HISTORICAL AND GENETIC APPROACHES TO

UNDERSTANDING INVASION SUCCESS OF

THE EUROPEAN STARLING (*STURNUS VULGARIS*)

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Natalie R. Hofmeister

August 2022

HISTORICAL AND GENETIC APPROACHES TO

UNDERSTANDING INVASION SUCCESS OF

THE EUROPEAN STARLING (*STURNUS VULGARIS*)

Natalie R. Hofmeister, Ph. D.

Cornell University 2022

How invasive species establish and spread is a central question in biology, but also one
that impacts public perceptions of and engagement with those species. I use the
invasive European starling (*Sturnus vulgaris*) as a case study to identify factors
supporting its establishment and expansion in the United States, and to consider its
evolutionary history in its native range. In my first chapter, I integrate recent work in
the evolutionary genetics of invasive European starling populations with ecological
studies over its residence in each region, identifying areas for future research. In my
second chapter, I use genomic methods to reconstruct demographic history and test for
natural selection in the North American invasion, and I find that starlings in North
America show evidence for local adaptation despite a genetic bottleneck upon invasion.
In my third chapter, I compare patterns of genetic variation between the concurrent
North American and Australian invasions: starling populations show remarkably high
differentiation from each other on a short evolutionary timescale, and this
differentiation is consistent with selection in at least a few regions of the genome. In my
fourth chapter, I consider starling invasions from the perspective of science and
technology studies (STS), tracing where human interference and potential biases shape
both the practice of invasion science. I consider how my own genomic analyses depend

on assumptions in both population genetic methods and invasion theory. Overall, this dissertation attempts to reconstruct the evolutionary history of starling invasions, with a focus on the invasion in North America.

BIOGRAPHICAL SKETCH

Natalie earned her B.A. in Biology with a minor in Gender Studies (then Women's Studies) from St. Olaf College in 2013. Her interests in pursuing graduate work began in a Vertebrate Biology course taught by Dr. Steven Freedberg, who later became her advisor for her honor's thesis on the spatial ecology of painted turtles in Minnesota waterways. Steve's passion for natural history was infectious, and his encouragement guided her to an M.A. in Conservation Biology from Columbia University, completed in 2015. At Columbia, Dr. Dustin Rubenstein taught Natalie how to critique her own interpretations, a skill developed in her Master's thesis that she leans on heavily in the latter portion of this dissertation. The training and support she received from Dustin led her to the Ecology and Evolutionary Biology program at Cornell University, where she completed the following dissertation.

ACKNOWLEDGMENTS

This dissertation represents eight years of personal and professional growth that cannot be contained in this document. For every page, it would take several more to explain what that work means to me, not because of its academic content, but because each page stands in for a series of lessons I learned about myself. I am finishing this dissertation hungry for more: a fortuitous attitude given where I was in the middle of my dissertation. To everyone I thank in the coming pages: each of you helped to build my path, and I aspire to share such energy and care in my future.

This dissertation would have been impossible for me to complete without the help of many colleagues and mentors. The non-model genomics journal club, its leaders Kelly Zamudio, Matt Hare, and Nina Therkildsen, and its many participants helped orient me to using genomic tools and interpreting at-times messy data. I learned how to embrace what I didn't know and identify lessons that I could apply in my own research, and I'm especially grateful to Nick Fletcher, Allison Tracy, Nick Mason, Coby McDonald, and Cinnamon Mittan for modeling how to learn together. In learning to apply genomic methods to my own data, Jennifer Walsh and Bronwyn Butcher were invaluable and patient teachers. Jen sat with me as I second-guessed each line of code, and has always responded promptly to any question I've had about study design, bioinformatic analysis, and crises of confidence. Bronwyn has been a constant source of wisdom in both lab work and work-life balance throughout my entire PhD. Both Bronwyn and Jen possess a dependability that I aspire to mimic, and each time I found myself in rough patches personally or professionally, they have each provided a landing spot when I most needed it. Lab post-docs and staff Dave Toews, Scott Taylor, Shawn Billerman, Rusty Ligon, Eliot Miller, Sarah Wagner, Gavin Leighton, Sabrina

McNew, Gemma Clucas, and Carrie Branch all taught me something about how to exist as a researcher and a person. Lovette Lab graduate students Amelia Demery, Stepfanie Aguillon, Jake Berv, Petra Deane, and Nick Mason are a community of brilliant and thoughtful humans, and I use lessons learned from each one of them in all the work I do. The EEB community more broadly is a buzzing hive where everyone I met could teach me skills for research and beyond, and I am leaving well-prepared for my future.

Two of my chapters are collaborations with Lee Ann Rollins and Katarina Stuart, who have been endlessly kind and helpful women that I am grateful to consider both friends and colleagues. Lee Ann is a generous, attentive researcher and person whose attitude towards science is one I admire and hope to emulate. The inimitable Kat Stuart possesses more organizational skill and determination than I may ever have, and I feel lucky to have worked closely with her over the last several years. Her attention to detail in our parallel projects on the genomics of European starlings has meant that I could always ask Kat to talk through a tricky bioinformatic problem or some perplexing data.

My dissertation committee, Kelly Zamudio, Maren Vitousek, Nina Therkildsen, and Suman Seth, have each encouraged me each year to develop my research and my advocacy. Their steady confidence in my work carried me through many years of re-analyzing and reflecting on my research, even as I maintained several parallel projects in advocacy and organizing work. Maren Vitousek and Nina Therkildsen have been dependable resources for any and all of my research- or career-related concerns over the years. Kelly Zamudio recognized how I eager I was for interdisciplinary scholarship long before I took that interest seriously, and has been a stalwart supporter in believing I could carry that interest into a future career. Suman Seth became a mentor to me only in the last few years of my PhD—almost exactly when he began as chair of Science and Technology Studies—and has been a kind and thoughtful role model. Suman's

insightful questions have always spurred me towards goals I thought were unreachable.

In teaching over the last several years, I learned from my students how to show up during tough times and how to reconsider my own assumptions. Students in every one of the semesters I taught asked me to reflect on how I wanted to engage with others in learning together, and frequently, students reminded me—either directly or implicitly—that each statement I made was an active choice of how to present myself and the academic community I represented. I find teaching to be both challenge and reward, and I've been lucky to learn from several teachers over the years. Abby Drake, Jeremy Searle, Bob Reed, Gregor Siegmund, Stepfanie Aguillon, Nick Fletcher, Nick Mason, Lizzie Lombardi, and Lina Arcila Hernandez have all been taught me something about how I want to show up in the classroom.

It is not immediately obvious how this dissertation was shaped by the advocacy and organizing projects I took part in. Aubrie James, Sue Pierre, and Coby McDonald were early inspirations as I watched their work in EEB, and I have been lucky to learn alongside them. Aubrie and Coby in particular inspire me to be brave and creative, and I would not be the person I am today without their encouragement. I worked with many groups over the eight years I spent at Cornell, and there is not a single person who didn't teach me something about collaborative organizing. Cornell's Intergroup Dialogue Project has grown immensely since I took its first graduate course in 2016, and after my own training I continued to facilitate dialogue sessions with them every year. Adi Grabiner-Keinan, Christine Barker, Alex Brown, Stephen Kim, Jumoke Warritay, Rachel Sumner, Jazlin Gomez Garner, Janani Hariharan, and many more facilitators and staff have restored my faith in our ability to grow across differences countless times, and I consider IDP and its lessons the ground where I began to grow as an organizer. I

worked with many other teams to both educate ourselves and advocate for each others'
needs, starting with Graduate Women in Science and then serving several different
roles in Cornell Graduate Students United. I've organized with many folks in many
spaces over the years, but Alec Pollak, Catie Ball, Xander Lacrampe, Monique Pipkin,
Amelia Demery, Maddie Ore, Dave Blatter, Maggie Pacheco, Katherine Quinn, Juhwan
Seo, Jacy Tackett, and so many more taught me what kindness looks like in community.
Labor organizing taught me to challenge myself in ways I could not have imagined, and
the courage I gathered from organizing led me to expand my dissertation and future
research beyond what I thought possible.

In this final section, I want to thank several people who are my pillars. The most
foundational of pillars are my parents, Michael and Brigitte, who have devoted their
lives to ensuring my siblings and I can thrive and have always encouraged me to
imagine a future in which we flourish. Both my parents committed themselves to
building a home and a family where each of us could depend on each other, and where
we might feel not only secure but confident in our capacity to succeed. Little did they
know that the summer backpacking trips of my childhood would lend me skills I used
in my future career, or that I would find a way to finagle my love of animals and of
people into an academic career. My parents' support has allowed to push beyond what
I thought possible countless times, and I am forever grateful for their belief in me.

Two friends in particular have been steadfast confidantes and wise role models
as I've worked through this PhD. For many years, Stepfanie Aguillon has been honest
with me when I most need a reality check. She has encouraged me to take risks in my
academic work, and to protect my time and energy when I needed to rest. Alec Pollak
inspires me to stand alongside her on all fronts of the work we undertake, and I trust
her to challenge whatever limitations I've placed on myself. Early in our friendship,

Alec noticed what I had not yet named to myself, and she has consistently held a vision for my future when my own confidence is rattled. I am lucky to count Stepfanie and Alec as both friends and colleagues.

My partner, Jeff Lapennas, and I met only in the final years of my dissertation, but he quickly became part of my core support in all of my projects. Where I am easily carried away by new ideas or problems demanding a rapid solution, Jeff is solid, generous, and reliable. Early in our relationship, he called himself a "slow-moving object" in jest, but to my surprise I've learned that is an asset much of the time. Jeff explains my research more compellingly than I ever can; whereas I can get carried away with whatever niche of my work seems exciting that day, people not only seem to understand the core of my work after he's explained it, but they also seem to want to hear more. Jeff's questions about my work continue to drive my growing research program post-PhD, and I hope I will always be able to remind him that he would make a stellar academic if he ever chose to pursue that career. I could not have dreamed a partner as kind and reliable as Jeff, and I'm endlessly thankful to spend each day with him.

My little dog Steve deserves all the smelly treats and chipmunk chases he can imagine after several years of putting off the adventure to spend just a few more hours on the computer. Steve is a joyful, impatient terrier, and when I adopted him, I expected a convalescing puppy who may have never been particularly mobile. To my surprise, he would like nothing better than to sprint around the woods for hours each day, followed by a long nap on my lap. He's aged into increasingly grouchy middle age during my PhD, but his reward for patiently providing comic relief and emotional support will be as many adventures and snuggles as he'll tolerate.

Finally, I want to thank Irby Lovette. "Advisor" is not a sufficient label for how Irby supported me through personal and professional challenges. A theme of my dissertation has been testing my own boundaries, and Irby has not only understood this but has facilitated my growth in whichever direction I choose. From the beginning of my PhD, Irby has encouraged me to take risks and experiment with my own work, encouraging a freedom and independence that I needed. His steadfast support no matter which barrier I decided to run up against has meant more to me than he can know, and I would be neither the researcher nor the person I am without his presence in my life.

My loved ones' unconditional support and belief in my ability to unite all my disparate interests into work that I truly love has meant that I can show up more wholly than I could have hoped. This dissertation represents years of figuring out how to be the person I wanted to become, and I am thankful to be leaving this PhD with a bit more courage and wisdom than I had.

TABLE OF CONTENTS

CHAPTER 1: A REVIEW OF THE GLOBAL INVASION HISTORY AND NATIVE DECLINE OF THE EUROPEAN STARLING

ABSTRACT

Few invasive birds are as globally successful as the European Starling (*Sturnus vulgaris*). Native to the Palearctic, the starling has been introduced to North and South America, southern Africa, Australia, and the Pacific Islands, and its global success allows us to explore species traits that may contribute to its invasion success. Coupling the rich studies of life history and more recent explorations of genomic variation among invasions, we illustrate how eco-evolutionary dynamics shape the invasion success of this long-studied and well-distributed species. Especially informative is the comparison between Australian and North American invasions, because these populations colonized novel ranges concurrently and exhibit shared signals of selection despite distinct population histories. In this review, we describe population dynamics in the native and invasive ranges, identify putatively selected traits that may influence the starling's spread, and suggest possible determinants of starling success world-wide. We also identify future opportunities to utilize this species as a model for avian invasion research.

1. THE STARLING AS AN ECO-EVOLUTIONARY MODEL

The ecological and economic impacts of invasive species are a growing concern in our

globalized world. Increasing travel among continents creates and reinforces invasion pathways, resulting in a great number of alien species becoming established and spreading in novel ranges (Turbelin *et al.* 2017). Despite decades of work, predicting which species might become invasive when others do not remains a challenge, and a thorough review of factors that promote invasion success may bring us closer to this aim. Genotypic variation, species niche, local abundance, and environmental features all influence invasion success (Colautti & Barrett 2013). Fundamental properties of invasions (e.g., propagule pressure, genetic bottlenecks and variation, etc.) and adaptive evolution of novel strategies (e.g., dispersal strategies, breeding behavior) may work in concert to facilitate successful invasions (Duncan *et al.* 2003; Redding *et al.* 2019; Fristoe *et al.* 2021). However, climate, ecosystem composition, and human activity may also influence how invasive species establish and spread (Liu *et al.* 2020; Miller *et al.* 2021). In reality, both intrinsic and extrinsic conditions are fundamental to the long-term success of an invasion (Colautti *et al.* 2017).

Few avian invaders have been as globally successful as the European or Common Starling (*Sturnus vulgaris*) (hereafter referred to simply the 'starling'; Box 1.1). Starlings are generalists that thrive in a wide array of environments—particularly those altered by humans— and have a costly impact on agriculture and native ecosystems (Linz *et al.* 2017) that has encouraged much of the interest in this invader. Native to the Palearctic, the starling has been introduced to North and South America, southern Africa, Australia, and the Pacific Islands, and has been listed as one of the world's 100 worst invasive alien species (Lowe *et al.* 2000). In addition, starlings are a widely-used model in laboratory studies (Asher & Bateson 2008), and linking such thorough studies

of starling traits with wild observations may help to clarify mechanisms that support invasiveness.

Box 1.1. Why the starling?

Despite its invasion success, the starling is not the only avian species to invade nearly every continent world-wide: in fact, the ubiquitous House Sparrow (*Passer domesticus*) is similarly successful in a wide range of environments (Hanson *et al.* 2020). Each species has its own advantages as a representative 'model' species (cf. Hanson *et al.*'s careful discussion of the consequences of labeling a species as a 'model'), To date, each species has been well-studied: a taxonomic search on NCBI Taxonomy indicates 761 "bio-samples" of *S. vulgaris* for the 179 bio-samples of *P. domesticus*, where a bio-sample indicates an independent sequencing project (Figure B1). A taxonomic search on the Web of Science yields 2,368 results for "*Sturnus vulgaris*" and 2,034 results for "*Passer domesticus*." Because both species are well-studied, comparisons among these avian invaders might yield insight into taxonomically-broad and/or species-specific strategies to promote invasion success.
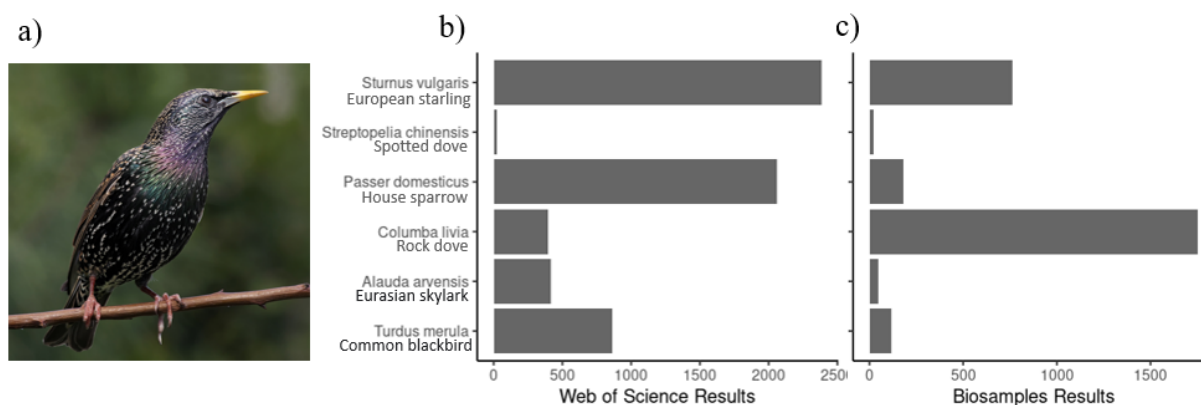
Figure B1: Common invasive avian species, and their Web of Science and NCBI bio-

sample search result counts (accessed on July 17, 2021).

Nearly all studies of invasion genetics examine the genetic diversity of an invasive population, addressing the paradox of invasion, where species thrive despite a loss of diversity (Dlugosch *et al.* 2015; Estoup *et al.* 2016). Repeated invasion success across starlings' many introduction sites presents an opportunity to examine patterns in how the species undergoes adaptation post introduction bottleneck. Sequencing advances over the last decade have made genomic approaches more accessible to non-model species such as the starling (North *et al.* 2021), enabling the use of genetic analyses across the starling's global range to examine the proximate mechanisms that may contribute to this invasive species' success. Often the focus of such studies is invasive species' ability to undergo rapid evolution in their novel range despite apparent low genetic diversity. Despite their invasion success, starling numbers are declining in both their native range (Smith *et al.* 2012; Heldbjerg *et al.* 2019) and in the invasive North American population (Rosenberg *et al.* 2019). Understanding how starlings thrive in invasive populations may inform conservation efforts in the native range and control in the invasive ranges.

Here we synthesize extensive research on starling life history with genetic and genomic evidence from the native and invasive range to identify factors influencing invasion success in the starling. We first describe the history of each native and invasive starling population, and then suggest how eco-evolutionary feedback might continue to shape range expansion and/or population declines. The starling's dynamic invasion history presents a wild system where concurrent, replicated invasions (Australia, North

America, New Zealand and South Africa) as well as more recent invasions (into Argentina) enable us to distinguish between population-specific and species-wide strategies to support invasion success. A holistic perspective on starling invasions may yield additional hypotheses that clarify how this particular invasive species came to thrive in nearly every continent. Despite their longstanding title as a prolific pest, the starling continues to decline in numbers globally. As a model of invasion, the starling is a dynamic system, and we articulate how factors supporting invasion success might interact as a step towards predicting future shifts in range and abundance of the starling world-wide.

2. NATIVE STARLING DISTRIBUTION AND POPULATION DYNAMICS

In its native range, *Sturnus vulgaris* comprises 11-13 subspecies, and its distribution is thought to be primarily a result of changes in forest coverage and aridity tied to major climate shifts at 16-14, 10, and 6-5 Mya (Fortelius *et al.* 2002; Zuccon *et al.* 2008). The European starling's native range extends across the Palearctic (Figure 1.1). The non-breeding range extends as far as Russia (Sandakova *et al.* 2018), whereas the breeding range extends southwest into Pakistan and further still into Israel (Mahmood *et al.* 2013).

Since the mid-19th century, this species has been slowly expanding its Eurasian range (Feare 1984). Colonisation of Iceland occurred in the late 1930's (Anderson 1992), and a small number of starlings have been reported to winter in Hong Kong since as early as the 1970's (Webster 1975). This range expansion is likely a result of increased

anthropogenic land alteration, climate change, and possibly a decrease in competing species in newly colonized ranges (Webster 1975). In one region of this recent expansion, *S. vulgaris* has come back into contact with its closest relative, the spotless starling (*Sturnus unicolor*), from which it only split ~40 Mya (Feare 1984). In this contact zone in the Iberian Peninsula, the two species seem to interbreed readily: in fact, allozyme studies show that genetic distances between *S. unicolor* populations are larger than genetic distances between the two species (Cruz-Cardiel *et al.* 1997). Because gene flow between two species that share essentially the same niche is maintained, competition may not play too great a role—at least in this one part of the starling's range—but to verify this hypothesis, further studies of potential hybridization between these sister species are needed.
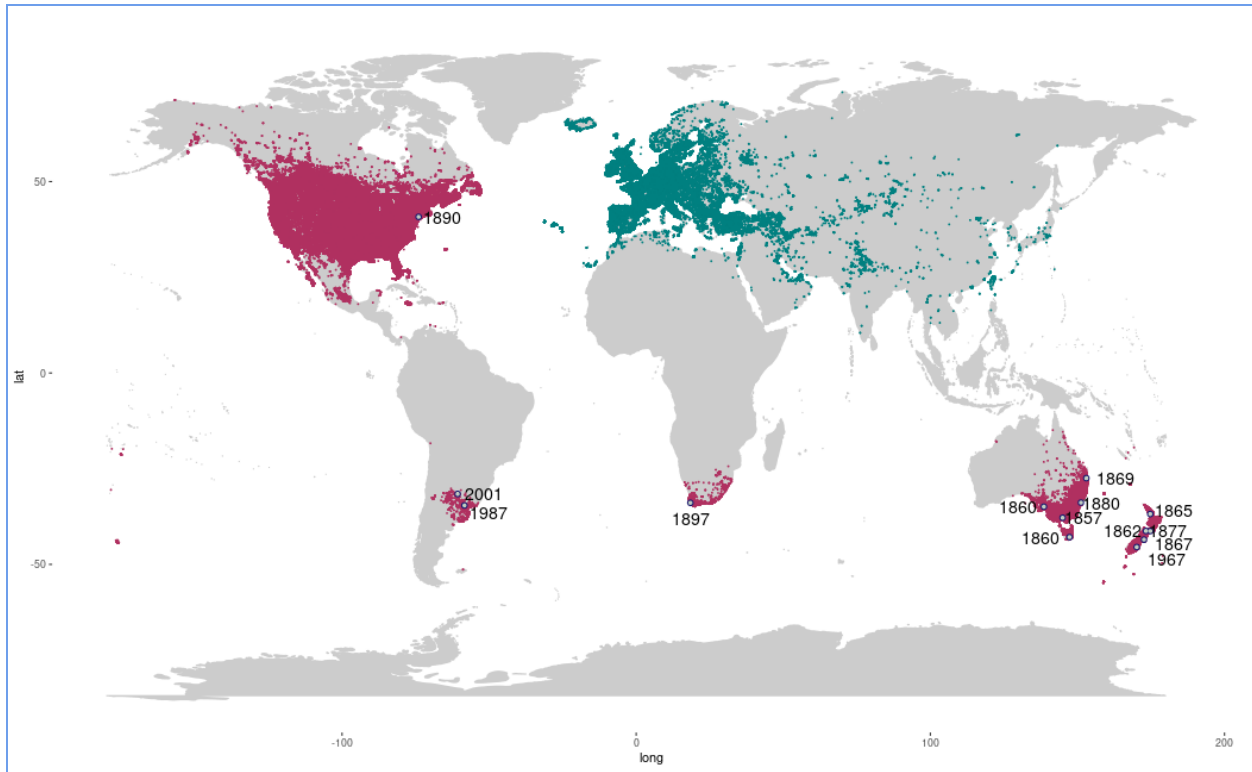
Figure 1.1. Starling distribution map according to eBird sightings data (Sullivan *et al.* 2009) (retrieved Feb 2018). Native marked in teal, invasive in maroon. First introduction date at an introduction site is marked with a blue circle.

One of the major puzzles in starling biology is that in the native range, starling numbers have dwindled even as its geographical range expands. Since 1964 starling population size in Great Britain has declined by more than 50% This decline is particularly notable in livestock farming areas of southwest Britain, with the breeding population estimated to be 8.5 million as of 2005 (Robinson *et al.* 2005). Finnish starlings underwent a dramatic 90% population decrease from 1970-1985, corresponding to a widespread abandonment of cattle farming across the country (Rintala *et al.* 2003). Similarly, a shift to indoor cattle husbandry in Denmark may have contributed to the 60% overall decline in starling abundance between 1976 and 2015 (Heldbjerg *et al.* 2016). Data on starling foraging behavior may explain this close association between cattle farming and starling population size: a move toward modern cattle rearing has resulted in changes to available pasture, which may indirectly influence availability of small invertebrate prey (Chamberlain *et al.* 2000; Wretenberg *et al.* 2006) on which starlings particularly rely during breeding season (Feare 1984). In addition, easy access to livestock feed may be a large component of the starling diet, and modern cattle rearing processes seek to minimize feed loss to pests such as starlings (Linz *et al.* 2017). The starling decline is not unique: recent decades have brought a decline in farmland birds across the UK and Europe (Gregory *et al.* 2002; Wretenberg *et al.* 2007).

Although availability of food may directly explain starling abundance, changes in starling foraging behavior may have also accelerated the decline of starling populations via greater mortality in hatch-year or juvenile starlings. Demographic studies on Netherlands' declining starling populations identified that the main driver of population growth or decline was juvenile survival (Versluijs *et al.* 2016). Similarly,

first-year survival rates were linked to British population decline from the 1970's to the present (Freeman *et al.* 2007). We can hypothesize that a reduction in invertebrate abundance within grazing environments drives birds to store more fat, leaving them more vulnerable to predation, with the loss of first-year birds that have not reached their first mating season driving overall population declines. Large scale changes in starling numbers across their native range are hence likely driven by demographic changes in breeding populations due to juvenile loss (Smith *et al.* 2012). However, variation in survivorship may also be explained by a starvation-predation risk trade-off: accumulation of fat leaves birds more vulnerable to predation (Macleod *et al.* 2008). Starlings maintain a lower body mass during favorable foraging conditions, when resources are abundant. Within the United Kingdom, starlings with a higher average mass displayed a greater decline in numbers (Macleod *et al.* 2008), which may indicate foraging environment quality impacts survivorship, or alternatively a causative factor (e.g. climate) drives both provisioning and survivorship.

Much of the existing literature that touches on starling genetic diversity in the native range, while all supporting higher genetic diversity in the native range as expected, have been conducted with a focus on invasive range genetics (Rollins *et al.* 2011; Bodt *et al.* 2020; Hofmeister *et al.* 2021a). Early allozyme sequencing work reported very little genetic variation across *S. vulgaris* in the native range (Evans 1980; Ross 1983; Neves *et al.* 2009), however allozymes have poor resolution and this should be reassessed with modern markers. Analyses of mitochondrial control region sequence data suggest very high diversity within the native range (17 haplotypes from 27 individuals sampled in one locality; (Rollins *et al.* 2011)). Early morphological studies

indicate some population structure within *Sturnus vulgaris*: subtle differences in adult

plumage iridescence among starling populations across the native range have led to the

classification of 11-13 subspecies (Pateff & Stresemann 1947; Feare 1984). Given the lack

of genetic and morphological evidence for within-species differentiation, most studies

of this species forgo subspecies identification. Thus, from this point onwards we will

focus on *S. vulgaris L.*, unless otherwise mentioned.

Additional studies to document population structure and genetic consequences

of changing population size in the native range would help to predict how the starling

population might respond to continued anthropogenic changes in land use. In

particular, additional studies of population size and demography changes across the

native range but especially outside Europe (i.e., Northern Africa, Middle East, and Asia)

would help resolve the dynamic relationship between environmental change and

starling demography. Certainly, environmental conditions and range expansion impact

the evolutionary potential of the native starling population, but perhaps the greatest

difference between native and invasive populations is management strategies (e.g.,

culling). The native range population thus serves as a point of comparison where

selective regimes and demographic shifts common in invasions operate differently.

3. INVASIVE STARLING DISTRIBUTION AND POPULATION DYNAMICS

3.1 Australia

In Australia, starlings were introduced to control invertebrate agricultural pests, and as

part of efforts of acclimatisation societies similar to those in the US (Woolnough *et al.*

2006). There were undoubtedly many small private undocumented introductions,

however the first officially documented starling import was the 1856 introduction of an unknown number of birds to New South Wales (Higgins *et al.* 2006). Over the next several years, there were several introductions in Melbourne, Victoria, and the species was described as 'well established' in Victoria by as early as 1963 (Jenkins 1977). Subsequently, there were numerous introductions during the late 19th century to both New South Wales and Victoria, as well as to South Australia, Queensland and Tasmania (Table 1: global starling introductions). The exact releases that contributed to starlings' widespread success in Australia are thought to be near the capitals of Victoria, New South Wales, or South Australia (Stuart & Cardilini *et al.* 2021a). From the early years of their introduction to Australia, starlings increased rapidly in both range and population numbers (Jenkins 1977).

By the mid 20th century, most of the south and eastern states of Australia were colonised, including Tasmania, with the birds being most prolific in Victoria (Long 1981). There have even been introductions to islands, including Macquarie Island which is situated equidistant from New Zealand and Antarctica (Raymond *et al.* 2011). Western Australia (WA) has remained largely starling free, due jointly to the natural barrier provided by the Nullarbor Plain, and ongoing control efforts by the WA Department of Primary Industries and Regional Development (DPIRD) since 1971 (Campbell *et al.* 2016), and which was ramped up in the mid 2000's (Woolnough *et al.* 2005). Some of these incursions occurred in regional areas far from the main invasion (e.g. Broome on the northwest coast of Western Australia); however, it is presumed that these individuals arrived via unintentional anthropogenic means (Rollins *et al.* 2009, 2011). While the state has remained largely starling free for over a decade, the last few

11

years have seen a sharp increase in the number of new 'founder' starlings spotted in the area Munglinup and surrounds. All birds captured were females in breeding condition (Rollins, pers. comm.); since starlings have a female-biased dispersal (Rollins *et al.* 2009), these individuals are likely to represent new incursions into this area. Cost-benefit analyses suggest the short-term costs of control are a good investment in Western Australia, given the large potential costs to agriculture (Campbell *et al.* 2016).

Australian starlings do not undertake large-scale seasonal migration in Australia as is known in North America and their native ranges. Instead, they have a relatively steady range edge expansion rate of 20.7 km/year since establishment (Hui & Richardson 2017). It is thought the range expansion in Australia, particularly at the present day range edge, is driven by birds seeking new nesting sites, rather than seasonal visitation (Long 1981). Within subpopulations, small-scale regional movement attributed to food seeking drives some movement between subpopulations, however, a small number of birds are found to disperse long distances; up to 1000 km (Waterman *et al.* 2008).

Population genetics studies of the Australian invasion using microsatellites suggests four populations exist: two small, distinct incursions into Western Australia and two large populations, one in South Australia, and another in Victoria, Tasmania, and New South Wales (Rollins *et al.* 2009). Reduced representation sequencing (genotyping-by-sequencing) across the Australian range paints a slightly different picture, with Victoria, South Australia, and the easternmost incursion in Western Australia forming one large genetic group, New South Wales and Queensland forming a second, and two geographically restricted groupings identified (one is the

westernmost incursion in Western Australia identified in (Rollins *et al.* 2011) and the second in inland New South Wales) (Stuart & Cardilini *et al.* 2021a). Across the total range, genetic diversity in Australia is estimated to be lower than in native range samples (Rollins *et al.* 2011).

Within Australia, patterns of genetic diversity are in accordance with residence time. The highest genetic diversity is reported from sampling sites near the three primary introduction sites, and lowest genetic diversity was found at the range edge (Rollins *et al.* 2009, 2011; Stuart & Cardilini *et al.* 2021a). Across this invasion, there is a significant relationship between time of establishment and genetic diversity (Rollins *et al.* 2009). In addition to the effects of genetic drift often found in small, range-expanding populations, there is evidence of selection occurring in this invasion: a rapid shift in frequencies of mitochondrial DNA variants on the range edge is best explained by selection acting within heteroplastic individuals and morphological studies indicate clinal variation suggesting adaptation has occurred (Cardilini *et al.* 2016; Rollins *et al.* 2016). Environmental correlation approaches have identified allelic differences in coding regions related to a range of biological functions (e.g. immune response, metabolism), indicating putative loci under selection (Stuart & Cardilini *et al.* 2021a). Additionally, the Australian population appears to be undergoing spatial sorting, where dispersal-enhancing traits accumulate at the leading edge of an expanding population (Phair *et al.* 2018), and mitochondrial sequencing shows mixed signals of expansion (Rollins *et al.* 2011).

3.2 New Zealand

Similar to the starling introductions of North America and Australia, the starlings of New Zealand were introduced from Britain due to acclimatisation efforts (Thomson 1922). The introductions were well documented, with records of 14 major introductions from 1862-1883 across New Zealand, totaling to around 650 individuals (Table 1: global starling introductions) (Thomson 1922). Starlings' success in New Zealand was assisted by large-scale translocations within the country (Pipek *et al.* 2019). Much like with the Australian introduction, interest in the species for pest control may have also led to smaller private introductions to farming properties.

The starling is now the most widespread bird species in New Zealand: the starlings' range covers the entire country, from the northern outlying island of Kermadecs, to the south outlying Macquarie Islands (Williams 1953; Flux & Flux 1981). In 1886, the species was recorded to occur in the "hundreds of thousands" (Thomson 1922). The following decades brought with them a small population decrease, which is thought to be at least in part due to a decrease in nest availability (Flux & Flux 1981). As with the Australian invasion, the New Zealand starlings show no evidence of large-scale migration throughout the country's range (Ross 1983). Populations in agriculturally heavy areas have been actively monitored due to their ongoing success in managing major crop pests (such as grass grubs) (Coleman 1972).

The New Zealand invasion has also served as a stepping-stone for subsequent invasions in the Pacific islands. The Fiji starling populations were thought to be established in the mid 1920's, and have now spread to Tonga (Watling & Talbot-Kelly 1982). Fijian starlings are thought to have dispersed via natural means from the

14

Kermadec Islands, which lie equidistant to New Zealand (Williams 1953; Flux & Flux 1981). Further spread of the starlings through these tropical islands may be impeded by their physiological limits (Watling & Talbot-Kelly 1982), a possibility we discuss in Section 5.3 of this review.

According to allozyme data, the New Zealand starling invasion has retained much of the genetic diversity found in the native range, and genetic distances between invasive and native populations fell within the range of distances among allopatric populations within other avian species (Ross 1983). There was only a slight loss of genetic variation during the colonisation of New Zealand, associated with the loss of rare alleles from within the native Britain population (Ross 1983). This mild genetic bottleneck is no doubt a result of the large founding populations and multiple introductions at many of the founding sites. There is however noteworthy divergence in the degree of inter-population differentiation in each country: native populations show lower genetic distances and F-statistics than the invasive (Ross 1983). Relatively high levels of differentiation across the New Zealand invasion suggests subpopulations face geographic isolation from one another, owing to both the country's mountainous terrain, and the bird's preference for agricultural and urban areas, coupled with a lack of migratory movements (Ross 1983). To date, updated methodologies have not yet been used to clarify population structure in this invasion, and higher-resolution tools could validate and/or extend these findings. A study of starling population genetics across New Zealand and the Pacific Islands would provide an interesting avenue to look at consecutive bottlenecks during island hopping and how this impacts adaption to these novel environments.

3.3 North America

Acclimatization societies and individuals attempted to introduce starlings several times in North America, although only one introduction succeeded (Table 1: global starling introductions) (Cooke 1928). Wealthy businessman Eugene Schieffelin released 80 individuals in March 1890 and 40 more in April 1891 to Central Park, where these individuals reproduced and began to spread almost immediately. This New York population was the first introduction to establish a breeding population, after which the species spread rapidly throughout the continent (Kalmbach *et al.* 1921; Linz *et al.* 2017).

Although starlings remain common in North America, the population has declined by 49% since the 1970s, with a current size of 85.1 million birds (Rosenberg *et al.* 2019). These estimates--based on a model using direct counts of individuals--contrast with the often-repeated North American population size of 140-200 million (Linz *et al.* 2017). Genetic evidence can tell us about consequences of this decline for evolutionary potential in the North American population. Demographic models indicate a slight decline in effective population size, based on both reduced-representation markers (Hofmeister *et al.* 2021b) and all variable sites in the genome (Hofmeister *et al.* 2021a). In contrast to models that use site-frequency spectra to reconstruct demographic history, faster-evolving mitochondrial evidence indicates that the North American population may now be expanding (Bodt *et al.* 2020).

Comparisons of genetic diversity between invasive and native ranges provide another perspective on changing population sizes. Heterozygosity in allozyme data among the North American and native range populations suggests little evidence of a bottleneck (Cabe 1998). Mitochondrial analysis shows that both nucleotide and

haplotype diversity is lower in the North American range compared to the UK

population, further supporting the evidence for a slight genetic bottleneck in North

America (Bodt *et al.* 2020). Mitochondrial haplotype diversity is nevertheless higher in

North America in comparison to other invasive populations of Australia and South

Africa (Bodt *et al.* 2020). This higher level of genetic diversity in the North American

invasion may provide greater standing variation upon which selection can act

(Hofmeister *et al.* 2021b, a).

Starling expansion in North America likely relied on urban and agricultural

areas to support a large enough breeding population to facilitate expansion. Historical

records indicate that expansion accelerated after range-edge populations established in

a city (e.g., Philadelphia in 1910). Elevation also seems have imposed a serious barrier to

starling spread in North America, where population expansion stalled when birds

reached the Allegheny Mountains in 1911, the Adirondacks of New York and White

Mountains of Vermont in 1922 (Forbush 1915; Kalmbach *et al.* 1921; Cooke 1928), and

again when it reached the 1000m elevation mark in the Midwest of the US in 1930

(Hoffman 1930; Dickerson 1938). In accordance with this history, a genome-wide scan

did find genotypic associations between environmental characteristics, particularly

elevation, precipitation and temperature (Hofmeister *et al.* 2021b). Whether starling

expansion was in fact supported by adaptive evolution to novel elevational barriers

requires both more thorough genomic investigation and functional validation.

Outside of North America, starlings have now expanded into Cuba and the

Bahamas (Linz *et al.* 2017). Individuals are now as far north as the Arctic Circle and are

slowly continuing to expand their range southward (eBird 2021). Starlings were spotted

nesting in south-central Tamaulipas, Mexico for the first time in 2006 (Brush 2009).

Additionally, starlings were purposely introduced to Jamaica for crop damage

mitigation, likely sourced from the North America introduction (Lever 2010). By what

means starlings continue to expand into novel environmental conditions is a question

that we revisit in Section 5 of this review.

Starling expansion through the heterogeneous habitat of North America (as with

other invasive populations) depended on successful survival and reproduction; in later

sections we discuss how dispersal and migration (4.1), breeding strategies (4.2), and

potentially local adaptation to environmental conditions (5.1-5.3) support the starling's

spread. Here, we note that starlings' movement varies across this landscape: starlings in

the western U.S. tend to move regionally whereas birds sampled in the eastern U.S. may

migrate or disperse far greater distances (Werner *et al.* 2020). Banding efforts across

North America also indicate haphazard migratory patterns (Brewer 2010). These

patterns of movement likely contributed to the near-random mating across the North

American range which would eliminate geographical differentiation (Cabe 1999;

Hofmeister *et al.* 2021b).

Rapid evolution in this invasive population is further supported by

morphological differences in individuals sampled across the North American starling

invasion. For example, wing pointedness has decreased over the last 120 years since

colonisation, which might allow individual starlings to move greater distances (Bitton &

Graham 2015). Further, North American starlings sampled at intermediate latitudes

(which may best approximate the native range climate) had the greatest lipid reserves

overwinter, and was correlated with mean temperature in July and January (Blem 1981).

It is possible that the migration status may explain this relationship between morphology and environment, as larger northern individuals move south for the winter. While coarse morphology is conflated with a plethora of variables, these studies provide direction for more thorough molecular and experimental studies.

3.4 South Africa

The South African starling invasion resulted from 18 birds, introduced in 1897 to Cape Town by Cecil Rhodes, the then Prime Minister of the British Cape Colony (Cooper & Underhill 1991; Harrison & Cherry) (Table 1.1: global starling introductions). These birds were reportedly caught in Britain during winter months (Winterbottom & Liversidge 1954). From the introduction site, starlings spread eastwards across the Cape Flats. The natural mountainous barriers plausibly contributed to the initial slow expansion rate (Rensburg 2014). The range expansion rate increased, and starlings reached the Kwazulu Natal Province during the early 2000's; the species' present day range covers up to Johannesburg to the north and southern Namibia to the west (Berthouly-Salazar *et al.* 2013; Rensburg 2014). The range's eastward expansion has been largely enabled by the corridor provided by human habitation (Berthouly-Salazar *et al.* 2013). The rate of range spread in the South African starling invasion has increased since their introduction, from 6.1 km/year to 25.7 km/year (Hui *et al.* 2012). Despite this, mitochondrial sequencing of the South African population show no evidence of expansion (Bodt *et al.* 2020).

