

Concurrent invasions by European starlings (*Sturnus vulgaris*) suggest selection on shared genomic regions even after genetic bottlenecks

Natalie R. Hofmeister^{1,2*}, Katarina Stuart³, Wesley C. Warren⁴, Scott J. Werner⁵,
Melissa Bateson⁶, Gregory F. Ball⁷, Katherine L. Buchanan⁸, David W. Burt⁹, Adam
P.A. Cardilini¹⁰, Phillip Cassey¹¹, Tim De Meyer¹², Julia George¹³, Simone L. Meddle¹⁴,
Hannah M. Rowland^{15,16}, Craig D.H. Sherman¹⁰, William Sherwin³, Wim Vanden
Berghe¹⁷, Lee Ann Rollins³⁺, David F. Clayton¹³⁺.

¹ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA
14850

² Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Ithaca, NY, USA
14850

³ Evolution & Ecology Research Centre, School of Biological, Earth and Environmental
Sciences, UNSW Sydney, Sydney, New South Wales, Australia

⁴ Department of Animal Sciences and Surgery, Institute for Data Science and
Informatics, University of Missouri, MO, USA

⁵ United States Department of Agriculture, Animal and Plant Health Inspection Service,
Wildlife Services, National Wildlife Research Center, Fort Collins, CO, USA 80521

⁶ Biosciences Institute, Newcastle University, Framlington Place, Newcastle upon Tyne,
NE2 4HH, United Kingdom

⁷ Department of Psychology, University of Maryland, College Park MD 20742 USA

⁸ Centre for Integrative Ecology, Deakin University, Geelong, VIC 3216, Australia

⁹ Office of the Deputy Vice-Chancellor (Research and Innovation), The University of Queensland, Brisbane, Australia

¹⁰ School of Life and Environmental Sciences, Deakin University, Waurin Ponds, VIC, Australia

¹¹ Invasion Science & Wildlife Ecology Lab, University of Adelaide, Adelaide SA 5005, Australia

¹² Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent B9000, Belgium

¹³ Clemson University, Department of Genetics & Biochemistry, South Carolina, USA 29634

¹⁴ The Roslin Institute, The Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG, UK

¹⁵ Max Planck Institute for Chemical Ecology, Hans Knöll Straße 8, Jena, 07745, Germany

¹⁶ Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ

¹⁷ Department of Biomedical Sciences, University Antwerp, Antwerp B9000 Belgium

[†]joint last authors

* corresponding author: nrh44@cornell.edu

Abstract

A species' success during the invasion of new areas hinges on an interplay between demographic processes and the outcome of localized selection. Invasive European Starlings (*Sturnus vulgaris*) established populations in Australia and North America in the 19th century. Here, we compare whole-genome sequences among native and independently introduced European Starling populations from three continents to determine how demographic processes interact with rapid adaptive evolution to generate similar genetic patterns in these recent and replicated invasions. Our results confirm that a post-bottleneck expansion may in fact support local adaptation. We find that specific genomic regions have differentiated even on this short evolutionary timescale, and suggest that selection best explains differentiation in at least two of these regions. This infamous and highly mobile invader adapted to novel selection (e.g., extrinsic factors), perhaps in part due to the demographic boom intrinsic to many invasions.

Introduction

Some species can establish and spread in a novel environment more successfully than others, and defining what makes a species ‘invasive’ is hotly contested¹⁻⁴. Invasion biologists continue to debate whether an invasion’s success can be better attributed to an intrinsic property of the founding population or to extrinsic conditions experienced by the population. When invasive populations colonize a new environment, they often undergo genetic bottlenecks⁵. However, even populations with limited genetic diversity, including those subject to founder effects, can adapt quickly to novel environments⁶⁻⁸. For example, gene surfing during range expansion has led to adaptation in experiments in microbes⁹ and wild, invasive bank voles¹⁰. Simultaneous with demographic shifts, invasive populations may encounter environmental conditions that exert novel selective regimes. Although genetic bottlenecks often limit the genetic variation available to selection, local adaptation can occur even after a genetic bottleneck^{6,11-13}.

Recent and replicated invasions are particularly useful for exploring the eco-evolutionary dynamics of population expansion¹⁴⁻¹⁶, since any divergence after introduction likely reflects a combination of demographic and/or selective forces. One such recent invader is the Common or European Starling (*Sturnus vulgaris*), which was introduced across south-eastern Australia in 1856 and to New York, United States of America in 1890¹⁷. Both the American and Australian invasions were most likely founded by individuals from the United Kingdom^{18,19}, although multiple introductions in Australia might contribute to ongoing gene flow among continents.²⁰

The starling's ecology is well-studied in both invasions, enabling us to explore how environmental factors might impact genetic variation²¹. Native-range starlings thrive in open pastures and urban environments, but starlings' ecology and life history vary among populations. In range-wide studies of Australian²²⁻²⁴ and North American^{25,26} invasions, temperature and precipitation influence genetic differentiation, even after controlling for population structure. Migration and dispersal also vary among invasions. In North America, starlings can migrate long distances each season^{19,27}, but populations in the Western USA likely disperse and/or migrate shorter distances²⁸. In contrast, starlings in Australia exhibit strong population structure and likely migrate short distances in search of food, even though environmental conditions are much more arid than in the native range²². Finally, variation in the breeding cycle may also facilitate invasion success as invasive populations tend to lay more than one clutch, whereas the UK population generally lays only one clutch^{29,30}.

