Evolution, 2024, **78(11)**, 1774–1789
https://doi.org/10.1093/evolut/qpae118
Advance access publication 30 August 2024
Original Article

Go west, young bunting: recent climate change drives rapid movement of a Great Plains hybrid zone

Paul J. Dougherty[1,2,3] **and Matthew D. Carling**[1,2,3]

[1]Department of Zoology & Physiology, University of Wyoming, Laramie, WY, United States
[2]University of Wyoming Program in Ecology, Laramie, WY, United States
[3]University of Wyoming Museum of Vertebrates, Laramie, WY, United States

Corresponding author: Department of Zoology & Physiology, University of Wyoming, Laramie, WY 82071, United States. Email: pjdougherty@umass.edu

Abstract

Describing how hybrid zones respond to anthropogenic influence can illuminate how the environment regulates both species distributions and reproductive isolation between species. In this study, we analyzed specimens collected from the *Passerina cyanea* × *P. amoena* hybrid zone between 2004 and 2007 and between 2019 and 2021 to explore changes in genetic structure over time. This comparison follows a previous study that identified a significant westward shift of the *Passerina* hybrid zone during the latter half of the twentieth century. A second temporal comparison of hybrid zone genetic structure presents unique potential to describe finer-scale dynamics and to identify potential mechanisms of observed changes more accurately. After concluding that the westward movement of the *Passerina* hybrid zone has accelerated in recent decades, we investigated potential drivers of this trend by modeling the influence of bioclimatic and landcover variables on genetic structure. We also incorporated eBird data to determine how the distributions of *P. cyanea* and *P. amoena* have responded to recent climate and landcover changes. We found that the distribution of *P. cyanea* in the northern Great Plains has shifted west to track a moving climatic niche, supporting anthropogenic climate change as a key mediator of introgression in this system.

Keywords: hybridization, Great Plains, eBird, climate change

Introduction

Human activity has inflicted dramatic effects on the distributions and abundances of organisms throughout the world (Rosenberg et al. 2019). While anthropogenic climate change, landcover change, and species introductions have all unsurprisingly sparked grave conservation concerns, associated shifts in the kinds and frequencies of biological interactions present novel opportunities for investigating the processes of population divergence and speciation (Larson et al. 2019; Taylor et al. 2015). Hybrid zones, areas where two distinct evolutionary lineages come into contact and interbreed, in particular provide a lens through which researchers can observe the evolution of reproductive isolation in real time and identify environmental factors that mediate introgression (Harrison 1993; Grabenstein and Taylor 2018). Therefore, studying how human activity influences the structure of contemporary hybrid zones can illuminate the processes that led to existing patterns of diversity (McFarlane and Pemberton 2019), what future species assemblages will look like (Taylor et al. 2015), and possibly strategies to mitigate future biodiversity loss (Allendorf et al. 2001; McFarlane and Pemberton 2019).

The best way to definitively understand the effects of human activity on introgression among divergent populations is to describe changes in the genetic structure of hybrid systems over time (Buggs 2007; Carling and Zuckerberg 2011). Unfortunately, only a few hybrid zones have been repeatedly sampled over long enough time frames for significant changes to be detectable (Britch et al. 2001; Dasmahapatra et al. 2002; Krosby and Rohwer 2010; Carling and Zuckerberg 2011; Billerman et al. 2019; Ryan et al. 2018; Walsh et al. 2020; Suh et al. 2022). Moreover, even in rare cases when replicate sampling efforts allow for detection of changes in hybrid zone structure, identifying the causes of observed changes can be difficult (Buggs 2007). For instance, shifts in the geographic center of a hybrid zone can signal that parental taxa are adapted to different habitat types and are tracking shifting ecological conditions (Taylor et al. 2014), or may alternatively result from asymmetric abundances of parental taxa (Carling and Zuckerberg 2011), selective advantage of alleles from one parental taxon (Barton and Hewitt 1985), or social dominance of one parental taxon over the other (McDonald et al. 2001; Krosby and Rohwer 2010). Changes in the width of a hybrid zone can similarly arise from several different scenarios. For instance, if hybrids experience higher fitness than unadmixed individuals in intermediate or novel environmental conditions, changes in hybrid zone width will reflect the creation, expansion, or contraction of such habitat types (Carling and Zuckerberg 2011). Alternatively, endogenous factors (e.g., hybrid genetic incompatibilities or heterosis) may

Received October 26, 2023; revisions received July 22, 2024; accepted August 27, 2024
Associate Editor: Nina Therkildsen; Handling Editor: Tim Connallon

be more important in determining relative hybrid fitness and therefore also driving changes in hybrid zone width (Bierne et al. 2013; Galindo et al. 2014). If hybrid fitness is lower than that of unadmixed individuals, selection against heterospecific mating will cause the hybrid zone to become more narrow and may precipitate the establishment of parental taxa as completely reproductively isolated species (Ortiz-Barrientos et al. 2009; Irwin 2020).

Overall, the ways in which the center and width of a hybrid zone change over time depends on the specific characteristics of the system (Carling and Zuckerberg 2011; Natola et al. 2023). Understanding how human activity mediates gene flow among populations requires not only replicate sampling efforts of a hybrid system but also detailed investigation into the demographic history, ecology, and behavior of parental taxa as well as the relationship between ancestry and fitness.

The avian hybrid zones of the Great Plains

The Great Plains of North America are the site of a suture zone where many plant and animal taxa come into secondary contact and hybridize with a close relative from which they were repeatedly isolated by glaciations during the Pliocene and Pliestocene (Remington 1968; Swenson and Howard 2004; Swenson 2006; Carling and Zuckerberg 2011). It has been proposed that human forestation of riparian corridors in the Great Plains over the past century has greatly facilitated contact between many eastern and western lineage pairs, particularly for the five avian hybrid systems that occur in this suture zone. These include the hybrid zones between the eastern yellow-shafted and the western red-shafted Northern Flicker subspecies (*Colaptes auratus auratus* and *C. a. cafer*), between the Eastern Towhee (*Pipilo erythrophthalmus*) and the Spotted Towhee (*P. maculatus*), between the Baltimore Oriole (*Icterus galbula*) and the Bullock's Oriole (*I. bullocki*), between the Rose-breasted Grosbeak (*Pheucticus ludovicianus*) and the Black-headed Grosbeak (*P. melanocephalus*), and between the Indigo Bunting (*Passerina cyanea*) and the Lazuli Bunting (*P. amoena*). All of these species occur in forest, forest edge, or wooded riparian habitats (Anderson 1971; Carling and Zuckerberg 2011; Carling and Thomassen 2012). The resulting extensive hybridization of familiar bird taxa in the Great Plains has attracted repeated collecting efforts since the 1950s, making these hybrid zones excellent models for detecting the influence of anthropogenic factors on gene flow and introgression (Carling and Thomassen 2012; Sibley and Short Jr. 1959; Sibley and Short 1964; Rising 1996). Historic specimen series resulting from past collecting efforts represent an invaluable baseline against which modern specimen series can be compared (Walsh et al. 2020; Aguillon and Rohwer 2022).

The Great Plains suture zone occurs at an ecotone between mesic deciduous forest habitat in eastern North America and comparatively xeric woodlands in the west (Carling and Zuckerberg 2011). Thus, many systems are characterized by ecological differences that may render the distributions of parental taxa (and the occurrence of hybridization) highly sensitive to environmental shifts (Rohwer and Irwin 2011; Carling and Thomassen 2012). For instance, rising (1969) found that *Icterus bullocki* from western North America are more tolerant of high temperatures than the eastern *I. galbula*. Similarly, Swenson and Carling and Thomassen found associations between temperature and precipitation variables and the occurrence of *Passerina cyanea*, *P. amoena*, and hybridization between these 2 species (2006; 2012).

In this study, we estimate recent temporal changes in the hybrid zone between the Indigo Bunting (*P. cyanea*) and the Lazuli Bunting (*P. amoena*). These two species hybridize regularly where their ranges overlap in the Great Plains (Figure 1). The hybrid zone has been sampled along similar east-west transects sporadically beginning in 1955, with the most recent collecting effort in 2004-2007 (Sibley and Short Jr. 1959; Emlen et al. 1975; Kroodsma 1975; Carling et al. 2010). Carling and Zuckerberg compared specimens collected between 2004 and 2007 to historic specimen series and noted a westward shift in the *Passerina* hybrid zone over the previous 40 years (2011). They speculated the shift may be due to climatic changes and increased *P. cyanea* abundance in the Midwest. They also noted a narrowing of the hybrid zone, which they attributed to low hybrid fitness selecting for stronger prezygotic isolating mechanisms (Carling and Zuckerberg 2011). After resampling the same transect, we test if the trends in hybrid zone location and width described by Carling and Zuckerberg have continued under ongoing anthropogenic influence. We also incorporate previously unavailable eBird and environmental datasets to test potential mechanisms driving any observed changes. Given that anthropogenic climate change has accelerated in the past two decades (Watanabe et al. 2014; Smith et al. 2015), comparing recent shifts in the *Passerina* hybrid zone to the historic shifts described by Carling and Zuckerberg highlights the environmental factors most important for regulating gene flow in this system.

Materials and methods

Population sampling

We collected *Passerina* buntings from 21 sites between 2004 and 2007 and fourteen sites between 2019 and 2021 (Tables 1 and 2, Figure 2). For both collection times, we selected sites to enable sampling of putatively allopatric *P. cyanea* at the eastern sites, putatively allopatric *P. amoena* at the western sites, and sympatric, hybridizing populations at middle sites. All collection took place between late May and early July to ensure sampling of breeding individuals. We also included five *P. amoena* collected in southeastern Wyoming in the summer of 2018 as part of the 2019–2021 series. Overall, we collected 147 specimens between 2004 and 2007 and 82 specimens between 2018 and 2021. All collecting activity between 2018 and 2021 followed approved University of Wyoming IACUC protocols #20180720MC00319-01 and #20190416MC00358-01. All round skins, spread wings, and vouchered tissue specimens generated from 2004 to 2007 collecting are housed at the Louisiana State University Museum of Natural Science, while specimens from 2019 to 2021 collecting are housed at the University of Wyoming Museum of Vertebrates.

