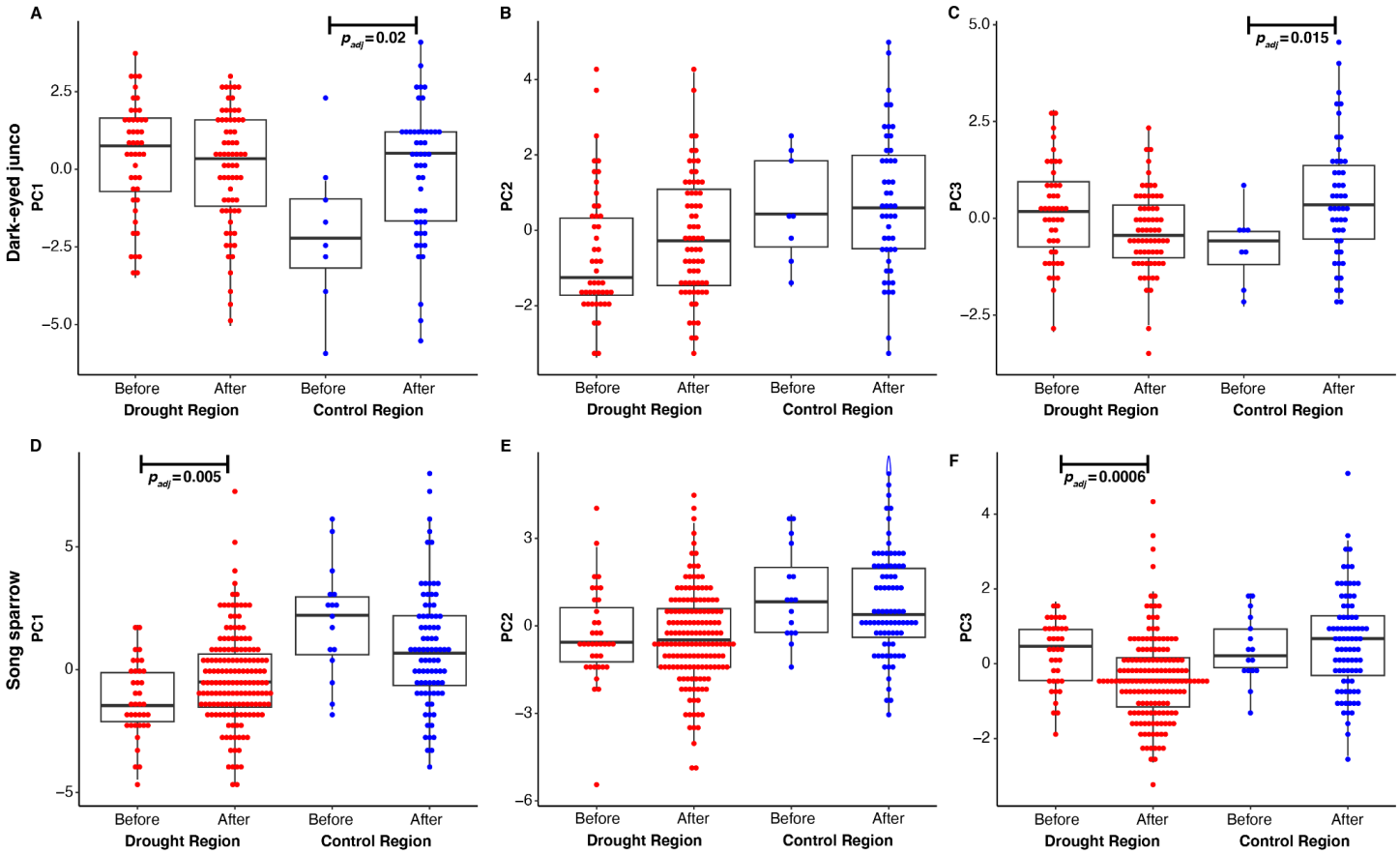

Figure S4: Distributions of the first three principal components of the song feature data for Song Sparrows and Dark-eyed Juncos. Statistical analyses performed as in the individual song features. For each set of Principal Component scores, we assessed whether there were differences between groups using a GLS model and post-hoc tests, as in **Figures 3** and **5** in the main text. We used the simple GLS model, which assumed equal weights for each group) unless an ANOVA indicated that the variable-weight model was a significantly better fit (variable-weights model used in panels C, D, F). Weights of song features in each dimension are provided in **Table S1**. Overall, there were significant regional differences in song features, with two of the three dimensions differing significantly between regions in Dark-eyed Juncos (ANOVA by Region: PC1 $p = 0.068$, PC2 $p = 0.0011$, PC3 $p = 0.020$), and three of the three dimensions differing by region in Song Sparrows (ANOVA by Region: PC1, PC2, PC3 $p < 1 \times 10^{-7}$).

