

Historical museum samples enable the examination of divergent and parallel evolution during invasion

Katarina C. Stuart¹  | William B. Sherwin¹  | Jeremy J. Austin²  | Melissa Bateson³  | Marcel Eens⁴  | Matthew C. Brandley⁵  | Lee A. Rollins¹ 

¹School of Biological, Earth and Environmental Sciences, Evolution & Ecology Research Centre, UNSW Sydney, Sydney, New South Wales, Australia

²School of Biological Sciences, Australian Centre for Ancient DNA (ACAD), University of Adelaide, Adelaide, South Australia, Australia

³Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK

⁴Behavioural Ecology and Ecophysiology Group, Department of Biology, University of Antwerp, Wilrijk, Belgium

⁵Section of Amphibians and Reptiles, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania, USA

Correspondence

Katarina C. Stuart, School of Biological, Earth and Environmental Sciences, Evolution & Ecology Research Centre, UNSW Sydney, Sydney, NSW, Australia.
Email: Katarina.Stuart@unsw.edu.au

Funding information

UNSW Scientia Fellowship

Abstract

During the Anthropocene, Earth has experienced unprecedented habitat loss, native species decline and global climate change. Concurrently, greater globalization is facilitating species movement, increasing the likelihood of alien species establishment and propagation. There is a great need to understand what influences a species' ability to persist or perish within a new or changing environment. Examining genes that may be associated with a species' invasion success or persistence informs invasive species management, assists with native species preservation and sheds light on important evolutionary mechanisms that occur in novel environments. This approach can be aided by coupling spatial and temporal investigations of evolutionary processes. Here we use the common starling, *Sturnus vulgaris*, to identify parallel and divergent evolutionary change between contemporary native and invasive range samples and their common ancestral population. To do this, we use reduced-representation sequencing of native samples collected recently in northwestern Europe and invasive samples from Australia, together with museum specimens sampled in the UK during the mid-19th century. We found evidence of parallel selection on both continents, possibly resulting from common global selective forces such as exposure to pollutants. We also identified divergent selection in these populations, which might be related to adaptive changes in response to the novel environment encountered in the introduced Australian range. Interestingly, signatures of selection are equally as common within both invasive and native range contemporary samples. Our results demonstrate the value of including historical samples in genetic studies of invasion and highlight the ongoing and occasionally parallel role of adaptation in both native and invasive ranges.

KEYWORDS

adaptation, divergent evolution, invasion, parallel evolution, *Sturnus vulgaris*

1 | INTRODUCTION

The ecological and economic impacts of invasive species are a growing concern in our globalized world. Increased intercontinental travel and trade is giving rise to new or reinforced invasion pathways (Turbelin et al., 2017), resulting in a great number of alien species becoming established and spreading within novel ranges (Hulme, 2009). The financial cost of invasive species within Australia is estimated to be in excess of \$13 billion annually (Hoffmann & Broadhurst, 2016). With habitat clearing and climate change expected to favour invasive species over native ones, the environmental and financial costs of invasive species are only expected to rise in the future (Dukes & Mooney, 1999). Many studies of invasive species' success involve examining evolutionary changes following introduction and focus on rapid adaptation to novel environments (Prentis et al., 2008). This information is vital for long-term management of invasive populations.

Understanding evolutionary trends across a species' native and invasive ranges will help determine important adaptive elements that aid species' persistence in a changing world. Species that are invasive present a contrariety when they face population decline within their native range (Bishop, 2011; Delibes-Mateos et al., 2009; Erfmeier & Bruelheide, 2010; Rogers et al., 2006). Research efforts should tackle ecological questions of conservation and invasion management concurrently, enabling us to understand how and why patterns of adaptation in a species' native and invasive populations may differ. It is possible that the translocation and establishment process itself may select for traits that enable an individual to overcome otherwise detrimental environmental instability or other novel stressors, increasing general fitness (Callaway & Ridenour, 2004; Liu & Trumble, 2007). Understanding how the invasion process may induce differences in population persistence is made even more pressing by the increasing anthropogenic impact on the natural world, including ongoing land alteration, environmental contamination and human-induced climate change (Hellmann et al., 2008).

Often, these adaptive changes are identified through contrasting present-day native and invasive populations (Hofmeister, Stuart, et al., 2021). However, such approaches exclude the temporal element of species' change, so that such studies assume native populations have not changed since the founders of the invasive population were collected. This would then lead to the conclusion that all similarities between native and invasive populations result from a common ancestral population and are not due to parallel change since separation. However, with global anthropogenic change impacting the natural world, it is reasonable to assume that altered or increased selection regimes have arisen during the post-industrialized world, shaping species worldwide (Siepielski et al., 2017; Sokolova & Lannig, 2008). Historical specimens therefore provide an unparalleled tool to better contextualize divergent vs. parallel evolution, providing phenotypic and, more recently, genotypic information that can be used to identify temporal changes in species ranges and traits (Ewart et al., 2019; Lopez et al., 2020). Studies focusing on rapid local adaptation in invasive species may now make use of historical DNA

alongside contemporary samples to understand the selective forces shaping both invasive and native ranges concurrently.

The common or European starling, *Sturnus vulgaris*, presents an ideal system to use historical samples to investigate both divergent and parallel genetic change within an invasive species. The European starling (hereafter starling) is a highly invasive pest, introduced and successfully establishing on every other continent except Antarctica (Higgins et al., 2006). Despite this, native range starlings are themselves a conservation focus, with declines of more than 50% in some countries (Versluijs et al., 2016) putatively associated with shifts in farming practice that are common in their native range (Freeman et al., 2007; Heldbjerg et al., 2016). Fortunately, due to the historical popularity of collecting bird skins, historical starling samples may be found scattered across many museums and institutions in both their native range and within invaded countries. These skins serve as untapped reserves of genetic information, which may be used to track temporal genetic changes across the native range, reveal information regarding historical population structure and provide context that enables us to better understand current patterns of native range starling decline.

Starlings present a prime example of how the combination of data from invasive, native and historical populations can clarify our understanding of evolution in both native and invasive contexts. Introduced deliberately and repeatedly by acclimatization societies into several Australian coastal cities during the 1860–1980s (Figure 1), the starlings' range now stretches across the continent's eastern and southern coasts (Long, 1981). Genetic analyses support strong population substructuring across the invasive Australian range (Rollins et al., 2009, 2011), with reduced-representation sequencing data indicating the two main subpopulations probably resulted from allelic differences in founding populations at different introduction sites (Stuart, Cardilini, et al., 2021). The historical specimens available for this species were collected within 15 years of the earliest documented introductions to Australia in 1856 from the same native range location (around London, UK) (Long, 1981), providing a snapshot of native starling populations at the time when founders were transported to Australia.

To better understand patterns of population structure and signatures of selection present in the invasive Australian range, we used a reduced-representation sequencing approach to compare contemporary Australian (AU) and native range (United Kingdom, UK; Belgium, BE) starlings to historical UK samples collected during the period when the Australian founders were collected. Moreover, this project explores proximate drivers of invasive species' evolution in the face of novel selection provided by new environments. Specifically, we compare population structure of native and invasive contemporary starling samples, and we explore genomic divergence between contemporary and historical samples and assess the putatively adaptive capacity of these genomic changes. Finally, we use historical samples as a basis of comparison to determine genomic regions of parallel change in both the contemporary native and invasive populations to better understand global shifts in selective forces.

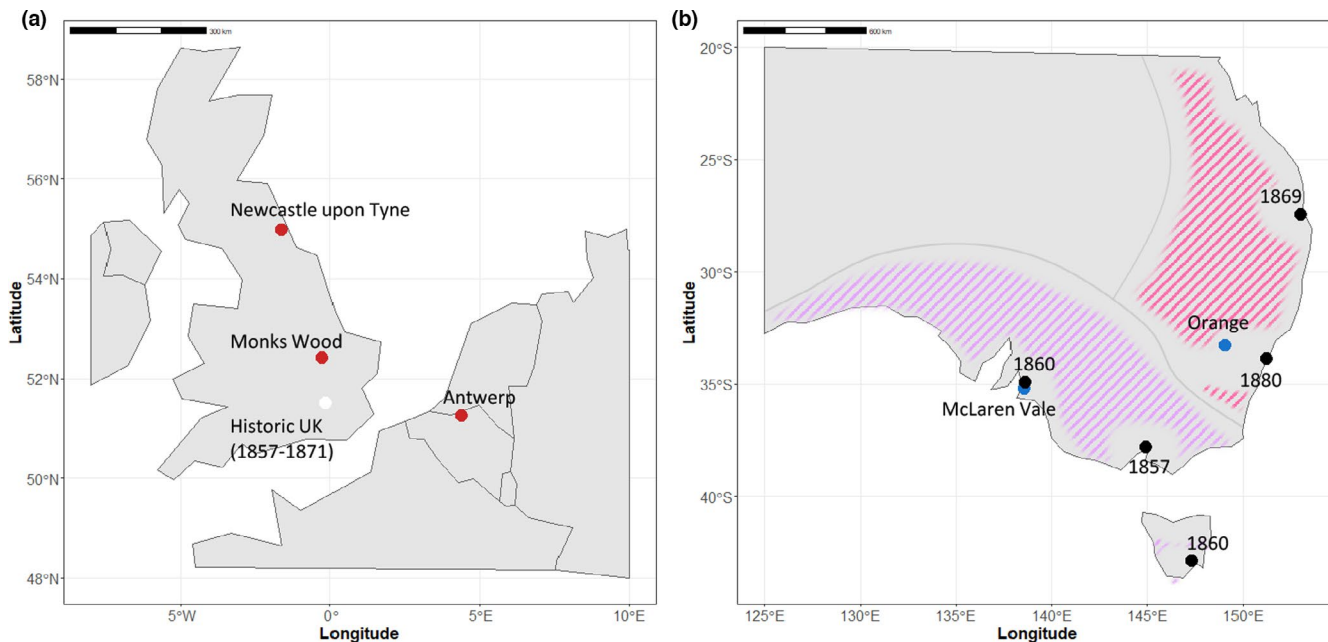


FIGURE 1 Geographical distribution of the *Sturnus vulgaris* collection sites in the United Kingdom and Belgium (native range, red points), Australia (invasive range, blue points) and historical samples (white point). The coloured shading on the Australian map denotes their Australian range, broken up into the two main subpopulations. Introduction sites are marked in black on the Australian map, with first introduction year listed adjacently

2 | METHODS

2.1 | Sample collection and extraction (historical starlings)

We sourced historical starling specimens (HS) from the Natural History Museum (NHM) in Tring, UK ($N = 15$). Historical samples were selected on the basis of sampling location (in the vicinity of London, thought to be where Australian founders were sourced; Jenkins, 1977), sample quality, completeness of the collection record and sample collection date (samples collected from 1857 to 1871, during the period when the Australian introductions took place; Higgins et al., 2006; Table S1), and specimen age (adult) (samples were of mixed sex).

