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

Running title: Starling genotype-environment associations

Natalie R. Hofmeister[1,2*], Scott J. Werner[3], and Irby J. Lovette[1,2]

[1] Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca, NY 14853

[2] Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850

[3] United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, 4101 LaPorte Avenue, Fort Collins, CO 80521

*corresponding author: nrh44@cornell.edu

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 (ddRADseq) of 17 winter-season starling populations. Consistent with this species' high dispersal rate and rapid expansion history, we found low geographic differentiation and few F_{ST} outliers even at a continental scale. Despite starting from a founding population of approximately 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. Local adaptation in North American starlings may have evolved rapidly even in this wide-ranging and evolutionarily young population. 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: panmixia, gene flow, local adaptation, invasion genetics

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 (Yeaman & Whitlock 2011; Tigano & Friesen, 2016) 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 locations 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 in Excoffier *et al*., 2009). After successful colonization of a new habitat, many invasive species show a demographic boom that may be facilitated by their ecological release in the new environment; ecological release is the concept that introduced species are often released from top-down limitations to their population growth, such as predators or pathogens in their native range (Riccardi *et al.,* 2013). Theory predicts that this rapid population growth will plateau as the population approaches carrying capacity in a region, but successful invasive species may continue to expand their range and thus maintain a high rate of overall population growth.

When population density increases and demographic rates change, introduced species may rapidly evolve traits that enable them to spread (Szűcs *et al.*, 2017): for example, invasive cane toads in Australia evolved a suite of morphological, physiological, and behavioral traits that facilitated their expansion in only 80 years (Rollins *et al.,* 2015).

Increased dispersal may evolve in concert with the demographic boom, such that certain traits are selected for in successful invasive species. Flexible 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). Invasions thus allow us to observe interactions between demography and the early processes of selection (Dlugosch *et al.,* 2015) as populations experience new environments. Importantly, these eco-evolutionary interactions depend in part on the genetic variation in the established population.

Recent work in invasion genetics aims to tease apart how selection and demography might resolve a paradox of invasion (Estoup *et al.*, 2016, Schrieber & Lachmuth, 2017): many invasive species experience genetic bottlenecks as a result of an initial founder effect, but often thrive and spread despite this loss of standing genetic diversity. Theory predicts that introductions will typically result in an initial contraction in population size and/or genetic diversity (Dlugosch & Parker, 2008). However, bottlenecks of genetic variation clearly do not limit the success of many invasive species (Schmid-Hempel *et al.,* 2007; Dlugosch & Parker, 2008; Facon *et al.,* 2011). Invaders might adapt through soft sweeps that reduce genetic diversity while selecting for adaptive variants from standing or novel genetic variation, which is especially likely in the case of ecological adaptation (Messer & Petrov, 2013). Furthermore, some invasions may increase rather than reduce genetic diversity, as when multiple invasions from different source populations introduce previously isolated alleles and thereby facilitate admixture (Dlugosch & Parker, 2008). The new conditions can also select among standing variation, where the presence of certain

genetic variants in the native range accelerates adaptation upon introduction (Tsutsui *et al.,* 2000; Schlaepfer *et al.,* 2009; Hufbauer *et al.,* 2011). In sum, although genetic diversity in introduced populations is often viewed as a pre-requisite to adaptation, changes in genetic diversity alone do not explain invasiveness (Uller & Leimu, 2011). Regardless, chance may be just as important as selection in an invasive species' establishment and spread (Gralka & Hallatschek, 2019).

The European starling (*Sturnus vulgaris*) stands out as an exceptionally successful avian colonist and invasive species. In North America, an estimated 200 million starlings currently range from northern Mexico to southern Alaska (Linz *et al.,* 2007). 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, OH (1872), Quebec, Canada (1875), Allegheny, PA (1897), and Springfield, MA (1897), with the second-most successful having been introduced to Portland, OR in 1889 (Forbush, 1915; Kalmbach & Gabrielson, 1921). 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 U.S. persisted without record (Kessel, 1953).

During the starling expansion, ongoing 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 (Burtt & Giltz, 1977). Models of molt origin indicate that starlings in the western U.S. may disperse or migrate shorter distances

(Werner, Fischer, & Hobson, 2020). The models used feathers from the same individuals sampled for this genetic survey, and they initially hypothesized that starlings in the south may be less likely to migrate long distances. Instead, they found that longitude better explains the differences in molt origin among starling populations: starling feathers collected west of -90° longitude were more likely to have originated nearby. In other words, starlings that migrate in eastern North America likely experience greater gene flow among sampling locations (states, in our study), but overall starlings tend to move only regionally and not continent-wide.

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; Cabe, 1999). Here we use robust genomic markers to explore the genomic and demographic patterns of range expansion in North American starlings with three specific aims: (1) to characterize genome-wide levels of diversity and differentiation among starlings; (2) to examine how genetic variation changes across the contiguous United States; (3) to test for a genetic bottleneck; and (4) to test for signatures of selection associated with environmental gradients. We also interpret our results in the context of range-wide data on starling migration and dispersal (Werner, Fischer, & Hobson, 2020), as this movement certainly influences population structure. Recent developments in genotype-environment association methods can identify polygenic traits which are locally adapted to overlapping environmental gradients (Forester *et al.,* 2018; Capblancq *et al.,* 2018), whereas traditional outlier-based methods may not recover evidence of local adaptation in a species that likely has low overall levels of genetic diversity. Especially with these more sensitive methods, the same low genetic diversity

resulting from a founder effect upon colonization could improve our ability to discriminate signatures of selection from low background divergence across the genome (Dlugosch *et al.,* 2015). This work employs modern genomic and analytical tools to examine the evolutionary history of this remarkably successful avian invasion.