Molecular methods, variant calling, and ancestry estimation

For all birds collected between 2018 and 2021, we extracted genomic DNA from pectoralis tissue using a Qiagen DNeasy blood and tissue kit and a QIAcube robot following the manufacturer's standard extraction protocol (Qiagen, Inc.). One library was prepared using a genotyping-by-sequencing protocol (Parchman et al. 2012). DNA samples were fragmented and prepared with the restriction endonucleases EcoRI and

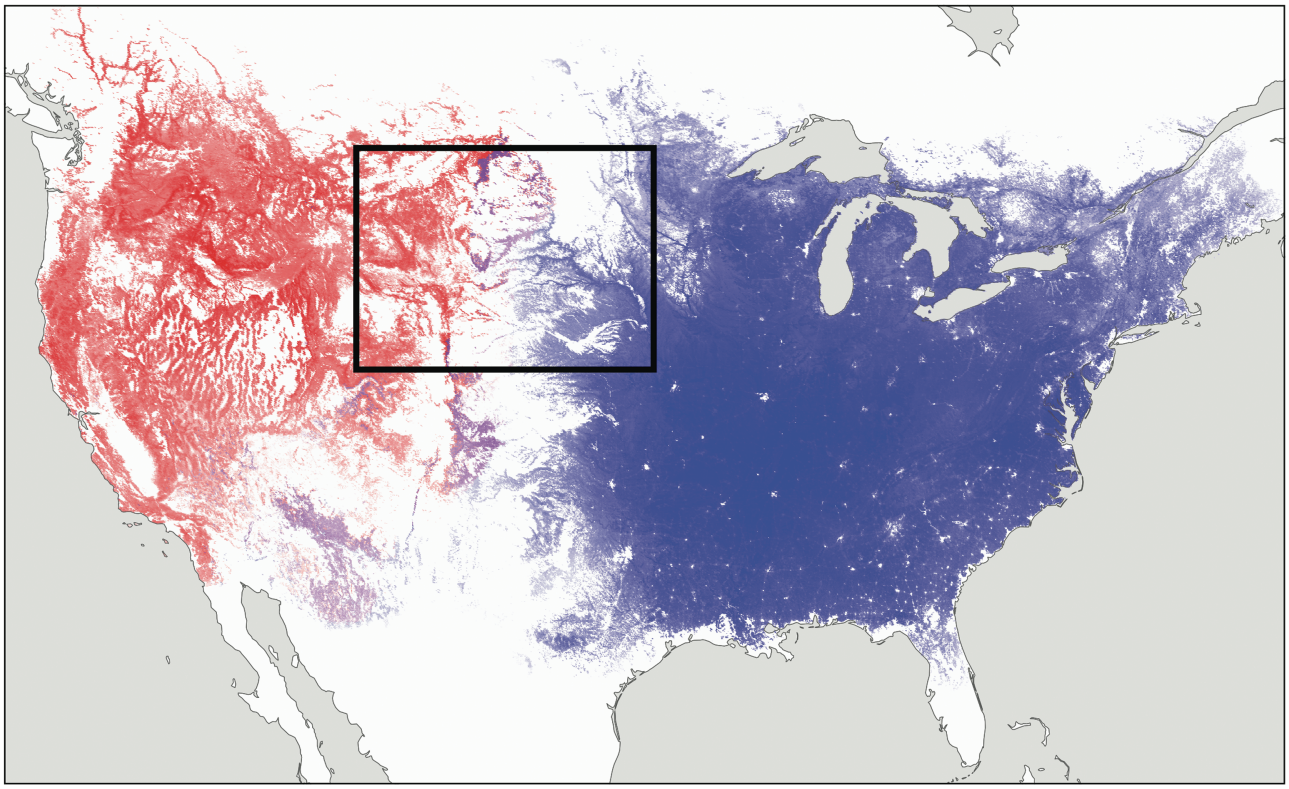

Figure 1. The current breeding distributions of *P. cyanea* (blue) and *P. amoena* (red) estimated from eBird data (http://www.ebird.org). Purple represents areas where both species are predicted to occur and hybridization is possible. The black box encompasses sites for 2004-2007 and 2019-2021 collecting efforts (Figure 2).

Table 1. Sampling localities, sample sizes, and mean hybrid index (0 = unadmixed *P. amoena*, 1 = unadmixed *P. cyanea*)†for populations sampled in 2004-2007. Table includes only individuals retained in final cline analyses.

Locality	Latitude	Longitude	*N*	Mean Hybrid Index
NE: Wiseman wildlife management area	42.75	-97.12	10	0.9997
SD: Newton Hills Game Production Area	43.23	-96.57	13	0.9997
ND: Pigeon Point Preserve (The Nature Conservancy)	46.50	-97.38	4	0.997
NE: Niobrara Valley Preserve (The Nature Conservancy)	42.78	-100.02	18	0.994
SD: Carpenter Game Production Area	43.70	-99.43	13	0.958
NE: Nenzel (private)	42.80	-101.10	4	0.750
NE: Ponderosa Wildlife Management Area	42.62	-103.32	7	0.470
SD: Ft. Meade National Recreation Area	44.23	-103.27	4	0.378
WY: Sand Creek (private)	44.53	-104.08	11	0.237
NE: White River (private)	42.62	-103.53	8	0.172
SD: Black Hills National Forest	43.72	-103.82	5	0.152
NE: Bordeaux Wildlife Management Area	42.75	-102.92	8	0.052
CO: Roosevelt National Forest	40.67	-105.23	7	0.0447
SD: Whitney Preserve (The Nature Conservancy)	43.33	-103.55	3	0.0203
WY: Medicine Bow National Forest	42.38	-105.32	3	0.0188
WY: Bighorn National Forest no. 2	44.57	-107.67	5	0.0172
ND: Little Missouri National Grassland	46.77	-103.52	11	0.0123
SD: Custer National Forest	45.83	-103.47	10	0.00887
MT: Custer National Forest no. 2	45.68	-106.03	5	0.00654
MT: Custer National Forest no. 1	45.08	-108.50	4	0.00556
WY: Bighorn National Forest no. 1	44.83	-107.30	3	0.000325

†Entropy estimates of individual hybrid index are never perfectly 0 or 1, even for unadmixed individuals, due to intraspecific genetic variation.

MseI. Fragments shorter than 275 base pairs and longer than 375 base pairs were excluded. Ten base pair barcodes denoting an individual bird were ligated to each fragment. Prior to sequencing, a 10% PhiX spike was added to ensure sufficient variability in the first 25 bases of neighboring clusters. Prepared fragments were then multiplexed and amplified by PCR. The library was sequenced on a NextSeq 2000 P3 cartridge to produce 1 billion 138 base pair single-end reads. All sample quality control, library preparation, and sequencing were performed at the University of Wyoming Genome Technologies Laboratory in Laramie, Wyoming.

For all birds collected between 2004 and 2007, we followed an identical protocol to generate genotyping-by-sequencing data. However, no PhiX spike was added to these samples.

Table 2. Sampling localities, sample sizes, and mean hybrid index (0 = unadmixed *P. amoena*, 1 = unadmixed *P. cyanea*) for populations sampled in 2019-2021. Table includes only individuals retained in final cline analyses.

Locality	Latitude	Longitude	N	Mean Hybrid Index
NE: Wiseman Wildlife Management Area	42.75	-97.12	4	0.997
NE: Bazille Creek Wildlife Management Area and Niobrara State Park	42.77	-98.07	3	0.954
NE: Redbird Wildlife Management Area	42.76	-98.44	1	0.999
NE: Keller Wildlife Management Area and Keller Park State Recreation Area	42.67	-99.78	4	0.973
NE: Niobrara Valley Preserve (The Nature Conservancy)	42.78	-100.02	10	0.979
NE: Smith Falls State Park	42.89	-100.31	4	0.986
NE: Anderson Bridge Wildlife Management Area	42.79	-100.93	4	0.967
NE: Chadron State Park	42.71	-103.01	1	0.9996
NE: Ponderosa Wildlife Management Area	42.62	-103.32	3	0.082
NE: Fort Robinson State Park	42.66	-103.46	27	0.471
NE: Gilbert Baker Wildlife Management Area	42.77	-103.93	2	0.137
WY: Tom Thorne / Beth Williams Wildlife Management Area	41.78	-105.37	12	0.024
WY: Medicine Bow National Forest, Murphy Canyon Road	42.22	-105.36	2	0.019
WY: Medicine Bow National Forest, Curtis Gulch Campground	42.41	-105.62	3	0.003

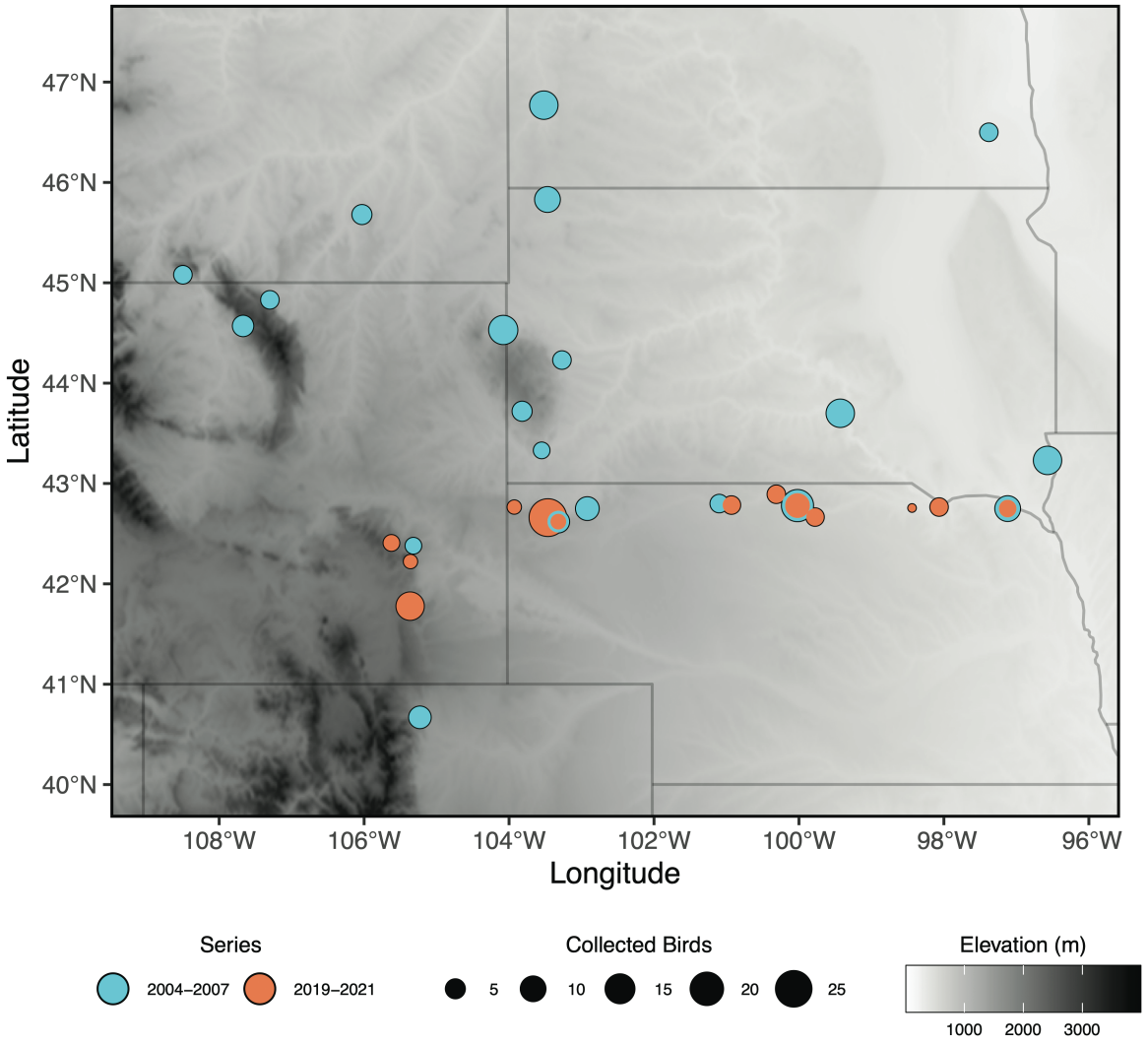

Figure 2. *Passerina* collecting sites colored by time series. The relative size of each point represents the total number of birds collected at each place and time.

The library was sequenced on an Illumina HiSeq 2000 at the National Center for Genome Resources in Santa Fe, New Mexico.

