1   Environment but not geography explains genetic variation in the invasive and largely

2   panmictic European starling in North America

3

4   Running title: Genotype-environment associations in starlings

5

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

7

8   [1] Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road,

9   Ithaca, NY 14853

10   [2] Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, 159

11   Sapsucker Woods Road, Ithaca, NY 14850

12   [3] United States Department of Agriculture, Animal and Plant Health Inspection Service,

13   Wildlife Services, National Wildlife Research Center, 4101 LaPorte Avenue, Fort Collins, CO

14   80521

15   *corresponding author: nrh44@cornell.edu

16

17   **ABSTRACT**

18   Populations of invasive species that colonize and spread in novel environments may

19   differentiate both through demographic processes and local selection throughout the

20   genome. European starlings (*Sturnus vulgaris*) were introduced to New York in 1890 and

21   subsequently spread throughout North America, becoming one of the most widespread and

22   numerous bird species on the continent. Genome-wide comparisons across starling

23   individuals and populations can identify demographic and/or selective factors that

24   facilitated this rapid and successful expansion. We investigated patterns of genomic

25   diversity and differentiation using reduced-representation genome sequencing

26   (ddRADseq) of 17 starling populations. Consistent with this species' high dispersal rates

27   and rapid expansion history, we found low genome-wide differentiation and few F_{ST}

28   outliers even at a continental scale. Despite starting from a founding population of

29   approximately 180 individuals, North American starlings do not seem to have undergone a

30   detectable genetic bottleneck: they have maintained an extremely large effective

31   population size since introduction. We find more than 200 variants that correlate with

32   temperature and/or precipitation. Genotype-environment associations (but not outlier

33   scans) identify these SNPs against a background of negligible genome- and range-wide

34   divergence. Such variants fall in the coding regions of genes associated with metabolism,

35   stress, and neurological function. This evidence for incipient local adaptation in North

36   American starlings suggests that it can evolve rapidly even in wide-ranging and

37   evolutionarily young populations. This survey of genomic signatures of expansion in North

38   American starlings is the most comprehensive to date and complements ongoing studies of

39   world-wide local adaptation in these highly dispersive and invasive birds.

40

41   Keywords: panmixia, gene flow, local adaptation, invasion genetics

42

43   **INTRODUCTION**

44   Studies of local adaptation have long bridged the interface between ecological and

45   evolutionary questions by exploring how populations adapt to differing environmental

46   conditions. Traditionally, high degrees of local adaptation were expected to be present only

47   in fairly isolated populations—those free from the homogenizing effects of high gene

48   flow—with a long history in those locations, providing the time thought to be necessary for

49   local adaptation to evolve (Lenormand 2002). In contrast, we now know that local

50   adaptation occurs frequently even in systems with high gene flow (Yeaman & Whitlock

51   2011; Tigano & Friesen 2016) and often rapidly after colonization of a novel environment

52   (Prentis et al. 2008). We continue to find evidence for rapid local adaptation in systems as

53   divergent as cane toads (Rollins et al. 2015), sticklebacks (Lescak et al. 2015), honeybees

54   (Avalos et al. 2017), steelhead trout (Willoughby *et al.* 2018), deer mice (Pfeifer *et al.*

55   2018), and many more. These studies show that many taxa can adapt rapidly to local

56   conditions in response to the new selection regimes they encounter as they expand their

57   range. Invasive species that have recently expanded into new locations provide tractable

58   opportunities to investigate local adaptation as it originates (Colautti & Lau 2015).

59   After successful colonization of a new habitat, many invasive species show a

60   demographic boom that is likely facilitated by their ecological release in the new

61   environment. Theory predicts that this rapid population growth will plateau as the

62   population approaches carrying capacity, but successful invasive species may continue to

63   expand their range and thus maintain a high rate of overall population growth. When

64   population density increases and demographic rates change, introduced species may

65   rapidly evolve traits that enable them to spread (Szűcs et al. 2017). Increased dispersal

66   may evolve in concert with the demographic boom, enabling populations to grow and

67   spread as individuals disperse. This dispersal can result in gene flow that counteracts

68   inbreeding depression and increases adaptive potential (Garant *et al.* 2007; Rius & Darling

69   2014). If particular traits enable individuals to disperse more easily to their preferred

70   habitat, gene flow may be directional and even adaptive (Edelaar & Bolnick 2012; Jacob *et*

71   *al.* 2017). Dispersal with habitat choice has thus facilitated range expansion in Western

72   bluebirds (Duckworth 2008) and invasive beetles (Lombaert *et al.* 2014; Ochocki & Miller

73   2017) among other species. For example, invasive European starlings in South Africa

74   disperse more frequently at the leading edge of their range expansion, where rates of

75   dispersal are determined by demographic changes and environmental quality (Hui et al.

76   2012). Invasions thus allow us to observe interactions between demography and the early

77   processes of selection (Dlugosch *et al.* 2015) as populations experience new environments.

78   Importantly, these eco-evolutionary interactions depend in part on the genetic variation in

79   the established population.

80   Recent work in invasion genetics aims to tease apart how selection and demography

81   might resolve a paradox of invasion (Estoup et al. 2016): many invasive species experience

82   genetic bottlenecks as a result of an initial founder effect, but often thrive and spread

83   despite this loss of standing genetic diversity. Theory predicts that introductions will

84   typically result in an initial contraction in population size and/or genetic diversity

85   (Dlugosch & Parker 2008). However, bottlenecks of genetic variation clearly do not limit

86   the success of many invasive species (Schmid-Hempel *et al.* 2007; Dlugosch & Parker 2008;

87   Facon *et al.* 2011). Invaders might adapt through soft sweeps that reduce genetic diversity

88   while selecting for adaptive variants from standing or novel genetic variation, which is

89   especially likely in the case of ecological adaptation (Messer & Petrov 2013). Furthermore,

90   some invasions may increase rather than reduce genetic diversity, as when multiple

91   invasions from different source populations introduce previously isolated alleles and

92   thereby facilitate admixture (Dlugosch & Parker 2008). The new conditions can also select

93   among standing variation, where the presence of certain genetic variants in the native

94   range accelerates adaptation upon introduction (Tsutsui *et al.* 2000; Schlaepfer *et al.* 2009;

95   Hufbauer *et al.* 2011). In sum, although genetic diversity in introduced populations is often

96   viewed as a pre-requisite to adaptation, changes in genetic diversity alone do not explain

97   invasiveness (Uller & Leimu 2011).

98   The European starling (*Sturnus vulgaris*) stands out as an exceptionally successful

99   avian colonist and invasive species. In North America, an estimated 200 million starlings

100   currently range from northern Mexico to southern Alaska. Introduced to New York City in

101   1890, starlings nearly covered the continent within a few generations by expanding up to

102   91 km each year (Bitton & Graham 2014). The starling population has grown steadily and

103   earned its reputation as a costly agricultural pest, as flocks often overwinter in feed lots or

104   swarm fruit orchards (Linz et al. 2007). Starlings also may compete for nest sites with

105   American Kestrels (*Falco sparverius*), Eastern Bluebirds (*Sialia sialis*), and other native

106   cavity-nesting birds, though this competition may not have substantial demographic

107   consequences for most native birds (Koenig 2003). Despite the starling's prominence as the

108   most successful avian invasive species in North America, not much is known about the

109   genetic consequences of its invasion history.

110   Genetic studies of other starling invasions suggest that North American starlings

111   could have differentiated or adapted even after a few generations. Previous genetic work in

112   North America indicated near-random mating at a continental scale, with large demes and

113   high dispersal rates (Cabe 1998; 1999). However, these studies relied on a handful of

114   microsatellite markers that sample a far smaller fraction of the genome than do current

115   genomic tools. Additional studies in the invasive Australian population, which colonized

116   that continent nearly concurrently with the North American invasion, used multiple genetic

117   approaches. Microsatellite evidence suggests that gene flow among Australian starling

118   populations is low (Rollins *et al.* 2009), and phylogeographic patterns of mitochondrial

119   sequence variation confirm that starlings on the edge of the expansion front in Western

120   Australia have differentiated from those still living in the introduction site (Rollins *et al.*

121   2011). In fact, starlings at the expansion front may have rapidly adapted during the

122   Australian invasion: the proportion of adult starlings in Western Australia carrying a novel

123   mitochondrial haplotype has increased rapidly only at this range edge (Rollins *et al.* 2016).