METHODS

Sample collection and processing

All starlings sampled were collected during the non-breeding 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'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. USDA Wildlife Services personnel collected birds from 2-5 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 at 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 U.S. Department of Agriculture, National Wildlife Research Center's Institutional Animal Care and Use Committee (QA-2572, S.J. Werner - Study Director; Werner, Fischer, & Hobson, 2020; Table S1).

161 Breast muscle tissue was sampled using biopsy punches (Integra Miltex, York, PA)
162 and frozen in 95% ethanol. Samples were shipped on dry ice, and DNA was extracted using
163 a Qiagen DNeasy kit following the manufacturer's protocol (Qiagen, New York, NY). DNA
164 concentration of each sample was quantified using a Qubit 2.0 fluorometer (Thermo Fisher
165 Scientific, New York, NY). Following the protocol of Peterson *et al.* (2012), we generated a
166 reduced-representation genomic dataset of double-digested, restriction-site associated
167 DNA markers as described in Thrasher *et al.* (2017) using the restriction enzymes SbfI and
168 MspI and adaptors P1 and P2. We sequenced 100bp, single-end reads on an Illumina HiSeq
169 2500. We trimmed and filtered for quality using the FASTX-Toolkit
170 (http://hannonlab.cshl.edu/fastx_toolkit). We then used the process_radtags commands in
171 STACKS v 1.19 (Catchen *et al.,* 2013) to demultiplex the remaining sequences. In subsequent
172 filtering steps, we retained reads only if the following conditions were met: reads passed
173 the Illumina chastity filter, contained an intact *Sbf*I RAD site, contained one of the unique
174 barcodes, and did not contain Illumina indexing adaptors.
175 We assembled sequences to an *S. vulgaris* reference genome (Hofmeister, Rollins *et*
176 *al., in prep*) using the ref-map option in STACKS. Individual reads were mapped to the
177 reference genome using BOWTIE2 version 2.2.8 (Langmead & Salzberg, 2012) using the
178 "very sensitive local" set of alignment pre-sets. The bioinformatics pipeline used for the
179 reference-based assembly has the advantage of using fewer similarity thresholds to build
180 loci. We required that a SNP be present in a minimum of 80% of the individuals (-r 0.8)
181 with a minimum stack depth of 10 reads at a locus within an individual (-m 10) for it to be
182 called. We removed two individuals, one with >50% missing data and one with >50%
183 relatedness (measured using the unadjusted AJK statistic and calculated within *vcftools*),

leaving 158 individuals remaining in the study. We also used *vcftools* –hwe option to remove any SNPs out of Hardy-Weinberg equilibrium (e.g., an exact test that compared expected and observed heterozygosity gave a P-value less than 0.001). About 6% of sequenced variants were out of HWE, and this filter left 14,134 variants remaining. 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 like the RAD loci analyzed here, this strategy of exporting one SNP per stack is often used as an indirect method of controlling for linkage.

(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 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 paired t-test in base R (R Core Team, 2019). 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 ("populations"). To determine significance, we compared observed variation at each hierarchical level to the randomly permuted distance matrices for that particular level using 1000 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 populations.

We tested for isolation by distance (IBD) using a simple Mantel test in *adegenet* (Jombart, 2008): for these data, the assumption of stationarity likely holds, given that North American starlings appear to be in mutation-migration-drift equilibrium (Guillot & Rousset, 2013). Geographic 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). Nevertheless, we report the results of a Mantel test with a Spearman correlation and 10000 permutations. Finally, we test for isolation by environment (IBE) using the R package *vegan* to compare environmental, geographic, and genetic distances (Oksanen *et al.,* 2019). Controlling for geographic distance, we ran partial Mantel correlograms using Euclidean environmental distance matrices built from all bioclimatic variables retained in the *Selection analyses* described in Section 4 below.

(2) Population structure

We first used non-parametric approaches to determine whether starlings 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 non-parametric tests described above. To select the best-supported

K, we used the Evanno method implemented in STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2011). We averaged results across the 10 runs using the greedy algorithm in the program CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007), and visualized results using DISTRUCT v1.1 (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 frequencies (MAF=0.3) and a more relaxed minimum frequency (MAF=0.1) (Linck & Battey, 2017). STRUCTURE results did not differ substantially among MAF thresholds, and we used a filtered dataset with a minimum minor allele frequency of 0.1 in the demographic analyses below.

(3) Demographic modeling

We used a traditional model-based approach (fastsimcoal2) to explicitly test for a bottleneck in North American starlings (Excoffier *et al.,* 2013). We first estimated the folded SFS using a Python script from Simon Martin (available at https://github.com/simonhmartin/genomics_general). We modified the simple 1PopBot20Mb.tpl and .est files provided with the software as follows: the size of the bottleneck (NBOT: 10-1000), start of bottleneck (TBOT: 10-300), ancestral size (ANC: 1000-100000), current size of the population (NCUR: 100-100000), 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; since 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

252 that this method does not explicitly test for a genetic bottleneck, we describe the method
253 and results in the Supplementary Information.
254
255 *(4) Selection analyses*
256 We first identified environmental variation at each sampling location using the R package
257 *raster* (Hijmans & van Etten, 2012), extracting all 19 bioclimatic variables for each set of
258 sampling coordinates from WorldClim 2.0 at a resolution of 5 min on June 16, 2018 (Fick &
259 Hijmans, 2017). We recognize that these environmental conditions represent those
260 experienced during the non-breeding season, and therefore they do not represent the full
261 range of conditions experienced by the starlings sampled; see the Supplementary
262 Information for a discussion of our assumptions about this environmental sampling. We
263 used redundancy analysis (RDA) to examine how loci may covary across multiple
264 environmental gradients (Forester *et al.,* 2018). RDA is especially powerful when testing
265 for weak selection, and detects true positives in large data sets more reliably than other
266 multivariate methods like Random Forest (Forester *et al.,* 2018). Because RDA requires no
267 missing data, we first imputed genotypes where missing sites were assigned the genotype
268 of highest probability across all individuals—a conservative but quick imputation method.
269 Before imputation and after filtering, about 31% of genotypes were missing from the
270 dataset of 15,038 SNPs (not filtered for MAF), and we required that all imputed SNPs were
271 present in at least 80% of individuals. We then used the R package *vegan* to run the RDA
272 model (Oksanen *et al.*, 2019); for a full description of this method, see Forester *et al.,* 2018.
273 Briefly, RDA uses constrained ordination to model a set of predictor variables, and
274 unconstrained ordination axes to model the response (genetic variation). RDA infers