Following sequencing, we demultiplexed raw reads to match sequence barcodes to individual birds. We then aligned reads to the Northern Cardinal (*Cardinalis cardinalis*) whole genome (GenBank accession GCA_014549065.1)(Sin et al. 2020) using the BWA-MEM algorithm (Li 2013), then identified variable sites using "SAMtools" and "BCFtools" (Danecek et al. 2021). We used "VCFtools" to retain only

biallelic loci, to remove loci with a minor allele frequency less than 1% and with data in fewer than 30% of individuals, and to retain only sites 20,000 base pairs from each other to reduce linkage disequilibrium. We also removed individuals missing data for at least 70% of retained sites (Danecek et al. 2011). We performed alignment, variant calling, and filtering for 2004-2007 samples and 2018-2021 samples together to generate a single VCF file for both series. We performed this and all following sequence processing on the University of Wyoming's Beartooth Computing Cluster (https://doi.org/10.15786/M2FY47).

We used "entropy," a hierarchical Bayesian model for describing population structure, to estimate the genomic hybrid index (q) and proportion of heterozygous sites (Q) for each bird (Gompert et al. 2014; Shastry et al. 2020) in the two series. We assumed two genetic clusters (k=2) (i.e., *P. cyanea* and *P. amoena*) and assigned a prior ancestry proportion of 0.5 for all individuals. We ran three replicate chains, each with a different number seed for the random number generator, for 120,000 steps, recording every 30th step, and discarding the first 30,000 steps as burn-in. We assessed chain convergence for q and Q values for each bird by examining trace plots.

We classified each bird into different ancestry classes based on estimates of q and Q. Specifically, we classified individuals with a proportion of *P. cyanea* ancestry (q) less than 0.1 and greater than 0.9 as unadmixed *P. amoena* and *P. cyanea*, respectively. We defined individuals as F1 hybrids if they had q between 0.4 and 0.6 and $Q > 0.8$, and F2 hybrids if they had q between 0.4 and 0.6 but Q between 0.4 and 0.6. We defined backcrosses to *P. amoena* as individuals with a value for $Q-2q$ between –0.1 and 0.1 and backcrosses to *P. cyanea* as individuals with a value for $Q+2q$ between 1.9 and 2.1. These cutoffs were developed by Mandeville et al. (2019) based on plausible values and uncertainty of "entropy" estimates (Lindtke et al. 2014).

Estimating change in the hybrid zone over time

We performed all data preparation and analyses for this and all subsequent sections in Program R (version 4.3.1). To estimate changes in the center location and width of the *Passerina* hybrid zone between the two sampling events, we used the "hzar" package to fit clines for each time series using the Metropolis–Hastings Markov chain Monte–Carlo (MCMC) algorithm (Derryberry et al. 2014). While the sampling locations of the two time series both occur primarily in northern Nebraska, southern South Dakota, and eastern Wyoming, there is considerably greater variation in sampling latitude in the 2004-2007 series than in the 2019-2021 series (Figure 2). To explore the extent to which ancestry varies with latitude, thereby identifying the most appropriate method to estimate the distance of each collection site for cline models (Wang et al. 2019), we modeled q as a function of latitude and longitude for all samples collected between 2004 and 2007 as a generalized additive mixed model (GAMM) with collection site as a random effect using the "mgcv" package (Wood 2001). We assumed a quasibinomial error distribution with a logit link function, assumed a cubic regression spline smooth function for latitude and longitude, and used cross-validation to automatically estimate the proper number of knots (Zuur et al. 2009). A generalized additive model framework allows for the likely possibility that ancestry varies nonlinearly with latitude and longitude. Because this model predicted q values for birds collected in 2004-2007 to vary little with sampling latitude (Supporting Information Figs. S1) and that collection sites in 2019-2021 followed a relatively straight east-west transect, we used sampling longitude as the distance predictor in all cline models (Wang et al. 2019; Walsh et al. 2020). We determined the mean latitude for all sample sites across both years and then calculated the distance for each site as:

$$d = (\text{long}_i - \text{long}_0) \times (111.320 \times cos(lat))$$

where d is the distance of the site along the transect in kilometers, long_i is the site longitude, long_0 is the longitude of the westernmost site, and lat is the mean latitude (Figure 2) (Carling and Zuckerberg 2011). While the model fit to samples collected in 2004-2007 predicted ancestry to vary little with latitude, *P. cyanea* was predicted to occur farther west in areas north and and south of the transect sampled in 2019-2021. Therefore, to more accurately compare genetic structure between the two series, we fit additional cline models to the 2004-2007 samples after retaining only individuals collected from sites between between 41.5°and 43.5°N, a much narrower latitudinal band that comprises all collection sites in the 2019-2021 dataset. Following Natola et al., we ran 16 different cline models for each specimen series independently (every combination of free, fixed, or no scaling; both, left, right, mirrored, or no exponential tails; and a null model) with a burn-in period of 100,000 and 500,000 Monte–Carlo iterations. For each time series, we selected the model with the lowest AICc score (2023). We evaluated chain convergence of top models by examining parameter trace plots.

Even when applying a stricter latitudinal filter to the 2004-2007 dataset, there are still few shared collection sites between the two time series (Figure 2). To further investigate change over time in a way that is less sensitive to differences in sampling locations, we also predicted the expected hybrid index at each collection site in the 2019-2021 series based on the GAMM fit to data from 2004-2007. These predictions represent the quantification of the null hypothesis that the *Passerina* hybrid zone has remained concordant and coincident between the two sampling times (Carling and Zuckerberg 2011). We then visually compared these predictions to the actual hybrid index values observed in 2019-2021 to determine if the observed hybrid index values were significantly higher than the predicted values at all sites, indicating a westward shift in *P. cyanea* alleles.

Investigating drivers of hybrid zone movement

To evaluate the importance of environmental variables in governing the distributions of *P. cyanea*, *P. amoena*, and hybrids across sampled locations, we modeled the relationships between interpopulation genetic distance and environmental distance and geographic distance as a Bayesian linear mixed model. We assumed

$$\text{logit}(y_{ij}) = \beta_0 + \beta_{\text{geo}} X_{ij}^{\text{geo}} + \beta_{\text{env}} X_{ij}^{\text{env}} + \lambda_i + \lambda_j + \epsilon_{ij}$$

where y_{ij} is the genetic distance (Nei's D) between two populations i and j, X_{ij}^{geo} is the geographic (great circle) distance between the populations, X_{ij}^{env} is the environmental distance (Euclidean distance based on 19 bioclimatic and 16 landcover variables (Supplementary material Tables S1 and S2)) between the populations, the β values are fixed-effect regression coefficients, the λ values are random effects for each population

pair, and ϵ_{ij} is the residual error (Parchman et al. 2016). To estimate environmental distance between each population pair, we first estimated the 19 bioclimatic variables by first using the "daymetr" package in R to download daily estimates of precipitation, minimum temperature, and maximum temperature at each collection location for 2004 to 2007 or 2019 to 2021, depending on the series, from the Daymet climate database (Version 4, R1) (Koen et al. 2018; Thornton et al. 2021; Koen et al. 2018). We then found the average values across the years for each series, and then used the "biovars" function from the "dismo" package to estimate the 19 bioclimatic variables at each checklist location based on the daily Daymet data (Carling and Thomassen 2012; Hijmans et al. 2023). We then quantified landcover at each collection site by using the MODIS package in R to extract MODIS landcover classification values from all 500 m x 500 m tiles within a 2.5 km × 2.5 km cell centered on the provided site latitude and longitude (MODIS MCD12Q1 v006 product (Sulla-Menashe et al. 2019). We then calculated the proportion of each landcover type within the cell. We chose to take into account landcover types surrounding a site, rather than simply record the landcover classification at a site, to more accurately represent the full territories of individual birds. For sample sites in the 2004-2007 dataset, we extracted landcover classification values for 2006. For sample sites in the 2019-2021 dataset, we extracted landcover classification values for 2020. We chose to include all bioclimatic variables and landcover types, rather than only those that previously have been identified to be important for *Passerina* occupancy (Carling and Thomassen 2012; Minor et al. 2022), to account for the possibility that the habitat associations of *P. cyanea* or *P. amoena* in the Great Plains have shifted between collection times. Before calculating Euclidean distance values between collection sites based on bioclimatic and landcover variables, we centered and scaled all variables. We also centered and scaled the final predictor variables (geographic distance and environmental distance) prior to fitting any models (Billerman et al. 2019). We assumed uninformative Gaussian priors for β coefficients ($\mu = 0, \sigma^2 = 1,000$), uninformative hierarchical Gaussian priors for random effects ($\mu = 0, \sigma^2 = \sigma^2_{\text{p op}}$), and uninformative gamma priors for the reciprocals of the random effect and residual variances ($\alpha = 1, \beta = 0.01$) (Gompert et al. 2014).

For both time series, we ran four different variants of this model to disentangle the relative contributions of geographic and environmental distance to genetic structure: (1) a global model in which we included both interpopulation geographic distance and environmental distance as predictors, (2) an alternative global model in which we included an interaction between these two predictors, (3) a model in which we included only geographic distance as a predictor, and (4) a model in which we included only environmental distance as a predictor. We compared the penalized deviance information criterion (DIC) among the models to identify the top performing combination of predictor variables, thereby determining the most important predictor of interpopulation genetic distance in the *Passerina* hybrid zone. We fit all models with Markov chain Monte–Carlo (MCMC) posterior estimation in JAGS using the "RJAGS" package (Plummer 2003) in R. We ran three independent MCMC chains with 10,000 iterations, a burn-in of 2,000 iterations, and a thinning interval of five for each model. We assessed chain convergence for all parameters by examining trace plots (Billerman et al. 2019).

To investigate the extent to which environmental change has driven any shifts in the distributions of *P. cyanea* and *P. amoena*, and therefore the location of the hybrid zone, we modeled the breeding distributions of both species in the Great Plains during the two sampling periods using data from eBird. eBird (http://www.ebird.org), a massive database of bird occurrence records submitted by volunteer amateur naturalists, has become an invaluable tool for estimating the distributions, abundances, population trends, and movements of bird species throughout the world (Walker and Taylor 2017; Sullivan et al. 2009; Minor et al. 2022). At the time of Carling and Zuckerberg's effort to describe changes in the *Passerina* hybrid zone, eBird activity was too infrequent and geographically sparse to accurately describe changes in the distribution and abundance of most bird species in the Great Plains. However, the dataset has grown to be enormous in recent years and now contains records from over 54,607,768 checklists submitted in the United States alone (http://www.ebird.org).

We downloaded all vetted eBird records for *P. cyanea* and *P. amoena* as well as sampling event data from the eBird basic dataset (Version ebd_relJul-2022). We used the auk package in R to filter these datasets to include only records from complete stationary and traveling checklists submitted between June 1 and July 16 (Strimas-Mackey et al. 2018). We excluded records submitted after July 16 because many *P. amoena* initiate fall migration in late July (Young 1991; Rohwer et al. 2005; Pyle et al. 2009). We further subset the datasets to include only checklists with fewer than 10 observers, shorter than 5 hr, and on which observers traveled fewer than 5 km (Johnston et al. 2019). We then merged the sampling event dataset with the *P. cyanea* and *P. amoena* datasets to generate a "zero-filled" dataset for each species, which identifies the checklists on which the species was recorded and sets the species abundance to zero for checklists on which the species was not recorded. We also retained only checklists submitted between 37.34°and 49.77°N and 111.5°and 92.91°W, an area that encompasses all historic *Passerina* collecting efforts in the Great Plains plus a 3°buffer in all directions (Sibley and Short Jr. 1959; Emlen et al. 1975; Carling and Brumfield 2008) (Figure 1). We excluded records from outside this area, as the habitat preferences of both parental species may differ outside of the Great Plains. We generated separate zero-filled datasets for *P. cyanea* and *P. amoena* for the years 2004, 2005, 2006, 2007, 2019, 2020, and 2021, then merged the datasets for the four earlier years and the three later years into two final datasets to estimate the breeding distributions of both species during the two sampling times.