DNA extractions of historical samples were conducted in a specialist ancient DNA laboratory at the Australian Centre for Ancient DNA, University of Adelaide. We rehydrated 3–4 mm³ dried tissue in 1 ml of 0.5 M EDTA for 2 h and extracted DNA using a Qiagen DNeasy Tissue Kit (Qiagen) as per the manufacturer's instructions. DNA was eluted twice with 40 μ l of EB buffer (+0.05% Tween-20) for a final elution volume of 80 μ l.

2.2 | Sample collection and extraction (contemporary starlings)

We sourced contemporary native range starling samples from two UK locations: Monks Wood (MW: $N = 15$, blood, ~100 km from London, where starlings were thought to be sourced for relocation to colonies

such as Australia) and rural sites around the city of Newcastle upon Tyne (NC: $N = 15$, blood, ~300 km to the north of Monks Wood), and one Belgian (BE) location in Antwerp (AW: $N = 15$, blood, ~350 km to the southeast of Monks Wood) (Figure 1a). We sourced contemporary Australian samples from two locations, previously shown to represent two separate populations within the Australian invasive range (Stuart, Cardilini, et al., 2021): McLaren Vale in South Australia (MV: $N = 15$, blood) and Orange in New South Wales (OR: $N = 15$, muscle tissue) (Figure 1b). Contemporary DNA extractions were performed using the Qiagen Genra Puregene Tissue kit as per the manufacturer's instructions.

2.3 | Sequencing and genome variant calling

We sequenced 75 contemporary and 15 historical samples using the DArTseq protocol (Kilian et al., 2012), using a restriction enzyme double digest of *Pst*I–*Sph*I. The sequencing was conducted on a HiSeq 2500, producing 312,907,523 single-end reads of raw data across the 90 samples (26,408,649 across the 10 successfully sequenced historical samples; Table S1).

We used the STACKS version 2.2 (Rochette et al., 2019) pipeline (Rochette & Catchen, 2017) to process the DArTseq raw data. We used the *process_radtags* function to clean the tags, discarding reads of low quality (-q), removing reads with uncalled bases (-c) and rescuing barcodes and radtags (-r). We used the Burrows–Wheeler aligner (BWA) version 0.7.15 (Li & Durbin, 2009) *aln* function to align the read data to the reference genome *S. vulgaris* vAU1.0 (Stuart, Edwards, et al., 2021). Using FASTQC, we identified

base sequence bias in the adapter region, and so the first five bases were trimmed (-B 5) during alignment. The reads were then processed through *BWA samse* and *SAMTOOLS* version 1.10 (Li et al., 2009), before single nucleotide polymorphism (SNP) variants were called through *STACKS gstacks* (default parameters) and then *populations* (parameter information below).

We produced an unfiltered SNP data set by running *STACKS populations* with no parameter thresholds specified. We used this data set to produce the unfiltered loci and site counts (Table S2), split the data into three separate files for further assessment of sequencing data (contemporary native range samples: MW, NC and AW; Australian samples: OR and MV; and historical samples: HS). We calculated variant base substitutions in *vcfstats* version 0.0.5 (Lindenbaum, 2015) and variant density mapped along the reference genome scaffolds using *SAMTOOLS bedtools* function (window size 1,000,000 bp). We used the *DARTR* version 1.1.11 (Gruber et al., 2018) function *glPlot* to create a smear plot of the mapped variant data across individuals and genomics sites. This resulted in a population genetic file of 239,538 SNPs. We also passed the raw single-end read data through the *STACKS BWA-MEM*, and the *BOWTIE-GATK* variant calling pipeline (Appendix S1), to compare the quantity of site variant data that was successfully mapped and assess how these alternative variant calling approaches performed for reduced representation sequencing data sets that contained degraded historical DNA.

We generated a “population genetics” variant file by running *STACKS populations*, filtering for a minimum per-population site call rate of 50% (-r 0.5), a minimum populations per-site of 2 (-p 2) and a minimum loci log likelihood value of -15 (-ln_{lim} -15), with one random SNP per tag retained (--write_random_snp). We used *vcftools* version 0.1.16 (Danecek et al., 2011) to filter the following parameters: maximum missingness per site of 10% (-max-missing 0.9), minor allele frequency of 2.5% (MAF; --maf 0.025), minimum loci depth of 2 (--minDP 2), minimum genotype quality score of 15 (--minGQ 15) and site Hardy-Weinberg equilibrium exact test minimum *p* value of .001 (--hwe 0.001). We chose a high threshold for missingness to not bias the population genetics analysis against the historical samples, which had much higher levels of missingness than the contemporary samples. MAF filtering helps remove misreads, and HWE filtering removed highly non-neutral loci, both of which are important for capturing neutral population substructure. After filtering, we calculated individual relatedness, and closely related individuals were removed so that there was only one representative from each cluster in the final data (Figure S1; five Monks Wood, five Newcastle and two Orange individuals removed). This resulted in a population genetic variant file of 3,840 SNPs used in the Section 2.4.

We generated a “selection” variant file by using *STACKS populations* to align the raw reads for all samples (with --ln_{lim} -15 --write_random_snp flags) and then used *vcftools* to filter out only SNPs present in at least 50% of the historical individuals (i.e., in at least five historical individuals), with additional quality filtering (--minGQ 15 --minDP 2), resulting in 12,219 SNP sites. Only these

sites were then retained to filter the original *populations* variant file, along with an MAF minimum of 2.5% to remove possible sequencing errors. This produced a data set that retained only SNPs sequenced in at least half the historical individuals, which would be necessary for selection analysis. We filtered the selection variant file to form five pairwise population SNP data comparisons: UK-HS (UK populations MW, NC, BE and 10 historical individuals); AU_{east}-HS (AU population OR, and 10 historical individuals); AU_{south}-HS (AU population MV, and 10 historical individuals); UK-AU_{east} (UK populations MW, NC, BE, and AU population OR); and UK-AU_{south} (UK populations MW, NC, BE, and AU population MV). While the native range population may contain a mix of resident and migratory individuals, because we see minimal population structure in the native range and very small *F_{ST}* values (0.003–0.008) we decided to include all contemporary native range samples in this analysis. Conversely, population genetics data from this paper (see Section 3.1) and Stuart, Cardilini, et al. (2021) support the existence of distinct subpopulations within Australia due to historical demographic processes, and hence these subpopulations were treated as separate populations for this analysis (AU_{east} and AU_{south}). Within each of these five variant file subsets, we retained SNP sites present in at least five contemporary individuals, because the file was already filtered for loci present in at least 50% (5/10) of historical individuals. This relatively lenient filtering was necessitated by the smaller number of genomic sample sites produced by the degraded DNA, and ensured minimal loss of the historical information that was present. This resulted in five pairwise population files used in the Section 2.5: UK-HS (4,997 SNPs), AU_e-HS (4,900 SNPs), AU_s-HS (4,961 SNPs), UK-AU_e (4,907 SNPs) and UK-AU_s (4,942 SNPs).

2.4 | Population structure analysis

We analysed the population genetics variant file in several ways to examine the population structure and differentiation across the native and invasive ranges, and between contemporary and historical native ranges. We used R version 3.5.3 (R Core Team, 2017) to run the *SNPRELATE snpgdsPCA* function to create a principal components analysis (PCA) of the loci. We used *ADMIXTURE* version 1.3.0 (Alexander et al., 2009) to determine individual ancestry proportions for each of the following three sample subsets: all samples, contemporary native range and historical, and contemporary Australian. We calculated marginal likelihood for model complexity (*K*, number of genetically distinct sources) 1–8 by averaging over 25 runs, and admixture proportion (*Q*) profiles were generated by *CLUMPAK* (Kopelman et al., 2015) (run on default settings) to obtain an average *Q* profile. We used the *STAMPP* function *stamppFst* and *stamppNeisD* to calculate pairwise *F_{ST}* and Nei's genetic distance (Nei, 1972) between sampling locations. Finally, we used the *DARTR* function *gl.tree.nj* to visualize the phylogeny of the six sampling groups (two contemporary Australian, three contemporary native range, and one historical sampling group).

2.5 | Selection analysis

To obtain outlier SNPs putatively under selection between populations, the five pairwise selection analysis files were examined using two approaches.

When looking for diversifying selection, allele frequency-based approaches are often used, such as *BAYESCAN* version 2.1 (Foll & Gaggiotti, 2008). *BAYESCAN* aims to identify SNPs subject to natural selection by assigning a per-site posterior probability estimated by comparison of explanatory models with and without selection. We conducted *BAYESCAN* SNP outlier analysis, with prior odds for the neutral model set to 10 ($-pr_odds$ 10), and a false discovery rate (FDR) of 0.05 (Figure S2).

Degraded historical DNA from older museum skins is known to have bias towards low-diversity SNPs (Ewart et al., 2019), which may impact single site approaches to examining divergent selection. Therefore, we used an F_{ST} sliding window approach to identify genomic regions of putative diversifying selection. The *vcftools* *weirfst-pop* function was used to analyse weighted F_{ST} in 900,000-bp windows (10,000 window step). We chose this window size primarily based on the ratio of variants to genome size, with the chosen windows putatively spanning three to four variants, with small step sizes then allowing shifts in F_{ST} patterns to be pinpointed more exactly. We selected site windows that reported a weighted F_{ST} value in the top 99th percentile for each pairwise population selection analysis, and analysed the F_{ST} of SNPs within these windows in rank order. As putative outlier SNPs, we retained any SNP within these outlier windows that lay above an F_{ST} threshold relevant for each pairwise data set (this was determined visually as a plateauing of the ranked F_{ST} values, see Figure S3).

From the two above identification processes, we pooled putative outlier SNPs within each pairwise population comparison, to be used for further variant analysis.

2.6 | Variant analysis and annotation

SNPs that were reported as outliers across either the UK-HS or AU-HS data set, as well as the UK-AU data set, were designated as sites under divergent selection. SNPs that were reported as outlier across both UK-HS and AU-HS data sets (but not divergent between UK-AU) were designated as sites under parallel selection. The remaining outlier SNPs we identified only in one data set: UK-HS (putative UK selection), AU-HS (putative AU selection) or UK-AU (putative UK-AU divergence). Classification of SNPs was conducted separately on the AU_{east} and AU_{south} Australian subpopulations, before being pooled. Further details are given in Section 3.3.

We analysed these five groups of SNPs for their functional roles and the nature of the mutation. We completed SNP analyses primarily using *VARIANT EFFECT PREDICTOR* (McLaren et al., 2016), using the genome annotation version released alongside the *S. vulgaris* vAU1.0 assembly (Stuart, Edwards, et al., 2021) to examine the functional consequences of the SNPs (processed to exclude multiple isoforms

using *AGAT* *agat_sp_keep_longest_isoform.pl*; Dainat, 2020). We used *BEDTOOLS* and the *AGAT* function *agat_sp_functional_statistics* to extract genes and transcripts that overlapped the putative loci under selection, and extracted gene ontology (GO) terms. We used *REVIGO* (Supek et al., 2011) to visually summarize GO terms, and we calculated allele frequencies at SNP sites using *BEDTOOLS*.

