

# Contrasting Genetic Diversity of European Starlings (*Sturnus vulgaris*) Across Three Invasive Continental Distributions

Louise Bodt<sup>1</sup>, Lee Rollins<sup>2</sup>, and Julia Zichello<sup>1</sup>

<sup>1</sup>American Museum of Natural History

<sup>2</sup>University of New South Wales

May 5, 2020

## Abstract

European starlings (*Sturnus vulgaris*) represent one of the most widespread and problematic avian invasive species in the world. Therefore, understanding their unique population history and current population dynamics can contribute to conservation efforts and clarify evolutionary processes over short timescales. European starlings were introduced to Central Park, New York in 1890, and from a founding group of about 100 birds, they have expanded across North America with a current population of approximately 200 million. There were also multiple introductions in Australia in the mid-19th century, and at least one introduction in South Africa in the late 19th century. These three independent introductions provide a robust evolutionary system to further understand invasion genetics. In this study, we compare mitochondrial diversity in European starlings from North America, Australia and South Africa, and a portion of the native-range in the United Kingdom. Of the three invasive ranges, the North American population shows the highest haplotype diversity and evidence of both sudden demographic expansion and current expansion. Comparatively, the Australian population shows the highest nucleotide diversity, and only evidence for sudden demographic expansion. We find no evidence of population structure in North America or South Africa, and three distinct haplogroups in the Australian population. Interestingly, none of the invasive ranges share any haplotypes with each other. The only shared haplotypes occur between the native-range and either North America or Australia. This suggests these three invasive populations represent independent subsamples of the vast diversity of the native range, and that novel haplotypes unique to each population may have emerged since introduction.

## Introduction

Here, we use mitochondrial DNA to track the diversity and population structure of the European starling (*Sturnus vulgaris*) in three geographically independent invasive ranges throughout the world. European starlings are native to Eurasia, yet over the last 163 years have been strikingly successful at living outside of its native range in North America, South Africa, and Australia (Jenkins, 1977). They are part of the family *Sturnidae* and order *Passeriformes* (Feare, 1984). Although this group is historically found in the Palearctic, their success as an invasive species and their multiple introductions have led to their existence on every continent, except for Antarctica (BirdLife International, 2016; Sullivan, 2009). Their invasion success likely results from a suite of life-history and behavioral traits that facilitate their ecological flexibility. For example, they are often classified as diet generalists, preferring insects, but they will eat most other foods depending on availability of resources (Cabe, 1993). There is some indication that they are not as successful in urban areas and not complete diet generalists but have been able to adapt to surroundings in order to exploit the environment (Mennechez and Clergeau, 2006). They are cavity nesters utilizing natural features such as crevasses and trees but can also nest in man-made structures allowing them to breed in rural and urban environments (Cabe, 1993). Starlings benefit from associations with other species; they are commensal with humans and are often found around livestock that disturb the soil with their hooves, exposing invertebrates to starlings (East and Pottinger, 1975). They normally lay one or two broods a year, which include between

three and six eggs (Cabe, 1993). Another unique feature that likely plays a role in the ability of the European starlings to expand into new localities is their ability to migrate (Cabe, 1993). Although not all starlings are migratory (e.g. in Australia and New Zealand, Higgins et al. 2006), it has been shown that there is a great deal of variation and that individuals can be differentially migratory from year to year (Cabe, 1993; Feare, 1984).

## Introduction History

European starlings were introduced to North America in 1890 as part of an American Acclimatization Society initiative to populate Central Park with the birds from Shakespeare’s plays (Cabe, 1993). There were many species of birds mentioned in Shakespeare, but not all were able to survive and thrive in the new North American climate. The initial introduction consisted of approximately 60 individuals released in 1890 and 40 more in 1891, leading to a total of ~100 individuals released into Central Park in New York City (Cabe, 1993). From this founding population, starlings have expanded their range across all of North America where their current population exceeds 200 million individuals, over one-third of the global population of this species (Feare, 1984). This range expansion that has taken place in the last 128 years, demonstrating that they are able to rapidly adapt to new surroundings and are incredibly versatile.