Despite the small size of the South African introduction, estimates from microsatellite data suggest this invasion has similar levels of genetic diversity to that of

the UK samples used in the study (Berthouly-Salazar *et al.* 2013). However, it must be noted here that all UK individuals were sourced from the one population and were unlikely to capture the full range of genetic diversity in the expansive native range. South Africa has moderate levels of mitochondrial haplotype diversity, less than of the native range and North America, but greater than Australia (Berthouly-Salazar *et al.* 2013; Bodt *et al.* 2020). This may indicate a heavily sex-biased introduction (more females than males), but this high mitochondrial diversity could also be due to undocumented introductions (as the documented introduction size is 18 individuals) or in situ mutations (Berthouly-Salazar *et al.* 2013). Analysis of mitochondrial control region sequence indicated no population structure but did find a subtle decrease in genetic diversity towards the range edge (Berthouly-Salazar *et al.* 2013). Despite the genetic patterns underlying the invasion gradient, the South African invasion displays no patterns of spatial sorting, unlike the Australian invasion (Phair *et al.* 2018). This may be due to long distance dispersal, which would maintain genetic homogeneity (Berthouly-Salazar *et al.* 2013). Despite a lack of spatial sorting, there is increased genetic distance with higher winter precipitation, indicating gene flow is limited where precipitation is high in winter and low in summer (Berthouly-Salazar *et al.* 2013). This has resulted in two subpopulations around George and Mossel Bay (300km east of introduction site), despite the lack of population subdivision elsewhere in the invasion. This area is associated with a sharp change in climate conditions, particularly winter precipitation (Berthouly-Salazar *et al.* 2013).

3.5 South America

More recently than the other introductions described above, starlings made their way to South America. In 1949, five starling individuals were brought over by ship from England and alighted in Lago de Maracaibo, Venezuela, though the success of these individuals remains unknown (Long 1981) (Table 1: global starling introductions). In 1987, starlings were spotted in Buenos Aires, Argentina, in the wooded areas of the Palermo district (Peris *et al.* 2005), thought to be an introduction via the pet trade using birds imported from North America (Navas 2002, Fiorini *et al.* 2021). Despite prompt eradication efforts, the species remained and further populations were spotted in 2001 around 400km north of Buenos Aires, near Sante Fe (Peris *et al.* 2005; Navas 2014). Range expansion within the South American invasive starling population is strongly associated with urban areas (Zufiaurre *et al.* 2016). Starlings make use of novel nesting sites available in the human modified environment though retaining a preference for natural nesting sites (Peris *et al.* 2005). The starling invasion stays roughly within 30 km of the coast (Peris *et al.* 2005), with small urban centers facilitating continual range expansion into regional areas (Zufiaurre *et al.* 2016).

The southernmost distribution stretches to the southern regions of Uruguay and central Argentina (Silva *et al.* 2017). Starlings reached Brazil in late 2016 (Silva *et al.* 2017), and are most abundant in grasslands, mirroring the habitat preference of functionally comparable native species (Palacio *et al.* 2016). The Brazilian range currently covers an area greater than 65,000 km^2 in the Pampas region, with the rate of range expansion having increased linearly from 7.5 km/year in 2005 to 22.2 km/year in 2016 (Zufiaurre *et al.* 2016). This acceleration of range expansion after establishment has

also been documented in the Australian and South African invasions (Peris *et al.* 2005; Zufiaurre *et al.* 2016). Mitochondrial analysis of birds collected in Buenos Aires, Argentina, were found to have reduced haplotype diversity compared to North America and the native range, though with several novel haplotypes identified (Fiorini *et al.* 2021). This same study also noted increased wing primary feather asymmetry within this secondary introduction, compared to the primary North American invasive population and to native birds sampled from the UK, which is hypothesized to result from destabilized developmental processes due to reduced genetic variation (Fiorini *et al.* 2021).

4. WHAT EXPLAINS INVASION SUCCESS IN THE STARLING?

Invasion theory recognizes that an invasive species' successful establishment and spread depends on a dynamic orchestration of ecological and evolutionary factors. Components of invasion success include but are not limited to: climate and environmental suitability, ecological interactions, social interactions, personality, demography, dispersal patterns and genetic factors such as pre-adaptation or invasion potential. Distinguishing among contributors to invasion success in wild systems is a technical challenge, but comparing recent and replicated invasions of the same species—in this case, the starling—may help identify which factors best explain invasion success in one wild bird. In the following sections, we place the burgeoning genomic studies of starlings in the context of modern invasion theory, to both highlight the utility of such genomic approaches, and to identify hypotheses yet to be tested in this species.
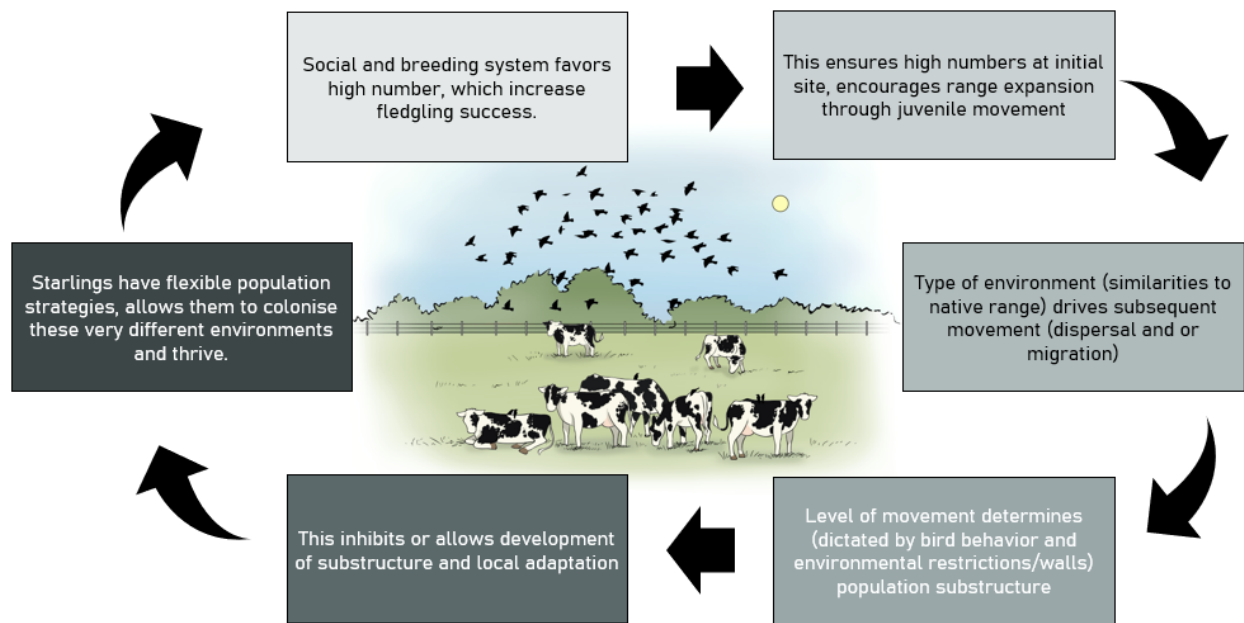
Figure 1.2. A proposed model of starling invasion success. This figure suggests how starlings might establish and spread in a variety of environmental conditions.

4.1 Dispersal and migration may evolve to support range expansion

Starlings may be resident (remaining in the same area year-round) or migratory (seasonal visitation to a location), with birds migrating up to 1,000-1,500 km (Linz *et al.* 2007). In general, starlings are migratory in the North and Eastern portions of their European range, and partially migratory and resident in the South and Western regions (due to warmer temperatures (Higgins *et al.* 2006). Within the North America range, rates of migration vary from 3-100% among populations (Kessel 1953; Blem 1981). Migratory behavior is frequently reported in North-East United States populations, though residency during colder winter months is enabled by urban landscape elements (Kessel 1953; Dolbeer 1982; Higgins *et al.* 2006). Within the other invasive populations including Australia and New Zealand there is no evidence of migration (Waterman *et al.* 2008); however, unconfirmed reports suggest migration may be a rare but nevertheless present phenomenon. How migratory behavior might support adaptation in other passerine birds is an active and fruitful area of research to extend into starlings (Chapman *et al.* 2011; Winger *et al.* 2019; Delmore *et al.* 2020).

In contrast to annual or partial migration, all starling populations experience variable dispersal strategies that impact range expansion. In every population, younger, juvenile birds or immature adults will form larger flocks in the non-breeding season, presumably as a means of additional protection during these more vulnerable periods of the bird's life cycle (Higgins *et al.* 2006; Figure 1.3). This stage is essential in the species' expansion: long-range dispersal is common when starlings are juveniles, before they have mated (Cabe 1999), which is common in many avian species (Paradis *et al.* 1998). Although dispersal itself is common across populations, the distance dispersed as

24

well as the timing of dispersal varies according to particular environmental conditions, whether population density or climatic conditions. Specifically, within the South African introduction, increased dispersal in this species is associated with population declines and relatively low abundance (Hui *et al.* 2012). Lower spread rates have been reported in areas with higher winter precipitation, indicating that unfavorable (low rainfall) conditions may trigger greater dispersal (Berthouly-Salazar *et al.* 2013). The same association has not explicitly been tested within other invasive ranges. However, isotopic evidence in North America suggests region-specific movement that may be related to population density, abiotic conditions, or any other number of factors (Werner *et al.* 2020). In Australia, long-distance dispersal events report heavy female bias (Rollins *et al.* 2009), which may suggest that juvenile female dispersal is enhanced by a drive to find a suitable nesting space, though this has not been tested. Considering this evidence, the starling's dispersal strategies differs dramatically between invasive ranges and the native one, indicating a flexible response of the species to spatial and temporal environmental variations (Hui *et al.* 2012).

4.2 The starling's breeding biology may facilitate expansion

Starlings depend on existing cavities for a space to nest. Because of this, starlings are resourceful when selecting nest sites, and regularly nest in manmade structures (Mainwaring 2015), or in cavities excavated by other birds or animals (Palacio *et al.* 2016). Beyond the benefits provided by cavity-nesting, starlings, like many bird species, breed synchronously during spring, informed by a number of abiotic and biotic cues (Figure 1.2). Breeding phenology shifts with short-term environmental changes such as

photorefractoriness (Dawson *et al.* 2001), and longer-term climatic shifts (McDermott & DeGroote 2016). Starlings also rely on social cues to facilitate breeding synchrony. When living in denser populations, starlings showed increased breeding synchrony (Evans *et al.* 2009). Higher population density is associated with an increase in reproduction-associated competition (for mates, nest sites, and/or prey) but also greater risks (increased predation), which presumably would encourage a decrease in breeding synchrony (Evans *et al.* 2009). It is likely then that the mass breeding achieved through synchrony provides group benefits such as collective predator awareness and defense in both parents and fledglings (Smith 2004). Hence the starling's social system actually facilitates the species' success, and if breeding success is positively related to high group density, then any strategy to increase local density (unseating other species, larger nests, use of anything natural or unnatural that may serve as a nest, etc.) all create a positive feedback loop, encouraging greater population expansion. Dramatic reproductive rates lead to increased density in the population, and this positive density dependence often triggers dispersal. Starlings may also accelerate population expansion by laying more eggs: averaged across the North American population, starlings in fact lay larger clutches than the average clutch size in the native population (Dawson 1983; Ball & Wingfield 1987). Finally, starlings are known for displaying a wide range of personality types across the species (Eens *et al.* 1993; Garamszegi *et al.* 2008; Thys *et al.* 2017), and starling parents have evolved many strategies (e.g. monogamous, polygamous, Intraspecific brood parasitism) for optimizing their effort in caring for young (Higgins *et al.* 2006). Breeding strategies may shift across native and invasive starling populations, which may support establishment and subsequent populaton expansion.

26

4.3 The starling's behavior and ecological interactions may also support invasion success

Variable ecological interactions may also affect invasion success and population characteristics. For instance bird invasion success is positively affected by the presence of other invasive species (Redding *et al.* 2019), a common occurrence in urban areas where starlings thrive. In areas where human urbanization is minimal, starlings will rely on natural nesting spaces, either preexisting or excavated by another species. Starlings may expand easily where equivalent niche or cavity-nesting species already reside, because there are already nesting sites available in these areas. Nest site availability is one of many limiting factors in starling survival: like any other species, starlings must find food and escape predators, and how exactly starlings adjust to new ecological contexts varies.

Starlings have a relatively big brain size compared to other birds of a similar size, which may plausibly play a role in their invasion success (Sol *et al.* 2002). Starlings' cognition may facilitate greater behavioral flexibility and innovation, of particular importance during initial invasive population establishment. This cognitive ability also enables great dietary flexibility, which impacts the species' persistence during invasion, and during times of stress such as food shortage (Van Berkel *et al.* 2018; Bateson *et al.* 2021). Such behavior may be heritable within families and developmentally modulated (Nettle *et al.* 2015), but regardless the starling's ability to learn and cope with changing conditions likely supports its invasion success.
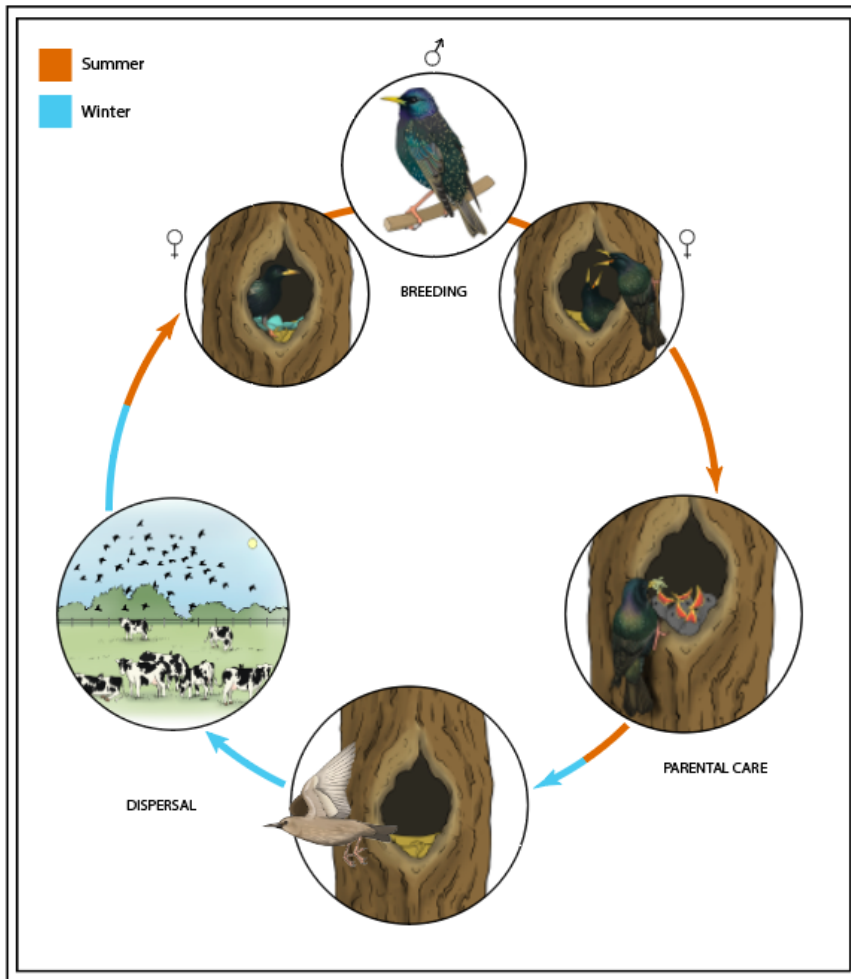
Figure 1.3. Starling life cycle, illustrating points in a starling's lifetime where variation in parental care strategy (for example) might impact invasion success.

5. RAPID ADAPTIVE EVOLUTION MAY FACILITATE EXPANSION IN NOVEL ENVIRONMENTAL CONDITIONS

5.1 Rapid expansion and evolution despite reduced genetic diversity

As in any invasion, a starling population's success may be constrained by its effective founding population size. A hallmark of invasive species is the rapid population

growth and expansion a species experiences upon "colonizing" a new environment. Nearly all invasions begin in a low-density population, and their subsequent expansion increases fitness of that population; by definition, these invasive populations are subject to an Allee effect (Allee 1931). The South African invasion is an outlier even among starling invasions: the founding population was only 18 individuals (Craig 2020, Table 1.1), which may explain the slow expansion speed in South Africa. Theory predicts that populations with smaller starting effective population sizes (N_e) should adapt more slowly than a population with higher N_e: effective population size impacts standing genetic variation available to selection, and adaptive variants are more likely to persist in larger populations (Baker & Stebbins 1965). However, even if the founding population is subject to a genetic bottleneck, rapid expansion can counteract diversity loss (Birzu *et al.* 2019); as the population expands, mutation may generate potentially adaptive variants (Gilbert *et al.* 2017; Gilbert & Whitlock 2017). Alternatively, expansion in South Africa may simply be slower because the population was not dense enough to trigger dispersal--a phenomenon we describe in the next section. In reality, population density and environmental conditions likely both shape range expansion in South Africa and other invasive populations.

Environmental conditions likely also shape expansion speed, perhaps even by promoting local adaptation. Even invasive species that have undergone severe bottlenecks are capable of rapid adaptive evolution in a novel environment (Dlugosch & Parker 2008; Facon *et al.* 2011; Rollins *et al.* 2013), perhaps via inbreeding x environment interactions (Schrieber & Lachmuth 2017). While genetic bottlenecking may explain why whole genome resequencing data reveals that the Australian

population remains very distinct from the North American and native UK populations (Hofmeister *et al.* 2021a), this would have been partially counteracted by the many and repetitious introductions to this range. Hence it is possible that, as the Australian climate is so dissimilar to that of the native range, rapid evolution may have been a prerequisite for establishment, and may be contributing to these differential genetic patterns. This hypothesis may also be supported by incidental evidence of establishment lag times, which are reported to be 4-5 years in North America, and considerably longer in Australia (Jenkins 1977; Woolnough *et al.* 2005; Higgins *et al.* 2006). In North America, historical records indicate that overcoming an elevational barrier may have slowed expansion speed (see Section 3.3; (Cooke 1928)). Given that genetic variation in the North American range is associated with elevation, it is possible that selection to cope with high-elevation conditions not experienced in the native range may also support North American expansion.

5.2 Interactions between adaptation and dispersal

Differences in genetic characteristics and substructure within each population are plausibly linked to differing dispersal characteristics. Environmental similarity to native range environments may explain part of the dispersal variation across populations, expanding our knowledge of native range environments and dispersal tendencies would further provide context to interpret these differences. Only within the eastern North American invasive range do we see migratory behaviour, facilitated possibly by environmental similarity and necessitated by the colder temperatures to the north. The presence of migration in this population undoubtedly has a massive impact on the

genetic substructure, encouraging greater genetic homogenisation. Contrasting this invasion to the older Australian and the marginally (7 years) younger South African one, the North American population covers an area many times larger than both. It is possible that migration enabled faster range expansion, and continues to enable genetic exchange across all but the more isolated areas of the range, diluting the effects of invasion gradient allele patterns (e.g. genetic drift and spatial sorting) and perhaps local adaption (Hofmeister *et al.* 2021b).

The Australian population shows spatial sorting, with wing length and loading strongly linked to distance from the introduction site, whereas this is not the case in the South African introduction (Phair *et al.* 2018). Why do some display spatial sorting and others not? Higher dispersal is associated with less desirable conditions, either due to the environment or high population density. Meanwhile, the native range shows a declining rate of spread with increased distance outside of Europe, which may be due to environmental suitability or population density (Hui *et al.* 2012). This 'good-stay, bad-disperse' hypothesis (Hui *et al.* 2012) may account for the introduction-sites to range-edge genetic gradients, and may encourage spatial sorting as seen in the Australian population (Phair *et al.* 2018). However, spatial sorting was not reported in South Africa, perhaps for two reasons: 1) founding size of the initial introduction was much smaller, providing less genetic variation to sort, and 2) the geographic range is much smaller, such that individuals disperse more readily from introduction site to range edge. Overall, this unidirectional dispersal movement may facilitate greater substructure in the population within these invasions compared to North America (though we cannot comment here on the native range due to the lack of genetic studies).

While many of these patterns of density-dependent growth and dispersal can at least in part explain the paradox of invasion success in starlings, determining the genetic basis for dispersal-related traits may clarify the eco-evolutionary feedback loops central to this invasion success. Empirical tests of dispersal evolution are underway in invasive systems like the cane toad (Perkins *et al.* 2013) and the ladybird beetle (Lombaert *et al.* 2014), and models that weight the contribution of both population densities and selection strength may yield insight into the relative importance of each factor (Lion 2018).

5.3 Persistence aided by environmental niche flexibility

The starling possesses great environmental flexibility as established above, with the characteristics of establishment and range expansion (e.g., introduction location, population density) varying dramatically among invasions. They are, however, restricted from northward expansion in the northern hemisphere due to colder temperatures, and expansions towards the equator are hampered in the southern hemisphere populations by heat and aridity extremes (e.g., inland Queensland, Australia). Recent analysis of global bird invasions indicates that climate suitability plays a major role in determining invasion success (Redding *et al.* 2019). Even within established and 'suitable' ecosystems, the nature of the environment holds great sway over starling population characteristics. Already we see local adaptation to environmental factors developing in the Australian population (Cardilini *et al.* 2016) and possibly even the North American population (Hofmeister *et al.* 2021b), though better understanding of these effects requires further insight into epigenetic variation

and phenotypic plasticity, and the role these play in facilitating adaptation and/or invasion success (Ghalambor *et al.* 2007; Gomez-Mestre & Jovani 2013; Murren *et al.* 2015).

Starlings in the native range have much lower thermal tolerance compared to the populations of North America (Dmi'el & Tel-Tzur 1985). Presumably, southern hemisphere starling populations are less restricted by cold environments, as they would be required to disperse a much farther distance to enter temperature limiting environments. Nevertheless, higher temperatures will likely restrict population expansion further towards warmer (more equatorial) regions (Feare & Craig 1999): this relationship is supported by the fact that starling populations become increasingly sparse approaching equatorial regions (Figure 1.1).

Niche opportunists make successful invaders, and the starling, as generalists, may successfully habituate in environments very different from those of their native range (Vall-llosera *et al.* 2016). Anthropogenic land alteration may counteract any restrictive effects of environment on range expansion. Humans may have assisted the starling's colonisation of cold extremes in the native range and North America, and arid areas of inland Australia, but in particular human land alteration likely impacts subsequent expansion. While starlings are successful in the urban environment, they prefer cleared agricultural and suburban areas to urban centers, and starlings have also been found to produce fewer young in more urbanised areas (Mennechez & Clergeau 2006). Starlings, however, do not require large habitats to settle, and are capable of colonising small remnant vegetation patches (Antos *et al.* 2006). Further population modeling and range estimates of this species should account for anthropogenic land

use, in particular land associated with agriculture, in any predictive models (Duncan *et al.* 2001; Baker & Bomford 2009). This is of particular importance as climate alone appears to not have any large scale macro-association with range distributions of many native European avian species (Beale *et al.* 2008). Examining further the link between anthropogenic land features and invasion success and expansion (e.g. Hill *et al.* 2005; Menon & Mohanraj 2016; Schmack *et al.* 2020) is an essential next step in understanding the interactions between this species and human populations.

Box 1.2. Starling as an excellent eco-evolutionary model system
The starling represents an excellent system in which to investigate questions around evolution and colonisation success. The multiple independent introductions of starlings bring with them a respectable amount of pre-existing literature characterising the patterns of genetic diversity (though to somewhat different degrees) across all the five invasive populations. There exists also multi tissue transcriptomic data available for the starling derived from both-short read and long-read RNA data (Richardson *et al.* 2017; Stuart & Edwards *et al.* 2021b), as well as two high quality genomes from two different contents (North America, and Australia, Stuart & Edwards et al. 2021), providing vital genomic references for future sequencing analysis. There is also much non-invasion related research into the starling, for example understanding their interactions with agriculture (Linz et al. 2017), patterns of migration and flocking (Piersma *et al.* 2020), and social behaviours, in particular its song (for example, Eens 1997) and extensive studies of hormone regulation of behavior (for example, Gwinner *et al.* 2002), which provide a wealth of background knowledge from which to interpret and contextualise

new results. Finally, the starling also features as one of the most abundant and globally collected avian skins in museums around the world, enabling temporal analysis to be conducted over a wide geographic area.

6. CONCLUSION

To summarize the future research directions this review has touched on, we present the below points as key knowledge gaps within starling population genetics research. First, future studies of starling demography and genetics within their native range are needed. We are currently witnessing range and demographic shifts in Europe, northern Africa, and Asia. How might climate change and anthropogenic land alteration shape these shifts, and what other factors might influence range shifts and changes in population size? Especially in light of these changes, conservation of native starling populations will require explicit studies of range-wide genetic diversity. To our knowledge, there is no record of starling population genetics and diversity within their native range, and this is a critical next step in starling population genetics.

Second, we must monitor of ongoing invasive starling range expansion. Documenting rapid evolution at expanding range edges may help to clarify whether and how populations adapt to local conditions. This work is critical in Australia, New Zealand, southern Africa and South America, where conservation managers actively work to control starling spread. As the newest major starling incursion, the Argentinean population will be a critical system to test the relative importance of intrinsic and extrinsic factors; specifically, we look forward to learning how genomic tools as well as careful tracking of the spatial expansion of this population clarify factors shaping

starling invasion success.

Next, comparative studies between starlings and other avian invaders might clarify mechanisms that support invasion success more broadly. There are a few other successful globally successful avian invaders, such as the common myna, the house sparrow, and the spotted dove. Cross-species investigation will allow more general examination of sources of genetic variation that predate rapid adaptation within invasive species, and elucidating shared properties of successful avian invasives.

Finally, starlings present a useful system for further empirical studies into the invasion paradox. The starlings' natural multi-continental introductions provide an ideal system to investigate the relative importance of introduction number on post-establishment genetic diversity, plasticity, and evolutionary capacity.

This review attempts a holistic survey of starling population genetics, bringing together literatures from both the native range and several invasions to identify drivers of starling invasion success. As an avian invader with a well-documented natural history across many continents, the starling provides an ideal testing ground for hypotheses about invasion success and post-introduction evolution.

Table 1.1: global starling introductions including introduction date, location, number of individuals, and other introduction relevant metadata. Fields with unknown values are indicated by a dash.

Country	Date of release	Introduction site	Number of individuals	Established?	Introduced by	Introduction source	Reference
North America	1850	Westchester, PA	-	No	-	-	Cooke 1928
	1872-3	Cincinnati, OH	-	No	-	-	Phillips 1928
	1875	Quebec, CA	-	No	-	-	Kalmbach and Gabrielson

							1921
	1889, 1892	Portland, OR	35 pairs	Yes, until 1900	Portland Songbird Club	-	
	1877	Central Park, NYC	-	No	Eugene Schieffelin	-	Phillips 1928
	1890	Central Park, NYC	80	Yes, common by 1895	Eugene Schieffelin	-	Phillips 1928, Cooke 1928
	1891	Central Park, NYC	80	Yes, common by 1895	Eugene Schieffelin	-	Phillips 1928, Cooke 1928
	1897	Springfield, MA	-	-	-	-	Phillips 1928
	1897	Bay Ridge, NY	-	-	-	-	Phillips 1928
Jamaica	1903	Jamaica	-	Yes	-	-	
Australia	1856	NSW	-	Unknown	Private introductions	England	Jenkins 1997, Long 1981
	1857	Melbourne	89	Yes, common by 1963	Private introductions	Britain	Long 1981
	1858	Melbourne	-	Presumably	Private introductions	Britain	Long 1981
	1860	Phillip island	6	Presumably	Presumed AS	-	Jenkins 1997
	1860s	SA	-	Probably	Presumed AS	-	Higgins et al 2005, Jenkins 1997
	1863	Melbourne	36	Presumably	Acclimatisation society	-	Jenkins 1997, Higgins et al 2005, Long 1981
	1864	Melbourne	6	Presumably	Acclimatisation society	-	Higgins et al 2005, Long 1981
	1865	Melbourne	120	Presumably	unknown	-	Jenkins 1997
	1866	Phillip island	6	Presumably	unknown	-	Jenkins 1997, Long 1981
	1866	Melbourne	15	Presumably	Acclimatisation society	-	Higgins et al 2005
	1869	Queensland	a 'batch'	Probably not - colonisation from range expansion most likely	Acclimatisation society	England	Jenkins 1997, Higgins et al 2005, Long 1981
	1871	Melbourne	20	Presumably	Presumed AS	-	Higgins et al 2005
	1880	New South Wales	2 small batches	Presumably	Presumed AS	Victoria or New Zealand	Higgins et al 2005, Jenkins 1997
	1880	Melbourne	-	Presumably	Presumed AS	nz	Higgins et al 2005
	1800/1860/1880 (1860 most	Tasmania	75	yes	D. L. Crowther	NZ	Higgins et al 2005, Long

	reported)						1981
	1881	South Australia	89	Presumably	Acclimatisation society	-	Higgins et al 2005, Long 1981
New Zealand	1862	Nelson	17	Presumably	The Nelson Society	Britain	Thompson 1922
	1967	Otago	3	Presumably	The Otago Society	Britain	Thompson 1922
	1868	Otago	81	Presumably	The Otago Society	Britain	Thompson 1922
	1869	Otago	85	Presumably	The Otago Society	Britain	Thompson 1922
	1867	Christchurch	20	Presumably	Canterbury Society	Britain	Thompson 1922
	1871	Christchurch	40	Presumably	Canterbury Society	Britain	Thompson 1922
	1865	Auckland	2	Presumably	Auckland Society	Britain	Thompson 1922
	1867	Auckland	15	Presumably	Auckland Society	Britain	Thompson 1922
	1868	Auckland	82	Presumably	Auckland Society	Britain	Thompson 1922
	1877	Wellington	60	Presumably	Wellington Society	Britain	Thompson 1922
	1878	Wellington	90	Presumably	Wellington Society	Britain	Thompson 1922
	1881	Wellington	14	Presumably	Wellington Society	Britain	Thompson 1922
	1882	Wellington	100	Presumably	Wellington Society	Britain	Thompson 1922
	1883	Wellington	34	Presumably	Wellington Society	Britain	Thompson 1922
South Africa	1897 (some dates list 1899)	Cape Town	18	Yes	Cecil Rhodes	Britain	Winterbottom and Liversidge 1954
South America	1949	Lago de Maracaibo, Venezuela	5	Unlikely	-	England	Long 1981
	pre-1987 (first spotted 1987)	Buenos Aires	-	Yes	-	North America	Perez 1988, Peris 2005
	pre-2001 (first spotted 2001)	Santa Fe	-	Yes	-	North America	Peris 2005

REFERENCES

Allee WC 1931 *Animal Aggregations: A Study in General Sociology.* Chicago: University of Chicago Press.

Anderson W 1992 Climates of opinion: acclimatization in nineteenth-century France and England. *Victorian Studies* **35** 135–157.

Antos MJ, Fitzsimons JA, Palmer GC & White JG 2006 Introduced birds in urban remnant vegetation: Does remnant size really matter? *Austral Ecology* **31** 254–261. (doi:10.1111/j.1442-9993.2006.01572.x)

Asher L & Bateson M 2008 Use and husbandry of captive European starlings (*Sturnus vulgaris*) in scientific research: a review of current practice. *Laboratory Animals* **42** 111–126. (doi:10.1258/la.2007.007006)

Baker HG & Stebbins GL 1965 *The Genetics of Colonizing Species: Proceedings of the First International Union of Biological Sciences Symposia on General Biology*. New York; London: Academic Press.

Baker J & Bomford M 2009 Opening the Climate Modelling Envelope. *Plant Protection Quarterly* **24** 88.

Ball GF & Wingfield JC 1987 Changes in Plasma Levels of Luteinizing Hormone and Sex Steroid Hormones in Relation to Multiple-Broodedness and Nest-Site Density in Male Starlings. *Physiological Zoology* **60** 191–199. (doi:10.1086/physzool.60.2.30158643)

Bateson M, Andrews C, Dunn J, Egger CBCM, Gray F, Mchugh M & Nettle D 2021 Food insecurity increases energetic efficiency, not food consumption: an exploratory study in European starlings. *PeerJ* **9** e11541. (doi:10.7717/peerj.11541)

Beale CM, Lennon JJ & Gimona A 2008 Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences* **105** 14908–14912. (doi:10.1073/pnas.0803506105)

Berthouly-Salazar C, Hui C, Blackburn TM, Gaboriaud C, Rensburg BJ van, Vuuren BJ van & Roux JJL 2013 Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Molecular Ecology* **22** 5793–5804. (doi:10.1111/mec.12538)

Birzu G, Matin S, Hallatschek O & Korolev KS 2019 Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecology Letters* **22** 1817–1827. (doi:10.1111/ele.13364)

Bitton P-P & Graham BA 2015 Change in wing morphology of the European starling during and after colonization of North America. *Journal of Zoology* **295** 254–260. (doi:10.1111/jzo.12200)

Blem CR 1981 Geographic Variation in Mid-Winter Body Composition of Starlings. *The Condor* **83** 370–376. (doi:10.2307/1367508)

Bodt LH, Rollins LA & Zichello J 2020 Genetic Diversity of the European Starling (*Sturnus vulgaris*) Compared Across Three Invasive Ranges.

Brewer D 2010 *Wrens, Dippers and Thrashers*. Bloomsbury Publishing.

Brush T 2009 Range Expansions and New Breeding Records of Birds in Tamaulipas, Mexico. *The Southwestern Naturalist* **54** 91–96. (doi:10.1894/MH-24.1)

Cabe PR 1998 The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity* **80** 519–525. (doi:10.1046/j.1365-2540.1998.00296.x)

Cabe PR 1999 Dispersal and Population Structure in the European Starling. *The Condor* **101** 451–454. (doi:10.2307/1370014)

Campbell S, Roberts EJ, Craemer R, Pacioni C, Rollins L & Woolnough AP 2016 Assessing the economic benefits of starling detection and control to Western Australia. *Australasian Journal of Environmental Management* **23** 81–99. (doi:10.1080/14486563.2015.1028486)

Cardilini APA, Buchanan KL, Sherman CDH, Cassey P & Symonds MRE 2016 Tests of ecogeographical relationships in a non-native species: what rules avian morphology? *Oecologia* **181** 783–793. (doi:10.1007/s00442-016-3590-9)

Chamberlain DE, Fuller RJ, Bunce RGH, Duckworth JC & Shrubb M 2000 Changes in the Abundance of Farmland Birds in Relation to the Timing of Agricultural Intensification in England and Wales. *Journal of Applied Ecology* **37** 771–788.

Chapman BB, Brönmark C, Nilsson J-Å & Hansson L-A 2011 The ecology and evolution of partial migration. *Oikos* **120** 1764–1775. (doi:10.1111/j.1600-0706.2011.20131.x)

Colautti RI & Barrett SCH 2013 Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342** 364–366. (doi:10.1126/science.1242121)

Colautti RI, Alexander JM, Dlugosch KM, Keller SR & Sultan SE 2017 Invasions and extinctions through the looking glass of evolutionary ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372** 20160031. (doi:10.1098/rstb.2016.0031)

Coleman JD 1972 The feeding ecology, productivity and management of Starlings in Canterbury, New Zealand.

Cooke MT 1928 *The Spread of the European Starling in North America (to 1928)*. Washington, D.C.: U.S. Dept. of Agriculture.

Cooper J & Underhill LG 1991 Breeding, Mass and Primary Moult of European Starlings *Sturnus Vulgaris* at Dassen Island, South Africa. *Ostrich* **62** 1–7. (doi:10.1080/00306525.1991.9639629)

Craig AJFK 2020 Common Starling (*Sturnus vulgaris Linnaeus,* 1758). In *Downs, C.T. and Hart, L.A (Eds) Invasive Birds: Global Trends and Impacts*, pp 9–24. Wallingford, UK: CAB International.

Cruz-Cardiel PJ de la, Deceuninck B, Peris SJ & Elena-Rosselló JA 1997 Allozyme polymorphism and interspecific relationships in the Common starling (*Sturnus vulgaris*) and Spotless starling (*S. unicolor*) (Aves: Sturnidae). *Journal of Zoological Systematics and Evolutionary Research* **35** 75–79. (doi:10.1111/j.1439-0469.1997.tb00406.x)

Dawson A 1983 Plasma gonadal steroid levels in wild starlings (*Sturnus vulgaris*) during the annual cycle and in relation to the stages of breeding. *General and Comparative Endocrinology* **49** 286–294. (doi:10.1016/0016-6480(83)90146-6)

Dawson A, King VM, Bentley GE & Ball GF 2001 Photoperiodic Control of Seasonality in Birds: *Journal of Biological Rhythms*. (doi:10.1177/074873001129002079)

Delmore KE, Van Doren BM, Conway GJ, Curk T, Garrido-Garduño T, Germain RR, Hasselmann T, Hiemer D, van der Jeugd HP, Justen H *et al.* 2020 Individual variability and versatility in an eco-evolutionary model of avian migration. *Proceedings of the Royal Society B: Biological Sciences* **287** 20201339. (doi:10.1098/rspb.2020.1339)

Dickerson LM 1938 The Western Frontier of the European Starling in the United States as of February, 1937. *The Condor* **40** 118–123. (doi:10.2307/1363826)

Dlugosch KM & Parker IM 2008 Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17** 431–449. (doi:10.1111/j.1365-294X.2007.03538.x)

Dlugosch KM, Anderson SR, Braasch J, Cang FA & Gillette HD 2015 The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology* **24** 2095–2111. (doi:10.1111/mec.13183)

Dmi'el R & Tel-Tzur D 1985 Heat balance of two starling species (*Sturnus vulgaris* and

Onychognathus tristrami) from temperate and desert habitats. *Journal of Comparative Physiology B* **155** 395–402. (doi:10.1007/BF00687484)

Dolbeer R 1982 Migration Patterns for Age and Sex Classes of Blackbirds and Starlings. *Journal of Field Ornithology* **53**.

Duncan RP, Bomford M, Forsyth DM & Conibear L 2001 High predictability in introduction outcomes and the geographical range size of introduced Australian birds: a role for climate. *Journal of Animal Ecology* **70** 621–632. (doi:10.1046/j.1365-2656.2001.00517.x)

Duncan RP, Blackburn TM & Sol D 2003 The Ecology of Bird Introductions. *Annual Review of Ecology, Evolution, and Systematics* **34** 71–98.

eBird 2021 *EBird: An Online Database of Bird Distribution and Abundance [Web Application]*. eBird, Cornell Lab of Ornithology, Ithaca, New York.

Eens M, Pinxten R & Verheyen RF 1993 Function of the Song and Song Repertoire in the European Starling (*Sturnus vulgaris*): An Aviary Experiment. *Behaviour* **125** 51–66.

Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M & Facon B 2016 Is There a Genetic Paradox of Biological Invasion? *Annual Review of Ecology, Evolution, and Systematics* **47** 51–72. (doi:10.1146/annurev-ecolsys-121415-032116)

Evans P 1980 Population Genetics of the European Starling (*Sturnus vulgaris*). University of Oxford.

Evans LE, Ardia DR & Flux JEC 2009 Breeding synchrony through social stimulation in a spatially segregated population of European starlings. *Animal Behaviour* **78** 671–675. (doi:10.1016/j.anbehav.2009.05.031)

Facon B, Hufbauer RA, Tayeh A, Loiseau A, Lombaert E, Vitalis R, Guillemaud T, Lundgren JG & Estoup A 2011 Inbreeding depression is purged in the invasive insect Harmonia axyridis. *Current Biology: CB* **21** 424–427. (doi:10.1016/j.cub.2011.01.068)

Feare CJ 1984 *The Starling*. Shire.

Feare C & Craig A 1999 *Starlings and Mynas*. Princeton University Press.

