Molecular Causes and Consequences of Sperm Competition in *Agelaius* Blackbirds

by

Irene Ai-Yin Liu

Department of Biology
Duke University

Date:_______________________
Approved:

____
Steve Nowicki, Supervisor

____
Mohamed Noor

____
Susan Alberts

____
Mark Rausher

____
Jerry Wilkinson

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2014

<u>ABSTRACT</u>

Molecular Causes and Consequences of Sperm Competition in *Agelaius* Blackbirds

by

Irene Ai-Yin Liu

Department of Biology
Duke University

Date:_______________________
Approved:

____
Steve Nowicki, Supervisor

____
Mohamed Noor

____
Susan Alberts

____
Mark Rausher

____
Jerry Wilkinson

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biology in the Graduate School of
Duke University

2014

Copyright by
Irene Ai-Yin Liu
2014

Abstract

Sexual selection has long been framed as a process that ends when copulation is achieved. However, in species with polyandry (multiple mating by females), competition persists after mating inside the female's reproductive tract, where sperm from multiple males must then compete to fertilize a female's eggs. This post-mating process, known as sperm competition, is thought to be just as strong as the competition to secure a mate. Because sperm competition has only recently been observed, its evolutionary role remains largely unknown. In this dissertation, I use field, laboratory and computational approaches to understand the evolution of sperm competition in two ways: (1) by testing a possible source of variation in sperm competition within species, and (2) by examining how variation in sperm competition results in DNA evolution across species. My study system is the *Agelaius* clade of New World blackbirds, a group of songbirds with predicted variation in the intensity of sperm competition. In the first half of the dissertation, I explore the factors that affect how intensely sperm competition is experienced in a population. In Chapter 1, I assess the relationship between genetic diversity and extra-pair paternity (EPP, a proxy for sperm competition) in seven continental and one island population of red-winged blackbird (*A. phoeniceus*). I find that while genetic diversity varies significantly across populations, the population with the lowest amount of genetic diversity exhibits similar rates of EPP as the more diverse populations, providing no support for a relationship between genetic diversity and EPP

rate. This result suggests that genetic diversity by itself is not an determining factor in EPP variation. In Chapter 2, I characterize the mating system of the endangered yellow-shouldered blackbird (*A. xanthomus*) and provide the first evidence that it, too, engages in EPP despite having low genetic diversity. I additionally present a conservation genetics profile of the species, showing that the yellow-shouldered blackbird's low effective population size and genetic diversity, both likely due to a recent bottleneck, may be increasing its vulnerability to extinction. I suggest ways in which future management decisions might account for the genetics of a small population. In the second half of the dissertation, I examine whether sperm competition itself can drive the molecular evolution of a species. I focus on the evolutionary patterns of seminal fluid proteins (Sfps), which are transferred with sperm during copulation and are known targets of sperm competition. I describe in Chapter 3 the transcriptomic and proteomic techniques I use to identify protein-coding genes in a non-model organism, presenting the first list of seminal fluid proteins in a songbird. I contrast the protein profile of the blackbird with the protein profile of insect and mammalian Sfps. Finally, in Chapter 4, I use eight of the proteins identified from the list to look for patterns of positive selection on these proteins. Specifically, I test whether Sfps evolve faster in species with mating systems featuring high levels of sperm competition than in species with mating systems featuring low levels of sperm competition. I first compare EPP rates measured from the previous two species with a third species, the tricolored blackbird (*A. tricolor*), and find

that all three experience similar levels of sperm competition. From the catalog of genes derived in Chapter 3, I select, sequence and search for evidence of rapid evolution in six candidate Sfps and two control genes. I find that not only is there no evidence for positive selection in any of these genes, there is strong evidence for purifying selection and furthermore very low levels of diversity within and divergence across species. Reasons for these unexpected preliminary findings could be both microevolutionary or macroevolutionary in nature and warrant larger-scale studies, especially across a broader sample of taxa and across species with greater variation in sperm competition. Taken together, this dissertation describes the relationship between mating systems, sperm competition and post-mating adaptations. By examining the effect of mating system on protein divergence, it links sexual selection with molecular evolution while generating behavioral, genetic, transcriptomic and proteomic resources for future comparative studies.

Dedication

To my family, who started it all

To Russ Greenberg, who led the way

Contents

Abstract........ 6

List of Tables........ 15

List of Figures........ 18

Acknowledgements........ 20

1. Introduction........ 24

1.1 Study system........ 26

1.2 Overview of chapters........ 29

2. Genetic diversity does not explain variation in extra-pair paternity in island and continental populations of a songbird........ 35

2.1 Introduction........ 35

2.2 Methods........ 44

2.2.1 Study system........ 44

2.2.2 Sampling........ 44

2.2.3 DNA extraction and amplification........ 47

2.2.4 Genetic diversity........ 49

2.2.5 Population structure........ 50

2.2.6 mtDNA analysis........ 51

2.2.7 Parentage analysis........ 52

2.2.8 Comparison of EPP rate across populations........ 53

2.3 Results........ 56

2.3.1 Genetic diversity........ 56

2.3.2 Microsatellite structure 60

2.3.3 mtDNA structure 62

2.3.4 Comparison of EPP rate across populations 64

2.3.5 Relationship between genetic diversity and EPP rate 67

2.4 Discussion 68

2.4.1 No difference between EPP in island and continental populations 69

2.4.2 Continental populations vary significantly in EPP rate 73

2.4.3 Adaptive explanations for EPP revisited 76

3. Conservation genetics and mating system of the endangered yellow-shouldered blackbird (*Agelaius xanthomus*) 79

3.1 Introduction 79

3.2 Methods 85

3.2.1. Sample collection 85

3.2.2. DNA extraction and amplification 87

3.2.3. Effective population size 89

3.2.4. Genetic diversity 90

3.2.5 Evidence of bottleneck 92

3.2.6. Parentage analysis 93

3.3 Results 95

3.3.1. Microsatellite quality 95

3.3.2. Effective population size 95

3.3.3. Genetic diversity and heterozygosity 97

3.3.4 Evidence of bottleneck 101

3.3.5. Mating system 102

3.3.6. Parentage assignment 106

3.4 Discussion 109

3.4.1 Low effective population size and genetic diversity in the yellow-shouldered blackbird 110

3.4.2. Mating system does not contribute adversely to N_e 112

3.4.3. Management implications 116

4. Transcriptomic and proteomic identification of seminal fluid proteins in a genetically polyandrous songbird 118

4.1. Introduction 118

4.2. Methods 121

4.2.1. Protein collection and protein identification with zebra finch proteome 121

4.2.2. RNA collection and transcriptome assembly 123

4.2.3. Protein identification with blackbird proteome 126

4.2.4. Differential expression analysis 127

4.2.5. Transcriptome annotation 129

4.3. Results 131

4.3.1. Transcriptome assembly 131

4.3.2. Protein identification with heterospecific and conspecific proteomes 133

4.3.3. Differential expression 137

4.4. Discussion 137

4.4.1. Comparison between proteome searches 138

4.4.2. Comparison with Sfp profiles in other species 140

5. No evidence for positive selection in *Agelaius* seminal fluid proteins 144

5.1. Introduction 144

5.2. Methods 148

5.2.1. Measurement of EPP rate 148

5.2.2. Selection and sequencing of candidate genes 152

5.2.3. Primer design 159

5.2.4. Sanger sequencing and exon concatenation 160

5.2.5. Summary statistics 163

5.2.6. Tests of selection 163

5.3. Results 166

5.3.1. Comparison of EPP 166

5.3.2. Gene polymorphism 170

5.3.3. Tests of selection 172

5.4. Discussion 175

5.4.1. No difference in EPP rates 176

5.4.2. No evidence for positive selection in candidate genes 177

5.4.3. Discussion by gene 179

5.4.4. Slow molecular evolutionary rate in birds 182

Appendix A 187

Appendix B 195

References 196

Biography 222

List of Tables

Table 1: Summary of samples used in this study. 46

Table 2: Genetic diversity compared across seven continental and one island population of red-winged blackbirds.N = sample size, N_r = raw number of alleles, N_s = sample-size-adjusted alleles, I = Shannon diversity index, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient. 57

Table 3: Measures of extra-pair paternity in each population. Asterisks indicate that the data were taken from published measurement of EPP. 64

Table 4: Summary of output for Model 2, the GLM incorporating population as a variable. "Estimate" indicates the maximum likelihood estimate of the intercept for each population. The z-score is the number of standard deviations away from the mean, and the P-value is derived from the z-score. 65

Table 5: Estimates of Ne in the yellow-shouldered blackbird vs. red-winged blackbird from NeEstimator. Estimates from the other two programs (ONeSAMP and COLONY) are reported in the text. 97

Table 6: Genetic diversity across nine loci for 63 yellow-shouldered blackbirds and 66 Bahamas red-winged blackbirds. N_r = number of raw alleles from genotyping, N_s = sample-size-adjusted allelic diversity, I = Shannon diversity index, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient. 99

Table 7: Comparison of genetic diversity at eight loci in one yellow-shouldered blackbird population and eight red-winged blackbird populations. 101

Table 8: Summary of output for Model 2, the GLM incorporating population as a variable, using the number of EPY as a measure for EPP rate. 104

Table 9: Multiple comparison against grand mean for EPP rate, measured by number of EPY. 104

Table 10: Summary of output for Model 2, the GLM incorporating population as a variable, using the number of nests containing at least one EPY as a measure for EPP rate. 105

Table 11: Multiple comparison against grand mean for EPP rate, measured by number of nests containing at least one EPY. 106

Table 12: Maximum-likelihood configuration of EPP broken down by nest. No. EPM = the number of extra-pair males thought to sire chicks in each nest. No. total males = the total number of genetic fathers at the nest. At two nests (101 and 96), the social male did not gain any paternity at his own nest. 108

Table 13: Maximum-likelihood configuration of paternity by ten extra-pair males (EPM) distributed across chicks and nests. For example, the first male fathered three chicks in two separate nests. IDs for the last two males are USFWS band numbers. 108

Table 14: Number of Illumina reads and number of transcripts assembled in Trinity for each tissue. "Leftover" refers to the second Trinity assembly using the unmapped reads that were identified with Bowtie. 132

Table 15: N50 statistics for all eight transcriptome assemblies (two per tissue). RBH = reciprocal best hit. Prop = number of RBH divided by the total number of proteins in the zebra finch reference assembly (n = 18,204). 132

Table 16: Proteins detected from MS-based identification using the zebra finch (ZEFI) *in silico* proteome and the empirical red-winged blackbird (RWBL) seminal fluid sample. Functional categories were determined from searches in Ensembl and NCBI. 135

Table 17: List of candidate and control genes. Transcript IDs are from the testis transcriptome. For each tissue, the approximate number of Illumina reads mapping to each transcriptome assembly is shown (the "expected count" metric in RSEM). SemGlom = seminal gomera. PostFC = posterior probability of fold change in reproductive vs. control tissues as predicted by EBSeq. 157

Table 18: Descriptions of functions for the six candidate Sfps and two control genes (ENO1 and PGK1). 159

Table 19: Predicted distribution of proportions of chicks identified as EPY. The empirical value for in red-winged and yellow-shouldered blackbird is included, as well as the predicted empirical value for tricolored blackbird (in parentheses). 168

Table 20: Predicted distribution of proportion of nests containing at least one EPY. 169

Table 21: Summary of polymorphism data for six candidate Sfps and two control genes in the three *Agelaius* species. ORF = total length of gene (longest open reading frame), Seq = length of sequenced region, Spp = species, n = number of individuals sampled, k = number of unique haplotypes, h = haplotype diversity, S = number of segregating sites, η = number of mutations, π = nucleotide diversity, SD = standard deviation. Genes in gray boxes are from Barker *et al.* (2012). Mean ± SD for π is shown for each species. 171

Table 22: Results of codeml analysis of eight genes across the three *Agelaius* species. The second column shows the test statistic, $(-2(\ln L_1 - \ln L_2))$, applied to each pair of models. All differences are extremely small and correspond to high P-values. For the positively selected sites, * corresponds to P > 70% and ** corresponds to P > 80% as predicted by Bayes Empirical Bayes analysis (Yang et al. 2005). 173

List of Figures

Figure 1: One of two equally parsimonious depictions of the evolution of social mating system of Agelaius blackbirds (Barker et al. 2008). The other tree predicts polygyny independently arising in red-winged and tricolored blackbirds, although Barker *et al.* (2008) argue that the present tree may be more feasible given the geographic distribution of the clade. 28

Figure 2: Map of field sites where red-winged blackbirds were sampled. 45

Figure 3: Bar graph of sample-size adjusted allelic diversity, by population, for each of ten loci. The Bahamas population (leftmost bar) had the lowest allelic richness for nine of ten loci. 58

Figure 4: Boxplot of simulated distributions of average genetic diversity across population. Whiskers show 95% confidence intervals. The interquartile range is the box, and the median is the thick line. No range is shown for Canada, because during each iteration the jackknife sampled 13 individuals, which is the entire sample size at this site. The Bahamas has significantly lower average diversity than the continental populations. 59

Figure 5: Rarefaction curve for allelic diversity. The Bahamas population (lowest curve) was estimated to have a lower maximum allelic diversity than the continental populations. 59

Figure 6: Likelihood plot from delta-K script showing that $K = 2$ is the optimum cluster number for this data set. 60

Figure 7: structure barplot showing proportion of the eight populations, showing the presence of population structure as measured from ten microsatellites. Each bar represents an individual. Color represents proportion of membership in either cluster. 61

Figure 8: Principal Coordinates Analysis showing distinct Bahamas cluster. 61

Figure 9: Neighbor-joining tree showing relationships between the eight populations from mtDNA. 63

Figure 10: Histograms of posterior samples for the eight population-specific intercepts (first eight panels) and the continental random effect mean (bottom right panel). The population-specific intercept histograms are overlaid with vertical lines indicating the maximum-likelihood estimates for μ from Model 2 (see Table 4), whereas the histogram

for the random-effects mean is overlaid with the estimate for the single parameter from Model 1. 67

Figure 11: Scatter plot showing relationship between average genetic diversity and proportion of EPY. The Bahamas population (left-most point) has less genetic diversity than the continental populations, yet its EPP rate falls in the middle of the continental populations' EPP rates. 68

Figure 12: Map of field site. Striped area refers to Boquerón State Forest. The circled region marks the area (Pitahaya) where sampling took place. From López-Ortiz *et al*. (2002). 86

Figure 13: Rarefaction curves for eight red-winged blackbird and one yellow-shouldered blackbird population. 100

Figure 14: Image of SDS-PAGE gel with ladder (left) and blackbird seminal fluid (right). Protein masses of different ladder bands are shown. The gel was cut into 24 sections, indicated by the horizontal lines, and each section was individually analyzed with MS/MS. 122

Figure 15: Distribution of inferred EPP rate across 1000 runs (reconstructed multilocus paternal genotypes). Solid line shows the empirically measured EPP rate for red-winged and yellow-shouldered blackbirds. 168

Figure 16: Distribution of inferred number of nests with EPP across 1000 runs. Solid line shows the empirically measured EPP rate for red-winged and yellow-shouldered blackbirds. 169

Acknowledgements

Junot Díaz writes in *This is How You Lose Her*: "And that's when I know it's over. As soon as you start thinking about the beginning, it's the end." Okay, he was narrating the horrible end of a failed relationship, but I'd argue that sentiment applies at the close of any life chapter. With my nostalgia turned up to full blast, I'd like to thank the many people who made my years at Duke the best I've had so far.

To Steve Nowicki, my advisor, I knew from the first paper Bernie gave me as an undergrad that I wanted to work in your lab. Being your student has been a real honor, and one hell of a ride. It can't have been easy guiding me as my project evolved away from the lab's focus, yet you always knew exactly how to help me face my fears. Thank you for giving me room to explore, question, and fall (hard) on my face. I hope one day I can be as astute and steady a teacher as you.

To the Nowicki lab, thank you for being wonderful colleagues. Susan Peters, you are the best lab mom. I aspire to have your radiant warmth and your ruthless organizational skills. Rindy Anderson, Casey Klofstadt, Rob Lachlan, Kendra Sewall, Jill Soha, and Kim Rosvall, thank you for the encouragement, the lunches and the conversations. It has been a pleasure to count myself as your labmate.

To Mohamed Noor, my unofficial co-advisor, my Ph.D. story wouldn't have been written without you. Literally! I don't know where I would have ended up if it hadn't been for those first few flies, and I definitely wouldn't be here if I didn't have a

bottomless supply of gloves and pipette tips to use. You, Caiti, Brenda, and Suzanne exemplify everything positive and open about this department. I wish for future students the same generosity you have always shown me.

To my committee members, Susan Alberts, Mark Rausher and Jerry Wilkinson, thank you for taking the time to think hard about my project and challenge me to be the best evolutionary biologist I can be. (To Susan especially, thank you for letting me cut my teeth in your lab.) Most of all, thank you for telling me I was ready to graduate!

Being in this department is like being in a giant hug. Thanks to the biograds for your contagious energy and enthusiasm, whether for BPCE or basketball. Thank you to Jon Shaw for letting me set up shop in your lab this past year and Nico Devos for being my bioinformatics buddy. Thanks to Dave Swofford and Steven Wu for additional computational guidance. Thank you to James Abe and Crystal Terry for being my outstanding undergraduate assistants. And to the unshakable staff of Biology: Thanks to Jim Tunney for keeping an eye on my complicated finances, Anne Lacey for keeping the wheels running deceptively smoothly, Jo Bernhardt for the keys and the kindness, Sönke Johnsen for looking out for us as DGS, Randy Smith for the impeccable management of this department, and the IT team (especially Andrew Turnier and Ryan Hardy) for the Mac/PC/Unix help.

People and facilities outside of the department and university were instrumental to this project. Thanks to the Pittsburgh Supercomputing Center and the Duke Shared

Cluster Resource for hosting my computational work. Thanks to the Duke Sequencing, Proteomics, and Microscopy Core Facilities for your expertise and analysis. Thanks to James Johndrow for the statistics help. Thanks to Dave Westneat, Stefan Lüpold, and Ken Yasukawa for the additional blackbird samples.

To my contacts at my various field sites, thank you for taking a chance on a random visiting student. You offered me your homes and field sites with open arms. In Pennsylvania, I thank the Pymatuning Lab of Ecology. In the Bahamas, I am grateful to the Bahamas National Trust, especially Cecilia Bodie, Lisa Wildgoose, Erika Gates, David Cooper. In Washington, I am humbled to have worked on Gordon Orians's old stomping grounds at Columbia National Wildlife Refuge. In Canada, I thank Queen's University Biological Station. In Puerto Rico, I thank Cabo Rojo NWR and the Department of Natural Resources, especially Roxie Medina-Miranda, Ricky López-Ortiz, and Katsí Ramos-Álvarez. In California, I thank Bob Meese and Conaway Ranch. Finally, I thank Ellen Paul of the Ornithological Council for her help with field permits.

I am thankful to have been independently supported during grad school. Thank you to my funding sources, especially the National Science Foundation, Graduate Women in Science, and the Duke Graduate School, that helped me get where I needed to be and then accomplish everything I wanted to do.

And now for some thanks to the people who helped me flourish in grad school. To Matt Johnson and Laurie Stevison, you are my friends and colleagues in every sense.

Let us look fondly back on our grad school days when we are old and gray and filled with more science (and questions) than ever. To my ladyfriends in Women in Science and Engineering, I'm so proud to know and work with you all. Keep being the savvy, ambitious, assertive advocates you are. To my friends from the University Scholars Program—Tori Lodewick, Sarah Heilbronner, Ben Hayden, Abhijit Mehta, Kristine Callan, Jacob Remes, Mari Armstrong-Hough, Brian Smithson, Chris Paul, Sam Bagg—thank you for the great fun and for the reminder (often needed) of the beauty and mystery of the world outside biology.

To the staff at CAPS, especially Jennie Dickson: You changed my life. You gave me the gift of mindfulness, and I hope to cultivate it always.

To the blackbirds, thank you for all the adventures. May this work help keep you safe. And may your inclusive fitness be off the charts.

To Russ, thank you for being my motivation. You are forever my teacher.

Finally, to my family: Thank you for the love and support I am so lucky to know. You've been there for me for all of the beginnings so far. Here's to many more.

1. Introduction

…the other day I got the curious case of a unisexual, instead of a hermaphrodite, cirripede [barnacle], in which the female had the common cirripedal character, and in two of the valves of her shell had two little pockets, in each of which she kept a little husband; I do not know of any other case where a female invariably has two husbands…

(Charles Darwin to Charles Lyell, ca. 1850)

Darwin's account of multiple mating in barnacles is a surprising counterpoint to his portrayal of sexual selection in "The Descent of Man, and Selection in Relation to Sex" (Darwin 1871). In that work, his descriptions of displaying males and choosy females portray sexual selection as a process that ends when copulation is achieved. Implicit in this scenario is the assumption that females mate only once. However, as Darwin's letter to Lyell reveals, females are not always monogamous. In fact, polyandry (multiple mating by females) occurs so routinely in so many species that approaches to sexual selection have been profoundly revised in recent years to accommodate this finding (Parker & Birkhead 2013). It is now known that in species with polyandry, sexual selection does not end at insemination but instead persists within the female's reproductive tract, shifting from the level of the organism to that of egg and sperm. In the same way that males compete for reproductive access to females, sperm from different males compete for fertilization access to eggs. These parallels are intuitive, yet

significant gaps remain in our understanding of how pre- and post-copulatory processes affect each other.

The process known as sperm competition, one of the two components of postcopulatory sexual selection, is thought to be just as strong as the competition to secure a mate. Because sperm competition has only recently been observed, many questions remain about its evolutionary role and its interaction with other forces. What factors affect how intensely sperm competition is experienced by a population? How strong is the link between behavioral and molecular responses to sperm competition? And how can sperm competition (and its counterpart, cryptic female choice) drive the evolution of a species?

I address these questions in the next four chapters using a genus of songbirds predicted to exhibit interspecific variation in sperm competition. Birds are a promising study system with which to explore questions of sperm competition, because their strong foundation in the behavioral ecology literature (especially for extra-pair paternity, EPP) provides the opportunity to examine behavioral determinants of, and responses to, sperm competition. At the same time, molecular profiles can be supplied by DNA, RNA and protein sequencing methods, which have now matured to the stage where they can be applied to non-model systems. The result is the opportunity to survey both genotypic and phenotypic patterns of evolution resulting from sperm competitive

interactions. This introduction will establish the basics of the *Agelaius* study system (common to all four chapters) and provide brief synopses of each chapter.

1.1 Study system

I compare the intensity of sperm competition (in the form of EPP) and its effects on reproductive proteins in three species from the *Agelaius* clade of New World blackbirds. These species belong to a monophyletic clade (Barker *et al.* 2008) consisting of five species, whose social mating systems vary between monogamy and polygyny and whose genetic mating systems are predicted to vary between monogamy and polyandry. (For ease of reference, I use common names throughout the dissertation.) Fig. 1 displays the phylogenetic relationships between species, depicted against one of two equally parsimonious hypotheses for social mating system evolution (adapted from Barker *et al.* 2008). Mating system diversity within this genus reflects the variation across Family *Icteridae*, in which social polygyny has independently arisen six times across 57 species (Searcy *et al.* 1999).

Two species in this clade are socially polygynous. ***A. phoeniceus*** (red-winged blackbird), one of the most common North American songbirds, occurs from Canada to Mexico and the Bahamas. Males defend territories with harems of one to four females, yet multiple mating by females is well established (Searcy & Yasukawa 1995). About 30% of young are sired by an extra-pair male and 50% of nests contain at least one extra-pair young (n = 8, SD = ± 5.4% and 7.6%) (Gibbs *et al.* 1990; Gray 1996; Westneat 1993; K.

Yasukawa pers. comm.; Chapter 2). ***A. tricolor*** (tricolored blackbird), restricted to California, is a colonially breeding species. I provide the first evidence for and measurement of EPP in the species (Chapter 5). Aside from sharing social and genetic mating systems, these two species both display other behavioral traits such as female-biased care and male-only song, and morphological traits such as pronounced sexual dimorphism and female-like (brown) juvenal plumage (Barker *et al.* 2008).

The remaining three congeners are socially monogamous species that occur on Caribbean islands. They share traits such as shared joint parental care, male and female song, reduced or absent sexual dimorphism, and black juvenal plumage (Garrido & Kirkconnell 1996; Whittingham *et al.* 1992; Whittingham *et al.* 1996, Post 1981). These traits are thought to be the ancestral values in icterids (Barker *et al.* 2008; Price 2009). ***A. assimilis*** (red-shouldered blackbird), endemic to Cuba, exhibits the traits described above, including vocal duetting. Its rate of EPP is unknown. The rate of EPP is also unknown for ***A. humeralis*** (tawny-shouldered blackbird), distributed across most of Cuba and a small portion of northwest Haiti. Finally, ***A. xanthomus*** (yellow-shouldered blackbird) is endemic to Puerto Rico. It is listed as endangered with a remaining population of ~400 individuals (see Chapter 3). In Chapter 3, I provide the first evidence for and measurement of EPP in this species.

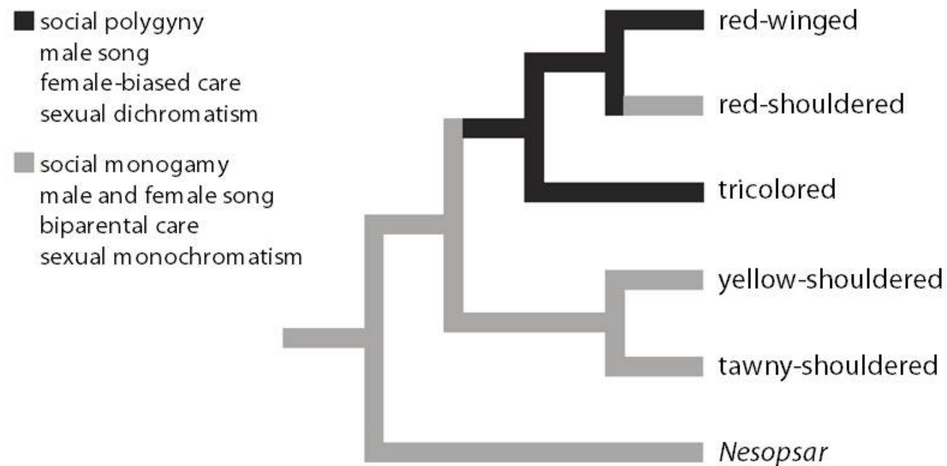

Figure 1: One of two equally parsimonious depictions of the evolution of social mating system of Agelaius blackbirds (Barker et al. 2008). The other tree predicts polygyny independently arising in red-winged and tricolored blackbirds, although Barker *et al.* (2008) argue that the present tree may be more feasible given the geographic distribution of the clade.

The behavior and morphology of the island-dwelling species give reasons to suspect that they may be genetically monogamous. Equal division of parental roles and sexual monomorphism suggest decreased intensity of sexual selection, which may be consistent with the reduced competition for fertilization characteristic of genetic monogamy. However, current knowledge about ancestral versus derived states in *Icteridae* suggests that the correlation between sex-specific traits and sexual selection may not be straightforward. For example, sexual dimorphism may actually be a derived trait, resulting from a loss of elaborate plumage in females that appears to accompany the change to a migratory or temperate-breeding life history (Friedman *et al.* 2009). Likewise, female song in icterids, among many other taxa, appears to be ancestral

(Odom *et al.* 2014). In orioles (genus *Icterus*), migration and temperate breeding predict the loss of this female song (Price *et al.* 2009), although different correlates accompany this loss in other icterid genera, such as the evolution of brood parasitism in cowbirds (genus *Molothrus*) (Price 2009). Given the direction of these changes (i.e., traits presumed to be sexually selected lost in females instead of gained in males), the traits that otherwise suggest genetic monogamy are, in *Agelaius*, unknown predictors of EPP and genetic mating system. Adding to the uncertainty of the island species' genetic mating system are the pervasive disparity between genetic and social mating systems in birds (Griffith *et al.* 2002), combined with conflicting evidence about differences in EPP rates in continental versus island populations and species (Conrad *et al.* 2001; García del Rey *et al.* 2012; Griffith 2000). Therefore, one of the goals of this dissertation was to empirically measure the rate of EPP from wild populations of each species via sampling and paternity analysis.

1.2 Overview of chapters

In **Chapter 2**, I first examine population genetic diversity as a possible cause of variation in EPP. Variation in EPP rate within and across species is a known phenomenon, but the sources of this variation remain unclear (Petrie & Kempenaers 1998). Several hypotheses for the function of EPP propose that females mate with extra-pair males to confer indirect genetic benefits to their offspring (Jennions & Petrie 2000). If these hypotheses are correct, then the genetic diversity of a population from which a

female selects mates could act as a source of variation by altering the benefits of engaging in EPP. I tested the relationship between genetic diversity and EPP using seven continental and one island population of the red-winged blackbird, a species with known intraspecific variation in EPP. I measured genetic diversity over ten microsatellite loci and found that, as predicted, the island population had the lowest genetic diversity at all but the most polymorphic locus, while the continental populations shared similar levels of diversity. Additional structure analyses with multilocus genotypes and mitochondrial ND2 sequences revealed the island population to be a distinct cluster from the genetically indistinguishable continental populations. However, contrary to my prediction that genetic diversity would explain variation in EPP, the island population's EPP rate fell in the middle of the continental populations' distribution, while the continental populations themselves showed significant variation in EPP rates. Thus, the population with the lowest amount of genetic diversity had similar rates of EPP as the more diverse populations, providing no support for a relationship between genetic diversity and EPP rate. This result suggests that genetic diversity by itself is not an determining factor in EPP variation. I discuss the possible interactions of genetic diversity with other sources of variation, such as local differences in payoff, and consider hypotheses for EPP that do not solely implicate female-driven behavior.

In **Chapter 3**, I shift focus to characterize the mating system and the population genetic profile of an endangered *Agelaius* congener. The yellow-shouldered blackbird, endemic to Puerto Rico, is threatened by habitat loss and severe brood parasitism by the invasive shiny cowbird (*Molothrus bonariensis*). The blackbird experienced a bottleneck in the early 1980s and has shown limited recovery following the establishment of an intensive management program. While management efforts incorporate the considerable natural history known for this species, little is known about its genetic profile, including its effective population size, genetic diversity and levels of inbreeding. Additionally, its genetic mating system is unknown but is of potential conservation importance, because mating systems directly affect variation in individual reproductive output and thus the effective population size of a species (Nunney 1993). I characterized the population genetics and the genetic mating system of a breeding population in southwest Puerto Rico. I first found that the blackbirds have low allelic diversity at nine microsatellite loci and a low effective population size (N_e ~ 65, 15% of the 2012 census size). I then tested whether genetic mating system contributes to low N_e, by conducting paternity tests on 30 sampled nests. I found that the yellow-shouldered blackbird engages in extra-pair mating at a rate similar to the red-winged blackbird but that this behavior was not found to contribute significantly to variation in individual reproductive output. This result indicates that the bottleneck is likely to have played a larger role than mating behavior in shaping the current effective population size. I also found little evidence of inbreeding

and no difference between expected and observed heterozygosity. Combined with the findings of low effective population size and allelic diversity, these data suggest that while genetic diversity on the individual level may be intact, the population as a whole remains vulnerable to adverse stochastic events.

In **Chapter 4,** I describe the field, molecular and bioinformatics methods I used to compile the first list of seminal fluid proteins in a songbird. Seminal fluid proteins (Sfps) are know for their rapid rates of evolution, especially in polyandrous species, and are thought to be under strong selection from both male-male competition and male-female antagonistic coevolution. Though these proteins have been documented in many species, they remain poorly characterized in birds, a taxon known for its polyandry but only recently studied at the molecular level. I first sequenced field-collected protein samples of red-winged blackbird seminal fluid and identified proteins using tandem mass spectrometry (MS/MS) by initially searching against a curated zebra finch proteome. To complement this heterospecific search, which may have failed to identify rapidly evolving proteins, I then sequenced field-collected RNA samples of four blackbird tissues (two reproductive and two control) and assembled *de novo* transcriptomes for each tissue. I translated the transcriptomes, searched the peptide mass spectra against these predicted proteomes, and identified proteins using differential expression analyses and an annotation pipeline. A search with the red-winged blackbird testis proteome yielded 157 unique proteins recovered from seminal

fluid, compared with 84 from the zebra finch proteome. Despite the difference in protein quantity, both searches showed similar proportions of proteins distributed across 14 functional categories, with most proteins playing a role in glycolysis and cell respiration. Strikingly, there were only three proteins (acrosin, acrosin-binding protein, and sperm-associated antigen 6) with explicitly reproductive roles. This protein profile is in sharp contrast to the seminal fluid protein profile characterized in mammals (and *Drosophila*), which have multiple proteins involved in fertilization and sperm-egg interactions, and raises questions about the conserved process of reproduction versus the highly specific processes within different taxa.

Finally, in **Chapter 5**, I examine the evolutionary consequences of sperm competition by testing whether polyandry intensifies selection on mechanisms that maximize sperm competitive ability in males. I investigated the relationship between mating behavior, sperm competition, and selection regime on candidate Sfps in the three *Agelaius* species to understand how differences in the strength of sexual selection shape DNA sequence evolution. I tested whether Sfps evolve faster in species with mating systems featuring high levels of sperm competition than in species with mating systems featuring low levels of sperm competition. I first measured the rate of extra-pair paternity (EPP) in three *Agelaius* congeners as a proxy for the intensity of sperm competition. Unexpectedly, all three species exhibited similar levels of EPP, including the socially monogamous yellow-shouldered blackbird. From the catalog of genes

derived from transcriptomic and proteomic analysis, I then selected and analyzed six candidate Sfps and two control genes in these three species. Tests of dN/dS with codon substitution models revealed that not only is there no evidence for positive selection in any of these genes, there is strong evidence for purifying selection and furthermore very low levels of diversity and divergence, although levels of diversity correspond with the census sizes of each species. Reasons for these findings could be both microevolutionary (the Sfps examined in this study could face mixed selective pressures, including selective constraints) or macroevolutionary (that birds have a slower tempo of molecular evolution than other taxa).

2. Genetic diversity does not explain variation in extra-pair paternity in island and continental populations of a songbird

2.1 Introduction

In many songbird species, males and females exhibiting monogamous social behavior often mate with individuals outside their pair bonds. Extra-pair paternity (EPP) has been extensively documented since DNA fingerprinting first enabled researchers to diagnosed mixed paternity in Eastern bluebirds (*Sialia sialis*, Gowaty & Karlin 1984). Across the nearly 90% of surveyed species found to engage in EPP, the frequency of EPP varies both within and across species (Griffith *et al.* 2002; Petrie & Kempenaers 1998). A commonly cited range of extra-pair young is from 0% (e.g., Old World warblers, *Acrocephalus* spp., Gyllensten *et al.* 1990) to 75% (fairy wrens, *Malurus* spp., Muller *et al.* 1994). One of the long-standing questions in EPP research is why this considerable variation exists and what factors might cause it. Identifying the sources of variation in reproductive strategies could greatly improve our understanding of the demographic and selective forces driving the evolution of animal mating systems.

Proposed hypotheses to explain the variation in frequency of EPP typically invoke ecological or genetic factors. In theory, ecological factors such as breeding density and breeding synchrony could predict inter- and intraspecific variation in genetic mating systems, akin to the way that ecologically shaped differences in mate availability and resource distribution have been argued to shape the evolution of social

mating systems (Emlen & Oring 1977). For example, greater breeding density could increase EPP by increasing the local pool of potential extra-pair mates, while greater synchrony could either increase EPP (by also enhancing opportunities for extra-pair copulations, EPCs) or decrease EPP (by restricting the time window for potential EPCs). However, empirical evidence has shown such factors have limited predictive ability of the genetic mating system of a population or species (reviewed in Griffith *et al.* 2002 and Petrie & Kempenaers 1998). A meta-analysis of 72 species showed that breeding density appears to be positively correlated with EPP in an intraspecific, but not interspecific, context (Westneat & Sherman 1997). Similarly, breeding synchrony in red-winged blackbirds (*Agelaius phoeniceus*) is an unreliable predictor of EPP rate because of its interaction with demographic factors such as age (Weatherhead 1997). Overall, these factors fail to explain a significant portion of the observed variation and are considered to be weak predictors of a species or population's frequency of EPP.

Most research has instead turned to considering genetic explanations that could drive variation in EPP rates. While hypotheses focusing on ecological determinants of EPP variation make no distinction between whether the advantages of EPP are direct (to the female) or indirect (to offspring), genetic explanations assume that that its adaptive function is to confer indirect genetic benefits to offspring (Jennions & Petrie 2000). Females are thought to obtain these benefits in one of two ways: by mating with genetically *superior* extra-pair mates to acquire additive genetic benefits (i.e., good genes)

(Johnsen *et al.* 2000; Kempenaers *et al.* 1997; Sheldon *et al.* 1997), or by mating with genetically *compatible* extra-pair mates to acquire non-additive genetic benefits (Løvlie *et al.* 2013; Neff & Pitcher 2005; Pryke *et al.* 2010; Richardson *et al.* 2005; Stapleton *et al.* 2007; Tregenza & Wedell 2000). Each hypothesis generates different predictions for the expected distribution of EPP among females, as well as the distribution of males chosen as extra-pair mates (Griffith *et al.* 2002). Under the good genes hypothesis, females mated to inferior social males are expected to be more likely to engage in EPP with genetically superior males to produce more fit offspring. Under the genetic compatibility hypothesis, females in general are expected to engage in EPP with genetically dissimilar males to maximize clutch diversity. Empirical tests have provided evidence for each of these hypotheses, although support has been inconsistent within and across species (e.g., Bollmer *et al.* 2012; Kleven *et al.* 2006; Wilk *et al.* 2008). In response, alternative frameworks describing EPP as a behavior that is not strictly female-driven, or one that carries more costs to females than benefits, have recently been advanced (Akcay & Roughgarden 2007; Arnqvist & Kirkpatrick 2005; Westneat & Stewart 2003).