Other starling introductions from the 19<sup>th</sup> century have been previously studied using genetics, including the mid-19<sup>th</sup> century Australian introduction (Rollins et al, 2009; Rollins et al, 2011; Rollins et al, 2016) and the late 19<sup>th</sup> century South African introduction (Berthouly-Salazar et al, 2013). During introductions, the often small number of founding individuals may result in a genetic bottleneck and, therefore, these populations are likely to have lower genetic diversity than those in the native range. This was demonstrated using data from the UK and Australia (Rollins et al, 2011). However, because multiple introductions were made to Australia (Jenkins 1959), and these occurred prior to and had a greater number of propagules than the New York introduction, we predict that the genetic diversity of the North American population will be lower than that of Australia. It is also of note that populations of North American and Australian birds increased exponentially following introduction (Bitton and Graham, 2014; Long, 1981), which may mitigate loss of genetic diversity from the founder population. In Australia, population expansion has been limited by large expanses of arid environment, which may have affected population growth and expansion on this continent.

## Why starlings are a problem and control strategies

There are several reasons that non-native starlings are an economic and environmental threat to introduced environments. For example, in the United States, starlings pose a challenge to farms and airports. Starlings are often present in large numbers at dairy feedlots throughout the U.S. to exploit corn-based food resources intended for cattle (Linz, 2007). At airports, they have the potential to interfere with jet engines, cause costly damage to planes and harm human life (Linz, 2007). Efforts to reduce their numbers have been conducted to prevent interference with air travel, including culling these birds close to airports and on farms. Furthermore, starlings cause substantial economic damage to agriculture, estimated at \$800 million in the United States each year (Pimentel et al., 2000). Starlings also carry many diseases (salmonellosis, avian tuberculosis, and chlamydiosis) which have additional effects on agricultural systems and the economy due to the transmission of diseases to livestock (Linz, 2007). Lastly, they have the potential to cause issues for native bird species due to competition for nesting sites and food. Though, studies into the effects of European starlings on native cavity-nesting birds have failed to show a significant effect on native populations, with the exception of sapsuckers (Koenig, 2003).

We use mitochondrial DNA (mtDNA) to characterize the population structure and diversity across North America, and compare this introduction to those in South Africa and Australia. We also compare all three invasions to native range starlings sampled in the UK. European starling populations across the US were previously investigated using allozyme data and no geographic structure to the population was found (Cabe, 1998). Due to the increased resolution of sequence data versus allozyme data, we anticipated that we would detect population structure that may not have been apparent in allozyme data. MtDNA provides a useful genetic marker to study how populations have changed, particularly over short-time periods

(Zink and Barrowclough, 2008). Mitochondrial studies have been the subject of debate because mtDNA is maternally-inherited, haploid DNA that is rapidly evolving, which provides only partial resolution into the evolutionary history of a species (Edwards and Bensch, 2009). However, although mtDNA cannot provide a complete evolutionary picture, it is especially useful as evidence to clarify recent changes in a population (Zink and Barrowclough, 2008). Given the timing of the introduction of starlings to North America, using mtDNA provides an important starting point for elucidating the population history of this recently introduced species. The control region is a non-coding region in the mitochondrial genome that has high nucleotide diversity concentrated in the two hypervariable regions within the control region (Saccone et al., 1991). This hypervariability allows us to track single nucleotide polymorphisms (SNPs) that help distinguish between populations that exist in the native range and identify putatively new haplotypes that are present in the non-native populations of North America, Australia and South Africa. We use these data to characterize the population structure across the North American introduction and revisit the absence of a population structure as previously determined using allozyme data (Cabe, 1998). We also compare population differences between the native range and the three replicated non-native populations. These analyses allow us to observe microevolutionary changes that may have occurred since the introduction of these populations across the world.

## Methods

### Samples and DNA extraction

Tissue samples were obtained from starlings culled by the United States Department of Agriculture Animal and Plant Health Inspection Service (USDA APHIS) between 2011-2018 at 14 localities across the US (Fig. 1, Table 1). Samples were shipped in ethanol and frozen at  $-20^{\circ}\text{C}$  upon arrival.