Fiorini VD, Domínguez M & Swaddle J 2021 Recent invasive population of the European starling *Sturnus vulgaris* has lower genetic diversity and higher fluctuating asymmetry than primary invasive and native populations. *Preprint*. (doi:10.21203/rs.3.rs-310175/v1)

Flux JEC & Flux MM 1981 Population dynamics and age structure of starlings (*Sturnus vulgaris*) in New Zealand. *New Zealand Journal of Ecology* **4** 65–72.

Forbush EH 1915 *The Starling*. Wright & Potter Printing Company.

Fortelius M, Eronen * Jussi, Jernvall J, Liu L, Pushkina D, Rinne J, Tesakov A, Vislobokova I, Zhang Z & Zhou L 2002 Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. *Evolutionary Ecology Research* **4** 1005–1016.

Freeman SN, Robinson RA, Clark JA, Griffin BM & Adams SY 2007 Changing demography and population decline in the Common Starling *Sturnus vulgaris:* a multisite approach to Integrated Population Monitoring. *Ibis* **149** 587–596. (doi:10.1111/j.1474-919X.2007.00684.x)

Fristoe TS, Chytrý M, Dawson W, Essl F, Heleno R, Kreft H, Maurel N, Pergl J, Pyšek P, Seebens H *et al.* 2021 Dimensions of invasiveness: Links between local abundance, geographic range size, and habitat breadth in Europe's alien and native floras. *Proceedings of the National Academy of Sciences of the United States of America* **118** e2021173118. (doi:10.1073/pnas.2021173118)

Garamszegi LZ, Eens M & Török J 2008 Birds Reveal their Personality when Singing. *PLoS ONE* **3**. (doi:10.1371/journal.pone.0002647)

Ghalambor CK, McKAY JK, Carroll SP & Reznick DN 2007 Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21** 394–407. (doi:10.1111/j.1365-2435.2007.01283.x)

Gilbert KJ & Whitlock MC 2017 The genetics of adaptation to discrete heterogeneous environments: frequent mutation or large-effect alleles can allow range expansion. *Journal of Evolutionary Biology* **30** 591–602. (doi:10.1111/jeb.13029)

Gilbert KJ, Sharp NP, Angert AL, Conte GL, Draghi JA, Guillaume F, Hargreaves AL, Matthey-Doret R & Whitlock MC 2017 Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. *The American Naturalist* **189** 368–380. (doi:10.1086/690673)

Gomez-Mestre I & Jovani R 2013 A heuristic model on the role of plasticity in adaptive evolution: plasticity increases adaptation, population viability and genetic variation. *Proceedings of the Royal Society B: Biological Sciences* **280** 20131869. (doi:10.1098/rspb.2013.1869)

Gregory RD, Wilkinson NI, Clark D, Robinson JA, Brown AF, Hughes JDAP, Procter DA, Gibbons DW & Galbraith CA 2002 The population status of birds in the United

Kingdom, Channel Islands and Isle of Man: An analysis of conservation concern 2002-2007. *British Birds* **95** 410–448.

Hanson HE, Mathews NS, Hauber ME & Martin LB 2020 The house sparrow in the service of basic and applied biology. *ELife* **9** e52803. (doi:10.7554/eLife.52803)

Harrison J & Cherry M *The Atlas of Southern African Birds*. Johannesburg: BirdLife South Africa.

Heldbjerg H, Fox AD, 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. (doi:10.1016/j.agee.2016.05.025)

Heldbjerg H, Fox AD, Lehikoinen A, Sunde P, Aunins A, Balmer DE, Calvi G, Chodkiewicz T, Chylarecki P, Escandell V *et al.* 2019 Contrasting population trends of Common Starlings (*Sturnus vulgaris*) across Europe. *Ornis Fennica* **96** 153–169.

Higgins PJ, Peter JM & Cowling SJ 2006 Handbook of Australian, New Zealand & Antarctic birds. Volume 7, Boatbill to starlings. Melbourne : Oxford University Press.

Hill SJ, Tung PJ & Leishman MR 2005 Relationships between anthropogenic disturbance, soil properties and plant invasion in endangered Cumberland Plain Woodland, Australia. *Austral Ecology* **30** 775–788. (doi:10.1111/j.1442-9993.2005.01518.x)

Hoffman EC 1930 The spread of the European starling in America. *Wilson Bull* **42** 80.

Hofmeister NR, Stuart K, Warren WC, Werner SJ, Bateson M, Ball GF, Buchanan KL, Burt DW, Cardilini APA, Cassey P *et al.* 2021a Concurrent invasions by European starlings (*Sturnus vulgaris*) suggest selection on shared genomic regions even after genetic bottlenecks. *BioRxiv* 2021.05.19.442026. (doi:10.1101/2021.05.19.442026)

Hofmeister NR, Werner SJ & Lovette IJ 2021b Environmental correlates of genetic variation in the invasive European starling in North America. *Molecular Ecology* **30** 1251–1263. (doi:10.1111/mec.15806)

Hui C & Richardson DM 2017 *Invasion Dynamics*. Oxford, New York: Oxford University Press.

Hui C, Roura-Pascual N, Brotons L, Robinson RA & Evans KL 2012 Flexible dispersal strategies in native and non-native ranges: environmental quality and the 'good–stay, bad–disperse' rule. *Ecography* **35** 1024–1032. (doi:10.1111/j.1600-0587.2012.07697.x)

Jenkins CFH 1977 *The Noah's Ark Syndrome: One Hundred Years of Acclimatization and Zoo Development in Australia*. Perth: Zoological Gardens Board of Western Australia.

Kalmbach ER, Gabrielson IN, United States, & Department of Agriculture 1921 *Economic Value of the Starling in the United States*. Washington, D.C.: U.S. Dept. of Agriculture.

Kessel B 1953 Distribution and Migration of the European Starling in North America. *The Condor* **55** 49–67. (doi:10.2307/1365026)

Lever C 2010 *Naturalised Birds of the World*. London: A&C Black.

Linz GM, Homan HJ, Gaulker SM, Penry LB & Bleier WJ 2007 European starlings: A review of an invasive species with far-reaching impacts. *Managing Vertebrate Invasive Species* **24**.

Linz G, Johnson R & Thiele J 2017 European Starlings. In *Ecology and Management of Terrestrial Vertebrate Invasive Species in the United States*, 1st ed, pp 311–332. Ed WC Pitt. Boca Raton : Taylor & Francis, 2018. | "A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.": CRC Press. (doi:10.1201/9781315157078-15)

Lion S 2018 Theoretical Approaches in Evolutionary Ecology: Environmental Feedback as a Unifying Perspective. *The American Naturalist* **191** 21–44. (doi:10.1086/694865)

Liu C, Wolter C, Xian W & Jeschke JM 2020 Most invasive species largely conserve their climatic niche. *Proceedings of the National Academy of Sciences* **117** 23643–23651. (doi:10.1073/pnas.2004289117)

Lombaert E, Estoup A, Facon B, Joubard B, Grégoire J-C, Jannin A, Blin A & Guillemaud T 2014 Rapid increase in dispersal during range expansion in the invasive ladybird Harmonia axyridis. *Journal of Evolutionary Biology* **27** 508–517. (doi:10.1111/jeb.12316)

Long JL 1981 *Introduced Birds of the World: The Worldwide History, Distribution, and Influence of Birds Introduced to New Environments*. Terrey Hills, Sydney, NSW: Universe Books.

Lowe S, Browne M & Boudjelas S 2000 *100 of the World's Worst Invasive Alien Species. A Selection from the Global Invasive Species Database*. Auckland, New Zealand: Invasive Species Specialist Group.

Macleod R, Clark J & Cresswell W 2008 The starvation–predation risk trade-off, body mass and population status in the Common Starling *Sturnus vulgaris*. *Ibis* **150** 199–208. (doi:10.1111/j.1474-919X.2008.00820.x)

Mahmood T, Usman-Ul-Hassan SM, Nadeem MS & Kayani A 2013 Population and diet of migratory Common Starlings *Sturnus vulgaris* wintering in agricultural areas of Sialkot district, Pakistan. *Forktail*.

Mainwaring MC 2015 The use of man-made structures as nesting sites by birds: A review of the costs and benefits. *Journal for Nature Conservation* **25** 17–22. (doi:10.1016/j.jnc.2015.02.007)

McDermott ME & DeGroote LW 2016 Long-term climate impacts on breeding bird phenology in Pennsylvania, USA. *Global Change Biology* **22** 3304–3319. (doi:10.1111/gcb.13363)

Mennechez G & Clergeau P 2006 Effect of urbanisation on habitat generalists: starlings not so flexible? *Acta Oecologica* **30** 182.

Menon M & Mohanraj R 2016 Temporal and spatial assemblages of invasive birds occupying the urban landscape and its gradient in a southern city of India. *Journal of Asia-Pacific Biodiversity* **9** 74–84. (doi:10.1016/j.japb.2015.12.005)

Miller AD, Inamine H, Buckling A, Roxburgh SH & Shea K 2021 How disturbance history alters invasion success: biotic legacies and regime change. *Ecology Letters* **24** 687–697. (doi:10.1111/ele.13685)

Murren CJ, Auld JR, Callahan H, Ghalambor CK, Handelsman CA, Heskel MA, Kingsolver JG, Maclean HJ, Masel J, Maughan H *et al.* 2015 Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity* **115** 293–301. (doi:10.1038/hdy.2015.8)

Navas J 2014 Las aves exóticas introducidas y naturalizadas en la Argentina. *Revista Del Museo Argentino de Ciencias Naturales Nueva Serie* **4** 191-202–202.

Nettle D, Andrews CP, Monaghan P, Brilot BO, Bedford T, Gillespie R & Bateson M 2015 Developmental and familial predictors of adult cognitive traits in the European starling. *Animal Behaviour* **107** 239–248. (doi:10.1016/j.anbehav.2015.07.002)

Neves VC, Griffiths K, Savory FR, Furness RW & Mable BK 2009 Are European starlings breeding in the Azores archipelago genetically distinct from birds breeding in mainland Europe? *European Journal of Wildlife Research* **56** 95–100. (doi:10.1007/s10344-009-0316-x)

North HL, McGaughran A & Jiggins C 2021 Insights into invasive species from whole-genome resequencing. *Molecular Ecology* **n/a**. (doi:10.1111/mec.15999)

Palacio FX, Maragliano RE & Montalti D 2016 Functional role of the invasive European

Starling, *Sturnus vulgaris*, in Argentina. *Emu* **116** 387–393. (doi:10.1071/MU16021)

Paradis E, Baillie SR, Sutherland WJ & Gregory RD 1998 Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology* **67** 518–536. (doi:10.1046/j.1365-2656.1998.00215.x)

Pateff P & Stresemann E 1947 On the Systematic Position of the Starlings inhabiting Bulgaria and the Neighbouring Countries. *Ibis* **89** 494–507. (doi:10.1111/j.1474-919X.1947.tb04367.x)

Peris S, Soave G, Camperi A, Darrieu C & Aramburú RM 2005 Range expansion of the European starling *Sturnus vulgaris* in Argentina. *Ardeola* **52** 359–364.

Perkins TA, Phillips BL, Baskett ML & Hastings A 2013 Evolution of dispersal and life history interact to drive accelerating spread of an invasive species. *Ecology Letters* **16** 1079–1087. (doi:10.1111/ele.12136)

Phair DJ, Roux JJL, Berthouly-Salazar C, Visser V, Vuuren BJ van, Cardilini APA & Hui C 2018 Context-dependent spatial sorting of dispersal-related traits in the invasive starlings (*Sturnus vulgaris*) of South Africa and Australia. *BioRxiv* 342451. (doi:10.1101/342451)

Pipek P, Blackburn TM & Pyšek P 2019 The ins and outs of acclimatisation: Imports versus translocations of skylarks and starlings in 19th century New Zealand. *Biological Invasions* **21** 1395–1413.

Raymond B, McInnes J, Dambacher JM, Way S & Bergstrom DM 2011 Qualitative modelling of invasive species eradication on subantarctic Macquarie Island. *Journal of Applied Ecology* **48** 181–191.

Redding DW, Pigot AL, Dyer EE, Şekercioğlu ÇH, Kark S & Blackburn TM 2019 Location-level processes drive the establishment of alien bird populations worldwide. *Nature*. (doi:10.1038/s41586-019-1292-2)

Rensburg MJ van 2014 Reconstructing the range expansion of the European starling in Southern Africa using a hybrid method of niche modelling and individual based modelling. Stellenbosch University.

Richardson MF, Sherwin WB & Rollins LA 2017 De Novo Assembly of the Liver Transcriptome of the European Starling, *Sturnus vulgaris*. *Journal of Genomics* **5** 54–57. (doi:10.7150/jgen.19504)

Rintala J, Tiainen J & Pakkala T 2003 Population trends of the Finnish starling *Sturnus vulgaris*, 1952—1998, as inferred from annual ringing totals. *Annales Zoologici Fennici* **40** 365–385.

Robinson RA, Siriwardena GM & Crick HQP 2005 Status and population trends of Starling *Sturnus vulgaris* in Great Britain. *Bird Study* **52** 252–260. (doi:10.1080/00063650509461398)

Rollins LA, Woolnough AP, Wilton AN, Sinclair R & Sherwin WB 2009 Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology* **18** 1560–1573. (doi:10.1111/j.1365-294X.2009.04132.x)

Rollins LA, Woolnough AP, Sinclair R, Mooney NJ & Sherwin WB 2011 Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology* **20** 2307–2317. (doi:10.1111/j.1365-294X.2011.05101.x)

Rollins LA, Moles AT, Lam S, Buitenwerf R, Buswell JM, Brandenburger CR, Flores-Moreno H, Nielsen KB, Couchman E, Brown GS *et al.* 2013 High genetic diversity is not essential for successful introduction. *Ecology and Evolution* **3** 4501–4517. (doi:10.1002/ece3.824)

Rollins LA, Woolnough AP, Fanson BG, Cummins ML, Crowley TM, Wilton AN, Sinclair R, Butler A & Sherwin WB 2016 Selection on Mitochondrial Variants Occurs between and within Individuals in an Expanding Invasion. *Molecular Biology and Evolution* **33** 995–1007. (doi:10.1093/molbev/msv343)

Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, Stanton JC, Panjabi A, Helft L, Parr M *et al.* 2019 Decline of the North American avifauna. *Science* **366** 120–124. (doi:10.1126/science.aaw1313)

Ross HA 1983 Genetic differentiation of starling (*Sturnus vulgaris*: Aves) populations in New Zealand and Great Britain. *Journal of Zoology* **201** 351–362. (doi:10.1111/j.1469-7998.1983.tb04281.x)

Sandakova SL, Kuksina DK-O, Saaya AT-O, Matveeva OA, Seveley SS, Toushkin AA, Toushkina AF & Tarazanova IS 2018 The fauna and nature of birds stay of residential landscapes of northern part of Central Asia. *Eurasian Journal of Biosciences* **12** 105–112.

Schmack JM, Schleuning M, Ward DF & Beggs JR 2020 Biogeography and anthropogenic impact shape the success of invasive wasps on New Zealand's offshore islands. *Diversity and Distributions* **26** 441–452. (doi:10.1111/ddi.13021)

Schrieber K & Lachmuth S 2017 The Genetic Paradox of Invasions revisited: the potential role of inbreeding × environment interactions in invasion success. *Biological Reviews* **92** 939–952. (doi:10.1111/brv.12263)

Silva CNS, McFarlane SE, Hagen IJ, Rönnegård L, Billing AM, Kvalnes T, Kemppainen

P, Rønning B, Ringsby TH, Sæther B-E *et al.* 2017 Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity* **119** 197–205. (doi:10.1038/hdy.2017.29)

Smith HG 2004 Selection for synchronous breeding in the European starling. *Oikos* **105** 301–311. (doi:10.1111/j.0030-1299.2004.10543.x)

Smith HG, Ryegård A & Svensson S 2012 Is the large-scale decline of the starling related to local changes in demography? *Ecography* **35** 741–748. (doi:10.1111/j.1600-0587.2011.06310.x)

Sol D, Timmermans S & Lefebvre L 2002 Behavioural flexibility and invasion success in birds. *Animal Behaviour* **63** 495–502. (doi:10.1006/anbe.2001.1953)

Stuart KC, Cardilini APA, Cassey P, Richardson MF, Sherwin WB, Rollins LA & Sherman CDH 2021a Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology* **30** 1419–1434. (doi:10.1111/mec.15601)

Stuart KC, Edwards RJ, Cheng Y, Warren WC, Burt DW, Sherwin WB, Hofmeister NR, Werner SJ, Ball GF, Bateson M *et al.* 2021b Transcript- and annotation-guided genome assembly of the European starling. *BioRxiv* 2021.04.07.438753. (doi:10.1101/2021.04.07.438753)

Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D & Kelling S 2009 eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation* **142** 2282–2292. (doi:10.1016/j.biocon.2009.05.006)

Thomson GM 1922 *The Naturalisation of Animals & Plants in New Zealand,*. Cambridge [Eng.] : The University Press,. (doi:10.5962/bhl.title.28093)

Thys B, Eens M, Aerts S, Delory A, Iserbyt A & Pinxten R 2017 Exploration and sociability in a highly gregarious bird are repeatable across seasons and in the long term but are unrelated. *Animal Behaviour* **123** 339–348. (doi:10.1016/j.anbehav.2016.11.014)

Turbelin AJ, Malamud BD & Francis RA 2017 Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography* **26** 78–92. (doi:10.1111/geb.12517)

Vall-llosera M, Llimona F, de Cáceres M, Sales S & Sol D 2016 Competition, niche opportunities and the successful invasion of natural habitats. *Biological Invasions* **18** 3535–3546. (doi:10.1007/s10530-016-1246-7)

Van Berkel M, Bateson M, Nettle D & Dunn J 2018 Can starlings use a reliable cue of

future food deprivation to adaptively modify foraging and fat reserves? *Animal Behaviour* **142** 147–155. (doi:10.1016/j.anbehav.2018.06.015)

Versluijs M, Turnhout CAM van, Kleijn D & Jeugd HP van der 2016 Demographic Changes Underpinning the Population Decline of Starlings *Sturnus vulgaris* in The Netherlands. *Ardea* **104** 153–165. (doi:10.5253/arde.v104i2.a7)

Waterman M, Fuller C & Murray MD 2008 Studies of roosting common starlings *Sturnus vulgaris* in South Australia. *Corella* **32** 25–29.

Watling D & Talbot-Kelly C 1982 *Birds of Fiji, Tonga and Samoa*.

Webster MA 1975 Hong Kong's trade in wildlife. *Biological Conservation* **8** 203–211.

Werner SJ, Fischer JW & Hobson KA 2020 Multi-isotopic (δ2H, δ13C, δ15N) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. *PLOS ONE* **15** e0237137. (doi:10.1371/journal.pone.0237137)

Williams GR 1953 The Dispersal from New Zeal- and Australia of Some Introduced European Passerines. *Ibis* **95** 676–692. (doi:10.1111/j.1474-919X.1953.tb01895.x)

Winger BM, Auteri GG, Pegan TM & Weeks BC 2019 A long winter for the Red Queen: rethinking the evolution of seasonal migration. *Biological Reviews* **94** 737–752. (doi:10.1111/brv.12476)

Winterbottom JM & Liversidge R 1954 The European Starling in the South West Cape. *Ostrich* **25** 89–96. (doi:10.1080/00306525.1954.9633410)

Woolnough AP, Massam MC, Payne RL & Pickles GS 2005 Out on the border: keeping starlings out of Western Australia. *Out on the Border: Keeping Starlings out of Western Australia.* 183–189.

Woolnough AP, Lowe TJ & Rose K 2006 Can the Judas technique be applied to pest birds? *Wildlife Research* **33** 449–455. (doi:10.1071/WR06009)

Wretenberg J, Lindström Å, Svensson S, Thierfelder T & Pärt T 2006 Population trends of farmland birds in Sweden and England: similar trends but different patterns of agricultural intensification. *Journal of Applied Ecology* **43** 1110–1120. (doi:10.1111/j.1365-2664.2006.01216.x)

Wretenberg J, Lindström Å, Svensson S & Pärt T 2007 Linking agricultural policies to population trends of Swedish farmland birds in different agricultural regions. *Journal of Applied Ecology* **44** 933–941. (doi:10.1111/j.1365-2664.2007.01349.x)

Zuccon D, Pasquet E & Ericson PGP 2008 Phylogenetic relationships among Palearctic–

Oriental starlings and mynas (genera Sturnus and Acridotheres: Sturnidae). *Zoologica Scripta* **37** 469–481. (doi:10.1111/j.1463-6409.2008.00339.x)

Zufiaurre E, Abba A, Bilenca D & Codesido M 2016 Role of landscape elements on recent distributional expansion of European Starlings (*Sturnus vulgaris*) in agroecosystems of the Pampas, Argentina. *The Wilson Journal of Ornithology* **128** 306–313. (doi:10.1676/wils-128-02-306-313.1)

CHAPTER 2: ENVIRONMENTAL CORRELATES OF GENETIC VARIATION IN THE INVASIVE AND LARGELY PANMICTIC EUROPEAN STARLING IN NORTH AMERICA

ABSTRACT

Populations of invasive species that colonize and spread in novel environments may differentiate both through demographic processes and local selection. European starlings (*Sturnus vulgaris*) were introduced to New York in 1890 and subsequently spread throughout North America, becoming one of the most widespread and numerous bird species on the continent. Genome-wide comparisons across starling individuals and populations can identify demographic and/or selective factors that facilitated this rapid and successful expansion. We investigated patterns of genomic diversity and differentiation using reduced-representation genome sequencing (ddRADseq) of 17 winter-season sampling sites. Consistent with this species' high dispersal rate and rapid expansion history, we found low geographic differentiation and few F_{ST} outliers even at a continental scale. Despite starting from a founding population of ~180 individuals, North American starlings show only a moderate genetic bottleneck, and models suggest a dramatic increase in effective population size since introduction. In genotype-environment associations we found that ~200 single-nucleotide polymorphisms are correlated with temperature and/or precipitation against a background of negligible genome- and range-wide divergence. Given this

evidence, we suggest that local adaptation in North American starlings may have evolved rapidly even in this wide-ranging and evolutionarily young system. This survey of genomic signatures of expansion in North American starlings is the most comprehensive to date and complements ongoing studies of world-wide local adaptation in these highly dispersive and invasive birds.

INTRODUCTION

Studies of local adaptation have long bridged the interface between ecological and evolutionary questions by exploring how populations adapt to differing environmental conditions. Traditionally, high degrees of local adaptation were expected to be present only in fairly isolated populations—those free from the homogenizing effects of high gene flow—with a long history in those locations, providing the time thought to be necessary for local adaptation to evolve (Lenormand, 2002). We now know that local adaptation occurs frequently even in systems with high gene flow (Yeaman & Whitlock 2011; Tigano & Friesen, 2016) and often rapidly after colonization of a novel environment (Prentis *et al.*, 2008). We continue to find evidence for rapid local adaptation in systems as divergent as cane toads (Rollins *et al.*, 2015), sticklebacks (Lescak *et al.*, 2015), honeybees (Avalos *et al.*, 2017), steelhead trout (Willoughby *et al.*, 2018), deer mice (Pfeifer *et al.*, 2018), and many more. These studies show that many taxa can adapt rapidly to local conditions in response to the new selection regimes they encounter as they expand their range.

Invasive species that have recently expanded into new environments provide tractable opportunities to investigate local adaptation as it originates (Colautti & Lau,

2015). Invasions typically expand from the founding population(s) following a predictable spatial and temporal pattern (reviewed in Excoffier *et al.*, 2009). Many invasive species experience genetic bottlenecks as a result of an initial founder effect, but this loss of standing genetic diversity does not necessarily limit the success of invasive species (Schmid-Hempel *et al.*, 2007; Dlugosch & Parker, 2008; Facon *et al.*, 2011); invasion biologists describe this phenomenon as the paradox of invasion (Estoup *et al.*, 2016, Schrieber & Lachmuth, 2017). After successful colonization of a new habitat, invasive species often show a demographic boom facilitated by ecological release in the new environment. Ecological release is the concept that introduced species are often released from top-down limitations to population growth, such as predators or pathogens in their native range (Riccardi *et al.,* 2013). At the same time, these novel ecological conditions may select among standing genetic variation, where the presence of certain genetic variants in the native range accelerates adaptation upon introduction (Tsutsui *et al.,* 2000; Schlaepfer *et al.,* 2009; Hufbauer *et al.,* 2011). Invasions thus allow us to observe interactions between demography and the early processes of selection (Dlugosch *et al.,* 2015) as populations experience new environments.

Spatial sorting on the invasion front may allow dispersing individuals to maximize spatiotemporal fitness and drive incipient adaptation, as in other starling invasions (Phair *et al.*, 2017), common mynas (Berthouly-Salazar *et al.*, 2012), pumpkinseed fish (Ashenden *et al.*, 2017), and cane toads (Brown *et al.*, 2014). More empirical work is needed to verify the conditions in which spatial sorting can lead to lasting shifts in fitness (Lee, 2011; Phillips & Perkins, 2019; Williams *et al.*, 2019). However, we do know that adaptive dispersal strategies can facilitate range expansion

in Western bluebirds (Duckworth, 2008), invasive beetles (Lombaert *et al.,* 2014; Ochocki & Miller, 2017), and other species. Certain dispersal strategies can result in gene flow that counteracts inbreeding depression and increases adaptive potential (Garant *et al.,* 2007; Rius & Darling, 2014). If particular traits enable individuals to disperse more easily to their preferred habitat, gene flow may be directional and even adaptive (Edelaar & Bolnick, 2012; Jacob *et al.,* 2017). Dispersal among populations will certainly impact population structure of an invasion, but could even support local adaptation.

The European starling (*Sturnus vulgaris*) stands out as an exceptionally successful avian colonist and invasive species. In North America, an estimated 200 million starlings currently range from northern Mexico to southern Alaska (Linz *et al.,* 2007). Introduced to New York City in 1890, starlings nearly covered the continent within a few generations by expanding up to 91 km each year (Bitton & Graham, 2014). The 1890 introduction is widely accepted as the first successful establishment of starlings in North America, but several populations were introduced in Cincinnati, Ohio (1872), Quebec, Canada (1875), Allegheny, Pennsylvania (1897), and Springfield, Massachusetts (1897), with the second-most successful having been introduced to Portland, Oregon in 1889 (Forbush, 1915; Kalmbach & Gabrielson, 1921). We note here that multiple invasions from different source populations may lead to admixture among previously isolated populations and thus the introduction of new alleles to a population (Dlugosch & Parker, 2008). Records indicate that none of the earlier starling introductions survived more than a few years after colonization, but it is possible that some populations in the western U.S. persisted without record (Kessel, 1953). During the starling expansion, ongoing seasonal migration and dispersal might have also influenced patterns of

genetic variation. In North America, some—but not all—starling populations migrate (Dolbeer, 1982). Previous studies indicate that there is considerable variation in migratory distances within flyways (Burtt & Giltz, 1977), and early genetic work based on a small set of allozyme markers indicated near-random mating at a continental scale in North American starlings, with large demes (subpopulations) and high dispersal rates (Cabe, 1998; Cabe, 1999).

Here we use genome-wide markers to explore the population genetics and possible genotype-environmental associations in North American starlings with four specific aims: (1) to characterize genome-wide levels of diversity and differentiation among starlings; (2) to examine how genetic variation changes across the contiguous United States; (3) to test for a genetic bottleneck; and (4) to test for signatures of selection associated with environmental gradients. We test for these genotype-environment associations using overwintering sites, and interpret our results in the context of range-wide data on starling migration and dispersal (Werner, Fischer, & Hobson, 2020), as this movement certainly influences population structure. Models of molt origin—which rely on feathers sampled from the same individuals in our genetic survey—suggest that starlings that migrate in eastern North America likely experience greater gene flow among sampling locations (states, in our study) compared to those in western North America, but overall starlings tend to move only regionally and not continent-wide (Werner, Fischer, & Hobson, 2020). This work employs modern genomic and analytical tools to examine the evolutionary history of this remarkably successful avian invasion.

METHODS

Sample collection and processing

All starlings sampled were collected during the non-breeding season, when large flocks aggregate at high-quality food sources. These samples therefore do not represent discrete breeding populations but rather a mixture of birds from the surrounding region, potentially including migrants from more northerly areas. Starlings ($N = 166$) were collected in January-March 2016 and 2017 from 26 dairies and feedlots by the U.S. Department of Agriculture (USDA) Wildlife Services personnel in Arizona, California, Colorado, Idaho, Illinois, Iowa, Kansas, Missouri, Nebraska, Nevada, New Hampshire, New Mexico, New York, North Carolina, Texas, Washington, and Wisconsin (Figure 2.1). USDA Wildlife Services personnel collected birds from 2-5 sites in each state, where each collection site was >5 km apart. USDA personnel then euthanized whole adult males by lethal gunshot, avicide, or live traps, stored these birds at 0°C until tissue sampling, and recorded the latitude and longitude of each collection site using a handheld GPS. The collection and use of starlings for this and related studies were approved by the U.S. Department of Agriculture, National Wildlife Research Center's Institutional Animal Care and Use Committee (QA-2572, S.J. Werner - Study Director; Werner, Fischer, & Hobson, 2020; Table A1).
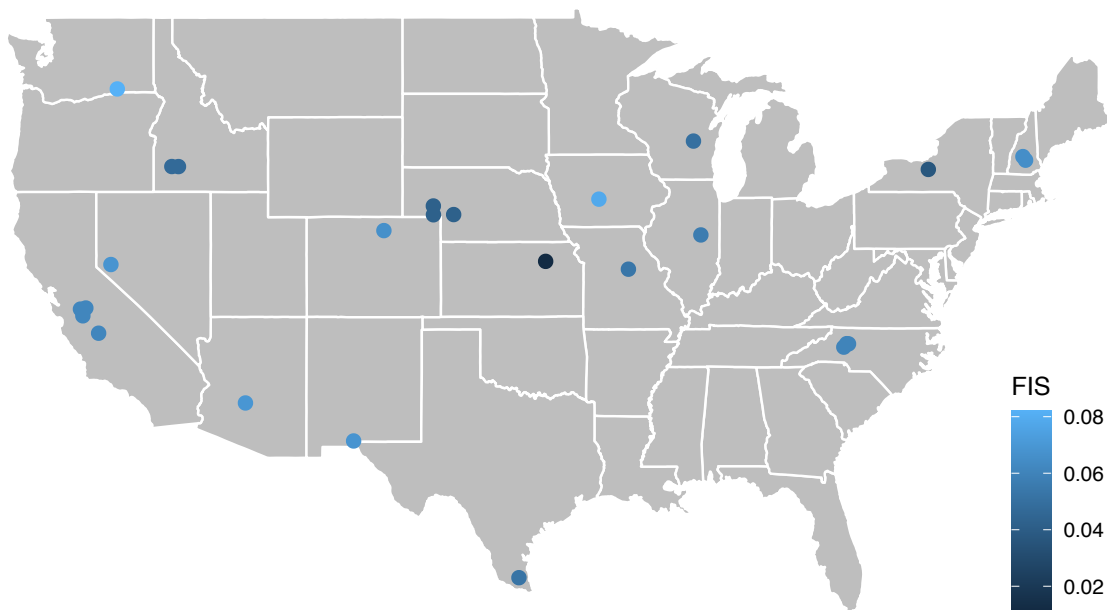
Figure 2.1. Sampling map of all locations, where color indicates inbreeding (F_{IS}) within each sampling location.

Breast muscle tissue was sampled using biopsy punches (Integra Miltex, York, PA) and frozen in 95% ethanol. Samples were shipped on dry ice, and DNA was extracted using a Qiagen DNeasy kit following the manufacturer's protocol (Qiagen, New York, NY). DNA concentration of each sample was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, New York, NY). Following the protocol of Peterson *et al.* (2012), we generated a reduced-representation genomic dataset of double-digested, restriction-site associated DNA markers as described in Thrasher *et al.* (2017) using the restriction enzymes SbfI and MspI and adaptors P1 and P2. We sequenced 100bp, single-end reads of the 160 best-quality libraries on an Illumina HiSeq 2500. We trimmed and filtered for quality using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). We then used the process_radtags commands in STACKS v 1.19 (Catchen *et al.,* 2013) to demultiplex the remaining sequences. In subsequent filtering steps, we retained reads only if the following conditions were met: reads passed the Illumina chastity filter, contained an intact *Sbf*I RAD site, contained one of the unique barcodes, and did not contain Illumina indexing adaptors.

Individual reads were mapped to an *S. vulgaris* reference genome (Hofmeister, Rollins *et al., in prep*) using BOWTIE2 version 2.2.8 (Langmead & Salzberg, 2012) using the "very sensitive local" set of alignment pre-sets, and then assembled sequences into 'stacks' using the ref-map option in STACKS. Compared to a reference-free approach, the bioinformatics pipeline used for the reference-based assembly has the advantage of using fewer similarity thresholds to build loci. We required that a SNP be present in a minimum of 80% of the individuals (-r 0.8) with a minimum stack depth of 10 reads at a

locus within an individual (-m 10) for it to be called. We removed two individuals, one with >50% missing data and one with >50% relatedness (measured using the unadjusted AJK statistic and calculated within *vcftools*), leaving 158 individuals remaining in the study. A total of 15,038 SNPs were identified. We used *vcftools* –hwe option to remove any SNPs out of Hardy-Weinberg equilibrium (e.g., an exact test that compared expected and observed heterozygosity in polymorphic sites only gave a P-value less than 0.001). About 6% of sequenced variants (904 variants) were out of HWE across all sampling sites; given that 1) we are particularly interested in SNPs that may be specific to certain populations, and 2) filtering for Hardy-Weinberg equilibrium did not change the results described in sections (1) and (2) below, we retain all 15,038 SNPs for the results presented here. For all genotype-environment analyses we used all SNPs in a given stack, but for STRUCTURE and other analyses sensitive to linkage disequilibrium we used only the first SNP in each stack; for unphased genomic data like the RAD loci analyzed here, this strategy of exporting one SNP per stack is often used as an indirect method of controlling for linkage disequilibrium.

(1) Patterns of genetic diversity and differentiation

We estimated per-locus measures of genetic diversity and genome-wide differentiation using the *populations* option in STACKS (Catchen *et al.*, 2013). We used *vcftools* to calculate F_{ST} among population pairs and heterozygosity and nucleotide diversity (π) within sampling sites ("populations") (Danecek *et al.*, 2011). We determined expected heterozygosity at the population level using the *Hs()* function in the R package *adegenet* (Jombart, 2008), and tested for differences in observed and expected heterozygosity of

each locus using a paired t-test in base R (R Core Team, 2019). We also used *adegenet* to calculate ϕ_{ST} between populations. We investigated genetic structure within North American starlings using an analysis of molecular variance (AMOVA) in the R package *poppr* (Kamvar *et al.*, 2014). We tested whether most genetic variation was observed among individuals or among sampling sites. To determine significance, we compared observed variation at each hierarchical level to the randomly permuted distance matrices for that particular level using 999 simulations in the function *randtest()* in the R package *adegenet* (Jombart, 2008), hypothesizing that the observed variance is greater than expected within individuals and less between individuals and between sampling sites.

   We tested for isolation by distance (IBD) using a simple Mantel test in *adegenet* (Jombart, 2008): for these data, the assumption of stationarity likely holds, given that North American starlings appear to be in mutation-migration-drift equilibrium (Guillot & Rousset, 2013). Geographic distances among sampling locations follow a bimodal distribution where locations are either very near or far from each other, and thus we caution that the Mantel test may not be an appropriate test of isolation by distance (Appendix A.2). Nevertheless, we report the results of a Mantel test with a Spearman correlation and 9999 permutations.

(2) Population structure

We first used non-parametric approaches to determine whether 6,287 SNPs clustered by sampling location, using principal components analysis in *SNPRelate* (Zheng *et al.*, 2012) and discriminant analysis of principal components (DAPC) in *adegenet* (Jombart, 2008).

We then tested for population structure using STRUCTURE (Pritchard *et al.*, 2000) by simulating 10 runs of each *K*. Although we sampled starlings from 17 locations, we hypothesized that North American starlings would cluster into at most eight populations (*K*=1-8) based on the non-parametric tests described above. To select the best-supported K, we used the Evanno method implemented in STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2011). We averaged results across the 10 runs using the greedy algorithm in the program CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007), and visualized results using DISTRUCT v1.1 (Rosenberg, 2003). Given that evidence of population structure depends on the filtering thresholds selected, we ran this model-based approach using a very strict minimum minor allele frequencies (MAF = 0.3, N SNPs = 3,568) and a more relaxed minimum frequency (MAF = 0.1, N SNPs = 6,287) (Linck & Battey, 2017). STRUCTURE results did not differ substantially among MAF thresholds, and we used a filtered dataset with a minimum minor allele frequency of 0.1 in the demographic analyses below.

(3) Demographic modeling

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

(TENDBOT: TBOT + 500). We performed 100 runs of this model and assessed model fit

using delta-likelihood for each run of the model; since all models use the same numbers

of parameters, the Akaike information criteria do not change between models. We also

ran a demographic model that takes linkage into account, the Stairway plot method, but

given that this method does not explicitly test for a genetic bottleneck, we describe the

method and results in the Appendix (Appendix A3).

(4) Selection analyses

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

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

each set of sampling coordinates from WorldClim 2.0 at a resolution of 5 min on June

16, 2018 (Fick & Hijmans, 2017). We test for isolation by environment (IBE) using the R

package *vegan* to compare environmental, geographic, and genetic distances (Oksanen

et al., 2019). Controlling for geographic distance, we ran partial Mantel correlograms

using Euclidean environmental distance matrices built for each bioclimatic variable. To

examine how loci may covary across multiple environmental gradients, we used

redundancy analysis (RDA) (Forester *et al.*, 2018). RDA is especially powerful when

testing for weak selection, and detects true positives in large data sets more reliably

than other multivariate methods like Random Forest (Forester *et al.*, 2018). Because RDA

requires no missing data, we first imputed genotypes where missing sites were

assigned the genotype of highest probability across all individuals—a conservative but

quick imputation method. Before imputation and after filtering, about 31% of genotypes

were missing from the dataset of 15,038 SNPs (not filtered for MAF), and we required

that all imputed SNPs were present in at least 80% of individuals. We then used the R package *vegan* to run the RDA model (Oksanen *et al.*, 2019); for a full description of this method, see Forester *et al.*, 2018. Briefly, RDA uses constrained ordination to model a set of predictor variables, and unconstrained ordination axes to model the response (genetic variation). RDA infers selection on a particular locus when it loads heavily onto one or more unconstrained predictor axes, and we retained five variables with relatively low variance inflation factors: BIO1 or Annual Mean Temperature (VIF=3.54), BIO7 or Temperature Annual Range (4.55), BIO12 or Annual Precipitation (8.69), BIO16 or Precipitation of Wettest Quarter (7.91), and elevation (2.19). Each of these variables appeared normally distributed, and correlations among predictors are all lower than 0.57, except for the correlation between BIO1 and BIO7 where R=-0.79. We tested for significance using the *anova.cca* function within the *vegan* package, and also permuted predictor values across individuals to further check significance of the model. We identify candidate loci as those that are 3 or more standard deviations outside the mean loading. The R scripts for all RDA analyses and figures were written by Brenna Forester (available at https://popgen.nescent.org/2018-03-27_RDA_GEA.html).

RESULTS

(1) Patterns of genetic diversity and differentiation

We identified 15,038 SNPs at a mean of 27X sequencing depth across 17 sampling locations. Using a dataset filtered by a minor allele frequency of 0.1, we find that genome-wide F_{ST} is extremely low (0.0085), and measures of genetic diversity do not vary substantially among sampling locations (Figure 2.1, Table 2.1). Across all of these

wintering starling sampling sites, the highest pairwise F_{ST} was 0.0106 between birds

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

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

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

explained by variation within individuals, and the remaining variance reflects

differences among individuals from the same sampling site, with negligible variation

explained at the between-population level (Figure 2.2C-D). Across the genome, F_{ST} and

nucleotide diversity are exceptionally low (Table 2.1, Table 2.2). Genome-wide

heterozygosity is moderate at 0.339, and across loci observed heterozygosity differs

significantly from expected (t = 66.6, df = 3569, P<0.001, Table 2.1).