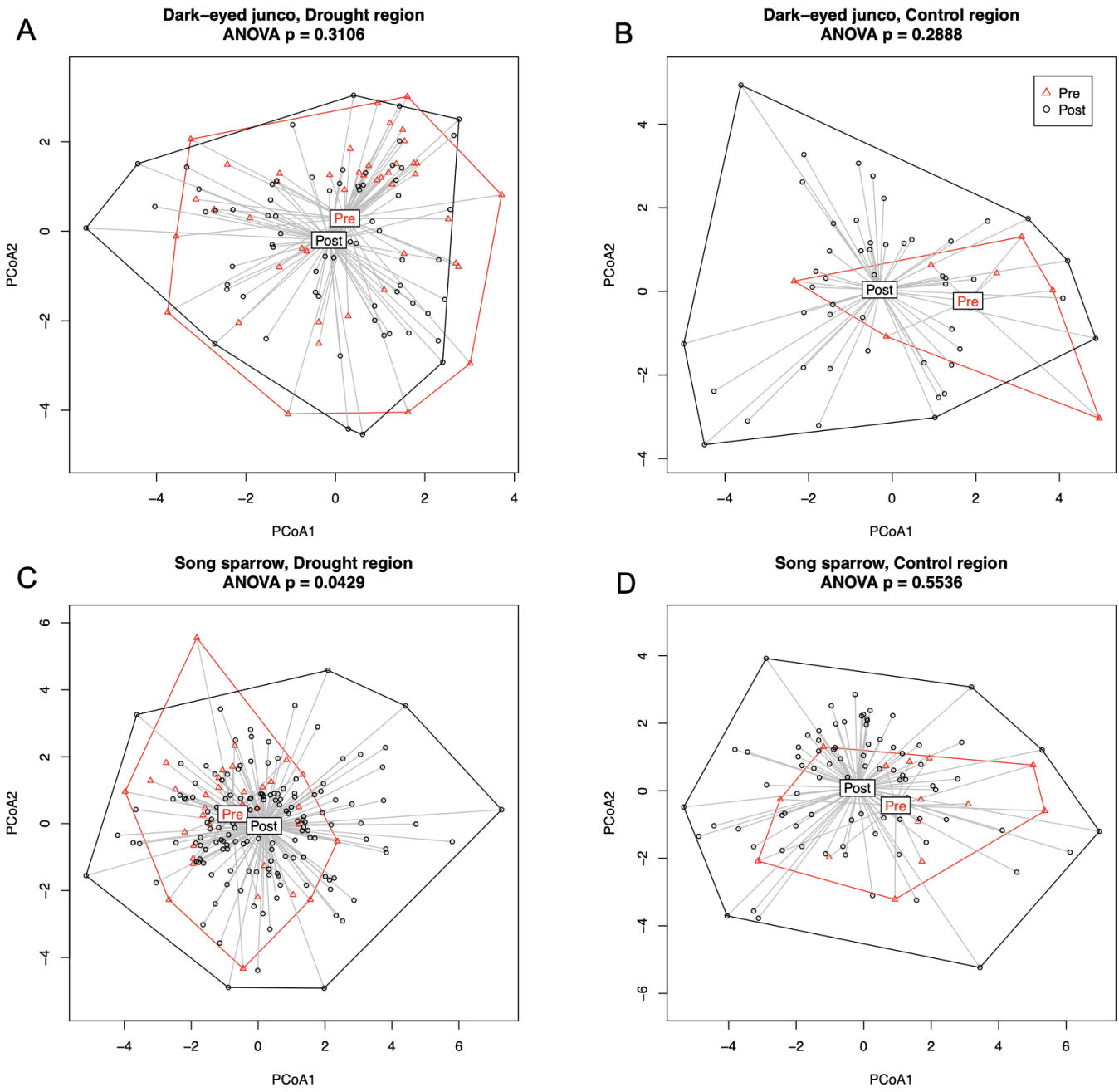

Figure S5: Multivariate test of homogeneity of song feature variances by Region and Era. This analysis was conducted with the function betadisper (R package "vegan") on the PCoA-transformed song-feature data. The ANOVA of this output tests for differences in the distances from members of each group to its group's centroid. Only the variances in Song Sparrow songs in the drought region were significantly different before versus after 2016 (Panel C).