For fresh samples collected by the USDA, DNA extraction was performed using the Qiagen blood and tissue kit. Extraction were performed at room temperature, with an overnight lysis at  $56^{\circ}\text{C}$  to completely break up muscle tissue. Elution buffer was warmed to  $56^{\circ}\text{C}$  and incubated for 30 minutes before the final spin and elution. Extracted DNA samples were stored at  $-20^{\circ}\text{C}$ .

### Amplification and Sequencing

The primers used to amplify the control region in North American specimens were initially designed to analyze the genetic diversity of the Australian population (Rollins et al, 2011). These two primers (svCRL1 and svPheH3) amplify the majority of the mitochondria control region (Table S1). These primers also were used to amplify DNA from the starling population in South Africa (Berthouly-Salazar et al., 2013). The thermocycling conditions used here were identical to those described in the original paper (Rollins et al, 2011). This includes a 5-min step at  $94^{\circ}\text{C}$ , 30 cycles of  $94^{\circ}\text{C}$  for 30s,  $53^{\circ}\text{C}$  for 15s and  $72^{\circ}\text{C}$  for 30s and a final extension step for 10 min at  $72^{\circ}\text{C}$ .

Rollins et al. (2011) also designed a series of nested primers to be utilized in the amplification of museum specimens or highly degraded samples (Table S1). We utilized these primers in order to sequence the control region in four overlapping segments to eliminate possibilities for error in the sequencing process and verify with confidence the base in each position. All sequences were sequenced in the forward direction. This was necessary due to a repeat region at the end of the sequence that would not amplify in the reverse direction. All sequences were sent to GENEWIZ, Inc (South Plainfield, NJ) for PCR cleanup via an enzymatic purification and to be sequenced. Primer extension sequencing was performed by GENEWIZ, Inc using Applied Biosystems BigDye version 3.1. The reactions were then run on Applied Biosystem's 3730xl DNA Analyzer.

### Population and Expansion Analysis

Contiguous sequences were aligned using the software *Geneious*11.1.2 (Kearse et al., 2012) to generate a consensus sequence for each individual. Furthermore, all subsequent alignments were generated on *Geneious* using the standard settings and the *Geneious*alignment algorithm (Kearse et al., 2012). Median joining haplotype networks were created on *Popart* (Population Analysis with Reticulate Trees) (Leigh & Bryant,

2015). The total sequence length differs for each of the four regions, therefore the data were trimmed to the overlapping segment of 922bp in order to create the median-joining network with data from all four regions (Figure 2). When the full sequences from each data set are aligned, there are fewer similarities than shown in the network (Figure 3).

Fixation indices ( $F_{ST}$  values) were determined using *Genepop* (Raymond & Rousset, 1995; Rousset, 2008). Pairwise nucleotide diversity and haplotype diversity for each population were calculated using *DNAsp* (Rozas et al., 2017). Mismatch analyses, Fu's  $F$  and Tajima's  $D$  were calculated in Arlequin (Excoffier & Lischer 2010) using combined data for each of the three invasions. Sequences from birds from the native range were downloaded from Genbank (Berthouly-Salazar et al., 2013; Rollins et al, 2011). Australian (Rollins et al, 2011; GenBank FJ542126.1–FJ542131.1, FJ542133.1, HQ2636230–HQ263630) and South African (Berthouly-Salazar et al., 2013) sequences were downloaded for comparison (Genbank KF638591–617). The total number of individuals included in the study was 956. The sample size was largest for Australian samples and smallest for the UK (Table 2).

## Results

### Sequences

A portion of the control region (1181 base-pairs) about one-kilobase long was sequenced in 95 specimens from across North America (Table 2). Additionally, 942 base-pair sequences from the same region were downloaded for 597 individuals from Australia (Rollins et al, 2011; GenBank FJ542126.1–FJ542131.1, FJ542133.1, HQ2636230–HQ263630) and the 1092 base-pair sequence from 219 specimens from South Africa (Berthouly-Salazar et al., 2013; Genbank KF638591–617) (Table 2). A total of 20 haplotypes were found with 26 polymorphic sites in the North American population.