Sampling location	N ind.	Private	H_{obs}	H_{exp}	π	F_{IS}
Arizona	10	1	0.32	0.24	0.35	0.069
California	11	0	0.33	0.25	0.35	0.061
Colorado	8	1	0.33	0.24	0.35	0.066
Idaho	8	0	0.33	0.22	0.35	0.047
Illinois	7	1	0.33	0.26	0.35	0.056
Iowa	9	0	0.33	0.24	0.36	0.078
Kansas	9	2	0.35	0.23	0.35	0.012
Missouri	9	0	0.33	0.23	0.35	0.053
Nebraska	11	0	0.34	0.22	0.35	0.042
Nevada	10	0	0.33	0.31	0.36	0.068
New Hampshire	11	1	0.33	0.26	0.35	0.065
New Mexico	7	2	0.33	0.29	0.36	0.068
New York	10	0	0.34	0.25	0.35	0.037
North Carolina	11	1	0.33	0.25	0.35	0.060
Texas	9	1	0.33	0.24	0.35	0.052
Wisconsin	8	0	0.33	0.22	0.35	0.050
Washington	10	2	0.32	0.30	0.35	0.082

Table 2.1. Population genetic summary statistics. For each sampling location, the table shows the number of individuals, number of private alleles, observed and expected heterozygosity, nucleotide diversity, and the inbreeding coefficient F_{IS}.

	AZ	CA	CO	IA	ID	IL	KS	MO	NC	NE	NH	NM	NV	NY	TX	WA	WI
Arizona	28	0.008 5	0.009 0	0.009 3	0.009 0	0.008 2	0.009 12	0.01 2	0.009 3	0.009 0	0.01 25	0.011 10	0.01 0	0.01 0	0.01 2	0.01 0	0.01 5
California	5	40	0.008 0	0.006 5	0.005 0	0.005 0	0.005 10	0.005 0	0.007 5	0.006 0	0.007 0	0.005 10	0.006 0	0.006 0	0.007 0	0.006 0	0.006 10
Colorado	0	7	43	0.006 0	0.008 7	0.008 2	0.007 0	0.007 0	0.006 0	0.006 11	0.007 7	0.01 4	0.007 2	0.006 0	0.007 0	0.007 2	0.009 0
Iowa	4	16	0	8	0.006 0	0.005 14	0.008 4	0.007 14	0.005 2	0.009 0	0.007 6	0.007 14	0.006 0	0.006 2	0.007 0	0.007 0	0.008 6
Idaho	0	0	4	0	23	0.007 0	0.007 0	0.006 0	0.007 0	0.005 4	0.007 0	0.008 0	0.004 8	0.006 0	0.007 0	0.006 19	0.007 0
Illinois	5	2	5	2	0	12	0.007 5	0.006 17	0.007 0	0.006 0	0.006 7	0.01 17	0.005 0	0.006 0	0.006 2	0.007 0	0.008 12
Kansas	2	12	2	0	0	5	7	0.005 12	0.006 29	0.005 2	0.006 5	0.006 2	0.005 0	0.006 2	0.007 2	0.006 0	0.006 10
Missouri	5	5	0	0	0	5	5	11	0.005 16	0.006 0	0.006 5	0.007 26	0.006 0	0.006 0	0.006 0	0.007 0	0.007 11
North Carolina	0	0	0	0	0	0	16	0	32	0.006 0	0.007 0	0.007 0	0.006 0	0.006 0	0.007 44	0.008 0	0.006 4
Nebraska	0	8	24	0	16	0	0	0	0	12	0.008 12	0.006 4	0.006 8	0.007 0	0.007 0	0.007 0	0.006 0
New Hampshire	28	6	0	3	0	0	3	3	0	6	9	0.007 6	0.007 3	0.006 0	0.007 0	0.007 0	0.007 13
New Mexico	11	0	0	0	0	0	0	0	11	11	0	67	0.006 0	0.008 0	0.007 0	0.007 0	0.007 0
Nevada	0	3	4	1	18	0	1	0	0	11	3	1	14	0.005 0	0.007 0	0.006 14	0.006 1
New York	3	10	0	1	0	15	1	10	0	1	6	1	1	1	0.007 0	0.006 0	0.005 21
Texas	5	0	0	0	0	0	0	3	26	0	0	2	0	0	65	0.007 0	0.008 0
Washington	0	0	0	0	0	0	0	0	0	7	0	0	60	0	0	7	0.006 0
Wisconsin	10	0	2	2	2	8	6	4	0	2	2	0	0	10	0	0	31

Table 2.2. Differentiation and movement among sampling locations. This table shows the percentage of birds assigned to each sampling location, where columns indicate state of origin. Note that these values are directional: although 12 starlings that were collected in AZ originated in KS, only 2 starlings collected in KS originated in AZ. The black diagonal indicates the percentage of birds assigned to that collection state according to a discriminant function analysis of molt-origin presented in Werner et al. For example, 28% of birds collected in Arizona originated in that location. F_{ST} among locations is presented above the diagonal, and darker gray colors indicate higher F_{ST}.
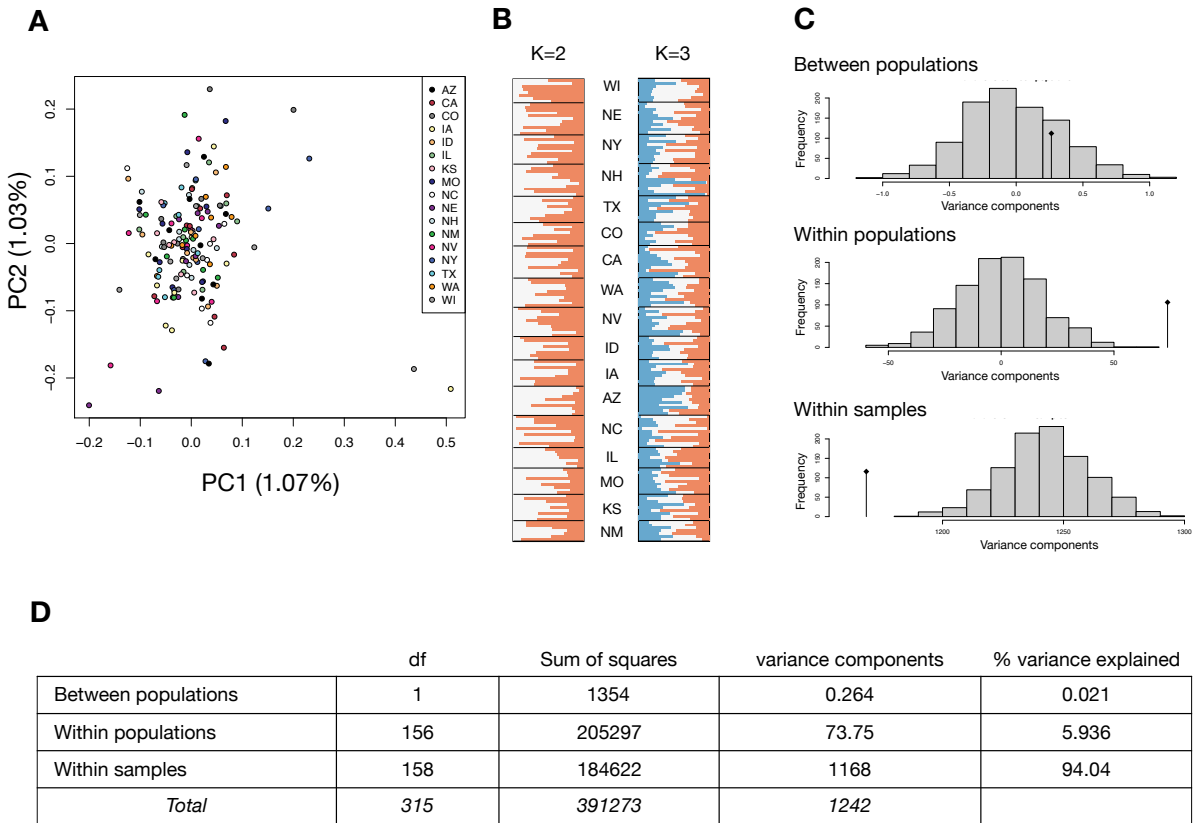
Figure 2.2. Population structure. A) Principal components analysis on 6,287 SNPs, with PC1 explaining 1.07% and PC2 explaining 1.03% of genetic variation observed. B) STRUCTURE analyses with K=2 and K=3 (best supported). C) Significance testing of hierarchical AMOVAs: the histogram shows expected variance components based on 999 simulations, and the black diamond is the observed variance component. D) AMOVA results.

	df	Sum of squares	variance components	% variance explained
Between populations	1	1354	0.264	0.021
Within populations	156	205297	73.75	5.936
Within samples	158	184622	1168	94
Total	*315*	*391273*	*1242*	

(2) Population structure

F_{ST} among sampling sites is relatively low overall (Table 2.2): starlings sampled in

Arizona show the highest differentiation from other sampling locations (F_{ST} = 0.008-

0.011), and only a few other pairwise comparisons (NM-IL, NM-CO, and CO-WI) show

an F_{ST} of 0.009 or higher. However, we do not recover clear population structure. The

first two principal components each explain about 1% of variation among individuals

(Figure 2.2A), and although STRUCTURE identified three populations at the best-

supported value of K, we do not observe obvious differences in ancestry proportions

among predicted "populations" (Figure 2.2B, Figure A1). Controlling for shared

ancestry does not resolve population structure, and instead provides support for

uniform gene flow among individuals (Figure A2). K-means clustering within DAPC

also does not identify biologically relevant clusters.

Starling "populations" (sampling sites) follow a clear pattern of isolation-by-

distance (Mantel r = 0.139, P < 0.0001, Figure A3). Spatially explicit models of isolation-

by-distance suggest a fairly uniform rate of migration range-wide, where local increases

in migration rate are likely a model artifact (Appendix A.2, Figure A4). However,

models of isolation by environment (IBE) show that the relationship between

environmental and genetic distances is stronger than the simple geographic-genetic

distance model. After controlling for geographic distance, we find that all bioclimatic

variables tested show non-zero correlations between environmental and genetic

distances (Appendix A.2, Table A1). There is a strong positive relationship between

genetic distance and both precipitation in the wettest quarter (BIO16, Mantel r = 0.282, P

= 0.001, Table A1) and elevation (Mantel r = 0.146, P = 0.001, Table A1).

(3) Demographic modeling

A traditional SFS-based model recovers a bottleneck in population size, as expected: the estimated effective population size during the bottleneck is ~16,000 individuals (1% of the ancestral population size). However, our demographic model suggests that the current population size of North American starlings is nearly identical to its pre-bottleneck size (Appendix A.3, Figure A5). In fact, every run of the model finds that starlings' current effective population size (1.6 million individuals) is considerably higher than the estimated N_e of the founding population. This model suggests that starlings experienced rapid population growth after the initial founder event, which may contribute to the overall lack of evidence for inbreeding. We do detect low levels of inbreeding within some populations (Table 2.1, highest F_{IS} = 0.082 in Washington). Taken together, these models do not suggest a classical founder effect, whereby effective population size remains low for many generations post-bottleneck.

(4) Selection analyses

Starlings encounter a range of precipitation, temperature, and elevation across their range and redundancy analyses revealed the strongest signatures of local adaptation compared to all selection testing methods (see Appendix A.5-A.7, Figures A6-8). The RDA model showed that 178 variants are correlated with environmental differences among sampling sites (F = 1.022, P = 0.002, Figure 2.3). Starlings living in warmer climates tend to cluster more closely in the left quadrant and high elevation individuals cluster in the middle right quadrant. However, starlings do not cluster based on

geographic distance: for example, starlings from TX and WA cluster closely due to shared genetic variants, even though the two sampling sites differ substantially in precipitation and temperature and are geographically separated. After controlling for population structure, candidates for selection are equally distributed among elevation, temperature- and precipitation-associated predictors. Importantly, when environmental predictors are randomly shuffled, the RDA model is not significant. Mean annual temperature (BIO1) opposes selective pressure related to the range of temperatures experienced each year (BIO7), annual precipitation (BIO12), precipitation in the wettest quarter (BIO16) and elevation (Figure 2.3). Genes under selection tend to have lower allele frequencies, and function in several biological processes (Appendix A.8, Table A2).
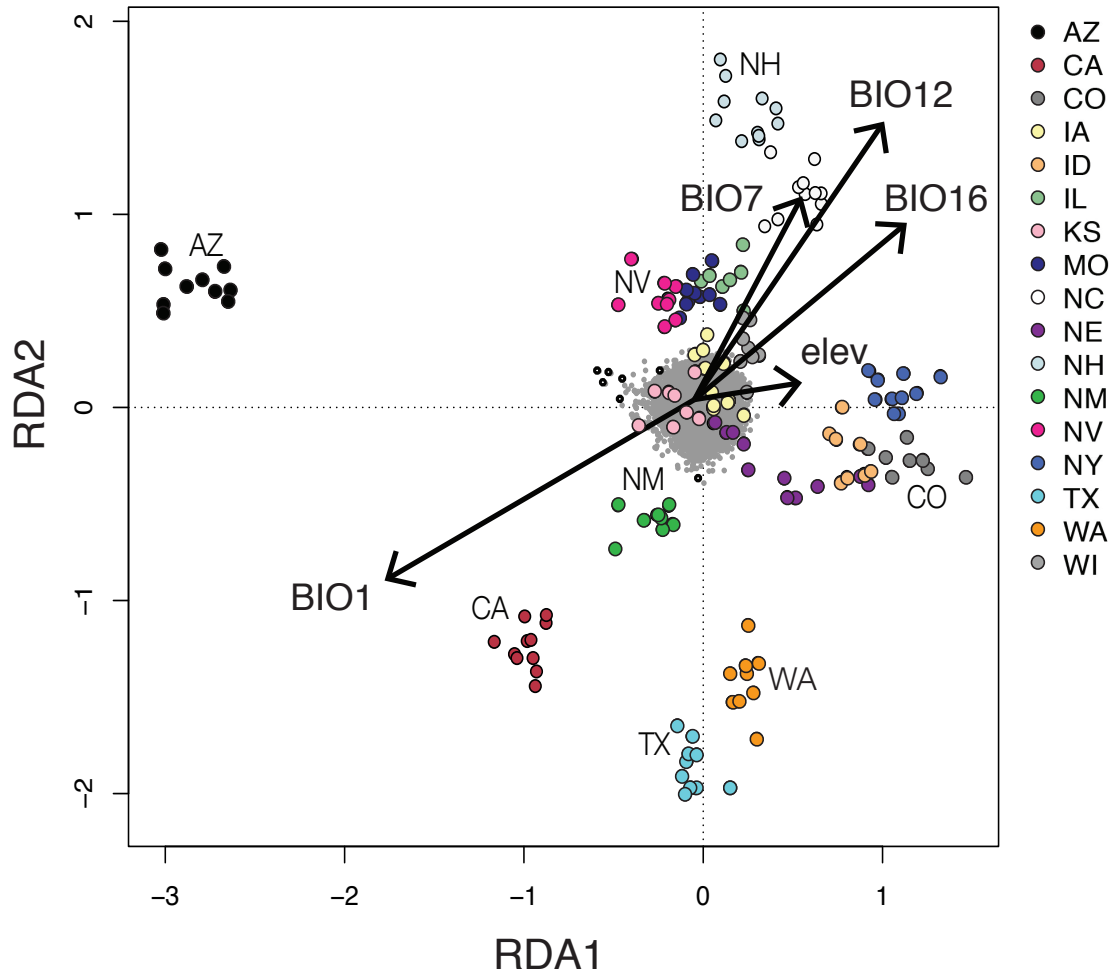
Figure 2.3. Evidence for incipient local adaptation. Redundancy analyses indicate that 191 SNPs (small gray points) are associated with bioclimatic predictors (vectors). BIO1: mean annual temperature; BIO7: temperature annual range; BIO12: annual precipitation; BIO16: precipitation of wettest quarter. Significant associations are shown in black dots.

DISCUSSION

Our genome-wide data reveal that genetic variation in invasive starlings is associated with environmental parameters, and we suggest that these associations might result from adaptive processes. Although there is a significant (but low magnitude) signal of isolation-by-distance, hierarchical AMOVAs find that variation within and among individuals explains observed differentiation better than variation among sampling sites. There is no evidence of population structure, and while models indicate some subtle spatial patterns of genetic variation, these model-based inferences likely reflect sampling artifacts (Appendix A.1, A.2, Figures A2 and A4). Finally, after the initial founder effect, the effective population size has grown by a ratio of 1:100. These patterns are consistent with the expectation that extensive gene flow—as shown by low F_{ST} among sampling sites (Table 2.2)—maintains high connectivity among North American starling "populations." When interpreted within the context of a complementary exploration of movements inferred from feather isotope assays (Werner, Fischer, & Hobson, 2020), this genomic survey confirms that dispersal and migration continue to influence the distribution of genetic variation as a result of just over a century of range expansion.

Evidence of local adaptation in North American starlings

This species' low genome-wide differentiation across the continent allows for tests of selection on loci that may be involved in local adaptation. We find that almost 200 of the 15,038 RAD loci appear to be under selection using a redundancy analysis. We discuss additional tests of selection in Appendix A.5, Figures A7 and A8, but only

13 of the SNPs identified in the RDA model overlap with the SNPs identified by another selection scan described in Appendix A (a latent-factor mixed model, Figure A8). It is unsurprising that each test identifies different candidates for selection, because the assumptions underlying each differ (for more details on selection testing, see Appendix A.5). Rather than making inferences based on the genes identified by these scans, we instead propose that genotype-environment associations show that changes in precipitation and temperature can explain genetic variation in North American starlings. Specifically, we hypothesize that aridity and cold temperatures that are not experienced in the starling's native range exert enough selective pressure on North American starlings to result in incipient local adaptation. This finding suggests that local adaptation may explain genetic variation within the North American starling invasion, and we now discuss relevant caveats to this conclusion.

Geography alone does not fully explain genetic structure

When we compare the relative importance of geographic and environmental distances in partial Mantel tests, we find that environmental conditions better explain genetic variation. We note that concordance among environmental and genetic distances indicates that spatial autocorrelation complicates our selection inferences. Under these conditions, any evidence for selection is likely to be weak, and these selection scans can generate false positives. However, of comparable genotype-environment association methods, RDA has the highest rate of true positives and lowest of false positives, and although this method has not been tested in such recent expansions, RDA is well-suited to systems where F_{ST} is very low (Meirmans, 2015, Forester *et al.*, 2018, Appendix A.5).

In addition, when we shuffle environmental predictors randomly, we find no evidence for genotype-environment associations, lending additional support to the RDA model shown here.

The environmental conditions that we expect to drive selection—precipitation and temperature—vary most substantially in the southwestern region: for example, the sampling location in Arizona is consistently warmer (BIO1) and drier (BIO12 and BIO16) than other locations (Figure 2.3, Appendix A.4, Figure A6). Starlings in the Western region experience these environmental conditions year-round, which could allow selection to drive advantageous alleles toward fixation. Elsewhere in the U.S., starlings move more freely among states: individuals within each sampling location may come from different breeding populations, and additional sampling could reveal stronger population structure among true breeding populations. However, if our sampling overlooked some true populations, we would expect some signal of population structure. Individual-based tests of population structure—e.g., those that do not define possible populations *a priori*—do not recover any signals of population structure.

Movement among sampling locations may influence genetic variation
Our study focuses on birds collected during the winter, which limits our inferences about selection to overwintering sites. We interpret the isotopic evidence presented in Werner, Fischer, and Hobson (2020) as showing that starlings in the western U.S. tend to move regionally whereas birds sampled in the eastern U.S. undertake longer movements (Table 2). This in turn suggests that starlings overwintering in the western

U.S. are more likely to breed nearby, and thus the environmental conditions that starlings experience may not change as dramatically among wintering and breeding ranges. However, we recognize that these environmental conditions represent those experienced during the non-breeding season, and therefore they do not represent the full range of conditions experienced by the starlings sampled; see Appendix A.4 for a discussion of our assumptions about this environmental sampling.

We note that lower rates of gene flow in the western U.S.—of which we find no evidence—could explain allele frequency shifts that we infer to be selection. Starling movement (e.g., migration and/or dispersal) should influence genetic differentiation among sampling locations: we expect high F_{ST} where gene flow among sites is lower, e.g., where starlings are assigned to a nearby sampling location according to Werner, Fischer, & Hobson's (2020) model of molt origin. Starlings in the western U.S. appear to have differentiated subtly from their eastern counterparts based on the higher F_{ST} between Arizona—and to some degree, New Mexico and Colorado—and all other sampling locations (Table 2). Birds collected in these southwestern states are also assigned to those states by discriminant function analysis (Werner, Fischer, & Hobson, 2020); for example, 67% of starlings collected in New Mexico were also assigned to that state (Table 2). We suggest that birds assigned to the same state where they were collected may reside in that state year-round, but we note that collecting feathers once in the lifespan of the bird does not allow us to determine the bird's lifelong migration and dispersal. This model indicates that dispersal and seasonal migration among some sampling locations could lead to gene flow among geographically distant "populations," and as expected, movement is not uniform across the starling's North

American range. Given this evidence, we suggest that environmental conditions may also influence starling movement across the landscape.

Rapid expansion and allele surfing may contribute to putative local adaptation

In North America, the effective population size (N_e) of the present-day invasive population has expanded dramatically, with models indicating that current N_e is even larger than the N_e of the founding population (Appendix A.3, Table A2, Figure A5). Given this demographic history of rapid expansion after a genetic bottleneck, it remains possible that allele surfing at the range edge could mimic these patterns of local adaptation (Excoffier *et al.*, 2009; Slatkin & Excoffier, 2012). Theory predicts that allele surfing should drive alleles closer to fixation, but no putative variant under selection has an allele frequency greater than 0.28. Furthermore, given that the earliest North American starling population was established in the mid 19th-century—which allows only a short time for mutations to accumulate—it is likely that any allele frequency shifts reflect evolutionary processes acting on standing, ancestral variation. Future work might investigate whether particular alleles are shared with the ancestral population(s) or generated via mutation. Here we do not explicitly control for linkage, and we note that some allele frequency shifts could also be explained by linkage or by genetic hitchhiking (Cruickshank & Hahn, 2014).

We suggest that explosive population growth from hundreds to millions of individuals may have encouraged long-distance dispersal away from the dense populations in eastern North America. We know that long-distance dispersal is common in introduced starlings in South Africa (Hui *et al.,* 2012) and Australia (Phair *et*

al., 2017), where rates of dispersal may be determined by demographic changes and environmental quality. Invasive starlings in South Africa disperse when their natal environment becomes crowded or unsuitable, e.g., at the leading edge of their range expansion (Hui *et al.,* 2012). In general, multiple lines of evidence have shown that the dispersal rates and distances of juvenile starlings in North America are remarkably high (Dolbeer, 1982; Cabe, 1999; Werner, Fischer, & Hobson, 2020). Whether and how dispersal might influence local adaptation of expanding starling populations remains an open question.

Local adaptation in other starling invasions and more

A similar project on starlings in the Australian invasion—which colonized that continent nearly concurrently with the North American invasion—found that geographic but not environmental distance explains genetic patterns there (Stuart & Cardilini *et al.,* 2020). Starlings in the Australian range show substantial population structuring and significant patterns of isolation-by-distance. Earlier work had shown that gene flow among Australian starling populations is low (Rollins *et al.,* 2009), and phylogeographic patterns of mitochondrial sequence variation confirm that starlings on the edge of the expansion front in Western Australia have differentiated from those still living in the introduction site (Rollins *et al.,* 2011). In fact, starlings at the expansion front may have rapidly adapted during the Australian invasion (Rollins *et al.,* 2016): the proportion of adult starlings in Western Australia carrying a novel mitochondrial haplotype has increased rapidly only at this range edge, which could indicate allele surfing. The genotyping-by-sequencing survey referenced above now indicates three

population subdivisions in Australia, and global F_{ST} across all Australian populations is an order of magnitude higher than the equivalent F_{ST} index across North America, despite similar areas sampled (Stuart & Cardilini *et al.*, 2020). In Australia, the strong evidence for isolation-by-distance and founder effects complicate attempts to disentangle selection from drift, yet despite their differences in invasion dynamics, genotype-environment associations reveal signatures of selection in both invasions. We note that preliminary results of whole-genome resequencing of native (UK) and introduced (Australian and North American) populations suggest that variability in temperature and precipitation may shape observed genetic variation in starlings world-wide (Hofmeister et al., *in prep*).

Our results contribute to the growing evidence of rapid adaptation in some expanding populations, even in young systems. Some studies of rapidly expanding invasions find little evidence that adaptation may facilitate this expansion, as in corals (Leydet et al. 2018). However, other work suggests a role for selection in supporting rapid range expansion, such as in experimental studies of flour beetles (Szucs *et al.,* 2017) and empirical work in guppies (Baltazar-Soares *et al.,* 2019). Invasion biologists have long highlighted propagule pressure as a driver of invasion success, but the genetic composition may be just as important as the size of the establishing population (Briski *et al.,* 2018). For example, genetic bottlenecks in monk parakeets, another avian invader now distributed world-wide, do not seem to inhibit invasion success (Edelaar *et al.,* 2015). Pre-adaptation in the native range or selection during transport may facilitate the spread of invasive species, and human commensalism may support establishment and spread, as shown in house sparrows (Ravinet *et al.,* 2018) and common mynas

(Cohen *et al.,* 2019), and reviewed across alien bird species (Cardador & Blackburn, 2019). Empirical studies of invaders like the ones described here also show how, in addition to genetic variation, epigenetic shifts and/or plastic changes in gene expression may support the establishment and expansion of invasive species (Marin *et al.,* 2019). In the well-studied house sparrow—a system quite similar to starlings— epigenetic shifts may have supported invasions in Africa (Liebl *et al.,* 2013) but not necessarily in Australia (Sheldon *et al.,* 2018). Taken together, recent work suggests that we should consider a much wider range of demographic and ecological processes that lead to adaptive evolution in invading populations.

Invasive populations allow us to explore the genetic consequences of colonization and establishment in novel environments. On a background of low genetic differentiation and diversity, we find evidence of incipient genotype-environment associations in North American starlings. Here we explore how genetic variation changes across the landscape, but we cannot fully understand gene flow without studies of dispersal and migration of the individuals that carry genes. Our results complement other recent studies that reveal associations between climate variables and particular loci in North American vertebrates (Schweizer *et al.,* 2015; Bay *et al.,* 2018). Finally, we suggest that our study adds to those suggesting that rapid local adaptation can evolve even in dispersive and young populations.

REFERENCES

Ashenden T, Rooke AC, Fox MG (2017) Boldness and dispersal tendency of native and invasive pumpkinseed (*Lepomis gibbosus*): is spatial sorting creating superior invaders? *Aquatic Invasions*, 12, 311-320.

Avalos A, Pan H, Li C *et al.* (2017) A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nature Communications*, 1–9.

Baltazar-Soares M, Blanchet S, Cote J, Tarkan AS, Záhorská, Gozlan RE, Eizaguirre C (2019) Genomic footprints of a biological invasion: introduction from Asia and dispersal in Europe of the topmouth gudgeon (*Pseudorasbora parva*). *Molecular Ecology*, **29**, 71-85.

Bay RA, Harrigan RJ, Le Underwood V *et al.* (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, **359**, 83–86.

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

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

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

Brown GP, Phillips BL, Shine R (2014) The straight and narrow path: the evolution of straight-line dispersal at a cane toad invasion front. *Proceedings of the Royal Society B*, **281**, 20141385.

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

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

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

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

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

Cohen, TM, McKinney, M, Kark S, Dor R (2019) Global invasion in progress: modeling the past, current, and potential global distribution of the common myna. *Biological Invasions*, **21**, 1295-1309.

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

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-3157.

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

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

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

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

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

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

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

Edelaar P, Roques S, Hobson EA, Da Silva AG, Avery ML, Russello MA, Senar JC, Wright TF, Carrete M, Tella JL (2015) Shared genetic diversity across the global invasive range of the monk parakeet suggests a common restricted geographic origin and the possibility of convergent selection. *Molecular Ecology*, **24**, 2164-2716.

Estoup A, Ravigné V, Hufbauer R *et al.* (2016) Is There a Genetic Paradox of Biological Invasion? *Annual Review of Ecology, Evolution, and Systematics*, **47**, 51–72.

Excoffier L, Foll M, Petit, RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481-501.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLOS Genetics*, **9**, e1003905.

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

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

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

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

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

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

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

Hufbauer RA, Facon B, Ravigné V *et al.* (2011) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications*, **5**, 89–101.

Hui C, Roura-Pascual N, Brotons L, Robinson RA, Evans KL (2012) Flexible dispersal strategies in native and non-native ranges: environmental quality and the "good-stay, bad-disperse" rule. *Ecography*, **35**, 1024–1032.

Jacob S, Legrand D, Chaine AS *et al.* (2017) Gene flow favours local adaptation under habitat choice in ciliate microcosms. *Nature Ecology & Evolution*, 1–4.

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

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

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

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

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

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

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

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

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

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

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

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

Lombaert E, Estoup A, Facon B *et al.* (2014) Rapid increase in dispersal during range expansion in the invasive ladybird *Harmonia axyridis*. *Journal of Evolutionary Biology*, **27**, 508–517.

Marin P, Genitoni J, Barloy D, Maury S, Gibert P, Ghalambor CK, Vieira C (2019) Biological invasion: The influence of the hidden side of the (epi)genome. Functional Ecology, 34, 385-400.

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

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

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

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species (L Orlando, Ed,). *PLoS ONE*, **7**, e37135–11.

Pfeifer SP, Laurent S, Sousa VC *et al.* (2018) The Evolutionary History of Nebraska Deer Mice: Local Adaptation in the Face of Strong Gene Flow (N Singh, Ed,). *Molecular Biology and Evolution*, **35**, 792–806.

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

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

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

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

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

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

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

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

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

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

Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling

(*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology*, **18**, 1560–1573.

Rosenberg NA (2003) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Schlaepfer DR, Glättli M, Fischer M, van Kleunen M (2009) A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. *New Phytologist*, **185**, 1087–1099.

Schmid-Hempel P, Schmid-Hempel R, Brunner PC, Seeman OD, Allen GR (2007) Invasion success of the bumblebee, Bombus terrestris, despite a drastic genetic bottleneck. *Heredity*, **99**, 414–422.

Schrieber K, Lachmuth L (2017) The Genetic Paradox of Invasions revisited: the potential role of inbreeding X environment interactions in invasion success. *Biol Rev*, **92**, 939-952.

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

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

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analogue of genetic drift. *Genetics*, **191**, 171-181.

Stuart KC, Cardilini APA, Cassey P, Richardson MF, Sherwin W, Rollins LA, Sherman CDH (2020) Signatures of selection in a recent invasion reveals adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, **00**, 1-16. doi: 10.111/mec.15601.

Szűcs M, Vahsen ML, Melbourne BA *et al.* (2017) Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences*, **114**, 13501–13506.

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

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

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

Werner SJ, Fischer JW, Hobson KA (2020) Multi-isotopic (δ^2H, δ^{13}C, δ^{15}N) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. *PLoS One*, **15**, e0237137.

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

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

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

CHAPTER 3: CONCURRENT INVASIONS BY STARLINGS ILLUSTRATE GENOMIC CONSEQUENCES OF INVASION

ABSTRACT

A species' success during the invasion of new areas hinges on an interplay between the demographic processes common to invasions and the specific ecological context of the novel environment. Evolutionary genetic studies of invasive species can investigate how genetic bottlenecks and ecological conditions shape genetic variation in invasions, and our study pairs two invasive populations that are hypothesized to be from the same source population to compare how each population evolved during and after introduction. Invasive European Starlings (*Sturnus vulgaris*) established populations in both 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 evolution to generate similar genetic patterns in these recent and replicated invasions. Demographic models indicate that both invasive populations experienced genetic bottlenecks, and we find that specific genomic regions have differentiated even on this short evolutionary timescale. Despite these bottlenecks, we suggest that genetic drift alone cannot explain differentiation in at least two of these regions, given that within-population nucleotide diversity and neutrality statistics indicate a role for purifying and/or diversifying selection. Instead, the patterns of genetic variation we find are

consistent with the hypothesis that this infamous and highly mobile invader adapted to novel selection (e.g., extrinsic factors), perhaps in part due to the demographic boom intrinsic to many invasions.

INTRODUCTION

Some species can establish and spread in a novel environment more successfully than others, and defining what makes a species 'invasive' is hotly contested (Ricciardi and Ryan 2017; Russell and Blackburn 2017; Sagoff 2019; Davis 2020). One open question is whether invasion success is more accurately explained by intrinsic properties of the founding population and/or by extrinsic conditions experienced by the population. Intrinsic properties such as propagule pressure or standing genetic variation certainly influence a population's establishment (Dlugosch et al. 2015). At the same time, recent work indicates that local, extrinsic factors are central in predicting establishment (Redding et al. 2019). Theoretical frameworks that emphasize eco-evolutionary feedbacks may facilitate more accurate predictions of invasion success: a population's expansion depends not only on local ecological conditions but also the evolutionary potential of traits such as dispersal ability (Williams et al. 2019; Miller et al. 2020). However, if we are to predict how evolutionary history figures in the rapid expansion of an invasive population, we first need an accurate measure of genetic variation upon establishment.

When invasive populations colonize a new environment, they often undergo genetic bottlenecks (Lowe et al. 2017). However, even populations with limited genetic diversity—including those subject to founder effects—can adapt quickly to novel environments (Kolbe et al. 2004; Dlugosch and Parker 2008; Marques et al. 2018). For

example, genetic drift (or gene surfing) during range expansion has led to adaptation in experiments in microbes (Gralka et al. 2016) and wild, invasive bank voles (White et al. 2013). Simultaneous with demographic shifts, invasive populations may encounter environmental conditions that exert novel selective regimes. Although genetic bottlenecks often limit the genetic variation available to selection, local adaptation can occur even after a genetic bottleneck (Tsutsui et al. 2000; Kolbe et al. 2004; Verhoeven et al. 2011; Willoughby et al. 2018). Any divergence after introduction likely reflects a combination of demographic and/or selective forces since rapid adaptive evolution after establishment depends on sufficient genetic variation in the invasive population.

In practice, distinguishing among evolutionary mechanisms that explain shifts in genetic variation presents several analytical challenges. Any summary statistic of genetic diversity or differentiation is built on certain assumptions about the demographic history, genomic architecture, and ongoing ecological context that a population experiences (Cruickshank and Hahn 2013). Studies that use genomic scans draw upon multiple summary statistics to explore how violations of these assumptions might lead to inappropriate conclusions about the evolutionary mechanisms operating in a population. Even so, geographic sampling and sequencing quality can introduce additional biases (Korneliussen et al. 2014), and genomic studies of wild invasive populations require methods sensitive to such limitations (North et al. 2021). Recent and replicated invasions may be particularly useful for identifying the relative contributions of ancestral variation and rapid evolution in expanding invasive populations (Dlugosch et al. 2015; Bock et al. 2018; Lu-Irving et al. 2018). One such recent invader is the Common or European Starling (*Sturnus vulgaris*), which was

92

introduced to New York, United States of America in 1890 and across south-eastern Australia from 1856 onwards (Higgins et al. 2006; Linz et al. 2007). Both the American and Australian invasions were most likely founded by individuals from the United Kingdom (Forbush 1915; Kessel 1953), although evidence supports the potential for low levels of ongoing gene flow to Western Australia (Rollins et al. 2009).

The starling's ecology is well-studied in both invasions, enabling us to explore how environmental factors might impact genetic variation (Feare 1984). Native-range starlings thrive in open pastures and urban environments alike, but starlings' ecology and life history vary among populations (Stuart et al., *in review*). In range-wide studies of Australian (Cardilini et al. 2016; Cardilini et al. 2018; Stuart et al. 2020) and North American (Bodt et al. 2020; Hofmeister et al. 2021) invasions, temperature and precipitation influence genetic differentiation, even after controlling for population structure. Migration and dispersal also vary among invasions. In North America, starlings can migrate long distances each season (Kessel 1953; Dolbeer 1982), but populations in the Western USA likely disperse and/or migrate shorter distances (Werner et al. 2020). In contrast, starlings in Australia exhibit strong population structure and likely move short distances in search of food, even though environmental conditions are much more arid than in the native range (Stuart et al. 2020, but see Waterman et al. 2008). Finally, variation in the breeding cycle may also facilitate invasion success as invasive populations tend to lay more clutches than the UK population, and it is possible that environmental differences might also generate differences in clutch size and hatching success between native and invasive populations (Dawson 1983; Feare 1984; Ball and Wingfield 1987). In summary, previous studies of

the ecology and life history of the European starling indicate that local adaptation in invasive populations may support the species' ongoing expansion throughout each invasive range (Stuart et al., *in review*).

We use concurrent starling invasions in Australia (AU) and North America (NA) to examine the evolutionary and genetic consequences of invasion. Complementing more shallow sampling across the expanding invasions in North America and Australia (Stuart et al. 2020; Hofmeister et al. 2021), we compare population genetic variation in the hypothesized introduction sites of both invasions with the putative ancestral population in the United Kingdom (UK). Both starling invasions rapidly expanded from small founding populations in the late 19th century (1890 in North America, Forbush 1915; Kessel 1953; 1856 in Australia, Higgins et al. 2006). Given this natural experiment, we take advantage of a rare opportunity to compare intrinsic and extrinsic drivers of invasion success. Founder effects and other intrinsic demographic properties of invasions certainly influence establishment (Briski et al. 2018; Irwin et al. 2018), and we predict some divergence from the native, ancestral population (represented by UK samples here) due to genetic drift. We combine demographic models with local measures of genomic diversity and differentiation to determine whether drift alone can explain observed genetic variation, and we find evidence that extrinsic factors such as novel selective regimes may generate differentiation in both invasions specific to only a few genomic regions.

METHODS

Sampling and whole-genome re-sequencing

Eight individuals from each continent were collected in 2016: 8 from New York City, USA (abbreviated as NA), 8 from Northumberland, UK, and 8 from southeastern Australia (coordinates for each individual sample listed in Table S1). Libraries for each individual starling were constructed using a TruSeq DNA PCR-free High Throughput Library Prep Kit (Illumina, San Diego, CA). All individuals passed the initial quality check with FASTQC (Babraham Bioinformatics, Cambridge, UK). Adapters were removed using AdapterRemoval (Schubert et al. 2016) and reads mapped to the reference *S. vulgaris* vNA genome (GCF_001447265.1) (Stuart et al. 2021) genome using BOWTIE2 (Langmead and Salzberg 2012) and checked for mapping quality using qualimap (Okonechnikov et al. 2015). Sequencing quality was relatively high: 96.4% of reads mapped to the *S. vulgaris* genome with a coverage of 18.4X and a mapping quality of 26.9. Reads were also mapped to a pseudo-chromosome-level *S. vulgaris* vNA genome, where scaffolds were assigned to chromosomes based on orthology to the zebra finch reference genome (GCF_000151805.1) (Grabherr et al. 2010). Assuming orthology, we were able to predict centromere positions based on the known genomic architecture of the zebra finch (Knief and Forstmeier 2016), but this study does not directly identify centromere position.

We called variants using GATK's HaplotypeCaller in GVCF mode and flagged low-quality variants using GATK Best Practices (QD<2, FS>60, MQ<40, and SOR>3, accessed March 21, 2018 (McKenna et al. 2010). We filtered sites for missing data, depth, and quality using vcftools (parameters: --max-missing 0.8 --min-meanDP 2 --max-meanDP 50 --remove-filtered-all), which removed 4.1 million sites from the original SNP set and left a total of 23.4 million sites for downstream analyses. Starting from the

mapped reads used in the GATK pipeline, we also called SNPs based on a minimum p-value of the correct genotype probability at each site using ANGSD (Kim et al. 2011; Korneliussen et al. 2014). Filtering for SNP p-value (0.0001), depth (between 60 and 400 sequences), and mapping quality (>20) left 16,151,007 sites. All scripts used in read processing and filtering are available on GitHub:

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

Population structure

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

We estimated variance among and between individuals using a principal components analysis in SNPRELATE (Zheng et al. 2012) (GATK) and a covariance matrix built in ngsTools (Fumagalli et al. 2014) (ANGSD). We used ADMIXTURE (Alexander and Lange 2011) (variant-called) and NGSADMIX (Skotte et al. 2013) (likelihood-based) to examine shared ancestry among individuals, and we also measured pairwise genetic distances using ngsDist for the likelihood-based dataset only.