We use concurrent starling invasions in Australia (AU) and North America (NA) to examine the evolutionary and genetic consequences of invasion. Both starling invasions rapidly expand from small founding populations in the late 19th century^{18,19}. Given this natural control, we take advantage of a rare opportunity to compare intrinsic and extrinsic drivers of invasion success. Founder effects and other intrinsic demographic properties of invasions certainly influence establishment^{31,32}, and we predict some divergence from the native, ancestral population (represented by UK samples here) due to genetic drift. We combine demographic models with fine-scale measures of genomic diversity and differentiation to determine whether drift alone can explain observed genetic variation, and we find strong evidence that extrinsic factors such as novel

selective pressures generate differentiation in both invasions specific to only a few genomic regions.

Results

Differentiation and population structure

We sequenced whole genomes of eight native-range individuals from Northumberland, UK and eight from each of the two invasions (NA: all 8 from New York City, USA; AU: 2 each from four locations in New South Wales, Australia, Table S1). After filtering, we obtained more than 11 million SNPs with minimum 5X coverage (average genome-wide coverage = 18.44, for details on how choice of reference genome impacts variant-calling, see SI Section 1). All patterns identified using a variant-called dataset concur with those based on genotype likelihood (ANGSD); for details on how variant-calling impacts patterns, see SI Section 2. We use a variant set filtered for minor allele frequency, Hardy-Weinberg Equilibrium, and linkage disequilibrium for analyses of population structure, but for more accurate estimates of genetic diversity and differentiation, we report results from a genome scan of variants filtered only for quality and depth. Differentiation of the two invasive populations from the native population is low, which is expected given that these populations split less than two centuries ago. However, differentiation between AU and UK (F_{ST} AU vs. UK = 0.016) is almost twice that between NA and UK (F_{ST} NA vs. UK = 0.008). We examined this contrast in genetic differentiation using analyses of population structure.

All three populations are readily distinguished from one another by a principal component analysis of 40,488 unlinked SNPs (minor AF > 0.25): PC1 (6.3% of genetic

variation) separates the UK and NA populations from the Australian population (Figure 1A). This evidence complements previous work that showed extensive population structuring in Australia but nearly continuous gene flow across North America, based on reduced-representation genomic data^{22,33}. Principal component analysis in the genotype likelihood framework of ANGSD³⁴ shows nearly identical results (Figure S6). Furthermore, individuals are reliably assigned to clades based on pairwise genetic distances calculated in ANGSD (Figure 1B).

We note that the tight clustering of UK individuals in Figure 1A contrasts with the large distances between these same individuals in Figure 1B. In the genotype likelihood dataset, genetic distances between native UK individuals are much greater than distances among individuals within each invasive population (Figure 1B). Because these two datasets differ in variant-calling and filtering strategies, the genetic distance among UK individuals in Figure 1B may reflect rare alleles that were filtered out of the variant set in Figure 1A. However, PCs 3 and 4 in the PCA of the variant-called dataset do indicate additional structure within the UK population (Figure S7).

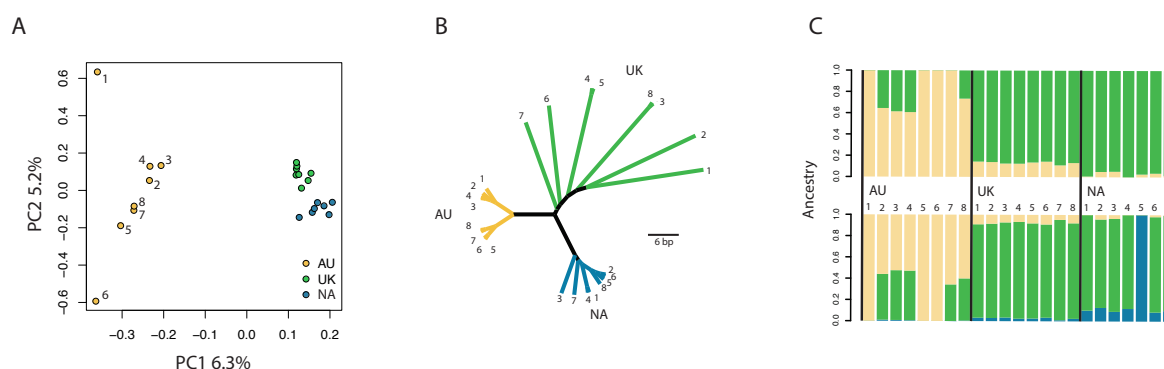


Figure 1. A) Principal components of 40,488 unlinked SNPs explain 6.3% (PC1) and 5.2% (PC2) of genetic variation. B) cladogram of genetic distances among samples based on genotype likelihoods of 16,151,007 sites. C) ADMIXTURE analyses showing K=2 (top) and K=3 (bottom row).

Admixture analyses revealed statistical support for a two-population model, which distinguished AU from NA+UK and was only slightly weaker (cross-validation error = 0.88) than the best-supported model of K=1 (cross-validation error = 0.73, Figure 1C, SI Section 3). Regardless of the number of populations hypothesized in models of admixture, the NA invasion consistently shares a higher proportion of its ancestry with the UK population. Individuals from the Australian population are distinguished from the other two populations in all tested values of K. Considered in concert, these tests of population structure show that Australian and North American populations differ in the amount of divergence from the native UK population. Founder effects likely contribute to the observed population structure, and below we describe explicit models of demographic processes.