When all four populations are included, the Australian and South African populations have a pairwise  $F_{ST}$  value of 0.264, the North American and Australian population have a pairwise  $F_{ST}$  value of 0.224 and the North American and South African populations have a pairwise  $F_{ST}$  value of 0.177 (Table S2). All three non-native populations have pairwise  $F_{ST}$  values closer to 0 when compared with the native population in the United Kingdom (Table S2). When the North American population is separated into three separate populations, Eastern, Middle and Western sampling sites, the  $F_{ST}$  values for are all  $<0.03$  (Table S3).

When all three introduced populations were compared, the sequences were truncated to a 922 bp segment to accommodate the shortest sequence, and these alignments were used to create the haplotype network in Figure 2. When compared there were a total of 70 haplotypes and 42 polymorphic sites (Table S4). The haplotype network for the North American specimens show that the UK native-range haplotypes UKC, UKK and UKE are found across North America in many of the localities included in this study (Figure 4). It also shows that there are a number of novel haplotypes that are present when comparing the North American dataset with both the haplotypes present in Australia and South Africa. In fact, there are no shared haplotypes between any combination of the three introduced regions (North America, Australia and South Africa). All shared haplotypes in the introduced regions are only shared with the UK native-range, with the exception of South Africa, which did not even share any haplotypes with the native-range data here (Fig. 4, S1, S2). When the full sequences from each data set are aligned, there are fewer similarities than shown in this network (refer to figure 2 for overlap).

### Expansion Analysis

The Tajima's  $D$  and Fu's  $F$  values were positive for the Australian and South African populations suggesting that these populations are not currently expanding (Table 2). The Fu's  $F$  value was negative for North America ( $-1.56808 \pm 1.53120$ ) indicating that the North American population is expanding (Table 2). The mismatch analyses for sudden (demographic) expansion was rejected for South Africa (SSD=0.081,  $p=0.04$ ) but not for Australia (SSD=0.107,  $p=0.07$ ) nor for North America (SSD=0.16,  $p=0.06$ ). The analyses for spatial expansion was not rejected for any of the three invasions (South Africa: SSD=0.075,  $p=0.06$ ; Australia: SSD=0.057;  $p=0.26$ ; North America: SSD=0.011,  $p=0.59$ ; Figure 5).

## Discussion

One highly successful invasive species occupying three different continents provides a complex evolutionary system and a unique window into molecular evolution and adaptation. This study provides intriguing preliminary data to continue to explore invasion genetics from a global population perspective. Overall our findings support that low nucleotide diversity, and small founder populations are not obstacles for rapid expansion and establishment in new environments (Rollins et al. 2013). We also find that none of the invasive populations share any haplotypes with each other, which suggests novel haplotypes may have evolved in these three regions since their introduction. This finding also highlights the need for a more comprehensive characterization of the population genetics of the native range for European starlings.

### North American Population

When three populations were defined in North America (Eastern US, Middle US and Western US) there is no distinct population structure based on the  $F_{ST}$  analysis. The comparison of these three regions shows little genetic differentiation (Table 3). These findings are consistent with an earlier investigation of this population, which found that there was no population structure based on allozyme data in the North American starling population (Cabe, 1998). Banding efforts in North America have shown that the starlings are found to migrate in unpredictable ways, not always in the North and South direction, but also in the East and West directions (Brewer, 2000). The absence of the population subdivision in North America is likely due to the high dispersal rates and the unpredictable and latitudinal migration patterns.

Since the introduction of European starlings into North America, novel haplotypes may have arisen, in addition to the native range haplotypes identified here (Figure 4). These haplotypes are UKK, UKC and UKE as described in the native population sampled from Monks Wood, UK (Rollins et al, 2011). However, because our native-range data are limited, it is likely that there are more shared haplotypes between North America and the UK than we have identified. Remarkably little is known about starling population genetics in the native range.