If indirect genetic benefits are indeed the primary function of EPP, then population genetic profiles could potentially act as a source of variation in EPP by varying the magnitude of those benefits. Regardless of the mechanism involved (i.e., additive or non-additive genetic benefits), hypotheses for EPP share in common the assumption that benefits to offspring are conferred when females mate with males that

are genetically distinct from their social mate (Petrie & Lipsitch 1994). The probability of a female's finding a suitable extra-pair mate—whether genetically superior or genetically compatible—increases as the population becomes more genetically diverse. (In the case of females engaging in EPP under a good genes hypothesis, there is likely to be an optimum level of population genetic diversity, as excessive variance in male fitness eventually increases the likelihood that a female will encounter genetically inferior males.) By altering the predicted benefits, genetic diversity could be a demographic trait with considerable influence on female mating strategies and thereby a driver of variation in EPP.

Thus, if the benefits of EPP are genetic in nature, a key prediction is that EPP will be more common in more genetically diverse populations. In a population with low genetic diversity, females face low odds of finding a suitable extra-pair mate and high costs that may outweigh whatever benefits EPP offers at all. These costs include the basic time and energetic demands of searching for an extra-pair mate, as well as additional physical or social costs such as male retaliation, male withholding of parental care, or exposure to pathogens (Kulkarni & Heeb 2007; Valera *et al.* 2003; Weatherhead *et al.* 1994; Westneat & Rambo 2000). By contrast, females in a diverse population should experience higher odds of finding a suitable extra-pair mate and lower costs associated with mate searching to give a more favorable cost-benefit ratio. EPP should therefore be

expected to occur more frequently in genetically diverse populations where the costs of finding a genetically distinct male are relatively low.

Two correlational studies examining interspecific variation in EPP have directly examined the relationship between genetic diversity and EPP rate. These studies differed, however, in their predictions for which variable drives the other. The first study (Petrie *et al.* 1998) hypothesized that increasing genetic diversity increases variance in male quality, which drives more females (especially those mated with poor-quality social mates) to seek EPP. In an analysis across 35 species that controlled for body size, sample size and sexual dichromatism, the authors found a positive relationship between allozyme diversity and EPP rate. This result supports the hypothesis that the benefits (and thus frequency) of EPP could be altered by population genetic profiles.

A second study (Gohli *et al.* 2013) took the opposite approach of asking whether genetic diversity is sustained *because* of female promiscuity. The authors sought to distinguish whether EPP is motivated by females searching for good genes (which would result in directional selection for a subset of male genotypes) or for genetic compatibility (which would result in balancing selection across most male genotypes). From a survey of 18 species, they reported a positive relationship between EPP and nucleotide diversity at loci under both neutral evolution and selection (MHC IIB), with diversity especially high at receptor-coding regions. This finding supported the

hypothesis that genetic compatibility was the mechanism driving female participation in EPP. However, the study left unaddressed the question of why EPP varies across species in the first place. Genetic diversity was used as a variable for testing the function of EPP (i.e., the indirect genetic benefits themselves), not for explaining its variation (i.e., the factors that alter those benefits). Instead, existing variation in EPP was taken for granted and compared against levels of genetic diversity to infer the benefits of EPP. Without identifying the variables shaping the relative payoffs of multiple mating, the evolutionary pressures promoting different rates of EPP in different species remain unidentified.

Despite differences in these two studies, with one study implicating genetic diversity and the other implicating EPP rate as the causal variable, both reported a direct relationship between genetic diversity and EPP rate. However, both studies overlooked two additional factors that could complicate the observed trends. First, by comparing only interspecific variation in EPP, these studies did not account for the variation contributed by intraspecific measurements (Garamszegi & Moller 2010; Spurgin 2013). In many cases samples from multiple populations are unavailable, but since within-species variation in EPP is often reported (Petrie & Kempenaers 1998), the error around a single representative measurement needs to be considered. Second, neither study included a species where genetic diversity was expected to be significantly lower. By using a sample set with a relatively narrow range of genetic diversity (and, in the case of

(Petrie *et al.* 1998), using low-resolution allozymes to measure such diversity), the results may not have documented whether EPP is affected by substantial changes in genetic profile. Therefore, comparisons accounting for these shortcomings are needed to expand on current findings.

Intraspecific tests provide a complementary view by testing whether finer-scale differences in genetic diversity, on the population level, also result in differences in EPP rates. In this context, a comparison of island versus continental populations of the same species provides a particularly good test of the relationship between genetic diversity and EPP (Griffith *et al.* 2002; Petrie & Kempenaers 1998). Island populations are well known to exhibit less genetic diversity than continental populations of the same species, due to founder events, smaller effective population sizes, increased effects of drift and higher odds of inbreeding (Charlesworth & Charlesworth 1987; Frankham 1997; Wright 1931). As for EPP, evidence of lower EPP on islands has been reported, with a meta-analysis of 74 populations from 54 species revealing a significant difference in EPP rates (17.6% continental vs. 8.2% island, Griffith 2000). A reason for this difference is that low levels of island genetic diversity lead to decreased variance in fitness and thus lower intensities of sexual selection. Exceptions to this trend have been noted from comparisons within species and between closely related species, although these studies did not additionally examine population genetic diversity (tree swallow, *Tachycineta*

bicolor, Conrad *et al.* 2001; Eurasian blue tit, *Cyanistes caeruleus*, Krokene & Lifjeld 2000; Eurasian blue tit vs. African blue tit, *Cyanistes teneriffae*, García del Rey *et al.* 2012).

To date, only one study has empirically tested the relationship between population genetic diversity and EPP rate between island and continental populations, comparing house sparrow (*Passer domesticus*) populations on mainland England and on an island 20 km off the coast (Ockendon *et al.* 2009). An earlier study had shown that EPP rates on this island were significantly lower than in continental populations (Griffith *et al.* 1999). Ockendon *et al.* (2009) confirmed that allelic diversity at microsatellite loci also was lower in the island population. In the same study, Ockendon *et al.* (2009) then introduced continental adults to the island to test if EPP would increase with experimentally increased genetic diversity. Though EPP rose immediately after the introduction, this increase was not due to island females mating with continental males to increase genetic diversity of their offspring, as was predicted. Instead, EPP increased because island females engaged in more frequent extra-pair mating with island males. Therefore, while the positive relationship between EPP and genetic diversity in this study followed prediction, the actual pattern of mating that led to its occurrence did not.

Evaluating population structure could offer additional insight by improving predictions of population-wide trends, such as mating preference (Pritchard *et al.* 2000). For instance, if the island population of house sparrows were isolated from the continental birds, then selection might favor assortative mating to maintain locally

adapted profiles over population admixture to boost genetic diversity. (Ockendon *et al.* (2009) report occasional migration to the island, suggesting the populations remain connected. However, it is unknown whether the migrants contribute to gene exchange between the two populations, especially considering that in the study, females chose island males as both within- and extra-pair mates.) Genetic differentiation could also contribute to divergent selective regimes experienced by the island versus continental population (Falconer & Mackay 1996). Therefore, examining the population structure of the island and continental populations could be important for studies using either correlative or experimental methods to test how populations adjust their reproductive strategies in response to changes in genetic landscape.

The present study investigates the relationship between levels of genetic diversity and EPP in seven continental and one island population of red-winged blackbird (*Agelaius phoeniceus*). EPP rates in this species have been extensively documented and are known to vary across continental populations (reviewed in Searcy & Yasukawa 1995), although it has never been documented in island populations. Genetic diversity has been compared in the MHC IIB region with individuals from a single population (Edwards *et al.* 1998; Gasper *et al.* 2001), but not with microsatellite data across multiple populations. Here I provide measures of genetic diversity for all populations. I also characterize EPP in an island (Bahamas) and both previously studied and unstudied continental populations. I then test for population structure across

populations. An early study found that continental populations do not show population structure (Ball *et al.* 1988), typical of New World birds that radiated across North America after the glacial retreat of the Pleistocene. However, it is unknown whether this pattern extends to island populations. I ask the following questions: (a) What are the levels of genetic diversity and EPP of each population? (b) Is there population structure distinguishing the continental populations from the island population? (c) Does population genetic diversity predict rates of EPP?

2.2 Methods

2.2.1 Study system

Red-winged blackbirds are a widespread species in North America, occurring from Canada to Costa Rica on the continent and on the Bahamas and western Cuba in the Caribbean (Yasukawa & Searcy 1995). They exhibit a socially polygynous mating system, in which males establish territories associated with one to four females (Orians 1969), but are genetically polyandrous (Bray *et al.* 1975; Searcy & Yasukawa 1995; Westneat 1993). Females frequently raise broods of mixed paternity, and EPP occurs despite male countermeasures such as repeated copulation and mate guarding (Searcy & Yasukawa 1995; Yasukawa & Searcy 1995).

2.2.2 Sampling

I analyzed data from seven continental populations and one island population (Fig. 2).

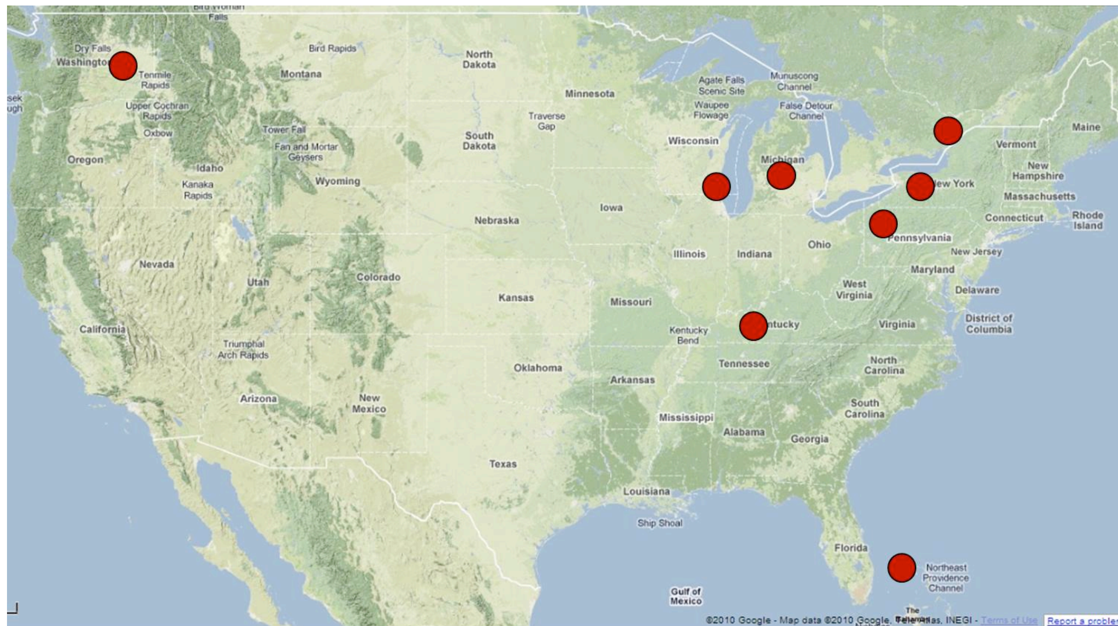

Figure 2: Map of field sites where red-winged blackbirds were sampled.

For three populations (Kentucky, New York and Washington), I used published EPP rates in my comparison (Gray 1996; Westneat 1993; Westneat & Mays 2005). For two populations (Wisconsin and Ontario, hereafter "Canada"), I used EPP rates from subsets of data shared by researchers (P. Weatherhead & K. Yasukawa, pers. comm.). Note that the data for Canada give rates of EPP that are different from published rates, which included data from additional years (e.g., Gibbs *et al.* 1990; Weatherhead & Boag 1995). To measure genetic diversity, I obtained blackbird blood samples either from other researchers or by collecting them myself from the study sites where EPP had previously been measured. Although it is ideal to measure EPP and genetic diversity during the same year to avoid potential confounding results from temporal variation (Petrie &

Kempenaers 1998), the constraints of seasonal field collection made it more practical to use readily available EPP data from past studies.

For the Michigan population, I analyzed samples collected by S. Lüpold in 2004 and 2005. For the Pennsylvania population, I collected samples at Conneaut Marsh in northwest Pennsylvania in 2009 to measure both EPP and genetic diversity. Finally, for the island population, I sampled red-winged blackbirds in 2011 around Freeport, Grand Bahama Island, 90 km from the coast of Florida. Sampling took place in mangroves in Lucayan National Park, Grand Bahama Shipyard, and the township of Pine Dale west of Freeport. Table 1 summarizes the origin of samples used for this study.

Table 1: Summary of samples used in this study.

Country/state	Site	Year samples collected EPP	Year samples collected diversity	Reference for EPP
Bahamas	Grand Bahama Island	2011	2011	Present study
Canada	Queen's University Biological Station	1987-89	2011	P. Weatherhead, pers. comm.
Kentucky	Muhlenberg County	1994-7	1996	(Westneat & Mays 2005)
Michigan	Unknown	2004-05	2004-05	S. Lüpold, pers. comm.
New York	Cornell University Experimental Ponds	1988-89	1991	(Westneat 1993)
Pennsylvania	Conneaut Marsh	2009	2009	Present study
Washington	Columbia National Wildlife Refuge	1990-92	2010	(Gray 1996)
Wisconsin	Newark Road Prairie	1992-94	2009	Yasukawa, pers. comm.

For samples I collected, I captured adults using mist nets, grain-baited walk-in traps, or walk-in traps placed over nests. I bled adults from the brachial vein using sterile 26G × ½ in. BD PrecisionGlide needles and collected five drops of blood (100 μl) onto Whatman FTA bloodstain cards treated with 1M EDTA. I then banded adults with USFWS numbered metal bands and three plastic colored leg bands. Chicks were bled between 0-7 days post-hatch.

Territory owners were assigned by behaviors such as singing and defense against intruders. Females in each harem were identified by their association with the territorial male. Occasionally territory assignments of females and chicks were unclear, especially when nests of females were on the boundary of two males' territories and were defended by both males when I approached. Because determination of EPP hinges on reliable identification of the social male, these individuals were discarded from my analysis.

2.2.3 DNA extraction and amplification

I extracted DNA from dried blood with a Qiagen DNeasy Blood and Tissue Kit and evaluated DNA concentration and purity with a Nanodrop spectrophotometer. Samples with poor concentrations (<4.0 ng/μl) were re-extracted.

To genotype individuals, I amplified ten microsatellite loci that were either known to be polymorphic in red-winged blackbirds, or polymorphic in other species and successfully used in red-winged blackbirds in this study: Aph54 (Westneat & Mays

2005), FhU2 (Primmer *et al.* 1996), LTMR6 (McDonald & Potts 1994), Qm10 (Hughes *et al.* 1998), Pca3 (Dawson *et al*. 2000), Dpμ16 (Dawson *et al.* 1997), Ap79, Ap107, Ap144, and Ap146 (Barker *et al.* 2011). For each individual, I ran three multiplex PCR reactions, the first two containing four primer pairs and the third containing two primer pairs. The forward primer in each pair was fluorescently labeled with 6-FAM, HEX (Sigma-Aldrich) or NED (Applied Biosystems). Reactions consisted of 2.0 μl of DNA, 3.0 μl of Qiagen Type-It Multiplex PCR Master Mix, 1.6 μl RNAse-free water, and 1.0 μl of 100 μM primer mix. PCR cycles were initiated at 95°C for 5 minutes to activate the HotStarTaq Plus DNA polymerase, followed by ten touchdown cycles from 60°C to 50°C and 28 additional cycles at 50°C. Each cycle consisted of denaturation at 95°C for 0:30, annealing for 1:30, and extension at 72°C for 0:30. The final extension was at 68°C for 10 minutes.

Plates were processed using Applied Biosystem 3730xl DNA Analyzers, and genotypes were scored with GeneMarker v.1.8 (SoftGenetics, State College, PA) using size standard GS-500 to determine allele sizes. Homozygous loci were genotyped at least twice to account for the possibility of allelic dropout. I was unable to use Microchecker to scan for null alleles or dropout, because many loci had irregular alleles outside the intervals expected from the motif. These genotypes were verified with multiple runs as genuine alleles and not artifacts of pull-up or stutter.

2.2.4 Genetic diversity

For the adults in each population (range = 13 to 66), I calculated the raw number of alleles; observed and expected heterozygosity; Shannon diversity index; and the inbreeding coefficient (F_{IS}) using GenAlEx v.6.501 (Peakall & Smouse 2006; Peakall & Smouse 2012). Because the raw number of alleles depends on sample size, I separately calculated sample-size-adjusted allelic diversity using the jackmsatpop function of the R package PopGenKit v.1.0 (Paquette 2012). This function uses Genepop input files to determine allelic diversity for a given sample size. The program sampled 13 individuals (corresponding to the lowest *n*, from Canada) per iteration for 100 iterations.

I also used the jackmsatpop function to generate a rarefaction curve predicting cumulative population allelic diversity. This function measures the number of sampled alleles for a given constant increase in sample size for each population. The results can then indicate whether sampling was sufficient to capture population allelic diversity, as well as whether different populations have different maximum allelic diversities. For each population, I ran 100 repetitions using a stepwise increase of one individual up to that population's sample size.

I ran an ANOVA to measure variation in sample-size-adjusted genetic diversity averaged across all loci, first within the continental populations and then including the Bahamas population. Next, I examined variation in genetic diversity within individual loci. Because allelic diversity varied significantly at every locus (see Results), post-hoc

tests were needed to identify which populations had significantly different levels of mean allelic diversity from each other. However, because the jackmsatpop function did not provide the output from the 100 iterations it used to calculate sample-size-adjusted genetic diversity, Tukey post-hoc tests and other pairwise comparisons were not possible. Therefore, a mock data set was generated with R to simulate these runs and record the data for each run (J. Johndrow, pers. comm.). One thousand jackknife resamples were produced to "create" the final results from jackmsatpop. From these data, 95% confidence intervals were calculated for each estimate, and a boxplot was generated. Populations with non-overlapping confidence intervals were determined to be significantly different from each other.

2.2.5 Population structure

I conducted a cluster analysis using *structure* v.2.3.3 (Pritchard *et al.* 2000). For *K* clusters from 1-8, ten replicate runs were performed, each with a 100,000 generation burn-in followed by 1,000,000 generations. I then constrained the number of clusters to two, generated a single Q-matrix for *K* =2, and used CLUMPP v.1.2.2 (Jakobsson & Rosenberg 2007) to summarize and align clusters. Cluster assignment and admixture were visualized with custom R scripts, one of which calculated delta-*K* to evaluate the fit of each *K*-value (Evanno *et al.* 2005; M. Johnson, pers. comm.). Additionally, I generated a distance matrix in GenAlEx and ran a Principal Coordinates Analysis (PCoA), based

on pairwise genetic distances, to visualize the genetic relationships between populations.

2.2.6 mtDNA analysis

To characterize divergence between the Bahamas and the continental populations, I sequenced the ND2 region for 14 Bahamas birds and 14 continental birds (two from each of the seven study populations). Because of its length, the gene was split into two pieces and amplified with two primer sets, L5216-H5766 and L5758-H6313 (M. Sorenson, http://people.bu.edu/msoren/Bird.mt.Primers.pdf). Each primer pair amplified a ~500-bp fragment. For the initial PCR, reactions consisted of 2.0 μl of DNA, 8.9 μl of distilled water, 2.0 μl of 10X buffer, 3.2 μl of dNTPs, 1.0 μl each of 10 μM forward and reverse primer, 1.5 μl of bovine serum albumin (BSA), and 0.4 μl of Taq (Denville Scientific). PCR cycles were initiated at 95°C for 5 minutes, followed by 12 touchdown cycles from 58°C to 52°C. Poor results for the second primer pair (L5758-H6313) were repeated using touchdown cycles from 60°C to 54°C. Touchdown cycles were followed by 28 additional cycles at 52°C (or 54°C). Each cycle consisted of denaturation at 95°C for 0:30, annealing for 0:30, and extension at 72°C for 1:00. The final extension was at 72°C for 7 minutes.

I ran gels after each reaction to verify successful amplification, then purified the DNA template with ExoSAP. To each template I added 2.6 μl of distilled water, 0.2 μl of exonuclease I (ExoI), and 0.2 μl of shrimp alkaline phosphatase (SAP). The reaction was

initiated at 37°C for 30:00, followed by 80°C at 15:00 to deactivate ExoI. Plates were processed by Eton Bioscience and edited in Sequencher (Gene Codes).

Sequences from each primer pair were aligned in MEGA v.5.2 (Tamura *et al.* 2011) and trimmed with PhyDE v.0.9971 (Müller *et al.* 2010). The two pieces were then merged using a custom Python script (M. Johnson, pers. comm.). There was no overlap between the two sequences, indicating a middle portion of the gene was left unsequenced and that the reading frame was likely different for the two sequences. However, as the goal was to align sequences and not to analyze coding regions, this uncertainty was not problematic. The complete NEXUS file was imported to PAUP v.4.0a129 (Swofford 2003), and a neighbor-joining tree was constructed. Finally, mean between-group distances were calculated for Bahamas vs. continental individuals in MEGA.

2.2.7 Parentage analysis

To determine extra-pair paternity across nests, I compared genotypes of the social father and offspring to identify allelic incongruities. All inconsistencies involved at least two of the loci, minimizing the possibility of mistaking occasional single-locus mutations for genetic mismatches (Westneat & Mays 2005). An exclusion analysis on GenAlEx confirmed that, with both parents' genotypes available, the probability of paternity exclusion for all populations except Canada and the Bahamas reached 100% with only four of the loci used, thus increasing the confidence of exclusion when

considering all ten loci. Canada and the Bahamas required five and eight loci, respectively, to reach 100% confidence. I measured the frequency of EPP by calculating two proportions: the number of extra-pair young (EPY) out of the total number of chicks, and the number of nests containing least one EPY out of the total number of nests.

2.2.8 Comparison of EPP rate across populations

I first used chi-square tests to assess variation in EPP proportions (measured separately by number of EPY and by nests with EPY) across the seven continental, and then across all eight, populations. I then performed equivalent hypothesis tests using generalized linear models to test which models best explained the observed variation in EPP (measured by number of EPY). Specifically, logistic regression models were appropriate because the response variable (extra-pair versus within-pair young) was binary in nature. Finally, I ran a Bayesian random-effects model to compare results from treating the continental populations as fixed versus random effects. Analyses were performed in R v.3.0.2 (R Development Core Team 2013).

The simplest GLM (Model 1) tested whether there was significant variation in EPP across populations at all, by considering the possibility that variation in EPP rate was explained by a single intercept. This model can be written as

$$Pr(EP_i = 1) = \frac{e^{\mu}}{1 + e^{\mu}},$$

which is a logistic regression with a single parameter μ. The parameter μ is related to the probability that a chick is an EPY, where EP_i indicates that chick i is extra-pair. Determining the maximum likelihood estimate of μ given the data is the equivalent of performing an ANOVA for binary response with only a grand mean.

Next, I tested whether the model could be improved by adding population as a variable, thus allowing the intercepts, and therefore the probability of being EPY, to differ across populations (Model 2). The model can be written as

$$Pr(EP_i = 1 | x_i) = \frac{e^{\mu x_i}}{1 + e^{\mu x_i}},$$

where x_i is a categorical variable that takes one value for each population, and will produce eight population-specific values of μ instead of a global intercept. I then evaluated the fit of these two models using a likelihood ratio test, similar to an ANOVA but using a χ^2 distribution to determine significance, to calculate the difference in the deviance of the two models and test if the data justified the more complex Model 2.

Finally, I tested whether the Bahamas' EPP rate was significantly different from the average EPP rate across all continental populations (Model 3). This regression was estimated by re-pooling the continental populations into a single group and testing them against the Bahamas.

Based on this analysis (see Results), the models showed that (a) EPP rates differed significantly across populations and that (b) the Bahamas EPP rate was not significantly different from the continental average rate. However, conclusion (b) was

based on a model (Model 3, in which all continental populations were considered jointly) that would be rejected relative to the model that supported conclusion (a) (Model 2, in which continental populations were considered separately). To be certain that conclusion (b) was correct, and to confirm the Bahamas population's EPP rate fell within the distribution of the continental populations' EPP rates, an additional analysis with random-effects models was conducted using a Bayesian approach.

The random-effects model hypothesized that there was some underlying process generating the difference across populations, and that the population-specific means themselves were effects of this process. In other words, there are obviously more red-winged blackbird populations than the eight in this study, and the prediction in question was whether the Bahamian population's EPP rate differed significantly from some central measure of the continental rate. However, in the previous comparison between the Bahamas and continental populations' EPP rates (Model 3), the model was constrained such that the continental populations were restricted to having the same intercept. A Bayesian random-effects model allowed for both different intercepts for all the populations *and* a central measure of the EPP rate for the continental population, given by the random-effects mean. The Bahamas EPP rate could then be compared to all of these measures.

The model was given hierarchically by

$$Pr(EP_i = 1|x_i) = \frac{e^{\mu_{x_i}}}{1+e^{\mu_{x_i}}} \quad \mu_{2:8} \overset{iid}{\sim} N(\mu_{\text{Continental}}, \sigma^2_{\text{Continental}})$$

$$\mu_1 = \mu_{\text{Bahamas}} \sim N(0, \sigma^2_{0B}) \quad \mu_{\text{Continental}} \sim N(0, \sigma^2_{0C}) \quad \sigma^2_{\text{Continental}} \sim \text{Gamma}(a, b),$$

where $\mu_{2:8}$ are the population-specific intercepts for the seven continental populations and μ_{Bahamas} is given the designation μ_1 for convenience. Thus, the seven continental population intercepts were hypothesized to be *iid* realizations from the common distribution $N(\mu_{\text{Continental}}, \sigma^2_{\text{Continental}})$, which controls all extra-pair mating rates throughout continental North America. To place priors on the parameters $\mu_{\text{Continental}}$ and $\sigma^2_{\text{Continental}}$, the suggested parameters from Gelman et al. (1995) was used, in which the "weakly informative" prior $N(0,9)$ was chosen for $\mu_{\text{Continental}}$, and a and b were both set to equal 2, a standard weak prior on the variance. As an isolated population, the Bahamian intercept μ_{Bahamas} did not follow the random effect distribution and was thus given the independent prior $N(0, \sigma^2_{0B})$ with $\sigma^2_{0B} = 9$. The model was estimated using Markov Chain Monte Carlo and the latent variable method of (Polson *et al.* 2013) and run for 10,000 iterations after 1000 iterations burn-in.

2.3 Results

2.3.1 Genetic diversity

The continental populations did not vary significantly in average genetic diversity across all ten loci (ANOVA on sample-size-adjusted allelic richness (N_s), F = 0.39, df = 6, P = 0.88, Table 2). However, when the Bahamas population was added, this variation became significant (F = 7.66, df = 7, P < 0.0001). The Bahamas population also

had lower population diversity as measured by the Shannon diversity index (continental only: $F = 0.29$, df = 6, $P = 0.94$; all: $F = 2.84$, df = 7, $P = 0.007$) and lower individual diversity as measured by observed heterozygosity (continental only: $F = 0.64$, df = 6, $P = 0.70$; all: $F = 4.1$, df = 7, $P < 0.0001$). By contrast, the Bahamas population was not significantly more inbred than the continental populations ($F = 1.44$, df = 7, $P = 0.19$).

Table 2: Genetic diversity compared across seven continental and one island population of red-winged blackbirds.N = sample size, N_r = raw number of alleles, N_s = sample-size-adjusted alleles, I = Shannon diversity index, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient.

Site		N	N_r	N_s	I	H_o	H_e	F_{IS}
Bah	**Mean**	66	10.6	7.07	1.71	0.72	0.73	0.005
	SE		2.78	1.32	0.24	0.05	0.05	0.014
Can		13	10.8	10.8	2.07	0.84	0.83	-0.015
			1.09	1.09	0.14	0.04	0.03	0.032
KY		32	16.1	11.0	2.31	0.83	0.85	0.012
			2.31	1.17	0.18	0.03	0.03	0.018
MI		51	17.8	10.8	2.36	0.81	0.85	0.050
			2.71	1.11	0.19	0.03	0.03	0.019
NY		31	15.3	10.8	2.30	0.82	0.85	0.043
			1.88	1.06	0.16	0.03	0.02	0.026
PA		60	19.1	10.9	2.41	0.86	0.87	0.009
			2.85	1.12	0.17	0.02	0.02	0.015
WA		31	13.3	9.92	2.17	0.81	0.83	0.030
			2.01	1.16	0.18	0.03	0.03	0.016
WI		22	14.2	11.2	2.30	0.86	0.86	-0.005
			1.93	1.27	0.17	0.04	0.03	0.028

The trend of lower allelic diversity persisted when diversity was examined at individual loci. The Bahamas population had the fewest alleles at every locus except one (Ap107, Fig. 3). However, allelic diversity varied sufficiently across populations that every locus was determined to have significant variation in diversity (ANOVA, $P < 0.001$

for all loci). A boxplot of the simulated runs from jackmsatpop revealed the Bahamas was the primary driver of this variation, possessing by far the lowest average diversity (Fig. 4). Likewise, the cumulative curve showed that the Bahamas population plateaued in average genetic diversity at 10.6 alleles (Fig. 5). This maximum was lower than any other population's, including Canada. At only 13 samples, Canada did not approach its own plateau, but it was following the same trajectory as the other continental populations and likely had comparable allelic diversity (mean plateau = 16.0 ± 0.9 alleles).

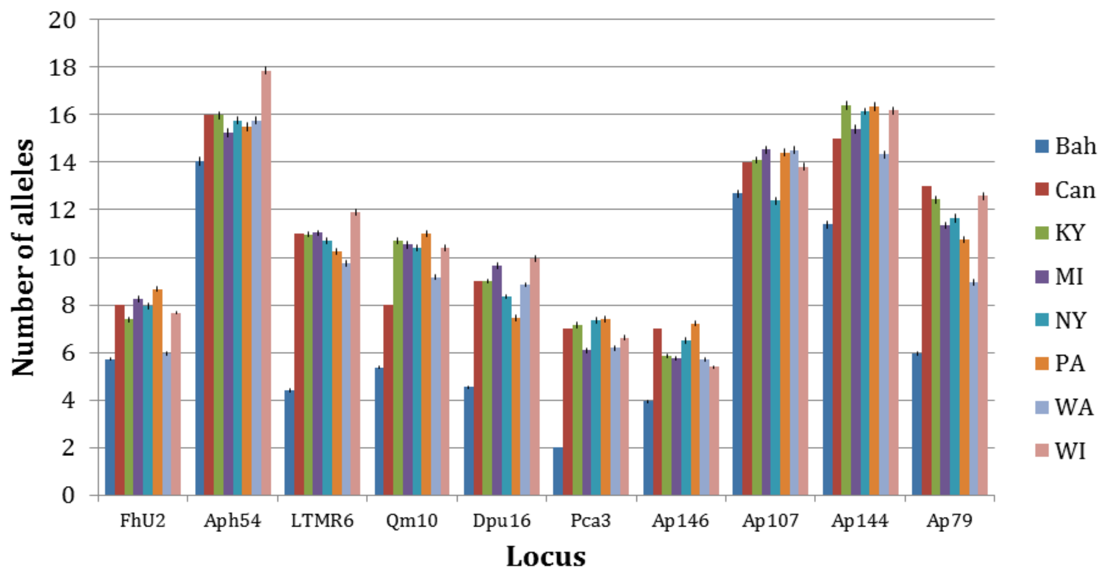

Figure 3: Bar graph of sample-size adjusted allelic diversity, by population, for each of ten loci. The Bahamas population (leftmost bar) had the lowest allelic richness for nine of ten loci.

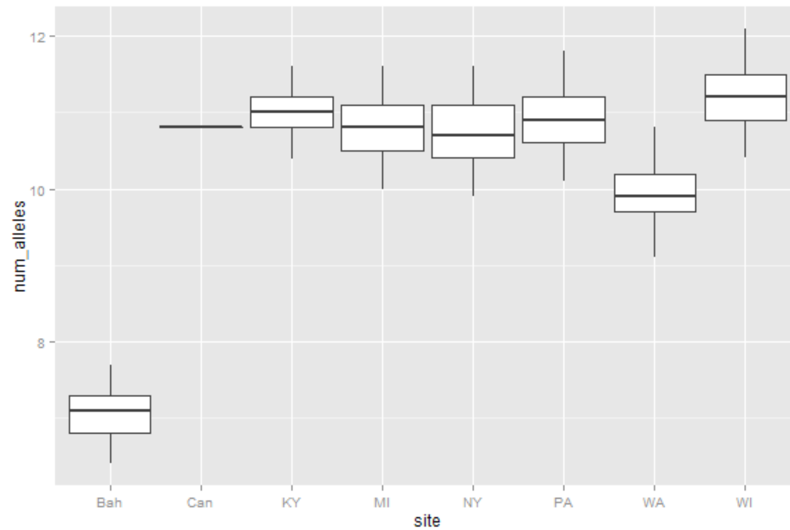

Figure 4: Boxplot of simulated distributions of average genetic diversity across population. Whiskers show 95% confidence intervals. The interquartile range is the box, and the median is the thick line. No range is shown for Canada, because during each iteration the jackknife sampled 13 individuals, which is the entire sample size at this site. The Bahamas has significantly lower average diversity than the continental populations.

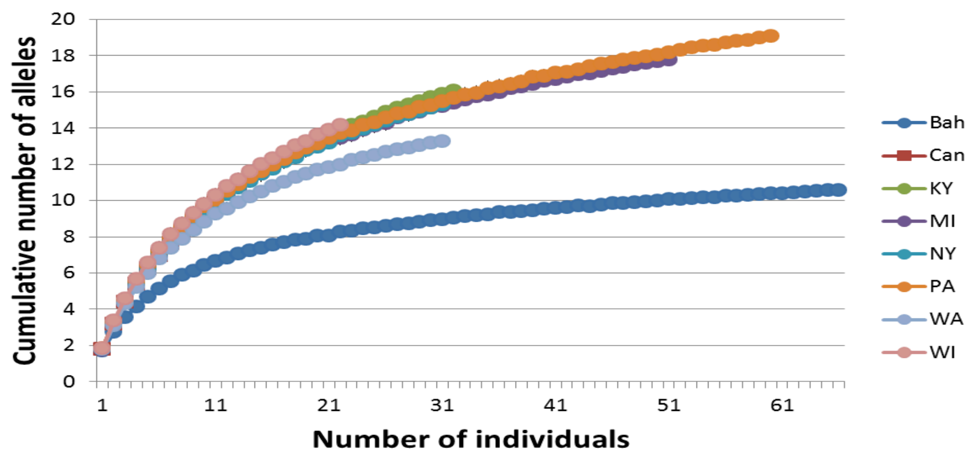

Figure 5: Rarefaction curve for allelic diversity. The Bahamas population (lowest curve) was estimated to have a lower maximum allelic diversity than the continental populations.

2.3.2 Microsatellite structure

The delta-*K* script determined the optimal *K*-value to be 2 (Fig. 6). Cluster assignment at $K = 2$ showed the continental populations were essentially a single population that differentiated strongly from the Bahamas population (Fig. 7). These results were supported by the PCoA results. Although the first three axes explained only 14.09% of the observed variation, the Bahamas genotypes emerged as a largely distinct cluster from the indistinguishable continental populations (Fig. 8).

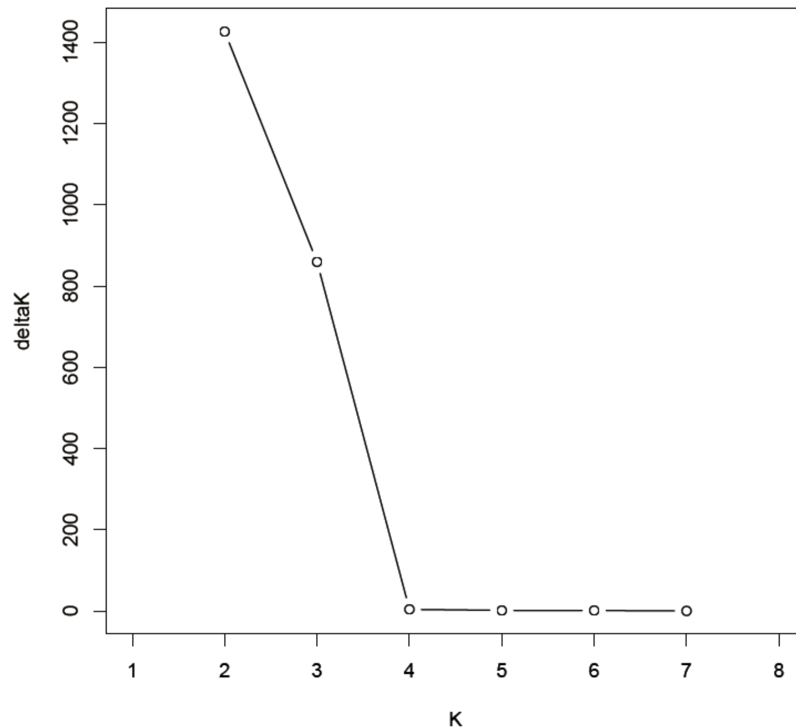

Figure 6: Likelihood plot from delta-*K* script showing that $K = 2$ is the optimum cluster number for this data set.