275 selection on a particular locus when it loads heavily onto one or more unconstrained
276 predictor axes, and we retained five variables with relatively low variance inflation factors:
277 BIO1 or Annual Mean Temperature (VIF=3.54), BIO7 or Temperature Annual Range (4.55),
278 BIO12 or Annual Precipitation (8.69), BIO16 or Precipitation of Wettest Quarter (7.91), and
279 elevation (2.19). We tested for significance using the *anova.cca* function within the *vegan*
280 package, and also permuted predictor values across individuals to further check
281 significance of the model. We identify candidate loci as those that are 3 or more standard
282 deviations outside the mean loading. The R scripts for all RDA analyses and figures were
283 written by Brenna Forester (available at https://popgen.nescent.org/2018-03-
284 27_RDA_GEA.html).
285

286 **RESULTS**

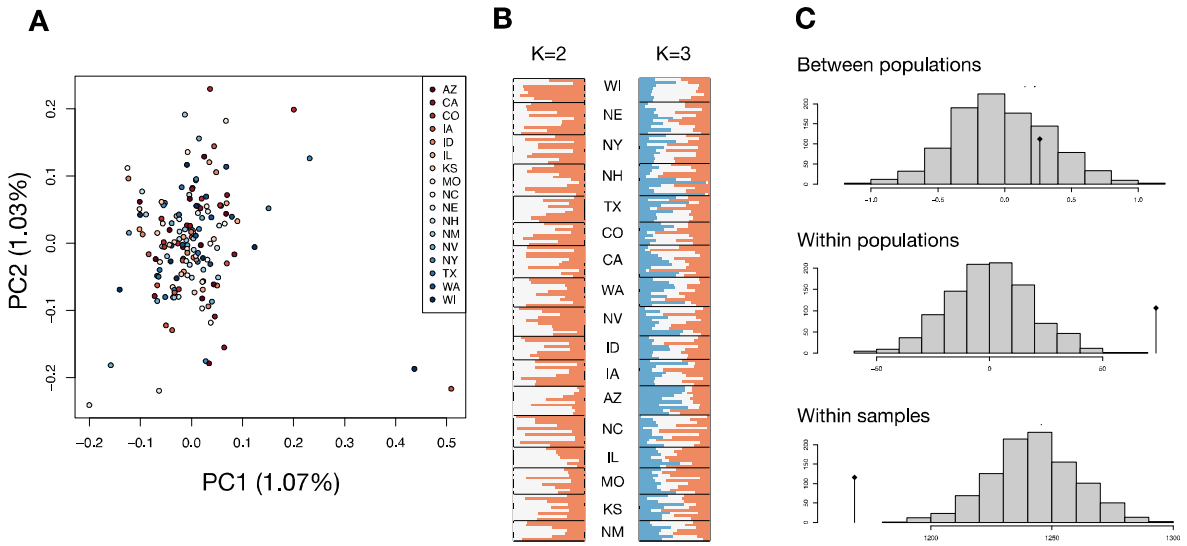
287 *(1) Patterns of genetic diversity and differentiation*

288 We identified 15,038 SNPs at a mean of 27X sequencing depth across 17 sampling
289 locations. Using a dataset filtered by a minor allele frequency of 0.1, we find that genome-
290 wide F_{ST} is extremely low (0.0085), and measures of genetic diversity do not vary
291 substantially among sampling locations (Table 1). Across all of these wintering starling
292 populations, the highest pairwise F_{ST} was 0.0106 between birds from the adjacent states of
293 Arizona and New Mexico. Using a haplotype-based statistic of differentiation, ϕ_{ST} among
294 populations shows an absence of genetic structure (ϕ_{ST} = 0.0002). Hierarchical AMOVAs
295 reveal that 94% of the observed genetic variance is explained by variation within
296 individuals, and the remaining variance reflects differences among individuals in the same
297 population, with negligible variation explained at the between-population level (Figure 1C-

298 D). Across the genome, F_{ST} and nucleotide diversity are exceptionally low (Table 1,2).

299 Genome-wide heterozygosity is moderate at 0.339, and across loci observed heterozygosity

300 differs significantly from expected (t = 66.6, df = 3569, P<0.001, Table 1).