124   A genotyping-by-sequencing survey employing a much greater number of SNP markers

125   indicates three populations in Australia, and geographic distance explains genetic

126   differentiation in starlings better than does environmental variation in the Australian

127   invasion (Cardilini 2016). Studies of the Australian invasion suggest incipient ecological

128   specialization at the range edge; thus, our study of the North American starling invasion

129   begins by examining the relationship between genetics and geography prior to testing for

130   environmental associations (Sexton *et al.* 2014).

131   Here we explore the genomic and demographic patterns of range expansion in

132   North American starlings with three specific aims: (1) to characterize genome-wide levels

133   of diversity and differentiation among starlings; (2) to examine how genetic variation

134   changes across the North American range; and (3) to test for signatures of selection

135   associated with environmental gradients. Although traditional outlier-based methods may

136   not recover evidence of local adaptation in a species that likely has low overall levels of

137   genetic diversity, recent developments in genotype-environment association methods can

138   identify polygenic traits which are locally adapted to overlapping environmental gradients

139   (Forester *et al.* 2018; Capblancq *et al.* 2018). Especially with these more sensitive methods,

140   the same low genetic diversity resulting from a founder effect could improve our ability to

141   discriminate signatures of selection from low background divergence across the genome

142   (Dlugosch *et al.* 2015). This study thereby leverages modern genomic and analytical tools

143   to examine the evolutionary history of this infamous avian invasion.

144

145   **METHODS**

146   *Sample collection and processing*

147   Starlings (*N* = 266) were collected in January-March 2016 and 2017 from 26 dairies and

148   feedlots by the U.S. Department of Agriculture's Wildlife Services personnel in Arizona,

149   California, Colorado, Idaho, Illinois, Iowa, Kansas, Missouri, Nebraska, Nevada, New

150   Hampshire, New Mexico, New York, North Carolina, Texas, Washington, and Wisconsin.

151   Whole birds were euthanized and stored at 0°C until tissue sampling. Breast muscle tissue

152   was sampled using biopsy punches (Integra Miltex, York, PA) and frozen in 95% ethanol.

153   Samples were shipped on dry ice, and DNA was extracted using a Qiagen DNeasy kit

154   following the manufacturer protocol (Qiagen, New York, NY). DNA concentration of each

155   sample was quantified using a Qubit fluorometer (Life Technologies, New York, NY). The

156   collection and use of starlings for this and related studies were approved by the U.S. Department

157   of Agriculture, National Wildlife Research Center's Institutional Animal Care and Use

158   Committee (QA-2572, S.J. Werner- Study Director; Werner et al., *in prep*; Supplementary

159   Table 1).

160   Following the protocol of Peterson *et al.* 2012, we generated a reduced-

161   representation genomic dataset of double-digested, restriction-site associated DNA

162   markers as described in Thrasher et al. 2017 using the restriction enzymes SbfI and MspI

163   and adaptors P1 and P2. We trimmed and filtered for quality using the FASTX-Toolkit

164   (http://hannonlab.cshl.edu/fastx_toolkit). We then used the process_radtags commands in

165   STACKS v 1.19 (Catchen et al. 2013) to demultiplex the remaining sequences. In subsequent

166   filtering steps, we retained reads only if the following conditions were met: reads passed

167   the Illumina chastity filter, contained an intact *Sbf*I RAD site, contained one of the unique

168   barcodes, and did not contain Illumina indexing adaptors. We also removed two individuals

169   with >50% missing data and >50% relatedness (measured using the unadjusted AJK

170   statistic and calculated within *vcftools*), leaving 158 individuals remaining in the sample.

171   We assembled sequences to an *S. vulgaris* reference genome (Hofmeister, Rollins et

172   al., *in prep*) using the ref-map option in STACKS. Individual reads were mapped to the

173   reference genome using BOWTIE2 version 2.2.8 (Langmead & Salzberg 2012) using the "very

174   sensitive local" set of alignment pre-sets. The bioinformatics pipeline used for the

175   reference-based assembly has the advantage of using less similarity thresholds to build

176   loci. We required that a SNP be present in a minimum of 80% of the individuals with a

177   minimum stack depth of 10 for individuals at a locus for it to be called. For analyses such as

178   BayeScan we used all SNPs in a given stack, but for STRUCTURE and other analyses

179   sensitive to linkage disequilibrium, we used only the first SNP in each stack.

180

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

182   We estimated per-locus measures of genetic diversity and genome-wide

183   differentiation using the *populations* option in Stacks (Catchen et al. 2013). We used

184   *vcftools* to calculate FST among population pairs and heterozygosity and nucleotide

185   diversity (π) within populations (Danecek et al. 2011). We investigated genetic structure

186   within North American starlings using an analysis of molecular variance (AMOVA) in the R

187   package *poppr* (Kamvar et al. 2014). We tested whether most genetic variation was

188   observed among individuals or among sampling sites ("populations"). To determine

189   significance, we compared observed variation at each hierarchical level to the randomly

190   permuted distance matrices for that particular level using 1000 simulations in the function

191   *randtest()* in the R package *adegenet* (Jombart 2008), hypothesizing that the observed

192   variance is greater than expected within individuals and less between individuals and

193   between populations. We tested for isolation by distance (IBD) using a simple Mantel test

194   in *adegenet* (Jombart 2008): for these data, the assumption of stationarity likely holds,

195   given that North American starlings appear to be in mutation-migration-drift equilibrium

196   (Guillot & Rousset 2013).

197

198   *(2) Population structure*

199   We first tested for population structure using STRUCTURE (Pritchard *et al.* 2000) by

200   simulating 10 runs of each *K*, hypothesizing that North American starlings would cluster

201   into at most eight populations (*K*=1-8). To select the best-supported K, we used the Evanno

202   method implemented in STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt 2011). We

203   averaged results across the 10 runs using the greedy algorithm in the program CLUMPP

204   v1.1.2 (Jakobsson & Rosenberg 2007), and visualized results using DISTRUCT v1.1

205   (Rosenberg 2003). Given that evidence of population structure depends on the filtering

206   thresholds selected, we ran this model-based approach using a very strict minimum minor

207   allele frequencies(MAF=0.3) and a more relaxed minimum frequency (MAF=0.1) (Linck &

208   Battey 2017). STRUCTURE results did not differ substantially among MAF thresholds, and

209   we used a filtered dataset with a minimum minor allele frequency of 0.1 in subsequent

210   analyses. We also used non-parametric approaches to determine whether starlings

211   clustered by sampling location, using principal components analysis in SNPRelate (Zheng et

212   al. 2012) and discriminant analysis of principal components in adegenet (Jombart 2008).

213   Because we expect population structure to be fairly low given the recent expansion

214   of North American starlings, we used fineRADstructure to test for more subtle patterns of

215   structure (Malinsky et al. 2016). This program calculates shared ancestry using a

216   coalescent model to determine haplotype linkage among sampled individuals. The resulting

217   coancestry matrix controls for similarity among individuals to infer fine-scale patterns of

218   population structure.

219   To identify potential geographic barriers, we used the program EEMS (Estimated

220   Effective Migration Surfaces, (Petkova et al. 2015). EEMS estimates how quickly genetic

221   similarity decays across the landscape, allowing us to pinpoint geographic regions that

222   depart from continuous IBD. Because the number of hypothesized demes (subpopulations)

223   can influence model sensitivity, we ran EEMS using polygons covering the entire North

224   American range and only the areas sampled, and also tested each map using different

225   numbers of demes (N=200, N=500, N=750 and N=1000). We adjusted the variance of the

226   proposal distribution for both migration and diversity parameters to ensure that all

227   parameters were accepted between 10 and 15% of the time, with the input proposal

228   variances as follows: mSeedsProposalS2 = 0.15, mEffctProposalS2 = 1.2, qSeedsProposalS2

229   = 1.8, qEffctProposalS2 = 0.1, and mrateMuProposalS2 = 0.01.