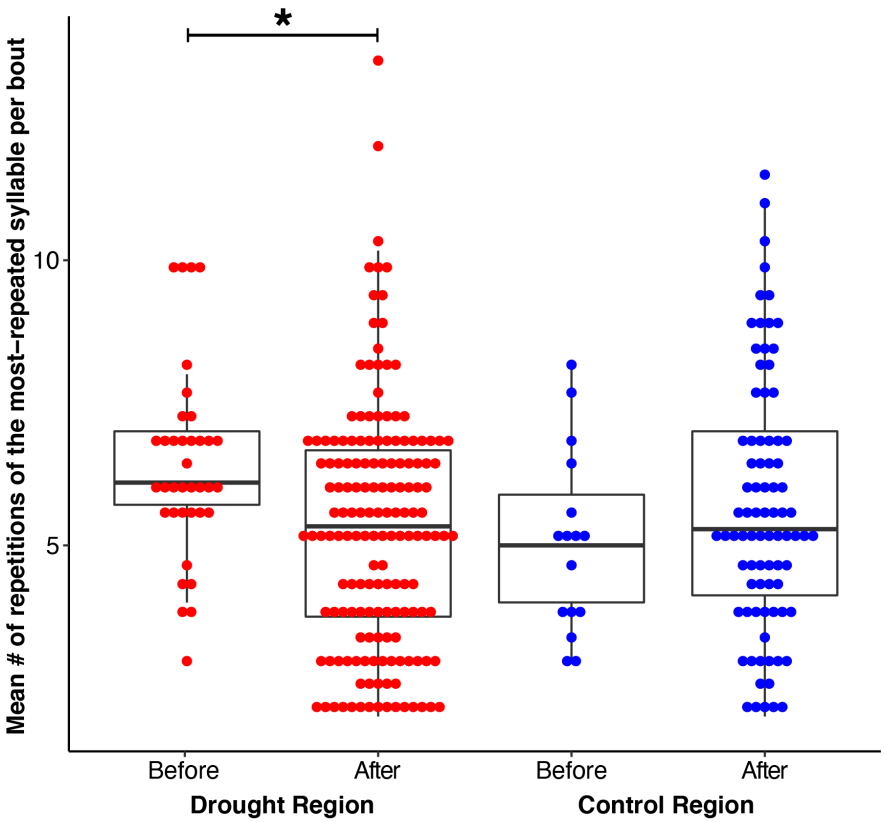

Figure S6: Distributions of average counts of the most-repeated syllable per bout per recording within Song Sparrow populations before and after the 2016 drought. For each bout, we found the most repeated syllable type based on the syllable type assignment from the Chipper analysis, counted the number 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 (GLS ANOVA: $p = 0.016$; nlme::gls with no weights: Drought region pre vs post $p = 0.009$, Control region pre vs post $p = 0.542$) .