Demographic inference

We used FASTSIMCOAL2 (Excoffier et al. 2013) to explicitly test for genetic bottlenecks in each population. FASTSIMCOAL2 takes a site-frequency spectrum (SFS) as input, and we used the SFS estimated from ANGSD given that likelihood-based estimates are more robust to sequencing error (Nielsen et al. 2012). Demographic models in FASTSIMCOAL2 used priors on the time (TBOT = 10 to 300) and size of the bottleneck (NBOT = 10 to 1000). This prior was chosen because historical records of starling introductions indicate an initial transport of 180 individuals in the NA population (Forbush 1915), and we expect a similar introduction effort in the AU population (Rollins et al. 2009). The command line arguments were as follows: -M -n 1000000 -L 50 -q -k 100000. Each run began with a randomly generated seed (-r), and the -k flag simply writes polymorphic sites to a temporary file to cope with the high memory usage of this analysis. Scripts for demographic analyses can be found at https://github.com/nathofme/global-RESEQ/blob/master/demography.sh. To verify this demographic model, we also estimate inbreeding coefficients (F-statistics) using the –het command in VCFTOOLS and calculate relatedness among each pair of individuals using the –relatedness method of Yang et al. (2010) in VCFTOOLS.

Sliding window scans

For scans of genetic divergence and diversity, we used a variant dataset filtered only for depth and quality: we kept variants that had less than 20% missing data across all sites but did not apply minor allele frequency (MAF) filters, filter for HWE, linkage, or any

other factors. Given that rare alleles likely provide the strongest evolutionary signals in this system, we did not want to filter out any alleles that might have been rare in one population (e.g., the native UK population) and increased in frequency in another population (e.g., the AU or NA invasions). Nevertheless, we test for sensitivity to this filtering choice in Appendix B.2.B.

We calculate F_{ST} and nucleotide diversity (π) using overlapping 50-kb sliding windows with a step size of 10kb using VCFTOOLS (Danecek et al. 2011). We calculate nucleotide diversity separately for each population. We then calculate d_{xy} using a Python script by Simon Martin (accessed at https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py on February 12, 2020). This analysis includes all confident variant calls (parameters: --output-mode EMIT_ALL_CONFIDENT_SITES in GATK's HaplotypeCaller); only with this additional parameter do we recover levels of d_{xy} similar to other systems. We also measured F_{ST} and π in overlapping 10-kb windows to localize elevated F_{ST} to an even smaller region of the genome. To visualize relationships between diversity metrics, we plotted the mean values of each metric in a 50-kb window. All scripts for plotting are stored on GitHub: https://github.com/nathofme/global-RESEQ/.

Identifying candidate genes under selection

We first identify specific genes underlying all regions of elevated F_{ST} using BLAST (E-value = 40), and manually curate the gene lists for each window by selecting genes previously identified in avian species. We then use these gene lists to conduct network analyses of gene ontology (GO) terms, which can provide a more holistic and objective

98

method of identifying shared functions. Network analyses can also provide more statistical power, correcting for the usual problem of multiple testing. To identify functions of candidate regions, we quantified the uniqueness and dispensability of each GO term using REVIGO, a method that quantifies semantic similarity (Supek et al. 2011). This analysis by default emphasizes GO terms that are rare in the list of candidates provided. We include a figure of raw REVIGO output in Appendix A that groups these GO terms into broader categories to determine the general functions underlying F_{ST} peaks.

RESULTS

Differentiation and population structure

After filtering, we obtained more than 11 million SNPs with minimum 5X coverage (average genome-wide coverage = 18.44, for details on how choice of reference genome impacts variant-calling, see Appendix B.2). All patterns identified using a variant-called dataset concur with those based on genotype likelihood (ANGSD); for details on how variant-calling impacts patterns, see Appendix B.3. We use a variant set filtered for minor allele frequency (MAF), Hardy-Weinberg Equilibrium, and linkage disequilibrium for analyses of population structure, but for more accurate estimates of genetic diversity and differentiation, we report results from a genome scan of variants filtered only for quality and depth. Differentiation of the two invasive populations from the native population is low, which is expected given that these populations split less than two centuries ago. However, genome-wide mean differentiation between AU and UK (F_{ST} AU vs. UK = 0.029) is almost twice that between NA and UK (F_{ST} NA vs. UK =

0.016; based on the dataset of 868,685 SNPs filtered for a minor allele count (MAC) of 2 and pruned to one SNP per 1000 Kb). We examined this contrast in genetic differentiation using analyses of population structure.

All three populations are readily distinguished from one another by a principal component analysis of the same set of 868,685 unlinked SNPs (MAC > 2): PC1 (6.3% of genetic variation) separates the UK and NA populations from the Australian population (Figure 3.1A). This evidence complements previous work that showed extensive population structuring in Australia but nearly continuous gene flow across North America, based on reduced-representation genomic data (Stuart et al. 2020; Hofmeister et al. 2021). Principal component analysis in the genotype likelihood framework of ANGSD (Korneliussen et al. 2014) shows nearly identical results (Figure B6). Furthermore, individuals are reliably assigned to clades based on pairwise genetic distances calculated in ANGSD (Figure 3.1B).

We note that the tight clustering of UK individuals in Figure 3.1A contrasts with the large distances between these same individuals in Figure 3.1B. In the genotype likelihood dataset, genetic distances between native UK individuals are much greater than distances among individuals within each invasive population (Figure 3.1B). Because these two datasets differ in variant-calling and filtering strategies, the genetic distance among UK individuals in Figure 3.1B may reflect rare alleles that were filtered out of the variant set in Figure 3.1A. However, PCs 3 and 4 in the PCA of the variant-called dataset do indicate additional structure within the UK population (Figure B7).
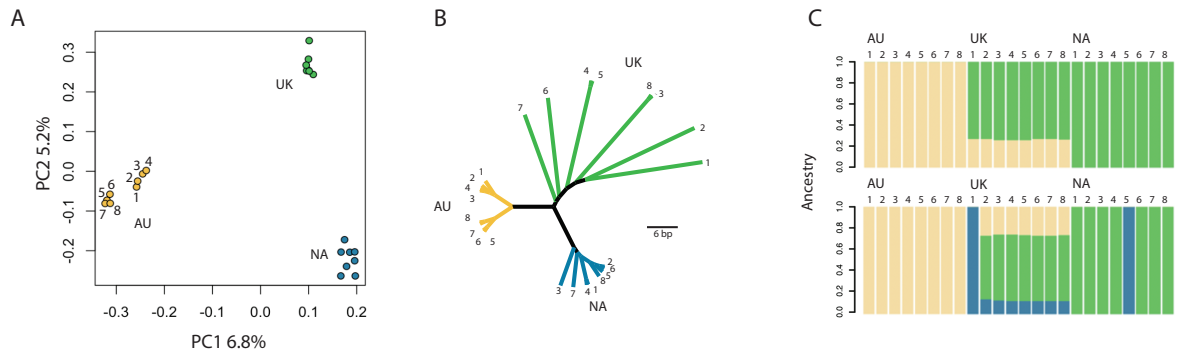
Figure 3.1. A) Principal components of 868,685 unlinked SNPs explain 6.8% (PC1) and

5.2% (PC2) of genetic variation. B) cladogram of genetic distances among samples based

on genotype likelihoods of 16,151,007 sites. C) ADMIXTURE analyses showing K=2

(top) and K=3 (bottom row).

101

Admixture analyses revealed statistical support for a two-population model, which distinguished AU from NA+UK and was only slightly weaker (cross-validation error = 0.88) than the best-supported model of K=1 (cross-validation error = 0.73, Figure 3.1C, Appendix B.3). Regardless of the number of populations hypothesized in models of admixture, the NA invasion consistently shares a higher proportion of its ancestry with the UK population. Individuals from the Australian population are distinguished from the other two populations in all tested values of K. Considered in concert, these tests of population structure show that Australian and North American populations differ in the amount of divergence from the native UK population. Founder effects likely contribute to the observed population structure, and below we describe explicit models of demographic processes.

All populations experienced bottlenecks and subsequent expansion
To address the impact of demographic processes in generating the observed patterns, we used the site-frequency spectrum built from genotype likelihoods to construct models of changes in the effective population size over time. We examined the demographic history of all three populations using FASTSIMCOAL2 (Excoffier et al. 2013). The demographic model shows that both invasions experienced a bottleneck upon introduction (Figure 2; Table S4), which is expected given historical records of a small number of founding individuals in both AU and NA. Each invasion appears to have recovered quickly by expanding in effective population size, and at present, our data suggest similar effective population sizes in both the AU and NA invasions (Ne AU = 60,306; Ne NA = 64,411). We note that this demographic inference could be biased: the

allele frequency distribution draws only from the individuals sampled from the putative origin of each invasion, and our sampling likely does not capture population-wide variation across each continent.

Bottlenecks often lead to inbreeding within a population, and we find that inbreeding is negligible but slightly higher in the Australian population than in the NA population (Table S3). We do identify two individuals with remarkably high inbreeding (AU1 = 0.32, and US3 = 0.49), confirming that bottlenecks may shape much of the genetic variation observed here. However, relatedness among individuals is generally low, where a zero inbreeding value indicates no shared alleles among individuals (maximum AJK statistic = 0.06). Much of the ancestral variation appears to be shared among invasions, given that the genome-wide average F_{ST} between AU and NA is 0.04, although we note that genetic differentiation among invasions confirms the expectation that different variants will make it through each genetic bottleneck. The results presented here concur with range-wide sampling that indicates genetic bottlenecks followed by rapid expansion with little evidence of inbreeding in both Australia (Stuart et al. 2020) and North America (Hofmeister et al. 2021).
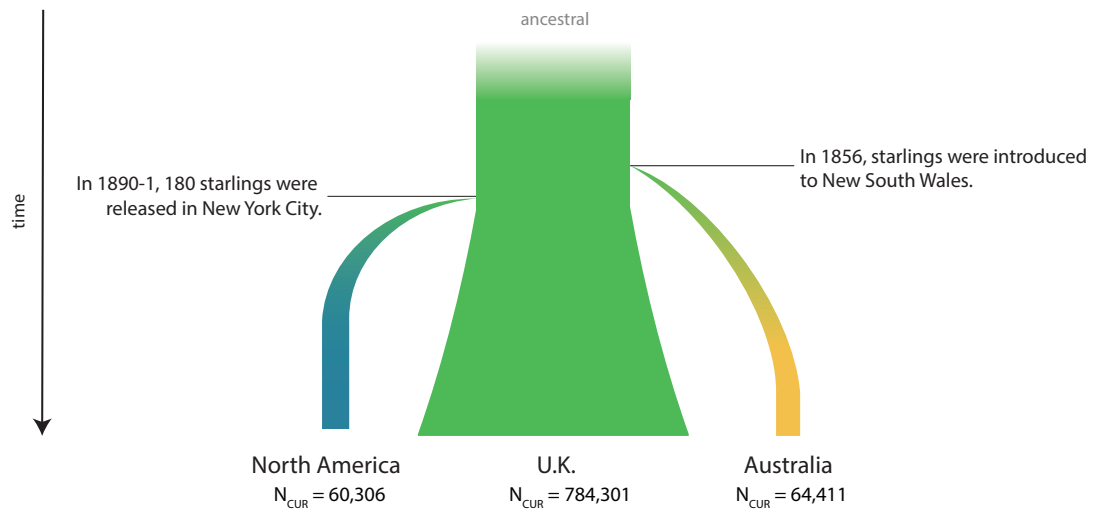
Figure 2. Demographic model of effective population size based on the site-frequency spectrum. Schematic approximates population growth based on model output from FASTSIMCOAL2, where current population estimates are labeled under each population.

Differentiation of a few genomic regions reveal consequences of invasion

Starling populations colonized their invaded ranges less than two centuries ago, and the age of these populations makes it somewhat surprising to find loci specific to the derived populations with fixation indices as high as 0.57 in a single 10-kb window (Figure 3). However, no putative outlier windows approach fixation in any comparison across all individuals in a population. We consider only the top 0.1% of 50-kb windows to be F_{ST} outliers (AU vs. UK: $F_{ST} > 0.21$; NA vs. UK: $F_{ST} > 0.18$). We report and interpret all regions of elevated F_{ST} (in the top 0.1% of windows) between any of the three populations in Appendix B.8: using this threshold, we find 21 50-Kb windows spread across Chromosomes 1, 2, 3, 6, and the Z chromosome (Table S6; Figures S13-14). High levels of differentiation paired with a reduction in diversity—such as those observed here—may result from suppressed recombination (e.g., proximity to the centromere or a structural rearrangement), or from alleles approaching fixation or loss due to drift or selection (e.g., a selective sweep). We find that most putative outlier regions are distant from the centromere location (predicted via homology with the zebra finch (*Taeniopygia guttata*) genome, Table S5), although we note that future work in the system should identify how recombination varies across each chromosome. Recombination rates are not measured in this study, as we have neither a chromosome-level assembly nor sufficient sampling to test for linkage (e.g., family groups). Nevertheless, we expect that genomic architecture—including linkage disequilibrium independent of the centromere—plays a role in how differentiation among populations is generated. We also note that absolute genetic differentiation (d_{xy}) could reveal whether genetic differences have accumulated after invasive populations split from the ancestral

population, as this test statistic is independent of the level of diversity within each population (Cruickshank and Hahn 2014). However, we find that changes in d_{xy} in these data provide no additional insight (Figure B15): there is no evidence that d_{xy} departs from F_{ST} under differentiated peaks, and instead d_{xy} remains unchanged under peaks.

We use population genetic test statistics (F_{ST}, π, and Tajima's D) to evaluate the relative contributions of drift, selection, and recombination in regions of elevated F_{ST} in all six regions of elevated F_{ST} to provide preliminary insight into whether replicated invasions in fact show common signatures of evolutionary processes beyond the expected genetic drift. Two regions of elevated F_{ST} indicate more easily interpreted patterns: (1) Chromosome 2, which shows high differentiation between the Australian invasion and both NA and UK populations (Figure 3C), and (2) Chromosome 6, where both invasions have differentiated from the UK population but F_{ST} between AU and NA is remarkably low (Figure 3D).
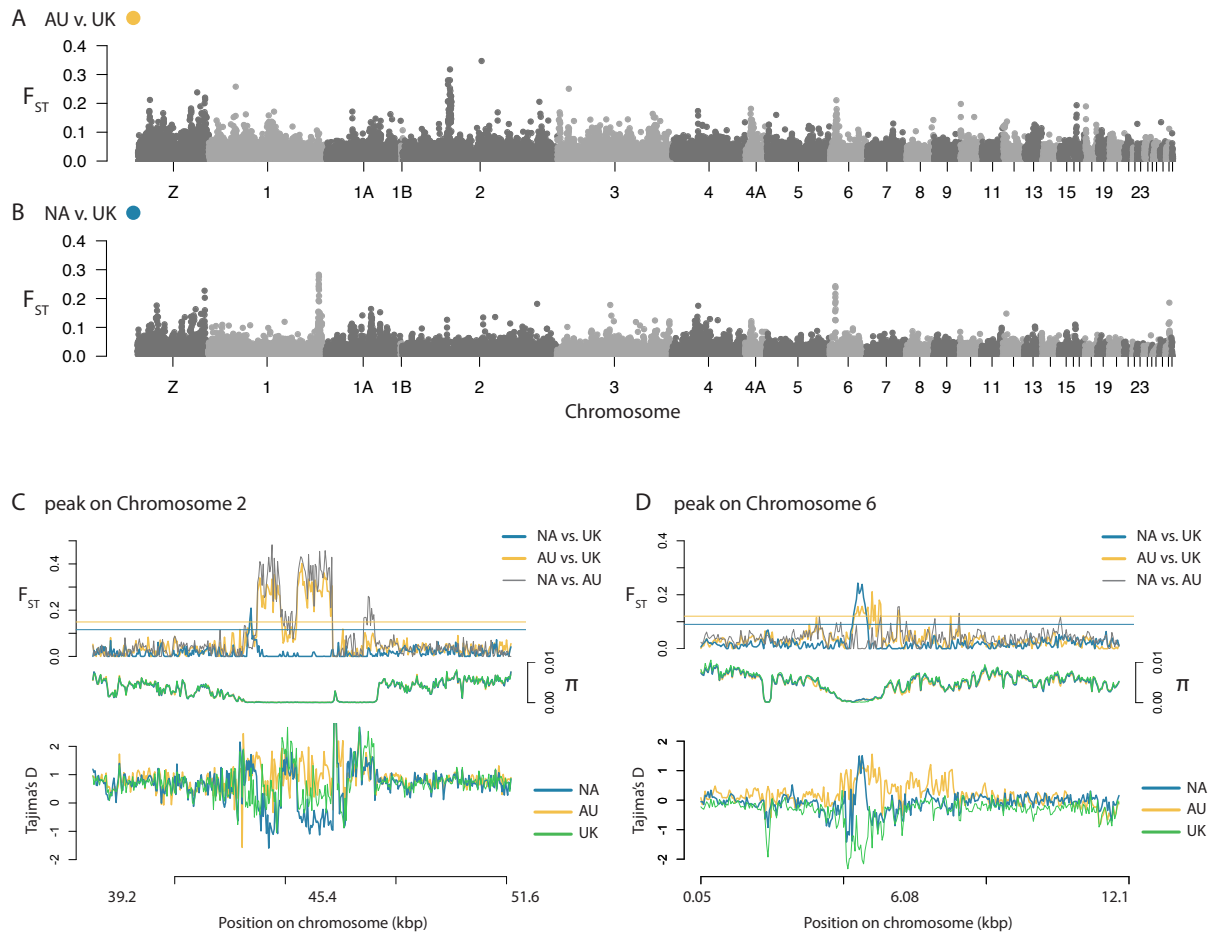
Figure 3. A-B) Manhattan plots show 50-kb windowed F_{ST} between AU & UK (A) and NA & UK (B) populations. C-D) 50-kb windowed F_{ST} (top row), π within each population (middle), and Tajima's *D* within each population (bottom row), centered on the elevated F_{ST} regions of Chromosome 2 (C) and Chromosome 6 (D). Color represents each population, except in F_{ST} plot where yellow indicates F_{ST} AU vs. UK, blue indicates F_{ST} NA vs UK, and gray indicates F_{ST} NA vs AU.

Differentiation on Chromosome 2 may result from evolutionary processes specific to the

Australian invasion

Where F_{ST} is highest on Chromosome 2, we find strong evidence of both purifying and balancing selection in all three populations (Figure 3C). We find that nucleotide diversity is very low within every population, and immediately after the block of elevated F_{ST}, we see a sharp increase in nucleotide diversity in all three populations (Figure 3C). If directional selection were driving differentiation between an invading population and its native ancestral population, we would expect to see a decline in nucleotide diversity specific to the invading population but given that π remains low across all three populations, low levels of within-population diversity may simply indicate low recombination. But, as described above, local reductions in π could also result from population bottlenecks experienced during founder events. This explanation, however, does not address the low diversity in the UK population. Because within-population diversity is relatively low in not only this region but across the genome, we suggest that most patterns in within-population diversity reflect either genetic bottlenecks or the sampling strategy in this study.

Nonetheless, it remains possible that selection may operate simultaneously to generate differentiation specific to the Australian invasion. To clarify the relative impacts of a bottleneck with selection, we look to the neutrality test statistic Tajima's *D*, which tests for a departure from the neutral model of evolution by standardizing the difference between observed genetic diversity (π) and expected genetic diversity (θ_w). Where the difference is negative, the observed diversity in a population is lower than expected, and there is an excess of rare alleles: in this case, negative Tajima's *D* indicates

a possible selective sweep to purge diversity. We find exactly this pattern on Chromosome 2: where F_{ST} and π are low, Tajima's D is also negative in all three populations—indicating an excess of low-frequency variants and perhaps purifying selection—but this statistic climbs to high positive values immediately before and after the block of elevated F_{ST}.

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

One region on Chromosome 6 reveals how population expansion could interact with selection
Both invasions have differentiated from the native range in a 4-Mb region of Chromosome 6 (Figure 3D). As a preliminary check, we note that the large distance

109

between this F_{ST} peak and the centromere suggests that low recombination is unlikely to explain differentiation. We suggest that the clearest explanation of this peak invokes selection on previously rare variants, based on three lines of evidence.

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

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

110

event in the UK population, but we note that this shift is specific to only a small region of the genome.

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

Differentiation on the remaining chromosomes may be explained by genetic drift
Genetic drift explains the moderate levels of differentiation on Chromosomes 1 (Figure B13) and 3 (Figure B14): on both chromosomes, we observe no decline in within-population diversity (π) under these F_{ST} peaks, and if anything, Tajima's *D* trends slightly positive. A positive value of Tajima's *D* results when observed diversity in that region is higher than expected and suggests diversifying selection.

Genes under putative selection may aid in invasion success
The region under putative selection on Chromosome 6 overlaps with the coding regions of four genes with dramatically different functions (*JMJD1C, RTKN2, NRBF2,* and

ARID5B), and we suggest that selection on one of these genes might explain the differentiation in the region, with the other genes remaining in linkage disequilibrium with the possible candidate. Among these candidates, we can speculate that *ARID5B* has the most intuitive link to hypothesized selective regimes: this protein is required for adipogenesis and involved in smooth muscle differentiation. The first exon of this gene lies directly under the F_{ST} peak between AU and UK starlings, and muscle growth and fat storage may have been key to dispersal ability. The three other genes that overlap this window are involved in the DNA-damage response (*JMJD1C*), lymphopoiesis (*RTKN2*), and regulating autophagy (*NRBF2*). For details on GO term enrichment in these outlier regions, see Appendix B.6. Regardless of the mechanism driving these loci toward an intermediate frequency, it remains possible that variation at one or more of these loci influenced invasion success by maintaining heterozygosity.

DISCUSSION

An open question in invasion biology is whether an invasive population's success is better attributed to intrinsic properties of the invasive species or to extrinsic factors specific to the novel environment. Invasion success results when a population clears the thresholds of transport, establishment, and subsequent expansion, and evolutionary genetic studies can offer some insight into the relative importance of each invasion stage for a given invasive species (Dormontt et al. 2011; Blakeslee et al. 2020). In invasions, the null hypothesis is that neutral processes—including genetic bottlenecks and/or changes in recombination rate—explain much of the genetic variation measured for studies such as this one.

The Australian and North American starling invasions colonized each continent around the same time (1856 and 1890 respectively), and experienced similar contractions in population size that led to classical founder events upon establishment. The shared decline in genetic diversity represents a shared intrinsic determinant of invasion success, although it is likely that several evolutionary mechanisms work in concert to shape the genetic variation observed in each population of starlings. In this study, we find that population expansion after a genetic bottleneck yields several regions of elevated F_{ST} between native and invasive populations, exactly as expected. We find that differentiation in many regions of the genome can be explained by genetic drift (Table S13), but that examining fine-scale shifts in diversity and differentiation suggest interactions between drift, selection, and recombination.

Demographic models show that both the AU and NA invasions rapidly expanded in population size after the initial bottleneck, which leads to the prediction that many loci would be lost upon establishment of the invasive populations. Founder effects might thus lead to lower within-population diversity (π) and higher between-population differentiation (F_{ST}). If demographic processes explained genome-wide differentiation, then we would expect to find that regions where within-population diversity differs are distributed across many chromosomes. However, shifts in diversity and differentiation occur only in a few narrow regions of the genome, which suggests evolutionary dynamics specific to these regions. We find evidence for heterogeneous evolutionary mechanisms operating across the genome, which provides indirect support for the possibility that recombination and/or selection might generate local shifts in diversity and differentiation independent of demographic factors. Although

113

some regions of elevated F_{ST} (e.g., on Chromosomes 1 and 3) clearly result from genetic drift alone, others may arise via an elevated recombination rate or even directional selection in one invasion (Chromosomes 2 and 6). Whether linked or purifying selection better explains such localized shifts in diversity remains an active area of research (Stapley et al. 2010, Kawakami et al. 2014). Comparisons between the two recent and replicated invasions sampled here indicate several regions where differentiation between invasive populations is remarkably low ($F_{ST} < 0.01$) even as each invasion has differentiated from the native range, suggesting a shared genomic architecture or even a similar response to extrinsic conditions during one or more stages of invasion. Accurate estimates of recombination rate are needed to better address the extent to which intrinsic properties determine invasion success.

Comparisons of the North American and Australian invasions with the UK population indicate that the sampled populations in Australia retain higher levels of within-population genetic diversity relative to the NA population. First, higher diversity in the AU population could be a sampling artifact: sampling in Australia may have captured higher levels of genetic diversity because individuals are drawn from multiple sampling sites in contrast to the single sampling location in both NA and UK (Table S1). However, previous genomic studies of starling invasions indicate panmixia in the NA population (Hofmeister et al. 2021) and moderate population structure in the AU population (Stuart et al. 2020). Sampling across a similar geographic gradient in North America may reveal higher within-population diversity than recovered here, although previous studies of genome-wide variation across the North American range find similarly low levels of differentiation and diversity (Hofmeister et al. 2021). Second,

rare variants in the native range may have 'surfed' to a higher frequency in the invasions, and our sampling of the native range (in Northumberland, UK) likely represents only a small portion of genetic variants that could have been introduced in AU and NA. Finally, we also note that the UK individuals sampled here do not capture range-wide diversity in the native population, and therefore we expect that actual nucleotide diversity in the UK population is higher than what we have sampled here. Additional investigation of native range starlings will be needed to determine whether lower diversity in the UK is recovered with broader geographic sampling. Accurate genetic estimates of each population rest on an assumption of ongoing gene flow across each continent. Broader geographic sampling could show that the estimates of diversity presented here are limited to the sub-populations sampled, but results based on the sampling in this study largely concur with the estimates of genetic variation drawn from wider geographic but shallower depth sequencing in each invasion.

This evolutionary genetic study sought to document genetic variation in two invasions, and we note that even with similar invasion pathways in both expected source population and timing of introduction, other factors—including differences in propagule pressure and dispersal—likely influence the evolutionary trajectory of each invasive population (Simberloff 2009; Williams et al. 2019). Propagule pressure (also termed introduction effort) is a composite measure of the number of individuals released, and we note that founding population sizes may have varied slightly among AU and NA introductions; the exact size of each founding population could be identified only through historical documentation. In addition to the unknown founding population sizes, changes in dispersal itself likely shaped genetic diversity and thus

adaptive potential, as shown in a recent study of invasive plants (Smith et al. 2020). Starlings' dispersal and migration varies among populations: starlings in Australia are not known to undertake seasonal migration and typically move across shorter distances, but starlings migrate and/or disperse hundreds of kilometers in North America (Kessel 1953; Burtt and Giltz 1977) and South Africa (Berthouly-Salazar et al. 2013). Future work should consider local ecological conditions to elucidate how gene flow at various geographical scales impacts invasion success.

It is remarkable that despite this contrast in life history strategies and the stochastic nature of evolution during range expansion, we find F_{ST} peaks shared among invasions at only a few regions of the starling genome. Although these F_{ST} peaks could arise via drift, footprints of other population genetic metrics are consistent with selection. Furthermore, a recent comparison of methods for detecting selective sweeps in rapidly evolving systems found that F_{ST}-based scans perform nearly as well as scans based on beta diversity (Schneider et al. 2021). Another recent study found evidence for parallel selection between native and invasive starling populations, and this historical sampling suggests that the signatures of rapid evolution identified in this contemporary may indeed result from selection regimes during invasion (Stuart et al. 2022a). We note that mutations in specific chromosomal regions could also be accelerated by extrinsic environmental properties (climate, food availability, and more) through epigenetic CpG DNA hypermethylation events, which are known to increase frequency of genetic mutations by spontaneous deamination CG>TG transition (Sved and Bird 1990; Saxonov et al. 2006; Simmen 2008). For example, such epigenetic shifts supported the invasion of another avian species (the house sparrow) into Australia (Sheldon et al.

116

2018). Recent work in the Australian invasion also suggests that some ecogeographical patterns result from non-genetic factors (Stuart et al. 2022b).

Regardless of genetic mechanism, we suggest that differentiation in these genetic regions is simultaneously shaped by intrinsic and extrinsic drivers of invasion success. We find it notable that some differentiated regions (in particular, Chromosome 6) are shared among invasions despite differences in the selective environment as well as stochastic processes that shape the starling's evolution on each continent. The European starling invasions compared here suggest that rapid population growth may support rapid and potentially adaptive evolution.

REFERENCES

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:1–6.

Ball GF, Wingfield JC. 1987. Changes in plasma levels of luteinizing hormone and sex steroid hormones in relation to multiple-broodedness and nest-site density in male starlings. Physiological Zoology 60:191–199.

Berthouly-Salazar C, Hui C, Blackburn TM, Gaboriaud C, van Rensburg BJ, van Vuuren BJ, Le Roux JJ. 2013. Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. Mol Ecol 22:5793–5804.

Blakeslee AMH, Manousaki T, Vasileiadou K, Tepolt CK. 2020. An evolutionary perspective on marine invasions. Evolutionary Applications 13:479–485.

Bock DG, Kantar MB, Caseys C, Matthey-Doret R, Rieseberg LH. 2018. Evolution of invasiveness by genetic accommodation. Nat. ecol. evol. 2:991–999.

Bodt LH, Rollins LA, Zichello JM. 2020. Contrasting mitochondrial diversity of European starlings (Sturnus vulgaris) across three invasive continental distributions. Ecol Evol 10:10186–10195.

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

Burtt HE, Giltz ML. 1977. Seasonal directional patterns of movements and migrations of starlings. Bird-Banding 48:259–271.

Cardilini APA, Buchanan KL, Sherman CDH, Cassey P, Symonds MRE. 2016. Tests of ecogeographical relationships in a non-native species: what rules avian morphology? Oecologia 181:783–793.

Cardilini APA, Micallef S, Bishop VR, Sherman CDH, Meddle SL, Buchanan KL. 2018. Environmental Influences on Neuromorphology in the Non-Native Starling *Sturnus vulgaris*. Brain Behav Evol 92:63–70.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133–3157.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools.

Bioinformatics 27:2156–2158.

Davis MA. 2020. Let's welcome a variety of voices to invasion biology. Conservation Biology:cobi.13608–6.

Dawson A. Plasma gonadal steroid levels in wild starlings (*Sturnus vulgaris*) during the annual cycle and in relation to the stages of breeding. General and Comparative Endocrinology 49:286–294.

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

Dlugosch KM, Parker IM. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431–449.

Dolbeer RA. Migration patterns of age and sex classes of Blackbirds and Starlings. Journal of Field Ornithology 53:28–46.

Dormontt EE, Lowe AJ, Prentis PJ. 2011. Is rapid adaptive evolution important in successful invasions? *in* D. M. Richardson, ed. Fifty Years of Invasion Ecology.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust Demographic Inference from Genomic and SNP Data.Akey JM, editor. PLoS Genet 9:e1003905–e1003917.

Feare C. 1984. The Starling. Oxford: Oxford University Press

Forbush EH. 1915. The Starling. Boston: Wright and Potter Printing Co.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30:1486–1487.

Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26:1145–1151.

Gralka M, Stiewe F, Farrell F, Möbius W, Waclaw B, Hallatschek O. 2016. Allele surfing promotes microbial adaptation from standing variation. Ecology Letters 19:889–898.

Higgins PJ, Peter JM, Cowling SJ. 2006. Handbook of Australian, New Zealand & Antarctic birds. Volume 7, Boatbill to starlings. Melbourne: Oxford University Press.

Hofmeister NR, Werner SJ, Lovette IJ. 2021. Environmental correlates of genetic

variation in the invasive European starling in North America. Mol Ecol:mec.15806–mec.15833.

Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. Mol Ecol 511:83–17.

Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, et al. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology, 23:4035–4058.

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

Kim SY, Lohmueller KE, Albrechtsen A, Li Y, Korneliussen T, Tian G, Grarup N, Jiang T, Andersen G, Witte D, et al. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics 12:231.

Knief U, Forstmeier W. 2016. Mapping centromeres of microchromosomes in the zebra finch (*Taeniopygia guttata*) using half-tetrad analysis. Chromosoma:1–12.

Kolbe JJ, Glor RE, Rodriguez Schettino L, Chamizo Lara A, Larson A, Losos JB. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431:177–181.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

Linz GM, Homan HJ, Gaulker SM, Penry LB, Bleier WJ. 2007. European starlings: a review of an invasive species with far-reaching impacts. Available from: http://digitalcommons.unl.edu/nwrcinvasive/24

Lowe WH, Kovach RP, Allendorf FW. 2017. Population genetics and demography unite ecology and evolution. Trends in Ecology & Evolution 32:141–152.

Lu-Irving P, Marx HE, Dlugosch KM. 2018. ScienceDirect Leveraging contemporary species introductions to test phylogenetic hypotheses of trait evolution. Current Opinion in Plant Biology 42:95–102.

Marques DA, Jones FC, Di Palma F, Kingsley DM, Reimchen TE. 2018. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation.

Nature Ecol. Evol.:1–14.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303.

Miller TEX, Angert AL, Brown CD, Lee-Yaw JA, Lewis M, Lutscher F, Marculis NG, Melbourne BA, Shaw AK, Szúcs M, O. Tabares O, Usui T, Weiss-Lehman C, Williams JL. 2020. Eco-evolutionary dynamics of range expansion. Ecology 101(10):e03139.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data.Awadalla P, editor. PLoS ONE 7:e37558–11.

Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics:btv566–10.

Redding DW, Pigot AL, Dyer EE, Şekercioğlu ÇH, Kark S, Blackburn TM. 2019. Location-level processes drive the establishment of alien bird populations worldwide. Nature 571:103-106.

Ricciardi A, Ryan R. 2017. The exponential growth of invasive species denialism. Biological Invasions 20:549–553.

Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB. 2009. Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. Mol Ecol 18:1560–1573.

Russell JC, Blackburn TM. 2017. The Rise of Invasive Species Denialism. Trends in Ecology & Evolution 32:3–6.

Sagoff M. 2019. Fact and value in invasion biology. Conservation Biology 34:581–588.

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA 103:1412–1417.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes:1–7.

Schneider K, White TJ, Mitchell S, Adams CE, Reeve R, Elmer KR. 2021. The pitfalls and virtues of population genetic summary statistics: Detecting selective sweeps in

recent divergences. J Evol Biol 34:893-909.

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

Simberloff D. 2009. The Role of Propagule Pressure in Biological Invasions. Annu. Rev. Ecol. Evol. Syst. 40:81–102.

Simmen M. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. Genomics 92:33–40.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. Genetics 195:693–702.

Smith AL, Hodkinson TR, Villellas J, Catford JA, Csergő AM, Blomberg SP, Crone EE, Ehrlén J, Garcia MB, Laine A-L, et al. 2020. Global gene flow releases invasive plants from environmental constraints on genetic diversity. Proc Natl Acad Sci USA 117:4218–4227.

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, et al. 2010. Adaptation genomics: the next generation. Trends Ecol Evol 25:705–712.

Stuart KC, Cardilini APA, Cassey P, Richardson MF, Sherwin W, Rollins LA, Sherman CDH. 2020. Signatures of selection in a recent invasion reveals adaptive divergence in a highly vagile invasive species. Mol Ecol [Internet]. Available from: https://doi.org/10.1111/mec.15601

Stuart KC, Edwards RJ, Cheng Y, Warren WC, Burt DW, Sherwin WB, Hofmeister NR, Werner SJ, Ball GF, Bateson M, et al. 2021. Transcript- and annotation-guided genome assembly of the European starling. bioRxiv.

Stuart KC, Sherwin WB, Austin JJ, Bateson M, Eens M, Brandley MC, Rollins LA. 2022. Historical museum samples enable the examination of divergent and parallel evolution during invasion. Mol Ecol DOI: 10.1111/mec.16353.

Stuart KC, Sherwin WB, Cardilini APA, Rollins LA. 2022. Genetics and plasticity are responsible for climate induced ecogeographical patterns in a recent invasion. Frontiers in Genetics. DOI: 10.3389/fgene.2022.824424.

Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE 6:e21800.

Sved J, Bird A. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc Natl Acad Sci USA 87:4692–4696.

Tsutsui ND, Suarez AV, Holway DA, Case TJ. 2000. Reduced genetic variation and the success of an invasive species. Proc Natl Acad Sci USA 97:5948–5953.

Verhoeven KJF, Macel M, Wolfe LM, Biere A. 2011. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. Proceedings of the Royal Society B: Biological Sciences 278:2–8.

Waterman M, Fuller C, Murray MD. 2008. Studies of roosting common starlings *Sturnus vulgaris* in South Australia. Corella 32:25-29.

Werner SJ, Fischer JW, Hobson KA. 2020. Multi-isotopic (δ2H, δ13C, δ15N) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. PLoS ONE 15:e0237137.

White TA, Perkins SE, Heckel G, Searle JB. 2013. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Mol Ecol 22:2971–2985.

Williams JL, Hufbauer RA, Miller TEX. 2019. How Evolution Modifies the Variability of Range Expansion. Trends in Ecology & Evolution:1–11.

Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR. 2018. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. Mol Ecol 17:675.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328.

CHAPTER 4: SIMPLE CHOICES: TOWARDS A 'BETTER' INVASION BIOLOGY

ABSTRACT

Invasion success depends on the precarious coordination of ecological and evolutionary mechanisms, and only interdependence of these mechanisms enables a species to establish and spread in a new environment. Genomic sequencing and other readily-available analytical tools can tempt invasion biologists to simplify a dynamic system, even though no one trait or strategy facilitates invasion success. In this manuscript we consider adjustments to hypothesis framing and model specification in invasion genetics using a case study of the European starling. When biologists focus on isolating causal factors to identify a biological truth, we rely on simplifications that can be both scientifically inaccurate and socially dangerous. Here we focus on three simplifications of our own: (I) establishing a species as invasive relative to ancestral and/or resident lineages, (II) locating causality in genetic variants, and finally (III) identifying conditions that support an assemblage's continued survival and reproduction. In all three areas of study, careful hypotheses, models, and narratives could move invasion biology away from a fruitless search for intrinsic invasiveness and towards an accurate model of complexity.

INTRODUCTION

Invasion biology relies on binary framings: alien transplant vs. native victim, species invasiveness vs. community invasibility, intrinsic properties vs. extrinsic factors, and so on. Reducing complexity into two opposing forces manufactures urgency that, in turn, funds research and management of invasive species (Colautti and Richardson 2008; Simberloff 2009; Valéry et al. 2013; Subramaniam 2014a). Binarism is hardly unique to invasion biology: it has notoriously structured discourse around sex and gender (Fausto-Sterling 2012; Monk et al. 2019) and regularly influences conservation initiatives (Brister et al. 2021). At present, invasion biologists seek to determine whether invasion success is best explained by extrinsic factors and/or by some intrinsic property of the organism. This tension between two explanations mirrors a conflict among evolutionary biologists and ecologists, but also locates causality in an essential property of a species or a dynamic biological system. Few if any invasions' success can be attributed to a single actor; given this empirical evidence, contemporary invasion theory may benefit from greater precision in methods and narratives of invasion.

How to define and respond to invasive species has social and material consequences: shifting political tides have reinvigorated the decades-long debate. Scholars have long described parallels in language between human and non-human immigrants (Subramaniam 2001; reviewed in Coates 2006). In isolation, neither rhetorical shifts (Colautti and Richardson 2008; Simberloff 2012) nor philosophical reframings (Hattingh 2011; Guiaşu and Tindale 2018; Frank et al. 2019) can ease the underlying material conflict. Narratives about invasive species like the Asian carp can and do impact

conservation actions (Mando and Stack 2019). The debates around rhetoric rage, with continued calls to replace the term "invasive," "non-native," and "foreign" with "potential problem species" (Inglis 2020). Identifying the narratives and philosophies underlying invasion theory brings us closer to understanding the conflict (Hattingh 2011; Simberloff 2012).