Lastly, to test if there was an overrepresentation of SNPs located on the macrochromosomes (>20 Mb, as described in Backström et al., 2010), microchromosomes, or the Z sex chromosome, we used a Chi-square test to examine the frequencies of these SNP types across four different SNP groupings: the divergent SNPs, the parallel SNPs, the remaining SNPs under selection, and the SNPs that were not flagged as being under selection in any of the pairwise data sets. We analysed these data using the *chisq.test()* function in R. To ensure results from this analysis were not artefacts of the data (due to diploid variant calling on hemizygous ZW females), we considered the sex of samples in our analysis. Because some individuals in this study were not sexed morphologically, we bioinformatically sexed individuals by assessing the inbreeding coefficient on the autosomes and Z sex chromosome, and conducted an analysis of variance (ANOVA) test to see whether major allele frequency was significantly associated with an interaction between SNP type (selection vs. non-selection) and SNP location (sex chromosome vs. autosome) using the *avov()* function in R. We constructed this analysis as many outlier identification methods employ allele frequencies for statistical comparisons between sample groups, and an interaction between SNP location and its categorization as under selection or not would be cause for concern about bias due to data artefacts.

3 | RESULTS

3.1 | Population structure of contemporary and historical *S. vulgaris*

Our genetic data revealed strong differentiation between contemporary native and invasive range samples, as well as replicating the previously established subpopulation structure within Australia. PCA indicated that the historical samples were genetically clustered with contemporary UK populations, but were nevertheless the most similar native range population to the AU populations along the first PC axis (Figure 2a). Interestingly, the Australian samples appeared to be as divergent from each other as they were from native range samples (Figure 2a).

Through admixture analysis on the 73 sequenced samples (12 contemporary samples removed to due close relatedness, and five historical samples not included due to failed sequencing, see Section 3.4 "Sequencing and variant calling with historical samples") we determined that $K = 1$ and $K = 2$ had similar support (Figure S4a). Ancestry proportions when plotted for $K = 2$ (Figure 2b) support the historical UK clustering with the contemporary native range, with very little further substructure revealed when admixture analysis was conducted just on the native range samples (Figure 2b; Figure

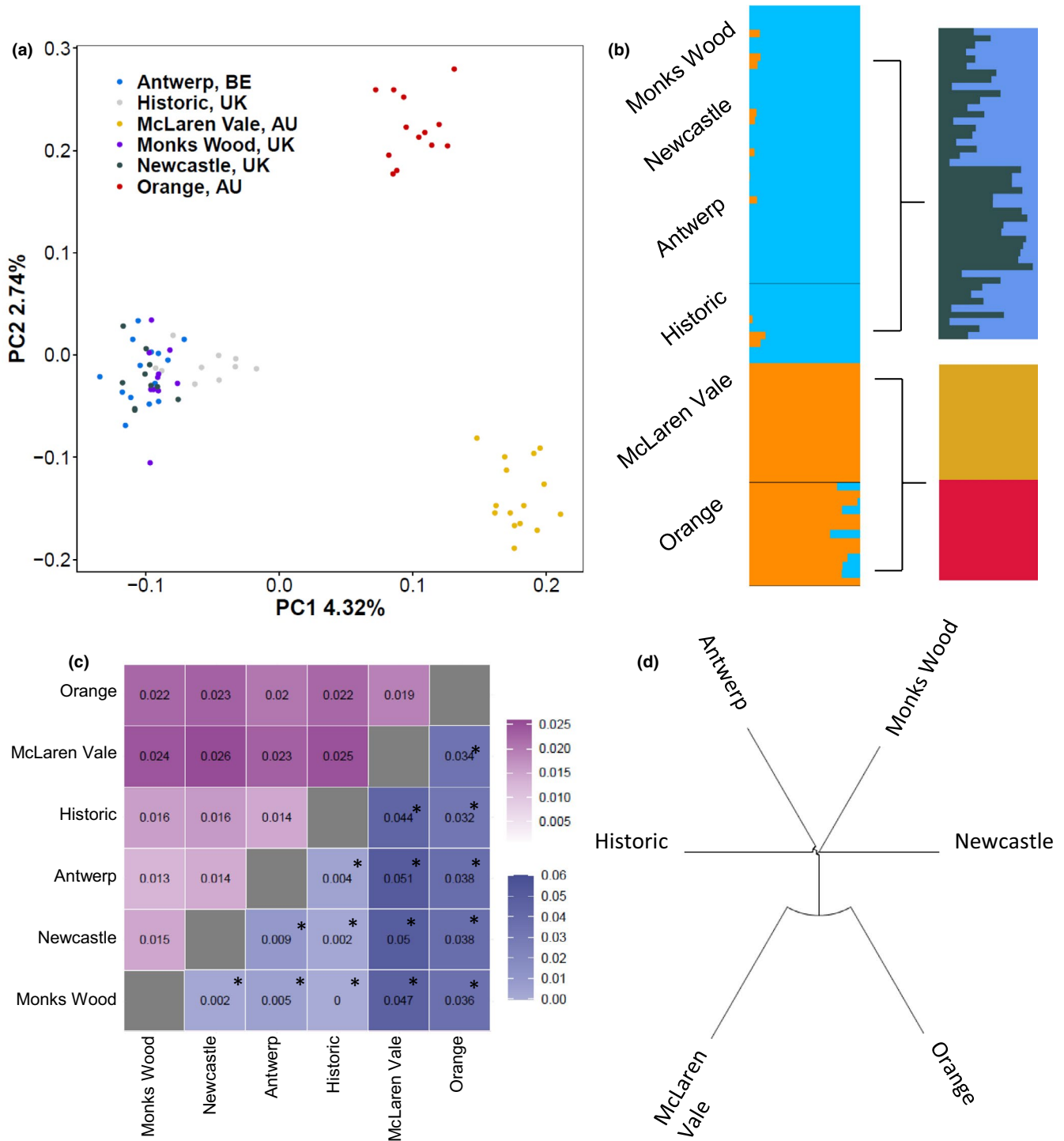


FIGURE 2 Population genetic analysis for contemporary and historical *Sturnus vulgaris* using the population genetics variant file. (a) PCA of the six sampling groups using SNPRERATE, and (b) ADMIXTURE ancestry Q profiles, averaged over 25 runs using CLUMPAK for all sample groups, all native range samples (contemporary and historical), contemporary native range samples, and invasive Australian samples. (c) Heatmap of pairwise analysis between each of the sample groups, with above the diagonal (purple) Nei's genetic distance, and below the diagonal (blue) pairwise F_{ST} (an asterisk * denoting a significant F_{ST} result), and (d) the phylogenetic relationships between sampling sites using a neighbour-joining tree

S4b,c). Again, this analysis identified strong substructure in the Australian samples (Figure S4c).

Of the invasive Australian samples, we found that those from McLaren Vale showed greater genetic differentiation from the native

range samples compared with those from Orange, evident in pairwise F_{ST} and genetic distance comparisons (Figure 2c) and corroborated by McLaren Vale admixture proportions lacking the UK cluster (Figure 2b). The historical samples were found to be most similar to

samples from Monks Wood when considering pairwise F_{ST} , PCA and ADMIXTURE, whereas Antwerp appeared closest when considering genetic distance and phylogeny (Figure 2). We also characterized the relationship amongst the native and invasive range samples by the neighbourhood-joining tree (Figure 2d), and found that the genetic differentiation between sampling sites across the native range was less than that across the invasive range.

3.2 | Genomic divergence between contemporary and historical *S. vulgaris*

Using the default BAYESCAN pipeline, we identified a total of 14 outlier loci across the five pairwise comparisons, eight of which were found in the UK-HS comparison, four in the UK-AU_e comparison, and one each in the AU_e-HS and UK-AU_s comparisons (Table 1; Figure S2a–e). The sliding window approach indicated roughly similar numbers of outlier SNPs across all the population comparisons, with the highest number of putative outlier loci at 40 for UK-AU_s, and the lowest number at 27 for UK-HS (Table 1; Figure S3). Interestingly, most but not all the SNPs identified by the BAYESCAN FDR approach were also identified in the F_{ST} sliding window approach (Table S3). We visualized the outlier sites pooled across both outlier detection methods against the starling genome, revealing a uniform spread throughout the genome (Figure 3). In plotting the outlier SNPs specific to either the AU_{east} and AU_{south} populations, we found that while there was some overlap, a majority of flagged outlier SNPs were different across these two subpopulations. Furthermore, in superimposing these identified outlier SNPs on top of single SNP site F_{ST} , we see that SNPs flagged as outliers within each population do not necessarily have high F_{ST} when the two populations are pooled and compared to either contemporary or historical native range samples.

We pooled putative outlier SNPs across the analyses (BAYESCAN and F_{ST}) done for each pairwise comparison, yielding a total of 31, 30, 32, 35 and 40 unique SNPs across identification methods for UK-HS, AU_e-HS, AU_s-HS, UK-AU_e and UK-AU_s respectively (Table 1).

3.3 | Putative adaptive selection

We identified SNPs under divergent and parallel selection, revealing a higher proportion of SNPs under parallel selection than divergent selection for AU_{east} comparisons (10 vs. six) and the reverse trend in comparisons involving AU_{south} (seven vs. 11; Figure 4a). We then

pooled the results from the two Australian subpopulation analyses and, of the 5,068 tested SNPs, a total of 137 SNPs were identified as under selection (one SNP appeared in both the divergent and parallel SNP lists), of which 15 were identified as resulting from divergent selection, and 12 from parallel selection (Figure 4a, Table 2). Of these SNPs, one appeared as under divergent selection in comparisons to AU_{south} but under parallel selection when compared to AU_{east} (Table S4). The remaining SNPs that did not fulfil the criteria for divergent or parallel selection were categorized as resulting from putative UK selection (24), putative AU selection (30), and putative UK and AU divergent selection (57) (Table 2, Figure 4a).

Across all five of these data sets, VARIANT EFFECT PREDICTOR analysis of the functional nature of the SNPs revealed that variants were mostly intergenic, and there were some within-intron variants (Figure 4b–f). Of the SNPs under divergent and parallel selection, 70% were intergenic variants, with the remainder largely being made up of intron variants, and one synonymous variant being flagged as under parallel selection. SNPs identified as under UK and AU putative selection (either UK-AU_e or UK-AU_s divergence) contained the lowest proportion of intergenic variants (Figure 4f), and the only other SNPs with predicted protein coding sequences were present in the UK putative selection SNP list (Figure 4d).