Invasive and native populations experienced bottlenecks and subsequent expansion

To address the impact of demographic processes in generating the observed patterns we used the site-frequency spectrum built from genotype likelihoods to construct models of changes in the effective population size over time. We examined the demographic history of all three populations using fastsimcoal2³⁵. The demographic

model shows that both invasions experienced a bottleneck upon introduction (Figure 2), which is exactly what we expect given the small number of founding individuals in both AU and NA. Bottlenecks often lead to inbreeding within a population, and we find that inbreeding is negligible but slightly higher in the Australian population than in the NA population (Table S2). In addition, relatedness among individuals is quite low, where zero indicates no shared alleles among individuals (maximum AJK statistic = 0.06). Each invasion appears to have recovered quickly by expanding in effective population size, and at present, our data suggest similar effective population sizes in both the AU and NA invasions. Much of the ancestral variation appears to be shared among invasions, given that the genome-wide average F_{ST} between AU and NA is 0.04, although we note that genetic differentiation among invasions confirms the expectation that different variants will make it through each genetic bottleneck. The results presented here concur with range-wide sampling that indicates genetic bottlenecks followed by rapid expansion with little evidence of inbreeding in both Australia²² and North America³³.

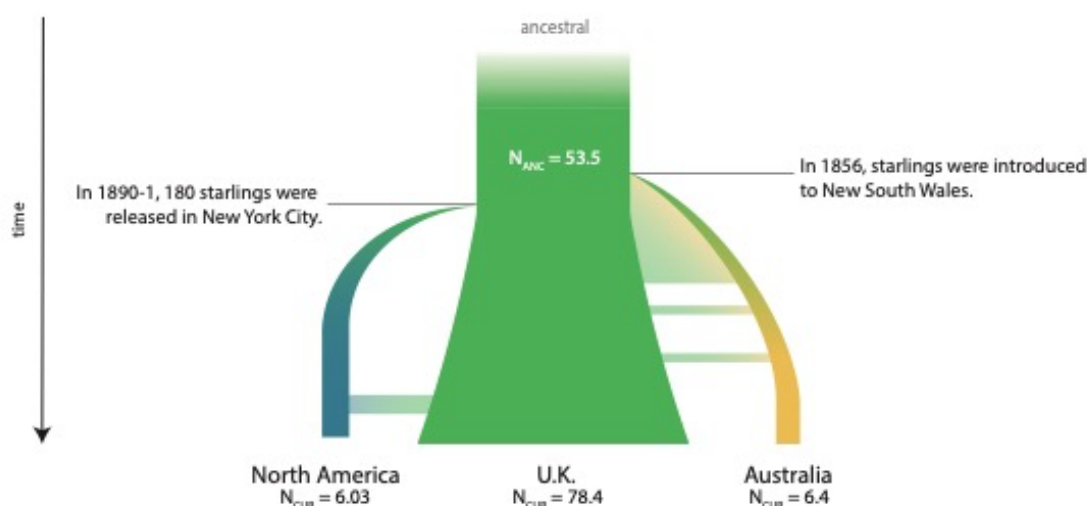


Figure 2. Demographic model of effective population size based on the site-frequency spectrum. Schematic approximates population growth based on model output from *fastsimcoal2*.

Population bottlenecks often lead to a loss of genetic diversity when genetic drift drives to fixation alleles once maintained at a moderate frequency. Since both the AU and NA invasions experienced a genetic bottleneck, we expect that nucleotide diversity (π) within each population should be higher in the older and larger native range population than in each invasive range. However, we find that 70% of 50-kb windows show greater nucleotide diversity in AU (compared to UK diversity), and 74% show greater diversity in NA. The filters applied should not lead to disproportional allelic dropout of rare alleles, because we did not filter for minor allele frequency prior to this scan. Higher invasive diversity may be a sampling artifact: rare variants in the native range may have 'surfed' to a higher frequency in the invasions, and our sampling of the native range (in Northumberland, UK) likely represents only a small portion of genetic

variants that could have been introduced in AU and NA. We also note that the UK individuals sampled here do not capture range-wide diversity in the native population, and therefore we expect that actual nucleotide diversity in the UK population is higher than what we have sampled here. Additional investigation of native range starlings will be needed to determine whether lower diversity in the UK is recovered with broader geographic sampling.