Invasion biologists recognize a pressing need for more integrative hypotheses and interpretations of invasion success (Heger et al. 2013; Crystal-Ornelas and Lockwood 2020; Fristoe et al. 2021). Conservation actions depend on accurate predictions of which species will become invasive, a project that demands more than just synthesis of scientific evidence (Elliott-Graves 2015). The science of invasion is marked by "theoretical exhaustions and empirical recapitulations that […] characterized the study of human redistributions of biota for over two centuries" (Chew 2006). On a practical level, many stakeholders believe that controlling the spread of invasive species is a urgent crisis that demands a response (Hobbs et al. 2011; Funk et al. 2020). However, practitioners' beliefs about how to describe and respond to invasive species vary dramatically (Humair et al. 2014; Kapitza et al. 2019; Shackleton et al. 2019; Gbedomon et al. 2020). Tumult within and around the discipline shows no signs of abating, and we offer yet another interpretation of the invasive species debate, this time centering on the power—and problem—of binary framing.

This paper focuses on what work isolating a species and/or positioning it in opposition to an other accomplishes for scientific and/or social causes. We critique our work on the European starling invasion to address biases embedded in our hypotheses, methods, and interpretation of invasion success. Making sense of a system in which the

assemblage of actors changes over space and time demands some simplifications, and we organize this manuscript into three simplifications of our own: (I) establishing a species as invasive relative to ancestral and/or resident lineages, (II) locating causality in genetic variants as a simplifying tool, and finally (III) identifying conditions that support an assemblage's continued survival and reproduction. In each simplification, we explore scientific challenges and their social implications, orienting this project towards a model that approximates the complexity of invasion.

1: DEFINING THE EUROPEAN STARLING AS AN INVASIVE SPECIES

Human values as a foundation in starling invasions

Much debate has centered on whether or not invasion biology is fundamentally value-oriented (Young and Larson 2011; Simberloff and Vitule 2013; Ricciardi et al. 2017) In the case of the starling in North America, its arrival is a direct consequence of European immigrants' desire to naturalize European flora and fauna in any region they colonize. Starlings may not have arrived on this continent if not for human desires and/or management. Much of the research on the starling only briefly mentions that humans facilitated starling invasions into North America (Forbush 1915), Australia (Rollins et al. 2009), New Zealand (Ross 1983), South Africa (Berthouly-Salazar et al. 2013), and most recently South America (Fiorini et al. 2021)—an irony considering the near-universal vitriol that conservationists launch against the bird today. It is no coincidence that starlings invaded these regions—mostly white settler colonies—given that merchants and immigrants likely carried the birds with them (Feare 1984). Despite a long and well-

documented history of human complicity, the European starling has been labeled one of the world's worst invaders (Lowe et al. 2000); a quick search online yields headlines such as "Get the flock out of here: Starlings are the worst. Birds. Ever" (Stopyra 2016). This attitude may be explained by starlings' dramatic agricultural and economic impacts (Linz et al. 2018). Due to these impacts, each year an estimated 1-3 million birds are killed in cattle feedlots and another 80-100 million birds die of natural causes (Linz et al. 2018). Human interference has shaped the starling's ecology and evolution, which demands that we acknowledge such interference in our hypotheses, methods, and interpretations in any studies of starling invasions.

Comparative methods draw a boundary

A successful invader is defined in opposition to the residents of both its new and ancestral communities (i.e., native species in either region). Biologists have long recognized that species delimitation is a moving target, in part due to changing spatial ranges (Sexton et al. 2009) and genetic exchange via hybridization (Pfennig et al. 2016). Even isolating the physical factors that constitute a species boundary is a major challenge: as Pfennig et al. point out, "hybridization is often an outcome of range expansion" (Pfennig et al. 2016). The boundary between invasive and native species is fixed only for the moment of observation, a fact that may explain why quantitative data cannot assuage fears of losing native diversity.

Evolutionary biologists might define *Sturnus vulgaris* as an invasive species by comparing *S. vulgaris* to its closest relative (spotless starling; *Sturnus unicolor*). The dramatic range expansion of *Sturnus vulgaris* is unique among the Sturnidae family

128

(Lovette et al. 2008); only in the last six decades has its sister lineage Sturnus unicolor expanded from its range in parts of the Iberian Peninsula and northern Africa. We might hypothesize that S. vulgaris' range can expand only to fill its realized niche, where competitors limit further expansion (Hutchinson 1957), a hypothesis supported by the fact that expansion of both S. vulgaris and S. unicolor slowed when the species came into contact (Ferrer et al. 1991). This evidence is not limited to the native range: ornithologists have conjectured that the European starling did not successfully establish when released in Portland, OR in part due to competition with another introduced relative, the Crested Myna (previously Aethiopsar, now Acridotheres cristatellus) (Wood 1924).

Given this evidence of competition between some starling species, we might test whether S. vulgaris evolved to outcompete its sister species, borrowing from the classic idea that competition among species reinforced a species' range limits and genetic boundaries (Mayr 1942). Although a species' environment itself changes over its evolutionary history, we biologists typically observe just one site in time and space (Siepielski et al. 2009). Each frame in time gives us a clearer image of the starling's spread in the last two centuries; we isolate that frame for ease of analysis, but that snapshot represents only the current ecological consequences (Závorka et al. 2018). Biogeographic and genetic studies of starlings suggest that the hypothesis based in competition might mislead. Genetic evidence shows that S. vulgaris and S. unicolor interbreed easily: allozyme studies show that S. unicolor populations are in fact more genetically distant from each other (Nei's genetic distance = 0.005) than from S. vulgaris (0.001-0.004) (la Cruz Cardiel et al. 1997). Further supporting the point that species

boundaries are constantly shifting, Feare (1984) suggests that S. unicolor may have only split from the ancestral lineage of both species 40 Mya, and as the landscape shifted post-glaciation, these lineages may have reintegrated as they came into contact naturally. How biologists present hybridization depends on how they perceive risk of a species' erasure.

Hybridization and genomic introgression certainly can impact invasive and native species, but whether and how these forces post a threat to diversity is an open question (Hirashiki et al. 2021). Introgression and/or admixture may pose risks to some native species (Quilodrán et al. 2018; Barker et al. 2019), but Hirashiki et al. (2021) show that only 35 of 870 invasive species studied may impact native taxa due to hybridization. This scientific evidence makes clear that hybridization may not pose as great a threat as previously assumed. At least one major fear about invasive species may thus relate more to social attitudes than to scientific evidence; for a detailed review of this topic, Banu Subramaniam's critiques thoroughly describe how values shape interpretations of interbreeding, even within the biological community (Subramaniam 2001; 2014a).

2: GENES AND ALLELES AS THE SITE OF INVESTIGATION

The rising popularity of genomic methods

Evolutionary genetic studies of invasion success are increasingly popular, perhaps in part because genomic data are more accessible and models that deploy such data are more advanced (Hoban et al. 2016; Blakeslee et al. 2020; North et al. 2021) However,

more so than ecological studies focused on interactions, evolutionary and genetic research tends to search for an intrinsic   property to explain a species' invasion success (Dormontt et al. 2011). At the same time, evolutionary genetic studies are often justified and funded by referencing ecological evidence that the invasive population harms some native community, reminding us that careful supporting work in ecology and molecular biology is essential . Cross-disciplinary work beyond the natural sciences shows that the beliefs embedded in our science might hinder even a management-focused project; a recent example of conservationists and social scientists working together on global orangutang conservation could inspire similar projects in other systems (Chua et al. 2020). Many of our stories assign blame to a single actor, rather than a dynamic orchestration of many interactions, and we argue that locating causal action in a genetic variant is an appealing but dangerous project.

Describing genes as actors has a long history in adaptationism

Focusing on genes and alleles as the locus of change simplifies our studies, but this analytical convenience cannot metamorphose into a biological truth. An evolutionary purist might consider an individual as a repository of genetic variants, where the body simply mediates between gene and environment. Scaling from gene to individual, individual to population, and population to species is another boundary problem, where the unit we observe is a construct of convenience. In this logic, genomic studies are tempting because they scale down to the causal unit: even in our own work, we conveniently scale from putatively adaptive allele to invasive species to create a more powerful story (Hofmeister et al. 2021). Alleles are not agents, which explains why

131

evolutionary biologists focus their stories on an intelligible subject, a visible body. Analyzing genes alone cannot explain why some species spread more easily or reproduce more readily. This point seems banal: any scientist knows that not all simplifications are equally good, and bad science can lead to poor social outcomes. Scientific integrity alone demands that we functionally validate putatively adaptive alleles, and—conveniently for us—this work aligns with more careful narratives about causality, as we illustrate in the next paragraph.

Most invasions are evolutionarily young, so any evidence of adaptation will often come from subtle shifts in allele frequencies. Demographic processes—specifically, genetic bottlenecks and/or founder events—shape the genetic variation available for other genetic mechanisms of evolution. Even if the founding population is subject to a genetic bottleneck, rapid expansion can counteract diversity loss (Birzu et al. 2019), and populations may even adapt (Dlugosch and Parker 2008; Facon et al. 2011), perhaps via inbreeding-by-environment interactions (Schrieber and Lachmuth 2017). Genetic variation available for selection may come from standing variation, novel mutation, or introgression (Visscher et al. 2008), and the source largely determines how we assign causality in invasion success. Adaptive variants are often rare in the native range, as shown in genomic studies of starling invasions (Hofmeister et al. 2021). In these cases, it is incredibly challenging to distinguish between selective sweeps, genetic bottlenecks, and recombination: all of these processes could yield variants at low-to-intermediate frequency surrounded by runs of homozygosity (Cruickshank and Hahn 2014). Models such as S/HIC and diploS/HIC may help to disentangle these processes via explicitly accounting for demographic history (Schrider and Kern 2016; Kern and

Schrider 2018). However, rapid expansions in population size can also lead to accumulation of novel mutations during expansion, commonly referred to as expansion load (Kirkpatrick and Barton 1997; Hallatschek and Nelson 2010; Peischl et al. 2015). Determining how expansion load impacts a population's viability remains an open question (Gilbert et al. 2017). Whether new mutations or ancestral alleles, adaptive variants may surf on the expanding range edge, further complicating the evolutionary dynamics of range expansion and/or adaptation.

Conditions that support range expansion are the same conditions that support an individual's reproduction and survival: these can be biotic or abiotic (Sexton et al. 2009), and also fluctuating or constant (Martín et al. 2019). Biologists often label these conditions as selective pressures, which are not biologically real phenomena but only metaphorical simplifications that ascribe agency to an inanimate environment (Coyne 2013). Adaptationist critique is best summarized in the classic Spandrels paper and surrounding literature (Gould and Lewontin 1979; Pigliucci and Kaplan 2000). Finding evidence of adaptation remains an enticing project, where the logic implicates some intrinsic, evolved strategies unique to a species like S. vulgaris, but many species thrive even without clear support for Darwinian natural selection on alleles. Despite centuries of work, the question of where and how evolution takes place remains an analytical and rhetorical puzzle (Laland et al. 2014). Here we exhume the extended evolutionary synthesis only to highlight that defining the actors in a system has never been a purely biological project. Stories about adaptation often lose precision, but in doing so, they incorrectly assign agency to a single actor.

3: MULTIPLE ACTORS COOPERATE TO YIELD AN INVADER'S SUCCESS

Bird-watchers' and ornithologists' records expose a close relationship between starlings and humans at nearly every stage of the population's expansion. Wealthy businessman and supposed Shakespeare enthusiast Eugene Schieffelin released European starlings into Central Park, NY, USA. Schieffelin released 80 pairs in 1891 and 40 in 1891; the population was truly established in 1895 (Forbush 1915; Cooke 1928). Some scientific and public narratives characterize Schieffelin as both an outsider (Carroll 2021) and the sole human culprit in driving starling expansions. However, the world-wide success of starling invasions even without assistance from acclimatization societies (e.g., in South Africa and South America) suggests that humans can play multiple roles in facilitating invasions. In the following section, we describe how the availability of human-engineered habitats—often agricultural land—influenced starling reproduction and thus population growth and decline. In making visible the multiple, cooperating architects of the starling invasion, we destabilize a narrative that may otherwise focus on intrinsic properties of this invasive species.

Starlings may produce more young where humans maintain suitable habitat
A commonly-held belief about starlings is that they adjust foraging strategy in response to readily-available food resources, facilitating their rapid expansion, but even this belief demands a few caveats. Stable foraging behaviors indicate a relationship between the starling and its food source, and in both the European and North American ranges, starlings prefer to forage in grasslands and pastures where soil invertebrates abound

(Devereux et al. 2006; Purcell 2015). However, in the last several decades, starlings have shifted from foraging on grasslands maintained for cattle grazing to gorging on the stockpiles of grain in large cattle feedlots now so common in North America (Linz et al. 2018). Starlings are touted as relatively intelligent birds that can acclimate and even adapt to novel conditions (Rodriguez et al. 2010), but we can just as easily blame humans for providing such an easy food source as we blame the birds who partake. Another line of evidence linking human management to starling population growth is the recent decline in population size: the North American population is now estimated at about 85.1 million birds, a 49% decline from the estimated 1970 population (Rosenberg et al. 2019). In the native range, starling populations are declining where pasture area has declined (Freeman et al. 2007; Heldbjerg et al. 2016). This evidence becomes especially critical when we note that starlings feed almost exclusively on these invertebrates during the breeding season (Feare 1984). Connecting these pieces of evidence yields a hypothesis that population size in starlings relates to the extent to which human-managed agriculture dominates the landscape, but simplifying starlings to focus only on foraging behavior overlooks changes in dispersal and migratory behavior that could also support invasion success.

Starling populations expand when individuals disperse or migrate due to environmental conditions

A historical reconstruction suggests that the starling population expanded based on some complex interaction of population size and environmental conditions (two variables that are themselves inter-related). Like most invasive species, the starling's

135

population grew dramatically in the center of its range (e.g., New York and Pennsylvania) (Wood 1924; Wing 1943), which suggests that starling expansion accelerated only after range-edge populations grew large enough to trigger dispersal. In other words, the starling's expansion may have functioned as a pushed wave, where individuals from the center of the range disperse due to density dependence (Gandhi et al. 2016; Deforet et al. 2019), especially if the original founding population depended on suitable habitat as described in the previous section (Dahirel et al. 2021). When populations are especially dense or breeding conditions are ideal that year, invasive starlings may lay more clutches than native birds (Dawson 1983; Ball and Wingfield 1987), such that individuals adjust their reproductive strategy and the population grows more quickly. Dispersal and migration are other life-history traits that can evolve when populations are especially dense (Miller et al. 2020) and facilitate range expansion. For many ecologists, the goal is to determine which of these life-history strategies (among migration, dispersal, and breeding behavior) best explain the starling's range expansion, but there is yet another layer of complexity to introduce.

Starlings expanded across a changing landscape: as we described above, the birds thrive in agricultural lands, and thus land use may best explain starling expansion. However, American ornithologists generally concurred that the starling's spread was limited by altitude, since expansion slowed near mountains, and European birdwatchers noted that the starling avoided mountains in the native range (Wood 1924; Wing 1943). Starling expansion slowed after each wave: after establishing in Philadelphia in 1910, after crossing the Allegheny Mountains in 1916, up to elevational limits in New York and Vermont in 1922 (Forbush 1915; Kalmbach and Gabrielson 1921;

Cooke 1928), and again when it reached the 1000m elevation mark by 1930 (Hoffman 1930; Dickerson 1937). In addition to expanding after crossing altitudinal thresholds, starlings may have expanded easily where they remain near large water bodies (based on their migration habits along Western range edge at the Mississippi River, and northern edge near Lake Ontario (Cooke 1928). This may be explained by the fact that starlings migrate along major water bodies, like many other birds (Kessel 1953), and these flyways also shape dispersal patterns (Cabe 1999; Werner et al. 2020). Elevational limits and water bodies may explain starling expansion, but farmland itself often abuts large bodies of water or mountain ranges. As biologists, we may seek to separate agricultural from climatic boundaries using statistical models to assign variable importance, but such a project has social implications.

Scientific accuracy guides us towards responsible action
As biologists, we are trained to organize actors—in this example, a bird and its surroundings—into the simplest, most accurate representation of reality possible. In much of this paper, but especially this section, we have relied on simplifications into two options: adaptation comes from new mutation or standing variation; starlings survive in a suitable climate or reproduce when they can access pasture. These oppositional hypotheses sanction the binary frame as not only a scientific tool but a biological truth. Some invasion ecologists have suggested a 'hierarchy of hypotheses' for understanding invasion success as a simplifying tool (Heger et al. 2013); this strategy serves simply to group connected hypotheses, rather than to position some hypotheses above others in importance. Organizing scientific knowledge creates an

easy entry point, but how we structure theories and models has social implications.

Binaries and hierarchies simplify biological complexity into a model we can understand, but also support political claims by naturalizing a discourse of "pure" natives vs. "aggressive" immigrant and/or invader. We take seriously the ongoing conflation of human immigrant and invasive species in public discourse (Subramaniam 2014b; Gardiner 2020). Public perceptions, in addition to the heated debate within invasion science (Simberloff and Vitule 2013; Crowley et al. 2017; Russell and Blackburn 2017), provide indirect evidence that nativism and xenophobia linger (see the first section in this manuscript for narratives about starlings). Whether or not nativist attitudes historically shaped invasion science (Coates 2006; Cardozo and Subramaniam 2013), such biases may still creep into scientific assumptions and interpretations, and we must be way of naturalizing the social assumptions built into our science.

Even recent advances in framing studies in invasion science rely on a binary simplification; our analysis of these frames is inspired by Karen Barad's reflections on scientific observation (Barad 2014). Leading invasion biologists reimagine analytical frames in a special issue on "Human influences on evolution, and the ecological and societal consequences" thus: "invasions and extinctions of closely related species may be like reflections in Lewis Carroll's looking glass [3], with similar elements reflecting opposite realities" (Colautti et al. 2017). Colautti et al. make the "looking glass" of eco-evolutionary theory visible, and also note that these "opposite realities" are in fact a theoretical simplification of how species respond to the same ecological and genetic factors. Naming such frames and simplifications ensures that we do not lose sight of the social reality of doing science, where our stories can be utilized and even weaponized

138

by the audience.

CONCLUSION

Reconstructing the starling expansion illustrates how challenging it is to isolate factors that might explain the starling's invasion success. Scientific evidence for invasion success alone demands that we not rely on simplifications (especially into dichotomies), but admitting complexity and uncertainty makes it harder to build the case for urgency. If we acknowledge this limitation in our science, we risk the public losing confidence in science—a major concern for scientists wanting to combat denialism in invasion biology (Ricciardi and Ryan 2017; Boltovskoy et al. 2018). However, privileging scientific—and especially genetic—evidence may distract from our shared project of valuing diversity, because such simplifications may also fuel an ideological conflict that percolates throughout broader society. Rather than pinning agency on single actors like an invasive species or even an 'adaptive' genetic variant, we argue for models that capture the complex interactions that shape an invasion's success.

REFERENCES

Ball, G. F., and J. C. Wingfield. 1987. Changes in plasma levels of luteinizing hormone and sex steroid hormones in relation to multiple-broodedness and nest-site density in male starlings. Physiological Zoology 60:191–199.

Barad, K. 2014. Diffracting diffraction: Cutting together-apart. Parallax 20:168–187.

Barker, B. S., J. E. Cocio, S. R. Anderson, J. E. Braasch, F. A. Cang, H. D. Gillette, and K. M. Dlugosch. 2019. Potential limits to the benefits of admixture during biological invasion. Molecular Ecology 28:100–113.

Berthouly-Salazar, C., Hui, C., Blackburn, T. M., Gaboriaud, C., van Rensburg, B. J., van Vuuren, B. J., & Le Roux, J. J. 2013. Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. Molecular Ecology 22:5793–5804.

Birzu, G., S. Matin, O. Hallatschek, and K. Korolev. 2019. Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. Ecology Letters 22:1817–1827.

Blakeslee, A. M. H., T. Manousaki, K. Vasileiadou, and C. K. Tepolt. 2020. An evolutionary perspective on marine invasions. Evolutionary Applications 13:479–485.

Boltovskoy, D., F. Sylvester, and E. M. Paolucci. 2018. Invasive species denialism: Sorting out facts, beliefs, and definitions. Ecology and Evolution 8:11190–11198.

Brister, E., J. B. Holbrook, and M. J. Palmer. 2021. Conservation science and the ethos of restraint. Conservation Science and Practice 3:e381.

Cabe, P. R. 1999. Dispersal and population structure in the European Starling. The Condor 101:451–454.

Cardozo, K., and B. Subramaniam. 2013. Assembling Asian/American Naturecultures: Orientalism and Invited Invasions. Journal of Asian American Studies 16:1–23.

Carroll, A. 2021. Fair is Fowl and Fowl is Fair: Shakespeare's Invasive Birds in North America. University of Minnesota Morris: Honors Capstone Project.

Chew, M. K. 2006. Ending with Elton. (M. D. Laubichler, ed.).

Chua, L., M. E. Harrison, H. Fair, S. Milne, A. Palmer, J. Rubis, P. Thung, et al. 2020.

Conservation and the social sciences: Beyond critique and co-optation. A case study from orangutan conservation. People and Nature 2:42–60.

Coates, P. 2006. American Perceptions of Immigrant and Invasive Species. University of California Press, Berkeley.

Colautti, R. I., and D. M. Richardson. 2008. Subjectivity and flexibility in invasion terminology: too much of a good thing? Biological Invasions 11:1225–1229.

Colautti, R. I., J. Alexander, K. M. Dlugosch, S. R. Keller, and S. E. Sultan. 2017. Invasions and extinctions through the looking glass of evolutionary ecology. Philosophical Transactions of the Royal Society B: Biological Sciences.

Cooke, M. T. 1928. The spread of the European starling in North America (to 1928) 40:1–11.

Coyne, J. 2013. "Selection pressures" are metaphors. So are the 'laws of physics.'. Why Evolution Is True.

Crowley, S. L., S. Hinchliffe, S. M. Redpath, and R. A. McDonald. 2017. Disagreement About Invasive Species Does Not Equate to Denialism: A Response to Russell and Blackburn. Trends in Ecology & Evolution 32:228–229.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133–3157.

Crystal-Ornelas, R., and J. L. Lockwood. 2020. The "known unknowns" of invasive species impact measurement. Biological Invasions 22:1513–1525.

Dahirel, M., A. Bertin, M. Haond, A. Blin, E. Lombaert, V. Calcagno, S. Fellous, et al. 2021. Shifts from pulled to pushed range expansions caused by reduction of landscape connectivity. Oikos 130:708–724.

Dawson, A. 1983. Plasma gonadal steroid levels in wild starlings (*Sturnus vulgaris*) during the annual cycle and in relation to the stages of breeding. General and Comparative Endocrinology 49:286–294.

Deforet, M., C. Carmona-Fontaine, K. S. Korolev, and J. B. Xavier. 2019. Evolution at the edge of expanding populations. The American Naturalist 194:291–305.

Devereux, C. L., M. J. Whittingham, J. R. Krebs, E. F. Juricic, and J. A. Vickery. 2006. What attracts birds to newly mown pasture? Decoupling the action of mowing from the provision of short swards. Ibis 148:302–306.

Dickerson, L. M. 1937. The Western Frontier of the European Starling in the United States as of February, 1937. The Condor 40:118–123.

Dlugosch, K. M., and I. M. Parker. 2008. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. Ecology Letters 11:701–709.

Dlugosch, K. M., S. R. Anderson, J. Braasch, F. A. Cang, and H. D. Gillette. 2015. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. Molecular Ecology 24:2095–2111.

Dormontt, E. E., A. J. Lowe, and P. J. Prentis. 2011. Is rapid adaptive evolution important in successful invasions? in D. M. Richardson, ed. Fifty Years of Invasion Ecology.

Elliott-Graves, A. 2015. The problem of prediction in invasion biology. Biology & Philosophy 31:373–393.

Estoup, A., V. Ravigné, R. Hufbauer, R. Vitalis, M. Gautier, and B. Facon. 2016. Is There a Genetic Paradox of Biological Invasion? Annual Review of Ecology, Evolution, and Systematics 47:51–72.

Facon, B., R. A. Hufbauer, A. Tayeh, A. Loiseau, E. Lombaert, R. Vitalis, T. Guillemaud, et al. 2011. Inbreeding Depression Is Purged in the Invasive Insect Harmonia axyridis. Current Biology 21:424–427.

Fausto-Sterling, A. 2012. Sex/gender. Routledge, New York.

Feare, C. 1984. The Starling. Oxford University Press, Oxford.

Ferrer, X., A. Motis, and S. Peris. 1991. Changes in the breeding range of starlings in the Iberian Peninsula during the last 30 years: competition as a limiting factor. Journal of Biogeography 18:631–636.

Fiorini, V. D., M. Domínguez, J. Reboreda, and J. Swaddle. 2021. Recent invasive population of the European starling *Sturnus vulgaris* has lower genetic diversity and higher fluctuating asymmetry than primary invasive and …. Biological Invasions in review.

Forbush, E. H. 1915. The Starling. Wright and Potter Printing Co., Boston.

Frank, D. M., D. Simberloff, J. Bush, A. Chuang, and C. Leppanen. 2019. Logical fallacies and reasonable debates in invasion biology: a response to Guiaşu and Tindale. Biology & Philosophy 34:1–11.

Freeman, S. N., R. A. Robinson, J. A. Clark, B. M. Griffin, and S. Y. Adams. 2007. Changing demography and population decline in the Common Starling *Sturnus vulgaris*: a multisite approach to Integrated Population Monitoring. Ibis 149:587–596.

Fristoe, T. S., M. Chytrý, W. Dawson, F. Essl, R. Heleno, H. Kreft, N. Maurel, et al. 2021. Dimensions of invasiveness: Links between local abundance, geographic range size, and habitat breadth in Europe's alien and native floras. Proceedings of the National Academy of Sciences 118.

Funk, J. L., I. M. Parker, V. Matzek, S. L. Flory, E. T. Aschehoug, C. M. D'Antonio, W. Dawson, et al. 2020. Keys to enhancing the value of invasion ecology research for management. Biological Invasions 22:2431–2445.

Gandhi, S. R., E. A. Yurtsev, K. S. Korolev, and J. Gore. 2016. Range expansions transition from pulled to pushed waves as growth becomes more cooperative in an experimental microbial population. Proceedings of the National Academy of Sciences 113:6922–6927.

Gardiner, B. 2020. White Supremacy Goes Green. The New York Times, p. 4. New York.

Gbedomon, R. C., V. K. Salako, and M. A. Schlaepfer. 2020. Diverse views among scientists on non-native species. NeoBiota 54:49–69.

Gilbert, K. J., N. P. Sharp, A. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, et al. 2017. Local adaptation interacts with expansion load during range expansion: maladaptation reduces expansion load. The American Naturalist 189.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm a critique of the adaptationist programme. Proceedings of the Royal Society B: Biological Sciences 205:581–598.

Guiaşu, R. C., and C. W. Tindale. 2018. Logical fallacies and invasion biology. Biology & Philosophy 33:1–24.

Hallatschek, O., and D. R. Nelson. 2010. Life at the front of an expanding population. Evolution 64:193–206.

Hattingh, J. P. 2011. Conceptual Clarity, Scientific Rigour and "The Stories We are": Engaging with Two Challenges to the Objectivity of Invasion Biology. in D. M. Richardson, ed. Fifty Years of Invasion Ecology.

Heger, T., A. T. Pahl, Z. Botta-Dukát, F. Gherardi, C. Hoppe, I. Hoste, K. Jax, et al. 2013. Conceptual Frameworks and Methods for Advancing Invasion Ecology. Ambio

42:527–540.

Heldbjerg, H., A. D. Fox, G. Levin, and T. Nyegaard. 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.

Hirashiki, C., P. Kareiva, and M. Marvier. 2021. Concern over hybridization risks should not preclude conservation interventions. Conservation Science and Practice 3:e424.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, et al. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist 188:379–397.

Hobbs, R. J., L. M. Hallett, P. R. Ehrlich, and H. Mooney. 2011. Intervention ecology: applying ecological science in the twenty-first century. BioScience 61:442–450.

Hoffman, E. C. 1930. Spread of the European Starling in America. The Wilson Bulletin 42:1–2.

Hofmeister, N. R., K. Stuart, W. C. Warren, S. J. Werner, M. Bateson, G. F. Ball, K. L. Buchanan, et al. 2021. Concurrent invasions by European starlings (*Sturnus vulgaris*) suggest selection on shared genomic regions even after genetic bottlenecks. bioRxiv 2021.05.19.442026.

Humair, F., P. J. Edwards, M. Siegrist, and C. Kueffer. 2014. Understanding misunderstandings in invasion science: why experts don't agree on common concepts and risk assessments. NeoBiota 20:1–30.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symp 22:415–427.

Inglis, M. I. 2020. Wildlife Ethics and Practice: Why We Need to Change the Way We Talk About "Invasive Species." Journal of Agricultural and Environmental Ethics 33:299–313.

Kalmbach, E. R., and I. N. Gabrielson. 1921. Economic value of the starling in the United States 868:1–59.

Kapitza, K., H. von Wehrden, B. Martín-López, and H. Zimmermann. 2019. Research on the social perception of invasive species: a systematic literature review. NeoBiota 43:47–68.

Kern, A. D., and D. Schrider. 2018. diploS/HIC: an updated approach to classifying selective sweeps. Genes, Genomes, Genetics 8:1959–1970.

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

Kirkpatrick, M., and N. H. Barton. 1997. Evolution of a Species' Range. The American Naturalist 150:1–23.

la Cruz Cardiel, de, P. J., B. Deceuninck, S. J. Peris, and J. A. E. Rosselló. 1997. Allozyme polymorphism and interspecific relationships in the Common starling (*Sturnus vulgaris*) and Spotless starling (*S. unicolor*) (Aves: Sturnidae). Journal of Zoological Systematics and Evolutionary Research 35:75–79.

Laland, K., T. Uller, M. Feldman, K. Sterelny, G. B. Müller, A. Moczek, E. Jablonka, et al. 2014. Does evolutionary theory need a rethink? Nature 514:161–164.

Linz, G. M., R. J. Johnson, and J. Thiele. 2018. European starlings: a review of an invasive species with far-reaching impacts. USDA National Wildlife Research Center - Staff Publications 2027.

Lovette, I. J., B. V. McCleery, A. L. Talaba, and D. R. Rubenstein. 2008. A complete species-level molecular phylogeny for the "Eurasian" starlings (Sturnidae: Sturnus, Acridotheres, and allies): Recent diversification in a highly social and dispersive avian group. Molecular Phylogenetics and Evolution 47:251–260.

Lowe, S., M. Browne, S. Boudjelas, and M. De Poorter. 2000. 100 of the world's worst invasive alien species: a selection from the global invasive species database.

Mando, J., and G. Stack. 2019. Convincing the Public to Kill: Asian Carp and the Proximization of Invasive Species Threat. Environmental Communication 13:820–833.

Martín, P. V., M. A. Muñoz, and S. Pigolotti. 2019. Bet-hedging strategies in expanding populations. PLoS computational biology 15:e1006529.

Mayr, E. 1942. Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press, Cambridge.

Miller, T. E. X., A. L. Angert, C. D. Brown, J. A. Lee-Yaw, M. Lewis, F. Lutscher, N. G. Marculis, et al. 2020. Eco-evolutionary dynamics of range expansion. Ecology 101:e03139.

Monk, J. D., E. Giglio, A. Kamath, M. R. Lambert, and C. E. McDonough. 2019. An alternative hypothesis for the evolution of same-sex sexual behaviour in animals. Nature Ecology & Evolution 3:1622–1631.

North, H. L., A. McGaughran, and C. Jiggins. 2021. Insights into invasive species from

whole-genome resequencing. Molecular Ecology.

Ogden, L. 2018. The beaver diaspora: A thought experiment. Environmental Humanities 10:63–85.

Peischl, S., M. Kirkpatrick, and L. Excoffier. 2015. Expansion load and the evolutionary dynamics of a species range. The American Naturalist 185.

Pfennig, K. S., A. L. Kelly, and A. Pierce. 2016. Hybridization as a facilitator of species range expansion. Proceedings of the Royal Society B: Biological Sciences 283:20161329.

Pigliucci, M., and J. Kaplan. 2000. The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. Trends in Ecology & Evolution 15:66–70.

Purcell, K. L. 2015. Foraging Behavior of European Starlings: Implications for Mitigating Their Impact on Native Species 1–9.

Quilodrán, C. S., F. Austerlitz, M. Currat, and J. I. Montoya-Burgos. 2018. Cryptic Biological Invasions: a General Model of Hybridization. Scientific Reports 8:1–12.

Ricciardi, A., and R. Ryan. 2017. The exponential growth of invasive species denialism. Biological Invasions 20:549–553.

Ricciardi, A., T. M. Blackburn, J. T. Carlton, J. T. A. Dick, P. E. Hulme, J. C. Iacarella, J. M. Jeschke, et al. 2017. Invasion Science: Looking Forward Rather Than Revisiting Old Ground – A Reply to Zenni et al. Trends in Ecology & Evolution 32:809–810.

Rodriguez, A., M. Hausberger, and P. Clergeau. 2010. Flexibility in European starlings' use of social information: experiments with decoys in different populations. Animal Behaviour 80:965–973.

Rollins, L. A., A. P. Woolnough, A. N. Wilton, R. Sinclair, and W. B. Sherwin. 2009. Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. Molecular Ecology 18:1560–1573.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, et al. 2019. Decline of the North American avifauna. Science 366:120–124.

Ross, H. A. 1983. Genetic differentiation of starling (*Sturnus vulgaris*: Aves) populations in New Zealand and Great Britain. Journal of Zoology 201:351–362.

Russell, J. C., and T. M. Blackburn. 2017. The Rise of Invasive Species Denialism. Trends

in Ecology & Evolution 32:3–6.

Schrider, D. R., and A. Kern. 2016. S/HIC: robust identification of soft and hard sweeps using machine learning. PLoS Genetics 12:e1005928.

Schrieber, K., and S. Lachmuth. 2017. The Genetic Paradox of Invasions revisited: the potential role of inbreeding × environment interactions in invasion success. Biological Reviews 92:939–952.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and Ecology of Species Range Limits. Annual Review of Ecology, Evolution, and Systematics 40:415–436.

Shackleton, R. T., D. M. Richardson, C. M. Shackleton, B. Bennett, S. L. Crowley, K. Dehnen-Schmutz, R. A. Estévez, et al. 2019. Explaining people's perceptions of invasive alien species: A conceptual framework. Journal of Environmental Management 229:10–26.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. Ecology Letters 12:1261–1276.

Simberloff, D. 2009. Moving Beyond Strawmen and Artificial Dichotomies: Adaptive Management When an Endangered Species Uses an Invasive One. Journal of Agricultural and Environmental Ethics 22:73–80.

Simberloff, D. 2012. Nature, Natives, Nativism, and Management: Worldviews Underlying Controversies in Invasion Biology. Environmental Ethics 34:5–25.

Simberloff, D., and J. R. S. Vitule. 2013. A call for an end to calls for the end of invasion biology. Oikos 123:408–413.

Stopyra, D. 2016. Get the flock out of here: Starlings are the worst. Birds. Ever. Salon.

Subramaniam, B. 2001. The Aliens Have Landed! Reflections on the Rhetoric of Biological Invasions. Meridians: feminism, race, transnationalism 2:26–40.

Subramaniam, B. 2014a. Ghost Stories for Darwin. University of Illinois Press, Urbana.

Subramaniam, B. 2014b. Spectacles of belonging. in J. Stanescu and K. Cummings, eds. The ethics and rhetoric of invasion ecology, (Un)documenting citizenship in a multispecies world. Lexington Books, New York.

Valéry, L., H. Fritz, and J.-C. Lefeuvre. 2013. Another call for the end of invasion biology. Oikos 122:1143–1146.

Visscher, P. M., W. G. Hill, and N. R. Wray. 2008. Heritability in the genomics era — concepts and misconceptions. Nature Reviews Genetics 9:255–266.

Werner, S. J., J. W. Fischer, and K. A. Hobson. 2020. Multi-isotopic (δ2H, δ13C, δ15N) tracing of molt origin for European starlings associated with U.S. dairies and feedlots. PLoS ONE 15:e0237137.

Wing, L. 1943. Spread of the Starling and English Sparrow. The Auk 60:74–87.

Wood, C. A. 1924. The Starling Family at Home and Abroad. The Condor 26:123–136.

Young, A. M., and B. M. H. Larson. 2011. Clarifying debates in invasion biology A survey of invasion biologists. Environmental Research 111:893–898.

Závorka, L., M. Buoro, and J. Cucherousset. 2018. The negative ecological impacts of a globally introduced species decrease with time since introduction. Global Change Biology 24:4428–4437.

APPENDIX A: CHAPTER 2

A.1 Tests of population structure

DeltaK = mean(|L''(K)|) / sd(L(K))
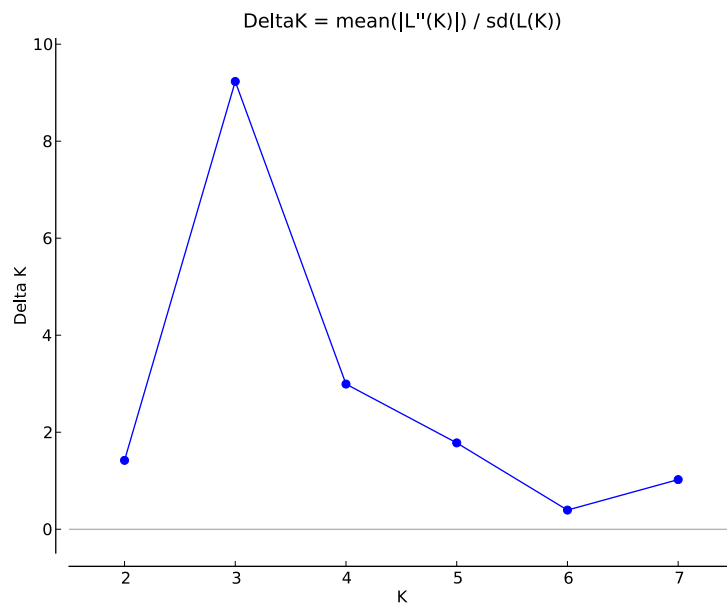
Figure A1. Delta-K values for the STRUCTURE runs.

Because we expect population structure to be fairly low given the recent expansion of North American starlings, we used fineRADstructure to test for more subtle patterns of structure (Malinsky et al. 2016). This program calculates shared ancestry using a coalescent model to determine haplotype linkage among sampled individuals. The resulting coancestry matrix controls for similarity among individuals to infer fine-scale patterns of population structure, and we found no evidence for subtle population structure even using this more sensitive test.
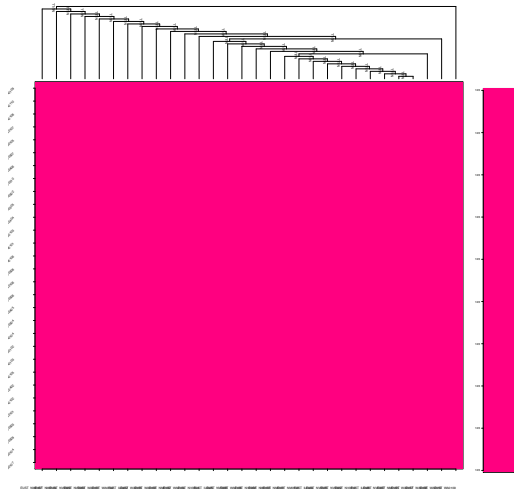
Figure A2. fineRADstructure results.

This clustered coancestry matrix shows that the estimated coancestry coefficient does not vary among sampled individuals. fineRADstructure identifies fine-scale patterns in population structure as a result of shared ancestry, indicating the range of coancestry coefficients using a heat map.