While many of the SNPs mapped to unannotated loci, we mapped eight SNPs in the divergent data set to annotated genes, along with six SNPs in the parallel data set (Table 2; Tables S4 and S5). One SNP that was flagged in both the divergent and parallel data set mapped to the gene *Transforming Growth Factor Beta Receptor 2* (*TGFBR2*). Only one gene was found to be divergent within the native range (*Ankyrin repeat and KH domain-containing protein 1*, *ANKKHD1*), while four were found to be divergent within the invasive Australian range: *Glutamate ionotropic receptor kainate type subunit 2* (*GRIK2*), *Neurexin-3* (*NRXN3*), *Voltage-dependent calcium channel subunit alpha-2/delta-3* (*CACNA2D3*), *Natterin-4* (*NATT4*) and *DNA methyltransferase 1* (*DNMT1*) (Table S5). A range of genes were flagged as being under parallel selection in both the native and invasive range: *E3 ubiquitin/ISG15 ligase TRIM25* (*TRIM25*), *Myb-binding protein 1A-like protein* (*MYBBP1A*), *CWC27 spliceosome associated cyclophilin* (*CWC27*), *Poly(A) RNA polymerase GLD2* (*TENT2*) and *Tripartite motif containing 28* (*TRIM28*) (Table S5). There was some gene ontology overlap between the two data sets (Figure S5).

We tested for any bias in chromosomal location of loci under selection, using Chi-squared analysis, and found that SNPs under putative selection were not proportionately distributed across macro-, micro- and Z sex chromosome ($\chi^2_6 = 33.23$, $p < .001$, $N = 5068$).

TABLE 1 Number of putative sites under selection in *Sturnus vulgaris* reported by the different selective scans for the pairwise comparisons of the UK-HS, AU_e-HS, AU_s-HS, UK-AU_e and UK-AU_s data sets

Outlier group		UK-HS	AU _e -HS	AU _s -HS	UK-AU _e	UK-AU _s
Data set SNP count		4997	4900	4961	4907	4942
BAYESCAN	FDR 0.05	8	1	0	4	1
F_{ST} sliding windows	Windowed F_{ST} > top 1% + SNP threshold	27	30	32	34	40
Total		31	30	32	35	40

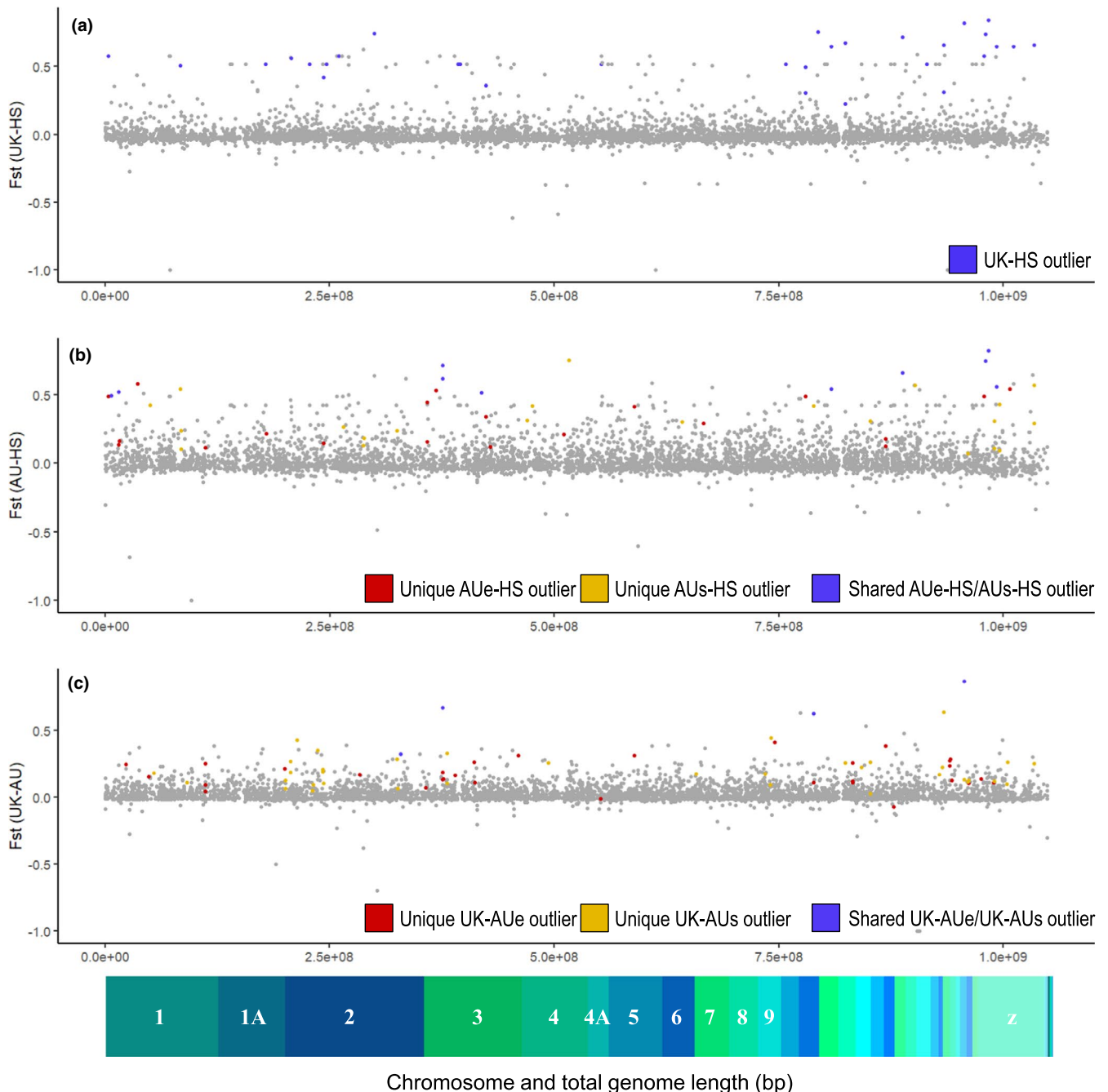


FIGURE 3 Pairwise population analysis of outlier loci across comparisons, mapped across the *Sturnus vulgaris* genome. Loci are plotted with respect to their position along the 1-Gb genome with major scaffolds of the *S. vulgaris* genome assembly noted along the x-axis, and with Weir and Cockerham F_{ST} calculated on a per-loci basis using *vcftools* between pairwise populations on the y-axis. (a) Flagged UK-HS F_{ST} and outliers, (b) AU-HS F_{ST} and outliers (red = unique AU_e -HS outliers, purple = unique AU_s -HS outliers, blue = shared AU_e -HS and AU_s -HS outliers), and (c) flagged UK-AU (red = unique UK- AU_e outliers, purple = unique UK- AU_s outliers, blue = shared UK- AU_e and UK- AU_s outliers). Nonoutlier SNPs are plotted in grey. The x-axis depicts the major scaffolds of the *S. vulgaris* genome assembly, sized for their representation in the SNPs plotted in (a)–(c)

There was an overabundance of micro- and Z sex chromosome SNPs for all groups of SNPs under selection (divergent, parallel, and the UK, AU, and UK and AU SNPs under putative selection, i.e., those SNPs that were reported as an outlier in only one pairwise population comparison; Figure 5; Table S6). We conducted additional analyses on potential biases in major allele frequency, as well as

bioinformatic sexing of individuals to determine if hemizyosity affected the results. We found no significant interaction between SNP location (sex chromosome or autosome) and SNP category (under selection or not) on major allele frequency ($F_{1,5064} = 0.001$, $p = .98$). Males and females were equally represented in our samples (Figure S6).

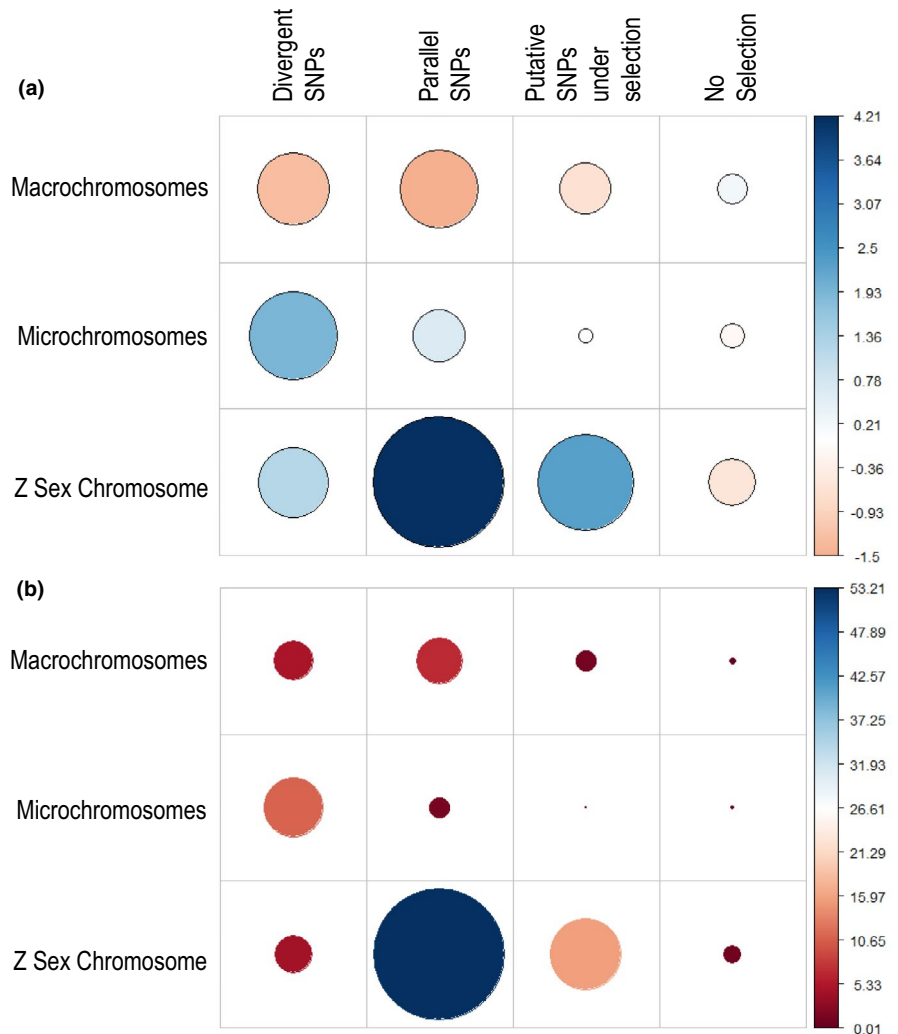


FIGURE 4 Summary of *Sturnus vulgaris* putative SNPs under selection. (a) Venn diagram of group categorization (left diagram using AU_{east} as the Australian comparison population, right diagram using AU_{south} as the Australian comparison population), and the remaining figures the VARIANT EFFECT PREDICTOR summary outputs of functional variation for each of the (pooled) five SNP groups of (b) SNPs under divergent selection, (c) SNPs under parallel selection, (d) UK SNPs under putative selection, (e) AU SNPs under putative selection, and (f) UK and AU SNPs under putative divergence

TABLE 2 SNP and gene counts of *Sturnus vulgaris* SNPs putatively under selection, for the five categorical groups based on pairwise comparisons

Outlier SNP group	Gene identification	Divergent selection	Parallel selection	Putative UK selection	Putative AU selection	Putative UK & AU selection
AU _{east}		6	10	20	15	29
AU _{south}		11	7	20	18	29
Pooled		15	12	24	30	57
Coding regions	Total	8	6	16	14	40
	Known (BLAST+)	7	6	10	14	30
	Unknown	1	0	6	0	10

FIGURE 5 Test of statistical association between SNPs categorized as under selection vs. the chromosome type they reside in for *Sturnus vulgaris* DArT-Seq, across SNP groupings of divergent SNPs, parallel SNPs, putative SNPs under selection (UK, AU, and UK and AU SNPs under putative selection), and SNPs not flagged as under selection. (a) Visualization of Pearson residuals, where the circle area is proportional to the amount of the cell contribution, positive residuals (indicating a positive correlation) are in blue, and negative residuals (indicating a negative correlation) are in orange; and (b) relative contribution of each cell to the total Chi-square score



3.4 | Sequencing and variant calling with historical samples

Of the 15 historical samples, 10 were successfully sequenced using DArTseq, a success rate (66.7%; Table S1) similar to that previously reported (62%) in a study using museum avian toe-pad samples ranging from 5 to 123 years old (Ewart et al., 2019). We found no trends related to DNA concentration, sample age or fragmentation between historical tissue samples that were successfully sequenced vs. those which were not (Table S1; Figures S7 and S8).