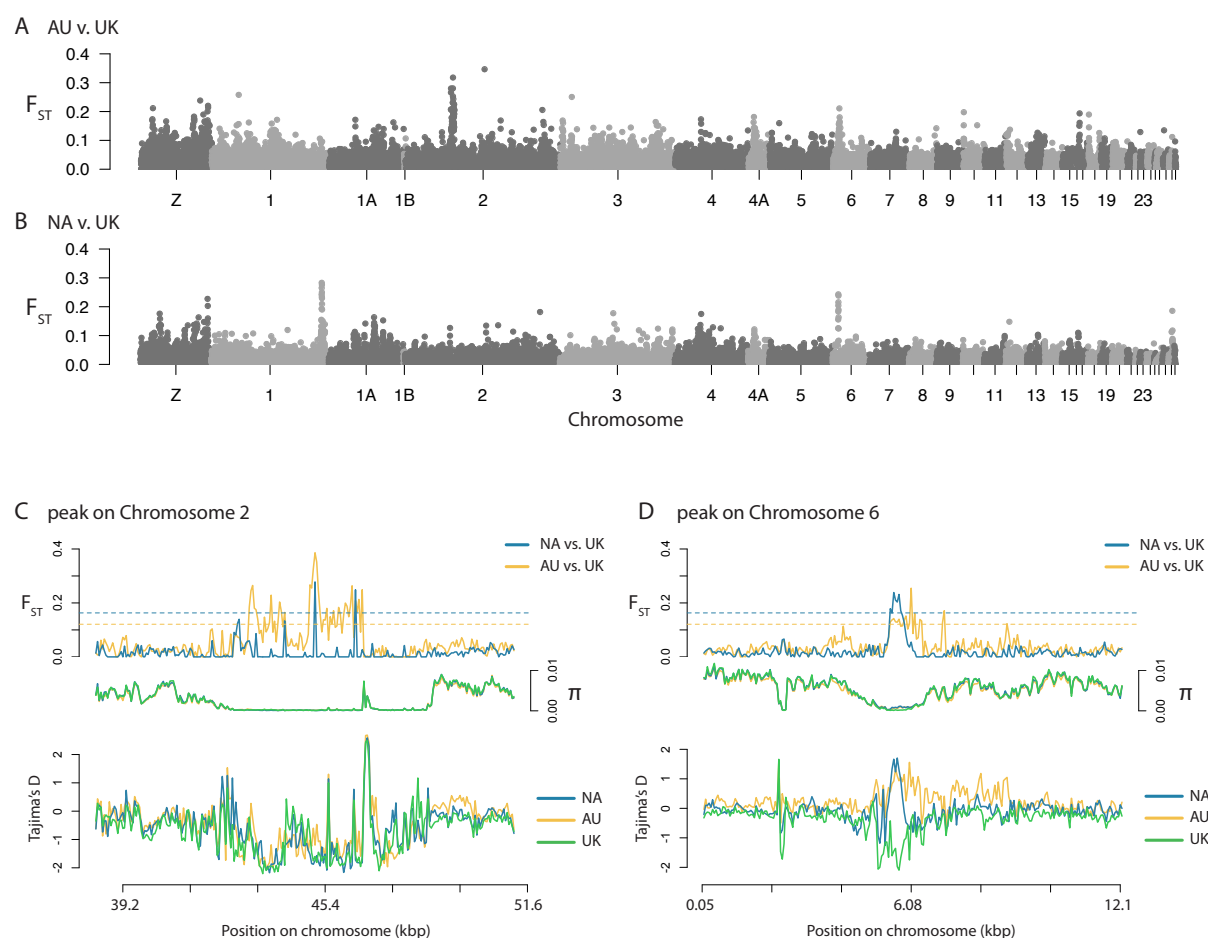
Our demographic models show that both the AU and NA invasions rapidly expanded in population size after the initial bottleneck, which predicts that many loci would be lost upon establishment of the invasive populations even as variants in other loci increase in frequency. If gene surfing facilitated by rapid population expansion explained genome-wide differentiation, then we would expect to find that regions where invasive diversity is higher than diversity in the native range are distributed across many chromosomes. However, shifts in diversity and differentiation that occur only in a few narrow regions of the genome would suggest evolutionary dynamics specific to that region. If regions that have differentiated (e.g., high F_{ST}) from the native range also show higher diversity in the invasions, this may suggest either a relaxation of purifying selection or an increase in diversifying selection at specific regions of the genome. We note here that the recombination landscape may also contribute to these patterns, and discuss this factor in the following section.

Differentiation of a few genomic regions reveal consequences of invasion

Starling populations colonized their invaded ranges less than two centuries ago, and the age of these populations makes it somewhat surprising to find loci specific to the

derived populations with fixation indices as high as 0.57 in a single 10-kb window (Figure 3). However, no putative outlier windows approach fixation in any comparison across all individuals in a population. We consider only the top 0.1% of 50-kb windows to be F_{ST} outliers (AU vs. UK: $F_{ST} > 0.30$; NA vs. UK: $F_{ST} > 0.22$); subsequent references to outliers indicate these windows. Outliers include Chromosomes 1, 2 and 6, and to a lesser extent, 1A, 4, and 4A. Examining localized shifts in diversity and differentiation at these regions can help to resolve the relative contributions of demography and selection.

Genome scans like this one tend to search for genomic islands of divergence, where recombination can explain divergence among populations³⁶. High levels of differentiation paired with a reduction in diversity may stem from suppressed recombination (e.g., proximity to the centromere or a structural rearrangement), or from alleles approaching fixation or loss due to drift or selection (e.g., a selective sweep). We find that most putative outlier regions are distant from the centromere location (predicted via homology with the zebra finch (*Taeniopygia guttata*) genome, Supplementary Information). Nevertheless, genomic architecture—including linkage disequilibrium independent of the centromere—plays a role in how differentiation among populations is generated. Below, we discuss how demography and selection drive divergence in invasions, considering the context of the starling's genomic architecture.



184
185 Figure 3. A-B) Manhattan plots show 50-kb windowed F_{ST} between AU & UK (A) and
186 NA & UK (B) populations. C-D) 50-kb windowed F_{ST} (top row), π within each population
187 (middle), and Tajima's D within each population (bottom row), centered on the elevated
188 F_{ST} regions of Chromosome 2 (C) and Chromosome 6 (D). Color represents each
189 population, except in F_{ST} plot where yellow indicates F_{ST} AU vs. UK and blue indicates
190 F_{ST} NA vs. UK.

191

192

193

Some F_{ST} outlier regions show clear signals of selection in one or both invasions

If directional selection were driving differentiation between an invading population and its native ancestral population, we would expect to see a decline in nucleotide diversity specific to the invading population. But, as described above, local reductions in π could also result from population bottlenecks experienced during founder events; to clarify the impact of selection, we look Tajima's D . We infer directional selection where most variants are either very rare or very common (e.g., negative Tajima's D). We examine these predictions by looking at two F_{ST} peaks on Chromosome 1A; in this region, we find very low nucleotide diversity and negative Tajima's D in both AU and NA, which is exactly the signature we would expect under purifying selection (Figure S12). Around that region, Tajima's D in the UK population varies between 0 and 1. We note that genetic distance (d_{xy}) could also clarify mechanisms, but as d_{xy} tracks F_{ST} perfectly, we argue that linked selection offers the clearest explanation. This pattern is echoed on Chromosomes 4 and 4A, but the remaining outlier regions show more complicated signals (Figures S13-14).