230

231   *(3) Demographic modeling*

232   Because starlings appear to be panmictic based on evidence from STRUCTURE and

233   PCA, we reconstructed demographic history using the Stairway plot method (Liu & Fu

234   2015). This method estimates recent population histories from hundreds of unphased, low-

235   coverage loci, which distinguishes the stairway plot from other demographic methods (e.g.,

236   PSMC) that can infer ancient population history more accurately. The stairway plot method

237   models changes in population size using the site frequency spectrum, where the null model

238   assumes constant size. We used this model-flexible method to determine whether starlings

239   experienced any genetic bottleneck after introduction: in the stairway plot, this result could

240   occur if an alternative model was accepted during one or more steps of the stairway plot.

241   We assumed a mutation rate of 1×10^{-9} and a generation length of 4.6 years (BirdLife

242   International). We estimated the folded SFS using a Python script from Simon Martin

243   (available at https://github.com/simonhmartin/genomics_general), and used the

244   recommended 67% of sites for training. The results presented here are averaged among

245   eight independent runs, each with 10 to 30 randomly generated breakpoints during the

246   reconstruction.

247

248   *(4) Selection analyses*

249   We first identified environmental variation at each sampling location using the R

250   package *raster* (Hijmans & van Etten 2012), extracting all 19 bioclimatic variables for each

251   set of sampling coordinates from WorldClim 2.0 at a resolution of 5 min on June 16, 2018

252   (Fick & Hijmans 2017). For the univariate method (LFMM), environmental variation was

253   modeled as the first three principal components of bioclimatic variation across the range of

254   North American starlings. The multivariate methods (RDA) used raw values of bioclimatic

255   variables, retaining five variables with relatively low variance inflation factors: BIO1

256   (VIF=3.54), BIO7 (4.55), BIO12 (8.69), BIO16 (7.91), and elevation (2.19).

257   As a univariate test of selection, we used the *lfmm* function (Frichot *et al.* 2013) to

258   test for associations with climatic gradients and to decrease the number of false positives.

259   We used the R package *LEA* (Frichot & François 2015) to prepare input files and run a

260   model where genotypic variation is considered a response variable in a linear regression

261   that controls for latent factors (e.g., population structure and/or background variation) in

262   estimating the association between the genotypic response and the environmental

263   predictor. For each of three models—including 1, 2, and 3 latent factors—we ran 30 MCMC

264   chains of 10,000 cycles each, discarding a burn-in of 5,000 cycles. Z-scores were combined

265   across all 30 runs and p-values readjusted to calibrate the null hypothesis and increase

266   power using the Fisher-Stouffer method as suggested in the LEA and LFMM manuals. We

267   used the Benjamini-Hochberg algorithm to control for false discoveries. Since we identified

268   1315 candidates using a q-value cut-off of 0.01—which is already 10-fold more stringent

269   than the recommended value—only loci that were identified in all three runs (K=1-3 latent

270   factors) and were more than five standard deviations from the mean log10p value were

271   considered candidates under selection (FDR<0.05, TP>25).

272   To complement the univariate methods (Bayescenv and LFMM), We also used

273   redundancy analysis (RDA) to examine how loci may covary across multiple environmental

274   gradients (Forester *et al.* 2018). RDA is especially powerful when testing for weak

275   selection, and detects true positives in large data sets more reliably than other multivariate

276   methods like Random Forest (Forester *et al.* 2018). Because RDA requires no missing data,

277   we first imputed genotypes where missing sites were assigned the genotype of highest

278   probability across all individuals—a conservative but quick imputation method. Out of over

279   1.6 million data points, about 750,000 were missing. We then used the R package vegan to

280   run the RDA model; for a full description of this method, see (Forester *et al.* 2018). Briefly,

281   RDA uses constrained ordination to model a set of predictor variables, and unconstrained

282   ordination axes to model the response (genetic variation). RDA infers selection on a

283   particular locus when it loads heavily onto one or more unconstrained predictor axes; we

284   retained five predictors (BIO1, BIO7, BIO12, BIO16, and elevation) with relatively low

285   variance inflation factors (range: 2.2—8.7). We tested for significance using the *anova.cca*

286   function within the *vegan* package, and also permuted predictor values across individuals

287   to further check significance of the model. We identify candidate loci as those that are 3 or

288   more standard deviations outside the mean loading. The R script for all RDA analyses and

289   figures were written by Brenna Forester (available at https://popgen.nescent.org/2018-

290   03-27_RDA_GEA.html).

291   Differentiation methods can identify loci that have undergone strong selective

292   sweeps, but these methods may be inappropriate in systems like this one with low overall

293   differentiation like this one. On the other hand, F_{ST}-based genome-scan approaches may

294   identify loci that stand out against a low background level of differentiation. To test these

295   more traditional methods against the model-based approaches described above, we also

296   used BayeScEnv, which incorporates environmental differentiation when identifying

297   outlier loci (de Villemereuil & Gaggiotti 2015). BayeScEnv includes a term to explicitly

298   model environmental differentiation in the framework used in BayeScan, which enables us

299   to pinpoint loci that may be associated with environmental variation. For full details on this

300   method, see the Supplementary Information.

301   To functionally annotate the contigs that contain these candidate SNPs, we assessed

302   homology to the starling (*Sturnus vulgaris*) genome using BLASTN (Altschul et al. 1990).

303   We first used the bedtools "getfasta" option to extract 10kb regions surrounding each

304   candidate SNP (Quinlan & Hall 2010). We identified genes by choosing the match with the

305   lowest E-value and highest identity: most matches showed 100% identity and an E-value

306   approximated to zero, and all matches have an E value of $< 1.0 \times 10^{-5}$ and >90% identity.

307   Gene IDs of all candidates were uploaded to PANTHER to test for overrepresentation of

308   particular gene ontology terms using a Fisher's exact test and displaying the false discovery

309   rate (Thomas *et al.* 2003; Mi *et al.* 2009).

310

311   **RESULTS**

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

313   We identified 15,038 SNPs at a mean of 27X coverage across 17 sampling locations

314   of European starlings in North America. Genome-wide F_{ST} is extremely low (0.0085), and

315   measures of genetic diversity do not vary substantially among sampling locations (Table 1).

316   Across all populations, the highest pairwise F_{ST} was 0.0106, differentiating birds from the

317   adjacent states of Arizona and New Mexico. Using a haplotype-based statistic of

318   differentiation, ϕ_{ST} among populations shows an absence of genetic structure ($\phi_{ST} =$

319   0.0002). Hierarchical AMOVAs reveal that 94% of the observed genetic variance is

320   explained by variation within individuals, and the remaining variance reflects differences

321   among individuals in the same population, with negligible variation explained at the

322   between-population level (Figure 1C-D). Across the genome, F_{ST} and nucleotide diversity

323   are exceptionally low (Figure 2A-B). Genome-wide heterozygosity is moderate at 0.339,

324   and observed heterozygosity differs significantly from expected: there are more loci with a

325   minor allele frequency near zero than might be expected by chance (t = 66.6, df = 3569,

326   P<0.001; Figure 2D), although the genetic mechanism generating this excess of rare alleles

327   is unknown.