Of the three variant calling pipelines (Table S2), we found that BOWTIE2-GATK reported the highest percentage of successfully mapped reads over both contemporary and historical samples, followed closely by BWA-MEM. BWA-ALN resulted in much lower mapped read percentages. However, BOWTIE2-GATK returned the smallest numbers of variant sites in the unfiltered and filtered data set, with BWA-MEM returning slightly higher values than BWA-ALN. These results are in agreement with previous assessments of these software performances for reads of approximately this length (Li, 2013; Li & Durbin, 2009). BWA-ALN is generally reported

to map more conservatively than BWA-MEM (Robinson et al., 2017), leading to the much smaller mapped reads percentage, but this did not have a very large effect on the site counts or missing data per individual. The biggest difference between these two was the difference between the number of filtered variant sites for the historical samples, indicating that the lower quality reads produced by the historical samples were most impacted by the change in aligning algorithm. BWA-ALN was used as the variant calling pipeline in this paper because our read length fell on the border of what was recommended for BWA-ALN and BWA-MEM (70 bp), and a more conservative mapping and variant calling approach is suitable for population and selection analyses (when approaches are based on per-site allele frequencies).

We assessed the unfiltered data, and the base substitution plots per population revealed that though the historical samples reported lower SNP counts, base substitution frequencies were similar across the three population groupings (Figure S9a–c). When we mapped and aligned these reads to the genome assembly alongside Illumina whole genome variant data for the species (Hofmeister, Stuart, et al., 2021), similar patterns were found between the two sequencing approaches, and across all three population groupings, although with lower resolution in the historical individuals (Figure S10). Our sequencing and mapping of the historical samples indicate that, despite the lower quality and fragmentation of DNA, the overall patterns of base substitutions resembled that of the higher quality fresh tissue used from the contemporary samples, and that the reduced representation approach reflected variant densities seen in whole genome sequencing analyses. A smear plot of data revealed that missing data are relatively evenly spaced along the genome for historical samples (Figure S11), and not centred on particular genomic regions or chromosomes. Finally, the MAF plots revealed slightly differing patterns of MAF across the six sample groupings sequenced but, importantly, the historical samples did not appear to contain an unusually high number of sites with very low MAF (Figure S12).

4 | DISCUSSION

This study demonstrates that *Sturnus vulgaris* has not only undergone divergent selection within the invasive Australian range, but that the native and invasive ranges are undergoing parallel selection, possibly in response to global environmental changes. We note that contemporary native range populations, when considering comparative numbers of divergent SNPs, have undergone a similar amount of genetic change when compared to invasive range populations, despite the latter presumably being exposed to radically different and novel selection regimes. Moreover, we identified several genes related to immune function and pollution that appear to be under parallel selection in the contemporary native and invasive range samples, which may reflect global environmental changes over the last century and a half. The genes reported as divergent between the populations capture differing

selection regimes driving evolution within the two contemporary populations. We also identified a bias for selection on the Z chromosome in comparison to the autosomes.

Importantly, this study has successfully used reduced representation sequencing of historical and contemporary specimens to examine selection in *S. vulgaris* within the native and Australian invasive ranges, demonstrating the utility of museum collections in aiding evolutionary studies. While the success rate and quality of the historical specimen sequencing reads was less than with their contemporary counterparts, the method nevertheless yielded sufficient SNP data to enable examination of population structure and description of temporal patterns of genomic change in starling populations. Museum resources remain largely untapped in genomic studies across a variety of biological systems, but may serve as an invaluable source of information that will significantly extend knowledge about evolutionary processes and change over time.

4.1 | Population structure

Very few genetic data exist for native range *S. vulgaris*, and hence our study provides much needed insight into the population structure and genetic variation of the northwestern region of the native range. We identified low levels of genetic differentiation across the native range localities sampled. Some native range starlings are migratory (Feare, 1984), and this large-scale dispersion undoubtedly helps to maintain genetic diversity and suppress local differentiation. As expected, the historical starlings bear a stronger genetic resemblance to contemporary samples from the native range than those from the invasive Australian population. The historical samples are most differentiated from their contemporary counterparts in the PCA as compared to analyses of admixture, F_{ST} , genetic distance and phylogeny.

Records identify that the historical samples were taken from around London (Jenkins, 1977). The different population genetic analyses conducted indicated that either Monks Wood or Antwerp bears the strongest resemblance to the historical samples; these differences probably result from different statistical approaches underlying these analyses (e.g. F_{ST} considers the loci individually and ignores haplotypes, while PCA considers loci simultaneously and will include effects such as linkage disequilibrium between loci). The genetic differentiation between the two invasive Australian populations concurs with the two previously described Australian genetic subclusters (Stuart, Cardilini, et al., 2021), and further reinforces the idea that there were slight but distinct genetic differences in the founding individuals. Comparing the contemporary Australian sample sites to genotyping by sequencing (GBS) sequencing data from the same regions (Stuart, Cardilini, et al., 2021) suggests that the sample sizes in this study were sufficient to be representative of the genetic variation at sampling locations, and that between-sample site genetic divergence is higher within this invasive population when compared to the native population over comparable geographical distance (Stuart, Cardilini, et al., 2021).

4.2 | Genomic divergence

Only a few outliers were flagged by the *BAYESCAN* approach, probably due to the small sample sizes giving low statistical power to this stricter analysis (i.e., the program was unable to pick up the low signals of selection in these recently diverged populations; Al-Breiki et al., 2018). Nevertheless, pooling SNPs across both outlier methods enabled us to highlight a few key results regarding allele frequencies and outlier SNPs facing divergent or parallel selection across the starlings' range.

Our results indicate that when assessing native and invasive genetic differences, supposed evolutionary divergence should not be attributed solely to novel invasive range selection pressures or the processes involved in invasion itself. Many studies utilize genetic comparisons between invasive and native ranges to examine evolutionary divergence (Leger & Rice, 2007; Liu et al., 2020; Querns et al., 2020). Research into evolution within invasive systems often focuses on the divergent evolution within the invasive range in response to invasion processes or new selection regimes (Lee, 2002). Despite the apparent conservatism over all loci (including neutral ones) in the native range, roughly equal numbers of SNPs were found to be under divergent and parallel selection, and native range divergence occurred at a similar rate to the number of diverging SNPs within the invasive range. It is apparent that neutral similarities or differences between sample groups (Figure 2, close genetic similarity between contemporary and historical native range) may not necessarily be indicative of the number of selective differences (Figures 3 and 4).

When examining the SNPs that were flagged as outliers against Australia-wide calculated F_{ST} (Figure 3b,c), as well as subpopulation allele frequencies (Table S4), it is apparent that signatures of selection vary dramatically across the two subpopulations, particularly so for SNPs facing divergent selection. Divergence within the AU population may be a result of invasion bottleneck processes biasing allelic variation for or against rare variants (depending on their representation in the translocated individuals). However, given that several hundred individuals were introduced to Australia (Higgins et al., 2006; Jenkins, 1977), it would be very unlikely for common allelic variants to be lost through such random processes. A more likely explanation is that observed allele frequency shifts in AU are a result of factors that occurred after introduction, such as genetic drift, or selection against nonlocal maladaptation in which an invader's trait may be poorly matched to the new environment (e.g., Ward-Fear et al., 2009). This result also emphasizes the importance of analysing selection separately for subpopulations within highly structured and demographically complicated invasions with multiple putative introduction sites, as is the case with many invasive populations (LaRue et al., 2011; Tay et al., 2020; Xia et al., 2020). In comparison, patterns of parallel selection are far more consistent across the two AU subpopulations, with similar frequency shifts in both AU_{east} and AU_{south} (Tables S4 and S5). In instances where the divergent allele frequency shifts have occurred in both AU subpopulations, this may indicate independent instances of selection, or alternately, as there is gene flow across the population, an allele could have reached high proportions in one subpopulation and spread to the other.

Parallel change across invasive and native ranges is difficult to detect without an outgroup population or historical samples, and thus has been little studied within a species' invasive and native range. Here, we found that SNPs under putative parallel selection had higher allelic diversity in the historical samples, with both contemporary populations becoming fixed (or nearly so) for the same allelic variant (Table S4). This may be indicative of beneficial/non-deleterious alleles shifting towards fixation within contemporary populations, linkage with a nearby variant, or be a result of random processes such as drift (though this last explanation would be quite unlikely for parallel processes). This over-representation of fixation in contemporary populations is unsurprising, because parallel evolution is more likely to have been based on standing genetic variation in the ancestral population (variants that were already there) than on the same novel variants arising independently in the two descendant populations. We do see apparent fixation in historical populations and more mixed allele frequencies within contemporary populations, in a small number of cases within the parallel SNP data.

Across all SNPs under selection, only a few resided in coding regions, with only a couple of these residing in protein coding regions. Other than intergenic variants, intron variants made up the highest proportion of SNPs. Despite not being transcribed gene regions, introns may function as gene regulatory regions, so polymorphisms may still elicit functional changes (Shaul, 2017), and even synonymous variants can show codon usage bias (Zeng & Bromberg, 2019).