A.2 Isolation-by-distance and isolation-by-environment

The Mantel test is widely used to compare geographic and genetic distances, even though this test may not be appropriate in all such comparisons (Legendre and Fortin 2010). Where there is weak differentiation among populations—as is the case in North American starlings—Mantel's R may be artificially inflated at intermediate migration rates (Meirmans 2012, 2015). The distribution of geographic distances among sampling locations may complicate our use of a Mantel test: as shown in the figure below, there is not a continuous distribution of geographic distances.

150
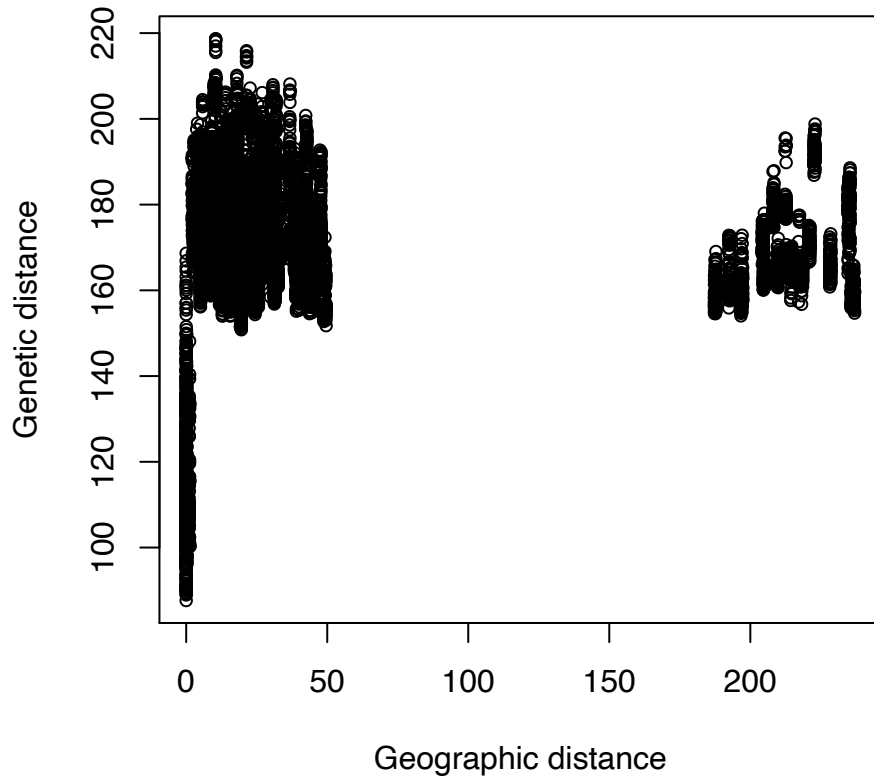
Figure A3. Mantel test of isolation by distance.

As described in the main text of the Methods, we also ran partial Mantel tests control for geographic distance when testing for the relationship between environmental and genetic distances. We report all Spearman correlations here, where the P-value is the two-sided P-value after randomizing the genetic distance matrix 999 times.

Table A1. Partial Mantel tests.

VARIABLE	DEFINITION	MANTEL R	P-VALUE
BIO1	Annual mean temperature	-0.07	0.98
BIO12	Annual precipitation	0.06	0.02
BIO16	Precipitation of wettest quarter	0.28	0.01
BIO7	Temperature annual range	0.15	0.001
ELEVATION	-	0.15	0.001

To identify potential geographic barriers, we also used the program EEMS (Estimated Effective Migration Surfaces, (Petkova *et al.* 2015). EEMS estimates how quickly genetic similarity decays across the landscape, allowing us to pinpoint geographic regions that depart from continuous IBD. Because the number of hypothesized demes (subpopulations) can influence model sensitivity, we ran EEMS using polygons covering the entire North American range and only the areas sampled, and also tested each map using different numbers of demes where the number of demes is limited by the number of individuals (N=50, N=100, and N=150). We adjusted the variance of the proposal distribution for both migration and diversity parameters to ensure that all parameters were accepted between 10 and 15% of the time as suggested in the EEMS documentation, with the input proposal variances as follows: mSeedsProposalS2 = 0.15, mEffctProposalS2 = 1.5, qSeedsProposalS2 = 1.5, qEffctProposalS2 = 0.1, and mrateMuProposalS2 = 0.001. We ran three chains to check convergence.
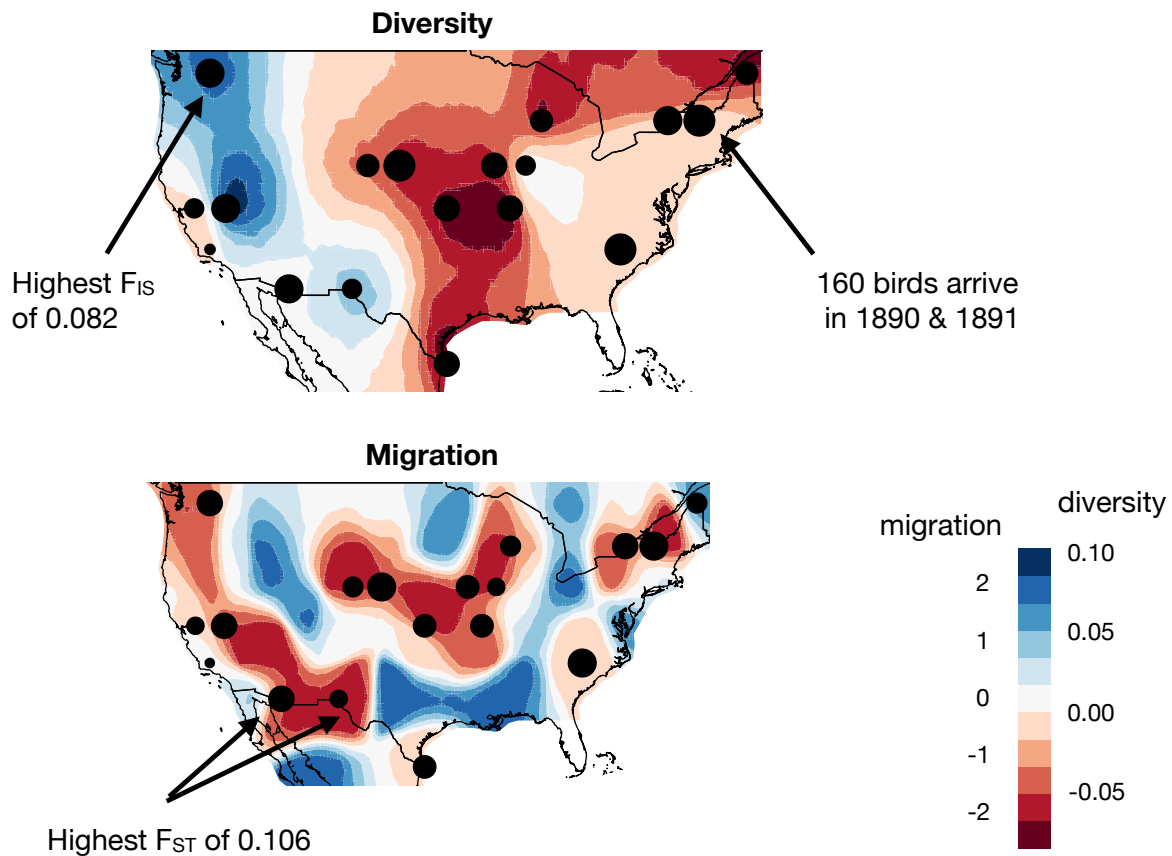
Figure A4. Estimated effective migration surfaces (EEMS) model. Warmer (red) colors indicate lower migration or diversity rates, whereas cooler (blue) colors suggest higher diversity (top) or higher migration (bottom).

The EEMS model recovers fairly uniform rates of migration among sampling locations. Although migration rates appear to be higher in some areas, these areas have not been sampled, and thus we suggest that high migration rates in this model are an artifact of sampling. Even when we decrease the number of demes, we still recover similar results. These results, combined with previous genetic studies (Cabe 1998; 1999), support the

hypothesis that the Rocky Mountains may have imposed an altitudinal barrier to starlings' spread: spatially explicit models indicate a decreased migration rate on the eastern front of the mountains, and an increase in genetic diversity west of the mountain range (Figure 2). In other words, Western starling populations on the range edge are more diverse than populations nearer to the introduction site. Historical records complement this evidence, as the starling expansion slowed only when reaching these mountains (Jernelov 2017). We suggest that elevation may impose a barrier even across this species' worldwide range: both in this study and in a parallel study of Australian populations, patterns of genetic variation can be attributed to a montane barrier (Cardilini et al. 2020).

A.3 Demographic models

The Stairway plot method estimates recent population histories from hundreds of unphased, low-coverage loci, which distinguishes the stairway plot from other demographic methods (e.g., PSMC) that can infer ancient population history more accurately (Liu & Fu 2015). The stairway plot method models changes in population size using the site frequency spectrum, where the null model assumes constant size. We used this model-flexible method to determine whether starlings experienced any genetic bottleneck after introduction: in the stairway plot, this result could occur if an alternative model was accepted during one or more steps of the stairway plot. We assumed a mutation rate of 1×10^{-9} and a generation length of 4.6 years (BirdLife International) and used the recommended 67% of sites for training. The results presented here are averaged among eight independent runs, each with 10 to 30

randomly generated breakpoints during the reconstruction.
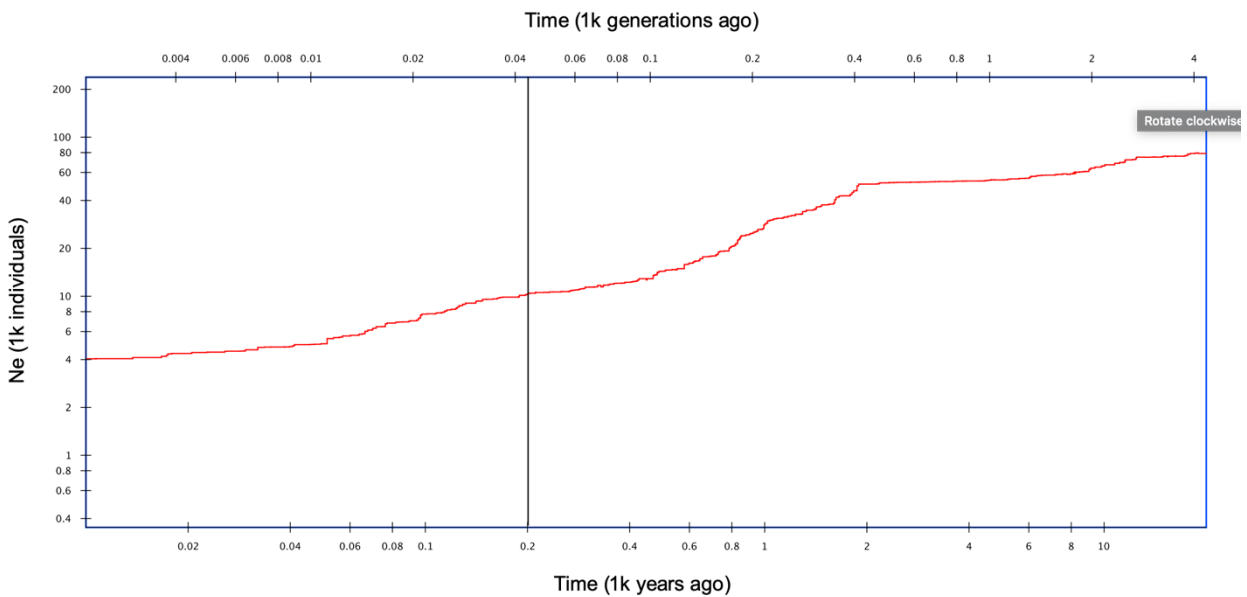
Figure A5. Stairway plot. This model indicates a gradual decline in effective population size over time. Black line indicates approximate colonization date of starlings according to historical records (1890).

The stairway plot method finds that upon introduction approximately 130 years ago, effective population size was 10,000 individuals, and population size has gradually declined to 4,000 individuals. Importantly, the decline in Ne in the most recent time steps—the last 100 years—may be a spurious pattern resulting from known uncertainties in the final steps of this stairway plot method (Liu & Fu 2015).

A.4 Rationale for GEA methods

There are several assumptions underlying the genotype-environment associations presented in this manuscript, including that: (1) RAD loci are an effective tool for

155

exploring true genetic variation; (2) bioclimatic variation at the sampling location is a possible selective pressure that drives genetic variation in starlings; and (3) these data are best-suited to population-level inferences.

(1) We assume that RAD loci can represent actual genetic variation well enough to make inferences about population structure, demography, and even selection. Population geneticists have clearly demonstrated that the genetic diversity recovered by RAD-sequencing tends to be lower than true values (Cariou et al. 2016). In addition, RAD loci are limited to the cut sites of the enzymes used, and thus do not reflect a random sampling of genes (DaCosta & Sorenson 2014). These realities limit the utility of RAD markers, but RAD-sequencing is nonetheless a cost-effective method for exploring genetic diversity and differentiation prior to more thorough sampling of the genome.

(2) Our study assumes that the environmental conditions experienced in the collection location and season represent the conditions that a bird experiences throughout its lifetime. In the main text, we present isotopic evidence that starlings in some sampling locations appear to permanently reside in or near that collection state, but here we focus on the environmental variation across time and space in our dataset. We hypothesized that conditions at the sampling location may drive selection, but environmental conditions do vary between breeding and wintering ranges. Starlings collected in the western U.S. tend to remain in the same region during both breeding and overwintering, but elsewhere in the U.S., starlings may not experience uniform environmental conditions across their lifetime. Importantly, the environmental variation that we use in our genotype-environment associations represents an average of conditions at that location between 1970-2018 and not conditions experienced at the

time of sampling. We assume that this average is a more accurate representation of potential selective pressures than a single point estimate of the conditions experienced by that bird at the time of collection.
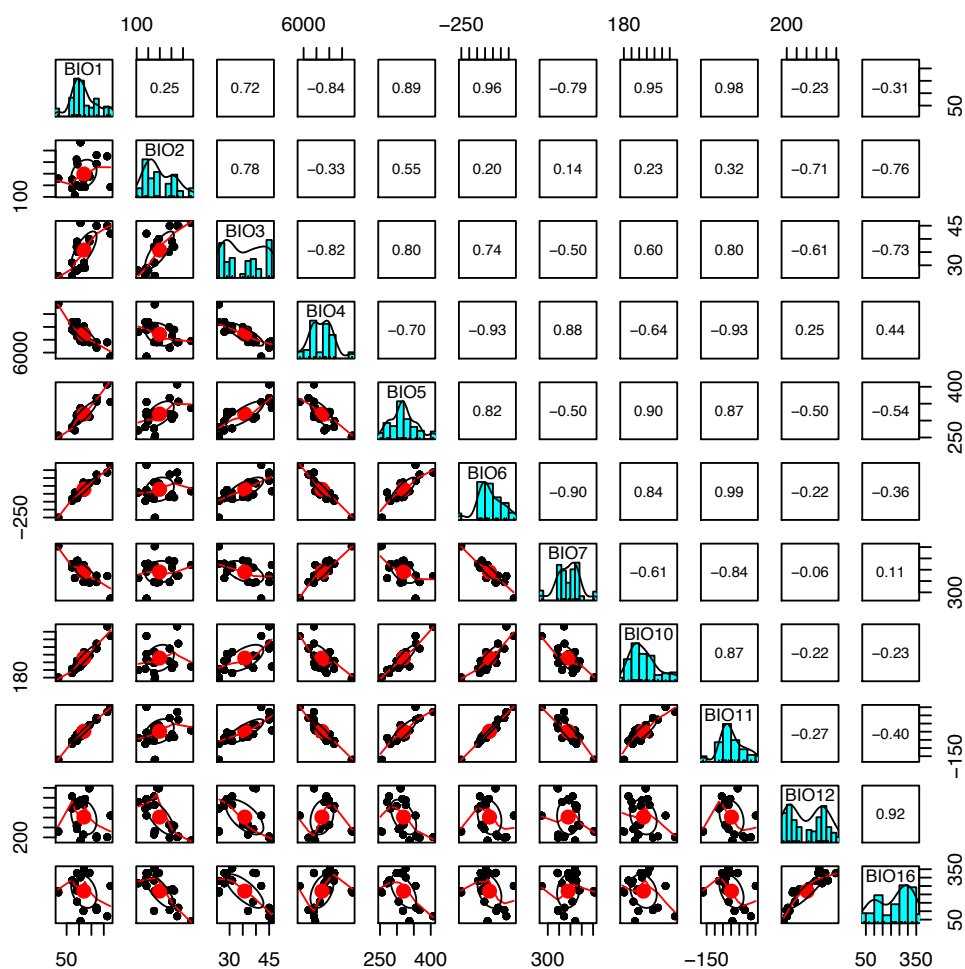
Figure A6. Covariation among and distribution of all bioclimatic variables.

We note that 'winter climate' variables (e.g., BIO6: min temperature of coldest month, BIO11: mean temperature of coldest quarter, and BIO19: precipitation of coldest

quarter) are correlated with the range of temperatures experienced across the year (BIO7), which is retained in the RDA model. We chose to retain BIO7 because it does not correlate significantly with other predictors, and because it is approximately normally distributed, whereas other wintering variables are not (Figure A5).

(3) This study focuses on potential local adaptation across North America, but we acknowledge that individual dispersal and migration may complicate our ability to make inferences about selection. We cannot compare individuals' molt origin and RAD-sequencing data side-by-side, since the original collection of starlings did not identify individuals and instead pooled samples within a population. Because we cannot trace individuals in this comparison, from the . Although migratory strategy can vary within the same sampling location, the molt-origin data presented in Werner et al. do enable us to determine whether individuals sampled at a given location are likely to experience consistent environmental conditions.

A.5 Comparing selection-scan methods

Differentiation methods can identify loci that have undergone strong selective sweeps, but these methods may be inappropriate in systems like this one with low overall differentiation. Testing for selection on such a short time-scale—and especially in a system where a genetic bottleneck may explain much of the variation observed—is a major challenge. We discuss demographic evidence in the discussion section of the main text, but given that the effective population size expanded so rapidly after the initial founder effect, we argue that the relative importance of selection in shaping genetic variation in North American starlings may be stronger than in other invasions.

However, any evidence for selection is likely to be subtle, and each of the selection tests used in this project have different assumptions that we outline in this section. Traits under selection—especially environmentally-driven selection—are likely to be polygenic, and thus we would expect to find low levels of differentiation at many loci. For this reason, we suggest that multivariate methods such as redundancy analyses may be more appropriate tests of selection in this system than univariate or differentiation-based methods. Each selection scan is best suited to a particular environmental context: RDA is generally better able to identify selection when the environment gradient is not correlated to population structure (Capblancq et al. 2018), and when selection is multi-locus and weak (Forester et al. 2018). However, where there is spatial autocorrelation and environmental conditions change on the same scale as geography, many genotype-environment association methods may struggle to disentangle geography and environment. Given that population structure is negligible within North American starlings, we suggest that controlling for structure—e.g., with LFMM—may be less important than ensuring sufficient statistical power to identify multilocus adaptation.

Results from the redundancy analysis are thus reported in the main text, whereas we report the methods and results of all other tests here. Briefly, there is little overlap among these tests: no SNPs were identified by all tests, 13 SNPs were identified by both RDA and LFMM, and 3 SNPs by LFMM and BayeScEnv. This low level of overlap is consistent with other studies that compare RDA and LFMM in the same system (Capblancq et al. 2018). Furthermore, we report only evidence from the selection scan method best suited to our assumptions, given that reporting only the overlap would

limit selection inference to the test with the least power (Forester et al. 2018).

A.6 Bayescan and BayeScEnv

FST-based genome-scan approaches are best suited to identify loci that stand out against a low background level of differentiation. To test these more traditional methods against the model-based approaches (LFMM & RDA), we used both BayeScan (Foll & Gaggiotti 2008) and BayeScEnv (de Villemereuil & Gaggiotti 2015). BayeScan identified no SNPs with a Q-value lower than 0.989, indicating that there is no evidence of selection based on this test (Table S2; excel file). BayeScEnv incorporates environmental differentiation when identifying outlier loci by including a term to explicitly model environmental differentiation in the framework used in BayeScan. BayeScEnv identified several SNPs that could be under selection (Table S2), and three of these were also found the LFMM analysis (below).
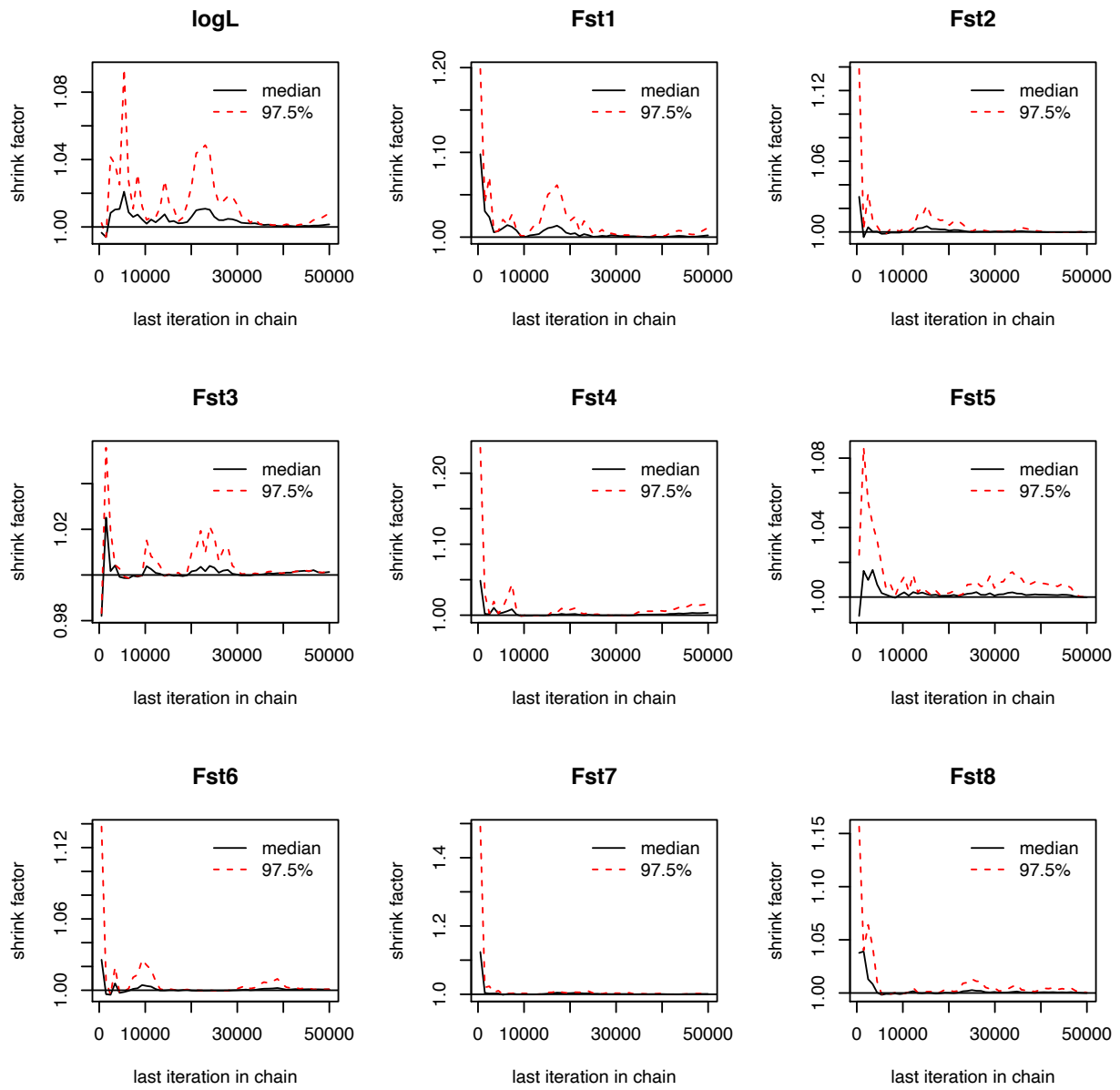
Figure A7. Gelman Plot for BayeScEnv. Here, the shrink factor demonstrates convergence of each model.

A.7 Latent-factor mixed models (LFMM)

As a univariate test of selection, we used the lfmm function (Frichot et al. 2013) to test for associations with climatic gradients and to decrease the number of false positives. For the univariate method (LFMM), environmental variation was modeled as the first three principal components of bioclimatic variation across the range of North American starlings. We used the R package LEA (Frichot & François 2015) to prepare input files and run a model where genotypic variation is considered a response variable in a linear regression that controls for latent factors (e.g., population structure and/or background variation) in estimating the association between the genotypic response and the environmental predictor. For each of three models—including 1, 2, and 3 latent factors—we ran 30 MCMC chains of 10,000 cycles each, discarding a burn-in of 5,000 cycles. Z-scores were combined across all 30 runs and p-values readjusted to calibrate the null hypothesis and increase power using the Fisher-Stouffer method as suggested in the LEA and LFMM manuals. We used the Benjamini-Hochberg algorithm to control for false discoveries.
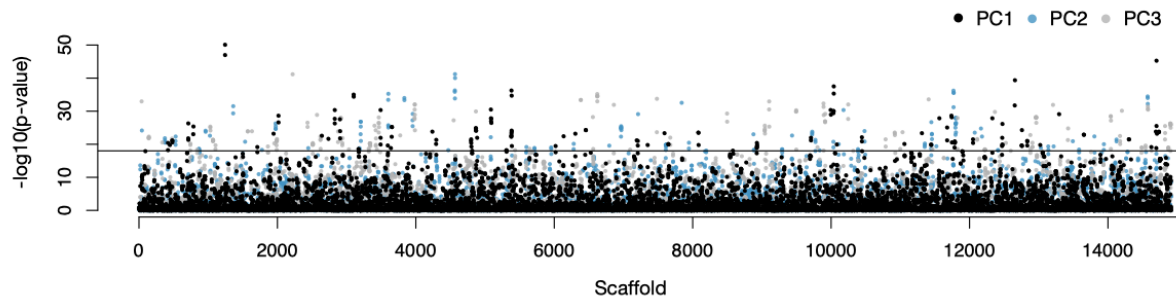
Figure A8. Latent-factor mixed model (LFMM) results. Each point reflects an association between a SNP and environmental variation (captured as a principal component of all possible bioclimatic variation across sites).

Latent-factor mixed models identified 2490 candidate variants associated with the first principal component of environmental variation, which explains 41.5% of the variation and loads with temperature-related variables. An additional 1315 variants were associated with precipitation-related PC2, and/or with PC3, a composite of temperature and precipitation variables. Since we identified many candidates using a q-value cut-off of 0.01, only loci that were identified in all three runs (K=1-3 latent factors) and were more than five standard deviations from the mean log10p value were considered candidates under selection (false discovery rate < 0.05, true positive >25). This filtering left 1218 remaining SNPs—or 8% of all SNPs—distributed across all three principal components of environmental variation (Table S2).

A.7 Functions of genes and gene ontology information

Although no gene ontology categories were significantly overrepresented, signaling and response to stimuli were particularly well-represented among GO terms, showing up to 48-fold enrichment (FDR-corrected P=0.12-0.98, Table S2). It is important to note that this analysis does not correct for gene size nor does it expect that any 'candidates' reported here are likely to drive adaptation in the North American starling. However, we find it useful to examine possible functions that would need to be verified by whole-genome data. Among signaling-related GO terms, neuron development, synaptic transmission and organization were particularly common (Table S2). Other common GO terms relate to kidney function, viral processing, metabolism, and regulation of growth factors. Among the top twenty variants under strong selection (r2 > 0.2 or log10p > 10), we find four genes related to growth factors (EOGT, GAB3, HBEGF,

STAT3), six involved in immune responses that do not directly involve growth factors (DNAJB14, FKBP4, ASB2), and three essential to muscle function (LIMCH1, HBEGF, CALD1). Putatively selected genes may play a role in physiological processes that support starlings' invasion success in North America.

Effective solute transport and kidney development are critical in dry habitats, which may explain why all but one of the genes related to kidney function correlate with precipitation (BIO16). Claudin 16 (CLDN16; $R2 = 0.23$) is one such protein that regulates ion concentrations in the kidney, while others maintain homeostasis and vasoconstriction (AVPR1B; $R2 = 0.18$) or transport iron (STEAP3; $R2 = 0.21$). Many invaders shift their diet upon colonization of a new habitat, and many candidates play a role in metabolism and/or digestion: for example, aridity may result in selection on proteins that process lipids (MTMR3; $R2 = 0.18$) and fatty acids (PEX5; $R2 = 0.17$), since organisms living in dry environments often depend on fat storage for proper hydration. Proteins that modify growth factors—key orchestrators of cellular growth and development—correlate with aridity but also temperature, and complexes that rely on ubiquitin ligases to degrade proteins are similarly strong candidates. Many of the putatively selected genes are critical to starlings' survival, and investigating a wider range of environmental conditions and sampling whole genomes may support this preliminary evidence of incipient local adaptation.

References

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

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

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

Cariou M, Duret L, Charlat S (2016) How and how much does RAD-seq bias genetic diversity estimates? *BMC Evol Bio,l* **16,** https://doi.org/10.1186/s12862-016-0791-0

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

DaCosta JM, Sorenson MD (2014) Amplification Biases and Consistent Recovery of Loci in a Double-Digest RAD-seq Protocol. *PLoS ONE*, **9**, e106713.

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.

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

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

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

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

Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831-44.

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

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

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839-2846.

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

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

Stuart KC, Cardilini APA, Cassey P, Richardson MF, Sherwin W, Rollins LA, Sherman CDH (2020) Signatures of selection in a recent invasion reveals adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, **00**, 1-16. doi: 10.111/mec.15601.

APPENDIX B: CHAPTER 3

B.1 Geographic sampling and sequencing quality

Table B1. Sampling for each individual.

Table includes latitude and longitude coordinates for each individual (ind.) sample.
Additional columns indicate the total number of bases sequenced (in Kb) for each
individual, coverage from two lanes of sequencing, average quality score (Qscore) for
each round of sequencing (R1 and R2), and percent of bases over a quality threshold of
30 for each round.

Ind.	Lat	Long	Site	Total Bases Kb	Coverage	Avg QScore (R1)	Avg QScore (R2)	% >Q30 (R1)	% >Q30 (R2)
uk1	54.99	-1.90	Northumberland	20,841,992	17.37	38.19	36.24	92	85
uk2	54.99	-1.90	Northumberland	29,361,687	24.47	38.26	36.51	92	86
uk3	55.09	-1.98	Northumberland	27,540,273	22.95	38.2	36.06	92	84
uk4	54.99	-1.90	Northumberland	24,594,124	20.50	38.18	36.08	92	84
uk5	54.97	-1.98	Northumberland	23,291,164	19.41	38.34	36.39	92	85
uk6	54.99	-1.90	Northumberland	24,423,207	20.35	38.34	36.44	92	85
uk7	55.09	-1.98	Northumberland	26,595,490	22.16	38.25	36.09	92	84
uk8	55.09	-1.98	Northumberland	25,575,830	21.31	38.23	36.19	92	84
au1	-40.17	151.5	Lemontree	29,066,476	24.22	38.28	36.54	92	86
au2	-40.17	151.5	Lemontree	25,044,625	20.87	38.29	36.34	92	85
au3	-44.68	151.0	Newcastle	21,233,948	17.69	38.01	36.16	91	84
au4	-44.68	151.0	Newcastle	22,054,657	18.38	38.29	36.43	92	85

au5	-48.08	155.6	Wonthaggi	21,087,222	17.57	38.17	36.19	92	84
au6	-48.08	155.6	Wonthaggi	22,758,978	18.97	38.19	36.27	92	85
au7	-42.37	139.7	Meningie	23,318,043	19.43	38.35	36.63	92	86
au8	-42.37	139.7	Meningie	28,061,300	23.38	38.15	36.21	92	84
us1	40.65	-73.82	NYC	29,347,953	24.46	38.32	36.34	92	85
us2	40.65	-73.81	NYC	24,843,058	20.70	37.65	36.04	90	84
us3	40.65	-73.82	NYC	19,679,896	16.40	37.94	36.39	91	85
us4	40.64	-73.81	NYC	22,273,753	18.56	38.24	36.42	92	85
us5	40.65	-73.82	NYC	18,268,966	15.22	38.34	36.53	92	86
us6	40.64	-73.75	NYC	24,266,344	20.22	38.25	36.48	92	85
us7	40.64	-73.76	NYC	26,614,217	22.18	38.22	36.55	92	86
us8	40.64	-73.76	NYC	22,832,374	19.03	38.29	36.45	92	85

B.2 Three different genome assemblies produce similar results

We first compared how different genome assembly versions may affect variant calling
and further downstream analysis. The raw data from a global whole genome starling
comparison project was realigned to each of the three genome assemblies. A scaffolded
version of *S. vulgaris* vNA (GCF_001447265.1, scaffolded using Zebra Finch genome
GCF_008822105.2 and Satsuma2), *S. vulgaris* vAU1.0 (Stuart et al. 2021), and lastly a
non-scaffolded version of *S. vulgaris* vAU that involved an initial SUPERNOVA (v2.1.1)
(Weisenfeld et al. 2017) assembly and polishing using PILON (v1.23) (Walker et al.) of
the 10x linked-read data in Stuart and Edwards *et al.* 2021. The variant calling pipeline
was similar to that used in the main manuscript, with some program version
differences. Briefly, eight *S. vulgaris* individuals from each of the United Kingdom (UK;

native range), North America (USA; invasive range), and Australia (AU; invasive range) were sequenced using short read Illumina sequencing. The raw reads were processed using SAMTOOLS (v1.9) (Li et al. 2009) and BEDTOOLS (v2.27.1) (Quinlan and Hall 2010), and adapters were removed using ADAPTERREMOVAL (v2.2.2) (Schubert et al. 2016). The processed reads were aligned to each of the three genome assemblies (Table 1) using BOWTIE2 (v 2.3.5.1) (Langmead and Salzberg 2012), and indexed using PICARD (v2.18.26) (Anon 2019). GATK'S (v4.1.0.0) (Poplin et al. 2018) HaplotypeCaller (GVCF mode) was used to call haplotypes for each sample, which were processed with CombineGVCFs and finally passed into GenotypeGVCFs to produce the initial VCF file. The VCF file was put through an initial filter step in GATK (QD<2.0, FS>60.0, MQ<40.0, SOR>3.0), and a secondary filter step (max missing count=4, min mean DP=2, max mean DP=50, min alleles=2, max alleles=2). VCFtools (v0.1.16) (Danecek et al. 2011) was then used to create three sets of filtered VCF files for each genome. A SNP data set was created by filtering for minor allele frequency (MAF) of 0.1, and one data set was created at MAF 0.05. The MAF 0.05 dataset was then further filtered; SNPs were pruned in BCFTOOLS (samtools v1.9) for linkage by removing sites with an $r^2 > 0.6$ within 1000 bp site windows (Linnér et al. 2019). This last data set was used for population genetics analysis.

The three variant data files were assessed using samtools (v1.9) *bcftools stats* function. The final data set (MAF 0.05 and LD filtering) was parsed through SNPRELATE (v1.22) (Zheng et al. 2012), and displayed in a PCA to illustrate individual clustering and population relatedness. ADMIXTURE (v1.2) (Alexander and Lange 2011) analysis was used to explore population relatedness and admixture.

