

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.

1 Introduction

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

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

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

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

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

48

49 **Results**

50 **Differentiation and population structure**

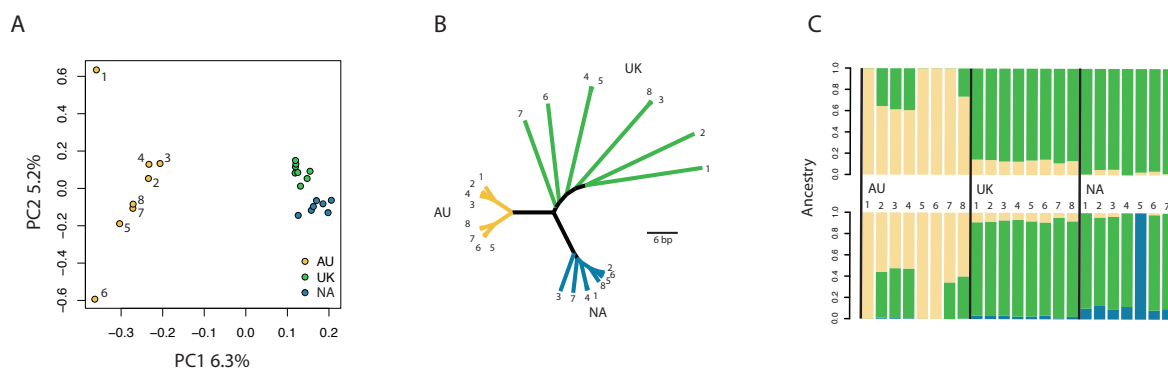
51 We sequenced whole genomes of eight native-range individuals from Northumberland,
52 UK and eight from each of the two invasions (NA: all 8 from New York City, USA; AU: 2
53 each from four locations in New South Wales, Australia, Table S1). After filtering, we
54 obtained more than 11 million SNPs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

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

69 variation) separates the UK and NA populations from the Australian population (Figure
70 1A). This evidence complements previous work that showed extensive population
71 structuring in Australia but nearly continuous gene flow across North America, based on
72 reduced-representation genomic data^{22,33}. Principal component analysis in the genotype
73 likelihood framework of ANGSD³⁴ shows nearly identical results (Figure S6).
74 Furthermore, individuals are reliably assigned to clades based on pairwise genetic
75 distances calculated in ANGSD (Figure 1B).

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

84



85

86 Figure 1. A) Principal components of 40,488 unlinked SNPs explain 6.3% (PC1) and
87 5.2% (PC2) of genetic variation. B) cladogram of genetic distances among samples
88 based on genotype likelihoods of 16,151,007 sites. C) ADMIXTURE analyses showing
89 K=2 (top) and K=3 (bottom row).