Understanding broad genetic patterns behind rapid evolution has long been a focus of fundamental evolutionary biology. Here, we observed a bias towards the larger sex chromosome (Z) in terms of SNPs that were categorized as being under parallel selection (and also within the UK, AU, and UK and AU SNPs under putative selection). It is possible that this may have arisen due to biases in aspects of the data (e.g., sequencing method, SNP variant calling pipeline). However, we found no differences in major allele frequency across SNPs categorized as either under selection or not, across sex chromosomes and autosomes. The conclusion does align with theory that suggests that sex chromosomes are capable of playing a disproportionate role in evolutionary divergence due to their haploid nature in one sex (e.g., the "faster-X effect"; Meisel & Connallon, 2013), and may be one of the first steps towards speciation (Oyler-McCance et al., 2015; Wilson Sayres, 2018). The relationship between sex chromosome evolution and local adaptation has been demonstrated theoretically (Lasne et al., 2017) and experimentally (Lasne et al., 2019), and has previously been observed in the invasive *Drosophila suzukii* (Ometto et al., 2013). The latter finding, in conjunction with our results, suggests that rapid sex chromosome evolution may be a widespread phenomenon across many invasive taxa.

4.3 | Genes undergoing putative adaptive selection

We identified a range of genes as under divergent or parallel selection in native and invasive range starlings, suggesting that a diverse range of biological processes are under selection across the species'

range. One gene, *TGFBR2*, was flagged as under parallel selection between UK and AU_{east} , but divergent from AU_{south} . *TGFBR2* has been well established as a top candidate gene for regulating differential beak morphology, and functional alterations in this gene are commonplace across many avian species (Abzhanov et al., 2006; Knief et al., 2012; Mallarino et al., 2011), suggesting a possible rapid response to changes in food source.

Examining genes under divergent selection across the starlings' contemporary ranges allows us to consider possible drivers of selection that differ between populations. Of the other genes found to be under divergent selection in AU, three have been associated with cognitive function and learning in birds, and may be linked to cognitive selection processes during urban colonization. First, *GRIK2* may be involved in the functional molecular organization of the avian cerebrum (Jarvis et al., 2013), *CACNA2D3* is involved in neurexin-mediated retrograde signalling (Tong et al., 2017) and may play an important role in pathways for learned avian vocalization (Friedrich et al., 2019; Wada et al., 2004), and finally *NRXN3* is related to brain connectivity (Mueller et al., 2020). *DNMT1* plays a central role in epigenetic inheritance because it copies methylation patterns following replication (Goyal et al., 2006), and may play an important role in rapid heritable responses within an invasion (Marin et al., 2020). In some cases, it is not easy to see how the function relates to possible drivers of selection; for example, *ANKHD1* was the only annotated gene we identified that appears to be undergoing divergent selection within the native range. This gene plays an important role in cell cycle progression and proliferation, and has been associated with cancers in humans and model organisms (Dhyani et al., 2012; Machado-Neto et al., 2014). Although the biological functions associated with these genes are broad, this result provides candidate genes for future studies investigating epigenetic inheritance and cognition, and their relationship to invasion success within this and other species.

There are few studies that focus on parallel selection between invasive and native ranges within the same species, with research often focused on parallel changes across independent introductions to better understand invasion mechanisms (Popovic et al., 2021; Stern & Lee, 2020). Here we examined signatures of parallel selection to better understand the types of selection pressures on a globally widespread species. The gene *MYBBP1A*, also putatively under parallel selection in our study, has previously been found to be up-regulated in response to pollution stress (Kumazawa et al., 2015; Mitra et al., 2020). Over the last 160 years, increased modernization globally has resulted in organisms being exposed to increased levels of pollutants and other immune system triggers (Capilla-Lasheras et al., 2017; Cummings et al., 2020; Watson et al., 2017). Even trace amounts of compounds may be detrimental to some organisms (Bucci et al., 2020; Kozlov et al., 2009), and starling eggs have been demonstrated to accumulate polluting organic compounds (Eens et al., 2013). There are less obvious putative links to potential global drivers of selection for other identified parallel genes *CWC27*, which is tentatively related to

inflammation and retinal degeneration (Busetto et al., 2020), and *TENT2*, possibly associated with post-transcriptional gene regulation and epitranscriptomics (Menezes et al., 2018). Of the genes that have possibly undergone parallel selection in both the native and invasive range, we identified two from the tripartite motif protein family (*TRIM25* and *TRIM28*), which is involved in pathogen recognition and host defence pathways in numerous avian species (Blaine, 2013; Wei et al., 2016). Nevertheless, with global bird numbers declining across both rare and common species (Gross, 2015; Li et al., 2020), understanding selective regimes may assist our understanding of species and range persistence patterns, and may shed light onto the various factors influencing native range starling declines.

The genes discussed above represent a short and analytically conservative list of those putatively under selection across the starling's global ranges. There are undoubtedly more genomic sites under selection, either only identified in one of the pairwise outlier data sets (and so not categorized as parallel or divergent), or in linkage with the identified SNPs (Brodie et al., 2016), or not sequenced at all using this reduced representation approach. However, these variants and genes serve as a shortlist of suitable targets for future gene expression studies, or as the basis of the development of further hypotheses regarding selection regimes of global avian populations.

4.4 | Historical sample sequencing

The success rate of historical sample sequencing was ~70%, and we found no correlation between sample properties and sequencing success. Patterns of variant density across the genome identified using DArTseq data appeared to follow similar patterns to the high-quality variant density data set provided by the whole genome comparison (Figure S10), although historical samples did result in a patchier variant distribution. However, prior simulations using data from historical samples indicated that though historical samples contain significantly more missing data when compared to fresh tissue samples, the level of genotyping error had a minimal effect on population structure inference (Ewart et al., 2019).

In sequence data processing, the BWA ALN-STACKS and BWA MEM-STACKS pipelines performed similarly, with BOWTIE-GATK performing comparatively much worse when considering total called SNPs. This large difference may be due to the large amounts of missing data (Catchen et al., 2013), but is different from the relative performance previously reported for STACKS and GATK variant calling pipelines (Wright et al., 2019), suggesting that variant calling success is data set-specific. Lastly, while processing of historical samples on their own resulted in lower levels of missingness, the comparatively lower number of called SNPs means that processing historical samples alongside contemporary counterparts resulted in a larger number of called variant sites.

4.5 | Future directions

Greater geographical coverage of native range starling genetics is a vital future step for evolutionary genomic studies on this species. Improved native range genetic data will help us better understand population structure and allelic shifts in the invasive ranges, and also shed light on native range dispersal dynamics. Further, the success of the museum sample sequencing demonstrated here gives hope that when future large-scale genomic studies are conducted in the native range, historical samples will be able to provide crucial background information regarding genetic diversity prior to the significant contemporary population declines.

Analyses of a greater number of historical samples will also aid in the categorization of adaptive SNPs, because this will reduce possible effects of random sampling bias and capture more rare alleles. The sequencing failure rate of this study is comparable to another study using similarly aged museum skins (Ewart et al., 2019), suggesting that future projects seeking to use museum samples might expect similar failure rates (30%–40%) and adjust their sampling design accordingly. Finally, similar analyses may be conducted between the historical UK and the well-studied contemporary North American starlings (because this population has a similar introduction time, range size and environmental variation: Bodt et al., 2020; Hofmeister, Werner, et al., 2021). Further studies into the temporal changes this species has undergone may be extended upon through comparisons with other invasive populations in New Zealand, South Africa and South America. Comparisons across parallel introductions will provide an invaluable opportunity to contrast concurrent species invasion and selection across multiple different environments, allowing for the discovery of broad evolutionary patterns in this invasive species.

5 | CONCLUSIONS

Overall, this study demonstrates that the combination of native, invasive and historical genetic data can lead to a more thorough understanding of global species shifts and adaptation during the Anthropocene. We used genetic sequencing of museum specimens to identify putatively adaptive genetic changes through reduced representation sequencing and outlier SNP identification analyses. We have described evidence of parallel and divergent evolution in native and invasive starlings since the mid-19th century. Finally, we identify an apparent bias towards putatively adaptive SNPs on the Z chromosome, suggesting that the major sex chromosome may play a major role in rapid evolution within this species.

ACKNOWLEDGEMENTS

Thanks are expressed to Hein Van Grouw at Tring NHM for assistance with sourcing historical starling specimens, and to the two anonymous reviewers for the constructive comments they provided, helping us to improve our manuscript. This work was

supported by the Equity Trustees Charitable Foundation (Holsworth Wildlife Research Endowment). L.A.R. was supported by a Scientia Fellowship from UNSW.

AUTHOR CONTRIBUTIONS

Project conception: K.C.S., L.A.R. Sample collection: M.B., M.E., M.C.B., L.A.R. Laboratory work: K.C.S., J.J.A. Data analysis: K.C.S. Manuscript writing: K.C.S. Manuscript Editing: all authors.

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at (<https://www.ncbi.nlm.nih.gov/bioproject/>).

DATA AVAILABILITY STATEMENT

The raw sequencing data have been deposited under BioProject accession no. PRJNA781785 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) (Stuart, Sherwin, et al., 2021). Processed genetic data files and basic code is available on Dryad (<https://doi.org/10.5061/dryad.dbrv15f2v>) and Zenodo (<https://doi.org/10.5281/zenodo.5839093>). More fully annotated and cleaned code, along with some project vignettes and any other relevant files or metadata for this project are available on GitHub (https://github.com/katarinastuart/Sv4_HistoricalStarlings).

ORCID

Katarina C. Stuart <https://orcid.org/0000-0002-0386-4600>
 William B. Sherwin <https://orcid.org/0000-0002-1578-8473>
 Jeremy J. Austin <https://orcid.org/0000-0003-4244-2942>
 Melissa Bateson <https://orcid.org/0000-0002-0861-0191>
 Marcel Eens <https://orcid.org/0000-0001-7538-3542>
 Matthew C. Brandley <https://orcid.org/0000-0002-3734-913X>
 Lee A. Rollins <https://orcid.org/0000-0002-3279-7005>

REFERENCES

- Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R., & Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, 442, 563–567. <https://doi.org/10.1038/nature04843>
- Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., Al-Abri, M. A., & Delghandi, M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (*Panulirus homarus*) along the Omani coastline. *BMC Genomics*, 19, 690. <https://doi.org/10.1186/s12864-018-5044-8>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., Webster, M. T., Öst, T., Schneider, M., Kempenaers, B., & Ellegren, H. (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*, 20, 485–495. <https://doi.org/10.1101/gr.101410.109>