Where F_{ST} is highest on Chromosome 2, we find strong evidence of both purifying and balancing selection in all three populations (Figure 3C). We find that nucleotide diversity is very low within every population, and immediately after the block of elevated F_{ST} , we see a sharp increase in nucleotide diversity in all three populations (Figure 3C). In addition, Tajima's D tracks these changes in diversity: where F_{ST} and π are low, Tajima's D is also negative in all three populations—indicating an excess of low-frequency variants and perhaps purifying selection—but this statistic climbs to high positive values immediately before and after the block of elevated F_{ST} , indicating

balancing selection. The concordance of Tajima's D in all three populations suggests a release of some kind, whether it is a relaxation of purifying selection or a recombination breakpoint. Even though the centromere is predicted to be 30-Mb downstream of this region, these signatures are consistent with linkage disequilibrium in this 4-Mb region: eukaryotes generally show suppression of recombination near the centromere, leading to a build-up of linkage disequilibrium if this suppression extends for 30-Mb. It is possible that a structural variant in the founding population could generate this pattern. However, we note that F_{ST} among AU and the other populations in fact declines dramatically (to around 0.1) in the middle of the 4Mb region, and in the same location, F_{ST} between NA and UK increases slightly. If this genomic region differentiated as a large linkage block, we would not expect such a decline in F_{ST} and a weakening of selective pressure (as evidenced by the increase in Tajima's D). For these reasons, we suggest that the peak on Chromosome 2 indicates both purifying and balancing selection in the AU invasion.

One region on Chromosome 6 reveals how population expansion could interact with selection

Both invasions have differentiated from the native range in a 4-Mb region of Chromosome 6 (Figure 3). As a preliminary check, we note that the large distance between this F_{ST} peak and the centromere suggests that low recombination is unlikely to explain differentiation. We suggest that the clearest explanation of this peak invokes selection on previously rare variants, based on three lines of evidence.

First, we suggest that rapid population expansion allowed previously rare variants to surf to a higher allele frequency in the invasions. In this 4-Mb region, invasive diversity (π_{AU} and π_{NA}) are each more than three times the native diversity. This shift in within-population diversity is not random; in fact, when we examine invasive nucleotide diversity directly under the F_{ST} peak, we find only three other windows ~4.2-Mb upstream of this peak show invasive diversity that is notably higher than native diversity. This evidence supports the hypothesis that upon establishment, starlings experienced either (1) balancing selection (strong positive Tajima's D) in both invasions due to novel selective pressures or (2) a release of purifying selection that led to an accumulation of variants and thus higher invasive diversity—but only in this specific region. These patterns could be driven by a small number of individuals, or they could indicate a population-wide shift, which leads us to our next point.

Second, in this region, we find that these higher-diversity alleles in both the NA and AU populations have increased in frequency relative to the native range. In the same region, we find strong positive values of Tajima's D in the invasions—indicating a moderate-to-high frequency of the alternative allele—and negative Tajima's D in the UK population at this peak, since these signatures suggest that previously rare variants have increased in frequency in the invasions only. Alternatively—or simultaneously—purifying selection may have driven these same variants to a lower frequency in the UK population. The most parsimonious explanation of these shifts in diversity is a single event in the UK population, but we note that this shift is specific to only a small region of the genome.

Third, and most importantly, these patterns are found in this region of the genome only, and it is notable that this shift in diversity co-occurs with one of the highest F_{ST} peaks. We would expect to see similarly high invasive diversity under other F_{ST} peaks if population expansion alone could explain these patterns. However, nowhere else in the genome do we find such high invasive diversity where native diversity is low. We suggest a selective explanation given that a genetic bottleneck is not likely to produce this pattern. Taken together, these results provide evidence that rapid expansion of these starling invasions may have facilitated selection to drive previously rare variants higher in frequency, independently in both NA and AU populations.

Genes under putative selection may aid in invasion success

The region under putative selection on Chromosome 6 overlaps with the coding regions of four genes with dramatically different functions (*JMJD1C*, *RTKN2*, *NRBF2*, and *ARID5B*), and we suggest that selection on one of these genes might explain the differentiation in the region, with the other genes remaining in linkage disequilibrium with the possible candidate. Among these candidates, we can speculate that *ARID5B* has the most intuitive link to hypothesized selective pressures: this protein is required for adipogenesis and involved in smooth muscle differentiation. The first exon of this gene lies directly under the F_{ST} peak between AU and UK starlings, and muscle growth and fat storage may have been key to dispersal ability. The three other genes that overlap this window are involved in the DNA-damage response (*JMJD1C*), lymphopoiesis (*RTKN2*), and regulating autophagy (*NRBF2*). For details on GO term enrichment in

these outlier regions, see SI Section 6. Regardless of the mechanism driving these loci toward an intermediate frequency, it remains possible that variation at one or more of these loci influenced invasion success by maintaining heterozygosity.

