University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

USDA National Wildlife Research Center - Staff
Publications

U.S. Department of Agriculture: Animal and
Plant Health Inspection Service

2021

Environmental correlates of genetic variation in the invasive European starling in North America

Natalie R. Hofmeister
Cornell University, nrh44@cornell.edu

Scott J. Werner
APHIS, scott.j.werner@aphis.usda.gov

Irby J. Lovette
Cornell University, ij12@cornell.edu

Follow this and additional works at: https://digitalcommons.unl.edu/icwdm_usdanwrc

Part of the [Natural Resources and Conservation Commons](#), [Natural Resources Management and Policy Commons](#), [Other Environmental Sciences Commons](#), [Other Veterinary Medicine Commons](#), [Population Biology Commons](#), [Terrestrial and Aquatic Ecology Commons](#), [Veterinary Infectious Diseases Commons](#), [Veterinary Microbiology and Immunobiology Commons](#), [Veterinary Preventive Medicine, Epidemiology, and Public Health Commons](#), and the [Zoology Commons](#)

Hofmeister, Natalie R.; Werner, Scott J.; and Lovette, Irby J., "Environmental correlates of genetic variation in the invasive European starling in North America" (2021). *USDA National Wildlife Research Center - Staff Publications*. 2409.

https://digitalcommons.unl.edu/icwdm_usdanwrc/2409

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Animal and Plant Health Inspection Service at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USDA National Wildlife Research Center - Staff Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Environmental correlates of genetic variation in the invasive European starling in North America

Natalie R. Hofmeister^{1,2} | Scott J. Werner³ | Irby J. Lovette^{1,2}

¹Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

²Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA

³United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, CO, USA

Correspondence

Natalie R. Hofmeister, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA.
Email: nrh44@cornell.edu

Funding information

This work was supported by the Cornell Lab of Ornithology Athena Grant, the Andrew W. Mellon Student Research Grant, the American Ornithological Society Research Award, and the Cornell University EEB Paul P. Feeny Graduate Student Research Fund (all to N.R.H.). N.R.H. was also supported by the following fellowships: Charles Walcott Graduate Fellowship, Treman-Williams Graduate Fellowship, Edsall Ornithology Graduate Fellowship, and the Ivy Fellowship.

Abstract

Populations of invasive species that colonize and spread in novel environments may differentiate both through demographic processes and local selection. European starlings (*Sturnus vulgaris*) were introduced to New York in 1890 and subsequently spread throughout North America, becoming one of the most widespread and numerous bird species on the continent. Genome-wide comparisons across starling individuals and populations can identify demographic and/or selective factors that facilitated this rapid and successful expansion. We investigated patterns of genomic diversity and differentiation using reduced-representation genome sequencing of 17 winter-season sampling sites. Consistent with this species' high dispersal rate and rapid expansion history, we found low geographical differentiation and few F_{ST} outliers even at a continental scale. Despite starting from a founding population of ~180 individuals, North American starlings show only a moderate genetic bottleneck, and models suggest a dramatic increase in effective population size since introduction. In genotype-environment associations we found that ~200 single-nucleotide polymorphisms are correlated with temperature and/or precipitation against a background of negligible genome- and range-wide divergence. Given this evidence, we suggest that local adaptation in North American starlings may have evolved rapidly even in this wide-ranging and evolutionarily young system. This survey of genomic signatures of expansion in North American starlings is the most comprehensive to date and complements ongoing studies of world-wide local adaptation in these highly dispersive and invasive birds.

KEYWORDS

gene flow, invasion genetics, local adaptation, panmixia

1 | INTRODUCTION

Studies of local adaptation have long bridged the interface between ecological and evolutionary questions by exploring how populations adapt to differing environmental conditions. Traditionally, high degrees of local adaptation were expected to be present only in fairly isolated populations—those free from the homogenizing effects of high gene flow—with a long history in those locations, providing

the time thought to be necessary for local adaptation to evolve (Lenormand, 2002). We now know that local adaptation occurs frequently even in systems with high gene flow (Tigano & Friesen, 2016; Yeaman & Whitlock, 2011) and often rapidly after colonization of a novel environment (Prentis et al., 2008). We continue to find evidence for rapid local adaptation in systems as divergent as cane toads (Rollins et al., 2015), sticklebacks (Lescak et al., 2015), honeybees (Avalos et al., 2017), steelhead trout (Willoughby et al.,

2018), deer mice (Pfeifer et al., 2018) and many more. These studies show that many taxa can adapt rapidly to local conditions in response to the new selection regimes they encounter as they expand their range.

Invasive species that have recently expanded into new environments provide tractable opportunities to investigate local adaptation as it originates (Colautti & Lau, 2015). Invasions typically expand from the founding population(s) following a predictable spatial and temporal pattern (reviewed by Excoffier et al., 2009). Many invasive species experience genetic bottlenecks as a result of an initial founder effect, but this loss of standing genetic diversity does not necessarily limit the success of invasive species (Dlugosch & Parker, 2008; Facon et al., 2011; Schmid-Hempel et al., 2007); invasion biologists describe this phenomenon as the paradox of invasion (Estoup et al., 2016; Schrieber & Lachmuth, 2017). After successful colonization of a new habitat, invasive species often show a demographic boom facilitated by ecological release in the new environment. Ecological release is the concept that introduced species are often released from top-down limitations to population growth, such as predators or pathogens in their native range (Ricciardi et al., 2013). At the same time, these novel ecological conditions may select among standing genetic variation, where the presence of certain genetic variants in the native range accelerates adaptation upon introduction (Hufbauer et al., 2011; Schlaepfer et al., 2009; Tsutsui et al., 2000). Invasions thus allow us to observe interactions between demography and the early processes of selection (Dlugosch et al., 2015) as populations experience new environments.

Spatial sorting on the invasion front may allow dispersing individuals to maximize spatiotemporal fitness and drive incipient adaptation, as in other starling invasions (Phair et al., 2018), common mynas (Berthouly-Salazar et al., 2012), pumpkinseed fish (Ashenden et al., 2017) and cane toads (Brown et al., 2014). More empirical work is needed to verify the conditions in which spatial sorting can lead to lasting shifts in fitness (Lee, 2011; Phillips & Perkins, 2019; Williams et al., 2019). However, we know that adaptive dispersal strategies can facilitate range expansion in western bluebirds (Duckworth, 2008), invasive beetles (Lombaert et al., 2014; Ochocki & Miller, 2017) and other species. Certain dispersal strategies can result in gene flow that counteracts inbreeding depression and increases adaptive potential (Garant et al., 2007; Rius & Darling, 2014). If particular traits enable individuals to disperse more easily to their preferred habitat, gene flow may be directional and even adaptive (Edelaar & Bolnick, 2012; Jacob et al., 2017). Dispersal among populations will certainly impact population structure of an invasion, but could even support local adaptation.