As a semi-structured dataset generated by community scientists, eBird exhibits massive spatial and temporal biases in survey effort. Most checklists are submitted near human population centers. To reduce these biases, we generated a grid of hexagonal cells with 5 km between each cell center across our study area, and then randomly sampled one checklist on which the focal species was detected and one checklist on which the focal species was not detected from each cell for each week from each zero-filled dataset. While this subsampling approach has the potential to artificially inflate downstream estimates of occupancy, it reduces class imbalance in the eBird dataset (for most species, there are many more checklists without detections than with detections) and improves model performance (Johnston et al. 2019).

For all checklist locations in these datasets, we calculated bioclimatic variables and the proportion of the 16 landcover

types in 2.5 km × 2.5 km cells centered on the provided latitude and longitude in the same manner as when we calculated interpopulation environmental distance for the genetic distance models described above. We extracted MODIS land cover and Daymet-derived bioclimatic data for each checklist from the year the checklist was submitted. At the time of this study, MODIS data for 2021 was not yet available, so we used land cover classification from 2020 for all eBird checklists submitted in 2021.

To predict the probability of occurrence of *P. cyanea* and *P. amoena* in the Great Plains between 2004 and 2007 and between 2019 and 2021, we fit random forest models exploring the relationships between species detection and environmental variables using the ranger package in R. We defined species detection (a binary response) as a function of all landcover and bioclimatic variables as well as all effort variables from the eBird sampling dataset (year, day of year, checklist start time, checklist duration, checklist distance, and checklist observer number). Prior to running each model, we split the two zero-filled checklists for each species into training and testing datasets, with the training dataset comprising 80% of the original records and the testing dataset comprising 20% of the original records. To correct for artificial inflation of detection rates from our checklist subsampling approach in the random forest models, we calibrated each model by fitting a generalized additive model with original, uncalibrated model predicted occurrence as the predictor and observed rate of occurrence as the response. We then compared mean square error and sensitivity between calibrated and uncalibrated models based on the models' performance when fit to the testing dataset. We found that in all cases the calibrated models allowed for more accurate estimates of *P. cyanea* and *P. amoena* occurrence (Strimas-Mackey et al. 2023).

We then used the calibrated model output for 2004-2007 and for 2019-2021 to predict the probability of occurrence of each species across our study area based on bioclimatic and landcover data for 2006 and for 2020, respectively. These prediction surfaces hold all effort covariates constant, enabling us to disentangle the role of environmental predictor variables from varying eBird effort in *P. cyanea* and *P. amoena* detection probability (Strimas-Mackey et al. 2023). We then used these predicted occurrence probabilities to create heat maps of each species' breeding distributions during the two collection series and visually compared these distributions to determine if they have changed over time. To investigate the extent to which climate change has driven any observed shifts in the breeding distributions of *P. cyanea* and *P. amoena*, we predicted the calibrated models of occurrence probability for 2019-2021 to the environmental prediction surface for 2006. We chose not to predict the 2004-2007 model to environmental data for 2020 due to this model's higher uncertainty.

Ancestry estimation from phenotype

Since eBird users identify birds on the basis of phenotype alone, discordance between genotype and phenotype in some hybrid systems may lead to rampant misidentification and a significant number of incorrect records around hybrid zones (Slager et al. 2020). To determine whether or not we could safely use eBird data to model the distributions of *P. cyanea* and *P. amoena* in and around the hybrid zone, we modeled the relationship between genomic and phenotypic hybrid index for male *Passerina* collected from 2018-2021 as a linear regression. We estimated phenotypic hybrid index for all birds using the plumage scoring system developed by Kroodsma (1975). This system estimates hybrid index for a specimen based on color in the upper breast, lower abdomen and lower breast, and secondary coverts. Color for upper breast and secondary coverts was scored on a scale from 0-4 (0 for orange or white like an unadmixed *P. amoena*, 4 for blue like an unadmixed *P. cyanea*) and color for lower breast and abdomen was scored on a scale from 0-6 (0 for white like an unadmixed *P. amoena*, 4 for blue like an unadmixed *P. cyanea*). The sum of phenotypic hybrid index scores for all individuals ranged from 0 to 16. PJD scored plumage hybrid index for all specimens to avoid potential inconsistency among observers.

Results

Molecular methods, variant calling, and ancestry estimation

Sequencing yielded averages of 218,489 reads per individual for samples collected between 2004 and 2007 and 3,022,704 reads per individual for samples collected between 2018 and 2021. Of those, an average of 1,019,862 reads per individual were successfully mapped to the *C. cardinalis* reference genome across both series. Mean read depth across sites was 2.01 for 2004-2007 samples and 8.36 for 2019-2021 samples. We removed seven individuals from the 2004-2007 dataset and two individuals from the 2018-2021 dataset due to missing data, leaving 187 individuals for 2004-2007 and 80 individuals for 2018-2021. We also excluded 31 additional individuals that were collected from localities far outside the hybrid zone (e.g., Washington, New York) from the 2004-2007 dataset, leaving 156 individuals from that series for analyses (236 individuals total). The final dataset included 19,536 SNPs for both time series.

Entropy error estimates were low for both collection series. For 2004-2007 samples, mean 95% credible interval widths were 0.0173 for q and 0.0402 for Q. For 2019-2021 samples, mean 95% credible interval widths were 0.0152 for q and 0.0286 for Q. In both time series, most adults were unadmixed. Only 7/156 and 2/80 collected individuals were classified as F1s for the 2004-2007 series and 2019-2021 series, respectively. In both datasets, most admixed individuals were classified as backcrosses to *P. amoena*, or at the very least had q estimates closer to that of unadmixed *P. amoena*. There were comparatively few backcrosses to *P. cyanea*. In the 2004-2007 series, twelve individuals were classified as backcrosses to *P. amoena* and two individuals were classified as backcrosses to *P. cyanea*. In the 2019-2021 series, nine individuals were classified as backcrosses to *P. amoena* and four individuals were classified as backcrosses to *P. cyanea* (Figure 3).

Estimating change in the hybrid zone over time

Estimates of cline parameters for the 2004-2007 and 2019-2021 series indicate rapid changes in the genetic structure of the *Passerina* hybrid zone between the two sampling times (Figure 4). For the full 2004-2007 series, the filtered 2004-2007 series, and the 2019-2021 series, the cline model with fixed scaling and no exponential tails had the lowest AICc value. The model fit to samples from the full 2004-2007 series was centered at 593.55 km along the transect (95% CI: 550.83 to 647.47 km) and had an estimated width of 292.93 km (95% CI: 210.90 to 426.81 km). The model fit to samples

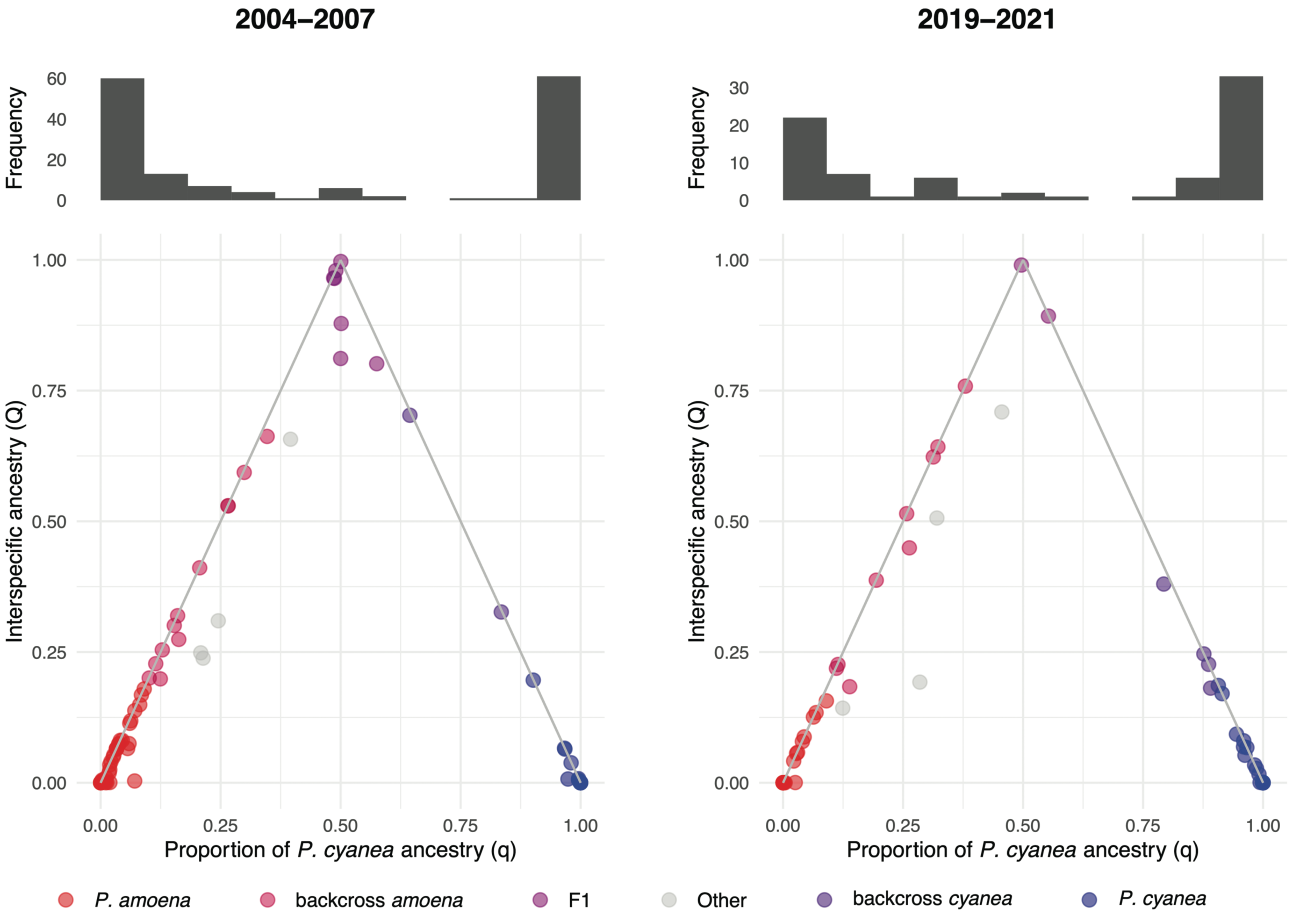

Figure 3. Results from "entropy" for a model with $k = 2$ for individuals collected between 2004 and 2007 and individuals collected between 2019 and 2021. The *x*-axis represents the proportion of *P. cyanea* ancestry (q) for each individual, and the *y*-axis show the proportion of sampled loci that are heterozygous (exhibiting the reference alleles for both *P. cyanea* and *P. amoena* for that individual (Q). Histograms above each plot show the distribution of q for each series.