90

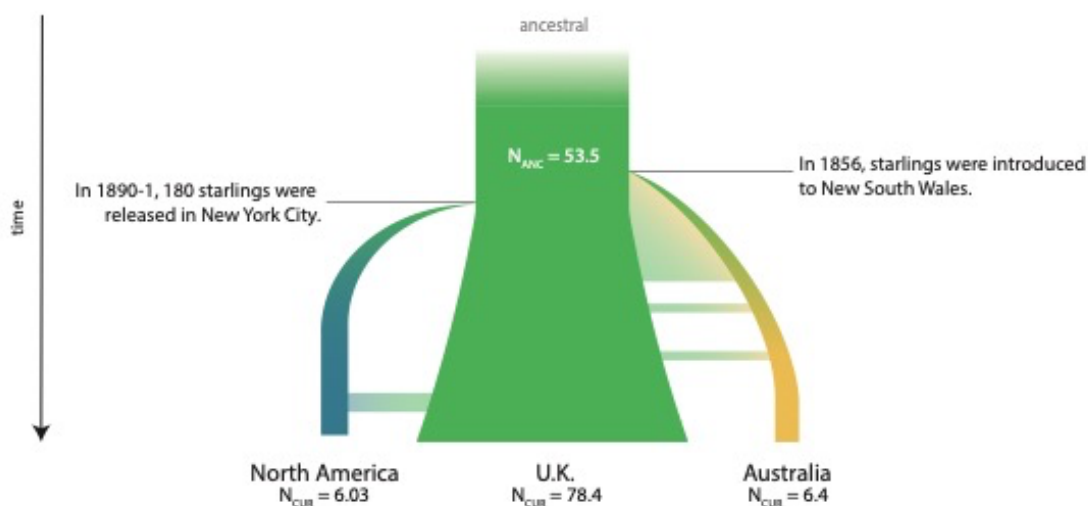
91 Admixture analyses revealed statistical support for a two-population model,
92 which distinguished AU from NA+UK and was only slightly weaker (cross-validation
93 error = 0.88) than the best-supported model of K=1 (cross-validation error = 0.73, Figure
94 1C, SI Section 3). Regardless of the number of populations hypothesized in models of
95 admixture, the NA invasion consistently shares a higher proportion of its ancestry with
96 the UK population. Individuals from the Australian population are distinguished from the
97 other two populations in all tested values of K. Considered in concert, these tests of
98 population structure show that Australian and North American populations differ in the
99 amount of divergence from the native UK population. Founder effects likely contribute to
100 the observed population structure, and below we describe explicit models of
101 demographic processes.

102

103 **Invasive and native populations experienced bottlenecks and subsequent** 104 **expansion**

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

109 model shows that both invasions experienced a bottleneck upon introduction (Figure 2),
110 which is exactly what we expect given the small number of founding individuals in both
111 AU and NA. Bottlenecks often lead to inbreeding within a population, and we find that
112 inbreeding is negligible but slightly higher in the Australian population than in the NA
113 population (Table S2). In addition, relatedness among individuals is quite low, where
114 zero indicates no shared alleles among individuals (maximum AJK statistic = 0.06).
115 Each invasion appears to have recovered quickly by expanding in effective population
116 size, and at present, our data suggest similar effective population sizes in both the AU
117 and NA invasions. Much of the ancestral variation appears to be shared among
118 invasions, given that the genome-wide average F_{ST} between AU and NA is 0.04,
119 although we note that genetic differentiation among invasions confirms the expectation
120 that different variants will make it through each genetic bottleneck. The results
121 presented here concur with range-wide sampling that indicates genetic bottlenecks
122 followed by rapid expansion with little evidence of inbreeding in both Australia²² and
123 North America³³.
124