Consistent with our findings based on mitochondrial data, a recent study using genome-wide SNPs to investigate North American starlings did not find evidence of population structure (Hofmeister et al., 2019). However, those data do indicate that there are genotypes associated with specific environmental features such as precipitation and/or temperature. This suggests that over time, population structure may develop in this invasive population, despite apparent high levels of dispersal.

### Comparison to other invasive starling populations

By comparing the three sites of invasive starling populations: Australia, South Africa and North America, we are able to begin to define the differences that have occurred in these three individual introductions. The  $F_{ST}$  values show that Australian and South African populations are most differentiated from one another, Australian and North American populations are moderately differentiated, while South African and North American populations are most similar (Table S2). This is likely due to differences in timing, random sampling from the native range and number of introduction attempts across these three localities, with the Australian introduction occurring earlier than the others and characterized by multiple attempts.

As expected, the invasive populations had lower diversity than the native range, likely caused by genetic bottlenecks at introduction. The highest haplotype diversity and richness is found in the UK (0.954, 30.0), although only 45 individuals were sampled, we found 26 haplotypes. Of the three introduced ranges, the haplotype diversity and richness values are highest in North America (0.884, 14.6) and lowest for the Australian population (0.713, 7.7). The North American population has 17 unique haplotypes, the South African population has 14 unique haplotypes, and the Australian population has 10 unique haplotypes. The North American population had the highest number of novel haplotypes despite having a smaller founding size than Australia and the smallest sample size in this study (Table 2).

Opposite to the patterns we observe with haplotype diversity, nucleotide diversity ( $\pi$ ) in the three invasive populations is highest in the Australian population and the lowest in the North American population. This

could be caused by several factors. Firstly, the Australian population is the result of multiple introductions that began earlier than the other introductions (Jenkins, 1977). Multiple introductions could have brought haplotypes which are differentiated from one another by more nucleotide differences than haplotypes that evolved recently from a single small founding population. Secondly, there is evidence of selection on the mtDNA of the Australian population, which would have driven certain haplotypes to higher frequencies and eliminated others (Rollins et al. 2016). If the Australian population was originally propagated with more nucleotide diversity than the others, selection could explain the pattern of lower haplotype diversity evident today. Lastly, the sample size for the Australian data was 549, for South Africa 219 and for North America only 95 (Table 2). Therefore, although the nucleotide diversity indicates that Australian population has the most genetic diversity, it could be the result of the disparity in sample sizes between the three data sets here.

### Expansion Analysis

Not surprisingly, we found genetic evidence of spatial expansion in all three invasive populations. While there was genetic support for demographic expansion in both North America and Australia, the mismatch analysis of South African data did not support the sudden expansion model (Figure 5). The positive Tajima's D and Fu's F values suggest that the South African and Australian populations are not currently expanding (Table 2). This analysis captures current population dynamics and does not measure expansion history (Tajima, 1989; Fu and Li, 1993). Comparatively, the North American population appears to be currently expanding given the negative Fu's F (Table 5).

Sudden demographic expansion and current expansion in the North American population provides a potential explanation for the discrepancy between haplotype diversity and nucleotide diversity. For example, rapid and recent expansion from a single small founding population may have resulted in many closely related haplotypes. This would have increased haplotype diversity, while nucleotide diversity remained lower than a population with multiple independent introductions such as Australia.