328

	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*	

329
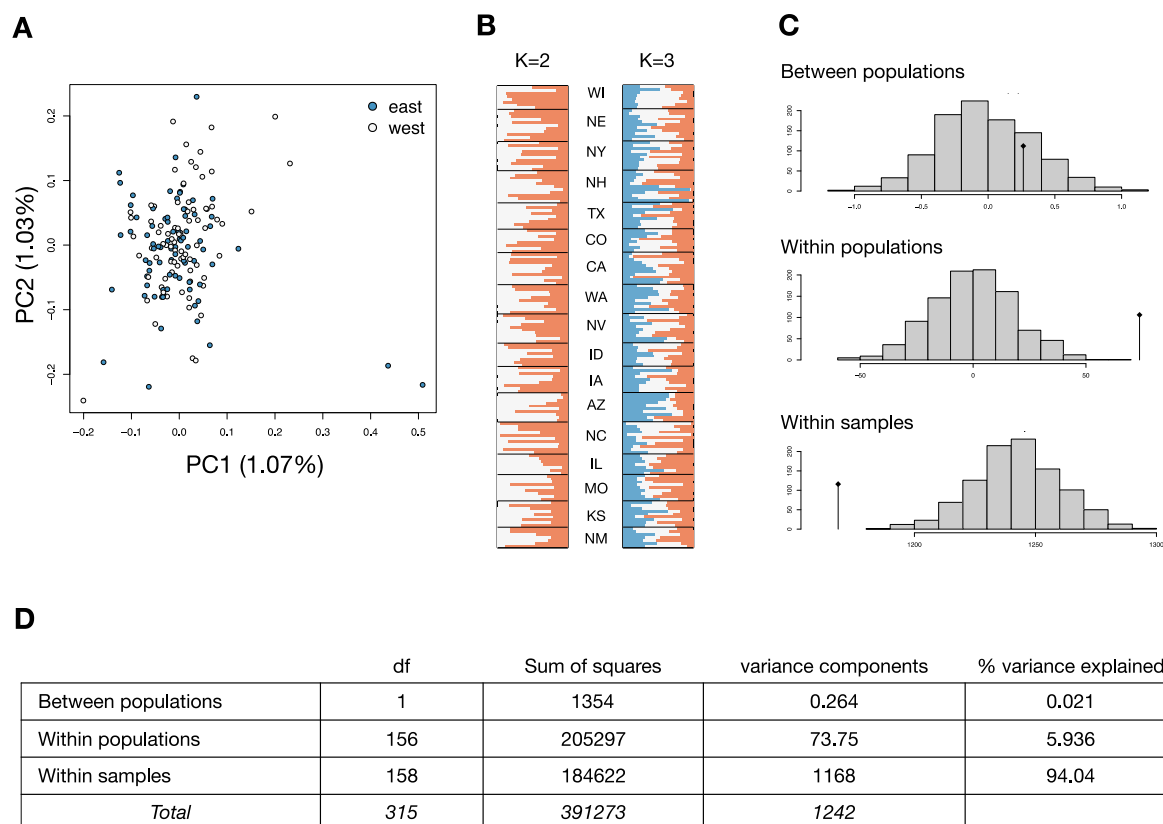
330   **Figure 1.** Population structure. A) Principal components analysis on 6,287 SNPs, explaining
331   1.03-1.08% of genetic variation observed. East and west indicate sampling locations
332   relative to the Rocky Mountains. B) STRUCTURE analyses with K=2 and K=3 (best
333   supported). C) Significance testing of hierarchical AMOVAs: the histogram shows expected
334   variance components based on simulations, and the black diamond is the observed
335   variance component. D) AMOVA results.

336 *(2) Population structure*

337 Principal components explain only 0.91% of variation among individuals (Figure 1A), and

338 although STRUCTURE identified three populations at the best-supported value of K, these

339 predicted populations do not show obvious differences in ancestry proportions (Figure

340 1B). Controlling for shared ancestry does not resolve population structure, and instead

341 provides support for uniform gene flow among individuals (Figure S1). K-means clustering

342 within DAPC also does not identify biologically relevant clusters.

343 There is no relationship between geographic and genetic distance (Mantel statistic =

344 -0.07943, P=0.904). Spatially explicit models of isolation-by-distance show that starlings

345 west of the Rocky Mountains are more diverse than their eastern counterparts (Figure 3).

346 In terms of migration, the model correctly recovers a higher rate of migration of starlings

347 moving north from the introduction site in New York, which matches what we know from

348 historical records (Jernelov 2017). Although the EEMS model estimates the highest

349 migration rate in the south of the North American range, we have few samples within the

350 area of highest inferred migration. The model of migration estimates the lowest rates of

351 migration among Midwest populations, where sampling is most thorough. The migration

352 rate is extremely low just east of the potential geographic barrier of the Rocky Mountains,

353 but increases strongly just west of the mountains.
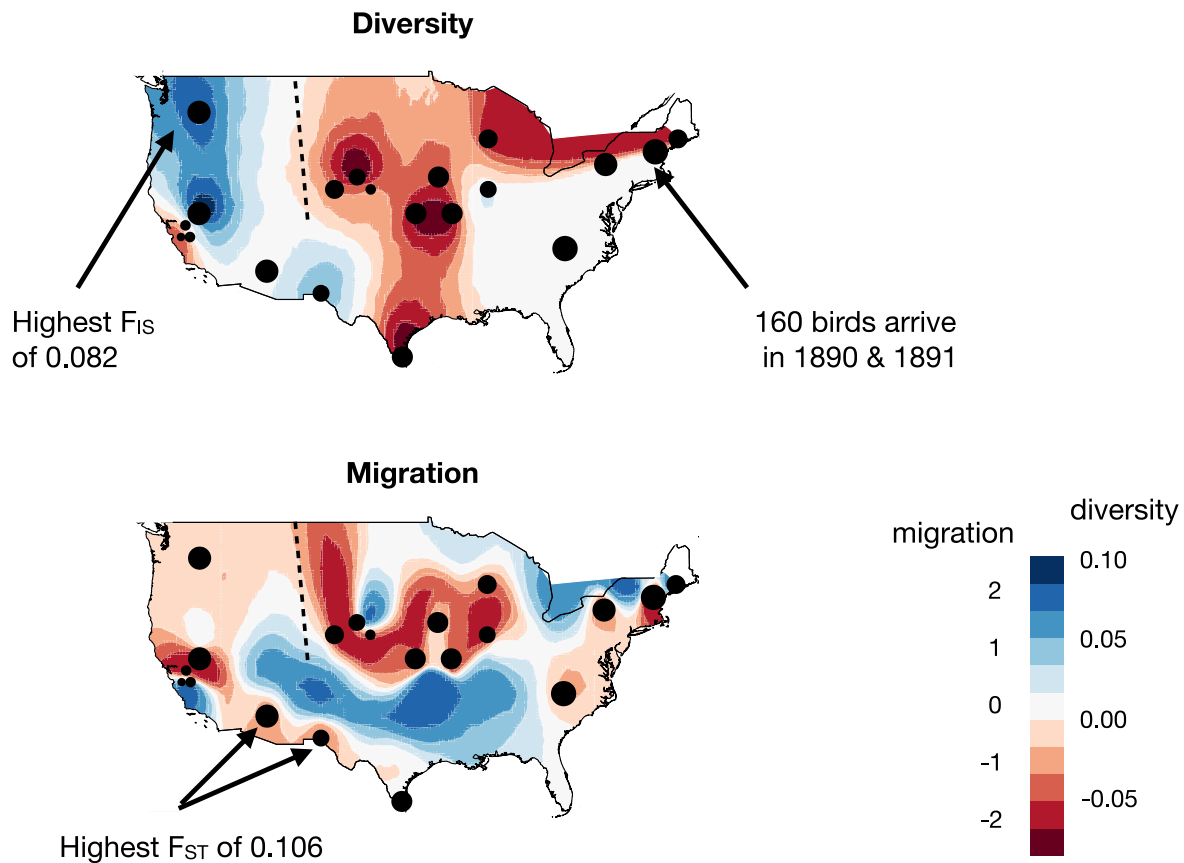
Diversity

Migration

Figure 2. Estimates of diversity and migration rates. Warmer (red) colors indicate lower levels of migration and diversity, whereas cool (blue) colors show high rates. The upper panel shows that diversity is estimated to be highest in the WA and NV populations, and diversity is much lower in the Midwest. The lower panel indicates low migration rates among nearly all populations sampled.