The European starling (*Sturnus vulgaris*) stands out as an exceptionally successful avian colonist and invasive species. In North America, an estimated 85 million starlings currently range from northern Mexico to southern Alaska; this estimate represents a dramatic decline since previous estimates of 160–200 million starlings (Linz et al., 2007; Rosenberg et al., 2019). Introduced to New York City in 1890, starlings nearly covered the continent within a few generations by expanding up to 91 km each year (Bitton &

Graham, 2014). The 1890 introduction is widely accepted as the first successful establishment of starlings in North America, but several populations were introduced in Cincinnati, Ohio (1872), Quebec, Canada (1875), Allegheny, Pennsylvania (1897), and Springfield, Massachusetts (1897), with the second-most successful having been introduced to Portland, Oregon, in 1889 (Forbush, 1915; Kalmbach, 1921). We note here that multiple invasions from different source populations may lead to admixture among previously isolated populations and thus the introduction of new alleles to a population (Dlugosch & Parker, 2008). Records indicate that none of the earlier starling introductions survived more than a few years after colonization, but it is possible that some populations in the western USA persisted without record (Kessel, 1953). During the starling expansion, ongoing seasonal migration and dispersal might have also influenced patterns of genetic variation. In North America, some—but not all—starling populations migrate (Dolbeer, 1982). Previous studies indicate that there is considerable variation in migratory distances within flyways (Burt & Giltz, 1977), and early genetic work based on a small set of allozyme markers indicated near-random mating at a continental scale in North American starlings, with large demes (subpopulations) and high dispersal rates (Cabe, 1998, 1999). More recently, a survey of mitochondrial diversity across several locations in the North American range also found little evidence of population structure (Bodt et al., 2020).

Here we use genome-wide markers to explore the population genetics and possible genotype–environment associations in North American starlings with four specific aims: (a) to characterize genome-wide levels of diversity and differentiation among starlings; (b) to examine how genetic variation changes across the contiguous United States; (c) to test for a genetic bottleneck; and (d) to test for signatures of selection associated with environmental gradients. We test for these genotype–environment associations using overwintering sites, and interpret our results in the context of range-wide data on starling migration and dispersal (Werner et al., 2020), as this movement certainly influences population structure. Models of moult origin—which rely on feathers sampled from the same individuals in our genetic survey—suggest that starlings that migrate in eastern North America probably experience greater gene flow among sampling locations (states, in our study) compared to those in western North America, but overall starlings tend to move only regionally and not continent-wide (Werner et al., 2020). This work employs modern genomic and analytical tools to examine the evolutionary history of this remarkably successful avian invasion.

2 | METHODS

2.1 | Sample collection and processing

All starlings sampled were collected during the nonbreeding season, when large flocks aggregate at high-quality food sources. These samples therefore do not represent discrete breeding populations but rather a mixture of birds from the surrounding region,

potentially including migrants from more northerly areas. Starlings ($N = 166$) were collected in January–March 2016 and 2017 from 26 dairies and feedlots by the U.S. Department of Agriculture (USDA) Wildlife Services personnel in Arizona, California, Colorado, Idaho, Illinois, Iowa, Kansas, Missouri, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, Texas, Washington and Wisconsin (Figure 1). USDA Wildlife Services personnel collected birds from two to five sites in each state, where each collection site was >5 km apart. USDA personnel then euthanized whole adult males by lethal gunshot, avicide or live traps, stored these birds at 0°C until tissue sampling, and recorded the latitude and longitude of each collection site using a handheld GPS. The collection and use of starlings for this and related studies were approved by the USDA, National Wildlife Research Center's Institutional Animal Care and Use Committee (QA-2572, S. J. Werner—Study Director; Werner et al., 2020; Table S1).

Breast muscle tissue was sampled using biopsy punches (Integra Miltex) and frozen in 95% ethanol. Samples were shipped on dry ice, and DNA was extracted using a Qiagen DNeasy kit following the manufacturer's protocol (Qiagen). DNA concentration of each sample was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific). Following the protocol of Peterson et al. (2012), we generated a reduced-representation genomic data set of double-digested, restriction-site associated DNA (RAD) markers as described in Thrasher et al. (2017) using the restriction enzymes *SbfI* and *MspI* and adaptors P1 and P2. We sequenced 100-bp, single-end reads of the 160 best-quality libraries on an Illumina HiSeq 2500. We trimmed and filtered for quality using the FASTX-TOOLKIT (http://hannonlab.cshl.edu/fastx_toolkit). We then used the `process_radtags` commands in STACKS version 1.19 (Catchen et al., 2013) to demultiplex the remaining sequences. In subsequent filtering steps, we retained reads only if the following conditions were met: reads passed the Illumina chastity filter, contained an intact *SbfI* RAD site, contained one of the unique barcodes, and did not contain Illumina indexing adaptors.

Individual reads were mapped to a *Sturnis vulgaris* reference genome (Hofmeister, Rollins et al., in prep) using BOWTIE2 version 2.2.8

(Langmead & Salzberg, 2012) using the “very sensitive local” set of alignment presets, and then assembled sequences into “stacks” using the `ref-map` option in STACKS. Compared to a reference-free approach, the bioinformatics pipeline used for the reference-based assembly has the advantage of using fewer similarity thresholds to build loci. We required that a single-nucleotide polymorphism (SNP) be present in a minimum of 80% of the individuals ($-r$ 0.8) with a minimum stack depth of 10 reads at a locus within an individual ($-m$ 10) for it to be called. We removed two individuals, one with >50% missing data and one with >50% relatedness (measured using the unadjusted AJK statistic and calculated within *vcftools*), leaving 158 individuals remaining in the study. A total of 15,038 SNPs were identified. We used the *VCFTOOLS* `-hwe` option to remove any SNPs out of Hardy–Weinberg equilibrium (HWE) (e.g., an exact test that compared expected and observed heterozygosity in polymorphic sites only gave a p -value less than .001). About 6% of sequenced variants (904 variants) were out of HWE across all sampling sites; given that (i) we are particularly interested in SNPs that may be specific to certain populations, and (ii) filtering for HWE did not change the results described in sections (1) and (2) below, we retain all 15,038 SNPs for the results presented here. For all genotype–environment analyses we used all SNPs in a given stack, but for *STRUCTURE* and other analyses sensitive to linkage disequilibrium we used only the first SNP in each stack; for unphased genomic data such as the RAD loci analysed here, this strategy of exporting one SNP per stack is often used as an indirect method of controlling for linkage disequilibrium.