125

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

129

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

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

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

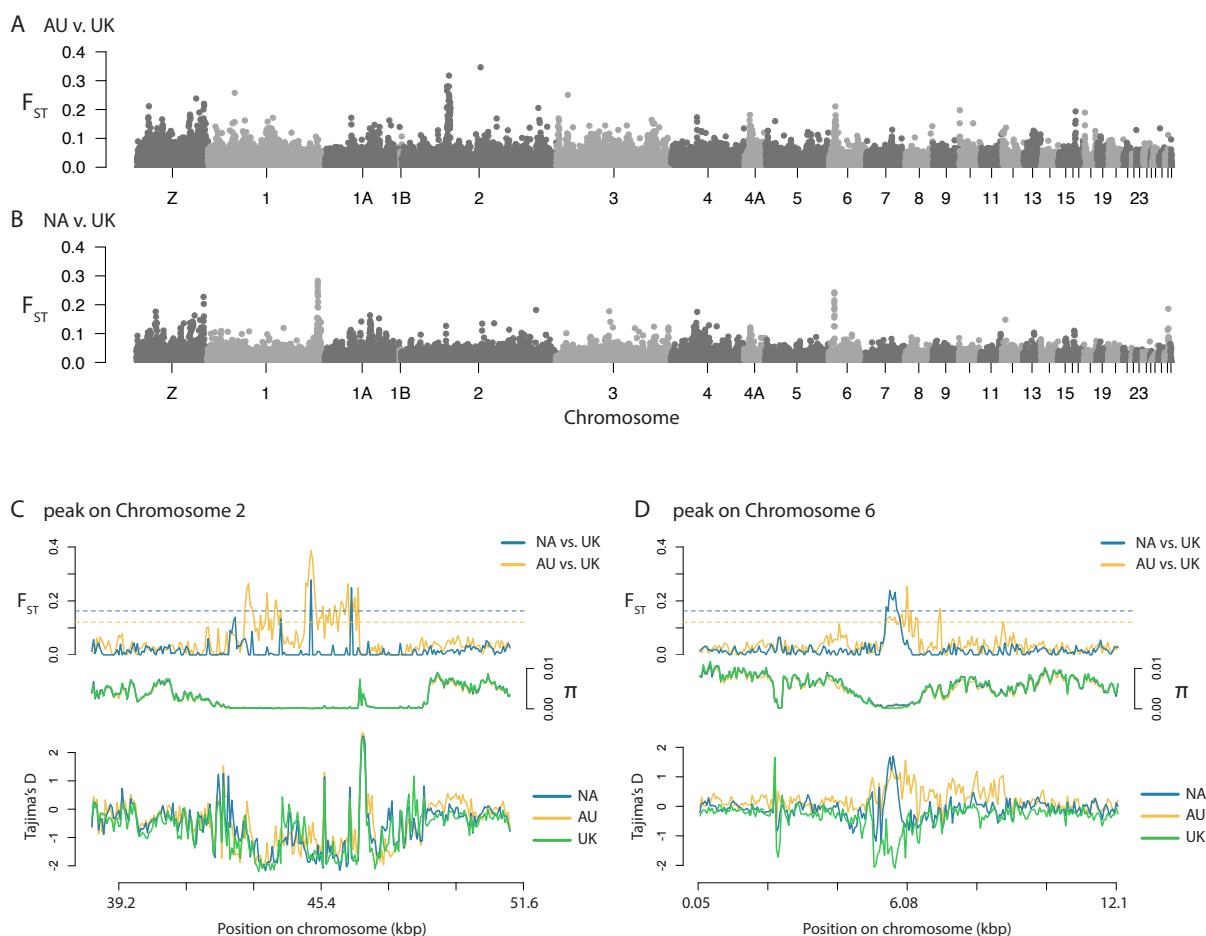
160

161 **Differentiation of a few genomic regions reveal consequences of invasion**

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

164 derived populations with fixation indices as high as 0.57 in a single 10-kb window
165 (Figure 3). However, no putative outlier windows approach fixation in any comparison
166 across all individuals in a population. We consider only the top 0.1% of 50-kb windows
167 to be F_{ST} outliers (AU vs. UK: $F_{ST} > 0.30$; NA vs. UK: $F_{ST} > 0.22$); subsequent
168 references to outliers indicate these windows. Outliers include Chromosomes 1, 2 and
169 6, and to a lesser extent, 1A, 4, and 4A. Examining localized shifts in diversity and
170 differentiation at these regions can help to resolve the relative contributions of
171 demography and selection.

172 Genome scans like this one tend to search for genomic islands of divergence,
173 where recombination can explain divergence among populations³⁶. High levels of
174 differentiation paired with a reduction in diversity may stem from suppressed
175 recombination (e.g., proximity to the centromere or a structural rearrangement), or from
176 alleles approaching fixation or loss due to drift or selection (e.g., a selective sweep). We
177 find that most putative outlier regions are distant from the centromere location
178 (predicted via homology with the zebra finch (*Taeniopygia guttata*) genome,
179 Supplementary Information). Nevertheless, genomic architecture—including linkage
180 disequilibrium independent of the centromere—plays a role in how differentiation among
181 populations is generated. Below, we discuss how demography and selection drive
182 divergence in invasions, considering the context of the starling's genomic architecture.
183