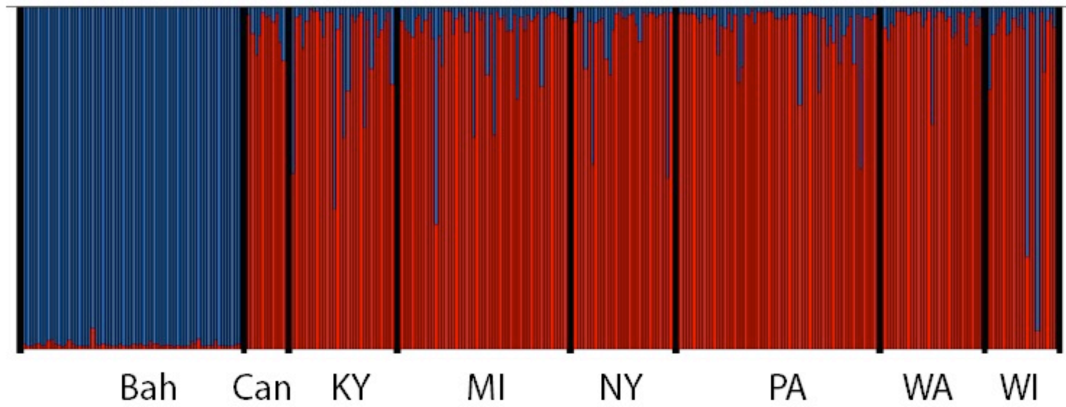

Figure 7: structure barplot showing proportion of the eight populations, showing the presence of population structure as measured from ten microsatellites. Each bar represents an individual. Color represents proportion of membership in either cluster.

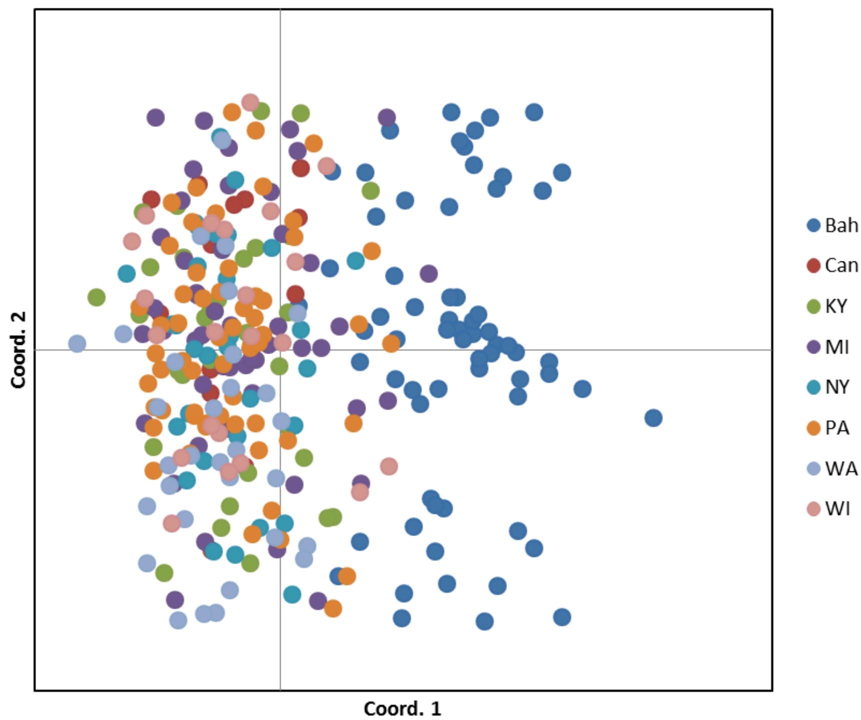

Figure 8: Principal Coordinates Analysis showing distinct Bahamas cluster.

2.3.3 mtDNA structure

Trimmed and concatenated ND2 sequences were 1119 bp long. A total of 18 sites were polymorphic, eleven of which were parsimony-uninformative. Of the seven parsimony-informative sites, three SNPs were fixed differences between Bahamas and continental individuals, equaling a between-group distance of 0.004. This result does not give temporal context to the island-continental divergence, since it could potentially arise from founder effects following the arrival of any individual harboring three singletons. However, it is consistent with the signal from the nuclear DNA that there has been absence of gene flow between the two groups, especially considering the slower evolutionary rate of mtDNA sequences relative to microsatellites (e.g., Brohede *et al.* 2004; Ramaiya *et al.* 2010).

The neighbor-joining tree showed that the 14 Bahamas individuals clustered together to form a polytomy with short branches, indicating few overall mutations across individuals (Fig. 9). Similarly, the 14 individuals across the seven continental populations formed their own polytomy and were not cleanly sorted. Instead, inconsistent sorting of the pairs from Michigan, Washington and New York reflects insufficient population-specific resolution and suggests that continental individuals are considered to be from the same genetic population. Unlike in the Bahamas individuals, however, the continental individuals were related to each other with varying branch

lengths, indicating greater nucleotide diversity within the continental population than within the Bahamas.

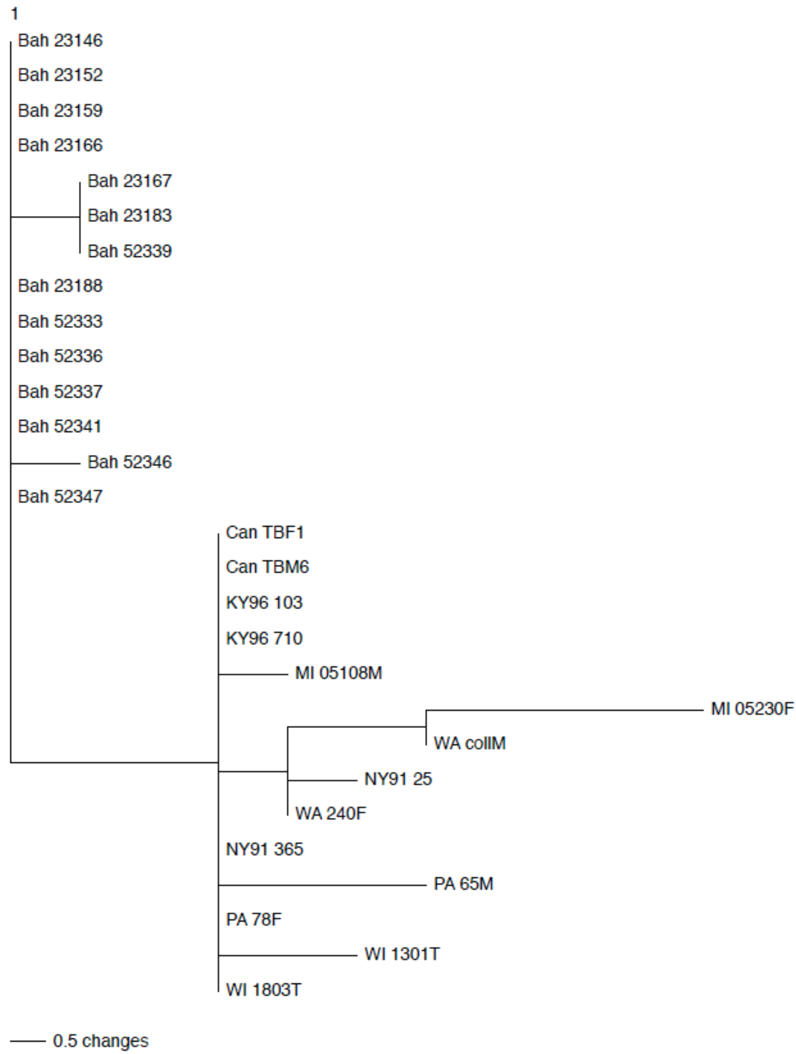

Figure 9: Neighbor-joining tree showing relationships between the eight populations from mtDNA.

2.3.4 Comparison of EPP rate across populations

The proportion of nests containing EPY was not significantly different in the Bahamas (Table 3). However, maximum clutch size in the Bahamas was 3 eggs, while average clutch size in the continental populations was 4 eggs. This difference could affect the measurement of EPP, since the larger the clutch size, the more likely it is for a nest to contain EPY. Thus, as a precaution, subsequent analyses beyond chi-square tests of proportion were conducted on the proportion of EPY and not on the proportion of nests with EPY.

Table 3: Measures of extra-pair paternity in each population. Asterisks indicate that the data were taken from published measurement of EPP.

	Extra-pair young (EPY)	%	Nests with ≥1 EPY	%
Bahamas	16/56	29	10/20	50
Canada	64/243	26	30/78	38
KY*	593/1479	40	295/537	55
MI	32/125	26	20/40	50
NY*	55/232	24	28/68	41
PA	23/87	26	13/27	48
WA*	136/403	34	72/134	54
WI	31/97	32	20/32	62

Continental populations varied significantly in proportion of EPY ($\chi^2 = 46.2$, df = 6, $P < 0.0001$) and in proportion of nests containing at least one EPY ($\chi^2 = 12.6$, df = 6, $P = 0.05$). Adding the Bahamas population did not alter the already significant variation for proportion of EPY ($\chi^2 = 47.3$, df = 7, $P < 0.0001$), but it did change to insignificant the variation in proportion of nests with EPY ($\chi^2 = 12.6$, df = 7, $P = 0.08$). EPP in the Bahamas

population was also not significantly different when all continental populations were pooled and considered as a single population (proportion of EPY: $\chi^2 = 0.74$, df = 1, $P = 0.39$; proportion of nests with EPY: $\chi^2 = 0.001$, df = 1, $P = 0.97$).

For the logistic regression models, Model 1 estimated the intercept μ to be -0.62 ± 0.04, indicating that the overall probability across all populations of being an EPY is $\frac{e^{-0.62}}{1+e^{-0.62}}$. However, the model allowing for population-specific intercepts (Model 2, Table 4) was a significantly better fit than the model considering a single intercept (ANOVA chi-square between Model 1 and Model 2, $\chi^2 = 48.54$, df = 7, $P < 0.0001$). This result identified population as a meaningful variable and was consistent with the chi-square tests (above) showing that EPP rate varied significantly by population.

Table 4: Summary of output for Model 2, the GLM incorporating population as a variable. "Estimate" indicates the maximum likelihood estimate of the intercept for each population. The z-score is the number of standard deviations away from the mean, and the P-value is derived from the z-score.

	Estimate	Std. error	z-score	Pr (>|z|)
$\mu_{Bahamas}$	-0.92	0.30	-3.10	0.0020
μ_{Canada}	-0.11	0.33	-0.34	0.73
μ_{KY}	0.51	0.30	1.71	0.087
μ_{MI}	-0.15	0.36	-0.42	0.68
μ_{NY}	-0.25	0.33	-0.76	0.45
μ_{PA}	-0.11	0.38	-0.28	0.78
μ_{WA}	0.24	0.31	0.77	0.44
μ_{WI}	0.16	0.37	0.44	0.66

Estimated population-specific intercepts were concordant between the Bayesian random-effects model and the fixed-effects model (Model 2) (Fig. 8, first eight panels).

By contrast, the mean continental intercept estimated by the Bayesian random-effects model was different from the mean continental intercept estimated from the fixed-effects model (Model 1 with Bahamas removed) (Fig. 8, final panel). In this case, the random-effects model controlled for heterogeneity in sample size (range = 13 in Canada to 1479 in Kentucky) better than the fixed-effects model, whose mean was skewed to the right given Kentucky's large sample size and high EPP rate (40%). Overall, the different mean intercepts derived from these two models suggest that different inferences are made when considering the continental populations as separate and random but united by a common underlying process, versus considering them as fixed populations with the same mean.

Finally, in the Bayesian random-effects model, the hypothesis that the Bahamas EPP rate was lower than the overall continental rate was tested by calculating the posterior probability that $\mu_{\text{Continental}} > \mu_{\text{Bahamas}}$. The MCMC results gave the value $P = 0.54$, confirming that the Bahamas EPP rate did not differ from the continental EPP rate. This result is visually supported by comparing the similarity between the estimated Bahamas intercept (from either the random- or fixed-effects model) and the estimated continental random-effects mean (Fig. 10).

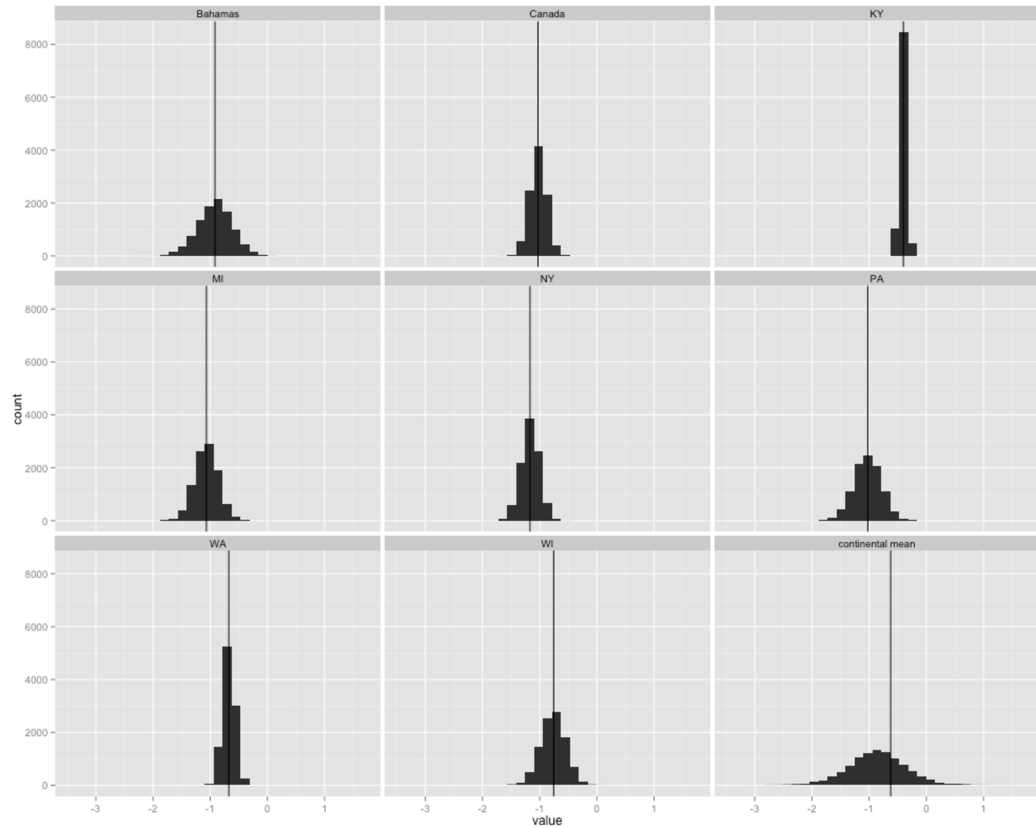

Figure 10: Histograms of posterior samples for the eight population-specific intercepts (first eight panels) and the continental random effect mean (bottom right panel). The population-specific intercept histograms are overlaid with vertical lines indicating the maximum-likelihood estimates for μ from Model 2 (see Table 4), whereas the histogram for the random-effects mean is overlaid with the estimate for the single parameter from Model 1.

2.3.5 Relationship between genetic diversity and EPP rate

A plot of proportion of EPY against genetic diversity confirms that, despite its lower genetic diversity, the Bahamas population has similar proportions of EPY to the continental populations (Fig. 11). A Spearman correlation test showed no relationship between genetic diversity and EPP rate ($S = 64$, $\varrho = 0.24$, $P = 0.58$).

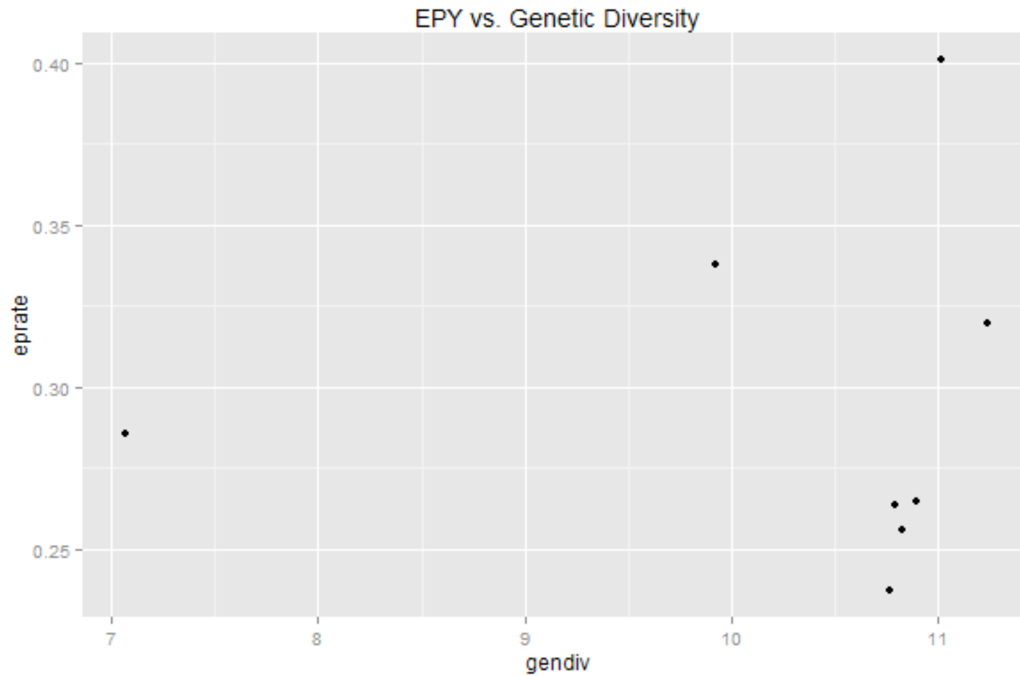

Figure 11: Scatter plot showing relationship between average genetic diversity and proportion of EPY. The Bahamas population (left-most point) has less genetic diversity than the continental populations, yet its EPP rate falls in the middle of the continental populations' EPP rates.

2.4 Discussion

The results presented here show that population genetic diversity does not explain variation in EPP in red-winged blackbirds. Based on the idea that increased genetic diversity could lead to higher payoffs of EPP for females, I predicted that populations from continental North America with high genetic diversity would exhibit higher rates of EPP than a genetically distinct population in the Bahamas with low genetic diversity. Contrary to this prediction, the Bahamas population's EPP rate was not significantly different from those of the continental populations. Furthermore,

genetic diversity did not account for the variation in EPP rate that *was* significant across the continental populations. These results raise questions of why the island and continental populations had similar levels of EPP, and why the continental populations showed significant variation in EPP rates despite being a genetically undifferentiated population.

2.4.1 No difference between EPP in island and continental populations

Findings from the present study contrast those of past studies reporting a positive correlation between genetic diversity and EPP rate in songbird species (Gohli *et al.* 2013; Petrie *et al.* 1998), as well as a study contrasting EPP rates in island and continental populations of one species (Griffith 2000). Several reasons could explain this difference. First, studies included in Griffith's (2000) meta-analysis detecting an "island effect" of EPP included analyses with low numbers of loci, including single-locus minisatellite and microsatellite genotyping, to infer paternity. Because paternity exclusion is sensitive to allele frequencies and diversity (Jamieson & Taylor 1997), the minimum number of loci needed to diagnose paternity for a less diverse population is higher than that needed for a population with high diversity (in the present study, eight vs. four loci, see Methods). Thus, in studies that used a low number of loci to diagnose EPP, EPP could be erroneously reported as lower in island populations, creating a methodological bias that can be resolved by including results from studies using a minimum number of loci to diagnose EPP in islands.

A second possibility is that, despite the lower levels of genetic diversity in the Bahamas population relative to the continental populations, the Bahamas population's genetic diversity may still have been sufficient to justify female participation in EPP. If this interpretation is true, then it raises the question of whether there is an absolute threshold for genetic diversity, above which EPP is adaptive and beneath which EPP's costs outweigh its benefits. This threshold could be similar to that seen in female preference functions, sexual ornaments or regulatory mechanisms (Emlen & Nijhout 2000; Jennions & Petrie 1997; Roff 1996). In this scenario, EPP loses its selective advantage once genetic diversity drops below a certain threshold, leading to an eventual decrease in its occurrence. Both inter- and intraspecific studies could be used to test this idea. Interspecific studies of genetic diversity could identify species with extremely low or high levels of EPP, expanding and updating the data reported in Petrie *et al.* (1998) and Griffith *et al.* (2002). From this spectrum, intraspecific studies of EPP rate comparing populations with different levels of genetic diversity could then test the predictive value of genetic diversity while accounting for within-species variation in EPP. If even the most genetically depauperate species still exhibit robust EPP levels, then genetic diversity could be ruled out as a reliable determinant of EPP.

A third explanation is that any decrease in EPP in the Bahamas, predicted to occur in response to the decreased variance in male genetic profiles (Griffith 2000; Petrie & Lipsitch 1994), was counteracted by an increase in EPP driven by inbreeding

avoidance. This hypothesis is supported by the finding that while the Bahamas population had significantly less observed heterozygosity (H_o) than the continental populations, its inbreeding coefficient (F_{IS}) was just as low as the continental populations, indicating similarly low levels of heterozygote deficit. Because heterozygosity erodes in the face of random mating, such a pattern suggests that the Bahamas population may have engaged in disassortative mating to sustain its standing heterozygosity. Such behavior is consistent with the genetic compatibility hypothesis, since inbreeding itself is an extreme form of genetic incompatibility, and could be especially important in populations of low diversity. In this case, the value of a rare, distinct male would increase as genetic diversity decreases. The advantage of multiple mating would be to locate genetically different males to produce a diverse clutch better able to withstand stochastic events (Reed & Frankham 2003).

This interpretation implies that multiple mating serves different adaptive functions under different contexts. Combined with the threshold hypothesis, it suggests that EPP is sensitive to genetic diversity, but in such a way that there appears not to be a change in the frequency of its occurrence. Instead of continuing to decrease once genetic diversity falls below a certain threshold, EPP levels instead may recover because the benefits shift from indirect genetic benefits to inbreeding avoidance. To test this hypothesis, pairwise comparisons would be required to show that extra-pair mates are nonrandomly chosen and are distinct relative to the population and/or to the female.

Interestingly, reproductive skew caused by disproportionate paternity of rarer genotypes would lead to a negative frequency-dependent dynamic (the "rare male effect," e.g., Kokko *et al.* 2002; Pemberton *et al.* 2003). If inbreeding avoidance drives EPP in a population of low genetic diversity, then the fitness of a male genotype will not be fixed but instead will be determined by its relative frequency to other genotypes in the population.

The primary shortcoming of this hypothesis is that it fails to account for escalating search costs of EPP, which could in fact be higher in populations with lower breeding densities (such as islands) than in high-density populations. Most studies that promote the idea of increased promiscuity for inbreeding avoidance use systems where the costs of multiple mating are relatively low (Madsen *et al.* 1992; Michalczyk *et al.* 2011). Ultimately, these findings will be placed in context when more examples are provided where EPP rates are the same in island and continental populations of the same species or of closely related species (e.g., Conrad *et al.* 2001; García del Rey *et al.* 2012).

Knowing the time since divergence would permit speculation of whether the behavior in the Bahamas population has had the opportunity to evolve independently. The fixed differences in clutch size between the Bahamas and continental population provides some evidence of differentiation. However, it was not possible in the present study to infer the temporal context of evolution. While nuclear and mtDNA of the island

and continental populations both exhibited signals of divergence, the three fixed SNPs in the mtDNA sequences did not provide sufficient molecular resolution to reveal the time of separation. Amplifying a longer region with higher divergence and applying the avian molecular clock (Weir & Schluter 2008; but see Pereira & Baker 2006) could clarify the time since the Bahamas population split from the continent. This result would shed light on whether behavior has been subjected to a separate selective regime. If divergence were recent, then the observed rate of EPP in the Bahamas may simply be a carryover of reproductive behavior in the continental populations, with the potential for future differentiation of selective regimes. By contrast, if divergence occurred much longer ago, then EPP in the Bahamas can be concluded to have remained unchanged from that of continental populations, even after the loss of gene flow.

2.4.2 Continental populations vary significantly in EPP rate

The other unexpected finding of this study is that, while the Bahamas population's EPP rate is not significantly different from that of the continental mean, the continental populations themselves differ in EPP rate. This result implies that local variation in payoffs existed that were not driven by population genetic profiles. Indeed, payoffs associated with extra-pair mating in red-winged blackbirds appear to vary geographically, possibly as a response to local opportunities or benefits of EPP. For example, females in a Washington population derive both genetic and material benefits from extra-pair males, in the form of improved fledging success and increased access to

food and nest defense, respectively (Gray 1997a, b). These direct benefits to females (*sensu* Kirkpatrick & Ryan 1991), as well as the high rates of female solicitation of EPC observed in the field (Gray 1996), have not been observed elsewhere, either because this payoff is truly isolated or because no subsequent research on this hypothesis has been conducted in other blackbird populations. In a New York population, fledging success also increases slightly with the number of sires in a brood, although EPC is resisted or at least never initiated by females (Westneat 1992). By contrast, increased EPP decreases fledging success in an Canada population, possibly through reduced nest defense by territorial males (Weatherhead *et al.* 1994). Although behavioral variables were not standardized across studies, the diversity in the reported behavior indicates dynamic cost-benefit calculations whose interactions cannot easily be conveyed by linear variables such as genetic diversity.

In fact, if differences in local landscape are sufficient, then females in separate populations could pursue EPP for different reasons (e.g., good genes or genetic compatibility). A feasible way to test this idea would be to survey the distribution of EPP across nests in target populations (Griffith *et al.* 2002). Where the good genes hypothesis prevails, EPP distribution should tend to be bimodal (females mated to social males of poor genetic quality would engage in EPP, while females mated to social males of high genetic quality would not). By contrast, if the genetic compatibility hypothesis drives female interactions, then EPP should be more equally distributed across clutches,

because it will pay all females to increase the diversity of their clutch. While populations with different mechanisms would appear to exhibit similar rates of EPP rates, variation in EPP would persist at the level of the nest.

Aside from genetic factors, variation in immediate payoffs could still be shaped by ecological factors such as breeding density, breeding synchrony and latitude. However, in red-winged blackbirds, density and synchrony appear not to influence EPP, and females additionally vary in choosing neighboring or distant males as extra-pair mates (Ball *et al.* 1988; Weatherhead 1997; Westneat & Mays 2005). As for latitude, while many studies have implicated latitude or migration as a predictor of EPP (Bonier *et al.* 2014; Douglas *et al.* 2012; Spottiswoode & Møller 2004; Stutchbury & Morton 2001), others have found that latitude does not influence the relative strength of sperm competition (Albrecht *et al.* 2013; Eikenaar *et al.* 2013), especially when accounting for differences in life-history traits like clutch size. One meta-analysis found that tropical and temperate species do not necessarily have as marked differences as have been reported in individual studies (Macedo *et al.* 2008), with tropical species having a much greater range of EPP rates that encompasses the range of EPP observed across most temperate species. Considering the results of the present study, in which a non-migratory, low-latitude population showed no difference in EPP rate with migratory temperate populations, latitude by itself does not seem to predict EPP rate in island vs. continental red-winged blackbirds.

More likely, interactions between ecological and genetic factors defining each population collectively drive variation in EPP rate (Arct *et al.* 2013; Arnold & Owens 2002). Future studies could evaluate models incorporating the multiple ecological and genetic variables that have been proposed to influence EPP. An integrative model would determine whether the effects of genetic diversity are enhanced or diminished by variables (such as breeding density and synchrony) influencing resource distribution and mate availability (Emlen & Oring 1977) to shape the relative strengths of post-copulatory sexual selection and thus the payoffs of EPP. In addition, an empirical comparison between the Bahamas and a Florida population (Prather & Cruz 2006) could potentially decouple the ecological and genetic factors experienced by lower-latitude populations of red-winged blackbirds. Although the Bahamas and continental populations have already been shown to have similar EPP rates, the inclusion of a continental population with predicted high genetic diversity, but similar latitude and ecology to the island population, could test whether EPP is impervious to genetic, ecological or both types of factors.

2.4.3 Adaptive explanations for EPP revisited

Results from the present study did not support the idea that females in populations with higher levels of genetic diversity engage in more frequent EPP for the increased probability of receiving indirect genetic benefits. Instead, the study joins a growing number where no clear results have emerged from considering EPP as a

female-driven behavior. An alternative framework with which to explore mating system variation is by considering it as the result of sexual cooperation and conflict between three parties: the female, the within-pair male, and the extra-pair male (Arnqvist & Kirkpatrick 2005; Westneat & Stewart 2003), with different "winners" and "losers" from each encounter. The proposed benefits conferred to extra-pair young would thus not be a reliable outcome of EPP. This scenario has been demonstrated in recent findings from a longitudinal study of song sparrows (*Melospiza melodia*) showing that, for at least one species, extra-pair young are less fit than within-pair young in term of their lifetime reproductive success (Sardell *et al.* 2012). In addition, non-genetic maternal effects in many taxa challenge the notion that variation in offspring fitness can be attributed exclusively to male-derived benefits.

In light of evidence that direct costs of multiple mating may outweigh the benefits of EPP (Arnqvist & Kirkpatrick 2005; but see Griffith 2007), alternative explanations integrating the indirect benefits of EPP and the costs of sexual conflict have been advanced to account for this variation. Fitness advantages and costs associated with EPP, measured with respect to all "players" involved, determine the net benefits of EPP and thereby the direction and strength of selection on extra-pair mating in a population. If this scenario is true, then the reasoning used in the present study (and others attributing the benefits of EPP largely to females) requires refinement beyond linear consequences for the female alone.

Finally, a non-adaptive explanation for the observed similarity in EPP rates between island and continental populations is that EPP is under phylogenetic control for all red-winged blackbirds. Although this explanation does not account for the significant variation observed across the continental populations, it refers to the argument that EPP is a fundamental reproductive strategy expressed broadly across passerines, with different evolutionary explanations valid at different scales of comparison. Variation in EPP at the level of species or higher has been shaped by differences in life history evolving over millions of years, while variation within species has been likely determined by differences in the local ecological and genetic factors discussed above (Arnold & Owens 2002). Further studies of *Agelaius* congeners could clarify the taxonomic breadth of EPP in this genus and highlight whether species are united or distinguished by their mating systems.

3. Conservation genetics and mating system of the endangered yellow-shouldered blackbird (*Agelaius xanthomus*)

3.1 Introduction

The yellow-shouldered blackbird (*Agelaius xanthomus*), endemic to Puerto Rico and once widespread across the island, has been listed as endangered since 1976 (Post 1981; USFWS 1976). Reasons for the species' decline include loss of feeding and nesting habitat, introduced predators, and disease (Post & Wiley 1976). By far the most serious threat to the blackbird, however, has been brood parasitism by the introduced shiny cowbird (*Molothrus bonariensis*), which reached Puerto Rico from South America in the 1940s (Post & Wiley 1976, 1977; Wiley *et al.* 1991). As the shiny cowbird's primary host in Puerto Rico, the yellow-shouldered blackbird has suffered severe decreases in reproductive success. In southwestern Puerto Rico, the largest known population shrank by 85% from 1,663 in 1975 to 226 in 1982 (Cruz *et al.* 2005).

To combat this decline, in 1980 the U.S. Fish and Wildlife Service (USFWS) and the Puerto Rico Department of Natural and Environmental Resources (PRDNER) implemented a management program consisting of artificial nest structures and active cowbird removal (López-Ortiz *et al.* 2002; USFWS 1996; Wiley *et al.* 1991). As a result of the intensive monitoring, the southwestern population recovered to 994 birds by 2007 (USFWS 2011). However, financial constraints have recently limited the scope of the cowbird control program, triggering a renewed increase in parasitism. This threat,

combined with ongoing habitat loss, has led to a continued decline, with 2012 PRDNER censuses counting fewer than 400 birds prior to breeding, and 650 birds after breeding, in their native mangrove habitat (Medina-Miranda *et al.* 2013).

In response to these combined threats, the U.S. Department of Agriculture (USDA), PRDNER, USFWS and the Conservation Breeding Specialist Group (Apple Valley, MN) conducted a population and habitat viability analysis (PHVA) in 2012 modeling the outcomes of different management strategies that prioritized different elements of the yellow-shouldered blackbird's survival. The PHVA used a novel metamodeling approach allowing for separate components of the yellow-shouldered blackbird and shiny cowbird models to interact, thus predicting the impact of different management plans on both species and the species' impacts on each other. A key finding was that, even with incumbent management practices, the yellow-shouldered blackbird is currently experiencing a negative growth rate with a 31% chance of eventual extinction (Medina-Miranda *et al.* 2013). Best-case scenario models predicting the lowest probability of extinction required both high rates of survival in adults (through increased protection of habitat and decreased anthropogenic disturbance) as well as recruitment in juveniles (through increased control of shiny cowbirds and predators), suggesting the long-term sustainability of the blackbird cannot be feasible without active management across all life stages, combined with aggressive control of the shiny cowbird.

Despite the extensive knowledge of the yellow-shouldered blackbird's natural history, the PHVA was conducted with limited knowledge of its genetic profile and its genetic mating system (USFWS 2011). (Genetic mating systems describing copulatory relationships between the sexes are distinct from social mating systems describing their behavioral associations (Griffith *et al.* 2002).) Understanding these traits is critical to improving the accuracy of both short- and long-term survival predictions and to informing future recovery plans (Elphick *et al.* 2007). Population genetic measures such as the degree of inbreeding reflect the immediate fitness consequences of mating, while allelic diversity, heterozygosity and effective population size indicate the long-term ability of the population to respond to environmental change (Frankham *et al.* 2010; Reed & Frankham 2003). Indeed, one of the stated goals from the PHVA was to maintain at least 95% of current levels of genetic diversity, but without a baseline measurement, assessing progress toward this goal is difficult.

As an island endemic, the yellow-shouldered blackbird is particularly vulnerable to the effects of decreased genetic diversity. Its limited population size accelerates the onset of inbreeding depression, and its isolation means that genetic diversity cannot be restored through gene flow from incoming migrants. Small populations additionally tend to exhibit exaggerated effects of drift and a decreased response to selection (Frankham 1997, 1998; Lande 1988). In extremely small populations with mating systems that require a threshold number of individuals, Allee and other social effects from

insufficient population density may lead to collapse (Stephens & Sutherland 2000). Together, these factors expose the blackbird to a greater risk of extinction than closely related species with larger sizes and multiple sources of genetic variation. Consequently, characterizing the population genetic profile of the yellow-shouldered blackbird is critical to designing plans that maximize its likelihood of survival. These measurements may be particularly important when assessing the necessity and benefits of more intensive management decisions such as outbreeding or translocation.

In addition to these population genetic traits, the genetic mating system of the yellow-shouldered blackbird is also of importance to predicting the long-term trajectory of the species. In birds, the discrepancy between genetic and social mating systems is well established, with almost 90% of socially monogamous species exhibiting genetic evidence of extra-pair paternity (EPP) (Griffith *et al.* 2002). EPP is an established occurrence in birds, but certain trends remain poorly understood. First, there is mixed evidence over whether island species exhibit significantly different levels of EPP compared with closely related species or with mainland populations of the same species. Initially, island species were thought to display consistently lower levels of EPP, possibly due to decreased sexual selection resulting from decreased genetic diversity (Griffith 2000). However, more recent studies have disputed this view, showing that certain elements of sperm competition are comparable across island and continental species (Albrecht *et al.* 2013). In general, EPP cannot be assumed to be especially high or

low on islands, and current work recommends that EPP for each species be measured empirically (Conrad *et al.* 2001; García del Rey *et al.* 2012). Verification of the genetic mating system in the socially monogamous yellow-shouldered blackbird is especially important, given the evolutionary history of icterids (Chapter 1).

Because it controls the intensity of sexual selection, genetic mating system is a life history trait with conservation implications. Genetic diversity, the degree of inbreeding, and effective population size are all shaped by the mating relationships between males and females. For instance, the amount of genetic diversity maintained across generations depends on the number of breeding adults, which can be skewed in mating systems such as harem and dominance polygyny where a few individuals of one sex (typically male) mate disproportionately with the majority of the other sex (Nunney 1993). Likewise, the amount of mating that occurs between close relatives directly determines the severity of inbreeding. Finally, the effective population size (N_e) is sensitive to unequal sex-ratios and variation in family size (Falconer & Mackay 1996; Frankham 1995; Wright 1938), both of which are dictated by the number of breeding adults and the identity of individual(s) with which they are mating. Overall, the genetic mating system of a species is an intrinsic life-history trait that plays a key role in determining the proportion of standing genetic variation transmitted to subsequent generations.

Genetic mating system can increase or decrease extinction risk depending on the exact mating relationships between males and females. Depending on whether multiple mating promotes or reduces skew, different predictions can be made about how a mating system alters population genetic diversity and effective population size across generations (Elphick *et al.* 2007; Frankham *et al.* 2010; Nunney 1993). For example, if a species exhibits harem polygyny in a lek system, then the reproductive skew is expected to decrease genetic diversity because certain individuals gain the majority of reproductive success (Kirkpatrick & Ryan 1991). Harem polygyny also has a much larger impact on effective population size than monogamy (Evans & Charlesworth 2013). By contrast, if multiple mating weakens reproductive skew by enabling *more* males to gain reproductive success, or if it is used as a strategy employed by females to reduce inbreeding, then polygamy may in fact sustain genetic diversity, potentially leading to more resilient populations.