(1) Patterns of genetic diversity and differentiation

We estimated per-locus measures of genetic diversity and genome-wide differentiation using the *populations* option in STACKS (Catchen et al., 2013). We used *VCFTOOLS* to calculate F_{ST} among population pairs and heterozygosity and nucleotide diversity (π) within sampling sites (“populations”) (Danecek et al., 2011). We determined expected heterozygosity at the population level using the *Hs()* function in the R package ADEGENET (Jombart, 2008), and tested for differences in observed and expected heterozygosity of each locus using a

FIGURE 1 Sampling map of all locations, where color indicates inbreeding (F_{IS}) within each sampling location [Colour figure can be viewed at wileyonlinelibrary.com]

paired *t* test in base R (R Core Team, 2019). We also used ADEGENET to calculate ϕ_{ST} between populations. We investigated genetic structure within North American starlings using an analysis of molecular variance (AMOVA) in the R package POPPR (Kamvar et al., 2014). We tested whether most genetic variation was observed among individuals or among sampling sites. To determine significance, we compared observed variation at each hierarchical level to the randomly permuted distance matrices for that particular level using 999 simulations in the function *randtest()* in the R package ADEGENET (Jombart, 2008), hypothesizing that the observed variance is greater than expected within individuals and less between individuals and between sampling sites.

We tested for isolation by distance (IBD) using a simple Mantel test in ADEGENET (Jombart, 2008): for these data, the assumption of stationarity probably holds, given that North American starlings appear to be in mutation-migration-drift equilibrium (Guillot & Rousset, 2013). Geographical distances among sampling locations follow a bimodal distribution where locations are either very near or far from each other, and thus we caution that the Mantel test may not be an appropriate test of isolation by distance (Supplementary Information Section 2). Nevertheless, we report the results of a Mantel test with a Spearman correlation and 9,999 permutations.

(2) Population structure

We first used nonparametric approaches to determine whether 6,287 SNPs clustered by sampling location, using principal components analysis in SNPRELATE (Zheng et al., 2012) and discriminant analysis of principal components (DAPC) in ADEGENET (Jombart, 2008). We then tested for population structure using STRUCTURE (Pritchard et al., 2000) by simulating 10 runs of each *K*. Although we sampled starlings from 17 locations, we hypothesized that North American starlings would cluster into at most eight populations (*K* = 1–8) based on the nonparametric tests described above. To select the best-supported *K*, we used the Evanno method implemented in STRUCTURE HARVESTER version 0.6.94 (Earl & vonHoldt, 2011). We averaged results across the 10 runs using the greedy algorithm in the program CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007), and visualized results using DISTRICT version 1.1 (Rosenberg, 2003). Given that evidence of population structure depends on the filtering thresholds selected, we ran this model-based approach using a very strict minimum minor allele frequency (MAF = 0.3, *N* SNPs = 3568) and a more relaxed minimum frequency (MAF = 0.1, *N* SNPs = 6287) (Linck & Battey, 2017). STRUCTURE results did not differ substantially among MAF thresholds, and we used a filtered data set with a minimum MAF of 0.1 in the demographic analyses below.

(3) Demographic modelling

We used a traditional model-based approach (FASTSIMCOAL2) to test for a bottleneck in North American starlings (Excoffier et al., 2013). We estimated the folded SFS using a PYTHON script from Simon Martin (available at <https://github.com/simonmartin/genom>

ics_general). We modified the simple 1PopBot20 Mb.tpl and.est files provided with the software as follows: the size of the bottleneck (NBOT: 10–1,000), start of bottleneck (TBOT: 10–300), ancestral size (ANC: 1000–10,0000), current size of the population (NCUR: 100–100,000), and end of bottleneck (TENDBOT: TBOT +500). We performed 100 runs of this model and assessed model fit using delta-likelihood for each run of the model; because all models use the same numbers of parameters, the Akaike information criteria do not change between models. We also ran a demographic model that takes linkage into account, the Stairway plot method, but given that this method does not explicitly test for a genetic bottleneck, we describe the method and results in the (Supplementary Information Section 3).

(4) Selection analyses

We first identified environmental variation at each sampling location using the R package RASTER (Hijmans & van Etten, 2012), extracting all 19 bioclimatic variables for each set of sampling coordinates from WorldClim 2.0 at a resolution of 5 min on June 16, 2018 (Fick & Hijmans, 2017). We test for isolation by environment (IBE) using the R package VEGAN to compare environmental, geographical and genetic distances (Oksanen et al., 2019). Controlling for geographical distance, we ran partial Mantel correlograms using Euclidean environmental distance matrices built for each bioclimatic variable. To examine how loci may covary across multiple environmental gradients, we used redundancy analysis (RDA) (Forester et al., 2018). RDA is especially powerful when testing for weak selection, and detects true positives in large data sets more reliably than other multivariate methods like Random Forest (Forester et al., 2018). Because RDA requires no missing data, we first imputed genotypes where missing sites were assigned the genotype of highest probability across all individuals—a conservative but quick imputation method. Before imputation and after filtering, about 31% of genotypes were missing from the data set of 15,038 SNPs (not filtered for MAF), and we required that all imputed SNPs were present in at least 80% of individuals. We then used the R package VEGAN to run the RDA model (Oksanen et al., 2019); for a full description of this method, see Forester et al. (2018). Briefly, RDA uses constrained ordination to model a set of predictor variables, and unconstrained ordination axes to model the response (genetic variation). RDA infers selection on a particular locus when it loads heavily onto one or more unconstrained predictor axes, and we retained five variables with relatively low variance inflation factors (VIF): BIO1 or Annual Mean Temperature (VIF = 3.54), BIO7 or Temperature Annual Range (4.55), BIO12 or Annual Precipitation (8.69), BIO16 or Precipitation of Wettest Quarter (7.91) and elevation (2.19). Each of these variables appeared to be normally distributed, and correlations among predictors are all lower than .57, except for the correlation between BIO1 and BIO7 where *R* = -.79. We tested for significance using the *anova.cca* function within the *vegan* package, and also permuted predictor values across individuals to further check significance of the model. We identify

candidate loci as those that are three or more standard deviations outside the mean loading. The R scripts for all RDAs and figures were written by Brenna Forester (available at https://popgen.nescent.org/2018-03-27_RDA_GEA.html).

3 | RESULTS

(1) Patterns of genetic diversity and differentiation

We identified 15,038 SNPs at a mean of 27× sequencing depth across 17 sampling locations. Using a data set filtered by an MAF of 0.1, we find that genome-wide F_{ST} is extremely low (0.0085), and measures of genetic diversity do not vary substantially among sampling locations (Figure 1, Table 1). Across all of these wintering starling sampling sites, the highest pairwise F_{ST} was 0.0106 between birds from the adjacent states of Arizona and New Mexico. Using a haplotype-based statistic of differentiation, ϕ_{ST} among sampling sites shows an absence of genetic structure ($\phi_{ST} = 0.0002$). Hierarchical AMOVAs reveal that 94% of the observed genetic variance is explained by variation within individuals, and the remaining variance reflects differences among individuals from the same sampling site, with negligible variation explained at the between-population level (Figure 2c,d). Across the genome, F_{ST} and nucleotide diversity are exceptionally low (Tables 1 and 2). Genome-wide heterozygosity is moderate at 0.339, and across loci observed heterozygosity differs significantly from expected ($t = 66.6$, $df = 3569$, $p < .001$, Table 1).