(3) Demographic modeling

For much of the starling's residence in North America, models suggest that its

effective population size has declined steadily (Figure S2). Upon introduction

approximately 130 years ago, effective population size was 10,000 individuals, and

population size has gradually declined to 4,000 individuals. This decline does not suggest a

classical founder effect, since population size remains fairly steady upon introduction to

New York, but we can detect very low levels of inbreeding within some populations

368   (highest F_{IS}=0.082 in Washington). Importantly, the decline in Ne in the most recent time

369   steps—the last 100 years—may be a spurious pattern resulting from known uncertainties

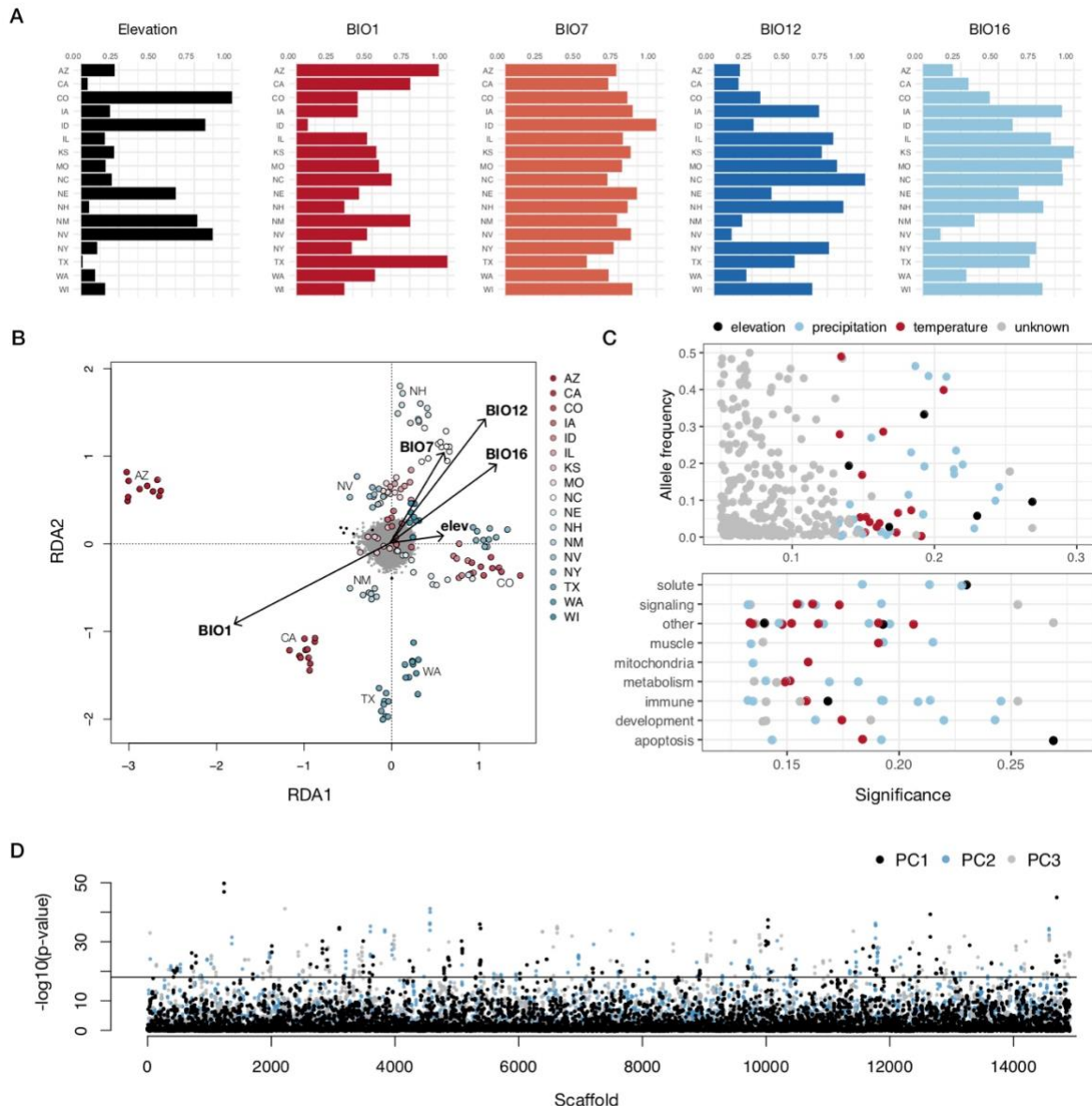370   in the final steps of this stairway plot method (Liu & Fu 2015).

371

372

Figure 3. Evidence for incipient local adaptation. A) Variation among sampling locations in
environmental predictors. B) Redundancy analyses indicate that 191 SNPs (small gray
points) are associated with bioclimatic predictors (vectors). BIO1: mean annual

376 temperature; BIO7: temperature annual range; BIO12: annual precipitation; BIO16:
377 precipitation of wettest quarter. C) Top panel illustrates the relationship between allele
378 frequency and strength of selection (R^2). Bottom panel assigns SNPs under selection to
379 gene ontology categories. Colors in both panels reflect environmental predictors. D) Latent-
380 factor mixed models show that 1315 SNPs are associated with PC1-3 of environmental
381 variation. Black line shows 5 s.d. (black line) above the mean $log_{10}(p\text{-}value)$.
382

383 *(4) Selection analyses*

384   Starlings encounter a range of precipitation, temperature, and elevation across their

385 range (Figure A), and about 10% of the ~15,000 SNPs sampled are associated with this

386 environmental variation (q-value < 0.01, FDR < 0.05, TPR > 15). Redundancy analyses

387 revealed the strongest signatures of local adaptation, showing that 191 variants are

388 correlated with environmental differences among populations (F = 1.022, P = 0.002, Figure

389 3B). Populations living in warmer climates tend to cluster more closely in the left quadrant

390 and high elevation populations cluster in the middle right quadrant. However, populations

391 do not cluster based on geographic distance: for example, starlings from TX and WA cluster

392 closely due to shared genetic variants, even though the two populations differ substantially

393 in precipitation and temperature. After controlling for population structure, candidates for

394 selection are equally distributed among elevation, temperature- and precipitation-

395 associated predictors (Figure 3C). Mean annual temperature (BIO1) opposes selective

396 pressure related to the range of temperatures experienced each year (BIO7), annual

397 precipitation (BIO12), precipitation in the wettest quarter (BIO16) and elevation. Latent-

398 factor mixed models identified 2490 candidate variants associated with the first principal

399 component of environmental variation, which explains 41.5% of the variation and loads

400 with temperature-related variables (Figure 3D). An additional 1315 variants were

401 associated with precipitation-related PC2, and/or with PC3, a composite of temperature

402   and precipitation variables. BayeScEnv identified six SNPs (Table S2), but none of these

403   variants were found in any other association analysis. Genes under stronger selection tend

404   to have lower allele frequencies (Figure 3C).

405   Genes under putative selection function in several biological processes (Figure 3C,

406   Table S2). Although no gene ontology categories were significantly overrepresented,

407   signaling and response to stimuli were particularly well-represented among GO terms,

408   showing up to 48-fold enrichment (FDR-corrected P=0.12-0.98, Table S2). Among

409   signaling-related GO terms, neuron development, synaptic transmission and organization

410   were particularly common (Table S2). Other common GO terms relate to kidney function,

411   viral processing, metabolism, and regulation of growth factors. Among the top twenty

412   variants under strong selection ($r^2 > 0.2$ or $\log_{10}p > 10$), we find four genes related to

413   growth factors (EOGT, GAB3, HBEGF, STAT3), six involved in immune responses that do

414   not directly involve growth factors (DNAJB14, FKBP4, ASB2), and three essential to muscle

415   function (LIMCH1, HBEGF, CALD1). Putatively selected genes play a role in physiological

416   processes that support starlings' invasion success in North America.

417

418   **DISCUSSION**

419   Our whole-genome data reveal that invasive starling populations show very low

420   levels of genetic differentiation and moderate genetic diversity across North America.

421   Genetic diversity is typically predicted to be low in invasive populations (Dlugosch &

422   Parker 2008), and although RADseq tends to estimate lower genetic diversity than the true

423   diversity, this bias is minimal for taxa known to be genetically depauperate (Cabe 1998;

424   Cariou *et al.* 2016). Genetic variation is not explained by geographic distance, as

425   hierarchical AMOVAs show that variation within and among individuals explains observed

426   differentiation better than variation among populations. These patterns are consistent with

427   the expectation that extensive gene flow—as shown by extremely low F_{ST} among

428   populations—maintains high connectivity across North American starling populations.

429   Perhaps surprisingly, neither demographic reconstructions nor patterns of genetic

430   diversity show evidence of a genetic bottleneck upon introduction, suggesting that many of

431   the approximately 180 birds known to have been translocated went on to contribute to the

432   invasive population.