Some minor differences in SNP counts and the levels of SNP missingness per individual was found between the non-scaffolded *S. vulgaris* vAU genome version and the other two scaffolded genome versions (Fig. 1, Fig. 2). This difference is likely due to smaller scaffold sizes affecting the mapping of the short whole genome sequencing reads in the non-scaffolded *S. vulgaris* vAU genome version. However, these discrepancies were minor, and no biologically meaningful differences were found in the population differentiation analysis using the three different genome versions (Table 1, Fig. 3, Fig. 4). These results indicate that neither the population from which the reference individual has been sourced, nor the level of genomic scaffolding (continuity of the genome), has an effect the SNP based population analysis conducted above. Further, non-scaffolded *S. vulgaris* vAU did not undergo chromosomal alignment to the chromosomes of the *T. guttata*, while the other two genome versions used here did undergo this scaffolding process. This indicated that the synteny alignment conducted had no biological impact on the population differentiation analysis conducted, and all analyses in the main text and the rest of this appendix use the vNA version.

a)

b)

c)
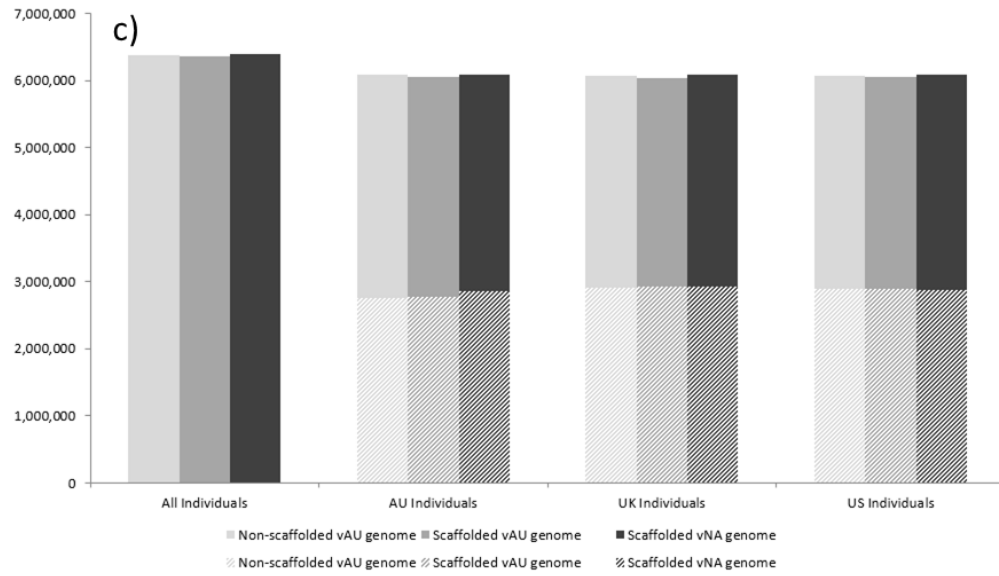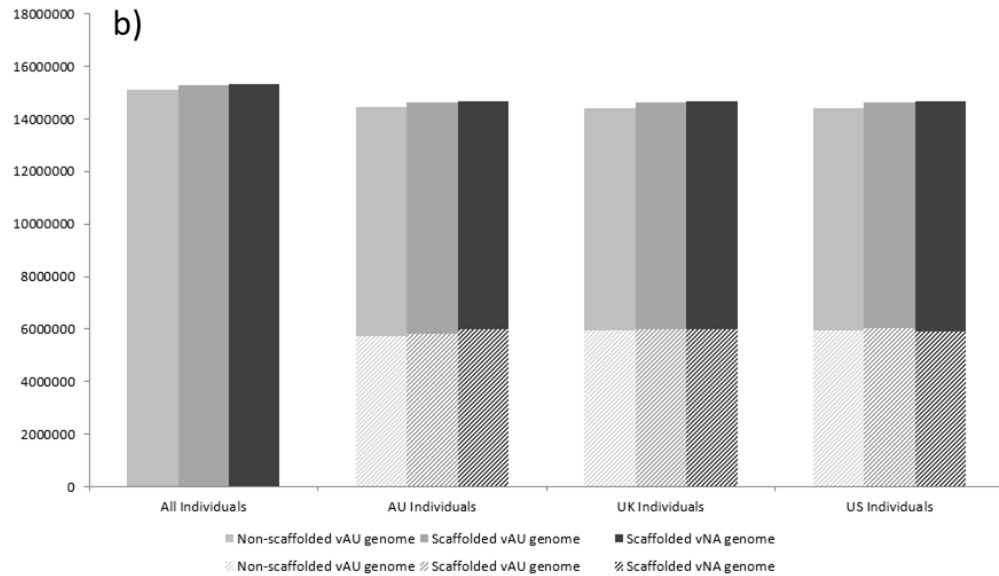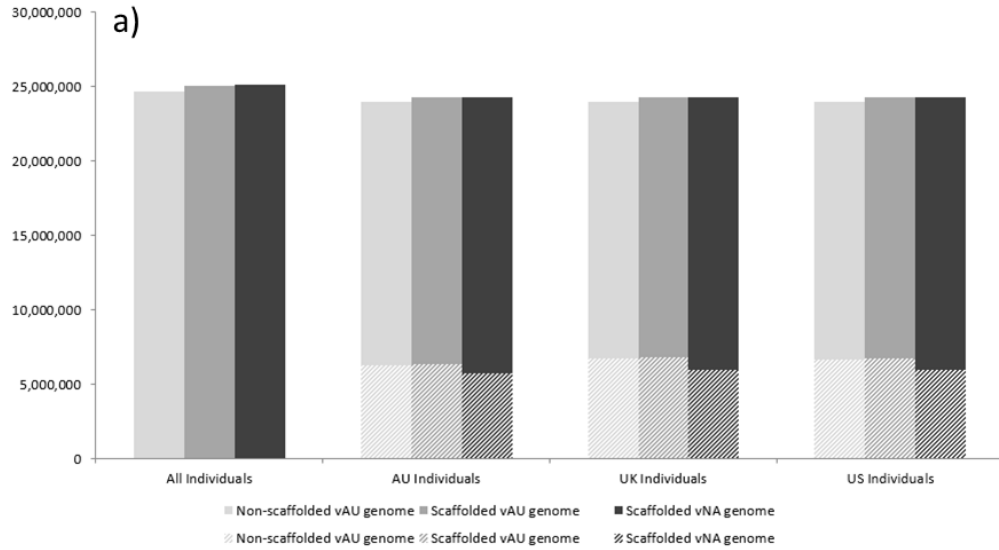
Figure B1. Number of alleles recovered for each of the three S. vulgaris genome versions. a) MAF of 1%, b) MAF 5%, c) MAF 5% and linkage filtered at r2<0.6 in 1000 bp sliding windows. Total alleles (solid) contains both reference and non-reference alleles (nRefHom + nNonRefHom + nHets), while non-reference SNP (stripped) is a nNonRefHom and nHets (nNonRefHom + nHets). Counts are obtained from BCFTOOLS, SNP counts for AU, UK and NA individuals were obtained by calculating the values per individual and averaging them across the population (n=8).
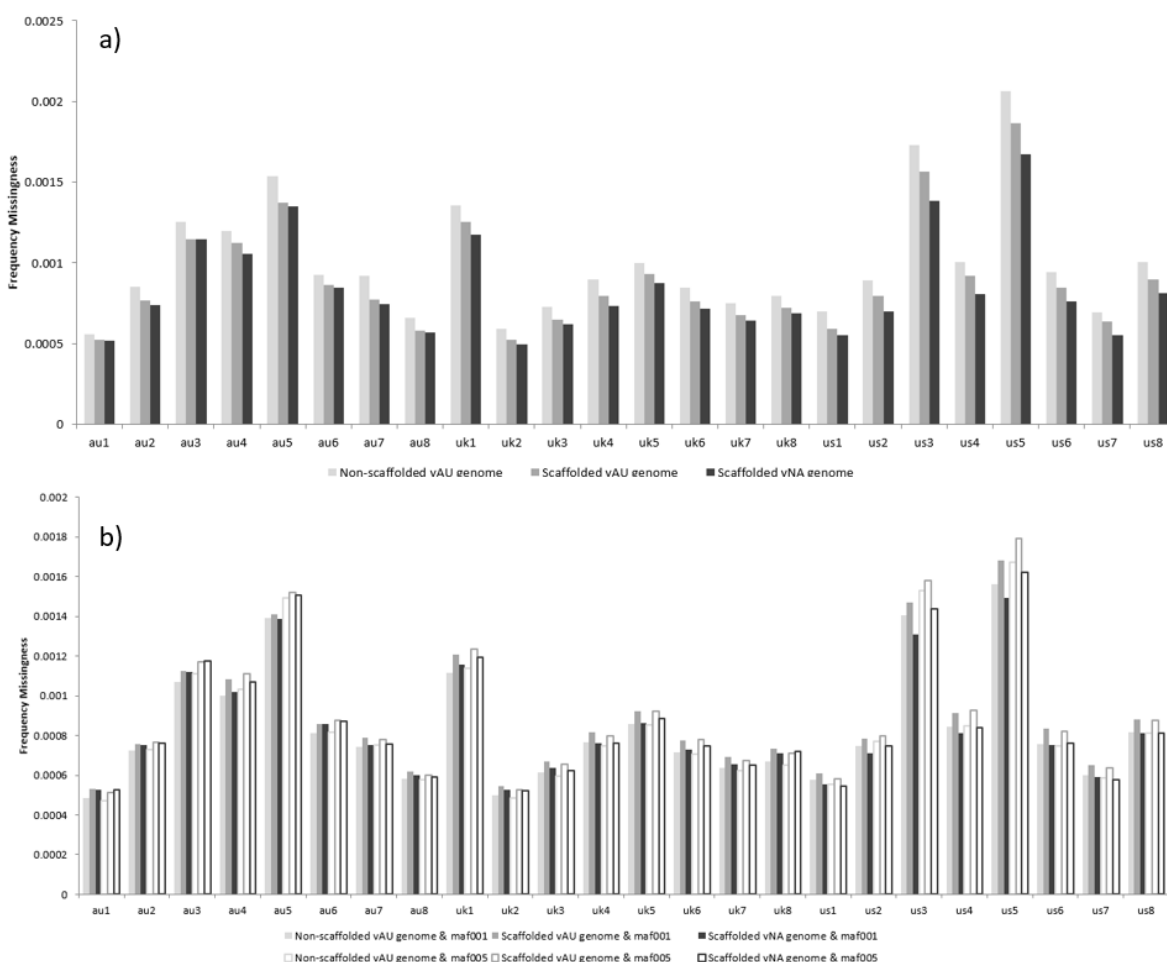
Figure B2. Missingness per individual for each genome assembly. a) MAF 5% and

linkage filtered at r2<0.6 in 1000 bp sliding windows, and b) MAF 1% (dark) and MAF 5% (light). Individual missingness calculated using VCFTOOLS.

Table B2. Differences in genome-wide FST among genome assemblies. FST calculated between pairwise population comparisons (n=8) for MAF = 0.05 and $r^2<0.6$.

Non-scaffolded vAU genome			
	AU	UK	NA
AU	-	0.0431668	0.045913
UK	-	-	0.0374013
Scaffolded vAU genome			
	AU	UK	NA
AU	-	0.0430631	0.045771
UK	-	-	0.0373351
Scaffolded vNA genome			
	AU	UK	NA
AU	-	0.043003	0.045816
UK	-	-	0.037413
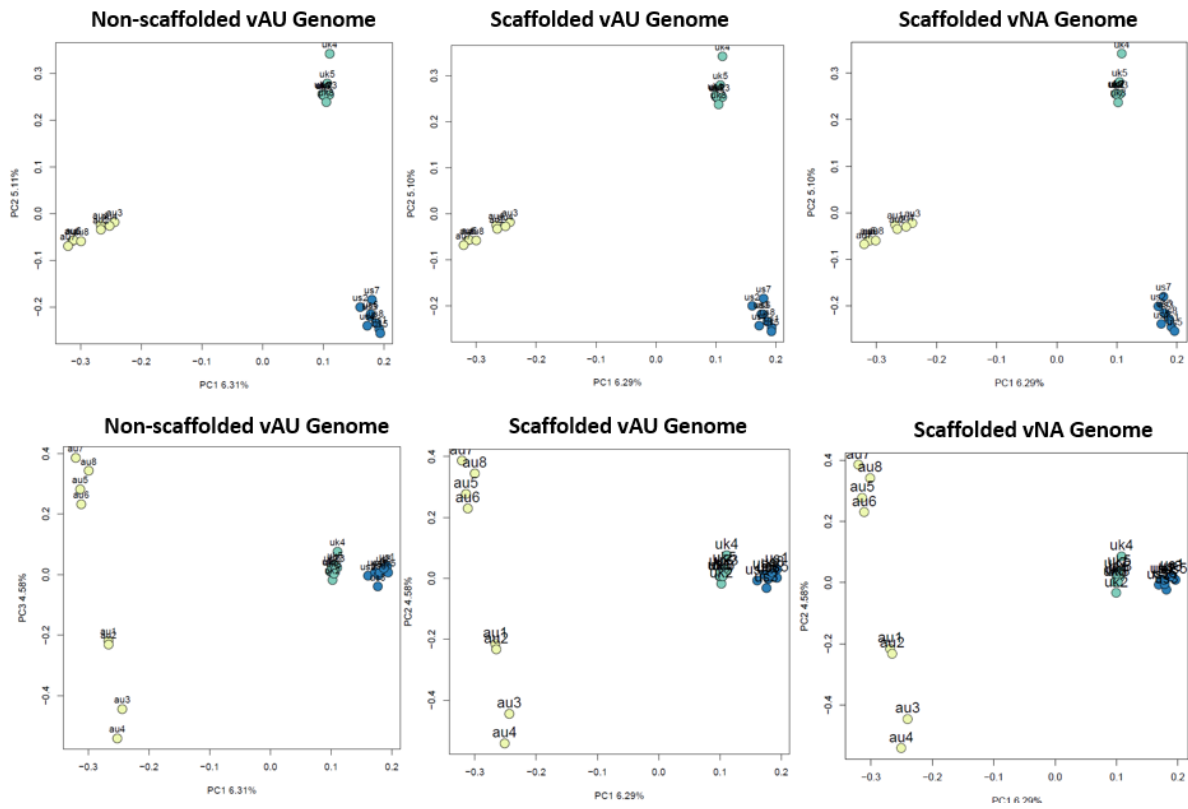
Figure B3. PCA comparisons among genome assemblies. The variant data are filtered using MAF 5% and linkage threshold of r2<0.6 in 1000 bp sliding windows. Produced using SNPRELATE.

175
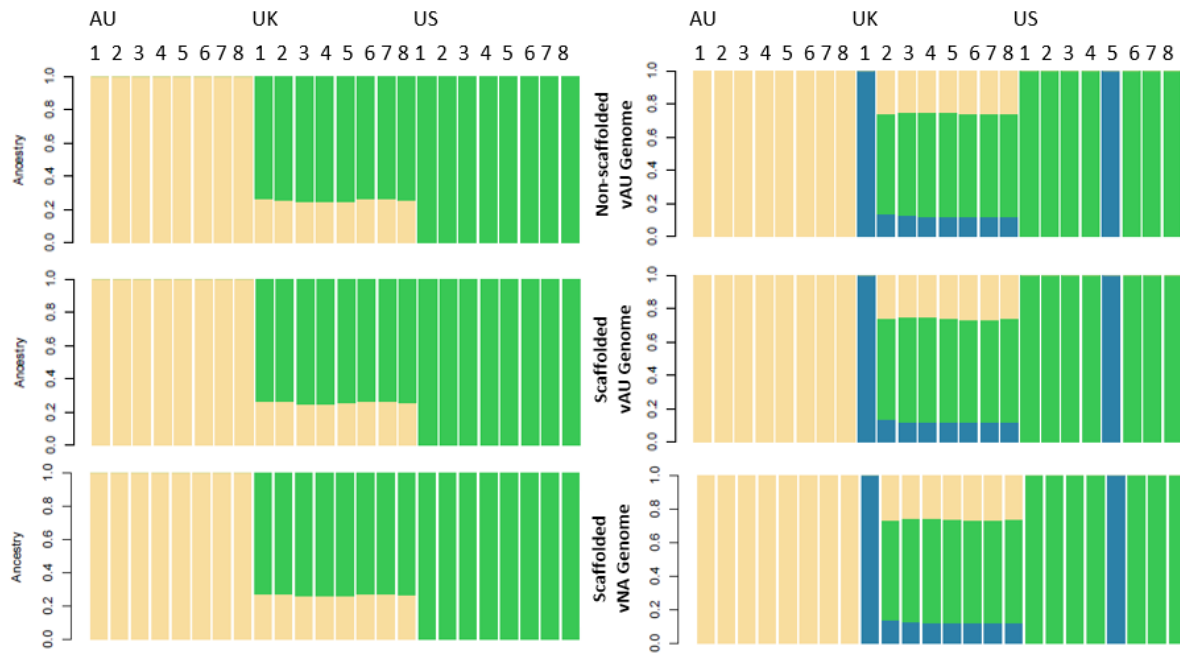
Figure B4. Admixture comparisons among the three S. vulgaris genome versions. K=2 is shown in the left column and K=3 in the right column. These data are filtered using MAF 5% and linkage filtered at r2<0.6 in 1000 bp sliding windows.

B.3 Variant-calling method does not change results dramatically

B.3.A Sequencing data quality

Recent advances in genomic methods urge caution when calling and filtering variants (Ahrens et al. 2021), and we note that the methods we use here represent only one of many possible strategies. Representing variants in the form of likelihoods captures the uncertainty inherent in sequencing data, and this approach is essential when working with low-coverage or low-quality data. However, when sequencing depth and quality is relatively high, we can be more confident that the called variant is the true variant at

that site. In these data, sequencing depth was relatively high even before filtering (Figure B5), but we describe sensitivity analyses of variant-calling approaches in Section B.2.B below.
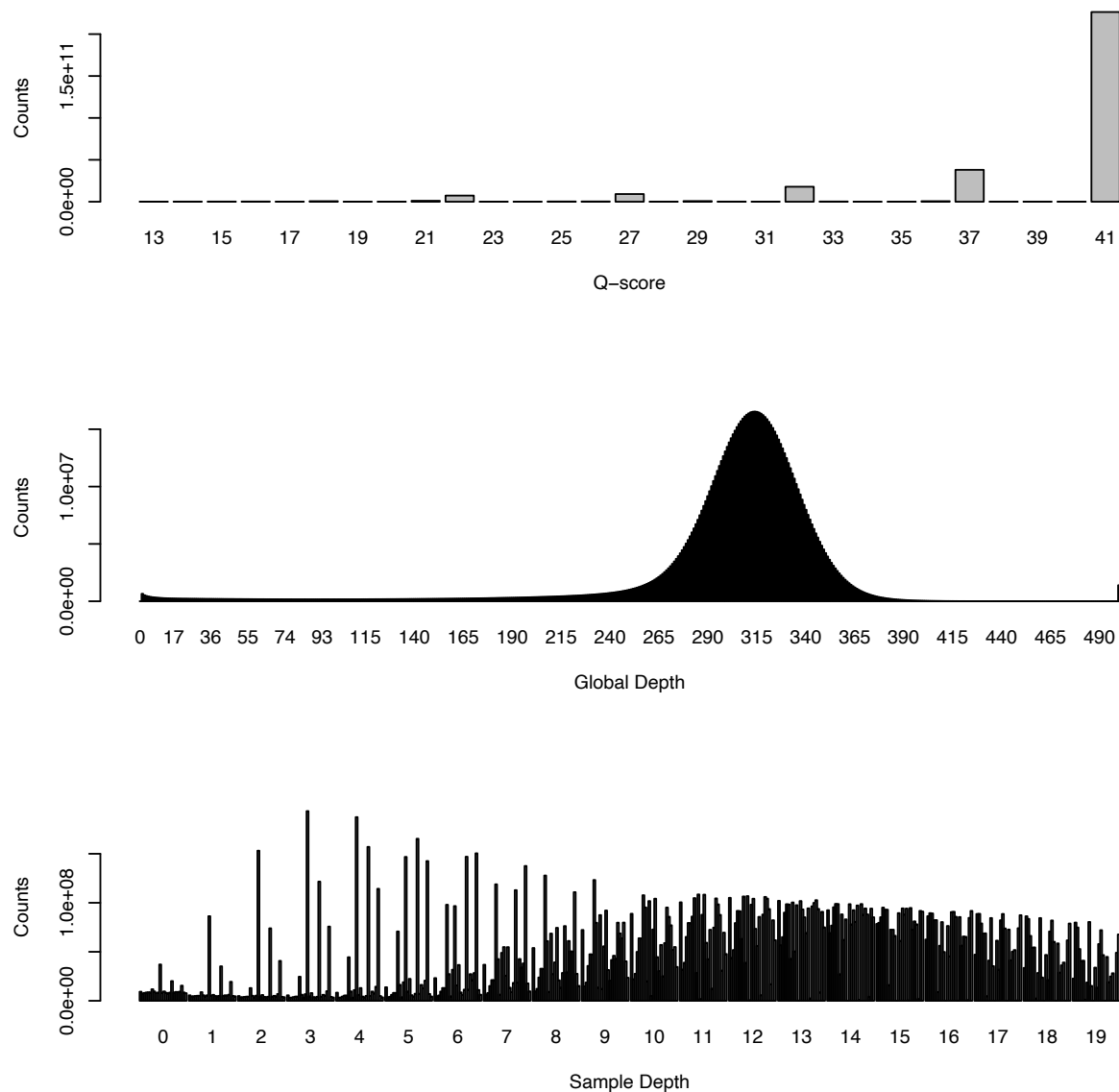
Figure B5. Quality of sequencing data as quantified in ANGSD. Top row indicates base quality score (Q-score), middle row shows read depth across all individuals (global

depth), and bottom row shows sequencing depth for each individual (sample depth.)

B.3.B Impact on diversity metrics

First, we note that genome-wide patterns are sensitive to whether or not invariant sites are included in genome scans: although the generalized results presented here are consistent regardless of which sites are included, F_{ST} can change dramatically when invariant sites are included or not. By way of example, here is a qualitative description of how patterns of differentiation on Chromosome 2 change: in a sliding scan of 50-kb windows, F_{ST} peaks in US vs. UK show up with variant sites only but not when we include invariant sites.

B.3.C. Impact on population structure

We first checked whether population structure results were sensitive to variant-calling method. We find that a principal components analysis using genotype likelihoods yields results consistent with the variant-called set in Figure 3.1 of the main text.
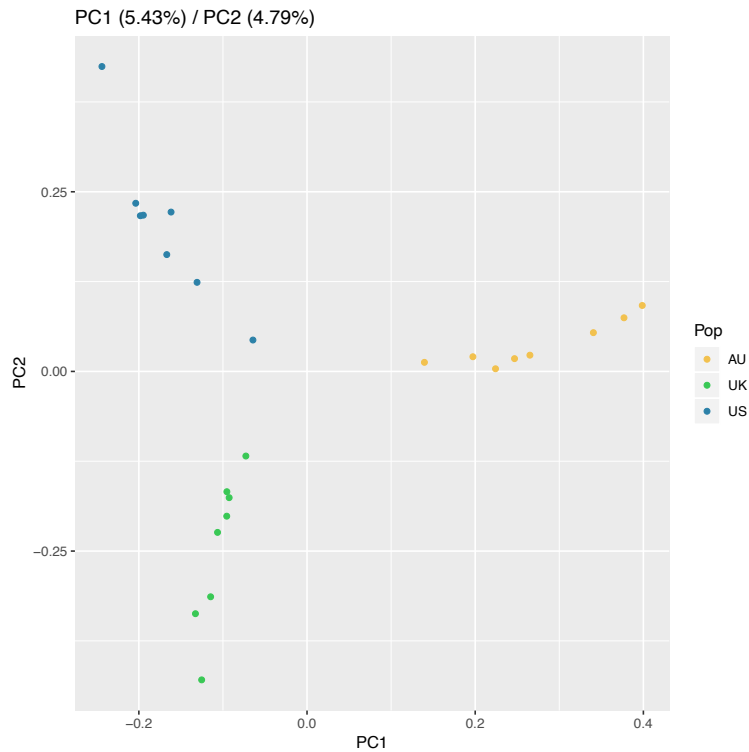
Figure B6. Principal components analysis of the same dataset using genotype likelihoods, to compare to called variants in the main text.

B.3.D. Impact on demographic model

Demographic models rely on subtle differences in allele frequencies to infer bottlenecks or other demographic shifts, but these differences can easily reflect sequencing errors. In particular, singleton frequency is highly dependent on sequencing quality. For this reason, any demographic methods presented in this manuscript use a site-frequency spectrum built within the ANGSD framework to avoid possible biases due to sequencing errors.

B.4 Additional tests of population structure

PC2 distinguishes among Australian individuals and explains nearly as much variation as the first component (5.23%). Additional components distinguish among populations nearly as effectively: PC3 (5.12%) shows that the Australian population falls in between UK and NA clusters, and PC4 (4.82%) indicates differences between the NA and other populations (Figure B5).
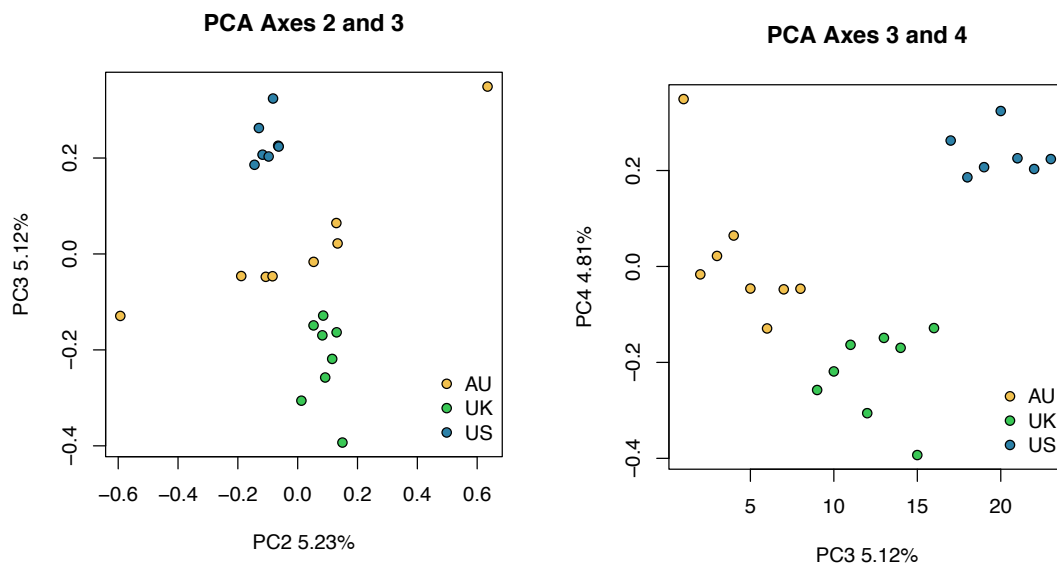
Figure B7. Additional PCs of a principal component analysis of the variant-called dataset.
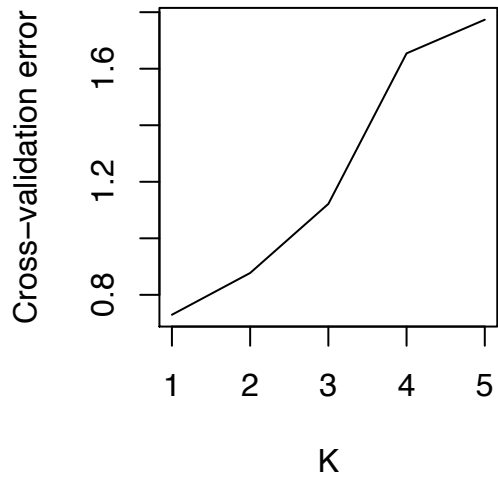
Figure B8. Cross-validation errors for ADMIXTURE tests of K=1-5.

B.5 Demography and inbreeding

Table B3. Inbreeding statistics (F-statistic) for each individual.

Individual	F-statistic
AU1	0.32043
AU2	0.03098
AU3	0.03386
AU4	0.05907
AU5	0.10412
AU6	0.25472
AU7	0.06493
AU8	0.0591
UK1	-0.01137

UK2	-0.02545
UK3	-0.0242
UK4	0.02273
UK5	-0.02165
UK6	-0.02398
UK7	0.06993
UK8	-0.01048
US1	0.00674
US2	-0.01957
US3	0.48915
US4	-0.00911
US5	0.01501
US6	-0.00107
US7	-0.00554
US8	0.00771

Table B4. Demographic model output from fastsimcoal2. NBOT indicates the size of the bottleneck (in individuals), TBOT the time of the bottleneck in years, NANC the size of the ancestral population, and NCUR the current estimated population size.

Population	NBOT	TBOT	NANC	NCUR
UK	31710.2	111497	514837.8	784301.2

AU	4394.4	178.4	541308.6	60306
NA	12299.2	72.6	529885.6	64410.8

B.6 Genomic architecture of the starling genome

Table B5. Distance between center of FST peaks shown in main text and approximate centromere position.

Chromosome	Centromere position (Mb)	Center of peak (Mb)
1	98.17	105
1A	62.54	49
2	76.29	48
4	16.82	23
4A	19.79*	6
6	0.89	5.5
Z	27.51	

*This position is opposite to Volker et al. 2010.
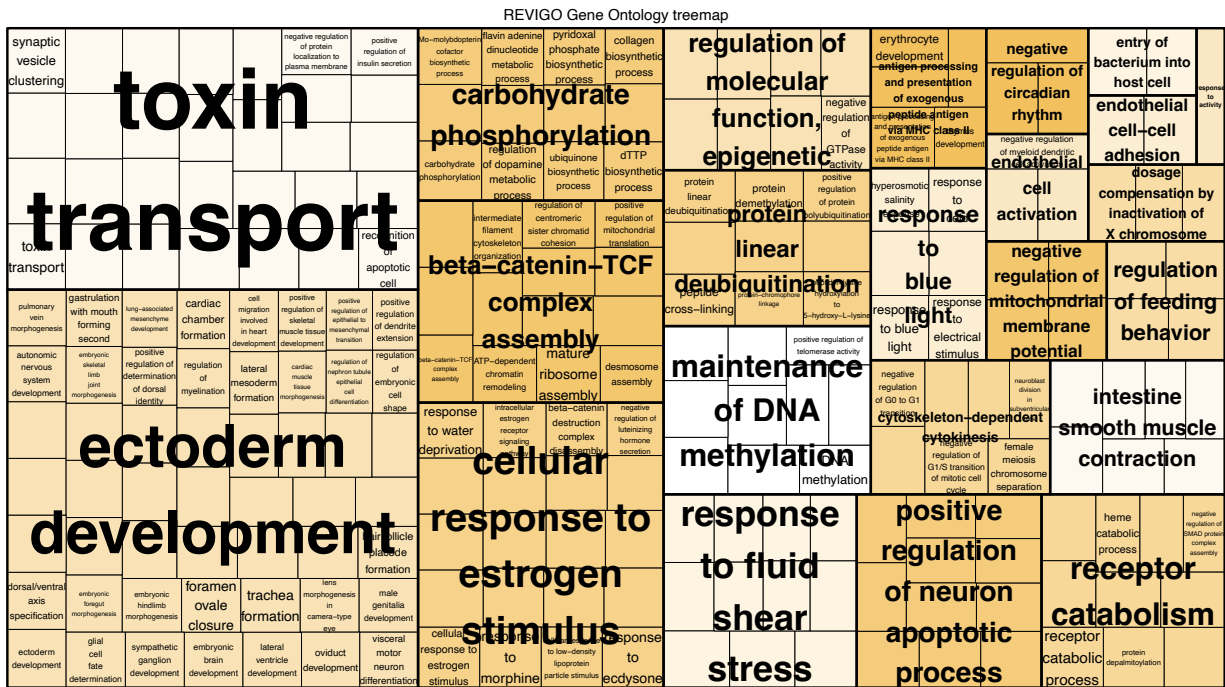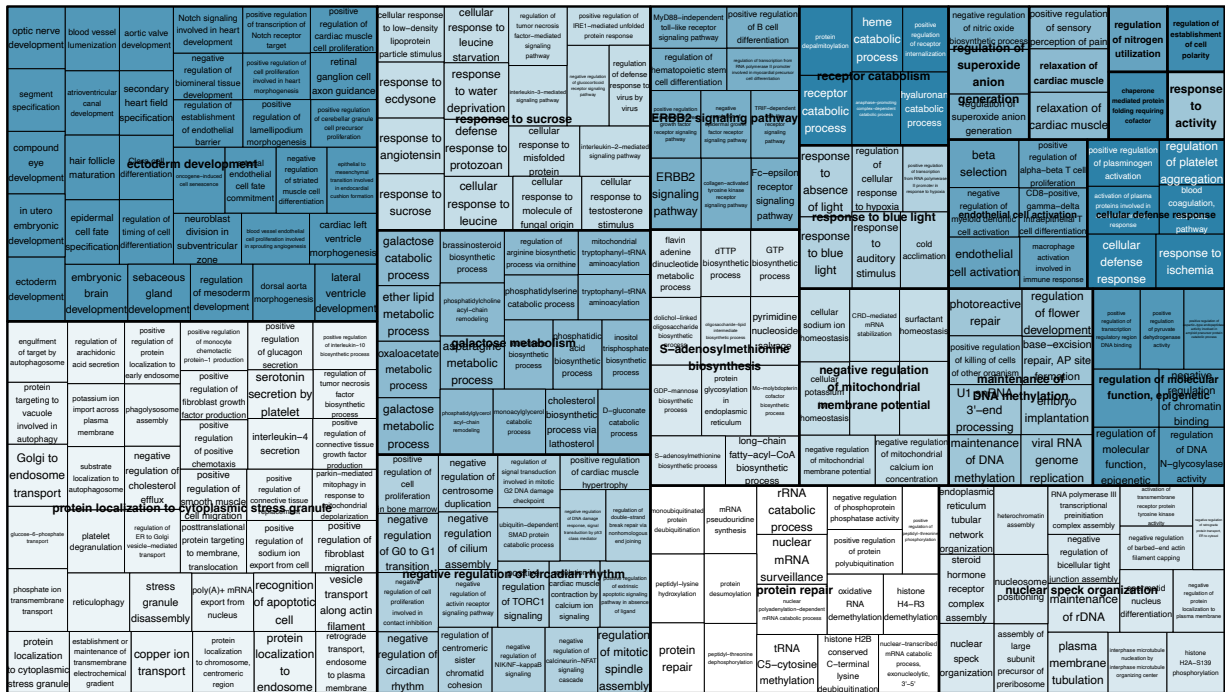
B.7 Gene ontology analyses

Figure B9. Common gene ontology (GO) categories represented in Australian vs. UK outliers.

Figure B10. Common gene ontology (GO) categories represented in NA vs. UK outliers.

B.8 Selection inference at additional FST peaks

Table B6. Estimates of population genetic statistics for all regions of elevated FST.

Comparisons with Australia

Chr.	Pos. (Mbp)	F_{ST}		π			Tajima's *D*		
		AU vs. UK	AU vs. NA	NA	AU	UK	NA	AU	UK
1	26.1	0.258	0.088	0.003	0.002	0.002	-0.23	-0.53	-0.31
2	83.7	0.346	0.281	~0	~0	~0	-0.22	0.43	-0.28
3	11.1	0.251	0.257	0.003	0.002	0.003	1.87	-0.75	1.70
6	6.05	0.211	0.078	~0	~0	~0	-0.86	1.55	-1.62
Z	57.3	0.238	0.274	0.002	0.001	0.002	-1.05	-0.93	-0.66

185

Comparisons with North America

Chr.	Pos.	F_{ST}		π			Tajima's D		
	(Mbp)	NA vs. UK	NA vs. AU	NA	AU	UK	NA	AU	UK
1	107.1	0.283	0.052	~0	~0	~0	-0.08	2.13	-0.08
2	130.7	0.181	0.043	~0	~0	~0	-0.47	0.86	-0.33
3	51.8	0.177	0.169	0.002	0.001	0.001	-0.46	-1.06	-1.49
6	55.5	0.243	0	~0	~0	~0	0.98	0.86	-1.55
Z	64.9	0.228	0	0.002	0.002	0.001	0.03	0.83	-0.79

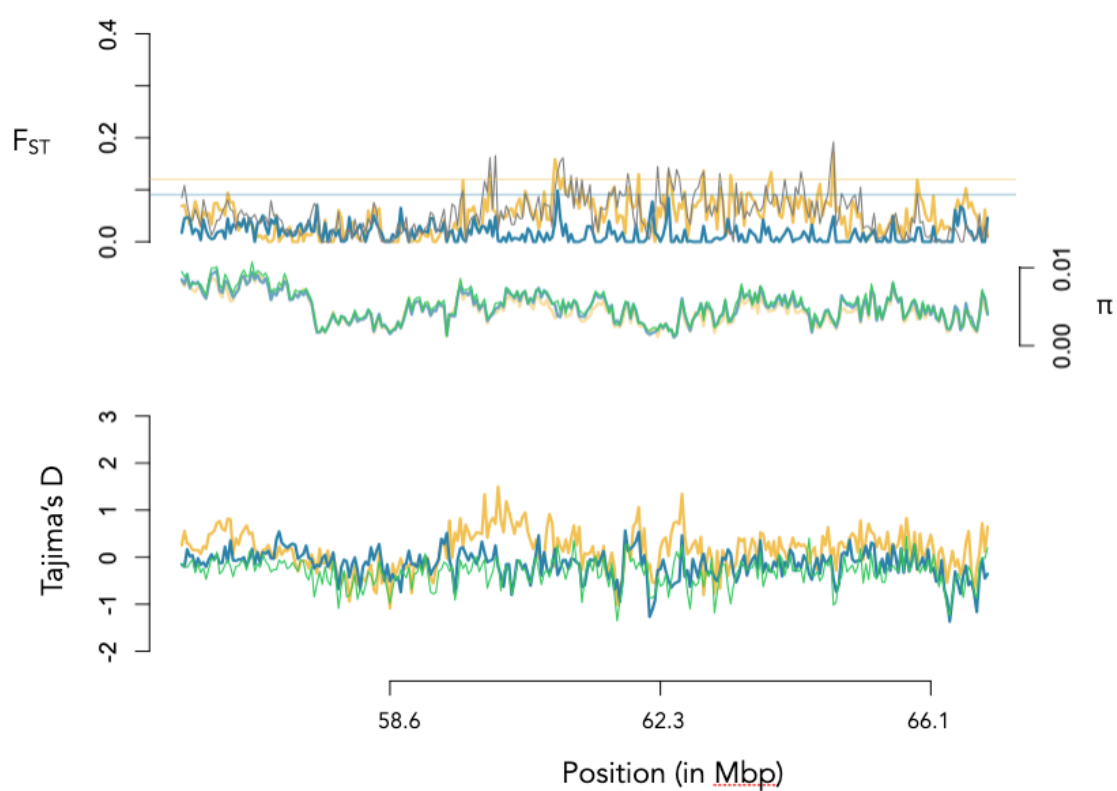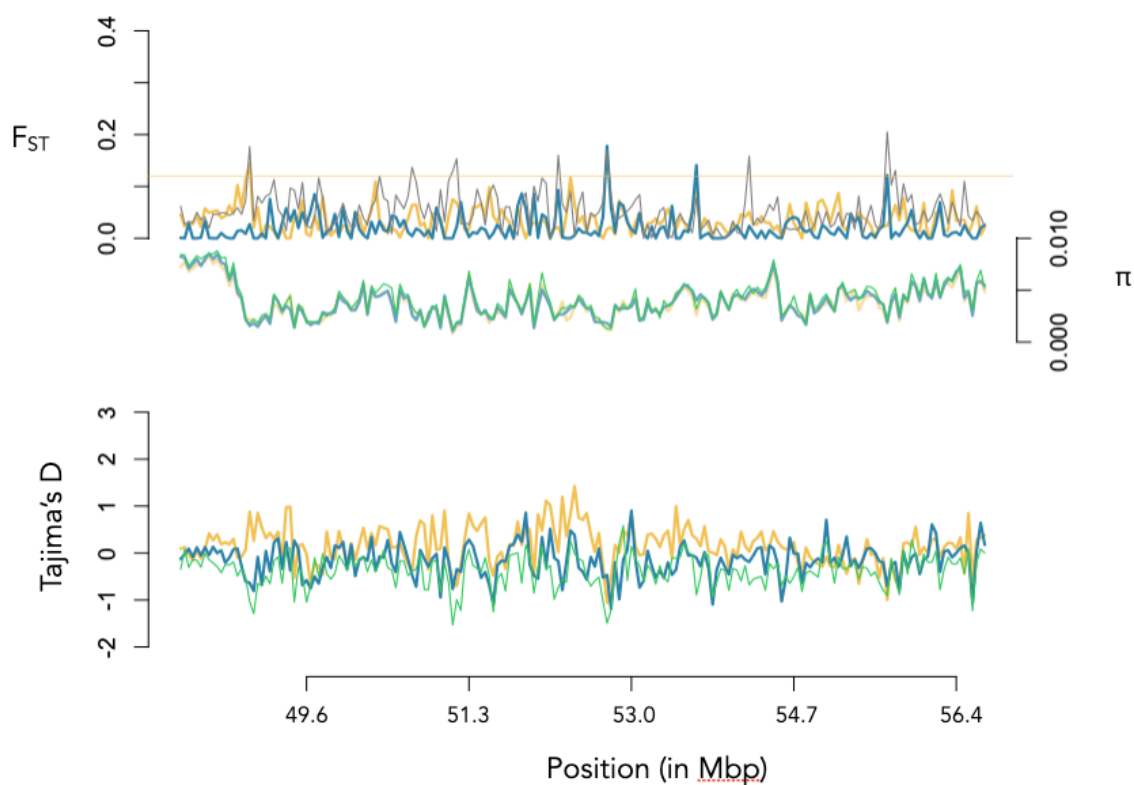**Figure B13.** Region of elevated FST on Chromosome 1. The top row shows a plot of FST,

where yellow indicates differentiation between AU and UK, blue between NA and UK, and grey between NA and AU. The middle row shows nucleotide diversity and bottom row Tajima's D, where yellow indicates the values in the AU population, blue in the NA, and green in the UK population.

Figure B14. Region of elevated FST on Chromosome 3. Colors are the same as in Figure 3.3 and the legend of Figure B13.
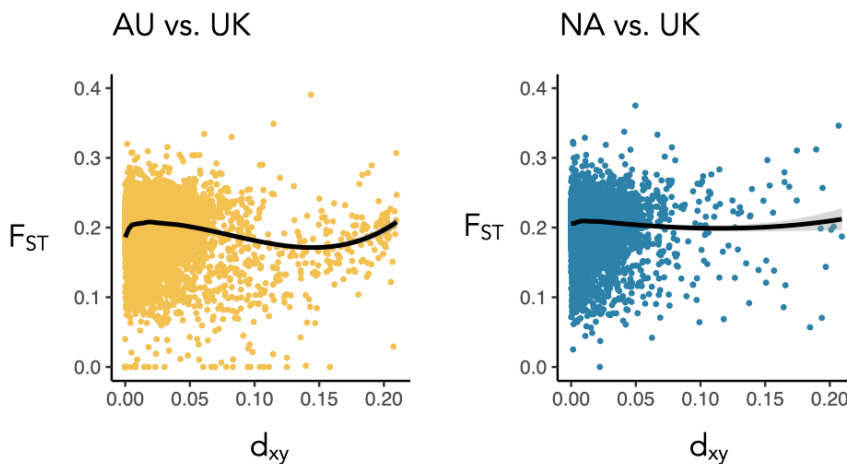
Figure B15. Comparisons of FST and dxy across the genome.

Each point indicates a single 50-Kb window, and the smoothing trendline shows that d_{xy} tends to change approximately linearly with F_{ST} in the comparison between NA vs. UK, and less so in AU vs. UK.

References

Ahrens CW, Jordan R, Bragg J, Harrison PA, Hopley T, Bothwell H, Murray K, Steane DA, Whale JW, Byrne M, et al. 2021. Regarding the F-word: The effects of data filtering on inferred genotype-environment associations. Mol Ecol Resour.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:1–6.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

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.

Linnér RK, Pietro Biroli, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M,

Tino SP, Abdellaoui A, Hammerschlag AR, et al. 2019. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 51:245–257.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv:201178.

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

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes 9:1–7.

Stuart KC, Edwards RJ, Cheng Y, Warren WC, Burt DW, Sherwin WB, Hofmeister NR, Werner SJ, Ball GF, Bateson M, et al. 2021. Transcript- and annotation-guided genome assembly of the European starling. bioRxiv.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9:e112963.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. Genome Research 27:757–767.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328.

2019. Picard Toolkit. Broad Institute Available from: http://broadinstitute.github.io/picard/