(2) Population structure

F_{ST} among sampling sites is relatively low overall (Table 2): starlings sampled in Arizona show the highest differentiation from other sampling locations ($F_{ST} = 0.008$ – 0.011), and only a few other pairwise comparisons (NM-IL, NM-CO and CO-WI) show an F_{ST} of 0.009 or higher. However, we do not recover clear population structure. The first two principal components each explain about 1% of the variation among individuals (Figure 2a), and although STRUCTURE identified three populations at the best-supported value of K , we do not observe obvious differences in ancestry proportions among predicted “populations” (Figure 2b; Figure S1). Controlling for shared ancestry does not resolve population structure, and instead provides support for uniform gene flow among individuals (Figure S2). K-means clustering within DAPC also does not identify biologically relevant clusters.

Starling “populations” (sampling sites) follow a clear pattern of IBD (Mantel $r = .139$, $p < .0001$, Figure S3). Spatially explicit models of IBD suggest a fairly uniform rate of migration range-wide, where local increases in migration rate are probably a model artifact (Supplementary Information Section 2, Figure S4). However, models of IBE show that the relationship between environmental and genetic distances is stronger than the simple geographical-genetic distance model. After controlling for geographical distance, we find that all bioclimatic variables tested show nonzero correlations between environmental and genetic distances (Supplementary Information Section 2, Table S1). There is a strong positive relationship between genetic distance and both precipitation in the wettest

TABLE 1 Population genetic summary statistics

Sampling location	N ind.	Private	H_{obs}	H_{exp}	π	F_{IS}
Arizona	10	1	0.32	0.24	0.35	0.069
California	11	0	0.33	0.25	0.35	0.061
Colorado	8	1	0.33	0.24	0.35	0.066
Idaho	8	0	0.33	0.22	0.35	0.047
Illinois	7	1	0.33	0.26	0.35	0.056
Iowa	9	0	0.33	0.24	0.36	0.078
Kansas	9	2	0.35	0.23	0.35	0.012
Missouri	9	0	0.33	0.23	0.35	0.053
Nebraska	11	0	0.34	0.22	0.35	0.042
Nevada	10	0	0.33	0.31	0.36	0.068
New Hampshire	11	1	0.33	0.26	0.35	0.065
New Mexico	7	2	0.33	0.29	0.36	0.068
New York	10	0	0.34	0.25	0.35	0.037
North Carolina	11	1	0.33	0.25	0.35	0.060
Texas	9	1	0.33	0.24	0.35	0.052
Wisconsin	8	0	0.33	0.22	0.35	0.050
Washington	10	2	0.32	0.30	0.35	0.082

Note: For each sampling location, the table shows the number of individuals, number of private alleles, observed and expected heterozygosity, nucleotide diversity, and the inbreeding coefficient F_{IS} .

FIGURE 2 Population structure. (a) Principal components analysis on 6,287 SNPs, with PC1 explaining 1.07% and PC2 explaining 1.03% of the genetic variation observed. (b) STRUCTURE analyses with $K = 2$ and $K = 3$ (best supported). (c) Significance testing of hierarchical AMOVAs: the histogram shows expected variance components based on 999 simulations, and the black diamond is the observed variance component. (d) AMOVA results [Colour figure can be viewed at wileyonlinelibrary.com]

quarter (BIO16, Mantel $r = .282$, $p = .001$, Table S1) and elevation (Mantel $r = .146$, $p = .001$, Table S1).

(3) Demographic modelling

A traditional SFS-based model recovers a bottleneck in population size, as expected: the estimated effective population size during the bottleneck is ~16,000 individuals (1% of the ancestral population size). However, our demographic model suggests that the current population size of North American starlings is nearly identical to its prebottleneck size (Supplementary Information Section 3, Figure S5). In fact, every run of the model finds that the current effective population size of starlings (1.6 million individuals) is considerably higher than the estimated N_e of the founding population. This model suggests that starlings experienced rapid population growth after the initial founder event, which may contribute to the overall lack of evidence for inbreeding. We do detect low levels of inbreeding within some populations (Table 1, highest $F_{IS} = 0.082$ in Washington). Together, these models do not suggest a classical founder effect, whereby effective population size remains low for many generations post-bottleneck.

(4) Selection analyses

Starlings encounter a range of precipitation, temperature and elevation across their range and redundancy analyses revealed the strongest signatures of local adaptation compared to all selection testing methods (see Supplementary Information Sections 5–7, Figures S6–Figure S8). The RDA model showed that 178 variants are correlated with environmental differences among sampling sites ($F = 1.022$, $p = .002$, Figure 3). Starlings living in warmer climates tend to cluster more closely in the left quadrant and high-elevation individuals cluster in the middle right quadrant. However, starlings do not cluster based on geographical distance: for example, starlings from Texas and Washington cluster closely due to shared genetic variants, even though the two sampling sites differ substantially in precipitation and temperature and are geographically separated. After controlling for population structure, candidates for selection are equally distributed among elevation-, temperature- and precipitation-associated predictors. Importantly, when environmental predictors are randomly shuffled, the RDA model is not significant. Mean annual temperature (BIO1) opposes selective pressure related to the range of temperatures experienced each year (BIO7),

TABLE 2 Differentiation and movement among sampling locations

	AZ	CA	CO	IA	ID	IL	KS	MO	NC	NE	NH	NM	NV	NY	TX	WA	WI
Arizona	28	0.008 5	0.009 0	0.009 3	0.009 0	0.008 2	0.009 1	0.01 20	0.009 3	0.009 0	0.01 25	0.011 1	0.01 00	0.01 00	0.01 20	0.01 0	0.01 50
California	5	40	0.008 0	0.006 5	0.005 0	0.005 0	0.005 10	0.005 0	0.007 5	0.006 0	0.007 0	0.005 1	0.006 0	0.006 0	0.007 0	0.006 0	0.006 1
Colorado	0	7	43	0.006 0	0.008 7	0.008 2	0.007 0	0.007 0	0.006 0	0.006 1	0.007 7	0.01 40	0.007 2	0.006 0	0.007 0	0.007 2	0.009 0
Iowa	4	16	0	8	0.006 0	0.005 1	0.008 4	0.007 1	0.005 2	0.009 0	0.007 6	0.007 1	0.006 0	0.006 2	0.007 0	0.007 0	0.008 6
Idaho	0	0	4	0	23	0.007 0	0.007 0	0.006 0	0.007 0	0.005 4	0.007 0	0.008 0	0.004 8	0.006 0	0.007 0	0.006 2	0.007 0
Illinois	5	2	5	2	0	12	0.007 5	0.006 2	0.007 0	0.006 0	0.006 7	0.01 17	0.005 0	0.006 0	0.006 2	0.007 0	0.008 1
Kansas	2	12	2	0	0	5	7	0.005 1	0.006 3	0.005 2	0.006 5	0.006 2	0.005 0	0.006 2	0.007 2	0.006 0	0.006 1
Missouri	5	5	0	0	0	5	5	11	0.005 2	0.006 0	0.006 5	0.007 3	0.006 0	0.006 0	0.006 0	0.007 0	0.007 1
North Carolina	0	0	0	0	0	0	16	0	32	0.006 0	0.007 0	0.007 0	0.006 0	0.006 0	0.007 4	0.008 0	0.006 4
Nebraska	0	8	24	0	16	0	0	0	0	12	0.008 1	0.006 4	0.006 8	0.007 0	0.007 0	0.007 0	0.006 0
New Hampshire	28	6	0	3	0	0	3	3	0	6	9	6	0.007 3	0.006 0	0.007 0	0.007 0	0.007 1
New Mexico	11	0	0	0	0	0	0	0	11	11	0	67	0.006 0	0.008 0	0.007 0	0.007 0	0.007 0
Nevada	0	3	4	1	18	0	1	0	0	11	3	1	14	0.005 0	0.007 0	0.006 1	0.006 1
New York	3	10	0	1	0	15	1	10	0	1	6	1	1	1	0.007 0	0.006 0	0.005 2
Texas	5	0	0	0	0	0	0	3	26	0	0	2	0	0	65	0.007 0	0.008 0
Washington	0	0	0	0	0	0	0	0	0	7	0	0	60	0	0	7	0.006 0
Wisconsin	10	0	2	2	2	8	6	4	0	2	2	0	0	10	0	0	31