Conclusion

An open question in invasion biology is whether an invasive population's success is better attributed to intrinsic properties of the invasive species or to extrinsic factors specific to the novel environment. These results demonstrate how demographic shifts during and immediately after any establishment may support a species' success and even lead to local adaptation to environmental conditions. The Australian and North American starling invasions colonized each continent around the same time, and experienced similar contractions in population size that led to classical founder events upon establishment. The shared decline in genetic diversity represents a shared intrinsic determinant of invasion success. However, differences in propagule pressure and dispersal likely influence the evolutionary trajectory of each population^{37,38}. Propagule pressure (also termed introduction effort) is a composite measure of the number of individuals released, and we note that founding population sizes may have varied slightly among AU and NA introductions. Even without dramatic variation in founding population sizes, dispersal itself shapes genetic diversity and thus adaptive potential, as shown in a recent study of invasive plants³⁹. Starlings' dispersal and migration varies among populations: starlings in Australia tend to migrate less frequently and across shorter distances, but starlings migrate and/or disperse hundreds of kilometers in North America^{19,40} and South Africa⁴¹.

It is remarkable that despite this contrast in life history strategies and the stochastic nature of evolution during range expansion, we find F_{ST} peaks shared among invasions at only a few regions of the starling genome. Although these F_{ST} peaks could arise via drift, footprints of other population genetic metrics are consistent with selection. We note that mutations in specific chromosomal regions could also be accelerated by extrinsic environmental properties (climate, food availability, and more) through epigenetic CpG DNA hypermethylation events, which are known to increase frequency of genetic mutations by spontaneous deamination CG>TG transition⁴²⁻⁴⁴. For example, such epigenetic shifts supported the invasion of another avian species (the house sparrow) into Australia⁴⁵. Regardless of genetic mechanism, we suggest that differentiation in these genetic regions is simultaneously shaped by intrinsic and extrinsic drivers of invasion success. We find it notable that some differentiated regions (in particular, Chromosome 6) are shared among invasions despite differences in the selective environment as well as stochastic processes that shape the starling's evolution on each continent. The European starling invasions compared here suggest that rapid population growth may support local adaptation.

Methods

Whole-genome re-sequencing

Libraries for each individual starling were constructed using a TruSeq DNA PCR-free High Throughput Library Prep Kit (Illumina, San Diego, CA). All individuals passed the initial quality check with FASTQC (Babraham Bioinformatics, Cambridge, UK). Adapters were removed using AdapterRemoval⁴⁶ and reads mapped to the reference S.

vulgaris vNA genome (GCF_001447265.1)⁴⁷ genome using BOWTIE2⁴⁸ and checked for mapping quality using qualimap⁴⁹. Sequencing quality was relatively high: 96.4% of reads mapped to the *S. vulgaris* genome with a coverage of 18.4X and a mapping quality of 26.9 (Table S1). Reads were also mapped to a pseudo-chromosome-level *S. vulgaris* vNA genome, where scaffolds were assigned to chromosomes based on orthology to the zebra finch reference genome (GCF_000151805.1)⁵⁰. Assuming orthology, we were able to predict centromere positions based on the known genomic architecture of the zebra finch⁵¹, but this study does not directly define centromere position.

We called variants using GATK's HaplotypeCaller in GVCF mode and flagged low-quality variants using GATK Best Practices (QD<2, FS>60, MQ<40, and SOR>3, accessed March 21, 2018⁵²). We filtered sites for missing data, depth, and quality using vcftools (parameters: --max-missing 0.8 --min-meanDP 2 --max-meanDP 50 --remove-filtered-all), which removed 4.1 million sites from the original SNP set and left a total of 23.4 million sites for downstream analyses. Starting from the mapped reads used in the GATK pipeline, we also called SNPs based on a minimum p-value of the correct genotype probability at each site using ANGSD^{34,53}. Filtering for SNP p-value (0.0001), depth (between 60 and 400 sequences), and mapping quality (>20) left 16,151,007 sites. All scripts used in read processing and filtering are available on GitHub:

<https://github.com/nathofme/global-RESEQ>.

Population structure

Population structure analyses used a dataset of biallelic SNPs in Hardy-Weinberg Equilibrium, where minor allele count (MAC) is > 2 and SNPs were pruned for LD by removing all sites with an $r^2 > 0.6$ within 1kb sliding windows. This filtering left 868,685 sites. Since some individuals showed much lower coverage (minimum 5.58X), all tests of population structure were run with both variant-called (GATK) and genotype probability (ANGSD) datasets. Scripts for variant-called analyses are stored at <https://github.com/nathofme/global-RESEQ/blob/master/filter-scan.sh>, and scripts for probability-based analyses at <https://github.com/nathofme/global-RESEQ/blob/master/angsd.sh>.

We estimated variance among and between individuals using a principal components analysis in SNPRELATE⁵⁴ (GATK) and a covariance matrix built in ngsTools⁵⁵ (ANGSD). We used ADMIXTURE⁵⁶ (GATK) and NGSADMIX⁵⁷ (ANGSD) to examine shared ancestry among individuals, and we also measured pairwise genetic distances using ngsDist for the ANGSD dataset only.

Demographic inference

We used FASTSIMCOAL2³⁵ to explicitly test for genetic bottlenecks in each population. FASTSIMCOAL2 takes a site-frequency spectrum (SFS) as input, and we used the SFS estimated from ANGSD given that likelihood-based estimates are more robust to sequencing error⁵⁸. Demographic models in *fastsimcoal2* used priors on the time (TBOT = 10 to 300) and size of the bottleneck (NBOT = 10 to 1000). The command line arguments were as follows: -M -n 1000000 -L 50 -q -k 100000. Each run began with a randomly generated seed (-r), and the -k flag simply writes polymorphic sites to a

temporary file to cope with the high memory usage of this analysis. Scripts for demographic analyses can be found at <https://github.com/nathofme/global-RESEQ/blob/master/demography.sh>.