433   Starlings' success depends on their ability to cope with highly variable

434   environmental conditions across North America, and this species now thrives in habitats

435   with dramatically different temperatures and levels of precipitation. Although there is no

436   evidence of population structure or of isolation by distance across the North American

437   range of this species, models indicate some subtle spatial patterns of genetic variation. Our

438   results, combined with previous genetic studies (Cabe 1998; 1999), support the hypothesis

439   that the Rocky Mountains may have imposed an altitudinal barrier to starlings' spread:

440   spatially explicit models indicate a decreased migration rate on the eastern front of the

441   mountains, and an increase in genetic diversity west of the mountain range (Figure 2). In

442   other words, Western starling populations on the range edge are more diverse than

443   populations nearer to the introduction site. Historical records complement this genetic

444   evidence, as the starling expansion slowed only when reaching these mountains (Jernelov

445   2017). We suggest that elevation may impose a strong barrier even across this species'

446   worldwide range: both in this study and in a parallel study of Australian populations,

447   patterns of genetic variation can be attributed to a montane barrier (Cardilini 2016).

448   This genome-wide panmixia across North American starlings allows for robust tests

449   of selection on loci that may be involved in local adaptation. Because genetic diversity and

450   differentiation are exceptionally low, it is relatively easy to identify sites that have

451   differentiated against the background of low genome-wide or population-specific

452   divergence (Dlugosch *et al.* 2015). We find only six candidates out of approximately 15K

453   SNPs using an outlier-scan method that includes environmental variation (BayeScEnv).

454   When we control for population structure and other confounding factors (Hoban *et al.*

455   2016; Forester *et al.* 2018; Capblancq *et al.* 2018), we identify 191 SNPs that directly

456   correlate with variation in temperature, precipitation, and/or elevation (Figure 3C). When

457   controlling for geographic or other confounding variation, we find strong support for local

458   adaptation when we explicitly model covariation among environmental and genetic

459   distances. The relationship among allele frequency and strength of selection supports the

460   hypothesis of incipient local adaptation, since the strongest candidates for selection tend to

461   be rarer among North American starlings (Figure 3C).

462   Our study suggests that environmentally-mediated selection best explains the

463   patterns of genetic variation observed. Although the rare variants we identify may have

464   arisen through genetic drift, there is no evidence of genetic bottlenecks that may drive

465   alleles towards fixation. Instead, two independent tests for genotype-environment

466   associations show that changes in precipitation and temperature explain the low levels of

467   genetic variation in North American starlings. Rapid evolution in genes underlying key

468   physiological processes may have supported the starling's spread across North America.

469   We suggest that aridity and cold temperatures that are not experienced in the starling's

470   native range exert enough selective pressure on North American starlings to result in

471   incipient local adaptation. Variants in genes that coordinate physiological responses are

472   most often correlated with precipitation in the wettest quarter (BIO16), which lends

473   additional support to the hypothesis that arid environments which receive little rainfall

474   year-round may generate the strongest selection on starlings. For example, effective solute

475   transport and kidney development are critical in dry habitats, which may explain why all

476   but one of the genes related to kidney function correlate with precipitation (BIO16).

477   Claudin 16 (CLDN16; R^2 = 0.23) is one such protein that regulates ion concentrations in the

478   kidney, while others maintain homeostasis and vasoconstriction (AVPR1B; R^2 = 0.18) or

479   transport iron (STEAP3; R^2 = 0.21). Many invaders shift their diet upon colonization of a

480   new habitat, and many candidates play a role in metabolism and/or digestion: for example,

481   aridity may result in selection on proteins that process lipids (MTMR3; R^2 = 0.18) and fatty

482   acids (PEX5; R^2 = 0.17), since organisms living in dry environments often depend on fat

483   storage for proper hydration. Proteins that modify growth factors—key orchestrators of

484   cellular growth and development—correlate with aridity but also temperature, and

485   complexes that rely on ubiquitin ligases to degrade proteins are similarly strong

486   candidates. Many of the putatively selected genes are critical to starlings' survival, and

487   investigating a wider range of environmental conditions and sampling whole genomes may

488   support this preliminary evidence of incipient local adaptation.

489   A similar study of starlings in the Australian invasion found that geographic but not

490   environmental distance explains genetic patterns there (Cardilini 2016). Starlings in the

491   Australian range show substantial population structuring and significant patterns of

492   isolation-by-distance. This contrast is especially remarkable because Australian

493   populations that are geographically close can experience dramatically different

494   environments, whereas environmental variation changes much more gradually across

495   North America (Hofmeister et al., *in prep*). Because Australia is a more variable

496   environment, starlings may have differentiated more rapidly within Australia than in North

497   America. Global F_{ST} across all Australian populations is an order of magnitude higher than

498   the equivalent F_{ST} index across North America, despite similar areas sampled. After

499   controlling for the considerable structuring among Australian populations, genotype-

500   environment associations reveal parallel signatures of selection among the two invasions.

501   Two independent tests of incipient local adaptation in Australia and North America suggest

502   that starlings may rapidly adapt to novel selective pressures. Both studies show that

503   genetic variation can be explained by extremes in temperature and precipitation, and

504   preliminary results of whole-genome resequencing of native and introduced populations

505   confirm that variability in temperature and precipitation may shape observed genetic

506   variation in starlings world-wide (Hofmeister et al., *in prep*).

507   This study explores how genetic variation changes across the landscape, but we

508   cannot fully understand gene flow without studies of dispersal and migration of the

509   individuals that carry genes. Juvenile dispersal promotes the high levels of gene flow

510   observed in North America (Cabe 1999). In addition, long-distance dispersal may be

511   common in introductions of this species: in South Africa, starlings disperse when their

512   natal environment becomes crowded or unsuitable (Hui *et al.* 2012). In both native and

513   introduced ranges, starlings are partial migrants: some overwintering individuals may

514   reside in that location year-round, whereas others migrate to that wintering location and

515   breed elsewhere (Kessel 1953). Because our samples were collected during winter,

516   individuals within each sampling location may come from different breeding populations. If

517   we have only a few individuals from each of these true populations, sampling in this study

518   may thus complicate population assignment. Although this study is the most thorough

519   sampling of the North American range to date, additional sampling could reveal stronger

520   population structure among true breeding populations, or divergence among resident and

521   migrant individuals. Detailed ecological studies of dispersal and migration in North

522   American starlings are needed to complement the genetic evidence shown here.

523   Our results contribute to the growing literature on rapid evolution in novel

524   environments, even in extremely young systems. Replicate invasive populations allow us to

525   explore the genetic consequences of colonization and establishment in novel environments.

526   On a background of low genetic differentiation and diversity, we find evidence of incipient

527   genotype-environment associations in North American starlings. Our results complement

528   other recent studies that reveal associations between climate variables and particular loci

529   in North American vertebrates (Schweizer *et al.* 2015; Bay *et al.* 2018). Finally, we suggest

530   that rapid local adaptation can evolve even in dispersive and young populations. Studies

531   such as this highlight the need for detailed eco-evolutionary studies of dispersal and

532   habitat choice as potentiating forces in local adaptation.

533

ACKNOWLEDGEMENTS

542

DATA ACCESSIBILITY STATEMENT

544

545 All scripts are archived on GitHub: https://github.com/nathofme/radseq-NAm

546 All data files that accompany the above scripts will be archived on Dryad upon journal
547 submission, and DNA sequences will be archived in NCBI SRA.
548
549 **REFERENCES**
550