The central question is whether the genetic mating system of the yellow-shouldered blackbird increases or decreases variation in male reproductive success. This trend can be determined by testing whether EPP is present in the species, and if so, by identifying the males that are chosen by females as extra-pair mates. Depending on the results, a few scenarios can be envisioned. First, if EPP is prevalent, then two outcomes are possible depending on the identities of the genetic fathers. (a) If genetic analyses reveal that only a few males claim the majority of extra-pair mating, then this skew will

decrease allelic diversity in subsequent generations, increasing the risk for genetic depletion. (b) If EPP is distributed evenly among territorial males, or if EPP gives floaters the opportunity to mate with females, then this behavior may be equalizing and have a neutral or positive effect on effective population size. Second, if EPP is rare or absent (and if most territory-holding males can be assumed to secure mates), then the effective population size is likely to be high. These latter two outcomes would indicate that mating system does not have a substantial effect on population viability, whereas the first scenario would be an issue of concern.

Here I characterize the genetic profile and mating system of a breeding population of yellow-shouldered blackbirds. This population is a subset of the southwestern population under long-term management. Using genotypes from nine microsatellite loci, I characterize the YSBL population profile and describe the species' genetic mating system. I compare findings with an island population of the red-shouldered blackbird (*Agelaius phoeniceus*), the YSBL's well-studied congener. I integrate my findings with the PHVA and discuss management implications in light of these results.

3.2 Methods

3.2.1. Sample collection

I sampled the complete families (female, social male and chicks) of 29 nests in a free-living population of yellow-shouldered blackbirds in the Pitahaya region of

Boquerón Commonwealth Forest in Lajas, Puerto Rico (Fig. 12, adapted from López-Ortiz *et al.* (2002)). Sampling occurred from 8 May to 14 June 2012. This population has been managed since 1984 and nests in artificial nest structures made from PVC pipes that allow for monitoring of shiny cowbird parasitism (López-Ortiz *et al.* 2002). The southwestern population is the major stronghold of the species; a smaller population in southern Puerto Rico contained 82 individuals in 2012, and an eastern population is likely extirpated. Additionally, a morphological subspecies, *A. xanthomus monensis* (Barnés 1945), inhabits Mona and Monito Islands 138 km to the west of the main island (n = 372 in 2010, Medina-Miranda *et al.* 2013).

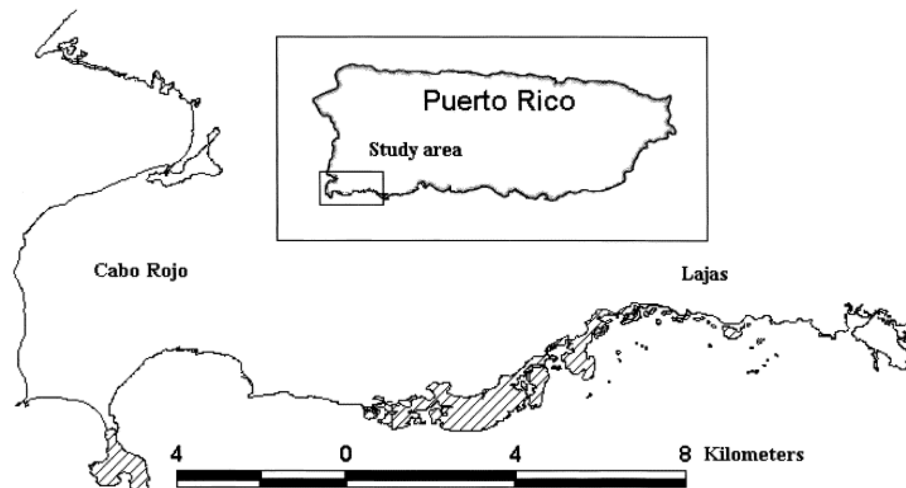

Figure 12: Map of field site. Striped area refers to Boquerón State Forest. The circled region marks the area (Pitahaya) where sampling took place. From López-Ortiz *et al.* (2002).

In one nest, I sampled the chicks and social male but not the female. I sampled an additional 19 chicks that were on the verge of fledging, as well as four randomly caught females. In total, I collected samples from 169 individuals (63 adults and 106 chicks).

However, I discarded the 19 near-fledglings from my analysis to avoid problems with non-independence due to their unknown maternity and paternity. Only the 87 chicks in 30 sufficiently sampled nests (including the nest with only the social male sampled) were used for subsequent analyses.

Both parents visited the nest to feed chicks. I captured adults using mist nets and walk-in nest-box traps. Traps consisted of a PVC pipe that fit into the mouth of the artificial nest structures, with a trapdoor constructed from cardboard and a drinking straw. I bled adults from the brachial vein using sterile 26G × ½ in. BD PrecisionGlide needles and collected five drops of blood (100 μl) onto Whatman FTA bloodstain cards treated with 0.5 mol EDTA. I then banded adults with USFWS numbered metal bands and three plastic colored leg bands. Chicks were bled from the tarsal vein if they were sampled from hatch to day 7 and from the brachial vein if they were sampled from day 7 to fledging at day 13 (Díaz-Rodríguez & Lewis 2006). Five drops of blood (25 μl) were collected from each chick.

3.2.2. DNA extraction and amplification

I extracted DNA from dried blood with a Qiagen DNeasy Blood and Tissue Kit and evaluated DNA concentration and purity with a Nanodrop spectrophotometer. Samples with poor concentrations (< 4.0 ng/μl) were re-extracted.

To genotype individuals, I amplified nine microsatellite loci previously shown to be polymorphic with red-winged blackbird DNA: Aph54 (Westneat & Mays 2005),

LTMR6 (McDonald & Potts 1994), Qm10 (Hughes *et al.* 1998), Pca3 (Dawson *et al.* 2000), Dpμ16 (Dawson *et al.* 1997), Ap38, Ap79, Ap107, and Ap144 (Barker *et al.* 2011). For each individual, I ran three multiplex PCR reactions. Each reaction combined three primer pairs, with the forward primer in each pair fluorescently labeled with FAM, HEX (Sigma-Aldrich) or NED (Applied Biosystems). Reactions consisted of 2.2 μl of DNA, 3.0 μl of Qiagen Type-It Multiplex PCR Master Mix, 1.6 μl RNAse-free water, and 1.0 μl of 100 μM primer mix (consisting of the three primer pairs). PCR cycles were initiated at 95°C for 5 minutes to activate the HotStarTaq Plus DNA polymerase, followed by ten touchdown cycles from 60°C to 50°C and 23 additional cycles at 50°C. Each cycle consisted of denaturation at 95°C for 0:30, annealing for 1:30, and extension at 72°C for 0:30, with one final extension at 68°C for 10 minutes.

Plates were processed using Duke Sequencing Facility's Applied Biosystem 3730xl DNA Analyzers, and genotypes were scored with GeneMarker v.1.8 (SoftGenetics, State College, PA) using size standard GS-500 to determine allele sizes. Homozygous loci were genotyped at least twice to account for the possibility of allelic dropout. Null alleles and allelic dropout were checked using Micro-Checker v.2.0 (Van Oosterhout *et al.* 2004). I tested all loci for Hardy-Weinberg equilibrium using GenAlEx v.6.5 (Peakall & Smouse 2006) and pairwise linkage disequilibrium using Genepop (Raymond & Rousset 1995). Dememorization was set to 10,000, and 100 batches were run, with 10,000 iterations per batch.

3.2.3. Effective population size

Yellow-shouldered blackbirds have discrete but overlapping generations, which violates an assumption of many equations that calculate N_e. Therefore, certain methods may instead estimate the effective number of breeding individuals (N_b) in a single reproductive season. For certain life histories (e.g., salmon), N_e is roughly a function of $N_b \times g$, where g = generation time (Waples 2010). However, in other systems, the relationship between these two values is more complex and depends on spatio-temporal demographic factors, such as age structure and migration, that require a long time frame to assess (Jorde & Ryman 1995; Waples 2010). Where generation times are overlapping but short (as in the yellow-shouldered blackbird), values of N_b can be expected to correspond with values of N_e. In particular, low values of N_b should represent low values of N_e (Phillipsen *et al.* 2010).

I estimated N_e and N_b using three programs that allowed analysis of single-sampled populations (versus samples over time). First, I used NeEstimator v.2 (Do *et al.* 2014) to calculate N_e using the rate of decay in linkage disequilibrium (Waples & Do 2008). NeEstimator also calculates N_b using heterozygosity excess (Pudovkin *et al.* 1996; Zhdanova & Pudovkin 2008) and molecular coancestry (Nomura 2008). To compare results with another population of island-dwelling blackbirds, I ran the program again to estimate N_e and N_b in the Bahamas red-winged blackbird population. Second, I used the online program ONeSAMP (Tallmon *et al.* 2008), which uses approximate Bayesian

computation to calculate N_e by measuring change in allele frequency due to drift. Finally, I used the feature in COLONY v.2.0.3.1 (Jones & Wang 2010) to calculate N_e from sibship assignments (Wang 2009). The red-winged blackbird data were not run with these latter two programs.

In NeEstimator, I used two separate critical values to define the minimum allele frequency allowed in analysis. As rare alleles can bias the estimate of N_e (Do *et al.* 2014), NeEstimator gives the option to set custom values. In the method using LD, alleles should not be used that are only featured once (i.e., in a single copy in a heterozygote). Therefore, the critical value needs to be greater than $1/(2S)$, where S is the number of individuals with data reported at both haplotypes. For both the red-winged and yellow-shouldered blackbird, this value was $1/(2 \times 66)$ and $1/(2 \times 63)$, respectively, corresponding to 0.008. Thus, I set one critical value as 0.01 and another, more stringent one at 0.05. The program also calculated 95% confidence intervals from both parametric and non-parametric (jackknife) methods.

3.2.4. Genetic diversity

I used GenAlex to test for Hardy-Weinberg equilibrium and calculate allelic diversity, observed and expected heterozygosity, Shannon diversity index, and the inbreeding coefficient (the fixation index F_{IS}) across all loci in the yellow-shouldered blackbird. I then compared the genetic diversity of the yellow-shouldered blackbird with all eight red-winged blackbird populations sampled in Chapter 2. However, this

comparison required me to use a subset of the genotyping data, because only eight of the nine loci used for genotyping (Aph54, LTMR6, Qm10, Dpμ16, Pca3, Ap107, Ap144, and Ap79) were shared across both species. I re-ran GenAlEx on the yellow-shouldered and red-winged blackbirds with these eight loci to compare population genetic measures. To standardize allelic diversity across populations with different sample sizes (see Table 7), I separately calculated sample-size-adjusted number of alleles (N_s) using the jackmsatpop function of the R package PopGenKit v.1.0 (Paquette 2012). This function uses Genepop input files to determine allelic diversity for a given sample size. The program sampled 13 individuals (corresponding to the lowest *n*, from Canada) per iteration for 100 iterations.

I first compared average allelic diversity in the Bahamas red-winged blackbird versus the yellow-shouldered blackbird simply by dividing the number of total alleles by the number of loci (n = 8). Sample size corrections were not required for this measure because of similar sample sizes in the two populations (n = 66 in the red-winged and 63 in the yellow-shouldered blackbird). I next ran a t-test comparing the genetic diversity in the yellow-shouldered blackbird with the diversity of the Bahamas red-winged blackbird population. Then, I ran a t-test comparing the genetic diversity of the yellow-shouldered blackbird to the seven continental red-winged blackbird populations. The seven populations were pooled in advance to give average levels of allelic diversity,

Shannon diversity index, observed and expected heterozygosity, and inbreeding coefficient.

To visualize the differences in genetic diversity across all three groups (continental red-winged blackbird, Bahamas red-winged blackbird, and yellow-shouldered blackbird), I used the jackmsatpop command in PopGenKit to generate rarefaction curves estimating cumulative allelic diversity. The curve was built by sampling the number of alleles recorded at stepwise increases (interval = 1 individual) until the sample size of each population was reached. At each interval, 100 jackknife replicates were performed.

3.2.5 Evidence of bottleneck

To look for evidence of a recent bottleneck in the yellow-shouldered blackbird, I used BOTTLENECK v.1.2.02 (Cornuet & Luikart 1996) to detect heterozygosity excess and deviations from expected allele frequency distribution. Populations expanding after a bottleneck are thought to show a temporary excess in heterozygosity because of deviation from mutation-drift equilibrium. Allelic diversity is reduced faster than heterozygosity, leading to higher observed levels of heterozygosity than expected under the observed allele frequencies. I used all three mutation models offered in BOTTLENECK to calculate expected heterozgosity: the infinite allele mutation (I.A.M), the two-phase (T.P.M.), and the stepwise mutation (S.M.M.) models. I tested for deviations between observed and expected heterozygosity with Wilcoxon, standardized

difference (H_{obs} – H_{exp} / SD) and sign-rank tests. I used the default settings for calculations: Variance for T.P.M. was set at 30%, the proportion of S.M.M. in T.P.M. was set at 70%, and the program was run for 1,000 iterations.

3.2.6. Parentage analysis

To determine extra-pair paternity across the 30 nests, I compared genotypes of the social father and offspring to identify all allelic incongruities. All inconsistencies involved at least two of the loci, minimizing the possibility of mistaking occasional single-locus mutations for genetic mismatches (Westneat & Mays 2005). An additional calculation on GenAlEx confirmed that, with both parents' genotypes available, the probability of paternity exclusion reached 95% with four of the loci used, thus increasing the confidence of exclusion when considering all nine loci. For the one nest where the female's genotype was unknown, I ruled out extra-pair paternity if one of the social male's alleles was present in each chick and if there were no more than four unique alleles per locus across all chicks. Extra-pair paternity was measured with two proportions: the number of extra-pair young (EPY) out of the total number of chicks, and the number of nests containing least one EPY out of the total number of nests.

I used COLONY to determine whether any of the 30 sampled males could be identified as extra-pair sires. Male and female mating systems were specified as polygamous. I ran two iterations with a 50% probability that the father was included

among the male candidates and a 95% probability that the mother was included among the female candidates.

I compared EPP rates against two different reference groups using R v.3.0.2 (R Development Core Team, 2013). First, defining EPP rate as the number of EPY, I compared the EPP rate of the yellow-shouldered blackbird with EPP rates in an island (Bahamas) population of red-winged blackbird. I used a chi-square test to compare differences in frequencies. Second, I compared EPP rate of the yellow-shouldered blackbird with EPP rates of the Bahamas plus seven continental red-winged blackbird populations (Chapter 2). I used general linear models to analyze the effect of population on variation in EPP rate (see equations, Chapter 2). I then used the R package multcomp (Hothorn *et al.* 2014) to perform multiple comparisons examining the relative placement of the yellow-shouldered blackbird's EPP rate under the distribution of red-winged blackbird EPP rates. Finally, I repeated the analyses, this time defining EPP rate as the number of nests containing at least one EPY. Comparisons were possible because maximum clutch size in the yellow-shouldered blackbird was four eggs, the same as in the continental populations but higher than the Bahamas' three eggs (Chapter 2).

To test where the yellow-shouldered blackbird's single measure of EPP fell relative to the red-winged blackbird distribution, I ran a generalized linear hypothesis test in multcomp. This test is a multiple comparison of means that measures the

contrasts of individual populations relative to a grand mean. I used a Tukey post-hoc test to examine pairwise differences between each population.

Finally, I calculated average individual reproductive output of males and females. For females, I simply calculated the average clutch size across all 30 nests, as all females were genetic mothers of their offspring. For males, I subtracted any paternity lost at the nest and added EPY sired, then divided by 38 (30 sampled males + 8 inferred extra-pair paternal genotypes from COLONY, see Results). However, this is an incomplete estimate of male reproductive success, because the extra-pair success of most sampled males, as well as the within-pair success of the inferred males, was unknown.

3.3 Results

3.3.1. Microsatellite quality

Micro-Checker confirmed there were no null alleles or allelic dropouts. Genepop found one locus pair, Qm10 and Ap107, with a pairwise LD probability that remained significant after Bonferroni correction. Since both loci were statistically independent with all other loci, they were kept in the analysis (Selkoe & Toonen 2006). One locus, Dpμ16, was not in Hardy-Weinberg equilibrium ($P < 0.05$).

3.3.2. Effective population size

Estimates for the three methods used in NeEstimator are shown in Table 5. In addition to those estimates, ONeSAMP calculated a mean N_e of 72.0 (95% CI = 64.6 and 82.8), while COLONY calculated an N_e of 92 (95% CI = 67 and 127). Collectively, the

average estimate of N_e is 62.6 ±10.6 SE. This quantity is 15% of the 2012 census size of 400 birds in southwest Puerto Rico.

With all methods except for heterozygosity excess, the estimated N_e for the Bahama red-winged blackbirds was several times larger than the estimated N_e for the yellow-shouldered blackbird. In the calculation with heterozygosity excess, the estimate for both species at both critical values (0.05 and 0.01) was infinity. The imprecision of this figure may have been due to the conditions assumed by the model using heterozygosity excess, which is best supported when the number of effective breeders is very small, the number of sampled progeny exceeds 200, and the cumulative number of independent alleles exceeds 80 (Zhdanova & Pudovkin 2008). The possibility of a larger N_b in the yellow-shouldered blackbird, combined with smaller than ideal sample sizes of progeny and alleles, may explain why the estimate could not be resolved.

Table 5: Estimates of Ne in the yellow-shouldered blackbird vs. red-winged blackbird from NeEstimator. Estimates from the other two programs (ONeSAMP and COLONY) are reported in the text.

Lowest allele frequency used	0.05		0.01	
	YSBL	**Bah RWBL**	**YSBL**	**Bah RWBL**
(1) LD				
Harmonic Mean Sample Size	63	66	63	66
Independent Comparisons	583	1149	879	2833
Overall r^2	0.022	0.017	0.022	0.017
Expected r^2 Sample	0.017	0.016	0.017	0.016
Estimated N_e	56.5	258.5	65	298.6
95% CIs for N_e:				
Parametric	37.1	118.1	44.5	162.2
	98.3	Infinite	106.2	1294.4
Jackknife on Loci	35.4	112.3	40.6	156.4
	106.6	Infinite	125.6	1709.9
(2) Heterozygosity excess				
Harmonic Mean Sample Size	63	66	63	66
Independent Alleles	37	51	46	82
Weighted Mean D	-0.012	-0.028	-0.0043	-0.025
Estimated N_b	Infinite	Infinite	Infinite	Infinite
95% CIs for N_b:				
Parametric	18.9	198.7	19.1	62.1
	Infinite	Infinite	Infinite	Infinite
(3) Molecular Coancestry				
Harmonic Mean Sample Size	63	66		
Overall f1^	0.018	-0.00044		
Estimated N_b	27.5	Infinite		
95% CIs for N_e:				
Jackknife on Loci	2	Infinite		
	85.7	Infinite		

3.3.3. Genetic diversity and heterozygosity

The Bahamas red-winged blackbird population had higher average allelic diversity across loci (95 alleles / 8 loci = 11.875 alleles/locus) than the yellow-shouldered

blackbirds (51 alleles / 8 loci = 6.375 alleles/locus, Table 6). Surprisingly, pairwise comparisons between the Bahamas red-winged blackbird and the yellow-shouldered blackbirds showed that the yellow-shouldered blackbird did not have significantly lower allelic diversity overall, whether measured by raw or sample-size-adjusted number of alleles (N_r: $t = 1.88$, df = 7, $P = 0.10$; N_s: $t = 1.73$, df = 7, $P = 0.13$; I: $t = 1.56$, df = 7, $P = 0.16$). This result is likely due to the fact that the yellow-shouldered blackbird actually had greater allelic diversity at two loci (Qm10 and Pca3, Table 6). When considering the cumulative diversity of the two species, however, the yellow-shouldered blackbird's population-wide diversity was lower than that of the Bahamas red-winged blackbird, which itself has significantly lower diversity than that of the seven continental populations (Fig. 13). Cumulative diversity across all eight loci plateaued at the average allelic diversity values calculated above: 6.375 alleles in the yellow-shouldered blackbird and at 11.875 alleles in the Bahamas red-winged blackbird.

Between the Bahamas red-winged blackbirds and yellow-shouldered blackbirds, expected and observed heterozygosity were not significantly different (H_o: $t = 1.03$, df = 7, $P = 0.34$; H_e: $t = 1.10$, df = 7, $P = 0.31$), nor was the degree of inbreeding (F_{IS}: $t = 0.30$, df = 7, $P = 0.77$). In fact, more loci showed signs of outbreeding (negative F_{IS} values) in the yellow-shouldered blackbird than the red-winged blackbird (five loci vs. three), though without further sampling it is difficult to infer whether this result is due to factors other than chance.

Table 6: Genetic diversity across nine loci for 63 yellow-shouldered blackbirds and 66 Bahamas red-winged blackbirds. N_r = number of raw alleles from genotyping, N_s = sample-size-adjusted allelic diversity, I = Shannon diversity index, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient.

Locus	N_r		N_s		I	
	YSBL	RWBL	YSBL	RWBL	YSBL	RWBL
Aph54	8	28	6.84	14.29	1.73	2.93
LTMR6	3	6	3.00	4.53	1.08	1.16
Qm10	13	6	9.88	5.30	2.30	1.37
Dpμ16	3	5	2.89	4.50	0.90	1.29
Ap38	5	-	4.70	-	1.35	-
Pca3	3	2	2.97	2.00	0.82	0.69
Ap107	12	22	8.73	13.07	2.18	2.70
Ap144	7	18	5.86	11.40	1.56	2.58
Ap79	2	8	2.00	5.85	0.69	1.70
Mean	**6.22**	**11.88**	**5.21**	**7.62**	**1.40**	**1.80**
SE	**1.36**	**3.35**	**0.94**	**1.62**	**0.20**	**0.29**
Locus	**H_o**		**H_e**		**F_{IS}**	
	YSBL	**RWBL**	**YSBL**	**YSBL**	**YSBL**	**RWBL**
Aph54	0.82	0.92	0.78	0.93	-0.055	0.006
LTMR6	0.70	0.54	0.66	0.58	-0.063	0.062
Qm10	0.84	0.68	0.88	0.65	0.046	-0.044
Dpμ16	0.51	0.65	0.55	0.66	0.069	0.010
Ap38	0.70	-	0.69	-	-0.016	-
Pca3	0.46	0.52	0.47	0.50	0.026	-0.031
Ap107	0.90	0.94	0.87	0.91	-0.036	-0.030
Ap144	0.68	0.82	0.72	0.91	0.050	0.102
Ap79	0.51	0.79	0.50	0.80	-0.016	0.008
Mean	**0.68**	**0.73**	**0.68**	**0.74**	**0.001**	**0.01**
SE	**0.05**	**0.06**	**0.05**	**0.06**	**0.016**	**0.02**

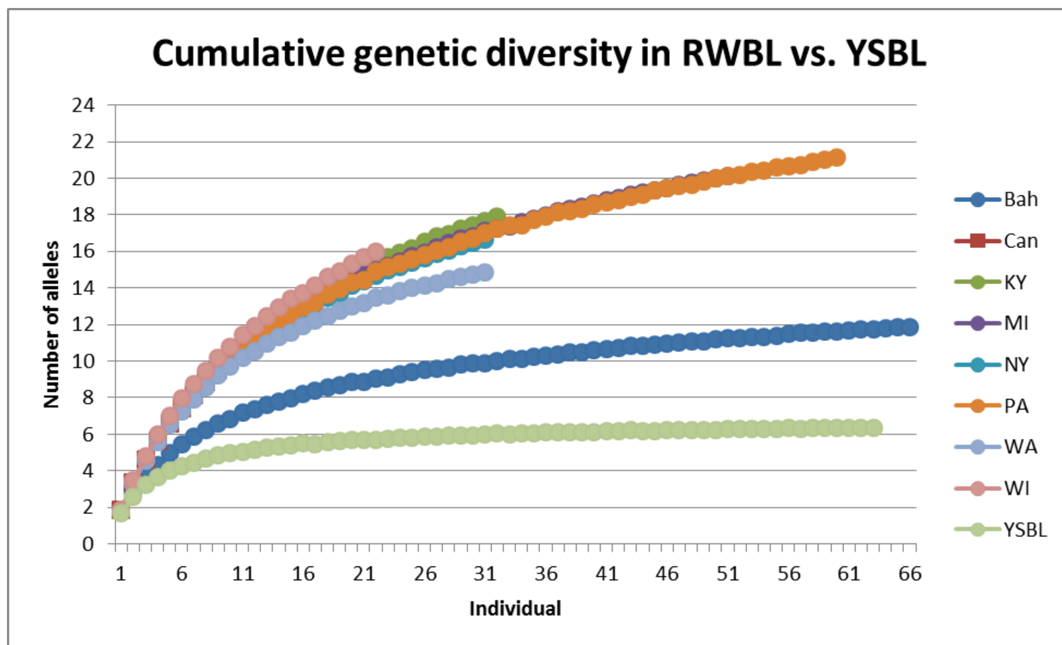

Figure 13: Rarefaction curves for eight red-winged blackbird and one yellow-shouldered blackbird population.

When compared to the seven continental red-winged blackbird populations, the yellow-shouldered blackbird's genetic diversity was significantly lower. Even using raw allele count as a metric, which included the Canada population's low allele count given its low sample size (n = 13), the decrease in allelic diversity was significant (N_r: t = 5.14, df = 7, P = 0.0013; N_s: t = 5.49, df = 7, P = 0.0009; I: t = 5.06, df =7, P = 0.0015, Table 7). This result is similar to the Bahamas red-winged blackbird populations' lower diversity relative to the continental populations (see Chapter 2). The yellow-shouldered blackbird's levels of expected and observed heterozygosity were also lower than those of the continental red-winged blackbird populations (H_o: t = 3.45, df = 7, P = 0.011; H_e: t = 4.05, df = 7, P = 0.005). However, the yellow-shouldered blackbirds did not have

significantly different levels of inbreeding relative to the continental red-winged blackbirds (F_{IS}: t = 1.17, df = 7, P = 0.28).

Table 7: Comparison of genetic diversity at eight loci in one yellow-shouldered blackbird population and eight red-winged blackbird populations.

Pop/Sp	N		N_r	N_s	I	H_o	H_e	F_{IS}
YSBL	63	Mean	6.38	5.24	0.68	1.41	0.68	0.003
		SE	1.53	1.05	0.06	0.22	0.06	0.018
Bahamas	66		11.88	7.62	0.74	1.80	0.73	0.010
			3.35	1.62	0.06	0.29	0.06	0.018
Canada	13		11.63	11.63	0.85	2.17	0.86	-0.008
			1.19	1.19	0.02	0.14	0.04	0.039
KY	32		17.88	12.11	0.88	2.49	0.86	0.021
			2.50	1.15	0.02	0.16	0.02	0.019
MI	51		20.13	11.92	0.88	2.54	0.83	0.063
			2.80	1.12	0.02	0.17	0.03	0.022
NY	31		16.63	11.60	0.88	2.44	0.82	0.067
			2.10	1.10	0.02	0.15	0.03	0.023
PA	60		21.13	11.69	0.89	2.53	0.87	0.014
			3.17	1.26	0.02	0.18	0.02	0.018
WA	31		14.88	10.92	0.87	2.33	0.83	0.037
			2.17	1.22	0.03	0.17	0.03	0.019
WI	22		16.00	12.43	0.89	2.46	0.89	0.003
			1.91	1.25	0.02	0.15	0.04	0.034

3.3.4 Evidence of bottleneck

BOTTLENECK detected increases in observed heterozygosity versus expected heterozygosity as calculated from all three models. Using the Wilcoxon test, these differences were significant at each model (I.A.M., P = 0.002, T.P.M., P = 0.002, S.M.M., P = 0.049). However, using the standardized differences and sign tests, the difference in observed heterozygosity and expected heterozygosity as predicted by the S.M.M. was

not significant (standardized differences: I.A.M., P = 0.00005, T.P.M., P = 0.001, S.M.M., P = 0.073; sign test: I.A.M., P = 0.004, T.P.M., P = 0.005, S.M.M., P = 0.19). Explaining the sign test's results is the observation that the S.M.M. was the only model to calculate greater values for expected vs. observed heterozygosity at two of the nine loci, whereas the other models calculated smaller values for expected vs. observed heterozygosity at all loci. BOTTLENECK reported a normal L-shaped distribution, indicating that there was no mode-shift in the allele frequency distribution (i.e., no deviation from allele frequencies predicted by drift-mutation equilibrium).

3.3.5. Mating system

Twenty of 87 chicks (0.23) were extra-pair, while eleven of 30 nests (0.37) contained at least one EPY. These proportions were not significantly different from EPP calculated from a population of island (Bahamas) red-winged blackbirds, where 16 of 56 chicks were extra-pair and 10 of 20 nests contained one EPY (chicks: Pearson's χ^2 = 0.56, df = 1, P = 0.45; nests: χ^2 = 0.88, df = 1, P = 0.35).

For the logistic regression models, Model 1 estimated the intercept μ to be -0.64 ± 0.04, indicating that the overall probability across all populations of being an EPY is $\frac{e^{-0.64}}{1+e^{-0.64}}$. However, the model allowing for population-specific intercepts (Model 2, Table 8) was a significantly better fit than the model considering a single intercept (ANOVA chi-square between Model 1 and Model 2, χ^2 = 54.2, df = 8, P < 0.0001). This result identified population as a meaningful variable and was consistent with the chi-square

tests (above) showing that EPP rate varied significantly by population (and, by extension, species). However, these models are almost identical to the model using only the eight red-winged blackbird populations (Chapter 2), suggesting the addition of the yellow-shouldered blackbird has a negligible contribution to the variation already present among red-winged blackbird populations.

The multiple comparison of means showed that the yellow-shouldered blackbird population was more than 0.5 standard deviations from the grand mean (Table 9). The New York population, which had an EPP rate of 0.24, was similarly far from the grand mean. However, neither deviation was found to be significant (P = 0.21 for yellow-shouldered blackbird), indicating the effect size was not sufficient to result in a significant contrast from the grand mean. In addition, the Tukey pairwise tests showed that the yellow-shouldered blackbird's mean EPP rate was significantly different only from the Kentucky red-winged blackbird population (data not shown). Given Kentucky's deviation from the other populations (Chapter 2; other pairwise comparisons in grand mean test), Kentucky is likely the population responsible for the difference.

Table 8: Summary of output for Model 2, the GLM incorporating population as a variable, using the number of EPY as a measure for EPP rate.

	Estimate	Std. error	z-score	Pr (>|z|)
$\mu_{Bahamas}$	-0.92	0.30	-3.10	0.0020
μ_{Canada}	-0.11	0.33	-0.34	0.73
μ_{KY}	0.51	0.30	1.71	0.087
μ_{MI}	-0.15	0.36	-0.42	0.68
μ_{NY}	-0.25	0.33	-0.76	0.45
μ_{PA}	-0.11	0.38	-0.28	0.78
μ_{WA}	0.24	0.31	0.77	0.44
μ_{WI}	0.16	0.37	0.44	0.66
μ_{YSBL}	-0.29	0.39	-0.75	0.45

Table 9: Multiple comparison against grand mean for EPP rate, measured by number of EPY.

	Estimate	Std. error	z value	Pr (>|z|)
Bahamas	-0.26	0.29	-0.89	0.98
Canada	-0.37	0.14	-2.70	0.06
KY	0.25	0.04	6.59	<0.001
MI	-0.41	0.20	-2.06	0.29
NY	-0.51	0.15	-3.50	0.004
PA	-0.37	0.24	-1.54	0.67
WA	-0.02	0.10	-0.20	1
WI	-0.10	0.21	-0.47	0.99
YSBL	-0.55	0.25	-2.22	0.21

The next series of tests defined EPP rate as the number of nests containing at least one EPY. Model 1 estimated the intercept μ to be 0.66 ± 0.06, with a significance of $P = 0.30$, indicating the number of nests containing at least one EPY did not differ significantly across all samples. An ANOVA chi-square on the logistic regression models

showed that Model 2 considering population as a variable was marginally significant compared to Model 1 considering only the intercept (χ^2 = 15.5, df = 8, P = 0.051, Table 10).

Similar to the measure of number of EPY, the multiple comparison for number of nests with EPY showed that the yellow-shouldered blackbird is > 0.5 standard deviations away from the grand mean but that the contrast was not significant (P = 0.59, Table 11). There were no pairwise comparisons where any population differed significantly from each other (data not shown).

Table 10: Summary of output for Model 2, the GLM incorporating population as a variable, using the number of nests containing at least one EPY as a measure for EPP rate.

	Estimate	Std. error	z-score	Pr (>|z|)
μ_{Bahamas}	-1.61E-14	4.47E-01	0	1
μ_{Canada}	-4.70E-01	5.04E-01	-0.93	0.35
μ_{KY}	1.98E-01	4.56E-01	0.44	0.66
μ_{MI}	7.49E-15	5.48E-01	0	1
μ_{NY}	-3.57E-01	5.11E-01	-0.70	0.48
μ_{PA}	-7.41E-02	5.90E-01	-0.13	0.90
μ_{WA}	1.50E-01	4.80E-01	0.31	0.76
μ_{WI}	5.11E-01	5.77E-01	0.88	0.38
μ_{YSBL}	-5.47E-01	5.86E-01	-0.93	0.35

Table 11: Multiple comparison against grand mean for EPP rate, measured by number of nests containing at least one EPY.

	Estimate	Std. error	z value	Pr (>|z|)
Bahamas	-0.066	0.44	-0.15	1
Canada	-0.54	0.22	-2.4	0.13
KY	0.13	0.058	2.28	0.18
MI	-0.066	0.31	-0.21	1
NY	-0.422	0.24	-1.78	0.49
PA	-0.14	0.38	-0.37	1
WA	0.08	0.16	0.52	1
WI	0.44	0.36	1.24	0.87
YSBL	-0.61	0.37	-1.64	0.60

3.3.6. Parentage assignment

COLONY found that ten paternal genotypes explained the 20 EPY. Eight were inferred genotypes, while two matched the genotypes of males included in the list of sampled males. The eight inferred genotypes explained paternity for 17 of the 20 extra-pair chicks. Because they did not match the genotypes of any of the candidate males, this result suggests that the genetic fathers were outside the pool of sampled males. For the remaining three chicks, COLONY matched their genetic sires to the genotypes of two sampled males. Two of the chicks, which were at separate nests, shared the same genetic father. Notably, that male lost all paternity at his own two-chick nest but (assuming the assignment is correct) was able to regain it at those other two nests. For the third chick, COLONY also matched the extra-pair male's genotype with that of a sampled male. Unlike the other identified male, this male sired his own two within-pair offspring in

addition to producing the extra-pair chick. At two of the 11 nests with extra-pair young, 100% of the chicks were EPY (Table 12).

All but one of the 30 sampled males fathered at least one chick. The exception was a male that did not sire any of his four social offspring and additionally was not found to be a genetic father elsewhere. At this male's nest (nest 96, Table 12), pooled offspring genotypes revealed five alleles at one of the loci, suggesting the female had mated with multiple extra-pair males. COLONY predicted that three extra-pair males were responsible for the chick genotypes.

COLONY's maximum-likelihood configuration predicted that females had between 1-3 extra-pair mates (Table 12). In addition, six of the ten putative genetic fathers sired extra-pair chicks across multiple nests (Table 13). A caveat to these inferences is that, given the low genetic diversity, errors in assignment could have occurred in inferring fathers of genetically similar chicks across different nests. Because the combination of multilocus genotypes is limited, the same paternal genotype could potentially explain paternity for chicks that were actually sired by different males. Therefore, the inferred assignments should be viewed as conservative estimates of the number of extra-pair males.

If these eight inferred males are accurate, then males sired an average of 2.29 ±1.01 chicks. Again, the caveat is that the within-pair success for the inferred males is

unknown, so there is considerable room for error in this estimate. Females raised an average of 2.9 ± 0.71 chicks.

Table 12: Maximum-likelihood configuration of EPP broken down by nest. No. EPM = the number of extra-pair males thought to sire chicks in each nest. No. total males = the total number of genetic fathers at the nest. At two nests (101 and 96), the social male did not gain any paternity at his own nest.

Nest ID	No. EPM	No. total males
69	1	2
139	2	3
22	1	2
101	2	2
108	1	2
28	1	2
106	1	2
88	2	3
96	3	3
44	1	2
131	2	3

Table 13: Maximum-likelihood configuration of paternity by ten extra-pair males (EPM) distributed across chicks and nests. For example, the first male fathered three chicks in two separate nests. IDs for the last two males are USFWS band numbers.

Putative EPM	No. EPY	No. nests
M1	3	2
M2	2	2
M3	2	2
M4	1	1
M5	3	3
M6	3	2
M7	2	1
M8	1	1
1222-38300	2	2
1292-36614	1	1

3.4 Discussion

The yellow-shouldered blackbird recovered three decades ago from a severe population decline, aided by a two-pronged conservation effort that promoted blackbird survival while controlling shiny cowbird brood parasitism. However, recent plateaus in recruitment and survival rates have resulted once more in decreasing population sizes. The effective population size of a managed population in southwest Puerto Rico is about 65 individuals, 15% of the current census size, and the genetic diversity is lower than that of an island-dwelling population of red-winged blackbirds. These low values may have been caused by a bottleneck in the 1980s brought on by the effects of the shiny cowbird and habitat loss, and they point to impending loss of evolutionary potential in the species. Fortunately, the population does not appear to exhibit immediate signs of inbreeding, and observed heterozygosity matches expected levels calculated from allele frequencies. In addition, genetic diversity and effective population size do not appear to be adversely impacted by the blackbird's genetically polyandrous mating system. These results suggest that the population appears to be genetically stable in the short-term, but that as a whole it remains vulnerable to stochastic change and other factors that can trigger extinction. The greatest concern is that the forces that led to the recent bottleneck continue to threaten the blackbird's survival. Efforts to stem the decline in population size are greatly needed to delay the onset of inbreeding depression and prevent further genetic deterioration.