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

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

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

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

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

231

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

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

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

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

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

271

272 **Genes under putative selection may aid in invasion success**

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

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

287

288 **Conclusion**

289 An open question in invasion biology is whether an invasive population's success
290 is better attributed to intrinsic properties of the invasive species or to extrinsic factors
291 specific to the novel environment. These results demonstrate how demographic shifts
292 during and immediately after any establishment may support a species' success and
293 even lead to local adaptation to environmental conditions. The Australian and North
294 American starling invasions colonized each continent around the same time, and
295 experienced similar contractions in population size that led to classical founder events
296 upon establishment. The shared decline in genetic diversity represents a shared
297 intrinsic determinant of invasion success. However, differences in propagule pressure
298 and dispersal likely influence the evolutionary trajectory of each population ^{37,38}.
299 Propagule pressure (also termed introduction effort) is a composite measure of the
300 number of individuals released, and we note that founding population sizes may have
301 varied slightly among AU and NA introductions. Even without dramatic variation in
302 founding population sizes, dispersal itself shapes genetic diversity and thus adaptive
303 potential, as shown in a recent study of invasive plants ³⁹. Starlings' dispersal and
304 migration varies among populations: starlings in Australia tend to migrate less frequently
305 and across shorter distances, but starlings migrate and/or disperse hundreds of
306 kilometers in North America ^{19,40} and South Africa⁴¹.

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

323

324 **Methods**

325 *Whole-genome re-sequencing*

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

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

339 We called variants using GATK's HaplotypeCaller in GVCF mode and flagged
340 low-quality variants using GATK Best Practices (QD<2, FS>60, MQ<40, and SOR>3,
341 accessed March 21, 2018⁵²). We filtered sites for missing data, depth, and quality using
342 vcftools (parameters: --max-missing 0.8 --min-meanDP 2 --max-meanDP 50 --remove-
343 filtered-all), which removed 4.1 million sites from the original SNP set and left a total of
344 23.4 million sites for downstream analyses. Starting from the mapped reads used in the
345 GATK pipeline, we also called SNPs based on a minimum p-value of the correct
346 genotype probability at each site using ANGSD^{34,53}. Filtering for SNP p-value (0.0001),
347 depth (between 60 and 400 sequences), and mapping quality (>20) left 16,151,007
348 sites. All scripts used in read processing and filtering are available on GitHub:
349 <https://github.com/nathofme/global-RESEQ>.

350

351 *Population structure*

352 Population structure analyses used a dataset of biallelic SNPs in Hardy-
353 Weinberg Equilibrium, where minor allele count (MAC) is > 2 and SNPs were pruned for
354 LD by removing all sites with an $r^2 > 0.6$ within 1kb sliding windows. This filtering left
355 868,685 sites. Since some individuals showed much lower coverage (minimum 5.58X),
356 all tests of population structure were run with both variant-called (GATK) and genotype
357 probability (ANGSD) datasets. Scripts for variant-called analyses are stored at
358 <https://github.com/nathofme/global-RESEQ/blob/master/filter-scan.sh>, and scripts for
359 probability-based analyses at [https://github.com/nathofme/global-](https://github.com/nathofme/global-RESEQ/blob/master/angsd.sh)
360 [RESEQ/blob/master/angsd.sh](https://github.com/nathofme/global-RESEQ/blob/master/angsd.sh).

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

366

367 *Demographic inference*

368 We used FASTSIMCOAL2³⁵ to explicitly test for genetic bottlenecks in each population.
369 FASTSIMCOAL2 takes a site-frequency spectrum (SFS) as input, and we used the SFS
370 estimated from ANGSD given that likelihood-based estimates are more robust to
371 sequencing error⁵⁸. Demographic models in *fastsimcoal2* used priors on the time (TBOT
372 = 10 to 300) and size of the bottleneck (NBOT = 10 to 1000). The command line
373 arguments were as follows: -M -n 1000000 -L 50 -q -k 100000. Each run began with a
374 randomly generated seed (-r), and the -k flag simply writes polymorphic sites to a

375 temporary file to cope with the high memory usage of this analysis. Scripts for
376 demographic analyses can be found at [https://github.com/nathofme/global-](https://github.com/nathofme/global-RESEQ/blob/master/demography.sh)
377 [RESEQ/blob/master/demography.sh](https://github.com/nathofme/global-RESEQ/blob/master/demography.sh).