551 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search
552   Tool. *Journal of Molecular Biology*, **215**, 403–410.
553 Avalos A, Pan H, Li C *et al.* (2017) A soft selective sweep during rapid evolution of gentle
554   behaviour in an Africanized honeybee. *Nature Communications*, 1–9.
555 Bay RA, Harrigan RJ, Le Underwood V *et al.* (2018) Genomic signals of selection predict
556   climate-driven population declines in a migratory bird. *Science*, **359**, 83–86.
557 Bitton P-P, Graham BA (2014) Change in wing morphology of the European starling during
558   and after colonization of North America. *Journal of Zoology*, **295**, 254–260.
559 Cabe PR (1998) The effects of founding bottlenecks on genetic variation in the European
560   starling (*Sturnus vulgaris*) in North America. *Heredity*, **80**, 519–525.
561 Cabe PR (1999) Dispersal and population structure in the European Starling. *The Condor*,
562   **101**, 451–454.
563 Capblancq T, Luu K, Blum MGB, Bazin E (2018) Evaluation of redundancy analysis to
564   identify signatures of local adaptation. *Molecular Ecology Resources*, **185**, 587–11.
565 Cardilini APA (2016) Phenotypic and genetic variation of non-native common starling,
566   *Sturnus vulgaris*.
567 Cariou M, Duret L, Charlat S (2016) How and how much does RAD-seq bias genetic
568   diversity estimates? *BMC Evolutionary Biology*, 1–8.
569 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool
570   set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
571 Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for
572   differentiation, natural selection, and local adaptation. *Molecular Ecology*, **24**, 1999–
573   2017.
574 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools.
575   *Bioinformatics*, **27**, 2156–2158.
576 de Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local
577   adaptation using environmental variables. *Methods in Ecology and Evolution*, **6**, 1248–
578   1258.
579 Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation,
580   adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–
581   449.
582 Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the
583   details: genetic variation in introduced populations and its contributions to invasion.
584   *Molecular Ecology*, **24**, 2095–2111.
585 Duckworth RA (2008) Adaptive Dispersal Strategies and the Dynamics of a Range
586   Expansion. *The American Naturalist*, **172**, S4–S17.
587 Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for
588   visualizing STRUCTURE output and implementing the Evanno method. *Conservation
589   Genetics Resources*, **4**, 359–361.
590 Edelaar P, Bolnick DI (2012) Non-random gene flow: an underappreciated force in
591   evolution and ecology. *Trends in Ecology & Evolution*, **27**, 659–665.

592   Estoup A, Ravigné V, Hufbauer R *et al.* (2016) Is There a Genetic Paradox of Biological
593   Invasion? *Annual Review of Ecology, Evolution, and Systematics*, **47**, 51–72.
594   Facon B, Hufbauer RA, Tayeh A *et al.* (2011) Inbreeding Depression Is Purged in the
595   Invasive Insect Harmonia axyridis. *Current Biology*, **21**, 424–427.
596   Fick SE, Hijmans RJ (2017) *WorldClim 2: New 1-km spatial resolution climate surfaces for*
597   *global land areas. International Journal of Climatology*.
598   Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting
599   multilocus adaptation with multivariate genotype-environment associations. *Molecular*
600   *Ecology*, **27**, 2215–2233.
601   Frichot E, François O (2015) LEA: An Rpackage for landscape and ecological association
602   studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**, 925–929.
603   Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for Associations between
604   Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology*
605   *and Evolution*, **30**, 1687–1699.
606   Garant D, Forde SE, Hendry AP (2007) The multifarious effects of dispersal and gene flow
607   on contemporary adaptation. *Functional Ecology*, **21**, 434–443.
608   Guillot G, Rousset F (2013) Dismantling the Mantel tests (L Harmon, Ed,). *Methods in*
609   *Ecology and Evolution*, **4**, 336–344.
610   Günther T, Genetics GC, 2013 Robust identification of local adaptation from allele
611   frequencies. *Genetics Soc America*
612   .
613   Hijmans RJ, van Etten J (2012) raster: Geographic analysis and modeling with raster data. R
614   package version 2.0-12.
615   Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local
616   Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*,
617   **188**, 379–397.
618   Hufbauer RA, Facon B, Ravigné V *et al.* (2011) Anthropogenically induced adaptation to
619   invade (AIAI): contemporary adaptation to human-altered habitats within the native
620   range can promote invasions. *Evolutionary Applications*, **5**, 89–101.
621   Hui C, Roura-Pascual N, Brotons L, Robinson RA, Evans KL (2012) Flexible dispersal
622   strategies in native and non-native ranges: environmental quality and the "good-stay,
623   bad-disperse" rule. *Ecography*, **35**, 1024–1032.
624   Jacob S, Legrand D, Chaine AS *et al.* (2017) Gene flow favours local adaptation under habitat
625   choice in ciliate microcosms. *Nature Ecology & Evolution*, 1–4.
626   Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program
627   for dealing with label switching and multimodality in analysis of population structure.
628   *Bioinformatics*, **23**, 1801–1806.
629   Jernelov A (2017) Starlings in North America. In: *The Long-Term Fate of Invasive Species*,
630   pp. 137–149. Jarpas.
631   Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.
632   *Bioinformatics*, **24**, 1403–1405.
633   Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of
634   populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281–14.
635   Kessel B (1953) Distribution and migration of the European Starling in North America. *The*
636   *Condor*, **55**, 49–67.

637 Koenig W (2003) European Starlings and Their Effect on Native Cavity-Nesting Birds.
638   *Conservation Biology*, **17**, 1134–1140.
639 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*
640   *Methods*, **9**, 357–359.
641 Lenormand T (2002) Gene flow and the limits to natural selection. **17**, 183–189.
642 Lescak EA, Bassham SL, Catchen J *et al.* (2015) Evolution of stickleback in 50 years on
643   earthquake-uplifted islands. *Proceedings of the National Academy of Sciences*, **112**,
644   E7204–E7212.
645 Linck EB, Battey CJ (2017) Minor allele frequency thresholds strongly affect population
646   structure inference with genomic datasets. *bioRxiv*.
647 Linz GM, Homan HJ, Gaulker SM, Penry LB, Bleier WJ (2007) *European starlings: a review of*
648   *an invasive species with far-reaching impacts*.
649 Liu X, Fu Y-X (2015) Exploring population size changes using SNP frequency spectra.
650   *Nature Genetics*, **47**, 555–559.
651 Lombaert E, Estoup A, Facon B *et al.* (2014) Rapid increase in dispersal during range
652   expansion in the invasive ladybird Harmonia axyridis. *Journal of Evolutionary Biology*,
653   **27**, 508–517.
654 Malinsky M, Trucchi E, Lawson D, Falush D (2016) RADpainter and fineRADstructure:
655   population inference from RADseq data. 1–6.
656 Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective
657   sweeps. *Trends in Ecology & Evolution*, **28**, 659–669.
658 Mi H, Dong Q, Muruganujan A *et al.* (2009) PANTHER version 7: improved phylogenetic
659   trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids*
660   *Research*, **38**, D204–D210.
661 Ochocki BM, Miller TEX (2017) Rapid evolution of dispersal ability makes biological
662   invasions faster and more variable. *Nature Communications*, **8**, 1–8.
663 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An
664   Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-
665   Model Species (L Orlando, Ed,). *PLoS ONE*, **7**, e37135–11.
666 Petkova D, Novembre J, Stephens M (2015) Visualizing spatial population structure with
667   estimated effective migration surfaces. *Nature Reviews Neuroscience*, **48**, 94–100.
668 Pfeifer SP, Laurent S, Sousa VC *et al.* (2018) The Evolutionary History of Nebraska Deer
669   Mice: Local Adaptation in the Face of Strong Gene Flow (N Singh, Ed,). *Molecular*
670   *Biology and Evolution*, **35**, 792–806.
671 Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in
672   invasive species. *Trends in Plant Science*, **13**, 288–294.
673 Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using
674   Multilocus Genotype Data. *Genetics*, **155**, 945–959.
675 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic
676   features. *Bioinformatics*, **26**, 841–842.
677 Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success
678   of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.
679 Rollins LA, Richardson MF, Shine R (2015) A genetic perspective on rapid evolution in cane
680   toads (Rhinella marina). *Molecular Ecology*, **24**, 2264–2276.