3.4.1 Low effective population size and genetic diversity in the yellow-shouldered blackbird

An estimated effective population size of 65 signifies that the population is losing heterozygosity (or becoming inbred or drifting) at the same rate as an ideal population of 65, regardless of the actual census size (Charlesworth 2009; Frankham *et al.* 2010). This value is barely above the conventional minimum N_e of 50 required to avoid the immediate effects of inbreeding depression (Soulé 1980). While F_{IS} values did not reveal current signatures of inbreeding at any of the ten loci, inbreeding inevitably occurs in a finite, closed population at an increase of $1/(2N_e)$ per generation (Johnson 1977). This equation corresponds to a rate of (1/130) = 0.008 for the yellow-shouldered blackbird. If all other factors are constant, then in only 18 generations, the inbreeding coefficient will increase from the current average of 0.003 (Table 7) to 0.15, reaching a threshold where reductions in reproductive and other fitness components are likely to occur (Charlesworth & Charlesworth 1987; Naish *et al.* 2013).

To permanently retain evolutionary potential, an N_e of 500-5000 is required (Frankham *et al.* 2010). In birds, minimum viable population size is thought to be exceed a few thousand individuals, ideally above 3000 individuals (Traill *et al.* 2007). The yellow-shouldered blackbird's population size is almost two orders of magnitude below this quantity. The Bahamas red-winged blackbird population size is also lower than the critical range (N_e ~300 at a minimum allele frequency of 0.01, Table 5), but it has the advantage of having a vast genetic source in the continental North American population

of red-winged blackbirds. Recolonization possibilities for the yellow-shouldered blackbird are limited but, if necessary, include outbreeding with the southern population or the subspecies in Mona and Monito islands to the west of Puerto Rico (Barnés 1945).

The small effective population size is likely a result of the bottleneck that led to the species' endangered listing in 1976. A bottleneck signature persists 0.2-4N_e generations after the bottleneck occurs (Hundertmark & Van Daele 2010). If N_e is approximately 65 individuals, then the signature can be expected to be visible 13-260 generations after the bottleneck. Given that the generation time of the blackbirds is one year, approximately 38 generations have passed since the known bottleneck of 200 individuals, indicating that the results fall within the timeframe and reflect the signature of that particular event.

The bottleneck likely also contributed to the low observed genetic diversity. While the yellow-shouldered blackbirds had greater allelic diversity than the Bahamas red-winged blackbirds at two loci, counts of cumulative diversity showed it had less diversity over all loci than the Bahamas population and significantly less diversity than seven continental red-winged blackbird populations (Fig. 13). Encouragingly, individual heterozygosity was intact, indicating that recessive deleterious alleles in the population have not yet been expressed at high levels. High heterozygosity could also be a footprint of the bottleneck, as it was the metric used in the program BOTTLENECK to detect

evidence of sudden population reduction and recovery (Cornuet & Luikart 1996). Combined with the negative F_{IS} values observed at five of the loci, the heterozygosity suggests individuals could be engaging in disassortative mating. This pattern is similar to that proposed for the Bahamas red-winged blackbirds, which also had low levels of genetic diversity but no evidence for inbreeding depression and no difference between observed and expected levels of heterozygosity (Chapter 2). However, the yellow-shouldered blackbird's expected heterozygosity is still lower in absolute terms than that expected in the red-winged blackbird populations (significantly so when compared to the continental populations). In addition, loss of heterozygosity is dependent on effective population size and experiences exponential decay (Crow & Kimura 1970). If half of initial heterozygosity is lost in $1.4N_e$ generations, then using the 2012 population as a beginning time point suggests that the population will lose half its current heterozygosity in $1.4 \times 65 = 91$ generations.

3.4.2. Mating system does not contribute adversely to N_e

I provided the first evidence that yellow-shouldered blackbirds engage in EPP and thus are genetically polyandrous. From the empirical data and the inferred paternal genotypes, there was no evidence for reproductive skew (i.e., no evidence that any of the sampled or inferred males fathered a disproportionate number of chicks). In fact, there was little overlap between within-pair and extra-pair males in my sample, with only two territory owners also identified as extra-pair sires. This finding suggests (reassuringly)

that many more males were successfully gaining copulations in the population than the 30 males I sampled. However, precisely because of the lack of overlap in identities, estimates of individual male reproductive success are incomplete. Except for the two sampled males that also sired EPY, extra-pair success is unknown for the 28 remaining sampled males, while within-pair success is unknown for the eight inferred extra-pair males.

A more quantitative approach to evaluating variance in reproductive output with the known sample is to use the model proposed in Webster *et al.* (1995), in which sources of variance within-pair and extra-pair success are calculated separately and then combined to measure an individual's total reproductive success. These sources of variance include the number of social (and extra-pair) mates; the number of young per social (and extra-pair) mate; and proportion of social (and extra-pair) mates' young sired. The 15 combinations of covariance between any two of these terms are also calculated. The goal is to partition and then rank the relative (co)variances in fitness components to see which terms show the greatest variance and are thus the most likely targets for sexual selection. Studies have applied this approach in many species with EPP, with some showing EPP does increase the opportunity for sexual selection (blue tits, *Cyanistes caeruleus*, Schlicht & Kempenaers 2013; red-winged blackbirds, Webster *et al.* 1995) and others showing its effect is limited (savannah sparrows, *Passerculus sandwichensis*, Freeman-Gallant *et al.* 2005; Mediterranean blue tits, García-Navas *et al.*

2014). Ultimately, this approach could help identify the importance of both individual and interactive fitness components in the yellow-shouldered blackbird. For example, positive covariance between number of within-pair and extra-pair young would suggest the potential for EPP to lead to reproductive skew.

At present, EPP does not seem to increase variance in male reproductive success, suggesting that it has a neutral effect on effective population size. As for genetic diversity, it is difficult to predict the effects of EPP from a single breeding season, but EPP could increase or decrease standing genetic variation in a variety of ways. For example, knowing whether the extra-pair males are territorial males or nonbreeding adult males (i.e., floaters) could shed insight on how permissive EPP allows the pool of breeding individuals to be and whether it overrides socially determined breeding opportunities. Enabling otherwise nonbreeding males to mate could either have positive effects on genetic diversity (by expanding the size of the breeding population), adverse effects on the population (if nonbreeding status is a reflection of poor genetic quality), or a mixture of both. Additionally, given the low inbreeding coefficient observed despite low genetic diversity, females could be using EPP as a strategy to increase heterozygosity in their offspring, thus maintaining genetic diversity (Richardson *et al.* 2005; Stapleton *et al.* 2007). This hypothesis is consistent with multiple mating as a form of evolutionary rescue or bet-hedging in unpredictable or unfavorable environments, although it is unlikely that EPP is a particularly plastic trait that could respond quickly

to such changes. Instead, EPP could simply be a life-history trait that happens to give a selective advantage over a system like monogamy that is more prone to extinction in stressful environments (Plesnar-Bielak *et al.* 2012; Saether *et al.* 2004). Finally, EPP could have heterogeneous effects across the genome, especially for paternally inherited components of the genome (Evans & Charlesworth 2013), although this idea is far from being documented in the species. Thus, while preliminary results do not demonstrate a visible effect of EPP on genetic diversity, EPP has potential in multiple ways to influence this population genetic trait.

From an evolutionary standpoint, EPP in the yellow-shouldered blackbird was not expected to occur at similar levels to the red-winged blackbird. As a socially monogamous species, the yellow-shouldered blackbird exhibited behavioral and morphological traits distinct from the two socially polygynous congeners (red-winged and tricolored blackbird) that suggested it might be genetically monogamous. However, genetic polyandry in the yellow-shouldered blackbird could be explained by the evolutionary history of icterids (Chapter 1). If the traits thought to illustrate the presence of strong sexual selection in males are actually results of evolutionary changes in females, such as gaining cryptic dimorphic plumage and losing song (Price 2009; Price *et al.* 2009), then the current traits seen in males are ancestral and will not predict the type of mating system exhibited in each species. Nevertheless, it is puzzling that changes in male-specific behavior, such as amount of male parental care, lack a correlation with

mating system. Males should not be expected to provision within-pair offspring at such high rates if there is a strong likelihood that some of the chicks are extra-pair, yet the male yellow-shouldered blackbirds fed as often as females, to the point that they were just as easy to capture as females with walk-in nest traps. Given this inconsistency, the genetic mating systems of the red-shouldered blackbird and tawny-shouldered blackbird, the two other island-dwelling congeners that share the same suite of traits with the yellow-shouldered blackbird, are now even more of an open question.

3.4.3. Management implications

The yellow-shouldered blackbird's depressed effective population size, and the continuing threats preventing it from recovering to pre-bottleneck levels, renews the urgency for policies that ensure the long-term maintenance of the species. No significant evidence of inbreeding depression or loss of heterozygosity is yet observed, but in the absence of targeted management, default rates of drift and heterozygosity erosion will outstrip the (much slower) rates of recovery from mutation and recombination, the only two sources of genetic innovation available for this island endemic. With the knowledge that reaching a "safe" effective population size is likely impossible, the next steps are to consider which policies can best prevent further declines in effective population size while monitoring the species' extinction risk over time.

Recommendations in the PHVA already emphasize continued management during the breeding season of both the yellow-shouldered blackbird and shiny cowbird.

For continued sampling of the blackbird's genetic profile, multigenerational sampling may be useful in tracking and increasing the precision of N_e. Longitudinal samples may also help determine the fates of extra-pair and within-pair young and tease out whether there are fitness differences between the two groups (Johnsen *et al.* 2000; Kempenaers *et al.* 1997; Sardell *et al.* 2012; Sheldon *et al.* 1997; but see Kleven et al. 2006). Additionally, genetic analyses of other populations may provide valuable information about the extent of population structure. With a small population in southern Puerto Rico and a population of a morphological subspecies on Mona Island, characterizing the profile of all extant populations can help evaluate each one's potential as a source population, should translocation, outbreeding or other options designed to increase gene flow become warranted and feasible.

With financial constraints an ongoing issue for the yellow-shouldered blackbird management program, periodic censuses and modeling may need to be an alternative to yearly genetic sampling and molecular analysis (Medina-Miranda *et al.* 2013). Nevertheless, the information in the present study provides a baseline measurement of the yellow-shouldered blackbird's population genetics and mating system that can be incorporated into future PHVAs assessing the blackbird's survival outlook.

4. Transcriptomic and proteomic identification of seminal fluid proteins in a genetically polyandrous songbird

4.1. Introduction

Advances in next-generation sequencing and bioinformatics pipelines have enabled transcriptomic and proteomic profiling of non-model animal species, including species studied exclusively in the wild. This technology has greatly increased the power to tackle fundamental questions in evolutionary biology, such as identifying the genetic bases of behavior (Weber *et al.* 2013; Zwarts *et al.* 2011), reconstructing divergence events such as speciation or domestication (Axelsson *et al.* 2013; Ellegren *et al.* 2012; Rubin *et al.* 2012), or describing fitness consequences of genetic variation (Hoekstra *et al.* 2006). Within the realm of behavioral ecology, next-generation sequencing has added a genetic dimension to tests of some of the best-known paradigms, such as sensory exploitation, heterozygote fitness, and the lek paradox (Alem *et al.* 2013; Hoffman *et al.* 2014; Johnston *et al.* 2013).

Birds are a clade that stand to benefit greatly from these integrative approaches. Traditionally studied from a behavioral perspective, they represent an exciting group in which to explore the intersection between selection at the organismal and genomic levels. Like many non-model vertebrate systems, they have a limited amount of available genetic and genomic resources, with the potential for evolutionary applications remaining largely untapped. However, transcriptomes of tissues from non-model

passerines are becoming increasingly available (Künstner *et al.* 2010; Peterson *et al.* 2012; Santure *et al.* 2011), as are proteomic approaches to identify and characterize proteins (Altelaar *et al.* 2013; Diz *et al.* 2012; Papakostas *et al.* 2012). Together, these two approaches can be used to identify potential genetic mechanisms underlying behavioral traits, even in organisms without annotated genomes.

Here I use a combination of transcriptomic and proteomic methods to produce the first catalog of seminal fluid proteins (Sfps) in the red-winged blackbird (*Agelaius phoeniceus*), a widespread North American species whose life history has been the focus of study for decades (Beletsky 1996; Orians 1969; Searcy & Yasukawa 1995). The blackbird's breeding biology, particularly with respect to male and female reproductive decisions, is well known, making it an excellent study species with which to examine the link between behavioral and molecular responses to sexual selection. However, like most songbirds, it lacks a reference genome and accompanying resources. Thus, any questions focusing on molecular evolutionary comparisons cannot be asked without first identifying the molecular targets of interest.

To circumvent the lack of genomic resources, genes can be identified by examining the products of transcription and translation. If both protein and RNA samples are available, then information about both can ultimately lead to identification of candidate genes for evolutionary analysis. RNA-Seq is now a popular high-throughput method with advantages over expressed sequence tag (EST) libraries, while

tandem mass spectrometry (MS/MS) has emerged as a powerful method to identify proteins (Aebersold & Mann 2003; Steen & Mann 2004). Sfps in particular have had consistent success with these joint transcriptomic (or EST) and proteomic methods (Andres *et al.* 2008; Dorus *et al.* 2006; Findlay & Swanson 2010; Findlay *et al.* 2008; Walters & Harrison 2010). Once these proteins are identified, downstream analyses such as gene differential expression and annotation are now also possible with open-source software.

In the present study, I identified Sfps from field-collected protein and RNA samples. I sequenced the transcriptomes from two reproductive and two control red-winged blackbird tissues, generated a custom database by translating the transcriptomes, and used the resulting blackbird proteome to identify mass spectra from proteins recovered in red-winged blackbird seminal fluid. I annotated the 612 testis-derived transcripts and performed differential expression (DE) analysis to assess the variation in expression across the four tissues. I then compared the blackbird's annotated Sfp profile with those of mammals and *Drosophila*. This catalog gives the opportunity to identify and sequence specific proteins for evolutionary analysis, including to examine molecular responses to sperm competition.

4.2. Methods

4.2.1. Protein collection and protein identification with zebra finch proteome

Seminal fluid was collected by S. Nowicki from a male red-winged blackbird at Conneaut Marsh in northwest Pennsylvania in June 2011. The male copulated with a female mount wired in a copulation solicitation pose and placed on his territory. Previous studies collected seminal fluid using a false cloaca, such as a Pasteur pipette bulb, inserted into the vent (Pellatt & Birkhead 1994; Westneat *et al.* 1998). However, during initial tests where I used a 0.6-ml microcentrifuge tube, the male did not consistently deposit seminal fluid into the false cloaca. Therefore, the entire ventral region was sterilized, and the male was allowed to deposit onto the undertail coverts. Avian seminal fluid is generally more condensed than mammalian seminal fluid, allowing the sample to remain intact during transfer and removal. Samples were removed with an stainless steel scoop and immersed in a 1.5-ml microcentrifuge tube containing 1M PBS solution treated with protease and phosphatase inhibitor tablets (Roche), then stored at -80°C.

Proteins were sequenced using tandem mass spectrometry (MS/MS) at the Proteomics Facility in October 2011. To separate sperm from seminal fluid, the solution was spun at 10,000g at 4°C for 10 minutes (Cornwallis & O'Connor 2009; Mohan *et al.* 1995). I then removed the supernatant, placed it in a new tube, and submitted the sample to Duke Proteomics Facility on dry ice. At the time of analysis, the facility spun

the sample again at 10,000g at 4°C for 5 minutes and drew off the top of the supernatant immediately after spinning. A Bradford assay was performed on the supernatant to measure total protein content. Up to 30-40 μg total protein was run on a polyacrylamide (SDS-PAGE) gel. The gel was imaged, then divided into 24 gel band samples (Fig. 14) for in-gel (long) digestion and MS/MS using the FastMS2 protocol.

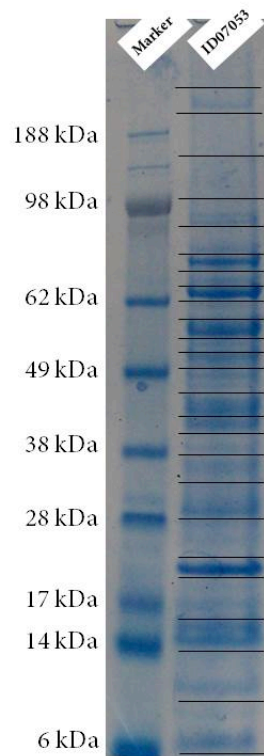

Figure 14: Image of SDS-PAGE gel with ladder (left) and blackbird seminal fluid (right). Protein masses of different ladder bands are shown. The gel was cut into 24 sections, indicated by the horizontal lines, and each section was individually analyzed with MS/MS.

The facility performed a complementary search using MASCOT (Matrix Science) to interpret the mass spectra. Mass spectra were matched against candidate proteins from two NCBI proteomic databases: all known proteins from Chordata, and all known

proteins in zebra finch (*Taeniopygia guttata*), predicted *in silico* from the zebra finch genome. Protein mass spectra and their highest-likelihood matches (in the form of NCBI accession numbers and names) were viewed on the program Scaffold v.4 (Proteome Software). I filtered hits at 90% minimum protein and peptide identification probabilities. Known contaminants such as keratin were removed.

To obtain more information about the zebra finch proteins, I used the NCBI (and, if necessary, Ensembl) accession numbers to perform a search on the data-mining tool BioMart, hosted on Ensembl. BioMart retrieves and compiles accession numbers and other data from multiple databases for user-submitted sequences. For each protein, I obtained GO terms and accession numbers for genes, transcripts and proteins from UniProt, NCBI and Ensembl. From the GO terms generated, I grouped the zebra finch proteins into broad functional categories.

4.2.2. RNA collection and transcriptome assembly

Two reproductive (testis and seminal glomera) and control (heart and liver) tissues were collected by S. Nowicki from a different male blackbird. The male was caught by mist net and sacrificed by decapitation. Samples were collected with RNAse-free tools and immediately immersed in RNAlater (Ambion, Life Technologies). Following manufacturer instructions, samples were incubated at 4°C for 24 hours, then stored at -20°C.

RNA was extracted and submitted in October 2011. Extractions were performed with K. Rosvall at Indiana University using Trizol (Invitrogen). Samples were analyzed by spectrophotometer for concentration and purity, then stored at -80°C. At least 1000 ng of each sample were submitted in 50 µl nuclease-free water to the Duke Sequencing Facility. Samples were sequenced on two lanes of Illumina HiSeq 2000 to produce 100-bp paired-end reads. Reproductive tissue samples were run on one lane and control tissue samples on the other. The sequencing facility inspected samples for RNA integrity (following a protocol by Agilent), then prepared Illumina mRNA libraries using the TruSeq sample preparation protocol. Sequences were received from the Duke Sequencing Facility in January 2012

For each tissue, the facility returned two files (R1 and R2) corresponding to the sequences that formed the paired-end reads. Reads were inspected with FastQC v0.10.1 (Andrews 2010). This program returns per-base and per-site statistics such as quality scores and GC content, as well as sequence length distributions and duplication levels. Fragments were then cleaned with the following steps. Using custom Python scripts (N. Devos, pers. comm.), I first joined corresponding paired-end (R1 and R2) reads, removed duplicate joined reads, and then split the fragments again. This step was completed to remove duplicates that were present due to PCR amplification and not to biological differences in abundance. I next clipped Illumina adapters and trimmed the remaining reads by Phred score (cutoff = 25). At this point, any fragments fewer than 25bp were

also discarded. This step resulted in the presence of "orphan" single-end reads, which were sorted from the remaining paired-end reads using the *compfastq* script (Newhouse & To 2011). The single-end reads were then concatenated with the paired-end reads to form two final, cleaned sets of reads for each tissue. With these files, I generated transcriptome assemblies for all four tissues with Trinity (http://trinityrnaseq.sourceforge.net/, version released 25 Feb. 2013, Grabherr *et al.* 2011). Assemblies were performed on the Blacklight server of Pittsburgh Supercomputing Center.

A potential concern was that Trinity's default settings, imposed to minimize false discovery rates (FDR) from spurious assemblies, would discard low-abundance reads and thus overlook rare transcripts. However, because I had a protein sample with which to validate the transcriptome, a slightly higher FDR was acceptable. Therefore, after the first assembly was produced, I used Bowtie v.1 (Langmead *et al.* 2009) to align the original reads against the transcripts. During this alignment, I treated all reads as single-end (instead of paired-end) reads. I then specified that Bowtie produce an output file containing all reads that did not map back to the Trinity assembly. With those unmapped reads, I ran a second assembly on Trinity with a lower minimum contig length, decreased from the default of 200 bp to 140 bp (and thus allowing up to a 20-amino-acid decrease in length). Because Trinity randomly assigns component IDs to transcripts, the chance existed that there could be duplicate IDs for transcripts across the

two files. Therefore, assemblies were not merged, but instead treated as separate files for the rest of the analysis.

To evaluate the quality of the reads, I measured N50, the length of the shortest contig in the top half of the total length of all contigs (Yandell & Ence 2012), for each transcriptome. I also used a custom Python script (M. Johnson, pers. comm.) to calculate the reciprocal best BLAST hit (O'Neil & Emrich 2013) using the zebra finch proteome as the query database. Each of the transcripts in my assemblies was BLASTed to the zebra finch proteome. The highest-scoring gene was then taken from the zebra finch proteome and BLASTed to my assembly. This method gives an idea of the percent of orthologs recovered from the assembly.

4.2.3. Protein identification with blackbird proteome

The two transcriptome assemblies per gene enabled me to make a custom, species-specific proteome with which to search the mass spectra. This database search complemented the search with the zebra finch proteome. Using two pipelines that took advantage of my having proteomic data (L. Pachtor, University of California-Berkeley, pers. comm.), I translated both the original Illumina reads and the finished transcriptomes and submitted two separate proteomic databases to the Proteomics Facility. The first database was ultimately unable to be run at the facility because of its size, but the second was run successfully.

(1) *Align-then-assemble* (unsuccessful): This process theoretically entailed six-way translation of the original Illumina reads using custom Python scripts (S. Wu, pers. comm.), followed by database searching against the mass spectra and finishing with transcriptomic assembly of the hits using Trinity. Because of the high number of reads (Table 14), the database produced was 214G, too large to be run at the facility but available for future searches. However, a caveat is that the risk of missed hits is high, because the small fragment length (max = 100 bp or 33 amino acids) decreased the likelihood that the Proteomics Facility's program could find reliable matches using their search parameters. Therefore, a complementary approach, below, used assembled transcripts (max = 1900 amino acids) to find potential matches between the RNA and protein samples.

(2) *Assemble-then-align* (successful): This process entailed six-way translation of the two Trinity assemblies per tissue (the original assembly and the assembly with unmapped reads) into tissue-specific proteomes, followed by database searching against the mass spectra. The proteomes were "barcoded" with tissue names, then concatenated to produce a 2G database that was searched against the protein mass spectra at the Proteomics Facility. Results were returned in July 2013.

4.2.4. Differential expression analysis

Once the results were returned, I again used Scaffold to view results from the database search. Hits between the custom database and the peptide mass spectra were

identified not by NCBI accession numbers (as they were in the zebra finch proteome), but instead by the Trinity-generated component IDs in the blackbird proteome.

The majority of mass spectra matched transcripts from all four tissues. A preliminary search using Blast2GO (Conesa *et al.* 2005) confirmed that many proteins were not tissue-specific but instead expressed across multiple tissues. However, abundance levels could still vary across tissues, and a quantitative comparison of protein expression levels could indicate their relative importance to each tissue. This step can be accomplished with gene differential expression (DE) analysis (Mortazavi *et al.* 2008). First, to select the proteins of interest, I narrowed my analysis to proteins originating from a reproductive tissue by searching the mass spectra against only the testis proteome. This filtering yielded 612 mass spectra matching to testis-derived transcripts (see Results). Future work will include proteins from seminal glomera. However, as the field-collected testis sample was cleaner than the seminal glomera sample, I chose to use the testis proteome first to avoid inadvertently profiling proteins that were actually contaminants.

For the 612 contigs, I used the program RNA-Seq by Expectation Maximization (RSEM) v.1.2.11 (Li & Dewey 2011) to map Illumina reads from each tissue to the transcripts. RSEM produced a matrix for each tissue showing the number of reads matching to each of the target transcripts. Number of reads was measured by expected count, FPKM, and TPM (transcripts per million). "Expected count" refers to the number

of expected fragments assigned to the transcript given maximum likelihood transcript abundance estimates, while FPKM is defined as the number of fragments per kilobase effective length per million fragments mapped to all transcripts (Trinity webpage, http://trinityrnaseq.sourceforge.net/trinity_rnaseq_tutorial.html). Downstream analyses use expected count as the metric.

The last step in RSEM was to generate gene and isoform count matrices across multiple tissues. From these matrices, it was already apparent that certain tissues contained many more reads mapping to a given transcript than others. (See Table 17 in Chapter 5 for examples of counts.)

The isoform count matrix was then exported to EBSeq v.1.1.5 (Leng *et al.* 2013), an R package that uses empirical Bayesian analysis to calculate the posterior fold-change in protein expression level between conditions. I analyzed DE under two separate conditions: by treating each sample as a separate condition (multiple conditions with no replicates), and by pooling the control and reproductive tissues together (two conditions with two replicates each).

4.2.5. Transcriptome annotation

As products of *de novo* assembly, the transcripts produced by Trinity carry no information in terms of structure and function. Therefore, I separately used the Trinotate pipeline (http://trinotate.sourceforge.net/) to functionally annotate the 612 testis-derived transcripts. Trinotate uses multiple databases and a SQLite platform to compile results. I

first used the Trinity plug-in Transdecoder to translate each transcript six ways. Transdecoder produced an output file identifying the longest open reading frame (ORF), and thus the predicted reading frame, for each transcript. These peptide sequences were used as input for databases that conducted protein alignments (in contrast to the cDNA sequences used for nucleotide alignments).

I first ran a BLASTX search on the cDNA transcripts and a BLASTP search on the Transdecoder-predicted proteins. Second, following the instructions on Trinotate, I ran HMMER (Finn *et al.* 2011), which uses a profile-Hidden Markov Model to identify the most likely proteins in the Pfam database. Third, I ran SignalP (Petersen *et al.* 2011) to detect the presence of signal peptides, which indicate that proteins are secreted and are a typical feature of Sfps in the extracellular matrix (Swanson *et al.* 2001). Fourth, I ran tmHMM (Krogh *et al.* 2001) to identify transmembrane regions. I omitted the step of running RNAmmer (Lagesen *et al.* 2007) because identification of rRNA genes was outside the scope of the study.

I exported all results into a pregenerated SQLite database that was pulled from the Trinity ftp site and contained SwissProt-related annotation. I first populated the database with a gene/transcript map (distinguishing genes from isoforms), the original transcripts, and the Transdecoder-generated peptide file. I then built the database with the results of all four above previous alignment searches. The end result was a single

spreadsheet that contained annotation information from multiple databases for each isoform.

The final step was to integrate the DE and annotation results. I ranked all annotation results (from Trinotate) by the posterior fold change values (from RSEM and EBSeq). I chose to use fold-change values for the two-treatments, two-replicates condition comparing expression in reproductive vs. control tissues, instead of fold-change values for each individual tissue, because elevated expression in either the testis or seminal glomera could help identify Sfps. I then removed all isoforms and duplicate genes to find the number of unique genes recovered. I used BioMart to retrieve GO terms for all unique genes and categorized them under the same functional categories as the ones I used to describe the zebra finch proteins.

4.3. Results

4.3.1. Transcriptome assembly

RNA extracted from each tissue type yielded samples of sufficiently high quality to be submitted to the sequencing facility (testis: 2996 ng/μl; seminal glomera: 1030 ng/μl; heart: 464 ng/μl; liver: 1619 ng/μl). Tables 14 and 15 provide summary statistics for the quantity and quality of the Illumina reads.

There was noticeable variation in the number of proteins recovered from each tissue. The seminal glomera transcriptome was the smallest, while the other three tissues had comparable number of reads. This difference was also pronounced when counting

the number of transcripts assembled by Trinity. Interestingly, the number of transcripts in the initial assembly did not predict the number of transcripts in the assembly with the leftover reads.

For the quality statistics, RBH was at acceptable levels (40-60% of the total number of transcripts in the reference file), and N50 lengths were high for all assemblies (Tables 14 and 15). N50 statistics are generally less informative for transcriptome assemblies than genome assemblies because of the high expected number of duplicate sequences that could bias the calculation of N50 (O'Neil & Emrich 2013). For comparison, N75 and N95 lengths are also shown.

Table 14: Number of Illumina reads and number of transcripts assembled in Trinity for each tissue. "Leftover" refers to the second Trinity assembly using the unmapped reads that were identified with Bowtie.

Tissue	No. of reads	No. of transcripts	No. of transcripts (leftover)
Testis	63.2m	236,333	186,812
Seminal glomera	37.6m	108,491	85,559
Heart	57.7m	115,653	230,723
Liver	60.3m	198,110	95,689

Table 15: N50 statistics for all eight transcriptome assemblies (two per tissue). RBH = reciprocal best hit. Prop = number of RBH divided by the total number of proteins in the zebra finch reference assembly (n = 18,204).

Tissue	Total length (bp)	No. of contigs	N50 (bp)	N75 (bp)	N95 (bp)	RBH	Prop
Testis	294,511,602	236,333	2734	1190	291	10,777	0.59
Testis (leftover)	147,915,535	186,812	2394	733	180	10,804	0.59
Seminal glomera	122,880,743	108,491	2300	999	285	10,616	0.58

Seminal glomera (leftover)	56,446,445	85,559	1682	484	171	10,683	0.59
Heart	117,292,729	115,653	2108	823	264	9,699	0.53
Heart (leftover)	121,369,755	230,723	913	331	167	11,264	0.62
Liver	257,157,794	198,110	3153	1347	285	10,126	0.56
Liver (leftover)	90,459,369	95,689	2831	1052	193	9,487	0.52

4.3.2. Protein identification with heterospecific and conspecific proteomes

Following submission of the seminal fluid sample to the Proteomics Facility, all spectra were searched against two reference databases: the NCBI zebra finch database and the NCBI chordate database, selected to give a more generalized view. In the Scaffold file using the NCBI zebra finch proteome, a total of 88 proteins with at least two unique peptides to match were identified across the 24 gel bands. With contaminants such as keratin removed, the total decreased to 84 proteins. When this criterion was lowered to one unique peptide, a total of 113 proteins were identified. In the Scaffold file using the NCBI Chordata search results, a total of 106 proteins with at least two unique peptides to match were identified across the 24 gel bands. When this criterion was once again lowered to one unique peptide, a total of 107 proteins were identified. I used the more conservative criterion of two unique peptides to lower the FDR. I additionally focused on results from the search with the zebra finch proteome to prevent biasing my search for candidate Sfps in proteins found across chordates, since they are likely to be conserved.

For the third search using the red-winged blackbird proteome, a total of 157 unique proteins were annotated from the 612 testis-derived transcripts. This observation suggests that splice variants were common (Dean *et al.* 2009). Because I removed all duplicate entries, the 157 protein names excluded isoforms (as intended) but also excluded different genes of the same name belonging to a gene family, thus leading to an artificially low estimate. With this normalized count, I was able to compare the breakdowns by functional category with those in the zebra-finch derived proteins. Notably, 17 proteins identified from the search with the zebra finch proteome were not included among the 157 proteins identified with the red-winged blackbird proteome. They are included in the general comparisons below but were not used for downstream analysis (Chapter 5). Table 16 shows a breakdown of the identified proteins in zebra finch and red-winged blackbird by functional category. A full list of proteins is shown in Appendices A and B. Appendix A shows the overlap in profile between the zebra finch and red-winged blackbird, while Appendix B lists the 17 proteins detected only in search using the zebra finch proteome.

When the red-winged blackbird was used as the reference proteome, Trinotate identified 40 transcripts that did not yield identification results from any of the databases. It is unknown whether these transcripts correspond with unique genes, are all transcripts of the same gene, or are variants of an intermediate number of genes. Given the distribution of 157 unique genes across the 612 transcripts, it is likely that the

40 unknown transcripts constitute a mixture of uncharacterized genes and isoforms of these uncharacterized genes.

Table 16: Proteins detected from MS-based identification using the zebra finch (ZEFI) *in silico* proteome and the empirical red-winged blackbird (RWBL) seminal fluid sample. Functional categories were determined from searches in Ensembl and NCBI.

	ZEFI		RWBL	
Functional category	**Detected**	**Proportion**	**Detected**	**Proportion**
Amino acid interactions	4	4.76	8	5.10
Carbohydrate interactions	14	16.67	17	10.83
Chaperone	5	5.95	6	3.82
Cytoskeletal	2	2.38	6	3.82
Defense/stress response	7	8.33	13	7.64
General metabolism	21	25.00	29	18.47
Lipid interactions	6	7.14	21	13.38
Nucleic acid interactions	5	5.95	15	9.55
Protease	1	1.19	4	3.18
Protease inhibitor	3	3.57	5	3.18
Protein modification	9	10.71	18	11.46
Sperm protein	2	2.38	2	1.27
Transport	1	1.19	5	3.18
Other	4	4.76	8	5.10
Total	**84**	**100%**	**157**	**100%**

Roughly twice as many Sfps were identified from the search with the red-winged blackbird proteome as from the search with the zebra finch proteome. However, the increase in proteins was not distributed evenly across classes. About three times as many proteins involved in lipid and nucleic acid interactions were detected with the red-winged blackbird search, while the number of chaperone proteins and sperm proteins remained the same (and were also the same proteins). The plurality of proteins in both searches were related to general metabolism, especially the TCA cycle. In

addition, many of the proteins classified as interacting with carbohydrates were directly related to glycolysis and included proteins that broke down the various intermediates at each step of the cycle. In the red-winged blackbird search, however, the proportion of proteins involved in carbohydrate modification was superseded by the proportion of proteins involved in both lipid interactions and protein modifications. In contrast to the high number of proteins involved in cell respiration, very few proteins in both searches were identified as explicitly related to reproductive functions. Only three were identified with functions exclusively used in reproduction: acrosin (a protease), acrosin-binding protein (a sperm protein), and sperm-associated antigen 6 (a sperm protein).

Although direct comparisons of Sfp profiles in different taxa are limited by methodological variation across studies, certain similarities and differences can be observed between the Sfp profile of the blackbird and those of mammals and insects. Notably, the number of unique genes recovered (157) was intermediate compared to other studies: higher than the number found in rams (41, Souza *et al.* 2012), lower than those in bull (419, Byrne *et al.* 2012) and mice (506, Dean *et al.* 2009), and similar to those in *Drosophila* (~160, Findlay *et al.* 2009; Findlay *et al.* 2008; Wolfner 2009) and human (161, Clark & Swanson 2005). Second, the proportion of proteases and protease inhibitors was consistently lower in blackbirds than in other organisms. Twelve of the 174 combined proteins (7%) were proteases or protease inhibitors in zebra finch and red-winged blackbird, which was a lower proportion than in *Drosophila*, ram, bull, mouse and

human (Byrne *et al.* 2012; Dean *et al.* 2009; LaFlamme *et al.* 2014; Souza *et al.* 2012; Utleg *et al.* 2003). By contrast, the blackbird Sfps shared with the other species a high representation of genes related to metabolism, including for carbohydrate metabolism, lipid metabolism, and ATP synthesis/catabolism. Proteins related to defense and stress response were at intermediate levels, between the numbers observed in mouse and bull.

4.3.3. Differential expression

Certain proteins were more highly expressed in reproductive tissues, while others were more highly expressed in control tissues. Differential expression values ranged from nearly a 5000-fold change in carbonic anhydrase 6, indicating higher expression in reproductive tissues, to a 0.001-fold change in anti-thrombin 3, indicating higher expression in control tissues. The average posterior fold-change was 42.9 (higher in reproductive tissues), although the SD of ± 253 demonstrates the wide range of expression levels. Interestingly, for some genes with globally important functions (e.g., creatine kinase b-type, ATP-binding cassette sub-family D member 4), different isoforms of the same gene varied widely in expression levels, including spanning all three possibilities: high expression in reproductive tissue, coexpression at similar levels across all tissues, and high expression in control tissues.

4.4. Discussion

Applying next-generation sequencing of non-model organisms, especially those with strong foundations in fields such as behavioral ecology, is a promising avenue for

evolutionary biology (Ekblom & Galindo 2011). In this study, I demonstrate that transcriptomic and proteomic methods are a successful way to extract and characterize both mRNA and proteins from field-collected samples. Both differential expression and annotation information were used to build the first catalog of seminal fluid proteins (Sfps) in songbirds. Together with the transcriptomic and proteomic data, this list is a valuable resource for future comparative studies.

Unexpectedly, the Sfp list in blackbirds does not resemble a group of reproductive proteins so much as it does a group of proteins vaguely associated with energetic demands in a stressful environment. Certain proteins do feature GO terms referring to reproduction, and proteins are present that modulate pH and respond to stress or immune challenges. Many proteins are also shared across insect and mammalian profiles, especially those related to metabolism. Viewed together, however, the absence of proteins explicitly related to gamete recognition and fertilization presents an intriguing puzzle, especially in the context of an organism known to engage in multiple mating.

4.4.1. Comparison between proteome searches

About twice as many proteins were identified when peptide mass spectra were matched to a conspecific proteome than to a heterospecific proteome. One reason for this difference is that Sfps may have diverged too widely to be strong matches to a heterospecific database. If so, then the categories with disproportionate increases in

protein count could be logical places to look for rapidly evolving Sfps. However, the chance exists that some of the protein matches to the red-winged blackbird proteome are spurious, because the database consisted of transcripts that were translated across all six reading frames and therefore contain incorrect predicted peptide sequences. The presence of these incorrect peptides in the red-winged blackbird database could lead to a greater number of false positives than when using the non-redundant, curated proteome of the zebra finch. A possible solution is to inspect the read counts to ensure a minimum number of reads map to each contig, thus defining a minimum level of confidence required to call a true identification.