Note: This table shows the percentage of birds assigned to each sampling location, where columns indicate state of origin. Note that these values are directional: although 12 starlings that were collected in Arizona originated in Kansas, only two starlings collected in Kansas originated in Arizona. The black diagonal indicates the percentage of birds assigned to that collection state according to a discriminant function analysis of moult-origin presented in Werner et al. For example, 28% of birds collected in Arizona originated in that location. F_{ST} among locations is presented above the diagonal, and darker grey colors indicate higher F_{ST} .

Percentage assigned to each state does not sum to 100 because Minnesota, Oregon and Vermont are not included in the genetic data set.

FIGURE 3 Evidence for incipient local adaptation. Redundancy analyses indicate that 191 SNPs (small grey points) are associated with bioclimatic predictors (vectors). BIO1: mean annual temperature; BIO7: temperature annual range; BIO12: annual precipitation; BIO16: precipitation of wettest quarter. Significant associations are shown with black dots [Colour figure can be viewed at wileyonlinelibrary.com]

annual precipitation (BIO12), precipitation in the wettest quarter (BIO16) and elevation (Figure 3). Genes under selection tend to have lower allele frequencies, and function in several biological processes (Supplementary Information Section 8, Table S2).

4 | DISCUSSION

Our genome-wide data reveal that genetic variation in invasive starlings is associated with environmental parameters, and we suggest that these associations might result from adaptive processes. Although there is a significant (but low-magnitude) signal of IBD, hierarchical AMOVAs find that variation within and among individuals explains observed differentiation better than variation among sampling sites. There is no evidence of population structure, and while models indicate some subtle spatial patterns of genetic variation, these model-based inferences probably reflect sampling artefacts (Supplementary Information Sections 1–2, Figures S2 and S4). Finally, after the initial founder effect, the effective population size has grown by a ratio of 1:100. These patterns are consistent with the expectation that extensive gene flow—as shown by low F_{ST} among sampling sites (Table 2)—maintains high connectivity among North American starling “populations.” When interpreted within the context of a complementary exploration of movements inferred from feather isotope assays (Werner et al., 2020), this genomic survey confirms that dispersal and migration continue to influence the distribution of genetic variation as a result of just over a century of range expansion.

4.1 | Evidence of local adaptation in North American starlings

This species' low genome-wide differentiation across the continent allows for tests of selection on loci that may be involved in local adaptation. We find that almost 200 of the 15,038 RAD loci appear to be under selection using RDA. We discuss additional tests of selection in Supplementary Information Section 5, Figure S7 and Figure S8, but only 13 of the SNPs identified in the RDA model overlap with the SNPs identified by another selection scan described in the Supplementary Information (a latent-factor mixed model, Figure S8). It is unsurprising that each test identifies different candidates for selection, because the assumptions underlying each differ (for more details on selection testing, see Supplementary Information Section 5). Rather than making inferences based on the genes identified by these scans, we instead propose that genotype–environment associations show that changes in precipitation and temperature can explain genetic variation in North American starlings. Specifically, we hypothesize that aridity and low temperatures that are not experienced in the starling's native range exert enough selective pressure on North American starlings to result in incipient local adaptation. This finding suggests that local adaptation may explain genetic variation within the North American starling invasion, and we now discuss relevant caveats to this conclusion.

4.2 | Geography alone does not fully explain genetic structure

When we compare the relative importance of geographical and environmental distances in partial Mantel tests, we find that environmental conditions better explain genetic variation. We note that concordance among environmental and genetic distances indicates that spatial autocorrelation complicates our selection inferences. Under these conditions, any evidence for selection is likely to be weak, and these selection scans can generate false positives. However, of comparable genotype–environment association methods, RDA has the highest rate of true positives and lowest of false positives, and although this method has not been tested in such recent expansions, RDA is well suited to systems where F_{ST} is very low (Meirmans, 2015; Forester et al., 2018, Supplementary Information Section 5). In addition, when we shuffle environmental predictors randomly, we find no evidence for genotype–environment associations, lending additional support to the RDA model shown here.

The environmental conditions that we expect to drive selection—precipitation and temperature—vary most substantially in the southwestern region: for example, the sampling location in Arizona is consistently warmer (BIO1) and drier (BIO12 and BIO16) than other locations (Figure 3; Supplementary Information Section 5, Figure S6). Starlings in the Western region experience these environmental conditions year-round, which could allow selection to

drive advantageous alleles toward fixation. Elsewhere in the USA, starlings move more freely among states: individuals within each sampling location may come from different breeding populations, and additional sampling could reveal stronger population structure among true breeding populations. However, if our sampling overlooked some true populations, we would expect some signal of population structure. Individual-based tests of population structure (e.g., those that do not define possible populations a priori) do not recover any signals of population structure.

4.3 | Movement among sampling locations may influence genetic variation

Our study focuses on birds collected during the winter, which limits our inferences about selection to overwintering sites. We interpret the isotopic evidence presented in Werner et al. (2020) as showing that starlings in the western USA tend to move regionally whereas birds sampled in the eastern USA undertake longer movements (Table 2). This in turn suggests that starlings overwintering in the western USA are more likely to breed nearby, and thus the environmental conditions that starlings experience may not change as dramatically among wintering and breeding ranges. However, we recognize that these environmental conditions represent those experienced during the nonbreeding season, and therefore they do not represent the full range of conditions experienced by the starlings sampled; see Supplementary Information Section 4 for a discussion of our assumptions about this environmental sampling.