To verify this demographic model, we also estimate inbreeding coefficients (F -statistics) using the `-het` command in VCFTOOLS, and calculate relatedness among each pair of individuals using the `-relatedness` method of Yang et al. (2010) in VCFTOOLS.

Sliding window scans

For scans of genetic divergence and diversity, we used a variant dataset filtered only for depth and quality: we kept variants that had less than 20% missing data across all sites, but did not apply minor allele frequency (MAF) filters, filter for HWE, linkage, or any other factors. Given that rare alleles likely provide the strongest evolutionary signals in this system, we did not want to filter out any alleles that might have been rare in one population (e.g., the native UK population) and increased in frequency in another population (e.g., the AU or USA invasions). Nevertheless, we test for sensitivity to this filtering choice in SI Section 2B.

We calculated F_{ST} and nucleotide diversity (π) using overlapping 50-kb sliding windows with a step size of 10kb using VCFTOOLS⁵⁹. We calculate nucleotide diversity separately for each population. We then calculate d_{xy} using a Python script by Simon Martin (accessed at https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py on February 12, 2020). This analysis includes all confident variant calls (parameters: `--output-mode EMIT_ALL_CONFIDENT_SITES` in GATK's HaplotypeCaller); only with

this modification do we recover levels of d_{xy} similar to other systems. We also measured F_{ST} and π in overlapping 10-kb windows to localize elevated F_{ST} to an even smaller region of the genome. To visualize relationships between diversity metrics, we plotted the mean values of each metric in a 50-kb window. All scripts for plotting are stored on GitHub: <https://github.com/nathofme/global-RESEQ/>.

Identifying candidate genes under selection

In contrast to identifying single genes, network analyses of gene ontology (GO) terms can provide a more holistic and objective method of identifying shared functions. Network analyses can also provide more statistical power, correcting for the usual problem of multiple testing. To identify functions of candidate regions, we quantified the uniqueness and dispensability of each GO term using a method that quantifies semantic similarity⁶⁰. This analysis by default emphasizes GO terms that are rare in the list of candidates provided; because we are interested in functions that are common across outlier regions, we manually curate category labels to choose GO terms that are less unique and more dispensable as representative titles.

Acknowledgments

SLM acknowledges Roslin Institute Strategic Grant funding from the UK Biotechnology and Biological Sciences Research Council (BB/P013759/1). HR acknowledges that sequencing costs were supported by a fellowship from The Winston Churchill Memorial Trust, and a Phyllis and Eileen Gibbs Fellowship from Newnham College, Cambridge. LAR was supported by a Scientia Fellowship from UNSW. This project was also

supported by discussions at Bird Sense meetings funded by the Royal Society of London, the Association for the Study of Animal Behaviour, and the Zoological Society of London. NRH acknowledges bioinformatic support from members of the Fuller Evolutionary Biology Program at the Cornell Lab of Ornithology, including Irby Lovette, Jennifer Walsh, and Leonardo Campagna.

Author Contributions

Project conception: all authors
Field work: Melissa Bateson, Lee Ann Rollins, Adam Cardilini, Scott J Werner
Lab work: Wes Warren and the McDonnell Group at WUSTL.
Data analysis and manuscript writing: Natalie R Hofmeister, Katarina Stuart
Manuscript editing: all authors

References

1. Russell, J. C. & Blackburn, T. M. The Rise of Invasive Species Denialism. *Trends in Ecology & Evolution* **32**, 3–6 (2017).
2. Ricciardi, A. & Ryan, R. The exponential growth of invasive species denialism. *Biological Invasions* **20**, 549–553 (2017).
3. Sagoff, M. Fact and value in invasion biology. *Conservation Biology* **34**, 581–588 (2019).
4. Davis, M. A. Let's welcome a variety of voices to invasion biology. *Conservation Biology* **cobi.13608–6** (2020). doi:10.1111/cobi.13608
5. Lowe, W. H., Kovach, R. P. & Allendorf, F. W. Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution* **32**, 141–152 (2017).
6. Kolbe, J. J. *et al.* Genetic variation increases during biological invasion by a Cuban lizard. *Nature* **431**, 177–181 (2004).
7. Dlugosch, K. M. & Parker, I. M. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* **17**, 431–449 (2008).

8. Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M. & Reimchen, T. E. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. ecol. evol.* 1–14 (2018). doi:10.1038/s41559-018-0581-8
9. Gralka, M. *et al.* Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters* **19**, 889–898 (2016).
10. White, T. A., Perkins, S. E., Heckel, G. & Searle, J. B. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Mol Ecol* **22**, 2971–2985 (2013).
11. Tsutsui, N. D., Suarez, A. V., Holway, D. A. & Case, T. J. Reduced genetic variation and the success of an invasive species. *Proc Natl Acad Sci USA* **97**, 5948–5953 (2000).
12. Verhoeven, K. J. F., Macel, M., Wolfe, L. M. & Biere, A. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2–8 (2011).
13. Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T. & Christie, M. R. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Mol Ecol* **17**, 675 (2018).
14. Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A. & Gillette, H. D. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Mol Ecol* **24**, 2095–2111 (2015).
15. Lu-Irving, P., Marx, H. E. & Dlugosch, K. M. ScienceDirect Leveraging contemporary species introductions to test phylogenetic hypotheses of trait evolution. *Current Opinion in Plant Biology* **42**, 95–102 (2018).
16. Bock, D. G., Kantar, M. B., Caseys, C., Matthey-Doret, R. & Rieseberg, L. H. Evolution of invasiveness by genetic accommodation. *Nat. ecol. evol.* **2**, 991–999 (2018).
17. Linz, G. M., Homan, H. J., Gaulker, S. M., Penry, L. B. & Bleier, W. J. *European starlings: a review of an invasive species with far-reaching impacts. Managing Vertebrate Invasive Species* 24 (2007).
18. Forbush, E. H. *The Starling*. (Wright and Potter Printing Co., 1915).
19. Kessel, B. Distribution and migration of the European Starling in North America. *The Condor* **55**, 49–67 (1953).
20. Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R. & Sherwin, W. B. Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Mol Ecol* **18**, 1560–1573 (2009).
21. Feare, C. *The Starling*. (Oxford University Press, 1984).
22. Stuart, K. C. *et al.* Signatures of selection in a recent invasion reveals adaptive divergence in a highly vagile invasive species. *Mol Ecol* (2020). doi:10.1101/mec.15601
23. Cardilini, A. P. A., Buchanan, K. L., Sherman, C. D. H., Cassey, P. & Symonds, M. R. E. Tests of ecogeographical relationships in a non-native species: what rules avian morphology? *Oecologia* **181**, 783–793 (2016).
24. Cardilini, A. P. A. *et al.* Environmental Influences on Neuromorphology in the Non-Native Starling *Sturnus vulgaris*. *Brain Behav Evol* **92**, 63–70 (2018).

25. Bodt, L. H., Rollins, L. A. & Zichello, J. M. Contrasting mitochondrial diversity of European starlings (*Sturnus vulgaris*) across three invasive continental distributions. *Ecol Evol* **10**, 10186–10195 (2020).
26. Hofmeister, N. R., Werner, S. J. & Lovette, I. J. Environmental correlates of genetic variation in the invasive European starling in North America. *Mol Ecol* **30**, 15806–33 (2021). doi:10.1111/mec.15806
27. Dolbeer, R. A. Migration patterns of age and sex classes of Blackbirds and Starlings. *Journal of Field Ornithology* **53**, 28–46
28. Werner, S. J., Fischer, J. W. & Hobson, K. A. Multi-isotopic ($\delta^2\text{H}$, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. *PLoS ONE* **15**, e0237137 (2020).
29. Ball, G. F. & Wingfield, J. C. Changes in plasma levels of luteinizing hormone and sex steroid hormones in relation to multiple-broodedness and nest-site density in male starlings. *Physiological Zoology* **60**, 191–199 (1987).
30. Dawson, A. Plasma gonadal steroid levels in wild starlings (*Sturnus vulgaris*) during the annual cycle and in relation to the stages of breeding. *General and Comparative Endocrinology* **49**, 286–294
31. Briski, E. *et al.* Beyond propagule pressure: importance of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment* **16**, 345–353 (2018).
32. Irwin, D. E. *et al.* A comparison of genomic islands of differentiation across three young avian species pairs. *Mol Ecol* **511**, 83–17 (2018).
33. Hofmeister, N. R., Werner, S. J. & Lovette, I. J. Environment but not geography explains genetic variation in the invasive and largely panmictic European starling in North America. **295**, 254–32 (2019).
34. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
35. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet* **9**, e1003905–17 (2013).
36. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**, 3133–3157 (2014).
37. Simberloff, D. The Role of Propagule Pressure in Biological Invasions. *Annu. Rev. Ecol. Evol. Syst.* **40**, 81–102 (2009).
38. Williams, J. L., Hufbauer, R. A. & Miller, T. E. X. How Evolution Modifies the Variability of Range Expansion. *Trends in Ecology & Evolution* 1–11 (2019). doi:10.1016/j.tree.2019.05.012
39. Smith, A. L. *et al.* Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proc Natl Acad Sci USA* **117**, 4218–4227 (2020).
40. Burt, H. E. & Giltz, M. L. Seasonal directional patterns of movements and migrations of starlings. *Bird-Banding* **48**, 259–271 (1977).
41. Berthouly-Salazar, C. *et al.* Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Mol Ecol* **22**, 5793–5804 (2013).

42. Sved, J. & Bird, A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci USA* **87**, 4692–4696 (1990).
43. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA* **103**, 1412–1417 (2006).
44. Simmen, M. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics* **92**, 33–40 (2009).
45. Sheldon, E. L., Schrey, A., Andrew, S. C., Ragsdale, A. & Griffith, S. C. Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. *Royal Society Open Science* **5**, 172185 (2018).
46. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* 1–7 (2016). doi:10.1186/s13104-016-1900-2
47. Stuart, K. C. *et al.* Transcript- and annotation-guided genome assembly of the European starling. *bioRxiv* (2021). doi:10.1101/2021.04.07.438753
48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
49. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **31**, 566–567 (2015). doi:10.1093/bioinformatics/btv566
50. Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151 (2010).
51. Knief, U. & Forstmeier, W. Mapping centromeres of microchromosomes in the zebra finch (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma* 1–12 (2016). doi:10.1007/s00412-015-0560-7
52. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
53. Kim, S. Y. *et al.* Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics* **12**, 231 (2011).
54. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
55. Fumagalli, M., Vieira, F. G., Linderöth, T. & Nielsen, R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487 (2014).
56. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 1–6 (2011).
57. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* **195**, 693–702 (2013).
58. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLoS ONE* **7**, e37558–11 (2012).

- 588 59. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–
589 2158 (2011).
- 590 60. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO Summarizes and
591 Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).
592