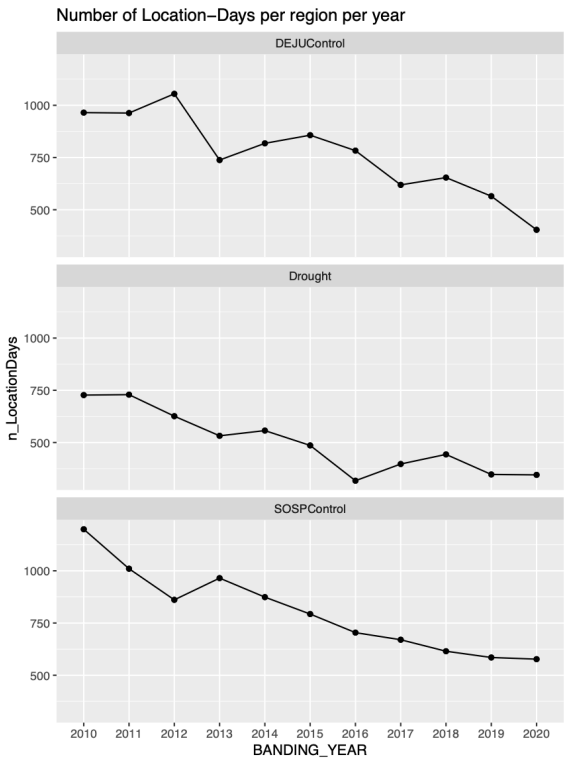

Figure S7: The number of unique banding location-days in the USGS database per year per region, used as a proxy for effort.

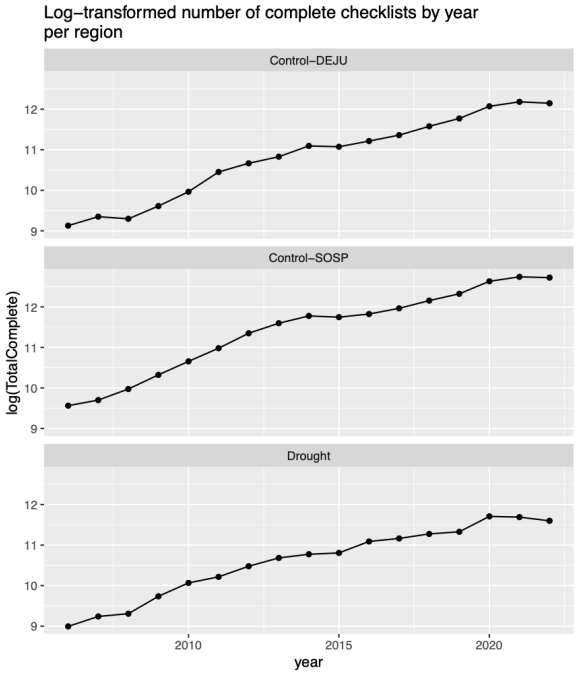

Figure S8: The number of complete checklists in the eBird database per year per region, used as a proxy for effort. Numbers were log-transformed (natural log) for ease of visualization.