The list of proteins detected with the zebra finch proteome contained 17 proteins not found with the red-winged blackbird proteome (Appendix B). Reasons for this discordance are unclear but could include the fact that I analyzed proteins derived only from testis and not from seminal glomera, leading to missed proteins that were specific to this second reproductive tissue. Given the dearth of testis-derived proteins observed to be expressed exclusively in testis, however, a scenario in which the seminal glomera harbor almost 20 tissue-specific proteins seems unlikely. Alternately, the discrepancy could be due to the use of different databases for protein annotation. The first search identified mass spectra by matching them to the zebra finch proteome from NCBI, while the second search did so by BLASTing blackbird transcripts to UniProt databases. Annotation pipelines are likely to be different between the two databases, potentially

causing proteins identified from one database to be overlooked in the other. Supporting this hypothesis is the observation that alternate names for the same protein were given in the two searches (e.g., protein NipSnap homolog 2 vs. glioblastoma-amplified sequence). Nevertheless, the red-winged blackbird proteome still appears to have yielded more successful matches than the zebra finch proteome in identifying mass spectra of blackbird Sfps.

4.4.2. Comparison with Sfp profiles in other species

Perhaps the most noticeable difference when comparing the blackbird Sfps to those of other organisms is the near absence of blackbird Sfps related to reproduction. No proteins analogous to mammalian spermadhesins, seminal vesicle proteins, and semenogelin were detected. Importantly, these proteins are all produced in mammalian accessory glands, which (as in *Drosophila*) are the sites of synthesis for most Sfps. Because birds lack accessory glands (Aire 2007), these proteins might be simply bypassed during ejaculate formation. Anecdotally, avian ejaculate is composed largely of sperm, with little seminal fluid relative to mammalian ejaculate (T. Birkhead, pers. comm.). The low amount of seminal fluid, combined with the absence of accessory glands, could help explain why no orthologs to such proteins were detected.

More difficult to explain is the low proportion of proteases and protease inhibitors. Proteases and protease inhibitors protect sperm while digesting membranes and proteins (including other Sfps), thereby playing roles in sperm competition, and are

often under positive selection (Wolfner 2002, 2009). Despite the genetic polyandry of red-winged blackbirds, these proteins were found in lower proportions in the blackbird than in any other organism examined. It is possible that these proteins could have been missed or dropped during transcriptome assembly, MS-based identification, and/or protein annotation. Many proteases are small (Chapman 2001) and may have been difficult to assemble from the Illumina fragments. Even in model organisms, accessory-gland proteins have been overlooked because of gene-specific features such as GC content and exon structure interfering with traditional search parameters (Findlay *et al.* 2009). Future annotation of the entire testis and seminal glomera transcriptomes, not just the portions matching to the mass spectra, would identify all transcripts assembled from the RNA-Seq data and produce a more comprehensive list of proteins to examine.

Alternately, the proteins were truly absent, which would suggest they are neither important in fertilization nor for competitive interactions leading to fertilization. This is a fascinating possibility that would suggest fundamentally different mechanisms at play in avian reproduction. Evidence for permissive sperm-egg interactions in birds is already known, such as the occurrence of polyspermy, decreased species specificity of sperm-perivitelline interactions, and high hybrid viability (Edwards *et al.* 2005b; Stepinska & Bakst 2007; Stewart *et al.* 2004; Chapter 5). If these events all correspond with relaxed molecular processes, then the proteins moderating sperm-egg interaction in birds may not be under as intense selection for gamete recognition as they are in insects

or mammals. However, reproduction as a whole is still a highly conserved process, and protein diversification (or attrition) is likely to operate within limited parameters. In addition, this scenario leaves open the major question of how sperm competitive interactions might unfold without the proteins most often associated with male-male competition.

Certain proteins do have reproduction-related GO terms, such as beta-hexosaminidase ("sexual reproduction"), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase ("egg activation"), and Protein DJ-1 ("plays a role in fertilization"). However, these terms were one of several in the results returned from BioMart. Without specifically knowing the exact targets of the proteins, it should not be assumed that the proteins are playing reproductive roles simply because they were recovered from seminal fluid. In addition, these genes are likely to be under purifying and not positive selection if their pleiotropic interactions stemming from expression elsewhere in the body are a source of selective constraint.

Of interest are the 40 uncharacterized proteins with no hits to any databases. These proteins reflect a limitation of the approach I used: that protein annotation relies on proteins that are already identified and archived. With improving annotation, perhaps their functional significance and their evolutionary patterns will come to light. The current results, however, demonstrate that for the majority of proteins, joint proteomic and transcriptomic methods are a feasible way to identify proteins collected

from a non-model organism. The catalog of blackbird Sfps produced from these pipelines is a valuable new resource and permits comparison of evolutionary patterns in a group of proteins known to experience unique and intense selective regimes.

5. No evidence for positive selection in *Agelaius* seminal fluid proteins

5.1. Introduction

Males and females invest unequally in the process of reproduction, leading to distinct optimal mating strategies in each sex (Bateman 1948; Darwin 1871; Parker 1979; Trivers 1972). In males, less costly gametic investment favors a strategy of fertilizing as many eggs as possible, promoting the occurrence of multiple mating in males (Bateman 1948). In species where multiple mating also occurs in females, selection on males to copulate with more than one female may lead to situations where ejaculates from multiple males are present in a single female (Parker 1979). The process by which sperm from different males compete for fertilization of a female's eggs thus constitutes an element of sexual selection exerting pressure on males to outcompete other males even after the act of mating (Birkhead & Pizzari 2002; Parker 1970).

In addition to sperm, elements in the seminal fluid transferred to the female may also be targets of selection (Adams & Wolfner 2007). Seminal fluid proteins (Sfps), which aid, among other things, in stimulating oogenesis, mediating sperm storage and inducing changes in the female to prevent or delay remating (Wolfner 2002, 2009), are known to evolve rapidly (Swanson & Vacquier 2002), particularly between closely related species (*Drosophila*, Swanson *et al.* 2001; *Gryllus* field crickets, Andres *et al.* 2008; *Heliconius*, Walters & Harrison 2010; rodents, Ramm *et al.* 2008; Turner *et al.* 2008; primates, Clark & Swanson 2005; Dorus *et al.* 2004). Accelerated divergence in these

protein-coding sequences is evidence for positive diversifying selection resulting from selection on male fertilization success (Chapman 2001). As variation in Sfps has been implicated in competitive ability (Clark *et al.* 1995; Clark *et al.* 1999; Clark & Swanson 2005; Fiumera *et al.* 2005), sperm competition may be partly responsible for this rapid evolution. (While it is true that rapid evolution may be attributed to either sperm competition, sexual conflict, or both forces (Arnqvist & Rowe 2005), partitioning the effects of sperm competition versus sexual selection is beyond the scope of this study. Therefore, this study focuses on testing the influence of sperm competition, which can be measured by proxy using the frequency of multiple mating.)

Sensitivity of Sfp evolution to the degree of sperm competition suggests that factors governing the intensity of sperm competition, such as an animal's mating system, could play an integral role in shaping patterns of protein evolution. Specifically, increasing levels of polyandry, or multiple mating by females, are expected to correspond with stronger selection on male sperm competitive ability (Dorus *et al.* 2004; Schumacher *et al.* 2014; Walters & Harrison 2010). Identifying clades with expected variation in sperm competition thus presents the opportunity to examine the causes of such variation and to characterize the selective regime operating on Sfps.

Songbirds are an excellent system with which to examine changes in molecular evolutionary response to sperm competition. The physiology of sperm competition in birds, including sperm characteristics in males and the mechanics of sperm storage,

fertilization and egg-laying in females, are well established (Birkhead & Møller 1992; Briskie & Montgomerie 2007), but in spite of this foundation, there have been few investigations of adaptations at the molecular level (see below). Researchers have been limited by the lack of reference sequences, but new methods such as proteomic analysis via mass spectrometry and RNA sequencing now overcome this difficulty (Ekblom & Galindo 2011).

Polyandry frequently occurs in the form of extra-pair paternity (EPP), a behavior in which males and females of many species form pair bonds during a breeding season but mate and produce offspring with individuals outside those pair bonds (Griffith *et al.* 2002; Chapter 2). Since the frequency of EPP, or any other form of polyandry, is a measure of multiple mating and thus a proxy for sperm competition, variation between species in EPP rates provides a natural opportunity to test the hypothesis that different genetic mating systems correspond with different evolutionary rates of Sfps.

Reproductive protein evolution has been well studied in mammals, primarily rodents and primates (Clark & Swanson 2005; Dorus *et al.* 2004; Ramm *et al.* 2008; Turner *et al.* 2008), but few trends have been identified in birds. Recent studies addressing the role of sexual selection in avian protein evolution have examined sequence evolution of zona pellucida (ZP, or egg coat) proteins, a female gamete-recognition protein, using comparative sequence analyses from galliform (chicken-like) species. Of seven ZP proteins analyzed from mammals (a group in which polyspermic fertilization is

detrimental) and birds (a group in which polyspermy routinely occurs), five avian proteins and two additional proteins (CD9 and acrosin) exhibited signs of positive selection, indicating that sperm competition or sexual conflict may sustain rapid protein evolution even in the absence of other selective pressures like polyspermy avoidance (Berlin *et al.* 2008; Calkins *et al.* 2007).

These studies and others testing for selection on avian protein-encoding genes (Ceplitis & Ellegren 2004) all compare rates of avian protein evolution against mammalian proteins to document broad-scale patterns of sequence evolution. (Outside of a reproductive context, studies of rapid immune protein evolution in birds have compared sequences within birds, although they are still limited to tests between distantly related model species like zebra finch (*Taeniopygia guttata*) and chicken (*Gallus gallus*) (Ekblom *et al.* 2010).) By contrast, no known studies have undertaken a finer-scale analysis strictly within closely related bird species. This point is of particular importance for studies of Sfps, since this rapidly evolving class of proteins is among the least likely to be well conserved between mammals and birds. Furthermore, the above studies stopped short of framing their results in the context of the study species' mating systems, making it difficult to draw conclusions about the relationship between naturally occurring differences in sperm competition and protein evolution. The lack of work on male seminal proteins, a group examined in many other species, additionally prevents a complete understanding of evolutionary patterns across different taxa.

Therefore, this study compares sperm competition and protein evolution in different *Agelaius* species (Chapter 1) to determine whether differences in the strength of selection on Sfps lead to distinct patterns of sequence evolution. The goals of the study were: (1) to measure sperm competition intensity in three species of *Agelaius* blackbirds; (2) to select candidate proteins from a catalog of Sfps (Chapter 4); and (3) to test for evidence of positive selection in the genes encoding these proteins.

5.2. Methods

5.2.1. Measurement of EPP rate

For red-winged blackbirds, EPP rates for eight populations were obtained from other researchers and from fieldwork I conducted myself from 2009-2011 (see Chapter 2 for full methods). For yellow-shouldered blackbirds, EPP rates were obtained from a population of breeding yellow-shouldered blackbirds in southwest Puerto Rico in 2012 (see Chapter 3 for full methods).

For tricolored blackbirds, I collected samples from a population breeding in a manmade cattail pond in Conaway Ranch (Yolo County), between Davis and Sacramento, California. Samples were collected from 5 to 17 June 2014. Tricolored blackbirds are itinerant breeders (Hamilton 1998), raising different broods in different locations during a single breeding season. This population had likely completed its first breeding attempt in Merced County in southern California before heading north to the Sacramento Valley (R. Meese, pers. comm.).

Tricolored blackbirds are highly colonial, with active nests built as closely as 8 cm from each other (Beedy & Hamilton 1999; pers. obs.). This particular colony had about 5,000 birds nesting in the pond, which is relatively small and possibly reflective of significant range-wide declines in the species (Cook & Toft 2005). Because entire areas flush in the presence of a human observer, territories are extremely difficult to define. Males defend only the immediate nesting area; early estimates of territory sizes range from 1.8m^2 to 3.25m^2 (Lack & Emlen 1939; Orians 1961), tending toward the lower measurement in denser areas.

The colony was extremely synchronous in its breeding cycle. During the twelve-day nestling period, I captured as many adults as possible and sampled the nestlings in their affiliated nests. I was unable to use mist nets due to the dense cattail growth and furthermore would have been unable to determine the territories of randomly captured adults. The sole successful method of capture was to place the Potter trap on the nest and wait for a provisioning parent to enter. Although both parents fed nestlings, this method was successful only with females. Most males fed less often than females (Barker *et al.* 2008; Payne 1969) and additionally avoided feeding as long as the trap was present. Consequently, I captured males at only six of the nests for which I sampled females and chicks. Two additional males were caught using baited traps on the banks of the pond. I sampled 152 chicks from 46 nests but discarded from my analysis 13

chicks from four nests where I did not capture the parents. For my analysis, I included a total of 50 adults (42 females and 8 males) and 139 chicks at 42 nests.

To calculate EPP in the red-winged and yellow-shouldered blackbirds, I compared genotypes of the chicks and social males to identify allelic incongruities (see Chapters 2 and 3). EPP was defined in two ways: as the number of extra-pair young (EPY) out of the total number of chicks, and as the number of nests containing least one EPY out of the total number of nests.

I was unable to use the same method for the tricolored blackbird because the genotypes for most of the social males were missing. Instead, I used the maximum-likelihood program COLONY v.2.0.4.7 (Jones & Wang 2010) to infer paternal genotypes from known parentage (i.e., maternity), known sibships, and allele frequencies. I ran one iteration with a 65% probability that the father was included among the male candidates (accounting for the known paternities) and a 95% probability that the mother was included among the female candidates. COLONY generated up to five putative genotypes at each locus for each male and gave posterior probabilities for each genotype. I set the mating system to monogamy to constrain the program to infer genotypes for one male per nest. This step inherently assumes that the inferred social male will be the genetic father of most of his social offspring. Because this assumption is violated with nests containing >50% EPY, the number of chicks identified as EPY using this method is expected to be lower than the true value. For example, in a nest that

contains 100% EPY, the program will accurately infer genotypes for the genetic father but have no way of identifying him as an extra-pair male. (I set the mating system to polygamy during initial runs, using mock yellow-shouldered blackbird data with all male genotypes removed, to test whether running simulations under a polygyny setting would yield more accurate results. Without the constraint of monogamy, the program inferred the presence of very high numbers of male genotypes, essentially tailored to the genotypes of each chick and female, that severely overestimated the number of extra-pair males and offspring.)

Because COLONY considers loci independently, it has limited ability to infer multilocus paternal genotypes. To reconstruct complete genotypes with which to estimate the rate of EPP, I used a custom Python bootstrapping script (M. Johnson, pers. comm.) to create 1000 pseudoreplicate multilocus paternal genotypes. The genotype chosen at each individual locus was determined by its probability of occurrence as calculated by COLONY. Genotypes of chicks were compared to each reconstructed genotype, and chicks were marked as EPY if they had any mismatches with the inferred paternal genotype. The program ignored maternal genotype and missing data.

To validate these results, I ran equivalent simulations in COLONY using the data from the yellow-shouldered blackbird and the Pennsylvania population of red-winged blackbirds. I removed all male genotypes, then applied the same bootstrapping script.

Distributions of the proportion of EPY over the 1000 runs for each species were visualized with a histogram in R.

5.2.2. Selection and sequencing of candidate genes

In Chapter 4, I described how I assembled the transcriptomes of four red-winged blackbird tissues and matched the transcripts of testis-expressed proteins with proteins detected in seminal fluid. I obtained expression levels for all 612 genes, then compared expression levels across all four tissues and across the two reproductive and two control tissues. I also created an annotation report for all genes, then combined the expression and annotation data by ranking annotated proteins by fold change between reproductive and control tissues. The end result was a list of testis-derived proteins present in seminal fluid, listed by fold-change expression. Here, I selected candidate Sfps and control genes for sequencing and tests of selection.

I used the annotation and expression information in multiple ways to identify candidate proteins. Criteria for possible candidates were:

- Higher expression in testis and/or seminal glomera than both heart and liver, for all isoforms detected. I did not use a gene if it had isoforms that had both high and low fold-changes in expression, implying that certain isoforms were co-expressed across the four tissues.

- Evidence of a signal peptide or transmembrane domain, as Sfps are expressed in the extracellular matrix
- Known importance in reproduction, as listed in the literature or on non-redundant databases (e.g., UniProt, GenBank, GeneCards)
- Previous discovery in mammalian Sfps, regardless of expression level (e.g., Byrne *et al.* 2012; Dean *et al.* 2009; Souza *et al.* 2012; Utleg *et al.* 2003)
- Ease of sequencing a unique gene. Genes with similar sequences, often belonging to a gene family (e.g., acrosin, heat shock proteins), were not used.
- Ease of sequencing exons (i.e., consecutive exons separated by short introns, or multiple individual exons longer than 100 bp)

These criteria, combined with the dearth of proteins with explicit reproductive functions (Chapter 4), were sufficient to filter out the majority of genes. I selected six candidate Sfps, two with explicit reproductive functions (acrosin-binding protein and sperm-associated antigen 6), two with probable roles in pH regulation within the reproductive tract (carbonic anhydrase 6 and regenerating islet-derived protein 4), and two with probable roles in sperm cell metabolism (hexokinase 3 and galactokinase 1). Descriptions of each protein are given below. Summary information is shown in Tables 17-18.

1) ***Acrosin-binding protein*** (ACRBP) was almost exclusively expressed in the testis transcriptome (Table 17). It is found in the sperm acrosome and facilitates the production of acrosin, the protease responsible for digesting the inner perivitelline layer of the egg (analogous to the mammalian zona pellucida) and allowing for sperm penetration (Digby & Howarth 1972). In mammals, ACRBP is needed to "activate" sperm to make them capable of fertilization (i.e., capacitation, Dubé *et al.* 2005). However, in birds, inseminated sperm are immediately capable of fertilization without this reaction (Howarth 1970). Nevertheless, avian sperm cells do switch between suppressed versus active states when entering and leaving sperm storage tubules prior to fertilization, which may be parallel to capacitation, and ACRBP could be important at this and other steps in sperm-egg interaction.

2) ***Sperm-associated antigen 6*** (SPAG6) was also highly expressed in testis relative to other tissues. It belongs to a group of proteins with structural roles in sperm. Mouse and human SPAG6 is found in the sperm tail and is thought to be important for the stability of the flagellar central apparatus and for flagellar motility (Neilson *et al.* 1999; Sapiro *et al.* 2000). In a knockout study, mice born without SPAG6 suffered increased mortality due to hydrocephaly within eight weeks of birth, and surviving male mice were infertile, with morphological and motility defects in sperm (Sapiro *et al.* 2002).

Interestingly, the human and mouse orthologs of SPAG6 were inferred from a gene involved in flagellar motility in *Chlamydomonas* algae (Neilson *et al.* 1999; Sapiro *et al.* 2000). The extreme conservatism and structural role of this gene suggests the gene may be under strong purifying selection. However, because sperm length and velocity are correlated with sperm competition in birds (Birkhead *et al.* 1999; Briskie *et al.* 1997), changes in this gene allowing for longer or faster sperm could be favored by sexual selection. Therefore, I was interested to assess the degree of sequence divergence and, by extension, the relative contributions of purifying versus positive selection.

3) ***Carbonic anhydrase 6*** (CA6) was highly expressed in both reproductive tissues, although the putative isoforms are present at different abundances in the testis vs. seminal glomera (see "TranscriptID" column in Table 17). CA6 is the sole secreted isoform in the family of carbonic anhydrases, which all catalyze the hydration of carbon dioxide to carbonic acid, and vice versa ($CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$) (Chegwidden & Carter 2000). This conversion regulates pH balance across different tissues. In mammals, CA6 is expressed in saliva and milk (Karhumaa *et al.* 2001; Parkkila *et al.* 1990), but surprisingly it was found to be absent in the human male reproductive tract (Kaunisto *et al.* 1990). Instead, carbonic anhydrases 2 and 4 have been found in the male reproductive system. Added to seminal fluid during spermatogenesis, these enzymes are thought to secrete bicarbonate ions that maintain sperm motility and protect sperm against the acidic environment of the female reproductive tract (Parkkila 2000).

One study exists on CA activity in the avian reproductive tract, although it appears to have been published before the standardized nomenclature of the different isoforms. In chickens, a significant correlation exists between CA activity in testis and the volume of sperm and seminal fluid (Harris & Goto 1984). Otherwise, the function of CA6 in the male reproductive tract is unknown. Using the collected information on CA function, and because CA6 was the sole carbonic anhydrase represented in the blackbird Sfps, I chose to use this enzyme as a candidate protein.

4) ***Regenerating islet-derived protein 4*** (REG4) was very highly expressed in two putative isoforms in seminal glomera. There is a chance that this protein was accidentally picked up because of the cloaca's physical proximity to the colon and its role as the final site in the digestive tract. REG4 is known to be highly expressed in the colon because of its importance in pH balance in the gastrointestinal environment (Hartupee *et al.* 2001). Nevertheless, because of its stated function of carbohydrate interaction in an acidic environment, as well as its predicted role in inflammatory responses (Hartupee *et al.* 2001), I considered it a possible Sfp.

5) and 6) While the two metabolic proteins may be associated with cell respiration across the body, they may both play particular roles in reproduction. ***Galactokinase 1*** modulates the metabolism of galactose, a carbohydrate that affects induction of the acrosome reaction prior to fertilization (Horrocks *et al.* 2000) and for which there are receptors on the sperm surface (Rivkin *et al.* 2000). Likewise, variants of

hexokinase 3 (Hex-1 and Hex-t) show testis- and accessory-gland specific expression in *Drosophila* (Cavener 1980; Murray & Ball 1967). HK3 in the blackbird shows higher but not exclusive expression in testis (Table 17).

7) and 8) Separate from the six candidate Sfps, I selected two control genes from the list of genes coexpressed equally across all four tissues. Both genes, ***phosphoglycerate kinase 1*** (PGK1) and ***alpha-enolase*** (ENO1), play critical roles in glycolysis and are expected to be under purifying selection. ENO1 is an isozyme of another gene (beta-enolase) used as a control in a previous study of mouse Sfps (Ramm *et al.* 2009).

Table 17: List of candidate and control genes. Transcript IDs are from the testis transcriptome. For each tissue, the approximate number of Illumina reads mapping to each transcriptome assembly is shown (the "expected count" metric in RSEM). SemGlom = seminal gomera. PostFC = posterior probability of fold change in reproductive vs. control tissues as predicted by EBSeq.

Gene	Abb.	Transcript ID	Testis	SemGlom	PostFC
Acrosin-binding protein	ACRBP	29636_c1_seq1	16154	252.33	1813.576
		56912_c0_seq1	7581.3	324.32	467.8122
Carbonic anhydrase 6	CA6	80084_c0_seq3	5919.6	17757.4	4844.49
		85129_c0_seq4	826.41	2768.71	206.9163
		85129_c0_seq2	525.03	0	57.9728
		85129_c0_seq3	1122.9	0	8.806228
Galactokinase 1	GALK1	24034_c0_seq1	58183	1903.67	1031.991
		43506_c0_seq1	197.41	2.48	12.55986
		55688_c0_seq1	3457.6	1902.84	3.369306
Hexokinase 3	HK3	80088_c0_seq1	178456	3716.05	100.7375
		80068_c0_seq1	510824	9491.71	91.46342
Regenerating islet-derived protein 4	REG4	71242_c0_seq1	50.46	204864	439.2539
		71242_c0_seq2	401.56	1025366	199.4784
Sperm-associated	SPAG6	37293_c0_seq1	2079.4	13.98	52.79939

antigen 6					
		77252_c0_seq1	338.98	0	18.5168
		83412_c0_seq1	556.58	10.02	4.264852
Alpha-enolase 1	ENO1	29236_c0_seq1	54283	38637.1	1.117991
		55456_c0_seq1	54283	38637.1	1.117991
		55445_c0_seq1	86.48	0	1.06328
Phosphoglycerate kinase 1	PGK1	72171_c0_seq1	30520	19656	0.701302
Gene	**Abb.**	**Transcript ID**	**Heart**	**Liver**	
Acrosin-binding protein	ACRBP	29636_c1_seq1	0	0	
		56912_c0_seq1	5.27	0	
Carbonic anhydrase 6	CA6	80084_c0_seq3	0	0	
		85129_c0_seq4	14.84	0	
		85129_c0_seq2	0	0	
		85129_c0_seq3	23.74	43.59	
Galactokinase 1	GALK1	24034_c0_seq1	31.66	0	
		43506_c0_seq1	2.49	1.96	
		55688_c0_seq1	825.84	493.04	
Hexokinase 3	HK3	80088_c0_seq1	917.09	193.77	
		80068_c0_seq1	2768.8	713.46	
Regenerating islet-	REG4	71242_c0_seq1	528.74	82.27	
derived protein 4		71242_c0_seq2	5800.2	980.46	
Sperm-associated antigen 6	SPAG6	37293_c0_seq1	10.91	7.15	
		77252_c0_seq1	5.93	0	
		83412_c0_seq1	60.16	16.85	
Alpha-enolase 1	ENO1	29236_c0_seq1	30030	40569	
		55456_c0_seq1	30030	40569	
		55445_c0_seq1	0	43.5	
Phosphoglycerate kinase 1	PGK1	72171_c0_seq1	33452	27213.2	

Table 18: Descriptions of functions for the six candidate Sfps and two control genes (ENO1 and PGK1).

Gene	Abb.	Function
Acrosin-binding protein	ACRBP	May be involved in packaging and condensation of the acrosin zymogen in the acrosomal matrix via its association with proacrosin
Carbonic anhydrase 6	CA6	Reversible hydration of carbon dioxide. Its role in saliva is unknown.
Galactokinase 1	GALK1	Major enzyme for galactose metabolism
Hexokinase 3	HK3	Major enzyme for hexose metabolism
Regenerating islet-derived protein 4	REG4	Calcium-independent lectin displaying mannose-binding specificity and able to maintain carbohydrate recognition activity in an acidic environment
Sperm-associated antigen 6	SPAG6	Important for structural integrity of the central apparatus in the sperm tail and for flagellar motility
Alpha-enolase 1	ENO1	Glycolytic enzyme expressed in most tissues
Phosphoglycerate kinase 1	PGK1	Present in all living organisms as one of the two ATP-generating enzymes in glycolysis

5.2.3. Primer design

Once I had selected the genes of interest, I retrieved the corresponding transcripts from either of the two testis transcriptomes (the original assembly, plus the second assembly composed of leftover reads identified in Bowtie). If a gene had multiple isoforms listed, I used the transcript with the highest fold change. For each transcript, I ran a BLAST search, restricting the search results to sequences from birds. I found the closest mRNA alignment across multiple species such as zebra finch (*Taeniopygia guttata*), medium ground finch (*Geospiza fortis*), white-throated sparrow (*Zonotrichia albicollis*), and ground tit (*Pseudopodoces humilis*).

Each mRNA hit from the BLAST search linked to a physical map of the gene and its DNA sequence. Most exons were short (<250 bp) and separated by long introns. In addition, all genes except REG4 were well over 1 kb (average: 1699 ± 659 bp) and unable to be sequenced with one primer pair. Therefore, for all genes except REG4, I sequenced multiple regions from different regions of the gene ranging from 200-800 bp, deleted any intronic sequences, and concatenated the exons. I attempted to sequence a total of about 600bp (200 amino acids) of exons per gene. For REG4, I sequenced the entire gene, which was relatively small (867 bp total, with <500 bp of coding region).

To design primers, I used the coordinate information on Genbank to find the targeted region from the genomic DNA sequence in zebra finch, white-throated sparrow or medium ground finch. I then found the orthologous sequence in my red-winged blackbird transcript and created species-specific primers from these orthologs. Primers were designed with Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and specified to be 19-24 bp long, with a melting temperature between 57-63°C. After Primer3 had identified candidate forward and reverse primers, I ran them in a BLAST search, again restricting results to sequences from birds, to confirm their specificity to one place in the genome. In total, I designed 19 primer pairs for the eight genes.

5.2.4. Sanger sequencing and exon concatenation

For each locus, I obtained sequences from three males per species. For the red-winged blackbird, I used the Trinity-generated transcript from the Pennsylvania male as

one sequence and sequenced two additional males, one from Michigan and one from Ontario. For the tricolored and yellow-shouldered blackbirds, I sequenced three sampled males.

I optimized the PCR protocol for all 19 primer pairs. Generally, a suitable annealing temperature was 5°C below the lower melting temperature of the two primers. For the initial PCR, reactions consisted of 2.0 μl of DNA, 8.9 μl of distilled water, 2.0 μl of 10X buffer, 3.2 μl of dNTPs, 1.0 μl each of 10 μM forward and reverse primer, 1.5 μl of bovine serum albumin (BSA), and 0.4 μl of Taq (Denville Scientific). PCR cycles were initiated at 95°C for 5 minutes, followed by 35 cycles at the optimal temperature. Each cycle consisted of denaturation at 95°C for 0:30, annealing for 0:30, and extension at 72°C for 1:00. The final extension was at 72°C for 7 minutes.

Gels were run after each reaction to verify successful amplification. The DNA template was then purified with ExoSAP. To each template I added 2.6 μl of distilled water, 0.2 μl of exonuclease I (ExoI), and 0.2 μl of shrimp alkaline phosphatase (SAP, New England Biolabs). The reaction was initiated at 37°C for 30:00, followed by 80°C at 15:00 to deactivate ExoI. Plates were processed by Eton Bioscience.

Multiple bands were produced in two loci (REG4 and one of the loci in ENO1). For REG4, the primers amplified excessive numbers of bands, even at high annealing temperatures. As the forward primer had an unusually high GC content and melting temperature, I redesigned the primer and successfully sequenced the target region. For

the locus in ENO1, I used a MinElute Gel Extraction kit (Qiagen) to excise and extract DNA from the target band before submitting for sequencing.

Sequences were edited in Sequencher (Gene Codes). I used ambiguity codes to call heterozygous SNPs. Once I had all nine sequences for a gene, I used PhyDE v.0.9971 (Müller *et al.* 2010) to edit and align sequences. I removed the sequence overlapping the primer, because any polymorphisms at those site would have been masked by the primer used to amplify the gene. I trimmed all sequences to the same length, deleting sites from the beginning and end if any of the eight sequenced individuals was missing data. I then found the correct frame using the longest ORF results predicted by Transdecoder, a program from the Trinotate pipeline (Chapter 4).

I chose the medium ground finch (*Geospiza fortis*) as my outgroup. Even though its genome is not as well annotated as the zebra finch, it is a New World species and thus more closely related to *Agelaius*. I found the orthologous sequences in *G. fortis* for each gene and included them in the alignment. The exception was REG4, which BLASTed to a different gene in *G. fortis*. Instead, I used the ortholog of the ground tit (*Pseudopodoces humilis*) as my outgroup for this gene.

For all genes except REG4, which was contained within one sequence, I concatenated the separately sequenced exons and ensured they were all in frame. The final product was one sequence in FASTA format per gene.

5.2.5. Summary statistics

I used DnaSP v.5.10.1 (Librado & Rozas 2009) to measure nucleotide diversity (π), number of haplotypes, haplotype diversity, and number of segregating sites and mutations. Haplotypes were reconstructed from 1000 iterations, with 100 iterations burn-in. For reference, I compared them to the statistics reported for two mitochondrial genes (control region and ND2), sequenced in red-winged and tricolored blackbirds in a separate study (Barker *et al.* 2012).

5.2.6. Tests of selection

For many genes, sequences within (and occasionally across) species were identical. Where sequences within species were identical, I removed duplicate sequences so that only the unique sequence remained. Where sequences across species were identical, I kept one sequence in each species as a representative sequences (but deleted any others within species that were also identical). Although this duplication was redundant for tree-building and led to branch lengths of zero, it ensured all three species were included in the tree.

Sequences were converted from FASTA to NEXUS format in PAUP v.4.0a134 (Swofford 2003), with additional instructions to constrain trees to the known phylogeny of the three species: ((red-winged, tricolored), (yellow-shouldered)) (Barker *et al.* 2008, Fig. 1). In addition, polytomies were not allowed. I first conducted a heuristic search under parsimony settings to find the best constraint-compatible tree. Once a tree was

found, I used the "automodel" command to find the best nucleotide substitution model (out of the standard 56 models) explaining this tree. Models were evaluated and ranked by small-sample-size-corrected Akaike Information Criterion (AIC_c), with the smallest AIC_c value corresponding to the best-fit model. Finally, using the best-fit model, a single maximum-likelihood (ML), constraint-compatible species tree was constructed for use in PAML.

Sequences were converted to PHYLIP, and the ML tree and the PHYLIP file were used as input files for the codeml program in PAML v.4.6 (Yang 2007). This program uses maximum likelihood methods to tests the fits of different codon substitution models (Yang & Bielawski 2000). Each model specifies different parameter for the distribution of dN/dS (ω)-values. Three model comparisons are commonly used to test for positive selection (Swanson *et al.* 2003; Wong *et al.* 2004). In these comparisons, one model is restricted to allowing only purifying or neutral selection ($0 < \omega \leq 1$), while the other allows for purifying selection ($\omega > 1$). Under both models, codeml assigns sites to each site class and calculates the proportions of sites in each class. A likelihood ratio test (LRT) is used to test the hypothesis that the model allowing for positive selection is a significantly better fit than the model allowing for only purifying and neutral selection. The null model is presented first in the descriptions below.

1) M1a (nearly neutral) vs. M2a (positive selection) (Nielsen & Yang 1998): M1a constrains ω to be one of two site classes corresponding to purifying selection and

neutral evolution. The lower-bound class can range between $0 < \omega < 1$, but the upper-bound class is fixed at $\omega = 1$ (neutral evolution). M2a allows for three site classes. The first two classes are the same as in M1a, but the third class now permits an estimate at $\omega \geq 1$.

2) M7 (beta) vs. M8 (beta + ω): M7 estimates the parameters of a beta-distribution (a probability distribution curve with ω on the x-axis, limited to $0 < \omega < 1$). It then calculates the proportion of sites belonging to each of eight discrete site classes that occupy equal areas under the curve (and thus occur with equal probability). M8 estimates the same parameters of the beta distribution, except it divides the curve into eleven site classes, including a site class allowing for $\omega \geq 1$.

3) M8a (beta + fixed ω) vs. M8 (beta + ω): M8a is the same as M7, except the model adds a ninth site class where $\omega = 1$, thus allowing $0 < \omega \leq 1$. M8 is the same as above, with an eleventh site class where $\omega \geq 1$. This model controls for false positives resulting from a poor fit of the data to the beta-distribution (Turner *et al.* 2008).

The program codeml calculated the log-likelihood score (lnL) for each model. Scores were then used to calculate the test statistic, $-2(\ln L_1 - \ln L_2)$, and the product was compared against critical values for the chi-square distribution at two degrees of freedom. At $\alpha = 0.05$, the critical value is 5.99. For the M8-M8a comparison, a mixed chi-square distribution with df = 1 was used to account for the two chi-square distributions produced because of a forced boundary in M8a (Self & Liang 1987). Under this

distribution, the critical value at $\alpha = 0.05$ is 2.71 instead of 3.84 under a normal distribution. A significant result indicated that the model allowing for positive selection was a better fit to the observed codon substitution patterns than the null model.

5.3. Results

5.3.1. Comparison of EPP

Results for EPP in red-winged and yellow-shouldered blackbirds are reported in Chapters 2 and 3. Across eight populations, red-winged blackbirds had an average proportion of 0.30 (±0.054 SD) of chicks identified as EPY and 0.50 (±0.076) nests containing at least one EPY. For the Pennsylvania population specifically, 23/87 (0.26) of chicks were EPY, while 13/27 (0.48) of nests contained at least one EPY. Yellow-shouldered blackbirds had a proportion of 20/87 (0.23) of chicks identified as EPY and 11/30 (0.37) nests containing at least one EPY.

The distribution of the bootstrapping runs for proportion of EPY is shown in Fig. 15 and Table 19. Both of the empirical proportions of EPY for the red-winged and yellow-shouldered blackbirds were approximately twice the inferred mean proportions (Fig. 2). Specifically, the empirical proportion of Pennsylvania red-winged blackbird EPY was 1.7 times the inferred mean, while the empirical number of yellow-shouldered blackbird EPY was 2.5 times the inferred mean. If the pattern is consistent, then the true proportion of EPY for the tricolored blackbirds is approximately 1.7 to 2.5 times the

inferred mean proportion of 0.17, or 0.29-0.42. Out of 139 sampled tricolored blackbird chicks, this proportion corresponds to 40-59 of chicks that are EPY.

The distribution of the runs for proportion of nests containing at least one EPY is shown in Fig. 16 and Table 20. The empirical proportion of Pennsylvania red-winged blackbird nests with EPY was about 1.5 times the estimate, while the empirical proportion in yellow-shouldered blackbirds was about 1.8 times the estimate. If the true proportion of tricolored blackbird nests containing EPY is between 1.5 to 1.8 times the inferred mean proportion of 0.35, then the proportion should range from 0.53 and 0.65. Out of 42 sampled nests, this proportion corresponds to 22-27 nests containing EPY.