### Geographic similarities

It is also important to note that differences in the environments of each of the three invasive ranges may have influenced expansion rates. In Australia and South Africa, starlings have not expanded to cover the same area that they have in a comparable amount of time in North America. In North America starlings spread from New York to Alaska from 1890-1970, which represents 80 years and a rate of 90km/year (Bitton and Graham 2014). In Australia, starlings rapidly expanded their range into south-eastern Australia and were in Western Australia by the 1970's. However, the center of the continent remains without starlings, due in part to ongoing invasive species management efforts (Rollins 2001). In South Africa, starlings spread primarily eastward from Cape Town, and are now reported only as far north as Kruger National Park (Berthouly-Salazar 2013; Sullivan, 2009). The regions that starlings have seen the most range expansion has been into temperate areas. Interestingly, the starling population in North America is at the same latitude as that of the natural range between about 40° – 55°N, whereas the invasive populations in Australia and South Africa occur at about 30° – 35°S (Sullivan, 2009). The arid center of Australia likely prevented expansion, with the highest temperatures and lowest rainfall associated with the middle of the country (Jones et al, 2009). These distinct climate zones may have differentially impacted invasion both spatially and temporally in each of these three non-native ranges. The center of Australia is classified as an arid, hot, desert with surrounding areas classified as arid, hot, steppe. The same is true for large areas of South Africa (Beck et al. 2019). On the other hand, the United Kingdom and surrounding parts of Europe (native-range) are largely classified as temperate with a hot or warm summer (Beck et al., 2019). The same classification is used for large portions of the northeastern United States near the initial introduction of starlings. This poses an explanation for the rapid and continued expansion of European starlings in North America, and suggests that the success of starlings in South Africa and Australia may have required adaptation to novel climatic conditions (Rollins et al, 2016).

### Implications for Invasive Bird Species

For future analyses, understanding the population structure from across the vast native range of European

starlings may enable the identification of the precise parent population for the US introduction, especially with the use of museum specimens from the native range around the time of the initial introductions. In another invasive bird, the monk parakeet, there was a similar global study on three separate introductions (Edelaar et al., 2015). They were able to pinpoint the parent populations for each invasion using mtDNA and microsatellites (Edelaar et al., 2015). They also found decreased genetic diversity within invasive populations, leading to the conclusion that low genetic diversity does not likely inhibit the success of an invasive species, a conclusion also reached by reviews of this topic (Dlugosch & Parker 2008; Rollins et al. 2013). In order to understand why certain invasive bird populations are more successful than others, it is useful to compare these large-scale invaders to one another. This is not the only study to determine the ancestral origins of an invader. Across a wide range of taxa, studies of mtDNA have successfully determined the ancestral origins of invasive populations (e.g. in birds Jackson et al., 2015; in invertebrates Perdereau et al., 2013). These studies further show that mtDNA data are useful for understanding both the origins of introductions and the role of genetic diversity in invasion success. There is still much to learn about the invasion history of European starlings across the world, but using mitochondrial data to begin to compare these successful invasions can help us to gain a better understanding of the evolutionary history and expansion potential of invasive species. Especially with continued global climate change, closely monitoring invasive species and understanding their outsized adaptive flexibility will be increasingly important in years to come.

### Acknowledgments

The following have been an integral part of this research:

Cecile Berthouly-Salazar, IRD

Rachel Welt, AMNH

Anthony Caragiulo, AMNH

Scott Werner, USDA APHIS

Paul Sweet, AMNH

Peter Capainolo, AMNH

Jeff Groth, AMNH

Science Research Mentoring Program Students, AMNH

### References

- Berthouly-Salazar, C., Hui, C., Blackburn, T.M., Gaboriaud, C., Rensburg, B.J., Vuuren, B.J., and Roux, J.J. (2013). Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Molecular Ecology* 22, 5793-5804.
- Bitton, P-P., and Graham, B.A., (2014). Change in wing morphology of the European starling during and after colonization of North America. *Journal of Zoology* 295 , 254 - 260
- BirdLife International. 2016. *Sturnus vulgaris* . The IUCN Red List of Threatened Species 2016: e.T22710886A87847984. <http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22710886A87847984.en>. Downloaded on 08 April 2018.
- Brewer, D. (2000). Canadian atlas of bird banding *Vol. 1* Doves, Cuckoos, and Hummingbirds through Passerines, 1921–1995 (Canadian Wildlife Service).
- Cabe, Paul R. (1993). European starling (*Sturnus vulgaris* ), version 2.0. in *The Birds of North America*. Cornell Lab of Ornithology, Ithaca, New York, USA.
- Cabe, P.R. (1998). The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity* 80, 519-525.

- Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular ecology* , 17 (1), 431-449.
- East, R., and Pottinger, R. P. (1975). Starling (*Sturnus vulgaris* L.) predation on grass grub (*Costelytra zealandica* (White), Melolonthinae) populations in Canterbury. *New Zealand journal of agricultural research* , 18 (4), 417-452
- Edelaar, P., Roques, S., Hobson, E.A., Goncalves da Silva, A., Avery, M.L., Russello, M.A., Senar, J.C., Wright, T.F., Carrete, M., and Tella, J.L. (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* 24,2164-2176.
- Edwards, S. and Bensch, S. (2009). Looking forwards or looking backwards in avian phylogeography? A comment on. *Molecular Ecology* 18 , 2930
- Excoffier, L. and Lischer H.E.L.(2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 10: 564-567.
- Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* , 133 (3), 693-709.
- Higgins, Peter and M Peter, J & J Cowling, S. (2006). Handbook of Australian, New Zealand & Antarctic Birds. Volume 7 Boatbill to Starlings. Published in two parts:  
 Volume 7A: Boatbill to Larks.  
 Volume 7B: Dunnock to Starlings
- Hofmeister, N.R., Werner, S.J., Lovette, I. J. (2019) Environment but not geography explains genetic variation in the invasive and largely panmictic European starling in North America  
 bioRxiv 643858; doi: <https://doi.org/10.1101/643858>
- Jackson, H., Strubbe, D., Tollington, S., Prys-Jones, R., Matthysen, E., and Groombridge, J.J. (2015). Ancestral origins and invasion pathways in a globally invasive bird correlate with climate and influences from bird trade. *Molecular Ecology* 24, 4269-4285.
- Jenkins, C.F.H. (1977) The Noah's ark syndrome, Western Australia, The Zoological Gardens Board.
- Jones, D.A., Wang, W., Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal* 58 . 233 - 248
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* 28, 1647-1649.
- Koenig, W. (2003). European Starlings and Their Effect on Native Cavity-Nesting Birds. *Conservation Biology*, 17 1134-1140
- Leigh, JW, Bryant D. (2015). PopART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6 , 1110–1116.
- Linz, G.M., Homan, H.J., Gaulker, S.M., Penry, L.B., and Bleier, W.J. (2007). European starlings: a review of an invasive species with far-reaching impacts. *Managing Vertebrate Invasive Species* 24
- Mennechez G. and Clergeau P. (2006). Effect of urbanization on habitat generalists: starlings not so flexible? *Acta Oologica*, 30 , 182-191
- Perdereau, E., Bagneres, A., Bankhead-Dronnet, S., Dupont, S., Zimmermann, M., Vargo, E. L. and Dedeine, F. (2013), Global genetic analysis reveals the putative native source of the invasive termite, *Reticulitermes flavipes*, in France. *Molecular Ecology* 22: 1105-1119.

Pimentel, D., Lach, L., Zuniga, R. & Morrison D. (2000). Environmental and economic costs of nonindigenous species in the United States. *Bioscience* *50*

Raymond M. and Rousset F. (1995). Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, *86* , 248-249

Rollins, L.A., Woolnough, A. P., Wilton, A. N., Sinclair, R. O. N., & Sherwin, W. B. (2009). Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology* , *18* (8), 1560-1573.

Rollins, L.A., Woolnough, A.P., Sinclair, R., Mooney, N.J., Sherwin, W.B. (2011). Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology* *20*, 2307-2317.

Rollins L.A., Woolnough A.P., Fanson B.G., Cummins M.L., Crowley T.M., Wilton A.N., Sinclair R., Butler A., and Sherwin W.B. (2016). Selection on Mitochondrial Variants Occurs between and within Individuals in an Expanding Invasion. *Molecular Biology and Evolution* *33*, 995-1007.

Rollins, L.A., Moles, A. T., Lam, S., Buitenwerf, R., Buswell, J. M., Brandenburger, C. R., ... & Thomson, F. J. (2013). High genetic diversity is not essential for successful introduction. *Ecology and Evolution* , *3* (13), 4501-4517.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* *8* , 103-106.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* *34* , 3299-3302.