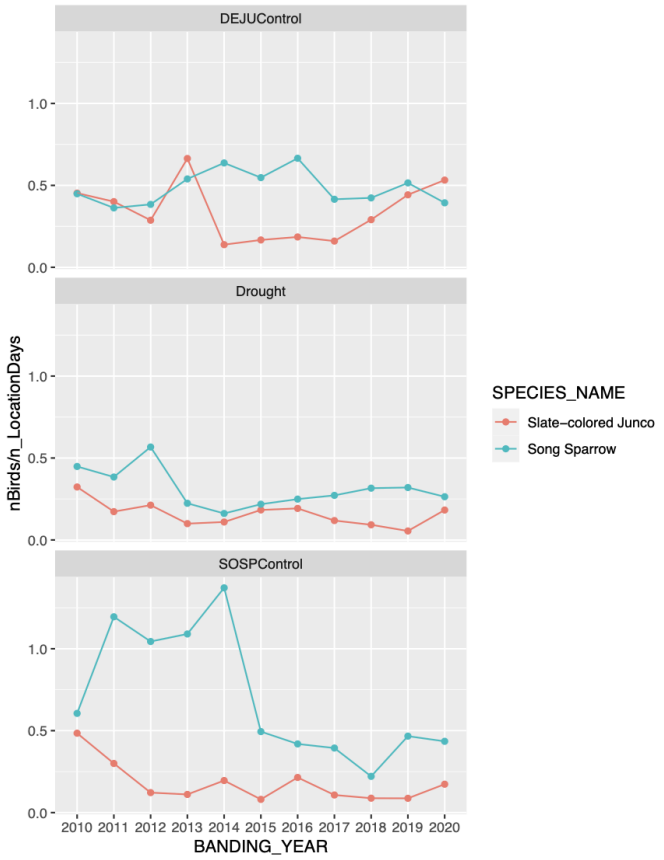

Figure S9: Relative species abundances in each of our three regions as reported by United States Geological Society (USGS) bird banding data. For each year between 2010 and 2020, we divided the number of Dark-eyed Juncos (*Junco hyemalis hyemalis*, Slate-colored Junco subspecies, shown in red) and Song Sparrows (shown in teal) of any age class banded in a region by the total number of days at each unique banding site that banding was conducted in that region. For example, a value greater than one indicates that, on average, more than one bird of that species was banded each time a banding effort was conducted in the region.

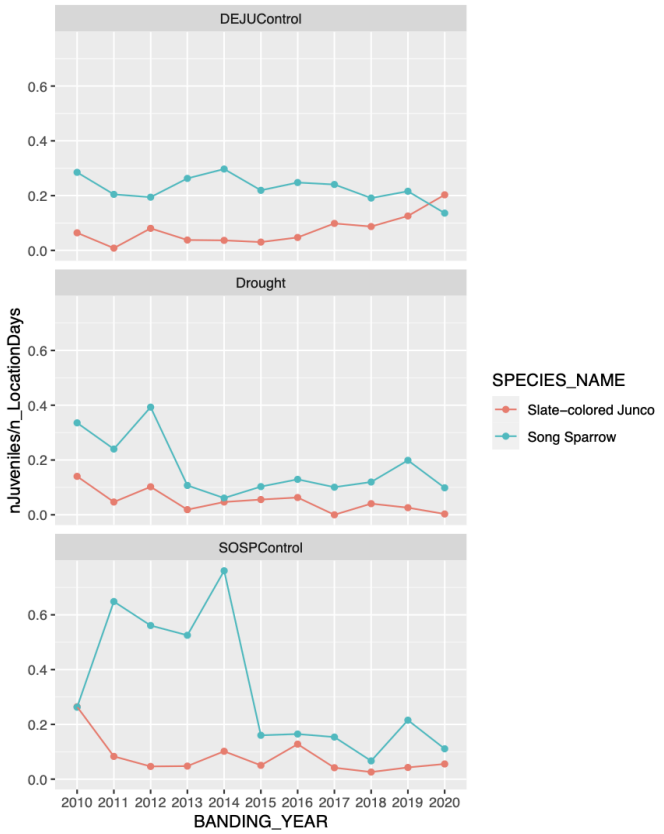

Figure S10: Juvenile bird abundances in each of our three regions as reported by United States Geological Society (USGS) bird banding data. For each year between 2010 and 2020, we divided the number of juvenile Slate-colored Juncos and Song Sparrows observed in a region by the total number of days at each unique banding site that banding was conducted in that region. Observations labeled "Hatch year" or "Local" (defined as "A nestling or young bird incapable of sustained flight") were counted as juvenile observations.