681 Rollins LA, Woolnough AP, Fanson BG *et al.* (2016) Selection on Mitochondrial Variants
682   Occurs between and within Individuals in an Expanding Invasion. *Molecular Biology*
683   *and Evolution*, **33**, 995–1007.
684 Rollins LA, Woolnough AP, Sinclair R, MOONEY NJ, Sherwin WB (2011) Mitochondrial DNA
685   offers unique insights into invasion history of the common starling. *Molecular Ecology*,
686   **20**, 2307–2317.
687 Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB (2009) Invasive species can't
688   cover their tracks: using microsatellites to assist management of starling (*Sturnus*
689   *vulgaris*) populations in Western Australia. *Molecular Ecology*, **18**, 1560–1573.
690 Rosenberg NA (2003) distruct: a program for the graphical display of population structure.
691   *Molecular Ecology Notes*, **4**, 137–138.
692 Schlaepfer DR, Glättli M, Fischer M, van Kleunen M (2009) A multi-species experiment in
693   their native range indicates pre-adaptation of invasive alien plant species. *New*
694   *Phytologist*, **185**, 1087–1099.
695 Schmid-Hempel P, Schmid-Hempel R, Brunner PC, Seeman OD, Allen GR (2007) Invasion
696   success of the bumblebee, Bombus terrestris, despite a drastic genetic bottleneck.
697   *Heredity*, **99**, 414–422.
698 Schweizer RM, Robinson J, Harrigan R *et al.* (2015) Targeted capture and resequencing of
699   1040 genes reveal environmentally driven functional variation in grey wolves.
700   *Molecular Ecology*, **25**, 357–379.
701 Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or
702   distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.
703 Szűcs M, Vahsen ML, Melbourne BA *et al.* (2017) Rapid adaptive evolution in novel
704   environments acts as an architect of population range expansion. *Proceedings of the*
705   *National Academy of Sciences*, **114**, 13501–13506.
706 Thomas PD, Campbell MJ, Kejariwal A *et al.* (2003) PANTHER: A Library of Protein Families
707   and Subfamilies Indexed by Function. *Genome Research*, **13**, 2129–2141.
708 Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ (2017) Double-digest RAD
709   sequencing outperforms microsatellite loci at assigning paternity and estimating
710   relatedness: a proof of concept in a highly promiscuous bird. 1–34.
711 Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular*
712   *Ecology*, **25**, 2144–2164.
713 Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the
714   success of an invasive species. *Proceedings of the National Academy of Sciences*, **97**,
715   5948–5953.
716 Uller T, Leimu R (2011) Founder events predict changes in genetic diversity during human-
717   mediated range expansions. *Global Change Biology*, **17**, 3478–3485.
718 Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018) Rapid genetic
719   adaptation to a novel environment despite a genome-wide reduction in genetic
720   diversity. *Molecular Ecology*, **17**, 675.
721 Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-
722   selection balance. *Evolution*, **65**, 1897–1911.
723 Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for
724   relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–
725   3328.
726

727
728   **Table 1.** Population genetic summary statistics.

Population	Private	N	P	H_{obs}	π	F_{is}
NY (10)	0	9.4	0.76	0.34	0.35	0.037
NH (11)	1	10.4	0.76	0.33	0.35	0.065
NC (11)	1	10.5	0.76	0.33	0.35	0.060
IL (7)	1	6.7	0.76	0.33	0.35	0.056
IA (9)	0	8.8	0.75	0.33	0.36	0.078
WI (8)	0	7.7	0.75	0.33	0.35	0.050
MO (9)	0	8.8	0.75	0.33	0.35	0.053
NE (11)	0	10.4	0.75	0.34	0.35	0.042
CO (8)	1	7.8	0.76	0.33	0.35	0.066
KS (9)	2	8.6	0.76	0.35	0.35	0.012
TX (9)	1	9.3	0.76	0.33	0.35	0.052
NM (7)	2	6.8	0.76	0.33	0.36	0.068
AZ (10)	1	9.5	0.76	0.32	0.35	0.069
NV (10)	0	9.5	0.75	0.33	0.36	0.068
CA (11)	0	10.4	0.75	0.33	0.35	0.061
ID (8)	0	7.7	0.76	0.33	0.35	0.047
WA (10)	2	9.5	0.76	0.32	0.35	0.082

729
730
731   **Figure S1.** This clustered coancestry matrix shows that the estimated coancestry
732   coefficient does not vary among sampled individuals. fineRADstructure identifies fine-scale
733   patterns in population structure as a result of shared ancestry, indicating the range of
734   coancestry coefficients using a heat map.
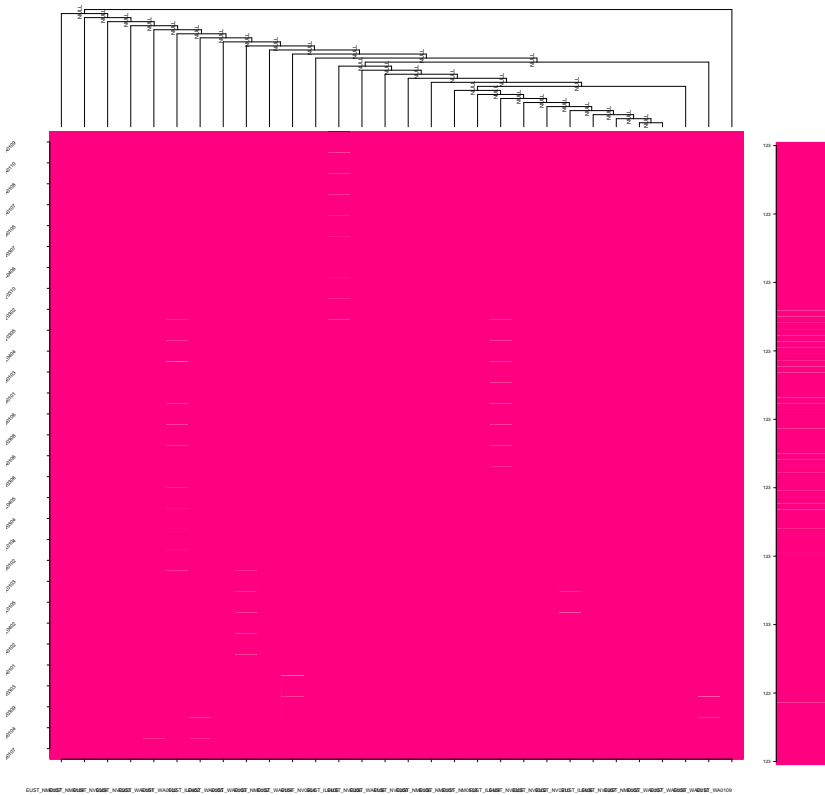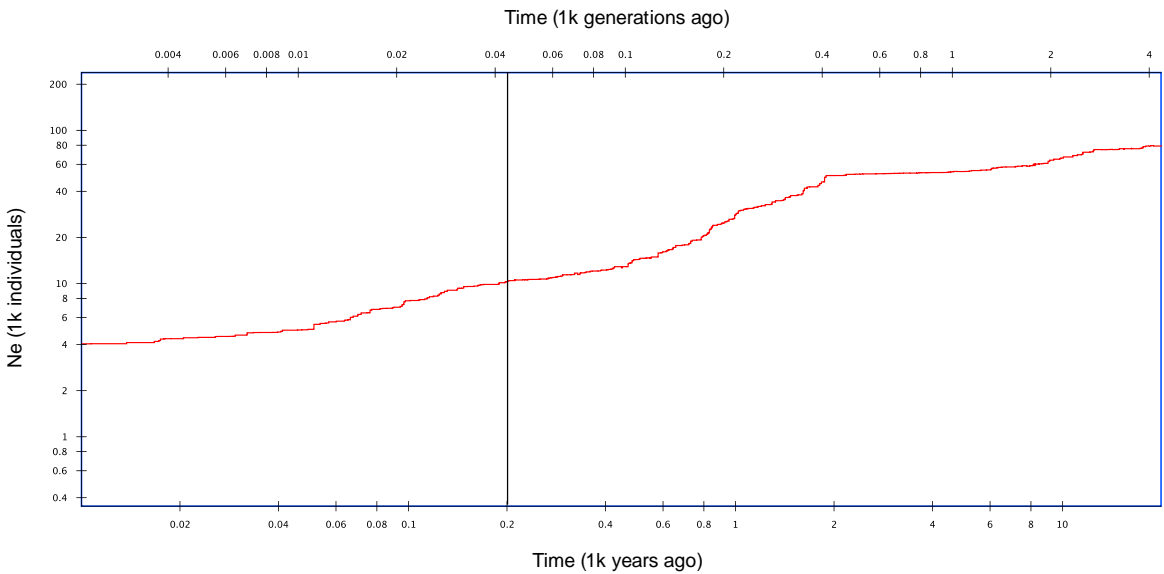
735
736
737 **Figure S2.** Stairway plot indicates a gradual decline in effective population size over time.
738 Black line indicates approximate colonization date of starlings according to historical
739 records (1890).

740
741

742   **Table S1.** Sampling locations and details.

743

744   **Table S2.** Gene and gene ontology information.

745

746