D

	df	Sum of squares	variance components	% variance explained
Between populations	1	1354	0.264	0.021
Within populations	156	205297	73.75	5.936
Within samples	158	184622	1168	94.04
Total	*315*	*391273*	*1242*	

301

302 **Figure 1.** Population structure. A) Principal components analysis on 6,287 SNPs, explaining
303 1.03-1.08% of genetic variation observed. B) STRUCTURE analyses with K=2 and K=3 (best
304 supported). C) Significance testing of hierarchical AMOVAs: the histogram shows expected
305 variance components based on 999 simulations, and the black diamond is the observed
306 variance component. D) AMOVA results.
307
308

309
310 **Table 1.** Population genetic summary statistics. For each sampling location, the table
311 shows the number of private alleles, observed and expected heterozygosity, nucleotide
312 diversity, and the inbreeding coefficient F_{IS}.
313

Pop.	No.	Private	H_{obs}	H_{exp}	π	F_{IS}
NY	10	0	0.34	0.25	0.35	0.037
NH	11	1	0.33	0.26	0.35	0.065
NC	11	1	0.33	0.25	0.35	0.060
IL	7	1	0.33	0.26	0.35	0.056
IA	9	0	0.33	0.24	0.36	0.078
WI	8	0	0.33	0.22	0.35	0.050
MO	9	0	0.33	0.23	0.35	0.053
NE	11	0	0.34	0.22	0.35	0.042
CO	8	1	0.33	0.24	0.35	0.066
KS	9	2	0.35	0.23	0.35	0.012
TX	9	1	0.33	0.24	0.35	0.052
NM	7	2	0.33	0.29	0.36	0.068
AZ	10	1	0.32	0.24	0.35	0.069
NV	10	0	0.33	0.31	0.36	0.068
CA	11	0	0.33	0.25	0.35	0.061
ID	8	0	0.33	0.22	0.35	0.047
WA	10	2	0.32	0.30	0.35	0.082

(2) Population structure

F_{ST} among populations 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 variation among individuals (Figure 1A), and although STRUCTURE identified three populations at the best-supported value of K, these predicted populations do not show obvious differences in ancestry proportions (Figure 1B, 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.

	AZ	CA	CO	IA	ID	IL	KS	MO	NC	NE	NH	NM	NV	NY	TX	WA	WI
Arizona	28	0.0085	0.0090	0.0093	0.0090	0.0082	0.00912	0.012	0.0093	0.0090	0.0125	0.01110	0.010	0.010	0.012	0.010	0.015
California	5	40	0.0080	0.0065	0.0050	0.0050	0.00510	0.0050	0.0075	0.0060	0.0070	0.00510	0.0060	0.0060	0.0070	0.0060	0.00610
Colorado	0	7	43	0.0060	0.0087	0.0082	0.0070	0.0070	0.0060	0.00611	0.0077	0.014	0.0072	0.0060	0.0070	0.0072	0.0090
Iowa	4	16	0	8	0.0060	0.00514	0.0084	0.00714	0.0052	0.0090	0.0076	0.00714	0.0060	0.0062	0.0070	0.0070	0.0086
Idaho	0	0	4	0	23	0.0070	0.0070	0.0060	0.0070	0.0054	0.0070	0.0080	0.0048	0.0060	0.0070	0.00619	0.0070
Illinois	5	2	5	2	0	12	0.0075	0.00617	0.0070	0.0060	0.0067	0.0117	0.0050	0.0060	0.0062	0.0070	0.00812
Kansas	2	12	2	0	0	5	7	0.00512	0.00629	0.0052	0.0065	0.0062	0.0050	0.0062	0.0072	0.0060	0.00610
Missouri	5	5	0	0	0	5	5	11	0.00516	0.0060	0.0065	0.00726	0.0060	0.0060	0.0060	0.0070	0.00711
North Carolina	0	0	0	0	0	0	16	0	32	0.0060	0.0070	0.0070	0.0060	0.0060	0.00744	0.0080	0.0064
Nebraska	0	8	24	0	16	0	0	0	0	12	0.00812	0.0064	0.0068	0.0070	0.0070	0.0070	0.0060
New Hampshire	28	6	0	3	0	0	3	3	0	6	9	0.0076	0.0073	0.0060	0.0070	0.0070	0.00713
New Mexico	11	0	0	0	0	0	0	0	11	11	0	67	0.0060	0.0080	0.0070	0.0070	0.0070
Nevada	0	3	4	1	18	0	1	0	0	11	3	1	14	0.0050	0.0070	0.00614	0.0061
New York	3	10	0	1	0	15	1	10	0	1	6	1	1	1	0.0070	0.0060	0.00521
Texas	5	0	0	0	0	0	0	3	26	0	0	2	0	0	65	0.0070	0.0080
Washington	0	0	0	0	0	0	0	0	0	7	0	0	60	0	0	7	0.0060
Wisconsin	10	0	2	2	2	8	6	4	0	2	2	0	0	10	0	0	31

*percent assigned to each state does not sum to 100 because Minnesota, Oregon, and Vermont are not included in the genetic dataset.

Table 2. Differentiation and movement among sampling locations. This table shows the percentage of birds assigned to each sampling location from every origin. Note that these values are directional: although 12 starlings that were collected in AZ originated in KS, only 2 starlings collected in KS originated in AZ. The black diagonal indicates the percentage of birds assigned to that collection state according to a discriminant function analysis of molt-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 gray colors indicate higher F_{ST}.

Starling populations follow a clear pattern of isolation-by-distance (Mantel r = 0.139, $P < 0.0001$). Spatially explicit models of isolation-by-distance suggest a fairly uniform rate of migration range-wide, where local increases in migration rate are likely a model artifact (Supplementary Information). However, models of isolation by environment (IBE) show that the relationship between environmental and genetic distances is stronger than the simple geographic-genetic distance model. After controlling for geographic distance, we find that all bioclimatic variables tested show non-zero correlations between

environmental and genetic distances (Supplementary Information). There is a strong positive relationship between genetic distance and both precipitation in the wettest quarter (BIO16, Mantel r = 0.282, P = 0.001) and elevation (Mantel r = 0.146, P = 0.001).