We note that lower rates of gene flow in the western USA—of which we find no evidence—could explain allele frequency shifts that we infer to be selection. Starling movement (e.g., migration and/or dispersal) should influence genetic differentiation among sampling locations: we expect high F_{ST} where gene flow among sites is lower, for example where starlings are assigned to a nearby sampling location according to Werner et al. (2020) model of moult origin. Starlings in the western USA appear to have differentiated subtly from their eastern counterparts based on the higher F_{ST} between Arizona—and to some degree, New Mexico and Colorado—and all other sampling locations (Table 2). Birds collected in these southwestern states are also assigned to those states by discriminant function analysis (Werner et al., 2020); for example, 67% of starlings collected in New Mexico were also assigned to that state (Table 2). We suggest that birds assigned to the same state where they were collected may reside in that state year-round, but we note that collecting feathers once in the lifespan of the bird does not allow us to determine the bird's lifelong migration and dispersal. This model indicates that dispersal and seasonal migration among some sampling locations could lead to gene flow among geographically distant "populations," and as expected, movement is not uniform across the starling's North American range. Given this evidence, we suggest that environmental conditions may also influence starling movement across the landscape.

4.4 | Rapid expansion and allele surfing may contribute to putative local adaptation

In North America, the effective population size (N_e) of the present-day invasive population has expanded dramatically, with models indicating that current N_e is even larger than the N_e of the founding population (Supplementary Information Section 3, Table S2, Figure S5). Given this demographic history of rapid expansion after a genetic bottleneck, it remains possible that allele surfing at the range edge could mimic these patterns of local adaptation (Excoffier et al., 2009; Slatkin & Excoffier, 2012). Theory predicts that allele surfing should drive alleles closer to fixation, but no putative variant under selection has an allele frequency greater than 0.28. Furthermore, given that the earliest North American starling population was established in the mid-19th-century—which allows only a short time for mutations to accumulate—it is likely that any allele frequency shifts reflect evolutionary processes acting on standing, ancestral variation. Future work might investigate whether particular alleles are shared with the ancestral population(s) or generated via mutation. Here we do not explicitly control for linkage, and we note that some allele frequency shifts could also be explained by linkage or by genetic hitchhiking (Cruickshank & Hahn, 2014).

We suggest that explosive population growth from hundreds to millions of individuals may have encouraged long-distance dispersal away from the dense populations in eastern North America. We know that long-distance dispersal is common in introduced starlings in South Africa (Hui et al., 2012) and Australia (Phair et al., 2018), where rates of dispersal may be determined by demographic changes and environmental quality. Invasive starlings in South Africa disperse when their natal environment becomes crowded or unsuitable, such as at the leading edge of their range expansion (Hui et al., 2012). In general, multiple lines of evidence have shown that the dispersal rates and distances of juvenile starlings in North America are remarkably high (Cabe, 1999; Dolbeer, 1982; Werner et al., 2020). Whether and how dispersal might influence local adaptation of expanding starling populations remains an open question.

4.5 | Local adaptation in other starling invasions and more

A similar project on starlings in the Australian invasion—which colonized that continent nearly concurrently with the North American invasion—found that geographical but not environmental distance explains genetic patterns there (Stuart et al., 2020). Starlings in the Australian range show substantial population structuring and significant patterns of IBD. Earlier work had shown that gene flow among Australian starling populations is low (Rollins et al., 2009), and phylogeographical patterns of mitochondrial sequence variation confirm that starlings on the edge of the expansion front in Western Australia have differentiated from those still living in the introduction site (Rollins et al., 2011). In fact, starlings at the expansion front may have rapidly adapted during the Australian

invasion (Rollins et al., 2016): the proportion of adult starlings in Western Australia carrying a novel mitochondrial haplotype has increased rapidly only at this range edge, which could indicate allele surfing. The genotyping-by-sequencing survey referenced above now indicates three population subdivisions in Australia, and global F_{ST} across all Australian populations is an order of magnitude higher than the equivalent F_{ST} index across North America, despite similar areas sampled (Stuart & Cardilini et al., 2020). In Australia, the strong evidence for IBD and founder effects complicate attempts to disentangle selection from drift, yet despite their differences in invasion dynamics, genotype–environment associations reveal signatures of selection in both invasions. We note that preliminary results of whole-genome resequencing of native (UK) and introduced (Australian and North American) populations suggest that variability in temperature and precipitation may shape observed genetic variation in starlings world-wide (Hofmeister et al., in prep).

Our results contribute to the growing evidence of rapid adaptation in some expanding populations, even in young systems. Some studies of rapidly expanding invasions find little evidence that adaptation may facilitate this expansion, as in corals (Leydet et al., 2018). However, other work suggests a role for selection in supporting rapid range expansion, such as in experimental studies of flour beetles (Szucs et al., 2017) and empirical work in guppies (Baltazar-Soares et al., 2019). Invasion biologists have long highlighted propagule pressure as a driver of invasion success, but the genetic composition may be just as important as the size of the establishing population (Briski et al., 2018). For example, genetic bottlenecks in monk parakeets, another avian invader now distributed world-wide, do not seem to inhibit invasion success (Edelaar et al., 2015). Pre-adaptation in the native range or selection during transport may facilitate the spread of invasive species, and human commensalism may support establishment and spread, as shown in house sparrows (Ravinet et al., 2018) and common mynas (Cohen et al., 2019), and reviewed across alien bird species (Cardador & Blackburn, 2019). Empirical studies of invaders like the ones described here also show how, in addition to genetic variation, epigenetic shifts and/or plastic changes in gene expression may support the establishment and expansion of invasive species (Marin et al., 2019). In the well-studied house sparrow—a system quite similar to starlings—epigenetic shifts may have supported invasions in Africa (Liebl et al., 2013) but not necessarily in Australia (Sheldon et al., 2018). Taken together, recent work suggests that we should consider a much wider range of demographic and ecological processes that lead to adaptive evolution in invading populations.

Invasive populations allow us to explore the genetic consequences of colonization and establishment in novel environments. On a background of low genetic differentiation and diversity, we find evidence of incipient genotype–environment associations in North American starlings. Here we explore how genetic variation changes across the landscape, but we cannot fully understand gene flow without studies of dispersal and migration of the individuals

that carry genes. Our results complement other recent studies that reveal associations between climate variables and particular loci in North American vertebrates (Bay et al., 2018; Schweizer et al., 2015). Finally, we suggest that our study adds to those suggesting that rapid local adaptation can evolve even in dispersive and young populations.

ACKNOWLEDGEMENTS

We are grateful to the U.S. Department of Agriculture's Wildlife Services personnel in Arizona, California, Colorado, Idaho, Illinois, Iowa, Kansas, Missouri, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, Texas, Washington and Wisconsin, who collected samples used here. Jennifer Walsh-Emond, Leonardo Campagna, Daniel Hooper, Jacob Berv and Stephanie Aguillon provided valuable advice on bioinformatic methods.

AUTHOR CONTRIBUTIONS

N.R.H. and I.J.L. designed the research, S.J.W. contributed samples, N.R.H. analysed data, and N.R.H. and I.J.L. wrote the paper.