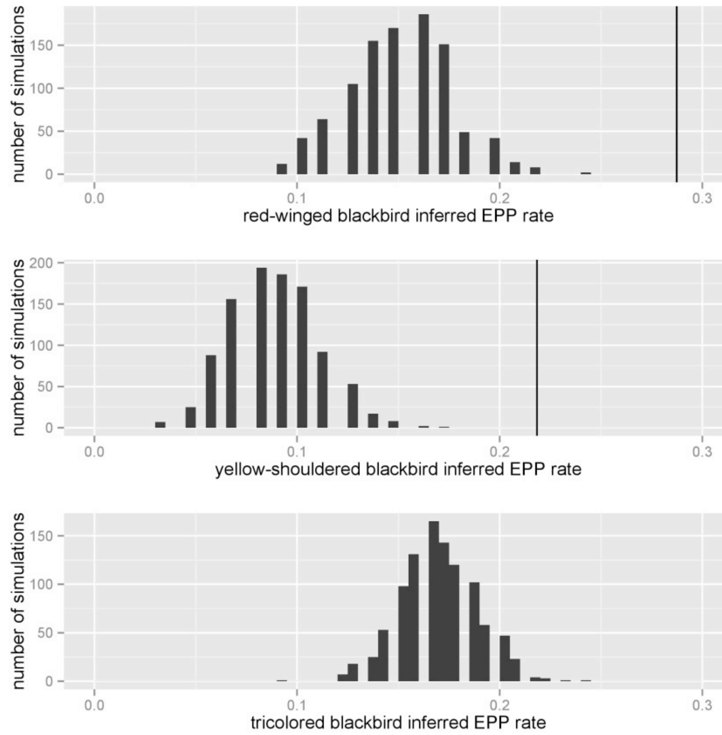

Figure 15: Distribution of inferred EPP rate across 1000 runs (reconstructed multilocus paternal genotypes). Solid line shows the empirically measured EPP rate for red-winged and yellow-shouldered blackbirds.

Table 19: Predicted distribution of proportions of chicks identified as EPY. The empirical value for in red-winged and yellow-shouldered blackbird is included, as well as the predicted empirical value for tricolored blackbird (in parentheses).

	2.50%	**50%**	97.50%	Emp. value
RWBL	0.10	**0.15**	0.20	0.26
YSBL	0.05	**0.092**	0.14	0.23
TRBL	0.13	**0.17**	0.21	(0.29-0.42)

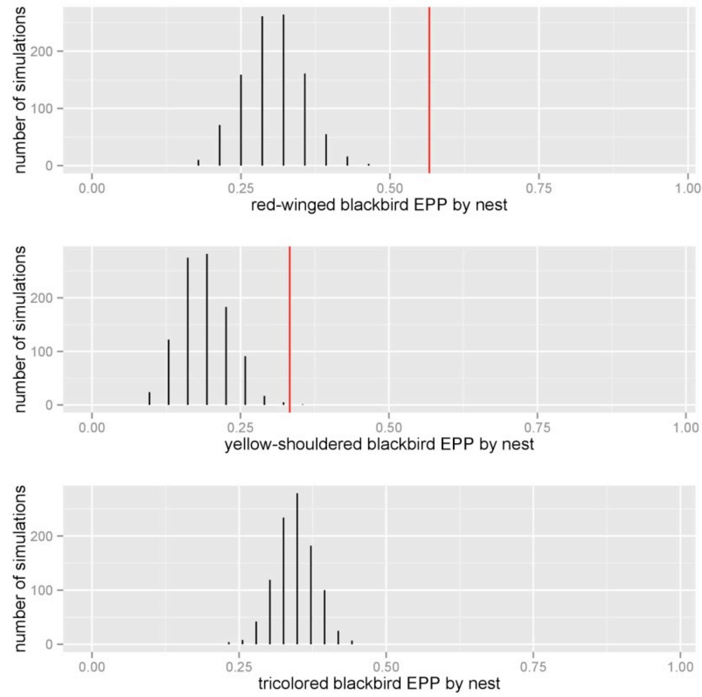

Figure 16: Distribution of inferred number of nests with EPP across 1000 runs. Solid line shows the empirically measured EPP rate for red-winged and yellow-shouldered blackbirds.

Table 20: Predicted distribution of proportion of nests containing at least one EPY.

	2.50%	**50%**	97.50%	Emp. value
RWBL	0.21	**0.32**	0.42	0.48
YSBL	0.11	**0.20**	0.28	0.37
TRBL	0.28	**0.35**	0.42	(0.53-0.65)

EPP was not significantly different across species, whether defined by number of chicks that were EPY ($\chi^2 = 0.02$, df = 1, $P = 0.89$) or by number of number of nests containing EPY ($\chi^2 = 0.09$, df = 1, $P = 0.76$). These results were consistent regardless of

whether the higher or lower estimate in tricolored blackbird was used. Therefore, there was no significant difference in EPP across the three species.

5.3.2. Gene polymorphism

I sequenced an average of 0.45 ± 0.12 SD (range 0.25 in HK3 – 0.53 in ENO1) of the total coding region of each gene. Summary statistics from DnaSP are shown in Table 21. Also shown are the mitochondrial genes (CR and ND2) characterized in Barker *et al.* (2012).

The nuclear genes were more conserved than the mitochondrial genes in the red-winged blackbirds (t = 3.16, df = 8, P = 0.013) but not in the tricolored blackbirds (t = 0.67, df = 8, P = 0.52). This result weakly supports the observation that animal mitochondrial DNA evolves more quickly than nuclear DNA (Edwards *et al.* 2005b). However, this comparison should be interpreted with caution because of the large difference in sample size. As in microsatellite diversity (Chapter 2), sample sizes influence observed levels of diversity.

Across species, diversity corresponded with the census sizes (and likely effective population sizes) of each species. Red-winged blackbirds exhibited the greatest diversity, while tricolored blackbirds had intermediate levels and yellow-shouldered blackbirds had identical haplotypes at every gene except one. Without greater sample sizes per species, the most that can be said is that the genes exhibit variation in evolutionary rate and that there appears to be a species effect. Overall lower levels of

diversity in nuclear DNA could be due to a mixture of insufficient sampling, purifying selection, or a slow rate of molecular evolution.

Table 21: Summary of polymorphism data for six candidate Sfps and two control genes in the three *Agelaius* species. ORF = total length of gene (longest open reading frame), Seq = length of sequenced region, Spp = species, n = number of individuals sampled, k = number of unique haplotypes, *h* = haplotype diversity, S = number of segregating sites, η = number of mutations, π = nucleotide diversity, SD = standard deviation. Genes in gray boxes are from Barker *et al.* (2012). Mean ± SD for π is shown for each species.

Spp	Gene	ORF (bp)	Seq (bp)	n	*k*	*h*	S	η	π	SD (π)
RWBL	ACRBP	1431	717	3	3	1	5	5	0.0046	0.0018
	CA6	1620	666	3	2	0.67	7	7	0.007	0.0033
	GALK1	1341	582	3	2	0.67	1	1	0.0012	0.00054
	HK3	3162	783	3	3	1	3	3	0.0026	0.0009
	REG4	477	321	3	3	1	2	2	0.0042	0.0014
	SPAG6	1617	666	3	2	0.67	1	1	0.001	0.00047
	ENO1	1476	792	3	2	0.67	1	1	0.0008	0.0004
	PGK1	1251	477	3	1	0	0	0	0	0
							Mean + SD		0.0027	0.0024
	CR		1208	31	25	0.98	65	67	0.009	0.0014
	ND2		1041	31	17	0.92	55	55	0.0077	0.0014
TRBL	ACRBP	1431	717	3	1	0	0	0	0	0
	CA6	1620	666	3	2	0.67	2	2	0.002	0.00094
	GALK1	1341	582	3	1	0	0	0	0	0
	HK3	3162	783	3	2	0.67	3	3	0.0026	0.0012
	REG4	477	321	3	2	0.67	2	2	0.0042	0.002
	SPAG6	1617	666	3	1	0	0	0	0	0
	ENO1	1476	792	3	3	1	2	2	0.0017	0.00056
	PGK1	1251	477	3	1	0	0	0	0	0
							Mean + SD		0.0013	0.0016
	CR		1208	10	8	0.93	12	12	0.0032	0.0005
	ND2		1041	10	3	0.69	3	3	0.0011	0.0003
YSBL	ACRBP	1431	717	3	1	0	0	0	0	0
	CA6	1620	666	3	3	1	2	2	0.002	0.00067

GALK1	1341	582	3	1	0	0	0	0	0
HK3	3162	783	3	1	0	0	0	0	0
REG4	477	321	3	1	0	0	0	0	0
SPAG6	1617	666	3	1	0	0	0	0	0
ENO1	1476	792	3	1	0	0	0	0	0
PGK1	1251	477	3	1	0	0	0	0	0
								0.00025	0.00082

5.3.3. Tests of selection

No genes were found to be under positive selection. In each of the three model tests (M1a-M2a, M7-M8, and M8-M8a), likelihood scores of paired models were nearly identical, leading to small test statistics ($-2(\ln L_1 - \ln L_2)$) and high *P*-values when examined on a chi-square distribution (Table 22). For three of the candidate genes (GALK, REG4, SPAG6) and both control genes (PGK1, ENO1), the ω-value in M2a at ω3, the site class where ω was permitted to exceed 1, remained at 1, indicating there was no evidence that ω exceeded 1 in any codon. Thus, there was instead strong evidence of purifying selection of candidate Sfps across the entire tree, including the *Geospiza* outgroup. Contrary to the results reported in Berlin *et al.* (2008), in which acrosin was found to be rapidly evolving across three orders (ducks, pigeons and chicken-like birds), ACRBP showed very few polymorphisms within the three *Agelaius* species.

In four of the six candidate genes (ACRBP, CA6, HK3, and REG4), the Bayes Empirical Bayes inference (Yang *et al.* 2005) identified a few positively selected sites, some with >70% confidence. The more sensitive M8-M8a comparison identified more

sites putatively under positive selection than the M1a-M2a comparisons. (Positively selected sites identified by M1a-M2a are not shown but were always subsets of or identical to the sites identified by M8-M8a.)

Given these null results in the site models, there was no reason to suspect branch or branch-sites models would be a better fit. In addition, the three species show similar levels of extra-pair mating and thus are inferred to experience similar levels of sperm competition, corresponding to little or no variation in trait value across the different lineages.

Table 22: Results of codeml analysis of eight genes across the three *Agelaius* species. The second column shows the test statistic, ($-2(\ln L_1 - \ln L_2)$), applied to each pair of models. All differences are extremely small and correspond to high P-values. For the positively selected sites, * corresponds to P > 70% and ** corresponds to P > 80% as predicted by Bayes Empirical Bayes analysis (Yang et al. 2005).

Gene	M1-M2 df =2	P	Parameter estimates under M2a
ACRBP	0.0467	0.98	p1: 0.80, w1: 0
			p2: 0, w2: 1
			p3: 0.20, w3: 1.14
CA6	0.17	0.92	p1: 0.84, w1: 0
			p2: 0, w2: 1
			p3: 0.16, w3: 1.36
GALK1	2E-05	1	p1: 1, w1: 0.16
			p2: 0, w2: 1
			p3: 0, w3: 1
HK3	0.41	0.81	p1: 0.96, w1: 0.17
			p2: 0, w2: 1
			p3: 0.04, w3: 4.66
REG4	2E-06	1	p1: 0.58, w1: 0
			p2: 0.33, w2:1
			p3: 0.09, w3:1

SPAG6	6E-06	1	p1: 1, w1: 0.22
			p2: 0, w2: 1
			p3: 0, w3:1
ENO1	2.26	0.32	p1: 1, w1:0
			p2: 0, w2: 1
			p3: 0, w3:1
PGK1	4E-06	1	p1: 1, w1: 0.06
			p2: 0, w2: 1
			p3: 0, w3: 1
Gene	**M7-M8**	**P**	**Parameter estimates**
	df = 2		**under M8**
ACRBP	0.194	0.91	p0 = 0.80 p = 0.005 q = 2.25
			(p1 = 0.20) w = 1.14
CA6	0.41	0.81	p0 = 0.84 p = 0.005 q = 0.99919
			(p1 = 0.16) w = 1.36
GALK1	0.000118	0.999945	p0 = 0.99999 p = 19.40 q = 99.00
			(p1 = 0.00001) w = 1.00
HK3	0.473	0.79	p0 = 0.96 p = 20.50 q = 99.00
			(p1 = 0.038) w = 4.66
REG4	0.0297	0.98	p0 = 0.95 p = 0.020 q = 0.037
			(p1 = 0.054) w = 1.00
SPAG6	4.4E-05	1	p0 = 0.99999 p = 28.81 q = 99.00
			(p1 = 0.00001) w = 1.00
ENO1	2.26	0.32	p0 = 0.99999 p = 0.005 q = 99.00
			(p1 = 0.00001) w = 1.00000
PGK1	0.000254	0.999875	p0 = 0.99999 p = 6.23 q = 99.00
			(p1 = 0.00001) w = 1.00

Gene	M8-M8a	P	Positively selected sites
	df = 1		
ACRBP	0.0467	P >> 0.05	53*, 133*, 214*, 226
CA6	0.17	P >> 0.05	35, 58*, 97, 129**, 167
GALK1	9.8E-05	P >> 0.05	
HK3	0.409	P >> 0.05	192,222
REG4	0.985	P >> 0.05	12,99
SPAG6	3.8E-05	P >> 0.05	
ENO1	8E-06	P >> 0.05	
PGK1	6E-06	P >> 0.05	

5.4. Discussion

Seminal fluid proteins (Sfps) show signatures in many surveyed insect and mammalian species of rapid evolution (Findlay *et al.* 2009; Ramm *et al.* 2008; Swanson *et al.* 2001; Turner *et al.* 2008), with mating system predicted to determine the strength of selection exerted on these proteins (Dorus *et al.* 2004; Walters & Harrison 2011). I searched for variation in protein evolution in a group of birds predicted to vary in genetic mating system and found no evidence either for variation in sperm competition across three species, or for positive selection across six candidate Sfps. These results contrast studies that have examined similar numbers of avian gamete-recognition genes and found evidence of positive evolution (Berlin *et al.* 2008; Calkins *et al.* 2007; Ceplitis &

Ellegren 2004). The high uniformity can be explained for some of the genes by pleiotropic constraints but is unexpected for one protein that plays an explicit role in sperm-egg interactions. Reasons for the uniformity of both EPP rate and gene sequences are discussed.

5.4.1. No difference in EPP rates

The evolutionary history of icterids (the family that includes *Agelaius*) suggests that the relationship between traits suggestive of increased sexual selection and the actual genetic mating system may be complex (Price 2009; Price *et al.* 2009). Given the result that the yellow-shouldered blackbird had comparable EPP levels to the red-winged and tricolored blackbird, extra-pair mating does not appear to be a life-history correlate that accompanied the change to temperate breeding and the gain of sexual dimorphism. Instead, because it appears in a tropical, sexually monomorphic species, EPP could potentially be pervasive across the genus. Ideally, sampling the two species in Cuba (tawny-shouldered and red-shouldered blackbird) would resolve the question of whether there is variation at all in EPP within the genus. A finding of EPP in the two Cuban species would implicate genetic polyandry as a shared derived trait, which would be substantially different from the hypothesized evolution of social polygyny in *Agelaius* (Barker *et al.* 2008; Searcy *et al.* 1999). This discordance raises intriguing questions about the amount of disparity between the evolution of genetic versus social mating systems in blackbirds and beyond.

Especially interesting to uncover would be the genetic mating system of the red-shouldered blackbird, the sister species to the red-winged blackbird that was once thought to be a subspecies but gained species status after fixed morphological, behavioral and genetic divergence was observed from field observations and phylogenetic analyses (Barker *et al.* 2008; Whittingham *et al.* 1992; Whittingham *et al.* 1996). If either species shows a decreased rate of EPP relative to its congeners, then a comparison of Sfp evolution along monogamous vs. polyandrous lineages could still be feasible to examine the effect of genetic mating system on protein evolution. If there is uniformity in EPP rates in the entire genus, then perhaps the study scope needs to be expanded to other genera or orders with greater variation in sperm competition. A wider scope could also increase the likelihood of finding patterns in protein-coding gene sequences, especially given the low overall divergence rates between birds (see below). Consistent with this observation, the only other studies finding evidence of positive selection in avian Sfps have surveyed a much broader sample of taxonomic orders, though these studies involved non-passerines (Berlin *et al.* 2008; Calkins *et al.* 2007; Ceplitis & Ellegren 2004).

5.4.2. No evidence for positive selection in candidate genes

Several reasons may explain why the models allowing for selection were rejected in all genes. First, a technical reason for these results is the limited power of the analysis due to a low number of species, number of genes, and number of nucleotides surveyed

per gene. The likelihood ratio tests used in codeml generally require gene sequences from a minimum of five to six species, with accuracy increasing to near 100% with 17 species (Anisimova *et al.* 2001). Such a sample size was clearly not present in a three-species comparison. The chi-square test also gives conservative results, especially when sequences are short and highly conserved (Anisimova *et al.* 2001), although given the data, it is very unlikely that the null models would be rejected by any statistical test.

Beyond issues of power, the fact remains that there simply were not that many candidate proteins to select from. The list of Sfps generated in Chapter 4 contained few proteins with explicit reproductive functions, making it difficult to select genes that were not pleiotropically acting and therefore likely to experience selective constraints elsewhere in the body. This dearth of reproductively related proteins may have been due to taxonomic differences in ejaculate composition or the sites of protein synthesis (such as accessory glands, Chapter 4), although a rigorous comparison has not been conducted. The only three listed proteins detected with functions exclusive to reproduction were acrosin, acrosin-binding protein, and sperm-associated antigen 6. Acrosin proteins were unable to be sequenced without cloning, because they belong to a large gene family with similar sequences that would have complicated attempts to sequence a single gene and reliably compare the same ortholog across all three species. Therefore, I chose to analyze ACRBP and SPAG6 as two genes implicated in fertility and sperm motility, respectively.

5.4.3. Discussion by gene

Interestingly, neither ACRBP nor SPAG6 showed evidence of positive selection. SPAG6 exemplifies a microevolutionary reason for the lack of variation: Not all Sfps are under positive selection. Instead, many Sfps experience selective constraints depending on their specific functions in reproduction (Dean *et al.* 2009; Dorus *et al.* 2006; Dorus *et al.* 2010; Schumacher *et al.* 2014). For proteins involved in roles such as spermatogenesis or sperm integrity (such as SPAG6), purifying selection may be especially strong. One study has provided especially strong support for this argument by examining the difference in selective pressures on proteins with different roles in the reproductive process. Schumacher *et al.* (2014) classified 169 primate reproductive proteins as having pre-mating functions (e.g., sperm composition and assembly) or post-mating functions (e.g., capacitation and fertilization). Included in the data set as a pre-mating protein was SPAG2, a similar gene to SPAG6. dN/dS tests revealed that pre-mating proteins had lower median ω-values than post-mating proteins (although both were well below ω = 0.5, reflecting inconclusive evidence of positive selection in the post-mating proteins as well). Pre-mating proteins additionally had more protein interaction partners per protein, suggesting their involvement in multiple pathways also serves a source of functional constraint. If results for SPAG2 can be extrapolated to SPAG6, then this result supports the argument that for structural proteins, the benefit of gaining selectively

advantageous substitutions is outweighed by the risk of interfering with a critical piece of cellular machinery.

Variation in dN/dS was observed for post-mating proteins as well (Schumacher *et al.* 2014). This finding is consistent with the observation that even in *Drosophila*, well known for its rapidly evolving accessory-gland proteins, certain Sfps such as sex peptide do not show signs of positive selection (Findlay *et al.* 2008). Most likely, the variation in evolutionary rate again reflects the tension between the benefits conferred by an advantageous change and the disadvantage of mistranslation or a deleterious substitution. Such reasoning could help explain the more unexpected result of the absence of positive selection on ACRBP. While codeml did identify codons under positive selection (Table 22), the sequence overall did not show any improvement in fit to the models allowing for positive selection. This finding stands in contrast to a previous study showing acrosin, the protease modulated by ACRBP, as one of four gamete-recognition genes rapidly evolving across ducks, pigeons, and chicken-like birds (Berlin *et al.* 2008). Of course, given the different selective pressures exerted on different Sfps, the positive selection seen on acrosin may have no direct bearing on the selection on ACRBP. It would be interesting to see whether acrosin in *Agelaius* does show evidence of positive selection, implicating it as a consistent target of selection in birds. However, a direct comparison between the two studies may require an expanded taxonomic scale beyond the *Agelaius* clade.

Looking beyond the proteins with explicit reproductive functions, the results suggest in hindsight that the methods I used to select the other candidate genes may have been problematic. Two of the genes (GALK1 and REG4) were likely not suitable candidates to begin with, because their central roles in glycolysis and digestion suggest they face strong purifying selection in other organ systems. Indeed, a previous study of mouse Sfps found that protein-encoding genes expressed in other tissues have significantly lower dN/dS values than genes expressed exclusively in male accessory glands (Dean *et al.* 2009). Functional constraints probably also apply to a third gene, HK3, because of its role in glycolysis, despite preliminary evidence that variants of this protein are testis-specific and play an explicit role in testis and male reproductive function (Cavener 1980; Murray & Ball 1967).

The main justification for selecting these genes was their high fold-change differences in expression in at least one reproductive tissue relative to control tissues (Table 17). However, there are a few issues in relying on gene expression levels as a proxy for the functional significance of a protein in a given tissue. First, the correlation between mRNA and protein expression levels is variable and actually quite modest, estimated in mammalian cells from 27% (Ghazalpour *et al.* 2011) to 40% (Schwanhausser *et al.* 2011; Vogel & Marcotte 2012). Second, gene expression does not necessarily correlate with a protein's relative importance. In fact, evidence exists that highly expressed proteins evolve slowly (Drummond *et al.* 2005), because their very abundance

indicates their importance and thus the deleterious effects that most nonsynonymous substitutions would have. In addition, different Sfps occur with different abundances in seminal fluid (Findlay *et al.* 2008; Sharma *et al.* 2013), so the number of reads mapping back to the transcriptomes is not always high in absolute values. (However, relative abundances should not be affected, as tissue-specific Sfps should still be highly expressed relative to control tissues.) Overall, using fold-change in expression may have been a less dependable strategy when choosing genes than I initially realized.

The last candidate gene, CA6, was thought to be absent in the reproductive tract (Kaunisto *et al.* 1990), although other carbonic anhydrases are known to modulate pH in mammalian reproductive systems. CA6's reproductive function, and thus the type of selective pressure it might experience, are unclear. Its evolutionary pattern is similar to ACRBP's, in that it possesses a few sites predicted to be under positive selection, but overall the models allowing for positive selection are no better fits than the models without it. Further functional characterization is needed to understand these results, but from the current tests, it appears that CA6 is also under purifying selection.

5.4.4. Slow molecular evolutionary rate in birds

Finally, the low power conferred by lack of sequence divergence is itself of interest, offering a macroevolutionary reason explaining the negative results: Lack of variation in the *Agelaius* sequences may be due to the slow molecular evolutionary rates in bird genomes. For all the genes in the present study, sequences were very nearly

identical, such that the number of synonymous mutations was very low, and the number of nonsynonymous mutations even lower. As a result, there were almost no segregating sites to analyze, much less strong evidence of positive selection. This low divergence has impeded other studies surveying the relationship between mating system and species divergence (primates, Wong 2010), where the conclusion was similarly that selection was impossible to detect.

Such low divergence could be common across coding regions in birds, especially compared with other areas of the genome. Nuclear DNA is known to evolve more slowly than mitochondrial DNA (Brown *et al.* 1979). While the polymorphism analysis in the current study lacked power to provide a clear comparison between nucleotide diversity in nDNA versus mtDNA in blackbirds (Table 21), the pattern of faster evolution in mtDNA is consistent in birds (Edwards *et al.* 2005a). In fact, nDNA and mtDNA sequences are applied to phylogenies at different scales, with the more slowly-evolving nDNA used to resolve higher-level phylogenies and the more quickly-evolving mtDNA used to resolve relationships of recently evolved species (Edwards *et al.* 2005a; Weibel & Moore 2002). For studies (such as the present study) wishing to test evolution of specific nuclear genes but facing conserved coding sequences, one alternative is to compare intronic sequences of nuclear DNA, since these noncoding regions experience relaxed constraints and exhibit higher overall mutation rates than coding regions (Prychitko & Moore 1997). From visual inspection of the candidate Sfps, introns were

indeed observed to vary within and across species, exhibiting both indels and point mutations (pers. obs.). Certain stretches were heterozygous, however, and would require cloning to resolve the haplotypes. Alignment of highly diverged introns could also pose a challenge (Weibel & Moore 2002). Despite these additional steps, comparing the introns of the candidate genes may be an appropriate strategy to assess divergence in the genes of three closely related species.

This observation of low divergence in nDNA supports the argument that in birds, speciation is able to occur without much genetic change (Nei 1987). Beginning with allozyme studies, research has shown that proteins diverge more slowly in bird congeners than they do in congeners of other taxa (*Geospiza*, Aquadro & Avise 1982). Indeed, most hypotheses for speciation mechanisms in birds attribute reproductive barriers to prezygotic, not postzygotic, mating barriers (Chapter 4). Postzygotic barriers are surprisingly weak, with one study showing that chicken sperm was capable of hydrolyzing (with decreasing efficiency) the inner perivitelline layer of ova from turkey, quail, pheasant, peafowl, goose, duck, zebra finch, and dove (Stewart *et al.* 2004). Paired with the knowledge that birds hybridize readily, this striking result supports the argument that precopulatory traits such as plumage and song are argued to be more effective at preventing hybridization than molecular mechanisms such as hybrid inviability or sterility (Edwards *et al.* 2005b; Grant & Grant 2002, 2004; Price 1998).

Ultimately, the challenge will be to distinguish between the contributions of both micro- and macroevolutionary forces leading to sequence divergence (or lack thereof). The current data are not likely to be informative about the relative significance of purifying selection versus evolutionary rate. For example, the finding that the *G. fortis* outgroup sequences were also not significantly divergent from *Agelaius* demonstrates that both forces could be at work. Most likely the only way to distinguish between these two causes is to widen the scope of analysis, either through increased taxon sampling, gene sampling, or both.

Several methods could clarify whether the observed results are due to purifying selection or a slow rate of molecular evolution. First, testing across a taxonomically broader sample might resolve the uncertainty around whether the sequences are simply non-variant and thus uninformative regardless of how many species are sampled, or if the sequences in the current study are conserved relative to more distantly related species. If the only way to detect positive selection is to survey species across diverse orders, as the current work on avian Sfps has done (Berlin *et al.* 2008; Calkins *et al.* 2007; Ceplitis & Ellegren 2004), then that would indicate the scope of molecular evolution is different from other vertebrate taxa, in which Sfps between closely related species show high levels of divergence (e.g., mice, Ramm *et al.* 2009). Second, sequencing genes and gene regions not considered in the present study, such as acrosin and the intronic regions of the candidate genes, could provide a complementary view of the current

results. Acrosin is a critical Sfp that has been demonstrated to display positive selection across bird orders (Berlin *et al.* 2008), and introns are expected to have greater information content due to higher levels of diversity (Prychitko & Moore 1997).

Finally, future studies can take advantage of available high-throughput approaches, such as genotyping by sequencing or bait capture, for targeted or whole-exome sequencing (Coffey *et al.* 2011; Elshire *et al.* 2011; Sulonen *et al.* 2011). While intronic information would be lost, whole-exome capture would circumvent any issues of incomplete gene sequencing. A combination of sequencing approaches over multiple taxonomic orders would yield the most potential to place the current findings in context and evaluate the signature of selection in avian Sfps.

Appendix A

Proteins in red-winged blackbird seminal fluid identified in Trinotate (n = 157). An asterisk indicates a protein was also detected in the search using the zebra finch proteome.

UniProt ID	UniProt Ref Spp	Functional category	Name of protein
Q5ZJ60	CHICK	Amino acid interactions	3-hydroxyisobutyryl-CoA hydrolase, mitochondrial*
Q28DB5	XENTR	Amino acid interactions	Alanine aminotransferase 2
P00504	CHICK	Amino acid interactions	Aspartate aminotransferase, cytoplasmic*
P00508	CHICK	Amino acid interactions	Aspartate aminotransferase, mitochondrial*
P54687	HUMAN	Amino acid interactions	Branched-chain-amino-acid aminotransferase, cytosolic*
Q2KIR8	BOVIN	Amino acid interactions	L-threonine 3-dehydrogenase, mitochondrial
B5FZA8	TAEGU	Amino acid interactions	LYR motif-containing protein 4
Q8VDG5	MOUSE	Amino acid interactions	Phosphopantothenate--cysteine ligase
Q9GKX6	PIG	Carbohydrate interactions	Aldose 1-epimerase*
P19140	ANAPL	Carbohydrate interactions	Alpha-enolase*
Q9TRY9	CANFA	Carbohydrate interactions	Beta-galactosidase
Q0V8R6	BOVIN	Carbohydrate interactions	Beta-hexosaminidase subunit alpha*
O00462	HUMAN	Carbohydrate interactions	Beta-mannosidase*
P05065	RAT	Carbohydrate interactions	Fructose-bisphosphate aldolase A
P51570	HUMAN	Carbohydrate interactions	Galactokinase*

P06745	MOUSE	Carbohydrate interactions	Glucose-6-phosphate isomerase*
O57479	COLLI	Carbohydrate interactions	Glyceraldehyde-3-phosphate dehydrogenase*
Q5R5V3	PONAB	Carbohydrate interactions	Glycerol-3-phosphate dehydrogenase 1-like protein*
P52790	HUMAN	Carbohydrate interactions	Hexokinase-3*
Q3U4H6	MOUSE	Carbohydrate interactions	Hexosaminidase D*
Q9PW07	COLLI	Carbohydrate interactions	L-lactate dehydrogenase A chain*
P00337	CHICK	Carbohydrate interactions	L-lactate dehydrogenase B chain*
O00754	HUMAN	Carbohydrate interactions	Lysosomal alpha-mannosidase
Q9BYZ8	HUMAN	Carbohydrate interactions	Regenerating islet-derived protein 4
P00940	CHICK	Carbohydrate interactions	Triosephosphate isomerase
Q90593	CHICK	Chaperone	78 kDa glucose-regulated protein
Q2HJ94	BOVIN	Chaperone	DnaJ homolog subfamily A member 2
P08106	CHICK	Chaperone	Heat shock 70 kDa protein*
Q71U34	SAGOE	Chaperone	Heat shock cognate 71 kDa protein
P11501	CHICK	Chaperone	Heat shock protein HSP 90-alpha
P34058	RAT	Chaperone	Heat shock protein HSP 90-beta*
P60707	TRIVU	Cytoskeletal	Actin, cytoplasmic 1*
A2BDB0	XENLA	Cytoskeletal	Actin, cytoplasmic 2
P53478	CHICK	Cytoskeletal	Actin, cytoplasmic type 5
P35556	HUMAN	Cytoskeletal	Fibrillin-2
Q9JJV2	MOUSE	Cytoskeletal	Profilin-2
Q7TQD2	MOUSE	Cytoskeletal	Tubulin polymerization-promoting protein*
O14681	HUMAN	Defense/stress response	Etoposide-induced protein 2.4 homolog
P04041	RAT	Defense/stress response	Glutathione peroxidase 1*
P46412	MOUSE	Defense/stress	Glutathione peroxidase 3

		response	
P04210	CHICK	Defense/stress response	Ig lambda chain V-1 region
P20768	SUNMU	Defense/stress response	Ig mu chain C region
B9A064	HUMAN	Defense/stress response	Immunoglobulin lambda-like polypeptide 5
P13796	HUMAN	Defense/stress response	Plastin-2*
P12815	MOUSE	Defense/stress response	Programmed cell death protein 6
P80566	CHICK	Defense/stress response	Superoxide dismutase [Cu-Zn]*
P41976	BOVIN	Defense/stress response	Superoxide dismutase [Mn], mitochondrial*
P20108	MOUSE	Defense/stress response	Thioredoxin-dependent peroxide reductase, mitochondrial*
Q9WUU8	MOUSE	Defense/stress response	TNFAIP3-interacting protein 1
Q90647	CHICK	General metabolism	V-type proton ATPase catalytic subunit A
Q9NUB1	HUMAN	General metabolism	Acetyl-coenzyme A synthetase 2-like, mitochondrial
P16276	PIG	General metabolism	Aconitate hydratase, mitochondrial*
P05081	CHICK	General metabolism	Adenylate kinase isoenzyme 1*
P81178	MESAU	General metabolism	Aldehyde dehydrogenase, mitochondrial*
P19483	BOVIN	General metabolism	ATP synthase subunit alpha, mitochondrial*
Q5ZLC5	CHICK	General metabolism	ATP synthase subunit beta, mitochondrial*
P05631	BOVIN	General metabolism	ATP synthase subunit gamma, mitochondrial
O89016	MOUSE	General metabolism	ATP-binding cassette sub-family D member 4
Q0GNE0	IGUIG	General metabolism	Citrate synthase, mitochondrial

P05122	CHICK	General metabolism	Creatine kinase B-type*
P70079	CHICK	General metabolism	Creatine kinase U-type, mitochondrial
P00018	DRONO	General metabolism	Cytochrome c*
Q2KJE4	BOVIN	General metabolism	Electron transfer flavoprotein subunit alpha, mitochondrial*
Q9DCW4	MOUSE	General metabolism	Electron transfer flavoprotein subunit beta*
Q4R502	MACFA	General metabolism	Isocitrate dehydrogenase [NADP], mitochondrial
Q5ZME2	CHICK	General metabolism	Malate dehydrogenase, cytoplasmic*
Q32LG3	BOVIN	General metabolism	Malate dehydrogenase, mitochondrial*
P23368	HUMAN	General metabolism	NAD-dependent malic enzyme, mitochondrial*
Q8BMF3	MOUSE	General metabolism	NADP-dependent malic enzyme, mitochondrial*
P51903	CHICK	General metabolism	Phosphoglycerate kinase*
O75323	HUMAN	General metabolism	Protein NipSnap homolog 2*
Q5RE79	PONAB	General metabolism	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial*
P00548	CHICK	General metabolism	Pyruvate kinase muscle isozyme*
Q9YGL9	CHICK	General metabolism	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3
P28568	CHICK	General metabolism	Solute carrier family 2, facilitated glucose transporter member 3*
Q9YHT1	CHICK	General metabolism	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial*
Q9YHT2	CHICK	General metabolism	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial*

Q9YI37	COLLI	General metabolism	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial
Q2VRL0	CHICK	Lipid interactions	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase zeta-1
Q6NU46	XENLA	Lipid interactions	Acetyl-CoA acetyltransferase A, mitochondrial
Q9H845	HUMAN	Lipid interactions	Acyl-CoA dehydrogenase family member 9, mitochondrial*
P0C7M7	HUMAN	Lipid interactions	Acyl-coenzyme A synthetase ACSM4, mitochondrial*
Q86TX2	HUMAN	Lipid interactions	Acyl-coenzyme A thioesterase 1
O42296	ANAPL	Lipid interactions	Apolipoprotein A-I
Q3LXA3	HUMAN	Lipid interactions	Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)*
Q13822	HUMAN	Lipid interactions	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2
P23965	RAT	Lipid interactions	Enoyl-CoA delta isomerase 1, mitochondrial*
Q8BH95	MOUSE	Lipid interactions	Enoyl-CoA hydratase, mitochondrial
Q9I9P7	COTJA	Lipid interactions	Extracellular fatty acid-binding protein
Q04791	ANAPL	Lipid interactions	Fatty acyl-CoA hydrolase precursor, medium chain
P79274	PIG	Lipid interactions	Long-chain specific acyl-CoA dehydrogenase, mitochondrial
Q5ZKR7	CHICK	Lipid interactions	Long-chain-fatty-acid--CoA ligase ACSBG2
Q32QL6	CALJA	Lipid interactions	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial
Q29RK2	BOVIN	Lipid interactions	Pyruvate carboxylase, mitochondrial
P15651	RAT	Lipid interactions	Short-chain specific acyl-CoA dehydrogenase, mitochondrial
A5PF10	PIG	Lipid interactions	Sialidase-1
Q9DA37	MOUSE	Lipid interactions	Sphingomyelin synthase-related

			protein 1
P55809	HUMAN	Lipid interactions	Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial
P50544	MOUSE	Lipid interactions	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial
A1YFU7	SAGLB	Nucleic acid interactions	Cbp/p300-interacting transactivator 2
E9Q1U1	MOUSE	Nucleic acid interactions	Coiled-coil domain-containing protein 171
Q9YGI5	CHICK	Nucleic acid interactions	Deoxyribonuclease-1*
Q805F9	CHICK	Nucleic acid interactions	DNA damage-binding protein 1
O95363	HUMAN	Nucleic acid interactions	Phenylalanine--tRNA ligase, mitochondrial
Q08623	HUMAN	Nucleic acid interactions	Pseudouridine-5'-monophosphatase
P28700	MOUSE	Nucleic acid interactions	Retinoic acid receptor RXR-alpha
P49743	RAT	Nucleic acid interactions	Retinoic acid receptor RXR-beta
Q5ZLR6	CHICK	Nucleic acid interactions	Rho guanine nucleotide exchange factor 6
Q9NW13	HUMAN	Nucleic acid interactions	RNA-binding protein 28
Q13535	HUMAN	Nucleic acid interactions	Serine/threonine-protein kinase ATR
Q13404	HUMAN	Nucleic acid interactions	Ubiquitin-conjugating enzyme E2 variant 1
Q9H0C1	HUMAN	Nucleic acid interactions	Zinc finger MYND domain-containing protein 12
Q6ZNC4	HUMAN	Nucleic acid interactions	Zinc finger protein 704
Q8N1W2	HUMAN	Nucleic acid interactions	Zinc finger protein 710
P23280	HUMAN	Other	Carbonic anhydrase 6*
Q8AXV0	CHICK	Other	Endophilin-A2
P70669	MOUSE	Other	Metalloendopeptidase homolog PEX
Q8WXI7	HUMAN	Other	Mucin-16

Q5ZKD5	CHICK	Other	RRP12-like protein
P19121	CHICK	Other	Serum albumin*
Q9Y6N5	HUMAN	Other	Sulfide:quinone oxidoreductase, mitochondrial
Q28520	MACMU	Other	Vitamin K-dependent protein S
Q2UVH8	MELGA	Protease	Acrosin
P04186	MOUSE	Protease	Complement factor B
Q8UW59	CHICK	Protease	Protein DJ-1*
Q6P1B1	MOUSE	Protease	Xaa-Pro aminopeptidase 1
P34065	CHICK	Protease	Proteasome subunit beta type-5
P12763	BOVIN	Protease inhibitor	Alpha-2-HS-glycoprotein
P32262	SHEEP	Protease inhibitor	Antithrombin-III
P10184	CHICK	Protease inhibitor	Ovoinhibitor
P13696	BOVIN	Protease inhibitor	Phosphatidylethanolamine-binding protein 1*
Q7Z7A4	HUMAN	Protease inhibitor	PX domain-containing protein kinase-like protein
P62262	SHEEP	Protein modification	14-3-3 protein epsilon
Q5ZKC9	CHICK	Protein modification	14-3-3 protein zeta
Q4R596	MACFA	Protein modification	Adenosylhomocysteinase*
O93477	XENLA	Protein modification	Adenosylhomocysteinase B
A6QL63	HUMAN	Protein modification	Ankyrin repeat and BTB/POZ domain-containing protein BTBD11
Q5RBM6	PONAB	Protein modification	Beta-ureidopropionase*
P62155	XENLA	Protein modification	Calmodulin*
Q6IQX7	MOUSE	Protein modification	Chondroitin sulfate synthase 2
Q5I0D1	RAT	Protein modification	Glyoxalase domain-containing protein 4
Q14C86	HUMAN	Protein modification	GTPase-activating protein and VPS9 domain-containing protein 1
Q9BSH5	HUMAN	Protein modification	Haloacid dehalogenase-like hydrolase domain-containing

			protein 3
Q5R4X0	PONAB	Protein modification	Inositol monophosphatase 1*
Q4R826	MACFA	Protein modification	Isochorismatase domain-containing protein 2, mitochondrial
Q5T700	HUMAN	Protein modification	Low-density lipoprotein receptor class A domain-containing protein 1
Q96S96	HUMAN	Protein modification	Phosphatidylethanolamine-binding protein 4
Q8JGM4	CHICK	Protein modification	Sulfhydryl oxidase 1
Q5F3K4	CHICK	Protein modification	WD repeat-containing protein 48*
Q8N4Q0	HUMAN	Protein modification	Zinc-binding alcohol dehydrogenase domain-containing protein 2
Q8NEB7	HUMAN	Sperm protein	Acrosin-binding protein*
Q9JLI7	MOUSE	Sperm protein	Sperm-associated antigen 6*
Q63HQ0	HUMAN	Transport	AP-1 complex-associated regulatory protein
Q5F448	CHICK	Transport	Golgi pH regulator
P42558	CHICK	Transport	GTP-binding nuclear protein Ran
Q5F334	CHICK	Transport	Leucine-rich repeat-containing protein 59
Q8WU76	HUMAN	Transport	Sec1 family domain-containing protein 2

Appendix B

Proteins identified only with the zebra finch proteome (n = 17).