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

381

382 *Sliding window scans*

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

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

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

403

404 *Identifying candidate genes under selection*

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

414

415 **Acknowledgments**

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

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

426

427

428 **Author Contributions**

429 Project conception: all authors

430 Field work: Melissa Bateson, Lee Ann Rollins, Adam Cardilini, Scott J Werner

431 Lab work: Wes Warren and the McDonnell Group at WUSTL.

432 Data analysis and manuscript writing: Natalie R Hofmeister, Katarina Stuart

433 Manuscript editing: all authors

434

435 **References**

- 436 1. Russell, J. C. & Blackburn, T. M. The Rise of Invasive Species Denialism. *Trends*
437 *in Ecology & Evolution* **32**, 3–6 (2017).
- 438 2. Ricciardi, A. & Ryan, R. The exponential growth of invasive species denialism.
439 *Biological Invasions* **20**, 549–553 (2017).
- 440 3. Sagoff, M. Fact and value in invasion biology. *Conservation Biology* **34**, 581–588
441 (2019).
- 442 4. Davis, M. A. Let's welcome a variety of voices to invasion biology. *Conservation*
443 *Biology* [cobi.13608–6](https://doi.org/10.1111/cobi.13608) (2020). doi:10.1111/cobi.13608
- 444 5. Lowe, W. H., Kovach, R. P. & Allendorf, F. W. Population Genetics and
445 Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution* **32**,
446 141–152 (2017).
- 447 6. Kolbe, J. J. *et al.* Genetic variation increases during biological invasion by a
448 Cuban lizard. *Nature* **431**, 177–181 (2004).
- 449 7. Dlugosch, K. M. & Parker, I. M. Founding events in species invasions: genetic
450 variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* **17**,
451 431–449 (2008).

- 452 8. Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M. & Reimchen, T. E.
453 Experimental evidence for rapid genomic adaptation to a new niche in an adaptive
454 radiation. *Nat. ecol. evol.* 1–14 (2018). doi:10.1038/s41559-018-0581-8
- 455 9. Gralka, M. *et al.* Allele surfing promotes microbial adaptation from standing
456 variation. *Ecology Letters* **19**, 889–898 (2016).
- 457 10. White, T. A., Perkins, S. E., Heckel, G. & Searle, J. B. Adaptive evolution during
458 an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in
459 Ireland. *Mol Ecol* **22**, 2971–2985 (2013).
- 460 11. Tsutsui, N. D., Suarez, A. V., Holway, D. A. & Case, T. J. Reduced genetic
461 variation and the success of an invasive species. *Proc Natl Acad Sci USA* **97**,
462 5948–5953 (2000).
- 463 12. Verhoeven, K. J. F., Macel, M., Wolfe, L. M. & Biere, A. Population admixture,
464 biological invasions and the balance between local adaptation and inbreeding
465 depression. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2–8
466 (2011).
- 467 13. Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T. & Christie, M.
468 R. Rapid genetic adaptation to a novel environment despite a genome-wide
469 reduction in genetic diversity. *Mol Ecol* **17**, 675 (2018).
- 470 14. Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A. & Gillette, H. D. The
471 devil is in the details: genetic variation in introduced populations and its
472 contributions to invasion. *Mol Ecol* **24**, 2095–2111 (2015).
- 473 15. Lu-Irving, P., Marx, H. E. & Dlugosch, K. M. ScienceDirect Leveraging
474 contemporary species introductions to test phylogenetic hypotheses of trait
475 evolution. *Current Opinion in Plant Biology* **42**, 95–102 (2018).
- 476 16. Bock, D. G., Kantar, M. B., Caseys, C., Matthey-Doret, R. & Rieseberg, L. H.
477 Evolution of invasiveness by genetic accommodation. *Nat. ecol. evol.* **2**, 991–999
478 (2018).
- 479 17. Linz, G. M., Homan, H. J., Gaulker, S. M., Penry, L. B. & Bleier, W. J. *European*
480 *starlings: a review of an invasive species with far-reaching impacts. Managing*
481 *Vertebrate Invasive Species* 24 (2007).
- 482 18. Forbush, E. H. *The Starling*. (Wright and Potter Printing Co., 1915).
- 483 19. Kessel, B. Distribution and migration of the European Starling in North America.
484 *The Condor* **55**, 49–67 (1953).
- 485 20. Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R. & Sherwin, W. B.
486 Invasive species can't cover their tracks: using microsatellites to assist
487 management of starling (*Sturnus vulgaris*) populations in Western Australia. *Mol*
488 *Ecol* **18**, 1560–1573 (2009).
- 489 21. Feare, C. *The Starling*. (Oxford University Press, 1984).
- 490 22. Stuart, K. C. *et al.* Signatures of selection in a recent invasion reveals adaptive
491 divergence in a highly vagile invasive species. *Mol Ecol* (2020).
492 doi:10.1101/mec.15601
- 493 23. Cardilini, A. P. A., Buchanan, K. L., Sherman, C. D. H., Cassey, P. & Symonds,
494 M. R. E. Tests of ecogeographical relationships in a non-native species: what
495 rules avian morphology? *Oecologia* **181**, 783–793 (2016).
- 496 24. Cardilini, A. P. A. *et al.* Environmental Influences on Neuromorphology in the Non-
497 Native Starling *Sturnus vulgaris*. *Brain Behav Evol* **92**, 63–70 (2018).