- Bishop, P. (2011). Bell frog populations in New Zealand - good news or bad news? *Australian Zoologist*, 34, 408–413. <https://doi.org/10.7882/AZ.2008.018>
- Blaine, A. H. (2013). *Tripartite-motif family members in the White Pekin duck (Anas platyrhynchos) modulate antiviral gene expression*. University of Alberta.
- Bodt, L. H., Rollins, L. A., & Zichello, J. (2020). *Genetic diversity of the European Starling (Sturnus vulgaris) compared across three invasive ranges*.
- Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Research*, 44, 6046–6054. <https://doi.org/10.1093/nar/gkw500>
- Bucci, K., Tulio, M., & Rochman, C. M. (2020). What is known and unknown about the effects of plastic pollution: A meta-analysis and systematic review. *Ecological Applications*, 30, e02044. <https://doi.org/10.1002/eap.2044>
- Busetto, V., Barbosa, I., Basquin, J., Marquet, É., Hocq, R., Hennion, M., Paternina, J. A., Namane, A., Conti, E., Bensaude, O., & Le Hir, H. (2020). Structural and functional insights into CWC27/CWC22 heterodimer linking the exon junction complex to spliceosomes. *Nucleic Acids Research*, 48, 5670–5683. <https://doi.org/10.1093/nar/gkaa267>
- Callaway, R. M., & Ridenour, W. M. (2004). Novel weapons: Invasive success and the evolution of increased competitive ability. *Frontiers in Ecology and the Environment*, 2, 436–443. [https://doi.org/10.1890/1540-9295\(2004\)002\[0436:NWISAT\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2004)002[0436:NWISAT]2.0.CO;2)
- Capilla-Lasheras, P., Dominoni, D. M., Babayan, S. A., O'Shaughnessy, P. J., Mladenova, M., Woodford, L., Pollock, C. J., Barr, T., Baldini, F., & Helm, B. (2017). Elevated immune gene expression is associated with poor reproductive success of urban blue tits. *Frontiers in Ecology and Evolution*, 5, 64. <https://doi.org/10.3389/fevo.2017.00064>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cummings, C. R., Khan, N. Y., Murray, M. M., Ellison, T., Welch, C. N., Hernandez, S. M., & Navara, K. J. (2020). Foraging in urban environments increases bactericidal capacity in plasma and decreases corticosterone concentrations in white ibises. *Frontiers in Ecology and Evolution*, 8, 449. <https://doi.org/10.3389/fevo.2020.575980>
- Dainat, J. (2020). AGAT: Another Gff Analysis Toolkit to Handle Annotations in Any GTF/GFF Format. *Zenodo*.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Delibes-Mateos, M., Ferreras, P., & Villafuerte, R. (2009). European rabbit population trends and associated factors: A review of the situation in the Iberian Peninsula. *Mammal Review*, 39, 124–140. <https://doi.org/10.1111/j.1365-2907.2009.00140.x>
- Dhyani, A., Duarte, A. S. S., Machado-Neto, J. A., Favaro, P., Ortega, M. M., & Olalla Saad, S. T. (2012). ANKHD1 regulates cell cycle progression and proliferation in multiple myeloma cells. *FEBS Letters*, 586, 4311–4318. <https://doi.org/10.1016/j.febslet.2012.10.037>
- Dukes, J. S., & Mooney, H. A. (1999). Does global change increase the success of biological invaders? *Trends in Ecology & Evolution*, 14, 135–139. [https://doi.org/10.1016/S0169-5347\(98\)01554-7](https://doi.org/10.1016/S0169-5347(98)01554-7)
- Eens, M., Jaspers, V. L. B., Van den Steen, E., Bateson, M., Carere, C., Clergeau, P., Costantini, D., Dolenc, Z., Elliott, J. E., Flux, J., Gwinner, H., Halbrook, R. S., Heeb, P., Mazgajski, T. D., Moksnes, A., Polo, V., Soler, J. J., Sinclair, R., Veiga, J. P., ... Pinxten, R. (2013). Can starling eggs be useful as a biomonitoring tool to study organohalogenated contaminants on a worldwide scale? *Environment International*, 51, 141–149. <https://doi.org/10.1016/j.envint.2012.11.003>
- Erfmeier, A., & Bruelheide, H. (2010). Invasibility or invasiveness? Effects of habitat, genotype, and their interaction on invasive *Rhododendron ponticum* populations. *Biological Invasions*, 12, 657–676. <https://doi.org/10.1007/s10530-009-9472-x>
- Ewart, K. M., Johnson, R. N., Ogden, R., Joseph, L., Frankham, G. J., & Lo, N. (2019). Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species. *Molecular Ecology Resources*, 19, 1578–1592. <https://doi.org/10.1111/1755-0998.13082>
- Feare, C. J. (1984). *The Starling*. Shire.
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Freeman, S. N., Robinson, R. A., Clark, J. A., Griffin, B. M., & Adams, S. Y. (2007). Changing demography and population decline in the Common Starling *Sturnus vulgaris*: A multisite approach to Integrated Population Monitoring. *Ibis*, 149, 587–596. <https://doi.org/10.1111/j.1474-919X.2007.00684.x>
- Friedrich, S. R., Lovell, P. V., Kaser, T. M., & Mello, C. V. (2019). Exploring the molecular basis of neuronal excitability in a vocal learner. *BMC Genomics*, 20, 629. <https://doi.org/10.1186/s12864-019-5871-2>
- Goyal, R., Reinhardt, R., & Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Research*, 34, 1182–1188. <https://doi.org/10.1093/nar/gkl002>
- Gross, M. (2015). Europe's bird populations in decline. *Current Biology*, 25, R483–R485. <https://doi.org/10.1016/j.cub.2015.05.057>
- Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18, 691–699. <https://doi.org/10.1111/1755-0998.12745>
- Heldbjerg, H., Fox, A. D., Levin, G., & Nyegaard, T. (2016). The decline of the Starling *Sturnus vulgaris* in Denmark is related to changes in grassland extent and intensity of cattle grazing. *Agriculture, Ecosystems & Environment*, 230, 24–31. <https://doi.org/10.1016/j.agee.2016.05.025>
- Hellmann, J. J., Byers, J. E., Bierwagen, B. G., & Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. *Conservation Biology*, 22, 534–543. <https://doi.org/10.1111/j.1523-1739.2008.00951.x>
- Higgins, P. J., Peter, J. M., & Cowling, S. J. (2006). *Handbook of Australian, New Zealand & Antarctic birds. Volume 7, Boatbill to starlings*. Oxford University Press.
- Hoffmann, B. D., & Broadhurst, L. M. (2016). The economic cost of managing invasive species in Australia. *NeoBiota*, 31, 1–18. <https://doi.org/10.3897/neobiota.31.6960>
- Hofmeister, N. R., Stuart, K., Warren, W. C., Werner, S. J., Bateson, M., Ball, G. F., Buchanan, K. L., Burt, D. W., Cardilini, A. P. A., Cassey, P., Meyer, T. D., George, J., Meddle, S. L., Rowland, H. M., Sherman, C. D. H., Sherwin, W., Berghe, W. V., Rollins, L. A., & Clayton, D. F. (2021). Concurrent invasions by European starlings (*Sturnus vulgaris*) suggest selection on shared genomic regions even after genetic bottlenecks. *BioRxiv*. <https://doi.org/10.1101/2021.05.19.442026>
- Hofmeister, N. R., Werner, S. J., & Lovette, I. J. (2021). Environmental correlates of genetic variation in the invasive European starling in North America. *Molecular Ecology*, 30, 1251–1263. <https://doi.org/10.1111/mec.15806>
- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46, 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Jarvis, E. D., Yu, J., Rivas, M. V., Horita, H., Feenders, G., Whitney, O., Jarvis, S. C., Jarvis, E. R., Kubikova, L., Puck, A. E. P., Siang-Bakshi, C., Martin, S., McElroy, M., Hara, E., Howard, J., Pfenning, A., Mouritsen, H., Chen, C.-C., & Wada, K. (2013). Global view of the functional molecular organization of the avian cerebrum: Mirror