DATA AVAILABILITY STATEMENT

All scripts are archived on GitHub: <https://github.com/nathofme/radseq-NAm>. All data files that accompany the above scripts are archived on Dryad (<https://doi.org/10.5061/dryad.j424f07>), and raw sequencing data archived in NCBI SRA (Bioproject PRJNA545151).

ORCID

Natalie R. Hofmeister <https://orcid.org/0000-0002-6875-412X>

REFERENCES

- Ashenden, T., Rooke, A. C., & Fox, M. G. (2017). Boldness and dispersal tendency of native and invasive pumpkinseed (*Lepomis gibbosus*): is spatial sorting creating superior invaders? *Aquatic Invasions*, 12, 311–320.
- Avalos, A., Pan, H., Li, C. Acevedo-Gonzalez, J. P., Rendon, G., Fields, C. J., Brown, P. J., Giray, T., Robinson, G. E., Hudson, M. E., & Zhang, G. (2017). A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nature Communications*, 8(1), 1–9.
- Baltazar-Soares, M., Blanchet, S., Cote, J., Tarkan, A. S., Záhorská, G. R. E., & Eizaguirre, C. (2019). Genomic footprints of a biological invasion: Introduction from Asia and dispersal in Europe of the topmouth gudgeon (*Pseudorasbora parva*). *Molecular Ecology*, 29, 71–85.
- Bay, R. A., Harrigan, R. J., Le Underwood, V., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359, 83–86.
- Berthouly-Salazar, C., van Rensburg, B. J., Le Roux, J. J., van Vuuren, B. J., & Hui, C. (2012). Spatial sorting drives morphological variation in the invasive bird, *Acridotheris tristis*. *PLoS One*, 7, e38145.
- Bitton, P.-P., & Graham, B. A. (2014). Change in wing morphology of the European starling during and after colonization of North America. *Journal of Zoology*, 295, 254–260.
- Bodt, L. H., Rollins, L. A., Zichello, J. M. (2020). Contrasting mitochondrial diversity of European starlings (*Sturnus vulgaris*) across three invasive continental distributions. *Ecology and Evolution*, 10(18), 10186–10195. <http://dx.doi.org/10.1002/eece3.6679>
- Briski, E., Chan, F. T., Darling, J. A., Lauringson, V., MacIsaac, H. J., Zhan, A., & Bailey, S. A. (2018). Beyond propagule pressure: importance

- of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment*, 16, 345–353.
- Brown, G. P., Phillips, B. L., & Shine, R. (2014). The straight and narrow path: the evolution of straight-line dispersal at a cane toad invasion front. *Proceedings of the Royal Society B: Biological Sciences*, 281, e20141385.
- Burt, H. E., & Giltz, M. L. (1977). Seasonal directional patterns of movements and migrations of Starlings and Blackbirds in North America. *Bird-Banding*, 48, 259–271.
- Cabe, P. R. (1998). The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity*, 80, 519–525.
- Cabe, P. R. (1999). Dispersal and population structure in the European Starling. *The Condor*, 101, 451–454.
- Cardador, L., & Blackburn, T. M. (2019). Human-habitat associations in the native distributions of alien bird species. *Journal of Applied Ecology*, 56, 1189–1199.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Cohen, T. M., McKinney, M., Kark, S., & Dor, R. (2019). Global invasion in progress: Modeling the past, current, and potential global distribution of the common myna. *Biological Invasions*, 21, 1295–1309.
- Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24, 1999–2017.
- Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157.
- Danecek, P., Auton, A., Abecasis, G., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, 24, 2095–2111.
- Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, 17, 431–449.
- Dolbeer, R. A. (1982). Migration patterns for age and sex classes of Blackbirds and Starlings. *Journal of Field Ornithology*, 53, 28–46.
- Duckworth, R. A. (2008). Adaptive dispersal strategies and the dynamics of a range expansion. *The American Naturalist*, 172, S4–S17.
- Earl, D. A., & von Holdt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.
- Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology & Evolution*, 27, 659–665.
- Edelaar, P., Roques, S., Hobson, E. A., Da Silva, A. G., Avery, M. L., Russello, M. A., Senar, J. C., Wright, T. F., Carrete, M., & Tella, J. L. (2015). Shared genetic diversity across the global invasive range of the monk parakeet suggests a common restricted geographic origin and the possibility of convergent selection. *Molecular Ecology*, 24, 2164–2176.
- Estoup, A., Ravnigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion? *Annual Review of Ecology, Evolution, and Systematics*, 47, 51–72.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLOS Genetics*, 9, e1003905.
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501.
- Facon, B., Hufbauer, R. A., Tayeh, A., Loiseau, A., Lombaert, E., Vitalis, R., Guillemaud, T., Lundgren, J. G., & Estoup, A. (2011). Inbreeding depression is purged in the invasive insect *Harmonia axyridis*. *Current Biology*, 21, 424–427.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.
- Forbush, E. H. (1915) *The starling*: Wright & Potter Printing Co.
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, 27, 2215–2233.
- Garant, D., Forde, S. E., & Hendry, A. P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, 21, 434–443.
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests (L. Harmon, Ed.). *Methods in Ecology and Evolution*, 4, 336–344.
- Hijmans, R. J., & van Etten, J. (2012). *raster: Geographic analysis and modeling with raster data*. R package version 2.0-12.
- Hufbauer, R. A., Facon, B., Ravnigné, V., Turgeon, J., Foucaud, J., Lee, C. E., Rey, O., & Estoup, A. (2012). Anthropogenically induced adaptation to invade (AIAI): Contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications*, 5, 89–101.
- Hui, C., Roura-Pascual, N., Brotons, L., Robinson, R. A., & Evans, K. L. (2012). Flexible dispersal strategies in native and non-native ranges: Environmental quality and the “good-stay, bad-disperse” rule. *Ecography*, 35, 1024–1032.
- Jacob, S., Legrand, D., Chaîne, A. S., Bonte, D., Schtickzelle, N., Huet, M., & Clobert, J. (2017). Gene flow favours local adaptation under habitat choice in ciliate microcosms. *Nature Ecology & Evolution*, 1(9), 1407–1410.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Kalmbach, E. R., & Gabrielson, I. N. (1921). Economic value of the starling in the United States. *USDA Bulletin*, 868, 1–59.
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281–e314.
- Kessel, B. (1953). Distribution and migration of the European Starling in North America. *The Condor*, 55, 49–67.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.
- Lee, M. S. Y. (2011). Macroevolutionary consequences of “spatial sorting”. *Proceedings of the National Academy of Sciences*, 108(31), E347–E347. <http://dx.doi.org/10.1073/pnas.1105702108>
- Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17, 183–189.
- Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, 112, E7204–E7212.
- Leydet, K. P., Grupstra, C. G. B., Coma, R., Ribes, M., & Hellberg, M. E. (2018). Host-targeted RAD-Seq reveals genetic changes in the coral *Oculina patagonica* associated with range expansion along the Spanish Mediterranean coast. *Molecular Ecology*, 27, 2529–2543.
- Liebl, A. L., Schrey, A. W., Richards, C. L., & Martin, L. B. (2013). Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integrative and Comparative Biology*, 53, 351–358.
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3), 639–647. <http://dx.doi.org/10.1111/1755-0998.12995>
- Linz, G. M., Homan, H. J., Gaulker, S. M., Penry, L. B., & Bleier, W. J. (2007). European starlings: a review of an invasive