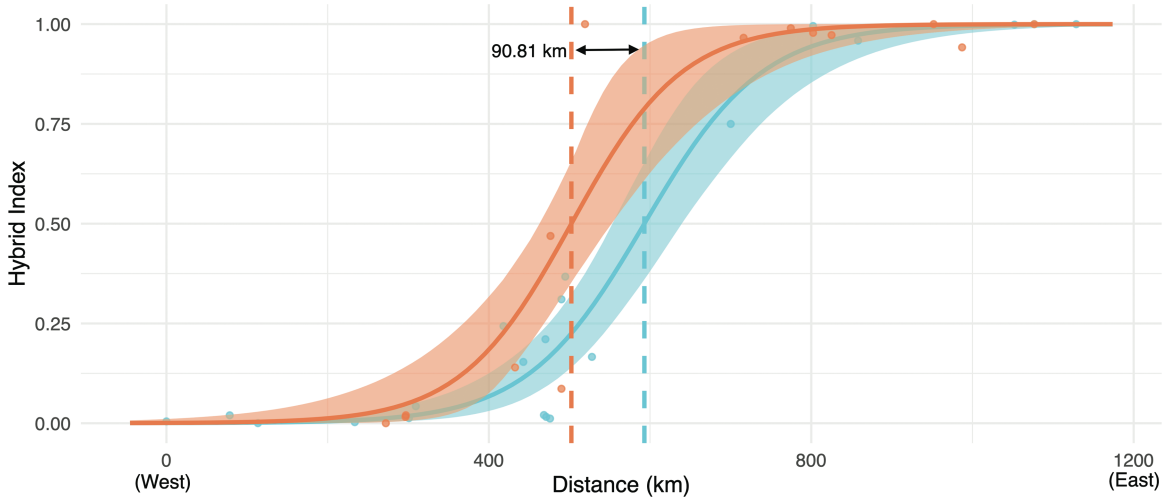

Figure 4. Top fitting cline models for all *Passerina* specimens collected between 2004 and 2007 (turquoise) and between 2019 and 2021 (orange). Solid lines represent the maximum likelihood clines for each series, and lighter shaded areas represent the 95% credible intervals around each maximum likelihood estimate. Vertical dashed lines represent the estimated center for each cline. We did not include the top cline model for the filtered 2004-2007 dataset in this plot as it appeared extremely similar to the cline for the unfiltered 2004-2007 dataset.

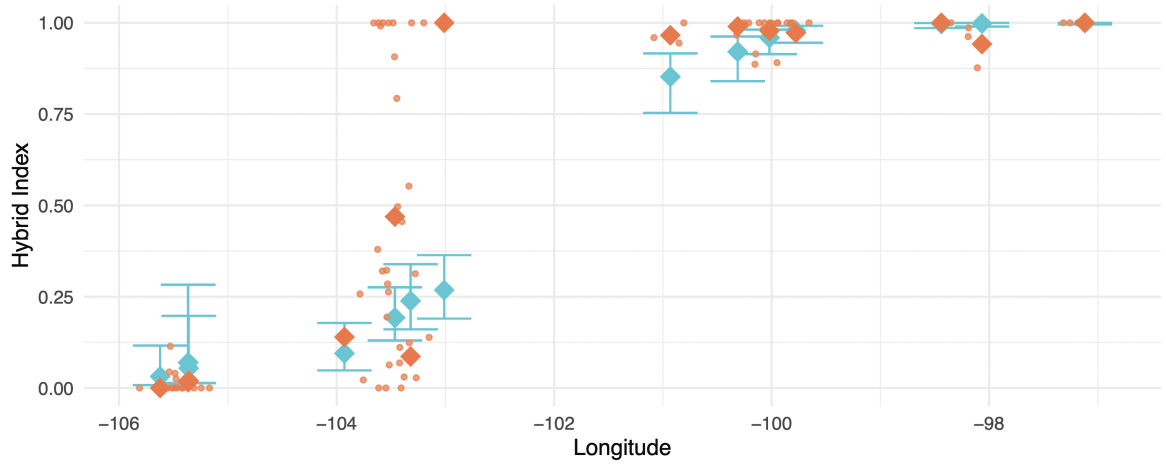

Figure 5. Predicted hybrid index values at 2019-2021 collection sites based on generalized additive mixed model fit to data from 2004-2007 (turquoise diamonds) compared to observed hybrid index values for birds collected at those sites in 2019-2021 averaged across sites (orange diamonds) and for individual birds (orange points).

from the 2019-2021 series was centered at 502.74 km along the transect (95% CI: 458.73 to 564.21 km) and had an estimated width of 275.19 km (95% CI: 136.32 to 505.86 km). The model fit to the 2004-2007 series filtered to include only sites at similar latitudes as the 2019-2021 dataset predicted a similar westward shift in the hybrid zone; this cline was centered at 586.26 km along the transect (95% CI: 533.87 to 653.47 km) with a predicted width of 246.17 km (95% CI: 156.98 to 455.26). Although the credible intervals for hybrid zone center in the two time series overlap slightly, these models suggest an approximate 90 km westward shift in the *Passerina* hybrid zone between sampling times, or approximately 4.12 to 5.83 km per year. This rate is much faster than that noted by Carling and Zuckerberg, who estimated that the zone moved only 2.28 to 2.57 km per year in the second half of the twentieth century (2011). Comparing the estimated cline widths between the filtered 2004-2007 series and the 2019-2021 series suggests that the *Passerina* hybrid zone has also become wider over time, and that sometime in the recent two decades the trend of the hybrid zone becoming more narrow as noted by Carling and Zuckerberg has reversed (2011). However, the credible intervals for the two width estimates overlap broadly. We therefore can not confidently describe changes in hybrid zone width between the two sampling times.

The generalized additive mixed model fit to data from 2004-2007 predicted hybrid index values to be consistently lower than observed values at many sites sampled in 2019-2021, supporting a westward shift in the *Passerina* hybrid zone over time (Figure 5). This comparison further supports a westward shift in the *Passerina* hybrid zone while accounting for differences in collection sites between the two time series. Contrary to this overall trend, birds collected at one site (Ponderosa Wildlife Management Area in Nebraska) in 2019-2021 had lower hybrid mean index values than predicted by the model. A higher than expected frequency of *P. amoena* alleles potentially resulted from the higher elevation ponderosa stands and narrow deciduous riparian corridors at this site closely resembling preferred breeding territories of *P. amoena* in western North America (Greene et al. 2000), suggesting matching habitat choice as a important mediator of gene flow in this system (Edelaar et al. 2008).

Table 3. Mean, penalized, and ΔDIC for Bayesian linear mixed models with different predictors of genetic distance.

	Mean DIC	Penalized DIC	ΔDIC
2004-2007			
Geo × Env	143.8	158.3	0
Geo-only	147	160	1.7
Geo + Env	146.4	160.8	2.5
Env-only	271.6	281.9	123.6
2019-2021			
Geo-only	109.8	121.4	0
Env-only	119.2	130.3	8.9

Investigating predictors of hybrid zone movement

We fit variants of a Bayesian linear mixed model to investigate potential predictors of interpopulation genetic distance in the *Passerina* hybrid zone. For the 2004-2007 dataset, the global model that included an interaction between geographic distance and environmental distance proved to be best supported, closely followed by the model that included only geographic distance as a predictor and the global model that did not allow for an interaction between the two predictors, then distantly followed by the model that included only environmental distance as a predictor. For the 2019-2021 dataset, the correlation between geographic distance and environmental distance was much stronger (0.937, compared to 0.358 for the 2004-2007 series), likely due to the narrow latitudinal spread of sampling sites. We therefore did not fit a global model to this dataset in which we include both geographic and environmental distance as predictors. Between the models that included only geographic distance or environmental distance as predictors, the geographic distance model again was the best supported (Table 3). In the top model for both datasets, geographic distance was predicted to have a weak but significant positive relationship with genetic distance (2004-2007: β = 0.00159, 95% CI = 0.00135 - 0.00184; 2019-2021: β = 0.00163, 95% CI = 0.00112 - 0.00215). The fact that our model selection approach identified geographic distance as the most important predictor of interpopulation genetic distance in both time series suggests that population demographics best

Table 4. β estimate with 95% credible interval for each predictor included in Bayesian linear mixed model with different predictors of genetic distance.

	β Geo	β Geo 95% CI	β Env	β Env 95% CI	β Geo × Env	β Geo × Env 95% CI
2004-2007						
Geo × Env	0.00159	0.00135 - 0.00184	-0.000871	-0.00174 - 0.000000880	0.00000345	0.000000120 - 0.00000679
Geo-only	0.00153	0.00130 - 0.00176	–	–	–	–
Geo + Env	0.00156	0.00133 - 0.00181	-0.000354	-0.00107 - 0.000347	–	–
Env-only	–	–	0.00106	0.000185 - 0.00193	–	–
2019-2021						
Geo-only	0.00163	0.00112 - 0.00215	–	–	–	–
Env-only	–	–	0.00651	0.00397 - 0.00903	–	–

explain the genetic structure of the *Passerina* hybrid zone. However, for both datasets, the environment-only models identified significant positive relationships between environmental and genetic distances (2004-2007: β = 0.00106, CI = 95% 0.000185 - 0.00193; 2019-2021: β = 0.00651, CI = 0.00397 - 0.00903) (Table 4). Overall, these models demonstrate that, in addition to geographic distance, environmental variation plays an important role in governing the distributions of *P. cyanea*, *P. amoena*, and hybrids in the Great Plains.

Our random forest models exploring relationships between environmental variables and *P. cyanea* and *P. amoena* probability of occurrence based on eBird data revealed consistent differences in breeding habitat preferences between the two species. After geographic filtering and spatiotemporal subsampling, the final training datasets included 127 detections and 1490 non-detections for *P. cyanea* from 2004 to 2007, 105 detections and 1491 non-detections for *P. amoena* from 2004 to 2007, 4,691 detections and 43,380 non-detections for *P. cyanea* from 2019 to 2021, and 2,458 detections and 44,699 non-detections for *P. amoena* from 2019 to 2021. The dramatically larger sample sizes for the 2019-2021 datasets reflects the rapid growth of eBird participation in the United States over time.

Random forest models identified precipitation variables as among the most important predictors for both species in both sampling times (Supporting Information Figs. S2 and S3, Supporting Information Tables S1 and S2). The relatively low importance of most landcover variables could be because the MODIS product classifies landcover at 500 meter resolution, and therefore may not accurately represent the narrow riparian corridors and forest edge habitat preferred by *P. cyanea* and *P. amoena*. Future studies exploring *Passerina* habitat preferences should incorporate finer scale landcover classification datasets. That being said, both models for *P. amoena* occurrence identified the proportion of landcover classified as grassland within 2.5 km of a checklist center as a top predictor. This relationship likely results not because *P. amoena* occurs in grasslands, but because the riparian corridors preferred by *P. amoena* in the western Great Plains are often surrounded by grasslands. For both species, the order of importance for different environmental predictors differs slightly between the models fit to the 2004-2007 dataset and the 2019-2021 dataset. We attribute this discrepancy to the smaller sample size of the 2004-2007 datasets, and assume that models fit to eBird data from 2019-2021 represent the habitat preferences of *P. cyanea* and *P. amoena* much more accurately.

The predicted distributions for both parental species matched our expectations. The models for both time periods predict *P. cyanea* to occur in the eastern Great Plains and *P. amoena* to occur in the western Great Plains. The strong differences in habitat associations between these species likely enabled us to reproduce the breeding distributions for both species even without including latitude, longitude, or elevation as predictors. When comparing the predicted distributions for both species between the two time series, *P. cyanea* shows a noticeable westward expansion in northern Nebraska and southern South Dakota, seemingly along the Missouri and Niobrara Rivers. In contrast, the breeding distribution for *P. amoena* appears relatively stable over time. When predicted to bioclimatic and landcover data from 2006, the model fit to *P. cyanea* eBird data from 2019-2021 produced occurrence probabilities visually similar to the model fit to eBird data from 2004-2007 (Figures 6 and 7).