Saccone, C., Pesole, G. and Sbisa, E. (1991). The Main Regulatory Region of Mammalian Mitochondrial DNA: Structure-Function Model and Evolutionary Pattern. *Journal of molecular evolution.* *33*. 83-91.

Sullivan, B.L., C.L. Wood, M.J. Iliff, R.E. Bonney, D. Fink, and S. Kelling. (2009). eBird: a citizen-based bird observation network in the biological sciences. *Biological Conservation* *142* : 2282-2292.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* , *123* (3), 585-595.

Zink, R.M. & Barrowclough, G.F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* *17* , 2107-2121

### **Data Accessibility Statement**

DNA sequences: Genbank accessions FJ542126.1–FJ542131.1, FJ542133.1, HQ2636230– HQ263630, KF638591–617

Novel DNA sequences will be uploaded to Genbank

### **Conflict of Interest Statement**

The authors declare that there is no conflict of interest.

### **Author Contributions**

Louise H Bodt (corresponding author): Acquisition, analysis and interpretation of data for the work; Drafting the work; Final approval of the version to be published; Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Lee Ann Rollins: Analysis and interpretation of data for the work; Revising the work critically for important intellectual content; Final approval of the version to be published; Agreement to be accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Julia M Zichello: Substantial contributions to the conception and design of the work; Revising it critically for important intellectual content; Final approval of the version to be published; Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Figure 1** Map of All North American Localities. Precise localities and number of samples are shown in Table 1.

| Locality        | Number of Samples |
|-----------------|-------------------|
| Westchester, NY | 10                |
| Queens, NY      | 10                |
| Albany, NY      | 13                |
| Eglin AFB, FL   | 13                |
| Brandon, NE     | 4                 |
| Ogallala, NE    | 1                 |
| Bruneau, ID     | 3                 |
| Hammett, ID     | 2                 |
| San Angelo, TX  | 5                 |
| Auxvasse, MO    | 5                 |
| Fort Morgan, CO | 5                 |
| Los Angeles, CA | 15                |
| Burbank, WA     | 5                 |
| Juneau, AK      | 4                 |
| Total           | 95                |

**Table 1** Summary of number of specimens analyzed in each locality indicated on the map in Figure 1.

**Figure 2** Median joining haplotype network constructed using 922bp of mtCR sequence and including North America, South Africa, Australia and United Kingdom samples. Population is indicated by color and node size indicates number of individuals. Number of mutations are indicated as hash marks on edges.

**Figure 3** Venn diagram showing the number of haplotypes and the haplotype diversity for the native-range population and the three invasive populations (US, Australia, South Africa). The overlaps indicate the haplotypes in common between two population . There are three native-range haplotypes found in the North American population and five native-range haplotypes in the Australian population and none from South Africa.

|   | North America          | Australia             | South Africa          | United Kingdom        | Other |
|---|------------------------|-----------------------|-----------------------|-----------------------|-------|
| Sample Size                             | 95                     | 597                   | 219                   | 45                    | 95    |
| Number of Haplotypes                    | 20                     | 15                    | 17                    | 26                    | 78    |
| Pairwise Nucleotide Diversity ( $\pi$ ) | 0.00439 $\pm$ 0.00020  | 0.00501 $\pm$ 0.00070 | 0.00474 $\pm$ 0.00017 | 0.00514 $\pm$ 0.00029 |       |
| Haplotype Diversity                     | 0.884 $\pm$ 0.020      | 0.703 $\pm$ 0.011     | 0.781 $\pm$ 0.014     | 0.954 $\pm$ 0.016     |       |
| Haplotype Richness                      | 14.648                 | 7.726                 | 9.950                 | 30.000                | 21    |
| Tajima's D                              | 0.03210 $\pm$ 0.68496  | 3.07904 $\pm$ 1.41219 | 2.06952               |                       |       |
| Fu's F                                  | -1.56808 $\pm$ 1.53120 | 7.20057 $\pm$ 4.38244 | 1.47781               |                       |       |

**Table 2** Comparison of number of haplotypes, nucleotide diversity and haplotype diversity in the UK, North America, Australia and South Africa, based on mtCR sequence data analyses.