- species with far-reaching impacts. In G. W. Witmer, W. C. Pitt, & K. A. Fagerstone (Eds.), *Managing Vertebrate Invasive Species: Proceedings of an International Symposium*, Fort Collins, Colorado, USA: USDA/APHIS Wildlife Services, National Wildlife Research Center.
- Lombaert, E., Estoup, A., Facon, B., Joubard, B., Grégoire, J.-C., Jannin, A., Blin, A., & Guillemaud, T. (2014). Rapid increase in dispersal during range expansion in the invasive ladybird *Harmonia axyridis*. *Journal of Evolutionary Biology*, 27, 508–517.
- Marin, P., Genitoni, J., Barloy, D., Maury, S., Gibert, P., Ghalambor, C. K., & Vieira, C. (2019). Biological invasion: The influence of the hidden side of the (epi)genome. *Functional Ecology*, 34, 385–400.
- Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24, 3223–3231.
- Ochocki, B. M., & Miller, T. E. X. (2017). Rapid evolution of dispersal ability makes biological invasions faster and more variable. *Nature Communications*, 8, 1–8.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for De Novo SNP discovery and genotyping in model and non-model species (L Orlando, Ed.). *PLoS One*, 7, e37135–e37211.
- Pfeifer, S. P., Laurent, S., Sousa, V. C., Linnen, C. R., Foll, M., Excoffier, L., Hoekstra, H. E., & Jensen, J. D. (2018). The evolutionary history of Nebraska Deer Mice: Local adaptation in the face of strong gene flow (N Singh, Ed.). *Molecular Biology and Evolution*, 35, 792–806.
- Phair, D. J., Le Roux, J., Berthouly-Salazar, C., Visser, V., Jansen, B., van Vuuren, A., & Cardilini, C. H. (2018). Context-dependent spatial sorting of dispersal-related traits in the invasive starlings (*Sturnus vulgaris*) of South Africa and Australia. *bioRxiv*, e342451. <https://doi.org/10.1101/342451>
- Phillips, B. L., & Perkins, T. A. (2019). Spatial sorting as the spatial analogue of natural selection. *Theoretical Ecology*, 12, 155–163.
- Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13, 288–294.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- R Core Team. (2019). *R: A language and environment for statistical computing*: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ravinet, M., Elgin, T. O., Trier, C., Aliabadian, M., Gavrillo, A., & Sætre, G.-P. (2018). Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B*, 285, e20181246.
- Ricciardi, A., Hoopes, M. F., Marchetti, M. P., Lockwood, J. L. (2013). Progress toward understanding the ecological impacts of nonnative species. *Ecological Monographs*, 83(3), 263–282. <http://dx.doi.org/10.1890/13-0183.1>
- Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, 29, 233–242.
- Rollins, L. A., Richardson, M. F., & Shine, R. (2015). A genetic perspective on rapid evolution in cane toads (*Rhinella marina*). *Molecular Ecology*, 24, 2264–2276.
- Rollins, L. A., Woolnough, A. P., Fanson, B. G., Cummins, M. L., Crowley, T. M., Wilton, A. N., Sinclair, R., Butler, A., & Sherwin, W. B. (2016). Selection on mitochondrial variants occurs between and within individuals in an expanding invasion. *Molecular Biology and Evolution*, 33, 995–1007.
- Rollins, L. A., Woolnough, A. P., Sinclair, R., Mooney, N. J., & Sherwin, W. B. (2011). Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology*, 20, 2307–2317.
- Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R., & Sherwin, W. B. (2009). Invasive species can't cover their tracks: Using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology*, 18, 1560–1573.
- Rosenberg, N. A. (2003). distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138.
- Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., Marra, P. P. (2019). Decline of the North American avifauna. *Science*, 366(6461), 120–124. <http://dx.doi.org/10.1126/science.aaw1313>
- Schlaepfer, D. R., Glättli, M., Fischer, M., & van Kleunen, M. (2009). A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. *New Phytologist*, 185, 1087–1099.
- Schmid-Hempel, P., Schmid-Hempel, R., Brunner, P. C., Seeman, O. D., & Allen, G. R. (2007). Invasion success of the bumblebee, *Bombus terrestris*, despite a drastic genetic bottleneck. *Heredity*, 99, 414–422.
- Schrieber, K., & Lachmuth, L. (2017). The Genetic Paradox of Invasions revisited: the potential role of inbreeding X environment interactions in invasion success. *Biological Reviews*, 92, 939–952.
- Schweizer, R. M., Robinson, J., Harrigan, R., Silva, P., Galverni, M., Musiani, M., Green, R. E., Novembre, J., & Wayne, R. K. (2015). Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology*, 25, 357–379.
- Sheldon, E. L., Schrey, A., Andrew, S. C., Ragsdale, A., & Griffith, S. C. (2018). Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. *Royal Society Open Science*, 5, e172185.
- Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: A spatial analogue of genetic drift. *Genetics*, 191, 171–181.
- Stuart, K. C., Cardilini, A. P. A., Cassey, P., Richardson, M. F., Sherwin, W. B., Rollins, L. A., Sherman, C. D. H. (2020). Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, <http://dx.doi.org/10.1111/mec.15601>
- Szűcs, M., Vahsen, M. L., Melbourne, B. A., Hoover, C., Weiss-Lehman, C., & Hufbauer, R. A. (2017). Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences*, 114, 13501–13506.
- Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, 18(5), 953–965. <http://dx.doi.org/10.1111/1755-0998.12771>
- Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25, 2144–2164.
- Tsutsui, N. D., Suarez, A. V., Holway, D. A., & Case, T. J. (2000). Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences*, 97, 5948–5953.
- Werner, S. J., Fischer, J. W., & Hobson, K. A. (2020). Multi-isotopic ($\delta^{2}\text{H}$, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. *PLoS One*, 15(8), e0237137.
- Williams J. L., Hufbauer R. A., Miller, T. E. X. (2019). How Evolution Modifies the Variability of Range Expansion. *Trends in Ecology & Evolution*, 34(10), 903–913. <http://dx.doi.org/10.1016/j.tree.2019.05.012>
- Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T., & Christie, M. R. (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, 17, 675.
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65, 1897–1911.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hofmeister NR, Werner SJ, Lovette IJ. Environmental correlates of genetic variation in the invasive European starling in North America. *Mol Ecol*. 2021;30:1251–1263. <https://doi.org/10.1111/mec.15806>