(3) Demographic modeling

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 pre-bottleneck size (Supplementary Information). In fact, every run of the model finds that starlings' current effective population size (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 effect, which may contribute to the overall lack of evidence for inbreeding. We do detect very low levels of inbreeding within some populations (Table 1, highest F_{IS} = 0.082 in Washington). Taken together, these models do not suggest a classical founder effect, whereby effective population size remains very 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, showing that 178 variants are correlated with environmental differences among populations (F = 1.022, P = 0.002, Figure 2). Populations living in warmer climates tend to cluster more

closely in the left quadrant and high elevation populations cluster in the middle right quadrant. However, populations do not cluster based on geographic distance: for example, starlings from TX and WA cluster closely due to shared genetic variants, even though the two populations 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), annual precipitation (BIO12), precipitation in the wettest quarter (BIO16) and elevation (Figure 2). Genes under selection tend to have lower allele frequencies, and function in several biological processes (Table S2).

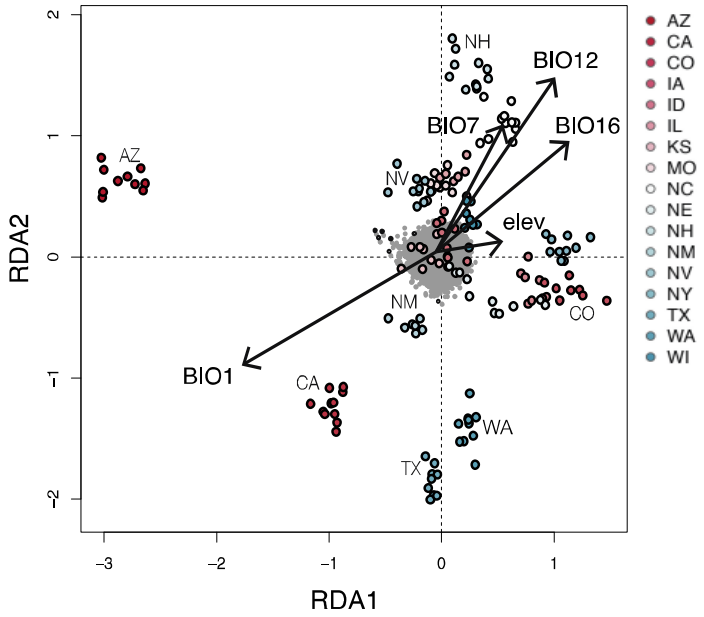


Figure 2. Evidence for incipient local adaptation. Redundancy analyses indicate that 191 SNPs (small gray points) are associated with bioclimatic predictors (vectors). BIO1: mean annual temperature; BIO7: temperature annual range; BIO12: annual precipitation; BIO16: precipitation of wettest quarter.

DISCUSSION

Our whole-genome data reveal that invasive starling populations show moderate genetic diversity across North America, rapid population growth after a genetic bottleneck, negligible geographically structured genetic differentiation, significant but quite weak isolation-by-distance, and intriguing associations of genetic variation with environmental parameters that might result from adaptive processes. Genetic diversity is typically predicted to be low in invasive populations (Dlugosch & Parker, 2008), and although RADseq tends to estimate lower genetic diversity than the true diversity, this bias is minimal for taxa known to be genetically depauperate (Cabe, 1998; Cariou *et al.,* 2016). Within invasion genetics, one area of recent focus involves the spatial dynamics of invasion, and we can explore how genetic variation varies across the post-invasion range of starlings in North America. Although there is a significant (but low magnitude) signal of isolation-by-distance, hierarchical AMOVAs find that variation within and among individuals explains observed differentiation better than variation among populations. There is no evidence of population structure, and while models indicate some subtle spatial patterns of genetic variation, these model-based inferences likely reflect sampling artifacts (Supplementary Information). 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 extremely low F_{ST} among populations (Table 2)—maintains high connectivity across North American starling populations. When interpreted within the context of a complementary exploration of movements inferred from feather isotope assays, this genomic survey confirms that dispersal and migration continue to influence the genetic variation generated in just over a century of range expansion.

Long-distance dispersal may be common in introduced starlings on multiple continents, where rates of dispersal are determined by demographic changes and environmental quality. In South Africa, invasive starlings disperse when their natal environment becomes crowded or unsuitable, as at the leading edge of their range expansion (Hui *et al.,* 2012). 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. This explosive population growth from hundreds to millions of individuals may have encouraged long-distance dispersal away from the dense populations in eastern North America. In general, multiple lines of evidence have shown that the dispersal rates and distances of juvenile starlings in North America are remarkably high (Dolbeer, 1982; Cabe, 1999; Werner, Fischer, & Hobson, 2020). Considered in concert with similar findings in the invasive South African and Australian starling populations (Hui *et al.,* 2012; Phair *et al.,* 2017), it is clear that dispersal similarly plays an ongoing and major role in the connectivity of North American starling populations.

Although it is generally important, dispersal has not been uniform across time or space in North American starlings. One open question in invasion genetics is whether expanding populations maximize spatiotemporal fitness via spatial sorting and incipient adaptation (Williams et al. 2019). Empirical studies of some invasions suggest that spatial sorting is common, including in 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).

429 However, we do know that adaptive dispersal strategies can facilitate range expansion in
430 Western bluebirds (Duckworth, 2008) and invasive beetles (Lombaert *et al.,* 2014; Ochocki
431 & Miller, 2017) among other species. For North American starlings, we can combine
432 historical records, isotopic and genetic evidence to make inferences about the possibility of
433 spatial sorting.