- 498 25. Bodt, L. H., Rollins, L. A. & Zichello, J. M. Contrasting mitochondrial diversity of
499 European starlings (*Sturnus vulgaris*) across three invasive continental
500 distributions. *Ecol Evol* **10**, 10186–10195 (2020).
- 501 26. Hofmeister, N. R., Werner, S. J. & Lovette, I. J. Environmental correlates of
502 genetic variation in the invasive European starling in North America. *Mol Ecol*
503 *mec.15806–33* (2021). doi:10.1111/mec.15806
- 504 27. Dolbeer, R. A. Migration patterns of age and sex classes of Blackbirds and
505 Starlings. *Journal of Field Ornithology* **53**, 28–46
- 506 28. Werner, S. J., Fischer, J. W. & Hobson, K. A. Multi-isotopic ($\delta^2\text{H}$, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$)
507 tracing of molt origin for European starlings associated with U.S. dairies and
508 feedlots. *PLoS ONE* **15**, e0237137 (2020).
- 509 29. Ball, G. F. & Wingfield, J. C. Changes in plasma levels of luteinizing hormone and
510 sex steroid hormones in relation to multiple-broodedness and nest-site density in
511 male starlings. *Physiological Zoology* **60**, 191–199 (1987).
- 512 30. Dawson, A. Plasma gonadal steroid levels in wild starlings (*Sturnus vulgaris*)
513 during the annual cycle and in relation to the stages of breeding. *General and*
514 *Comparative Endocrinology* **49**, 286–294
- 515 31. Briski, E. *et al.* Beyond propagule pressure: importance of selection during the
516 transport stage of biological invasions. *Frontiers in Ecology and the Environment*
517 **16**, 345–353 (2018).
- 518 32. Irwin, D. E. *et al.* A comparison of genomic islands of differentiation across three
519 young avian species pairs. *Mol Ecol* **511**, 83–17 (2018).
- 520 33. Hofmeister, N. R., Werner, S. J. & Lovette, I. J. Environment but not geography
521 explains genetic variation in the invasive and largely panmictic European starling
522 in North America. **295**, 254–32 (2019).
- 523 34. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
524 Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
- 525 35. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. Robust
526 Demographic Inference from Genomic and SNP Data. *PLoS Genet* **9**, e1003905–
527 17 (2013).
- 528 36. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of
529 speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**,
530 3133–3157 (2014).
- 531 37. Simberloff, D. The Role of Propagule Pressure in Biological Invasions. *Annu. Rev.*
532 *Ecol. Evol. Syst.* **40**, 81–102 (2009).
- 533 38. Williams, J. L., Hufbauer, R. A. & Miller, T. E. X. How Evolution Modifies the
534 Variability of Range Expansion. *Trends in Ecology & Evolution* 1–11 (2019).
535 doi:10.1016/j.tree.2019.05.012
- 536 39. Smith, A. L. *et al.* Global gene flow releases invasive plants from environmental
537 constraints on genetic diversity. *Proc Natl Acad Sci USA* **117**, 4218–4227 (2020).
- 538 40. Burt, H. E. & Giltz, M. L. Seasonal directional patterns of movements and
539 migrations of starlings. *Bird-Banding* **48**, 259–271 (1977).
- 540 41. Berthouly-Salazar, C. *et al.* Long-distance dispersal maximizes evolutionary
541 potential during rapid geographic range expansion. *Mol Ecol* **22**, 5793–5804
542 (2013).