Ancestry estimation from phenotype

We found plumage hybrid index score to be a strong predictor of genomic hybrid index for *Passerina* collected between 2018 and 2021. There was an exceptionally strong positive relationship between these two metrics ($\beta = 1.01308, R^2 = 0.9580172, p =< 2e-16$). While there were a small number of cryptic hybrids, the vast majority of study birds had breeding plumage that accurately reflected their ancestry (Figure 8). Likely because most admixed individuals were backcrosses to *P. amoena*, there was a small number of *P. amoena* backcrosses incorrectly identified as unadmixed *P. amoena* by plumage. There were extremely few birds that we classified as unadmixed *P. cyanea* based on plumage that proved to be admixed. However, there were some individuals that we estimated to be admixed based on plumage but who proved to be unadmixed *P. cyanea*. These turned out to be mostly second-year males who still had not achieved full adult plumage, showing a considerable amount of white in their bellies and undertail coverts and sometimes noticeable white or buff edging on their greater and median secondary coverts. These individuals demonstrate that second-year male buntings can not be reliably identified as hybrids unless they also have orange mixed with blue in the breast. Overall, breast coloration is likely the most reliable indicator of ancestry and hybrid status in *Passerina*.

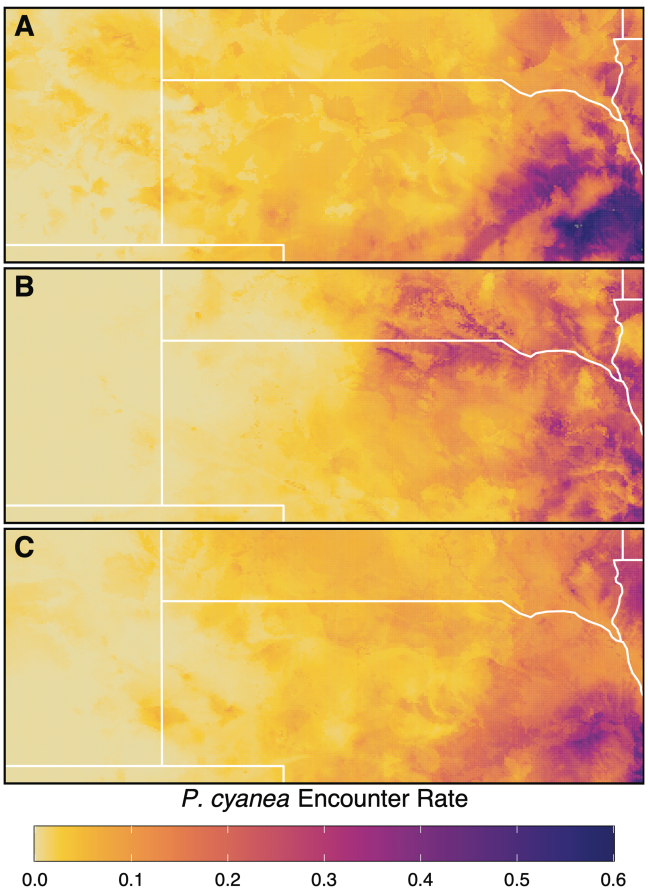

Figure 6. Predicted probability of occurrence of *P. cyanea* during the breeding season based on random forest models fit to eBird data and environmental data for (A) 2004-2007 predicted to environmental conditions for 2006, (B) 2019-2021 predicted to environmental conditions for 2020, and (C) 2019-2021 predicted to environmental conditions for 2006.

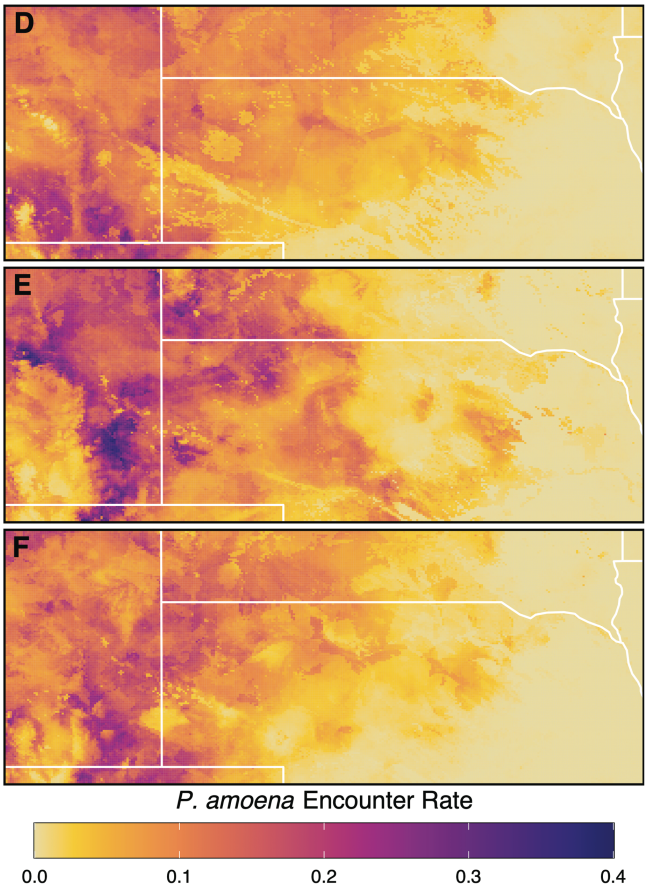

Figure 7. Predicted probability of occurrence of *P. amoena* during the breeding season based on random forest models fit to eBird data and environmental data for (A) 2004-2007 predicted to environmental conditions for 2006, (B) 2019-2021 predicted to environmental conditions for 2020, and (C) 2019-2021 predicted to environmental conditions for 2006.

Discussion

In this study, we generated genotyping-by-sequencing data and fit cline models for individuals sampled from the *Passerina* bunting hybrid zone in the Great Plains of North America between 2004 and 2007 and between 2019 and 2021. We aimed to test if the trends noted by Carling and Zuckerberg, who compared specimens collected from 2004-2007 to historic specimens collected from 1955-1957 and in 1969 and found that the hybrid zone had shifted west and become more narrow over time, have continued in the past two decades (2011). Our cline models demonstrated that the zone has indeed continued to move west, and that this westward movement has likely accelerated in recent decades. Although the credible intervals for our estimates of cline center in the two sime series overlap slightly, we present multiple lines of evidence supporting a substantial westward shift.

We identified variation in bioclimatic variables and landcover to be a significant predictor of genetic divergence across the hybrid zone during both sampling times. We interpret this relationship to mean that the breeding distributions of *P. cyanea*, *P. amoena*, and hybrids in the Great Plains is governed by environmental variation, and any environmental changes would alter these distributions. To test this hypothesis, we fit random forest models to eBird data to estimate the breeding distributions of *P. cyanea* and *P. amoena* between 2004-2007 and between 2019-2021 based on the same suite of environmental predictors. As expected, the models for both time periods predict *P. cyanea* to occur in the wetter, warmer, and more forested habitats in the eastern Great Plains and *P. amoena* to occur in the dryer, more open habitats in the western Great Plains. While the eBird models show no obvious changes in the breeding distribution of *P. amoena*, the model fit to data from 2019-2021 predicts *P. cyanea* to breed much farther west along the Niobrara and Missouri Rivers than the model fit to data from 2004-2007. This change in the breeding distribution of at least one of the two parental species mirrors the observed westward shift in hybrid zone center predicted by our cline models, and supports the conclusion that the hybrid zone has shifted west in recent decades. When we predicted the breeding distribution of *P. cyanea* based on the model fit to data from 2019-2021 to environmental data for 2004-2007, the predicted distribution appeared visually extremely similar to the distribution predicted by the model fit to data from 2004-2007. This similarity suggests that *P. cyanea* has maintained a similar climatic niche between the two sampling times, and the westward shift in it's breeding distribution has been driven, at least in part, by environmental change. Given that the random forest models for both time periods identified precipitation variables as the most important predictors

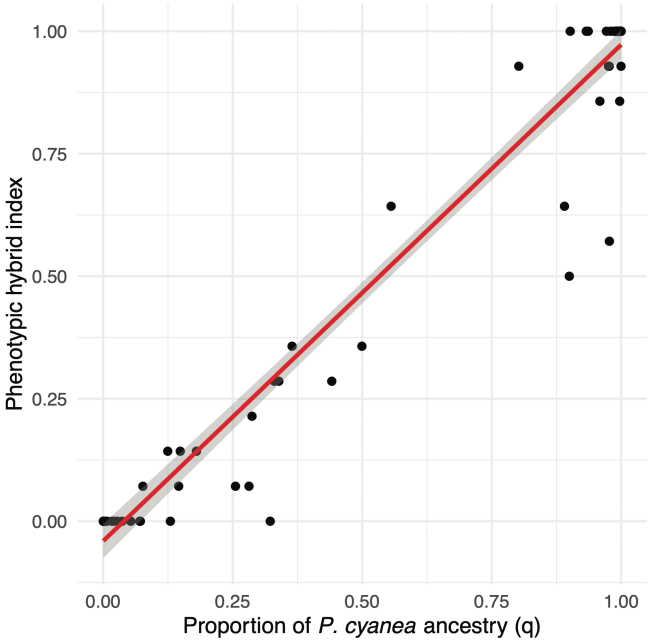

Figure 8. Phenotypic hybrid index plotted as a function of genomic hybrid index for 80 after hatch year male *Passerina* collected between 2018 and 2021. Phenotypic hybrid index scores derived from visual assessment of upper breast, lower abdomen and lower breast, and secondary covert coloration.

for *P. cyanea* occurrence, the westward shift in the species distribution has likely tracked changes in precipitation. However, the predicted distribution of *P. cyanea* in 2019-2021 suggests westward expansion along rivers in the northern Great Plains. This close association with riparian areas suggests that, while changes in precipitation have been the most important factor in enabling the westward expansion of *P. cyanea*, the presence of sufficient habitat also has been necessary.

Environmental changes driving westward movement of the *P. cyanea* and *P. amoena* hybrid zone is consistent with recent studies of some of the other Great Plains hybrid zones. Walsh et al. and Aguillon and Rohwer described similar westward shifts in the hybrid zone between *Icterus galbula* and *I. bullocki* and the hybrid zone between *Colaptes auratus auratus* and *C. a. cafer*, respectively (2020; 2022). The *Colaptes*, *Icterus*, and *Passerina* hybrid systems may vary in behavior, mating signals, or demographic trends, but all comprise a parental taxon that breeds primarily in mesic deciduous forest habitat in eastern North America and another parental taxon that occurs in the more xeric conditions in the west (Carling and Zuckerberg 2011). A consistent trend across at least these three systems suggests environmental change as a common driver of hybrid zone movement.