434 Feather isotope data from the same starlings sampled in this genetic survey
435 suggests that movements are not uniform continent-wide, and that both juvenile dispersal
436 and regional differences in migration could influence genetic variation (Werner, Fischer, &
437 Hobson, 2020). If starling movement was the primary driver of differentiation, then we
438 would expect F_{ST} among populations to be directly related to the number of birds collected
439 and assigned to that population. Starlings in the western U.S. appear to have differentiated
440 subtly from their eastern counterparts based on the higher F_{ST} between Arizona—and to
441 some degree, New Mexico and Colorado—and all other sampling locations (Table 2). Birds
442 collected in these southwestern states are also assigned to those states by discriminant
443 function analysis (Werner, Fischer, & Hobson, 2020); for example, 67% of starlings
444 collected in New Mexico were also assigned to that state (Table 2). We suggest that birds
445 assigned to the same state where they were collected may reside in that state year-round,
446 but we note that collecting feathers once in the lifespan of the bird does not allow us to
447 determine the bird's lifelong migration and dispersal. The highest rates of pairwise
448 differentiation occur between states where one of those "populations" shows a relatively
449 high collection-state assignment, which suggests that starling movement among sampling
450 locations does impact the genetic variation. This comparison is further supported by the
451 low but significant isolation-by-distance across the North American range; however,

452 environmental pressures explain genetic variation better than a simple model that
453 accounts for geographic distance. In addition, assignment to a state other than the
454 collection state seems to have only a minor effect on F_{ST}, even though we might expect that
455 movement among geographically distant sampling locations might lead to gene flow among
456 those locations. In Arizona—where F_{ST} is highest but still relatively weak—Werner et al.'s
457 model suggests that 25% of birds originated in New Hampshire, which is approximately
458 the percentage of birds assigned to Arizona (28%, Table 2). Unsurprisingly, geography,
459 environment, and demography all appear to influence genetic variation within North
460 American starling, and explicit tests for selection can help to clarify the relative weight of
461 each factor.

462 This species' general genome-wide panmixia across the continent allows for tests of
463 selection on loci that may be involved in local adaptation. Because genetic diversity and
464 differentiation are exceptionally low, it is relatively easy to identify sites that have
465 differentiated against this background of low genome-wide or population-specific
466 divergence (Dlugosch *et al.,* 2015). We find that almost 200 of the 15,038 RAD loci appear
467 to be under selection using a redundancy analysis. Only 13 of these SNPs overlap with the
468 SNPs identified by a latent-factor mixed model, and there is no overlap between the RDA
469 candidates and the differentiation-based scans (BayeScEnv and Bayescan). It is
470 unsurprising that each test identifies different candidates for selection, because the
471 assumptions underlying each are very different (for more details on our approach to
472 selection testing, see Supplementary Information). Rather than making inferences based on
473 the genes identified by these scans, we instead propose that genotype-environment
474 associations show that changes in precipitation and temperature can explain genetic

variation in North American starlings. However, we do suggest that rapid evolution in genes underlying key physiological processes may have supported the starling's spread across North America (Supplemental Information). Specifically, we hypothesize that aridity and cold 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. While this finding suggests that local adaptation may explain genetic variation within the North American starling invasion, there are several relevant caveats to this approach.

Because there is evidence of a genetic bottleneck in the North American starling invasion, it is possible that artifacts resulting from genetic drift could mimic these patterns of local adaptation. We cannot rule out allele surfing during range expansion as an explanation based on the methods currently available for testing genotype-environment associations in a young and expanding system like this one. However, given how short the timescale of divergence is in North American starlings, such drift effects are more likely if genetic variation is (1) ancestral and (2) structured among populations. Gene flow among starling populations is remarkably high, but we do not know if variants putatively under selection were present in the ancestral population. These alleles under putative selection do not approach fixation, as no putative outlier has an allele frequency greater than 0.28. We note that concordance among environmental and genetic distances (e.g., partial Mantel tests) indicate that spatial autocorrelation complicates our selection inferences. Under these conditions, any evidence for selection is likely to be weak, and as always these selection scans can generate false positives. However, 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). Finally, we do not explicitly control for linkage, and some allele frequency shifts could be explained by recombination or by genetic hitchhiking.

Our study focuses on birds collected during the winter, which may limit our inferences about population structure and selection to these wintering populations. As discussed above, isotopic evidence suggests that starlings in the western U.S. tend to move only regionally whereas birds sampled in the eastern U.S. undertake longer movements (Table 2). This in turn suggests that starlings overwintering in the western U.S. are more likely to breed nearby, and thus the environmental conditions may not change as dramatically among wintering and breeding ranges. In addition, 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 2, Supplementary Information). Western populations experience these environmental conditions year-round, which could allow selection to drive advantageous alleles toward fixation. Elsewhere in the U.S., 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. This sampling strategy uses the more vagile eastern populations as a comparison to more-resident western populations that may also be under stronger selection. This

521 framing may suggest that population structure in the western U.S.—of which we find no
522 evidence—could explain allele frequency shifts that we infer to be selection, but when we
523 compare the relative importance of geographic and environmental distances in partial
524 Mantel tests, we find that environmental conditions better explain genetic variation. This
525 evidence supports our interpretation of selection as a major driver of genetic variation in
526 North American starlings.

527 A similar project on starlings in the Australian invasion—which colonized that
528 continent nearly concurrently with the North American invasion—found that geographic
529 but not environmental distance explains genetic patterns there (Cardilini *et al.,* 2020).
530 Starlings in the Australian range show substantial population structuring and significant
531 patterns of isolation-by-distance. Earlier work had shown that gene flow among Australian
532 starling populations is low (Rollins *et al.,* 2009), and phylogeographic patterns of
533 mitochondrial sequence variation confirm that starlings on the edge of the expansion front
534 in Western Australia have differentiated from those still living in the introduction site
535 (Rollins *et al.,* 2011). In fact, starlings at the expansion front may have rapidly adapted
536 during the Australian invasion (Rollins *et al.,* 2016): the proportion of adult starlings in
537 Western Australia carrying a novel mitochondrial haplotype has increased rapidly only at
538 this range edge. A genotyping-by-sequencing survey employing a much greater number of
539 SNP markers indicates three population subdivisions in Australia, where geographic
540 distance explains genetic differentiation in starlings better than does environmental
541 variation (Cardilini *et al.*, 2020). Global F_{ST} across all Australian populations is an order of
542 magnitude higher than the equivalent F_{ST} index across North America, despite similar areas
543 sampled. In Australia, the strong evidence for isolation-by-distance and founder effects

544 complicate attempts to disentangle selection from drift, yet despite their differences in
545 invasion dynamics, genotype-environment associations reveal signatures of selection in
546 both invasions. On both continents, starling genetic variation can be explained by extremes
547 in temperature and precipitation, and preliminary results of whole-genome resequencing
548 of native and introduced populations confirm that variability in temperature and
549 precipitation may shape observed genetic variation in starlings world-wide (Hofmeister et
550 al., *in prep*).