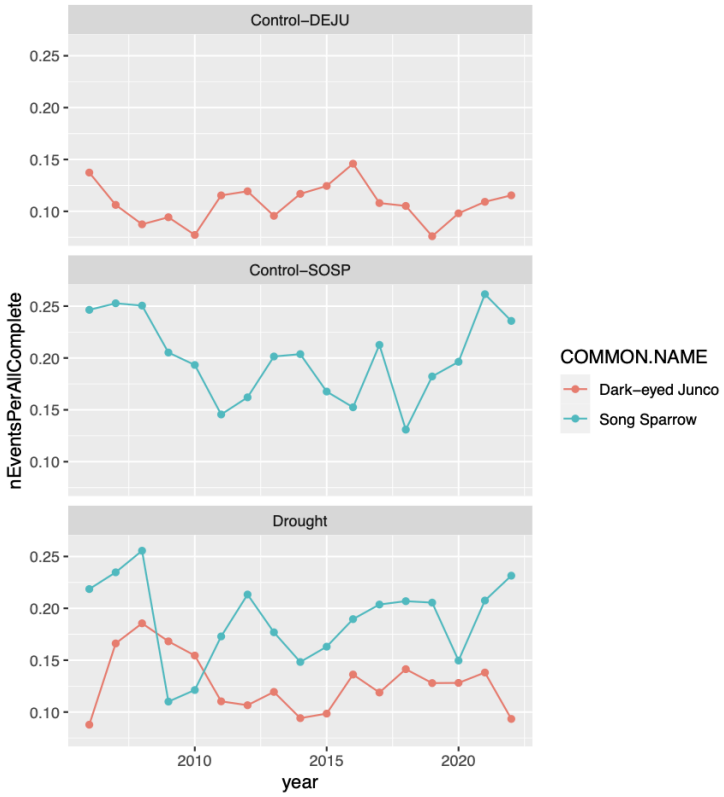

Figure S11: Relative species abundances in each of our three regions as reported by eBird user observation data. For each year between 2006 and 2022, we divided the number of completed eBird checklists that included Dark-eyed Juncos (red) and Song Sparrows (teal) by the total number of checklists that were completed in that region.

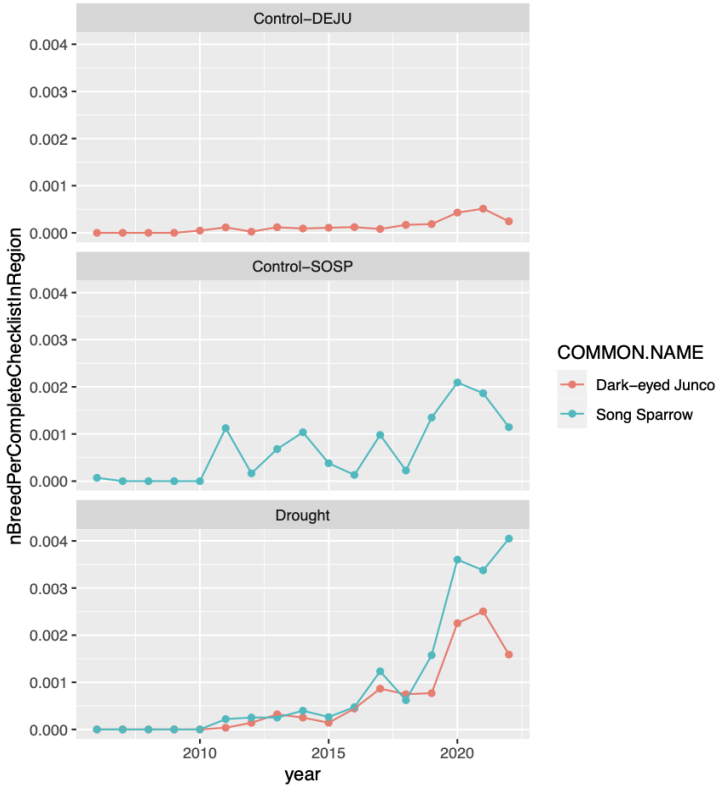

Figure S12: Breeding behavior observations in each of our three regions as reported by eBird user observation data. For each year between 2006 and 2022, we divided the number of completed eBird checklists that mentioned breeding behaviors in Dark-eyed Juncos (red) and Song Sparrows (teal) by the total number of checklists that were completed in that region.