While our estimates of cline widths had high uncertainty that prevented us from confidently describing changes over time, they suggest that the *Passerina* hybrid zone has become wider between the two sampling times. If this trend is accurate, it contradicts Carling and Zuckerberg's observation that the hybrid zone became more narrow in the second half of the twentieth century (2011). A widening of the *Passerina* hybrid zone would also be surprising given previous findings that the *Icterus* and *Pheucticus* hybrid zones in the Great Plains have similarly become more narrow over time (Mettler and Spellman 2009; Walsh et al. 2020). It is possible that rapid westward expansion and increased breeding densities of *P. cyanea* (see below) could cause the *Passerina* hybrid zone to become wider in recent decades (Hubbs 1955; Ryan et al. 2018; McEntee et al. 2020). Alternatively, recent changes in hybrid zone width could reflect changes in relative hybrid fitness. All studies documenting contractions in the avian hybrid zones of the Great Plains have speculated that hybrid fitness is low relative to unadmixed individuals, spurring the evolution of stronger prezygotic reproductive isolation throughout the twentieth century (Hudson and Price 2014). Our finer temporal sampling compared to these earlier studies may have allowed us to detect a potential reversal in these trends. It is possible that recent environmental changes have elevated the relative fitness of previously disadvantageous hybrid phenotypes, at least in *Passerina*. For example, *P. cyanea* and *P. amoena* differ markedly in their molt and migratory behavior, and hybrids may exhibit novel combinations of molt and migration phenology (Rohwer and Irwin 2011). Recent changes in the spatiotemporal distribution of resources, which are needed to supply molt, could now favor hybrid phenotypes. With the fitness of at least some hybrids now equaling or exceeding that of unadmixed parentals, prezygotic reproductive isolation may become weaker, causing the hybrid zone to expand (Coyne and Orr 2004). Importantly, this explanation implies that hybridization between *P. cyanea* and *P. amoena* could generate the adaptive variation needed for populations to survive rapid changes in the fitness landscape when mutation alone is too slow (Anderson and Stebbins Jr. 1954). Ultimately, however, continued sampling of the *Passerina* hybrid zone is necessary to assess the true current trend and mutability in cline width. More study is also needed to assess how adaptive divergence in molt and migration strategies influences hybrid fitness and mediates introgression between *P. cyanea* and *P. amoena* (Rohwer and Irwin 2011).

While we present evidence that climate change has driven shifts in the *Passerina* hybrid zone, there are likely other factors at play. The fact that the *Passerina* hybrid zone appears to be moving west faster than the other Great Plains avian hybrid zones in recent decades suggests that there are unique features of the system influencing zone movement (Walsh et al. 2020). For instance, declining *P. cyanea* habitat throughout eastern North America as abandoned farmland now progresses towards mature forest may be driving individuals from these areas westward into the Great Plains (Carey and Nolan 1979). In this scenario, westward expansion of *P. cyanea* is driven not only by climate change creating favorable habitat in the west, but also by large-scale demographic movement out of the east. Under the tension zone model, which assumes that hybrid zones occur where both parental taxa breed at low densities, increasing numbers of *P. cyanea* alone would push the abundance trough farther west (Key 1968; Buggs 2007; Carling and Zuckerberg 2011). To test this hypothesis, future studies could generate population genetic data from throughout the distribution of *P. cyanea* in eastern North America and determine the origin of Great Plains populations. Additionally, systematic surveys of *P. cyanea* and *P. amoena* in the Great Plains could reveal if *P. cyanea* truly is more abundant in and near areas of contact. It is also possible that hybrid zone movement may be driven by environmental change not only in the breeding distribution, but also in areas where *P. cyanea* and *P. amoena* occur during other parts of the year. As the two species differ in their overwintering and molting distributions (Rohwer and Irwin 2011), climate or landcover changes

in these areas could influence their breeding abundances and distributions differently. Future studies should describe migratory connectivity of *Passerina* populations that breed in the Great Plains and incorporate environmental variation in molting and overwintering areas in breeding distribution models (Kramer et al. 2018; Marra et al. 2015; Stevens et al. 2023).

Other potential drivers of hybrid zone movement include higher male aggression in *P. cyanea* and female preference for the traits of male *P. cyanea*. Baker extensively studied the behavior of sympatric *P. cyanea* and *P. amoena* and found that in both species, males respond aggressively to heterospecific songs and females prefer conspecific plumage and songs (Baker 1991; Baker and Baker 1990; Baker 1996). However, as a female's mating preferences are strongly influenced by the songs she hears during development, exposure to more singing male *P. cyanea* could lead to stronger preference for *P. cyanea* songs (Baker and Boylan 1999). Extensive extrapair matings between female *P. amoena* and male *P. cyanea* could increase introgression of *P. cyanea* phenotypes (Baldassarre and Webster 2013). Asymmetric abundances of parental species further accelerating hybrid zone movement by biasing female choice in favor of the more abundant species has never been documented in birds, but is consistent with our cline models. Future studies of female preference across the *Passerina* hybrid zone would be necessary to determine if this phenomenon is occurring.

While eBird data enabled us to identify climate change as a driver of hybrid zone movement, our study illustrates the limitations of the dataset for addressing questions related to hybrid systems. Ultimately, eBird users, although mostly highly skilled in bird identification and overseen by expert reviewers, identify individuals based on phenotype alone (Johnston et al. 2019). In and near hybrid zones, many individuals may exhibit phenotypes that do not accurately represent their ancestry (e.g., cryptic hybrids) (Minor et al. 2022). To evaluate whether or not we could safely use eBird data, we explored the relationship between phenotype and ancestry in *Passerina*. We found an extremely strong positive relationship between phenotype and ancestry in the *P. cyanea* x *P. amoena* hybrid zone and discovered few backcrosses to *P. cyanea*. We therefore inferred that almost all records of *P. cyanea* in the eBird dataset are correct. In contrast, we identified many backcrosses to *P. amoena* in our samples. Many of these individuals appeared similar to unadmixed *P. amoena*. If *P. amoena* backcrosses are as common in the hybrid zone as our samples suggest, many *P. amoena* records in the eBird database may actually be misidentified hybrids. These incorrect records could be why we observed relative stability in the breeding distribution of *P. amoena* over time, which conflicts with our findings that *P. cyeana* has expanded its range west and that the hybrid zone has shifted west without significantly expanding in width. It is possible that high numbers of *P. amoena* backcrosses misidentified as unadmixed *P. amoena* led us to estimate the breeding distribution of *P. amoena* as extending farther east than it does in reality. Overall, when applying eBird data to study hybrid systems, researchers should incorporate other datasets and make careful consideration to avoid inappropriate use (Minor et al. 2022).

In the future, we hope for not only increased eBird activity in the Great Plains, but also more genetic sampling of the *Passerina* hybrid zone. Given that the speed at which the hybrid zone is moving west seems to have increased in the past fifteen years, we predict it will continue to shift west in the following decades. However, our random forest models fit to eBird data identified annual precipitation as an important predictor of *P. cyanea* occurrence in the Great Plains, suggesting that intensifying droughts in western North America may eventually halt the species' westward expansion (Zhao et al. 2020). Describing how this dynamic hybrid zone continues to change under shifting environmental conditions will continue to yield insights into the processes that mediate speciation and ecological adaptation.

Conclusion

During the summers of 2019 and 2021, we sampled the *Passerina* hybrid zone to investigate how it has changed since the previous sampling effort between 2004 and 2007. Cline models demonstrated a significant westward shift in the center of the hybrid zone between the two sampling events. By incorporating eBird models, we present evidence that this westward shift has been driven, at least in part, by changes in precipitation. This work demonstrates the critical importance of scientific collection, particularly replicate sampling along transects.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

All eBird, Daymet, and MODIS data are available to download at https://ebird.org/data/download, https://daymet.ornl.gov/getdata, and https://modis.gsfc.nasa.gov/data/dataprod/mod12.php, respectively. All sequence reads used in the study are available at the National Institute of Health's Sequence Read Archive (PRJNA1137379) and all scripts are available and free to download at https://github.com/paul-dougherty/Passerina_hybrid_zone_changes.

Author contributions

P.J.D. and M.D.C. collected samples in the field and designed the study. P.J.D. performed analyses. P.J.D. and M.D.C. contributed to writing the manuscript.

Funding

This research was made possible thanks to generous support from the Society for the Study of Evolution, the American Ornithological Society, the Society of Systematic Biologists, the University of Wyoming Department of Zoology, the Laramie Audubon Society, and a grant from the National Science Foundation (1928870) to M.D.C. Equipment used for sequencing was funded by the National Institute of General Medical Sciences of the National Institutes of Health under Grants #P20GM103432 and #P20GM121310. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: The authors declare no conflict of interest.

Acknowledgments

We would like to thank the Nature Conservancy, the Nebraska Game and Parks Commission, Wyoming Game and Fish, and the National Forest Service for providing access to collection sites and supporting us during our field work. Doug Eddy, Sydnie Fossberg, Nick Minor, Ashlynn Ballard, Ashley Townsend, R. Franke, and Terri Higgins all provided essential assistance in the field. We would also like to thank the thousands of birders who report their sightings to eBird for generating the dataset we used in our models. Finally, this work would have been impossible without the following individuals: Elizabeth Wommack, who assisted with preparing collected birds as museum specimens, Josh Jahner and Will Rosenthal, who provided invaluable guidance on the genomic analyses used in this paper, Shawn Billerman, who provided code for the models investigating drivers of genetic distance, and members of the Senner lab at the University of Massachusetts Amherst, who provided valuable feedback on the manuscript.

References

Aguillon, S. M. & V. G. Rohwer. 2022. Revisiting a classic hybrid zone: Movement of the northern flicker hybrid zone in contemporary times. Evolution 76:1082–1090.

Allendorf, F. W., R. F. Leary, P. Spruell, & J. K. Wenburg. 2001. The problems with hybrids: setting conservation guidelines. Trends in Ecology and Evolution 16:613–622.

Anderson, B. W. 1971. Man's influence on hybridization in two avian species in South Dakota. The Condor 73:342–347.

Anderson, E. & G. Stebbins Jr. 1954. Hybridization as an Evolutionary Stimulus. Evolution 8:378–388.

Baker, M. C. 1991. Response of male Indigo and Lazuli Buntings and their hybrids to song playback in llopatric and Sympatric Populations. Behavior 119:225–242.

Baker, M. C. 1996. Female buntings from hybridizing populations prefer conspecific males. The Wilson Bulletin 108:771–775.

Baker, M. C. & A. E. M. Baker. 1990. Reproductive behavior of female buntings: isolating mechanisms in a hybridizing pair of species. Evolution 44:332–338.

Baker, M. C. & J. T. Boylan. 1999. Singing Behavior, Mating Associations and Reproductive Success in a Population of Hybridizing Lazuli and Indigo Buntings. The Condor 101:493–504.

Baldassarre, D. T. & M. S. Webster. 2013. Experimental evidence that extra-pair mating drives asymmetrical introgression of a sexual trait. Proceedings of the Royal Society B: Biological Sciences 280:20132175.

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

Bierne, N., P.-A. Gagnaire, & P. David. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Current Zoology 59:72–86.

Billerman, S. M., C. Cicero, R. C. K. Bowie, & M. D. Carling. 2019. Phenotypic and genetic introgression across a moving woodpecker hybrid zone. Molecular Ecology 28:1692–1708.

Britch, S. C., M. L. Cain, & D. J. Howard. 2001. Spatio-temporal dynamics of the Allonemobius fasciatus – A. socius mosaic hybrid zone: a 14-year perspective. Molecular Ecology 10:627–638.

Buggs, R. J. 2007. Empirical study of hybrid zone movement. Heredity 2007 99:3 99:301–312.

Carey, M. & V. Nolan. 1979. Population Dynamics of Indigo Buntings and the Evolution of Avian Polygyny. Evolution 33:1192.

Carling, M. D. & R. T. Brumfield. 2008. Haldane's rule in an avian system: Using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. Evolution 62:2600–2615.

Carling, M. D., I. J. Lovette, & R. T. Brumfield. 2010. Historical divergence and gene flow: coalescent analyses of mitochondrial, autosomal and sex-linked loci in Passerina buntings. Evolution 64:1762–1772.