551 Our results contribute to the growing evidence of rapid adaptation in some
552 expanding populations, even in extremely young systems. Some studies of rapidly
553 expanding invasions find little evidence that adaptation may facilitate this expansion, as in
554 corals (Leydet et al. 2018). However, other work suggests a role for selection in supporting
555 rapid range expansion, such as in experimental studies of flour beetles (Szucs *et al.,* 2017)
556 and empirical work in guppies (Baltazar-Soares *et al.,* 2019). Invasion biologists have long
557 highlighted propagule pressure as a driver of invasion success, but the genetic composition
558 may be just as important as the size of the establishing population (Briski *et al.,* 2018). For
559 example, genetic bottlenecks in monk parakeets, another avian invader now distributed
560 world-wide, do not seem to inhibit invasion success (Edelaar *et al.,* 2015). Pre-adaptation
561 in the native range or selection during transport may facilitate the spread of invasive
562 species, and human commensalism may support establishment and spread, as shown in
563 house sparrows (Ravinet *et al.,* 2018) and common mynas (Cohen *et al.,* 2019), and
564 reviewed across alien bird species (Cardador & Blackburn, 2019). Empirical studies of
565 invaders like the ones described here also show how, in addition to genetic variation,
566 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 (Schweizer *et al.,* 2015; Bay *et al.,* 2018). Finally, we suggest that our study adds to those suggesting that rapid local adaptation can evolve even in dispersive and young populations.

AUTHOR CONTRIBUTIONS

NRH, IJL designed research, SJW contributed samples, NRH analyzed data, NRH, IJL wrote paper.

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 Stepfanie Aguillon provided valuable advice on bioinformatic methods.

FUNDING

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.

DATA ACCESSIBILITY 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 (doi: 10.5061/dryad.j424f07), and raw sequencing data archived in NCBI SRA (Bioproject PRJNA545151).

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, 403–410.

Ashenden T, Rooke AC, Fox MG (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 *et al.* (2017) A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nature Communications*, 1–9.

Baltazar-Soares M, Blanchet S, Cote J, Tarkan AS, Záhorská, Gozlan RE, 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 RA, Harrigan RJ, Le Underwood V *et al.* (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, **359**, 83–86.

Berthouly-Salazar C, van Rensburg BJ, Le Roux JJ, van Vuuren BJ, Hui C (2012) Spatial sorting drives morphological variation in the invasive bird, *Acridotheris tristis*. *PLoS ONE*, **7**, e38145.

Bitton P-P, Graham BA (2014) Change in wing morphology of the European starling during and after colonization of North America. *Journal of Zoology*, **295**, 254–260.

Briski E, Chan FT, Darling JA, Lauringson V, MacIsaac HJ, Zhan A, Bailey SA (2018) Beyond propagule pressure: importance of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment*, **16**, 345-353.

Brown GP, Phillips BL, 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*, **281**, 20141385.

Burtt HE, Giltz ML (1977) Seasonal directional patterns of movements and migrations of Starlings and Blackbirds in North America. *Bird-Banding*, **48**, 259-271.

Cabe PR (1998) The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity*, **80**, 519–525.

Cabe PR (1999) Dispersal and population structure in the European Starling. *The Condor*, **101**, 451–454.

Capblancq T, Luu K, Blum MGB, Bazin E (2018) Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, **185**, 587–11.

Cardador L, Blackburn TM (2019) Human-habitat associations in the native distrbutions of alien bird species. *J Applied Ecology*, **56**, 1189-1199.

Cardilini APA, Stuart KC, Cassey P, Richardson MF, Sherwin W, Rollins LA, Sherman CDH (2020) Signatures of selection in a recent invasion reveals adaptive divergence in a highly vagile invasive species. bioRxiv, doi: https://doi.org/10.1101/643569.

Cariou M, Duret L, Charlat S (2016) How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology*, 1–8.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Cohen, TM, 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 RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, **24**, 1999–2017.

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

de Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, **6**, 1248–1258.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, **24**, 2095–2111.

Dolbeer RA (1982) Migration patterns for age and sex classes of Blackbirds and Starlings. *Journal of Field Ornithology*, **53**, 28-46.

Duckworth RA (2008) Adaptive Dispersal Strategies and the Dynamics of a Range Expansion. *The American Naturalist*, **172**, S4–S17.

Earl DA, vonHoldt BM (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 DI (2012) Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology & Evolution*, **27**, 659–665.

Edelaar P, Roques S, Hobson EA, Da Silva AG, Avery ML, Russello MA, Senar JC, Wright TF, Carrete M, Tella JL (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**, 2164-2716.

Estoup A, Ravigné V, Hufbauer R *et al.* (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 VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLOS Genetics,* **9**, e1003905.

Facon B, Hufbauer RA, Tayeh A *et al.* (2011) Inbreeding Depression Is Purged in the Invasive Insect Harmonia axyridis. *Current Biology*, **21**, 424–427.