- images and functional columns. *Journal of Comparative Neurology*, 521, 3614–3665. <https://doi.org/10.1002/cne.23404>
- Jenkins, C. F. H. (1977). *The Noah's Ark Syndrome: One hundred years of acclimatization and zoo development in Australia*. Zoological Gardens Board of Western Australia.
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, 888, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Knief, U., Schielzeth, H., Kempnaers, B., Ellegren, H., & Forstmeier, W. (2012). QTL and quantitative genetic analysis of beak morphology reveals patterns of standing genetic variation in an Estrildid finch. *Molecular Ecology*, 21, 3704–3717. <https://doi.org/10.1111/j.1365-294X.2012.05661.x>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15, 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Kozlov, M., Zvereva, E., & Zverev, V. (2009). *Impacts of point polluters on terrestrial biota: Comparative analysis of 18 contaminated areas*. Springer Science & Business Media.
- Kumazawa, T., Nishimura, K., Katagiri, N., Hashimoto, S., Hayashi, Y., & Kimura, K. (2015). Gradual reduction in rRNA transcription triggers p53 acetylation and apoptosis via MYBBP1A. *Scientific Reports*, 5, 10854. <https://doi.org/10.1038/srep10854>
- LaRue, E. A., Ruetz, C. R., Stacey, M. B., & Thum, R. A. (2011). Population genetic structure of the round goby in Lake Michigan: Implications for dispersal of invasive species. *Hydrobiologia*, 663, 71–82. <https://doi.org/10.1007/s10750-010-0555-6>
- Lasne, C., Sgrò, C. M., & Connallon, T. (2017). The relative contributions of the X chromosome and autosomes to local adaptation. *Genetics*, 205, 1285–1304. <https://doi.org/10.1534/genetics.116.194670>
- Lasne, C., Van Heerwaarden, B., Sgrò, C. M., & Connallon, T. (2019). Quantifying the relative contributions of the X chromosome, autosomes, and mitochondrial genome to local adaptation. *Evolution*, 73, 262–277. <https://doi.org/10.1111/evo.13647>
- Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, 17, 386–391. [https://doi.org/10.1016/S0169-5347\(02\)02554-5](https://doi.org/10.1016/S0169-5347(02)02554-5)
- Leger, E. A., & Rice, K. J. (2007). Assessing the speed and predictability of local adaptation in invasive California poppies (*Eschscholzia californica*). *Journal of Evolutionary Biology*, 20, 1090–1103. <https://doi.org/10.1111/j.1420-9101.2006.01292.x>
- Li, H. (2013). *Linking sequence reads, clone sequences and assembly contigs with BWA-MEM*. ArXiv:1303.3997 [q-Bio].
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Y., Miao, R., & Khanna, M. (2020). Neonicotinoids and decline in bird biodiversity in the United States. *Nature Sustainability*, 3, 1027–1035. <https://doi.org/10.1038/s41893-020-0582-x>
- Lindenbaum, P. (2015). *JVarkit: Java-Based Utilities for Bioinformatics*. figshare.
- Liu, D., & Trumble, J. T. (2007). Comparative fitness of invasive and native populations of the potato psyllid (*Bactericera cockerelli*). *Entomologia Experimentalis et Applicata*, 123, 35–42. <https://doi.org/10.1111/j.1570-7458.2007.00521.x>
- Liu, W., Zhang, Y., Chen, X., Maung-Douglass, K., Strong, D. R., & Pennings, S. C. (2020). Contrasting plant adaptation strategies to latitude in the native and invasive range of *Spartina alterniflora*. *New Phytologist*, 226, 623–634. <https://doi.org/10.1111/nph.16371>
- Long, J. L. (1981). *Introduced birds of the world: The worldwide history, distribution, and influence of birds introduced to new environments*. Universe Books.
- Lopez, L., Turner, K. G., Bellis, E. S., & Lasky, J. R. (2020). Genomics of natural history collections for understanding evolution in the wild. *Molecular Ecology Resources*, 20, 1153–1160. <https://doi.org/10.1111/1755-0998.13245>
- Machado-Neto, J. A., Lazarini, M., Favaro, P., Franchi, G. C., Nowill, A. E., Saad, S. T. O., & Traina, F. (2014). ANKHD1, a novel component of the Hippo signaling pathway, promotes YAP1 activation and cell cycle progression in prostate cancer cells. *Experimental Cell Research*, 324, 137–145. <https://doi.org/10.1016/j.yexcr.2014.04.004>
- Mallarino, R., Grant, P. R., Grant, B. R., Herrel, A., Kuo, W. P., & Abzhanov, A. (2011). Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 4057–4062. <https://doi.org/10.1073/pnas.1011480108>
- Marin, P., Genitoni, J., Barloy, D., Maury, S., Gibert, P., Ghalambor, C. K., & Vieira, C. (2020). Biological invasion: The influence of the hidden side of the (epi)genome. *Functional Ecology*, 34, 385–400. <https://doi.org/10.1111/1365-2435.13317>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biology*, 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>
- Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, 29, 537–544. <https://doi.org/10.1016/j.tig.2013.05.009>
- Menezes, M. R., Balzeau, J., & Hagan, J. P. (2018). 3' RNA uridylation in epi-transcriptomics, gene regulation, and disease. *Frontiers in Molecular Biosciences*, 5, 61. <https://doi.org/10.3389/fmolb.2018.00061>
- Mitra, T., Mahanty, A., Ganguly, S., & Mohanty, B. P. (2020). Transcriptomic responses to pollution in natural riverine environment in *Rita rita*. *Environmental Research*, 186, 109508. <https://doi.org/10.1016/j.envres.2020.109508>
- Mueller, J. C., Carrete, M., Boerno, S., Kuhl, H., Tella, J. L., & Kempnaers, B. (2020). Genes acting in synapses and neuron projections are early targets of selection during urban colonization. *Molecular Ecology*, 29, 3403–3412. <https://doi.org/10.1111/mec.15451>
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106, 283–292. <https://doi.org/10.1086/282771>
- Ometto, L., Cestaro, A., Ramasamy, S., Grassi, A., Revadi, S., Siozios, S., Moretto, M., Fontana, P., Varotto, C., Pisani, D., Dekker, T., Wrobel, N., Viola, R., Pertot, I., Cavalieri, D., Blaxter, M., Anfora, G., & Rota-Stabelli, O. (2013). Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biology and Evolution*, 5, 745–757. <https://doi.org/10.1093/gbe/evt034>
- Oyler-McCance, S. J., Cornman, R. S., Jones, K. L., & Fike, J. A. (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity*, 115, 452–459. <https://doi.org/10.1038/hdy.2015.46>
- Popovic, I., Bierne, N., Gaiti, F., Tanurđić, M., & Riginos, C. (2021). Pre-introduction introgression contributes to parallel differentiation and contrasting hybridization outcomes between invasive and native marine mussels. *Journal of Evolutionary Biology*, 34, 175–192. <https://doi.org/10.1111/jeb.13746>
- Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13, 288–294. <https://doi.org/10.1016/j.tplants.2008.03.004>
- Querns, A., Wooliver, R., Vallejo-Marín, M., & Sheth, S. (2020). The evolution of thermal performance in native and invasive populations of *Mimulus guttatus*. *bioRxiv*. <https://doi.org/10.1101/2020.09.10.291252>

- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Robinson, K. M., Hawkins, A. S., Santana-Cruz, I., Adkins, R. S., Shetty, A. C., Nagaraj, S., Sadzewicz, L., Tallon, L. J., Rasko, D. A., Fraser, C. M., Mahurkar, A., Silva, J. C., & Dunning Hotopp, J. C. (2017). Aligner optimization increases accuracy and decreases compute times in multi-species sequence data. *Microbial Genomics*, 3, e000122. <https://doi.org/10.1099/mgen.0.000122>
- Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12, 2640–2659. <https://doi.org/10.1038/nprot.2017.123>
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28, 4737–4754. <https://doi.org/10.1111/mec.15253>
- Rogers, D. L., Matheson, A. C., Vargas-Hernández, J. J., & Guerra-Santos, J. J. (2006). Genetic conservation of insular populations of Monterey Pine (*Pinus radiata* D. Don). *Biodiversity & Conservation*, 15, 779–798. <https://doi.org/10.1007/s10531-004-1066-4>
- Rollins, L. A., Woolnough, A. P., Sinclair, R., Mooney, N. J., & Sherwin, W. B. (2011). Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology*, 20, 2307–2317. <https://doi.org/10.1111/j.1365-294X.2011.05101.x>
- Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R., & Sherwin, W. B. (2009). Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology*, 18, 1560–1573. <https://doi.org/10.1111/j.1365-294X.2009.04132.x>
- Shaul, O. (2017). How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology*, 91, 145–155. <https://doi.org/10.1016/j.biocel.2017.06.016>
- Siepielski, A. M., Morrissey, M. B., Buoro, M., Carlson, S. M., Caruso, C. M., Clegg, S. M., Coulson, T., DiBattista, J., Gotanda, K. M., Francis, C. D., Hereford, J., Kingsolver, J. G., Augustine, K. E., Kruuk, L. E. B., Martin, R. A., Sheldon, B. C., Sletvold, N., Svensson, E. I., Wade, M. J., & MacColl, A. D. C. (2017). Precipitation drives global variation in natural selection. *Science*, 355, 959–962. <https://doi.org/10.1126/science.aag2773>
- Sokolova, I. M., & Lannig, G. (2008). Interactive effects of metal pollution and temperature on metabolism in aquatic ectotherms: Implications of global climate change. *Climate Research*, 37, 181–201. <https://doi.org/10.3354/cr00764>
- Stern, D. B., & Lee, C. E. (2020). Evolutionary origins of genomic adaptations in an invasive copepod. *Nature Ecology & Evolution*, 4, 1084–1094. <https://doi.org/10.1038/s41559-020-1201-y>
- Stuart, K. C., Cardilini, A. P. A., Cassey, P., Richardson, M. F., Sherwin, W. B., Rollins, L. A., & Sherman, C. D. H. (2021). Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, 30, 1419–1434. <https://doi.org/10.1111/mec.15601>
- Stuart, K. C., Edwards, R. J., Cheng, Y., Warren, W. C., Burt, D. W., Sherwin, W. B., Hofmeister, N. R., Werner, S. J., Ball, G. F., Bateson, M., Brandley, M. C., Buchanan, K. L., Cassey, P., Clayton, D. F., Meyer, T. D., Meddle, S. L., & Rollins, L. A. (2021). Transcript- and annotation-guided genome assembly of the European starling. *BioRxiv*. <https://doi.org/10.1101/2021.04.07.438753>
- Stuart, K. C., Sherwin, W. B., Austin, J. J., Bateson, M., Eens, M., Brandley, M. C., & Rollins, L. A. (2021). Using historical museum samples to examine divergent and parallel evolution in the invasive starling. *BioProject*, Accession number PRJNA781785.
- Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6, e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Tay, W. T., Rane, R., Padovan, A., Walsh, T., Elfekih, S., Downes, S., Nam, K., d'Alençon, E., Zhang, J., Wu, Y., Nègre, N., Kunz, D., Kriticos, D. J., Czapak, C., Otim, M., & Gordon, K. H. J. (2020). Global FAW population genomic signature supports complex introduction events across the Old World. *BioRxiv*. <https://doi.org/10.1101/2020.06.12.147660>
- Tong, X.-J., López-Soto, E. J., Li, L., Liu, H., Nedelcu, D., Lipscombe, D., Hu, Z., & Kaplan, J. M. (2017). Retrograde synaptic inhibition is mediated by α -neurexin binding to the $\alpha 2\delta$ subunits of N-type calcium channels. *Neuron*, 95, 326–340.e5. <https://doi.org/10.1016/j.neuron.2017.06.018>
- Turbelin, A. J., Malamud, B. D., & Francis, R. A. (2017). Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography*, 26, 78–92. <https://doi.org/10.1111/geb.12517>
- Versluijs, M., van Turnhout, C. A. M., Kleijn, D., & van der Jeugd, H. P. (2016). Demographic changes underpinning the population decline of starlings *Sturnus vulgaris* in The Netherlands. *Ardea*, 104, 153–165. <https://doi.org/10.5253/arde.v104i2.a7>
- Wada, K., Sakaguchi, H., Jarvis, E. D., & Hagiwara, M. (2004). Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *The Journal of Comparative Neurology*, 476, 44–64. <https://doi.org/10.1002/cne.20201>
- Ward-Fear, G., Brown, G. P., Greenlees, M. J., & Shine, R. (2009). Maladaptive traits in invasive species: in Australia, cane toads are more vulnerable to predatory ants than are native frogs. *Functional Ecology*, 23, 559–568. <https://doi.org/10.1111/j.1365-2435.2009.01556.x>
- Watson, H., Videvall, E., Andersson, M. N., & Isaksson, C. (2017). Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. *Scientific Reports*, 7, 44180. <https://doi.org/10.1038/srep44180>
- Wei, Y., Zhou, H., Wang, A., Sun, L., Wang, M., Jia, R., Zhu, D., Liu, M., Yang, Q., Wu, Y., Sun, K., Chen, X., Cheng, A., & Chen, S. (2016). TRIM25 identification in the Chinese goose: Gene structure, tissue expression profiles, and antiviral immune responses in vivo and in vitro. *BioMed Research International*, 2016, e1403984. <https://doi.org/10.1155/2016/1403984>
- Wilson Sayres, M. A. (2018). Genetic diversity on the sex chromosomes. *Genome Biology and Evolution*, 10, 1064–1078. <https://doi.org/10.1093/gbe/evy039>
- Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., & Grueber, C. E. (2019). From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*, 20, 453. <https://doi.org/10.1186/s12864-019-5806-y>
- Xia, L., Geng, Q., & An, S. (2020). Rapid genetic divergence of an invasive species, *Spartina alterniflora*, in China. *Frontiers in Genetics*, 11, 284. <https://doi.org/10.3389/fgene.2020.00284>
- Zeng, Z., & Bromberg, Y. (2019). Predicting functional effects of synonymous variants: A systematic review and perspectives. *Frontiers in Genetics*, 10, 914. <https://doi.org/10.3389/fgene.2019.00914>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Stuart, K. C., Sherwin, W. B., Austin, J. J., Bateson, M., Eens, M., Brandley, M. C., & Rollins, L. A. (2022). Historical museum samples enable the examination of divergent and parallel evolution during invasion. *Molecular Ecology*, 31, 1836–1852. <https://doi.org/10.1111/mec.16353>