Carling, M. D. & H. A. Thomassen. 2012. The Role of Environmental Heterogeneity in Maintaining Reproductive Isolation between Hybridizing Passerina (Aves: Cardinalidae) Buntings. International Journal of Ecology 2012:295463.

Carling, M. D. & B. Zuckerberg. 2011. Spatio-temporal changes in the genetic structure of the Passerina bunting hybrid zone. Molecular Ecology 20:1166–1175.

Coyne, J. A. & H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:1–4.

Dasmahapatra, K. K., M. J. Blum, A. Aiello, et al. 2002. Inferences from a rapidly moving hybrid zone. Evolution 56:741–753.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, & R. T. Brumfield. 2014. hzar: Hybrid zone analysis using an R software package. Molecular Ecology Resources 14:652–663.

Edelaar, P., A. M. Siepielski, & J. Clobert. 2008. Matching habitat choice causes directed gene flow: A neglected dimension in evolution and ecology. Evolution 62:2462–2472.

Emlen, S. T., J. D. Rising, & W. L. Thompson. 1975. A behavioral and morphological study of symaptry in the Indigo and Lazuli Buntings of the Great Plains. The Quarterly Magazine of Ornithology 87:145–302.

Galindo, J., M. J. Rivas, M. Saura, & E. Rolán-Alvarez. 2014. Selection on hybrids of ecologically divergent ecotypes of a marine snail: the relative importance of exogenous and endogenous barriers. Biological Journal of the Linnean Society 111: 391–400.

Gompert, Z., L. K. Lucas, C. A. Buerkle, et al. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Molecular Ecology 23:4555–4573.

Grabenstein, K. C. & S. A. Taylor. 2018. Breaking barriers : causes, consequences , and experimental utility of human-mediated hybridization. Trends in Ecology & Evolution 33:198–212.

Greene, E., B. E. Lyon, V. R. Muehter, et al. 2000. Disruptive sexual selection for plumage coloration in a passerine bird. Nature 407:1000–1003.

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press.

Hijmans, R. J., S. Phillips, J. Leathwick, and E. Jane. 2023. dismo: Species Distribution Modeling.

Hubbs, C. L. 1955. Hybridization between fish species in nature. Systematic Biology 4:1–20.

Hudson, E. J. and T. D. Price. 2014. Pervasive reinforcement and the role of sexual selection in biological speciation. Journal of Heredity 105:821–833.

Irwin, D. E. 2020. Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. The American Naturalist 195:E150–E167.

Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey et al. 2019. Best practices for making reliable inferences from citizen science data: Case study using eBird to estimate species distributions.

Key, K. H. 1968. The Concept of Stasipatric Speciation. Systematic Biology 17:14–22.

Koen, H., D. Basler, T. Milliman, E. K. Melaas, and A. D. Richardson. 2018. An integrated phenology modelling framework in R: Modelling vegetation phenology with phenor. Methods in Ecology & Evolution 9:1–10.

Kramer, G. R., D. E. Andersen, D. A. Buehler, et al. 2018. Population trends in Vermivora warblers are linked to strong migratory connectivity. Proceedings of the National Academy of Sciences 115:201718985.

Kroodsma, R. L. 1975. Hybridization in buntings (Passerina) in North Dakota and Eastern Montana. The Auk 92:66–80.

Krosby, M. & S. Rohwer. 2010. Ongoing movement of the Hermit Warbler X Townsend's Warbler hybrid zone. PLoS ONE 5:e14164.

Larson, E. L., R. M. Tinghitella, & S. A. Taylor. 2019. Insect hybridization and climate change. Frontiers in Ecology and Evolution 7:348.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [PhD thesis].

Lindtke, D., Z. Gompert, C. Lexer, & C. A. Buerkle. 2014. Unexpected ancestry of Populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. Molecular Ecology 23:4316–4330.

Mandeville, E. G., A. W. Walters, B. J. Nordberg, et al. 2019. Variable hybridization outcomes in trout are predicted by historical fish stocking and environmental context. Molecular Ecology 28:3738–3755.

Marra, P. P., E. B. Cohen, S. R. Loss, J. E. Rutter, & C. M. Tonra. 2015. A call for full annual cycle research in animal ecology. Biology Letters 11:20150552.

McDonald, D. B., R. P. Clay, R. T. Brumfield, & M. J. Braun. 2001. Sexual selection on plumage and behavior in an avian hybrid zone: Experimental tests of male-male interactions. Evolution 55:1443–1451.

McEntee, J. P., J. G. Burleigh, & S. Singhal. 2020. Dispersal predicts hybrid zone widths across animal diversity: Implications for species borders under incomplete reproductive isolation. American Naturalist 196:9–28.

McFarlane, S. E. & J. M. Pemberton. 2019. Detecting the true extent of introgression during anthropogenic hybridization. Trends in Ecology & Evolution 34:315–326.

Mettler, R. D. & G. M. Spellman. 2009. A hybrid zone revisited: Molecular and morphological analysis of the maintenance, movement, and evolution of a Great Plains avian (Cardinalidae: Pheucticus) hybrid zone. Molecular Ecology 18:3256–3267.

Minor, N. R., P. J. Dougherty, S. A. Taylor, & M. D. Carling. 2022. Estimating hybridization in the wild using citizen science data: A path forward. Evolution 76:362–372.

Natola, L., S. M. Billerman, M. D. Carling, S. S. Seneviratne, & D. Irwin. 2023. Geographic variability of hybridization between red-breasted and red-naped sapsuckers. Evolution 77:580–592.

Ortiz-Barrientos, D., P. Grealy, & P. Nosil. 2009. The genetics and ecology of reinforcement: Implications for the evolution of prezygotic isolation in sympatry and beyond. Annals of the New York Academy of Sciences 1168:156–182.

Parchman, T. L., C. A. Buerkle, V. Soria-Carrasco, & C. W. Benkman. 2016. Genome divergence and diversification within a geographic mosaic of coevolution. Molecular Ecology 25:5705–5718.

Parchman, T. L., Z. Gompert, J. Mudge, et al. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology 21:2991–3005.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling .

Pyle, P., W. A. Leitner, L. Lozano-Angulo, et al. 2009. Temporal, Spatial, and Annual Variation in the Occurrence of Molt-Migrant Passerines in the Mexican Monsoon Region. The Condor 111:583–590.

Remington, C. L. 1968. Suture-Zones of hybrid interaction between recently joined biotas. *In* T. Dobzhansky, M. K. Hecht, & W. C. Steere (Eds.), Evolutionary biology (pp. 321–428). Springer.

Rising, J. D. 1969. A comparison of metabolism and evaporative water loss of baltimore and bullock orioles. Comparative Biochemistry and Physiology 31:915–925.

Rising, J. D. 1996. The stability of the oriole hybrid zone in western Kansas. The Condor 98:658–663.

Rohwer, S., L. Butler, & D. Froehlich. 2005. Ecology and demography of east—west differences in molt scheduling of neotropical migrant passerines. *In* R. Greenberg and P. Marra (Eds.), Birds of two worlds: the ecology and evolution of migratory birds (pp. 87–105). Johns Hopkins University Press.

Rohwer, S. &D. Irwin. 2011. Molt, oreintation, and avian speciation. The Auk 128:419–425.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, et al. 2019. Decline of the North American avifauna. Science 366:120–124.

Ryan, S. F., J. M. Deines, J. Mark Scriber, et al. 2018. Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. Proceedings of the National Academy of Sciences of the United States of America 115: E2284–E2294.

Shastry, V., P. E. Adams, D. Lindtke, et al. 2020. Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. bioRxiv Page 2020.07.31.231514.

Sibley, C. G. & L. L. Short. 1964. Hybridization in the Orioles of the Great Plains. The Condor 66:130–150.

Sibley, C. G. and L. L. Short Jr. 1959. Hybridization in the buntings (Passerina) of the Great Plains. The Auk 76:443–463.

Sin, S. Y. W., L. Lu, and S. V. Edwards. 2020. De novo assembly of the Northern Cardinal (Cardinalis cardinalis) Genome Reveals Candidate Regulatory Regions for sexually dichromatic red plumage coloration. G3 Genes|Genomes|Genetics 10:3541–3548.

Slager, D. L., K. L. Epperly, R. R. Ha, et al. 2020. Cryptic and extensive hybridization between ancient lineages of American crows. Molecular Ecology 29:956–969.

Smith, S. J., J. Edmonds, C. A. Hartin, A. Mundra, & K. Calvin. 2015. Near-term acceleration in the rate of temperature change. Nature Climate Change 5:333–336.

Stevens, H. C., E. J. Williams, C. Q. Stanley, et al. 2023. Incorporating drivers of global change throughout the annual cycle in species distribution models for migratory birds: A gap in ecological forecasting. Frontiers in Bird Science 2:1230978.

Strimas-Mackey, M., W. M. Hochachka, V. Ruiz-Gutierrez, et al. 2023. Best Practices for Using eBird Data. Version 2.0. Cornell Lab of Ornithology.

Strimas-Mackey, M., E. Miller, & W. Hochachka. 2018. auk: eBird Data Extraction and processing with AWK. R package version 0.3. 0.

Suh, Y. H., R. A. Ligon, & V. G. Rohwer. 2022. Revisiting the Baltimore-Bullock's Oriole hybrid zone reveals changing plumage colour in Bullock's Orioles. Royal Society Open Science 9:221211.

Sulla-Menashe, D., J. M. Gray, S. P. Abercrombie, & M. A. Friedl. 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS collection 6 land cover product. Remote Sensing of Environment 222:183–194.

Sullivan, B. L., C. L. Wood, M. J. Iliff, et al. 2009. eBird : A citizen-based bird observation network in the biological sciences. Biological Conservation 142:2282–2292.

Swenson, N. G. 2006. Gis-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. Journal of Evolutionary Biology 19:717–725.

Swenson, N. G. & D. J. Howard. 2004. Do suture zones exist? Evolution 58:2391–2397.

Taylor, S. A., E. L. Larson, & R. G. Harrison. 2015. Hybrid zones: Windows on climate change. Trends in Ecology & Evolution 30:398–406.

Taylor, S. A., T. A. White, W. M. Hochachka, V. Ferretti, R. L. Curry, & I. Lovette. 2014. Climate-mediated movement of an avian hybrid zone. Current Biology 24:671–676.

Thornton, P. E., R. Shrestha, M. Thornton, et al. 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. Scientific Data 2021 8:1 8:190.

Walker, J. and P. D. Taylor. 2017. Using eBird data to model population change of migratory bird species. Avian Conservation and Evology 12.

Walsh, J., S. M. Billerman, V. G. Rohwer, B. G. Butcher, & I. J. Lovette. 2020. Genomic and plumage variation across the controversial Baltimore and Bullock's oriole hybrid zone. The Auk 137:1–15.

Wang, S., S. Rohwer, K. Delmore, & D. E. Irwin. 2019. Cross-decades stability of an avian hybrid zone. Journal of Evolutionary Biology 32:1242–1251.

Watanabe, M., H. Shiogama, H. Tatebe, et al. 2014. Contribution of natural decadal variability to global warming acceleration and hiatus. Nature Climate Change 4:893–897.

Wood, S. N. 2001. Mgcv: GAMs and generalized ridge regression for R. R News 1:20–25.

Young, B. E. 1991. Annual molts and interruption of the fall migration for molting in Lazuli buntings. The Condor 93:236–250.

Zhao, C., F. Brissette, J. Chen, & J. L. Martel. 2020. Frequency change of future extreme summer meteorological and hydrological droughts over North America. Journal of Hydrology 584: 124316.

Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, & G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Statistics for biology and health (1st ed.). Springer New York.