Fick SE, Hijmans RJ (2017) *WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas*. *International Journal of Climatology*.

Forbush EH (1915) The starling. Boston, Wright & Potter Printing Co.

Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, **27**, 2215–2233.

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Garant D, Forde SE, Hendry AP (2007) The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, **21**, 434–443.

Gralka M, Hallatschek O (2019) Environmental heterogeneity can tip the population genetics of range expansions. *eLife*, **8**, e44359.

Guillot G, Rousset F (2013) Dismantling the Mantel tests (L Harmon, Ed,). *Methods in Ecology and Evolution*, **4**, 336–344.

Günther T, Coop GC (2013) Robust identification of local adaptation from allele frequencies. *Genetics Soc America*, **195**, 205-220.

Hijmans RJ, van Etten J (2012) raster: Geographic analysis and modeling with raster data. R package version 2.0-12.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 379–397.

Hufbauer RA, Facon B, Ravigné V *et al.* (2011) 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 RA, Evans KL (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, Chaine AS *et al.* (2017) Gene flow favours local adaptation under habitat choice in ciliate microcosms. *Nature Ecology & Evolution*, 1–4.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jernelov A (2017) Starlings in North America. In: *The Long-Term Fate of Invasive Species*, pp. 137–149. Jarpas.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Kalmbach, ER, and IN Gabrielson (1921) Economic Value of the Starling in the United States *USDA Bulletin*, **868**, 1-59.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281–14.

Kessel B (1953) Distribution and migration of the European Starling in North America. *The Condor*, **55**, 49–67.

Koenig W (2003) European Starlings and Their Effect on Native Cavity-Nesting Birds. *Conservation Biology*, **17**, 1134–1140.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Lenormand T (2002) Gene flow and the limits to natural selection. **17**, 183–189.

Lescak EA, Bassham SL, Catchen J *et al.* (2015) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, **112**, E7204–E7212.

Leydet KP, Grupstra CGB, Coma R, Ribes M, Hellberg ME (2018) Host-targeted RAD-Seq reveals genetic changes in the coral Oculina patagonica associated with range expansion along the Spanish Mediterranean coast. *Mol Ecol*, **27**, 2529-2543.

Linck EB, Battey CJ (2017) Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *bioRxiv*.

Linz GM, Homan HJ, Gaulker SM, Penry LB, Bleier WJ (2007) *European starlings: a review of an invasive species with far-reaching impacts*.

Liebl AL, Schrey AW, Richards, CL, Martin LB (2013) Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integ Comp Bio*, **53**, 351-358.

Liu X, Fu Y-X (2015) Exploring population size changes using SNP frequency spectra. *Nature Genetics*, **47**, 555–559.

Lombaert E, Estoup A, Facon B *et al.* (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 CK, Vieira C (2019) Biological invasion: The influence of the hidden side of the (epi)genome. Functional Ecology, 34, 385-400.

Malinsky M, Trucchi E, Lawson D, Falush D (2016) RADpainter and fineRADstructure: population inference from RADseq data. 1–6.

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, **24**, 3223-3231.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, **28**, 659–669.

Mi H, Dong Q, Muruganujan A *et al.* (2009) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Research*, **38**, D204–D210.

Ochocki BM, Miller TEX (2017) Rapid evolution of dispersal ability makes biological invasions faster and more variable. *Nature Communications*, **8**, 1–8.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (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**, e37135–11.

Petkova D, Novembre J, Stephens M (2015) Visualizing spatial population structure with estimated effective migration surfaces. *Nature Reviews Neuroscience*, **48**, 94–100.

Pfeifer SP, Laurent S, Sousa VC *et al.* (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 DJ, Le Roux J, Berthouly-Salazar C et al (2018) Context-dependent spatial sorting of dispersal-related traits in the invasive starlings (Sturnus vulgaris) of South Africa and Australia. *bioRxiv,* 342451.

Phillips BL, Perkins TA (2019) Spatial sorting as the spatial analogue of natural selection. *Theoretical Ecology*, **12**, 155-163.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945–959.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ravinet M, Elgin TO, Trier C, Aliabadian M, Gavrilov A, Sætre, G-P (2018) Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B*, **285**, 20181246.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.

Rollins LA, Richardson MF, Shine R (2015) A genetic perspective on rapid evolution in cane toads (Rhinella marina). *Molecular Ecology*, **24**, 2264–2276.

Rollins LA, Woolnough AP, Fanson BG *et al.* (2016) Selection on Mitochondrial Variants Occurs between and within Individuals in an Expanding Invasion. *Molecular Biology and Evolution*, **33**, 995–1007.

Rollins LA, Woolnough AP, Sinclair R, Mooney NJ, Sherwin WB (2011) Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology*, **20**, 2307–2317.

Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB (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 NA (2003) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Schlaepfer DR, 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 PC, Seeman OD, Allen GR (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. *Biol Rev*, **92**, 939-952.

Schweizer RM, Robinson J, Harrigan R *et al.* (2015) Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology*, **25**, 357–379.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Sheldon EL, Schrey A, Andrew SC, Ragsdale A, Griffith SC (2018) Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. *R. Soc. Open Sci*, **5**, 172185.

Szűcs M, Vahsen ML, Melbourne BA *et al.* (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.

Thomas PD, Campbell MJ, Kejariwal A *et al.* (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, **13**, 2129–2141.

Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ (2017) Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. 1–34.

Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*, **25**, 2144–2164.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences*, **97**, 5948–5953.

Uller T, Leimu R (2011) Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology*, **17**, 3478–3485.

Werner SJ, Fischer JW, Hobson KA (in press) Multi-isotopic (δ^2H, δ^{13}C, δ^{15}N) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. *PLoS One*.

Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018) Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, **17**, 675.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.

Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.