- 543 42. Sved, J. & Bird, A. The expected equilibrium of the CpG dinucleotide in vertebrate
544 genomes under a mutation model. *Proc Natl Acad Sci USA* **87**, 4692–4696
545 (1990).
- 546 43. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG
547 dinucleotides in the human genome distinguishes two distinct classes of
548 promoters. *Proc Natl Acad Sci USA* **103**, 1412–1417 (2006).
- 549 44. Simmen, M. Genome-scale relationships between cytosine methylation and
550 dinucleotide abundances in animals. *Genomics* **92**, 33–40
- 551 45. Sheldon, E. L., Schrey, A., Andrew, S. C., Ragsdale, A. & Griffith, S. C.
552 Epigenetic and genetic variation among three separate introductions of the house
553 sparrow (*Passer domesticus*) into Australia. *Royal Society Open Science* **5**,
554 172185 (2018).
- 555 46. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter
556 trimming, identification, and read merging. *BMC Research Notes* 1–7 (2016).
557 doi:10.1186/s13104-016-1900-2
- 558 47. Stuart, K. C. *et al.* Transcript- and annotation-guided genome assembly of the
559 European starling. *bioRxiv* (2021). doi:10.1101/2021.04.07.438753
- 560 48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
561 *Methods* **9**, 357–359 (2012).
- 562 49. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-
563 sample quality control for high-throughput sequencing data. *Bioinformatics*
564 *btv566*–10 (2015). doi:10.1093/bioinformatics/btv566
- 565 50. Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence
566 alignment: Satsuma. *Bioinformatics* **26**, 1145–1151 (2010).
- 567 51. Knief, U. & Forstmeier, W. Mapping centromeres of microchromosomes in the
568 zebra finch (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma* 1–12
569 (2016). doi:10.1007/s00412-015-0560-7
- 570 52. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for
571 analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–
572 1303 (2010).
- 573 53. Kim, S. Y. *et al.* Estimation of allele frequency and association mapping using
574 next-generation sequencing data. *BMC Bioinformatics* **12**, 231 (2011).
- 575 54. Zheng, X. *et al.* A high-performance computing toolset for relatedness and
576 principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
- 577 55. Fumagalli, M., Vieira, F. G., Linderoth, T. & Nielsen, R. ngsTools: methods for
578 population genetics analyses from next-generation sequencing data.
579 *Bioinformatics* **30**, 1486–1487 (2014).
- 580 56. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for
581 individual ancestry estimation. *BMC Bioinformatics* **12**, 1–6 (2011).
- 582 57. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating Individual Admixture
583 Proportions from Next Generation Sequencing Data. *Genetics* **195**, 693–702
584 (2013).
- 585 58. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP Calling,
586 Genotype Calling, and Sample Allele Frequency Estimation from New-Generation
587 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