Functional category	**Name of protein**
Carbohydrate interactions	fucosidase, alpha-L- 1, tissue
Defense/stress response	lactotransferrin
General metabolism	fumarate hydratase
Lipid interactions	phospholipase C, zeta 1
Nucleic acid interactions	5'-nucleotidase, cytosolic IB autoimmune infertility related
Nucleic acid interactions	nucleoside diphosphate kinase B
Nucleic acid interactions	PHD finger protein 3
Nucleic acid interactions	RAN member RAS oncogene family variant 1
Protease inhibitor	alpha-1-antitrypsin
Protease inhibitor	cystatin variant 3 precursor
Protein modification	4-aminobutyrate aminotransferase
Protein modification	carboxylesterase 1 (monocyte/macrophage serine esterase 1)
Protein modification	transthyretin
Protein modification	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
Transport	importin 5
Other	AKR1B1 (aldo-keto reductase family)
Uncharacterized	uncharacterized

References

Adams EM, Wolfner MF (2007) Seminal proteins but not sperm induce morphological changes in the Drosophila melanogaster female reproductive tract during sperm storage. *Journal of Insect Physiology* **53**, 319-331.

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207.

Aire TA (2007) Anatomy of the testis and male reproductive tract. In: *Reproductive Biology and Phylogeny of Birds* (ed. Jamieson BGM). Science Publishers, Enfield, NH.

Akcay E, Roughgarden J (2007) Extra-pair paternity in birds: review of the genetic benefits. *Evolutionary Ecology Research* **9**, 855-868.

Albrecht T, Kleven O, Kreisinger J, *et al.* (2013) Sperm competition in tropical versus temperate zone birds. *Proceedings of the Royal Society B-Biological Sciences* **280**.

Alem S, Streiff R, Courtois B, *et al.* (2013) Genetic architecture of sensory exploitation: QTL mapping of female and male receiver traits in an acoustic moth. *Journal of Evolutionary Biology* **26**, 2581-2596.

Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48.

Andres JA, Maroja LS, Harrison RG (2008) Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proceedings of the Royal Society B-Biological Sciences* **275**, 1975-1983.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Anisimova M, Bielawski JP, Yang ZH (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**, 1585-1592.

Aquadro CF, Avise JC (1982) Evolutionary genetics of birds. 6. A re-examination of protein divergence using varied electrophoretic conditions. *Evolution* **36**, 1003-1019.

Arct A, Drobniak S, Podmokła E, Gustafson L, Cichoń M (2013) Benefits of extra-pair mating may depend on environmental conditions—an experimental study in the blue tit (Cyanistes caeruleus). *Behavioral Ecology and Sociobiology* **67**, 1809-1815.

Arnold KE, Owens IPF (2002) Extra-pair paternity and egg dumping in birds: life history, parental care and the risk of retaliation. *Proceedings of the Royal Society B-Biological Sciences* **269**, 1263-1269.

Arnqvist G, Kirkpatrick M (2005) The evolution of infidelity in socially monogamous passerines: The strength of direct and indirect selection on extrapair copulation behavior in females. *American Naturalist* **165**, S26-S37.

Arnqvist G, Rowe L (2005) *Sexual conflict.* Princeton University Press, Princeton, NJ.

Axelsson E, Ratnakumar A, Arendt M-L, *et al.* (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360-364.

Ball RM, Freeman S, James FC, Bermingham E, Avise JC (1988) Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 1558-1562.

Barker FK, Bell JJ, Bogdanowicz SM, *et al.* (2011) Permanent genetic resources added to Molecular Ecology Resources Database 1 June 2011–31 July 2011. *Molecular Ecology Resources* **11**, 1124-1126.

Barker FK, Benesh MK, Vandergon AJ, Lanyon SM (2012) Contrasting evolutionary dynamics and information content of the avian mitochondrial control region and ND2 gene. *PLoS ONE* **7**, e46403.

Barker FK, Vandergon AJ, Lanyon SM (2008) Species status of the Red-shouldered Blackbird (Agelaius assimilis): Implications for ecological, morphological, and behavioral evolution in agelaius. *Auk* **125**, 87-94.

Barnés V (1945) A new form of Agelaius from Mona Island, Puerto Rico. *Auk* **62**, 299-300.

Bateman AJ (1948) Intra-sexual selection in Drosophila. *Heredity* **2**, 349-368.

Beedy EC, Hamilton W, J. III (1999) Tricolored blackbird (Agelaius tricolor). In: *The Birds of North America Online* (ed. Poole A). Cornell Lab of Ornithology, Ithaca.

Beletsky L (1996) *The Red-Winged Blackbird: the Biology of a Stronly Polygynous Songbird.* Academic Press, New York.

Berlin S, Qu LJ, Ellegren H (2008) Adaptive evolution of gamete-recognition proteins in birds. *Journal of Molecular Evolution* **67**, 488-496.

Birkhead TR, Martinez JG, Burke T, Froman DP (1999) Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proceedings of the Royal Society B-Biological Sciences* **266**, 1759-1764.

Birkhead TR, Møller AP (1992) *Sperm Competition in Birds: Evolutionary Causes and Consequences.* Academic Press, New York.

Birkhead TR, Pizzari T (2002) Postcopulatory sexual selection. *Nature Reviews Genetics* **3**, 262-273.

Bollmer JL, Dunn PO, Freeman-Gallant CR, Whittingham LA (2012) Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society B-Biological Sciences* **279**, 4778-4785.

Bonier F, Eikenaar C, Martin PR, Moore IT (2014) Extrapair paternity rates vary with latitude and elevation in Emberizid sparrows. *The American Naturalist* **183**, 54-61.

Bray OE, Kennelly JJ, Guarino JL (1975) Fertility of eggs produced on territories of vasectomized red-winged blackbirds. *Wilson Bulletin* **87**, 187-195.

Briskie JV, Montgomerie R (2007) Testis size, sperm size and sperm competition. In: *Reproductive Biology and Phylogeny of Birds* (ed. BGM J), pp. 513-552. Science Publishers, Enfield, NH.

Briskie JV, Montgomerie R, Birkhead TR (1997) The evolution of sperm size in birds. *Evolution* **51**, 937-945.

Brohede J, Møller AP, Ellegren H (2004) Individual variation in microsatellite mutation rate in barn swallows. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **545**, 73-80.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1967-1971.

Byrne K, Leahy T, McCulloch R, Colgrave ML, Holland MK (2012) Comprehensive mapping of the bull sperm surface proteome. *Proteomics* **12**, 3559-3579.

Calkins JD, El-Hinn D, Swanson WJ (2007) Adaptive evolution in an avian reproductive protein: ZP3. *Journal of Molecular Evolution* **65**, 555-563.

Cavener DR (1980) Genetics of male-specific glucose-oxidase and the identification of other unusual hexose enzymes in Drosophila melanogaster. *Biochemical Genetics* **18**, 929-937.

Ceplitis H, Ellegren H (2004) Adaptive molecular evolution of HINTW, a female-specific gene in birds. *Molecular Biology and Evolution* **21**, 249-254.

Chapman T (2001) Seminal fluid-mediated fitness traits in Drosophila. *Heredity* **87**, 511-521.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205.

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237-268.

Chegwidden WR, Carter ND (2000) Introduction to the carbonic anhydrases. In: *The Carbonic Anhydrases: New Horizons* (eds. Chegwidden WR, Carter ND, Edwards YH). Birhäuser Verlag, Basel, Switzerland.

Clark AG, Aguade M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland protein loci in Drosophila melanogaster. *Genetics* **139**, 189-201.

Clark AG, Begun DJ, Prout T (1999) Female × male interactions in Drosophila sperm competition. *Science* **283**, 217-220.

Clark NL, Swanson WJ (2005) Pervasive adaptive evolution in primate seminal proteins. *PLoS Genetics* **1**, 335-342.

Coffey AJ, Kokocinski F, Calafato MS, *et al.* (2011) The GENCODE exome: sequencing the complete human exome. *European Journal of Human Genetics* **19**, 827-831.

Conesa A, Gotz S, Garcia-Gomez JM, *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.

Conrad KF, Johnston PV, Crossman C, *et al.* (2001) High levels of extra-pair paternity in an isolated, low-density, island population of tree swallows (Tachycineta bicolor). *Molecular Ecology* **10**, 1301-1308.

Cook LF, Toft CA (2005) Dynamics of extinction: population decline in the colonially nesting Tricolored Blackbird Agelaius tricolor. *Bird Conservation International* **15**, 73-88.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.

Cornwallis CK, O'Connor EA (2009) Sperm:seminal fluid interactions and the adjustment of sperm quality in relation to female attractiveness. *Proceedings of the Royal Society B-Biological Sciences* **276**, 3467-3475.

Crow JF, Kimura M (1970) *An introduction to population genetics theory* New York, Harper & Row [1970].

Cruz A, López-Ortiz R, Ventosa-Febles EA, *et al.* (2005) Ecology and management of shiny cowbirds (Molothrus bonariensis) and endangered yellow-shouldered blackbirds (Agelaius xanthomus) in Puerto Rico. *Ornithological Monographs* **57**, 38-44.

Darwin C (1871) *The Descent of Man, and Selection in Relation to Sex.* William Pickering, London.

Dawson DA, Hanotte O, Greig C, Stewart IRK, Burke T (2000) Polymorphic microsatellites in the blue tit Parus caeruleus and their cross-species utility in 20 songbird families. *Molecular Ecology* **9**, 1941-1944.

Dawson RJG, Lisle Gibbs H, Hobson KA, Yezerinac SM (1997) Isolation of microsatellite DNA markers from a passerine bird, Dendroica petechia (the yellow warbler), and their use in population studies. *Heredity* **79**, 506-514.

Dean MD, Clark NL, Findlay GD, *et al.* (2009) Proteomics and Comparative Genomic Investigations Reveal Heterogeneity in Evolutionary Rate of Male Reproductive Proteins in Mice (Mus domesticus). *Molecular Biology and Evolution* **26**, 1733-1743.

Díaz-Rodríguez A, Lewis AR (2006) Growth patterns of nestlings and other aspects of reproductive biology of Agelaius xanthomus in artificial nesting structures: Final report, p. 56. USFWS, Boquerón, PR.

Digby ST, Howarth B (1972) Evidence for penetration of vitelline membrane by a trypsin-like acrosomal enzyme. *Poultry Science* **51**, 1801-&.

Diz AP, Martínez-Fernández M, Rolán-Álvarez E (2012) Proteomics in evolutionary ecology: linking the genotype with the phenotype. *Molecular Ecology* **21**, 1060-1080.

Do C, Waples RS, Peel D, *et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N-e) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

Dorus S, Busby SA, Gerike U, *et al.* (2006) Genomic and functional evolution of the Drosophila melanogaster sperm proteome. *Nature Genetics* **38**, 1440-1445.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36**, 1326-1329.

Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL (2010) Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. *Molecular Biology and Evolution* **27**, 1235-1246.

Douglas SB, Heath DD, Mennill DJ (2012) Low levels of extra-pair paternity in a neotropical duetting songbird, the rufous-and-white wren (Thryothorus rufalbus). *Condor* **114**, 393-400.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14338-14343.

Dubé C, Leclerc P, Baba T, Reyes-Moreno C, Bailey JL (2005) The proacrosin binding protein, sp32, is tyrosine phosphorylated during capacitation of pig sperm. *Journal of Andrology* **26**, 519-528.

Edwards SV, Gasper J, March M (1998) Genomics and polymorphism of Agph-DAB1, an Mhc class II B gene in red-winged blackbirds (Agelaius phoeniceus). *Molecular Biology and Evolution* **15**, 236-250.

Edwards SV, Jennings WB, Shedlock AM (2005a) Phylogenetics of modern birds in the era of genomics. *Proceedings of the Royal Society B-Biological Sciences* **272**, 979-992.

Edwards SV, Kingan SB, Calkins JD, *et al.* (2005b) Speciation in birds: Genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences* **102**, 6550-6557.

Eikenaar C, Bonier F, Martin PR, Moore IT (2013) High rates of extra-pair paternity in two equatorial populations of rufous-collared sparrow, Zonotrichia capensis. *Journal of Avian Biology* **44**, 600-602.

Ekblom R, French L, Slate J, Burke T (2010) Evolutionary analysis and expression profiling of zebra finch immune genes. *Genome Biology and Evolution* **2**, 781-790.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.

Ellegren H, Smeds L, Burri R, *et al.* (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756-760.

Elphick CS, Reed JM, Delehanty DJ (2007) Applications of reproductive biology to bird conservation and population management. In: *Reproductive Biology and Phylogeny of Birds* (ed. Jamieson BGM), pp. 367-399. Science Publishers, Enfield, NH.

Elshire RJ, Glaubitz JC, Sun Q, *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

Emlen DJ, Nijhout HF (2000) The development and evolution of exaggerated morphologies in insects. *Annual Review of Entomology* **45**, 661-708.

Emlen ST, Oring LW (1977) Ecology, sexual selection, and evolution of mating systems. *Science* **197**, 215-223.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Evans BJ, Charlesworth B (2013) The effect of nonindependent mate pairing on the effective population size. *Genetics* **193**, 545-556.

Falconer DS, Mackay T (1996) *Introduction to Quantitative Genetics.* Longman, Harlow, England.

Findlay GD, MacCoss MJ, Swanson WJ (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. *Genome Research* **19**, 886-896.

Findlay GD, Swanson WJ (2010) Proteomics enhances evolutionary and functional analysis of reproductive proteins. *Bioessays* **32**, 26-36.

Findlay GD, Yi XH, MacCoss MJ, Swanson WJ (2008) Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. *PLoS Biology* **6**, 1417-1426.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**, W29-W37.

Fiumera AC, Dumont BL, Clark AG (2005) Sperm competitive ability in Drosophila melanogaster associated with variation in male reproductive proteins. *Genetics* **169**, 243-257.

Frankham R (1995) Effective population size/adult population size rations in wildlife: a review. *Genetical Research* **66**, 95-107.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity* **78**, 311-327.

Frankham R (1998) Inbreeding and extinction: Island populations. *Conservation Biology* **12**, 665-675.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK ; New York.

Freeman-Gallant CR, Wheelwright NT, Meiklejohn KE, *et al.* (2005) Little effect of extrapair paternity on the opportunity for sexual selection in Savannah sparrows (Passerculus sandwichensis). *Evolution* **59**, 422-430.

Friedman NR, Hofmann CM, Kondo B, Omland KE (2009) Correlated evolution of migration and sexual dichromatism in the New World orioles (Icterus). *Evolution* **63**, 3269-3274.

Garamszegi LZ, Moller AP (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. *Biological Reviews* **85**, 797-805.

García-Navas V, Ferrer ES, Bueno-Enciso J, *et al.* (2014) Extrapair paternity in Mediterranean blue tits: socioecological factors and the opportunity for sexual selection. *Behavioral Ecology* **25**, 228-238.

García del Rey E, Kleven O, Lifjeld JT (2012) Extrapair paternity in insular African Blue Tits (Cyanistes teneriffae) is no less frequent than in continental Eurasian Blue Tits (Cyanistes caeruleus). *Ibis* **154**, 862-867.

Garrido O, Kirkconnell A (1996) Taxonomic status of the Cuban form of the Red-winged Blackbird. *Wilson Bulletin* **108**, 372-374.

Gasper JS, Shiina T, Inoko H, Edwards SV (2001) Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (Agelaius phoeniceus). *Genomics* **75**, 26-34.

Ghazalpour A, Bennett B, Petyuk VA, *et al.* (2011) Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genetics* **7**, e1001393.

Gibbs HL, Weatherhead PJ, Boag PT, *et al.* (1990) Realized reproductive success of polygynous red-winged blackbirds revealed by DNA markers. *Science* **250**, 1394-1397.

Gohli J, Anmarkrud JA, Johnsen A, *et al.* (2013) Female promiscuity is positively associated with neutral and selected genetic diversity in passerine birds. *Evolution* **67**, 1406-1419.

Gowaty PA, Karlin AA (1984) Multiple maternity and paternity in single broods of apparently monogamous eastern bluebirds (Sialia sialis). *Behavioral Ecology and Sociobiology* **15**, 91-95.

Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652.

Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707-711.

Grant PR, Grant BR (2014) Synergism of natural selection and introgression in the origin of a new species. *American Naturalist* **183**, 671-681.

Gray EM (1996) Female control of offspring paternity in a western population of red-winged blackbirds (Agelaius phoeniceus). *Behavioral Ecology and Sociobiology* **38**, 267-278.

Gray EM (1997a) Do female red-winged blackbirds benefit genetically from seeking extra-pair copulations? *Animal Behaviour* **53**, 605-623.

Gray EM (1997b) Female red-winged blackbirds accrue material benefits from copulating with extra-pair males. *Animal Behaviour* **53**, 625-639.

Griffith SC (2000) High fidelity on islands: a comparative study of extrapair paternity in passerine birds. *Behavioral Ecology* **11**, 265-273.

Griffith SC (2007) The evolution of infidelity in socially monogamous passerines: Neglected components of direct and indirect selection. *American Naturalist* **169**, 274-281.

Griffith SC, Owens IPF, Thuman KA (2002) Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology* **11**, 2195-2212.

Griffith SC, Stewart IRK, Dawson DA, Owens IPF, Burke T (1999) Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow (Passer domesticus): is there an 'island effect'? *Biological Journal of the Linnean Society* **68**, 303-316.

Gyllensten UB, JakobssonSven, Temrin H (1990) No evidence for illegitimate young in monogamous and polygynous warblers. *Nature* **343**, 168-170.

Hamilton W, J. III (1998) Tricolored blackbird itinerant breeding in California. *The Condor* **100**, 218-226.

Harris GC, Goto K (1984) Carbonic anhydrase activity of the reproductive tract tissues of aged male fowls and its relationship to semen production. *Journal of Reproduction and Fertility* **70**, 25-30.

Hartupee JC, Zhang H, Bonaldo MF, Soares MB, Dieckgraefe BK (2001) Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV. *Biochimica et Biophysica Acta* **1518**, 287-293.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101-104.

Hoffman JI, Simpson F, David P, *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences* **111**, 3775-3780.

Horrocks AJ, Stewart S, Jackson L, Wishart GJ (2000) Induction of acrosomal exocytosis in chicken spermatozoa by inner perivitelline-derived N-linked glycans. *Biochemical and Biophysical Research Communications* **278**, 84-89.

Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A (2014) multcomp: Simultaneous Inference in General Parametric Models.

Howarth B (1970) An examination for sperm capacitation in the fowl. *Biology of Reproduction* **3**, 338-341.

Hughes CR, Kavlie R, Johnson K (1998) Characterization of polymorphic trinucleotide microsatellite loci in the great-tailed grackle, Quiscalus mexicanus. *Molecular Ecology* **7**, 783-784.

Hundertmark KJ, Van Daele LJ (2010) Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics* **11**, 139-147.

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.

Jamieson A, Taylor SS (1997) Comparisons of three probability formulae for parentage exclusion. *Animal Genetics* **28**, 397-400.

Jennions MD, Petrie M (1997) Variation in mate choice and mating preferences: A review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society* **72**, 283-327.

Jennions MD, Petrie M (2000) Why do females mate multiply? A review of the genetic benefits. *Biological Reviews* **75**, 21-64.

Johnsen A, Andersen V, Sunding C, Lifjeld JT (2000) Female bluethroats enhance offspring immunocompetence through extra-pair copulations. *Nature* **406**, 296-299.

Johnson DL (1977) Inbreeding in populations with overlapping generations. *Genetics* **87**, 581-591.

Johnston SE, Gratten J, Berenos C, *et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.

Jones OR, Wang JL (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**, 551-555.

Jorde PE, Ryman N (1995) Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* **139**, 1077-1090.

Karhumaa P, Leinonen J, Parkkila S, *et al.* (2001) The identification of secreted carbonic anhydrase VI as a constitutive glycoprotein of human and rat milk. *Proceedings of the National Academy of Sciences* **98**, 11604-11608.

Kaunisto K, Parkkila S, Tammela T, Rönnberg L, Rajaniemi H (1990) Immunohistochemical localization of carbonic anhydrase isoenzymes in the human male reproductive tract. *Histochemistry* **94**, 381-386.

Kempenaers B, Verheyren GR, Dhondt AA (1997) Extrapair paternity in the blue tit (Parus caeruleus): female choice, male characteristics, and offspring quality. *Behavioral Ecology* **8**, 481-492.

Kirkpatrick M, Ryan MJ (1991) The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33-38.

Kleven O, Jacobsen F, Izadnegahdar R, Robertson RJ, Lifjeld JT (2006) No evidence of paternal genetic contribution to nestling cell-mediated immunity in the North American barn swallow. *Animal Behaviour* **71**, 839-845.

Kokko H, Ranta E, Ruxton G, Lundberg P (2002) Sexually transmitted disease and the evolution of mating systems. *Evolution* **56**, 1091-1100.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Krokene C, Lifjeld JT (2000) Variation in the frequency of extra-pair paternity in birds: a comparison of an island and a mainland population of blue tits. *Behaviour* **137**, 1317-1330.

Kulkarni S, Heeb P (2007) Social and sexual behaviours aid transmission of bacteria in birds. *Behavioural Processes* **74**, 88-92.

Künstner A, Wolf JBW, Backström N, *et al.* (2010) Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Molecular Ecology* **19**, 266-276.

Lack D, Emlen JT, Jr. (1939) Observations on Breeding Behavior in Tricolored Red-Wings. *The Condor* **41**, 225-230.

LaFlamme BA, Avila FW, Michalski K, Wolfner MF (2014) A Drosophila protease cascade member, seminal metalloprotease-1, is activated stepwise by male factors and requires female factors for full activity. *Genetics* **196**, 1117-1129.

Lagesen K, Hallin P, Rødland EA, *et al.* (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* **35**, 3100-3108.

Lande R (1988) Genetics and demography in biological conservation. *Science* **241**, 1455-1460.

Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25.

Leng N, Dawson JA, Thomson JA, *et al.* (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035-1043.

Li B, Dewey C (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

López-Ortiz R, Ventosa-Febles EA, Reitsma LR, Hengstenberg D, Deluca W (2002) Increasing nest success in the yellow-shouldered blackbird Agelaius xanthomus in southwest Puerto Rico. *Biological Conservation* **108**, 259-263.

Løvlie H, Gillingham MAF, Worley K, Pizzari T, Richardson DS (2013) Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. *Proceedings of the Royal Society B-Biological Sciences* **280**.

Macedo RH, And JK, Webster MS (2008) Extrapair paternity and sexual selection in socially monogamous birds: are tropical birds different? *The Auk* **125**, 769-777.

Madsen T, Shine R, Loman J, Hakansson T (1992) Why do female adders copulate so frequently? *Nature* **355**, 440-441.

McDonald D, Potts W (1994) Cooperative display and relatedness among males in a lek-mating bird. *Science* **266**, 1030-1032.

Medina-Miranda R, López-Ortiz R, Ramos-Álavarez K, *et al.* (2013) Yellow-shouldered blackbird/shiny cowbird population and habitat viability assessment workshop final report. IUCN SSC Conservation Breeding Specialist Group, Apple Valley, MN.

Michalczyk L, Millard AL, Martin OY, *et al.* (2011) Inbreeding promotes female promiscuity. *Science* **333**, 1739-1742.

Mohan J, Saini M, Joshi P (1995) Isolation of a spermatozoa motility inhibiting factor from chicken seminal plasma with antibacterial property. *Biochimica et Biophysica Acta* **1245**, 407-413.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621-628.

Mulder RA, Dunn PO, Cockburn A, Lazenby-Cohen KA, Howell MJ (1994) Helpers liberate female fairy-wrens from constraints on extra-pair mate choice. *Proceedings of the Royal Society of London B-Biological Sciences* **255**, 223-229.

Müller J, Müller K, Neinhuis C, Quandt D (2010) PhyDE: Phylogenetic Data Editor.

Murray RF, Ball JA (1967) Testis-specific and sex-associated hexokinases in Drosophila melanogaster. *Science* **156**, 81-&.

Naish KA, Seamons TR, Dauer MB, Hauser L, Quinn TP (2013) Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (Oncorhynchus mykiss) population released in the wild. *Molecular Ecology* **22**, 1295-1309.

Neff BD, Pitcher TE (2005) Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology* **14**, 19-38.

Nei M (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Neilson LI, Schneider PA, Van Deerlin PG, *et al.* (1999) cDNA cloning and characterization of a human sperm antigen (SPAG6) with homology to the product of the Chlamydomonas PF16 locus. *Genomics* **60**, 272-280.

Newhouse S, To D (2011) cmpfastq: A simple perl program that allows the user to compare QC filtered fastq files.

Nielsen R, Yang ZH (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929-936.

Nomura T (2008) Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications* **1**, 462-474.

Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution* **47**, 1329-1341.

O'Neil S, Emrich S (2013) Assessing de novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* **14**, 465.

Ockendon N, Griffith SC, Burke T (2009) Extrapair paternity in an insular population of house sparrows after the experimental introduction of individuals from the mainland. *Behavioral Ecology* **20**, 305-312.

Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE (2014) Female song is widespread and ancestral in songbirds. *Nature Communications* **5**.

Orians GH (1961) Ecology of blackbird (Agelaius) social systems. *Ecological Monographs* **31**, 285-&.

Orians GH (1969) On the evolution of mating systems in birds and mammals. *American Naturalist* **103**, 589-&.

Papakostas S, Vasemägi A, Vähä J-P, *et al.* (2012) A proteomics approach reveals divergent molecular responses to salinity in populations of European whitefish (Coregonus lavaretus). *Molecular Ecology* **21**, 3516-3530.

Paquette SR (2012) PopGenKit: useful functions for (batch) file conversion and data resampling in microsatellite databases.

Parker GA (1970) Sperm competition and its evolutionary consequences in insects. *Biological Reviews of the Cambridge Philosophical Society* **45**, 525-&.

Parker GA (1979) Sexual selection and sexual conflict. In: *Sexual Selection and Reproductive Competition in insects* (eds. Blum MS, Blum NA). Academic Press, New York.

Parker GA, Birkhead TR (2013) Polyandry: the history of a revolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* **368**.

Parkkila S (2000) An overview of the distribution and function of carbonic anhydrase in mammals. In: *The Carbonic Anhydrases: New Horizons* (eds. Chegwidden WR, Carter ND, Edwards YH). Birkhäuser Verlag, Basel, Switzerland.

Parkkila S, Kaunisto K, Rajaniemi L, *et al.* (1990) Immunohistochemical localization of carbonic anhydrase isoenzyme VI, isoenzyme II, and isoenzyme I in human parotid and submandibular glands. *Journal of Histochemistry & Cytochemistry* **38**, 941-947.

Payne R (1969) Breeding seasons and reproductive physiology of Tricolored Blackbirds and Red-winged Blackbirds. *University of California Publications in Zoology* **90**, 1-137.

Peakall R, Smouse PE (2006) Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Pellatt EJ, Birkhead TR (1994) Ejaculate size in zebra finches (Taeniopygia guttata) and a method for obtaining ejaculates from passerine birds. *Ibis* **136**, 97-101.

Pemberton AJ, Noble LR, Bishop JDD (2003) Frequency dependence in matings with water-borne sperm. *Journal of Evolutionary Biology* **16**, 289-301.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Peterson M, Whittaker D, Ambreth S, *et al.* (2012) De novo transcriptome sequencing in a songbird, the dark-eyed junco (Junco hyemalis): genomic tools for an ecological model system. *BMC Genomics* **13**, 305.

Petrie M, Doums C, Moller AP (1998) The degree of extra-pair paternity increases with genetic variability. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9390-9395.

Petrie M, Kempenaers B (1998) Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology & Evolution* **13**, 52-58.

Petrie M, Lipsitch M (1994) Avian polygyny is most likely in populations with high variability in heritable male fitness. *Proceedings of the Royal Society B-Biological Sciences* **256**, 275-280.

Phillipsen IC, Bowerman J, Blouin M (2010) Effective number of breeding adults in Oregon spotted frogs (Rana pretiosa): genetic estimates at two life stages. *Conservation Genetics* **11**, 737-745.

Plesnar-Bielak A, Skrzynecka AM, Prokop ZM, Radwan J (2012) Mating system affects population performance and extinction risk under environmental challenge. *Proceedings of the Royal Society B-Biological Sciences* **279**, 4661-4667.

Polson NG, Scott JG, Windle J (2013) Bayesian inference for logistic models using Polyá-Gamma latent variables. *Journal of the American Statistical Association* **108**, 1339-1349.

Post W (1981) Biology of the Yellow-shouldered Blackbird: Agelaius on a tropical island. *Bulletin of the Florida State Museum, Biological Science* **26**, 123-205.

Post W, Wiley JW (1976) The Yellow-shouldered Blackbird: present and future. *American Birds* **30**, 13-20.

Post W, Wiley JW (1977) Reproductive interactions of shiny cowbird and yellow-shouldered blackbird. *Condor* **79**, 176-184.

Prather JW, Cruz A (2006) Breeding biology of red-winged blackbirds in south Florida. *Southeastern Naturalist* **5**, 547-554.

Price JJ (2009) Evolution and life-history correlates of female song in the New World blackbirds. *Behavioral Ecology* **20**, 967-977.

Price JJ, Lanyon SM, Omland KE (2009) Losses of female song with changes from tropical to temperate breeding in the New World blackbirds. *Proceedings of the Royal Society B-Biological Sciences* **276**, 1971-1980.

Price T (1998) Sexual selection and natural selection in bird speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **353**, 251-260.

Primmer CR, Moller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* **5**, 365-378.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Prychitko TM, Moore WS (1997) The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). *Molecular Phylogenetics and Evolution* **8**, 193-204.

Pryke SR, Rollins LA, Griffith SC (2010) Females use multiple mating and genetically loaded sperm competition to target compatible genes. *Science* **329**, 964-967.

Pudovkin AI, Zaykin DV, Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* **144**, 383-387.

R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramaiya M, Johnson MG, Shaw B, *et al.* (2010) Morphologically cryptic biological species within the liverwort Frullania asagrayana. *American Journal of Botany* **97**, 1707-1718.

Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P (2009) Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and Its functional significance in sperm competition. *Molecular Biology and Evolution* **26**, 189-198.

Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD (2008) Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Molecular Biology and Evolution* **25**, 207-219.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology* **17**, 230-237.

Richardson DS, Komdeur J, Burke T, von Schantz T (2005) MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B-Biological Sciences* **272**, 759-767.

Rivkin E, Tres LL, Kaplan-Kraicer R, Shalgi R, Kierszenbaum AL (2000) Molecular cloning of rat sperm galactosyl receptor, a C-type lectin with in vitro egg binding activity. *Molecular Reproduction and Development* **56**, 401-411.

Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly Review of Biology* **71**, 3-35.

Rubin C-J, Megens H-J, Barrio AM, *et al.* (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences* **109**, 19529-19536.

Saether BE, Engen S, Lande R, *et al.* (2004) Time to extinction in relation to mating system and type of density regulation in populations with two sexes. *Journal of Animal Ecology* **73**, 925-934.

Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J (2011) Characterisation of the transcriptome of a wild great tit Parus major population by next generation sequencing. *BMC Genomics* **12**, 1-18.

Sapiro R, Kostetskii I, Olds-Clarke P, *et al.* (2002) Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. *Molecular and Cellular Biology* **22**, 6298-6305.

Sapiro R, Tarantino LM, Velazquez F, *et al.* (2000) Sperm antigen 6 is the murine homologue of the Chlamydomonas reinhardtii central apparatus protein encoded by the PF16 locus. *Biology of Reproduction* **62**, 511-518.

Sardell RJ, Arcese P, Keller LF, Reid JM (2012) Are there indirect fitness benefits of female extra-pair reproduction? Lifetime reproductive success of within-pair and extra-pair offspring. *The American Naturalist* **179**, 779-793.

Schlicht E, Kempenaers B (2013) Effects of social and extra-pair mating on sexual selection in blue tits (Cyanistes caeruleus). *Evolution* **67**, 1420-1434.

Schumacher J, Rosenkranz D, Herlyn H (2014) Mating systems and protein–protein interactions determine evolutionary rates of primate sperm proteins. *Proceedings of the Royal Society B-Biological Sciences* **281**.

Schwanhausser B, Busse D, Li N, *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337-342.

Searcy WA, Yasukawa K (1995) *Polygyny and Sexual Selection in Red-winged Blackbirds.* Princeton University Press, Princeton, N.J.

Searcy WA, Yasukawa K, Lanyon S (1999) Evolution of polygyny in the ancestors of red-winged blackbirds. *Auk* **116**, 5-19.

Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605-610.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Sharma R, Agarwal A, Mohanty G*, et al.* (2013) Functional proteomic analysis of seminal plasma proteins in men with various semen parameters. *Reproductive Biology and Endocrinology* **11**, 38.

Sheldon BC, Merila J, Qvarnstrom A, Gustafsson L, Ellegren H (1997) Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character. *Proceedings of the Royal Society B-Biological Sciences* **264**, 297-302.

Soulé M (1980) Thresholds for survival: maintaining fitness and evolutionary potential. In: *Conservation Biology: An Evolutionary-Ecological Perspective* (eds. Soulé M, Wilcox B), pp. 151-169. Sinauer, Sunderland, MD.

Souza CEA, Rego JPA, Lobo CH*, et al.* (2012) Proteomic analysis of the reproductive tract fluids from tropically-adapted Santa Ines rams. *Journal of Proteomics* **75**, 4436-4456.

Spottiswoode C, Møller AP (2004) Extrapair paternity, migration, and breeding synchrony in birds. *Behavioral Ecology* **15**, 41-57.

Spurgin LG (2013) Comment on Gohli et al. (2013): “Does promiscuity explain differences in levels of genetic diversity across passerine birds?”. *Evolution* **67**, 3071-3072.

Stapleton MK, Kleven O, Lifjeld JT, Robertson RJ (2007) Female tree swallows (Tachycineta bicolor) increase offspring heterozygosity through extrapair mating. *Behavioral Ecology and Sociobiology* **61**, 1725-1733.

Steen H, Mann M (2004) The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology* **5**, 699-711.

Stephens PA, Sutherland WJ (2000) Vertebrate mating systems, Allee effects and conservation. In: *Vertebrate Mating Systems* (eds. Apollonio M, Festa-Bianchet M, Mainardi D), pp. 186-213. World Scientific Publishing, London.

Stepinska U, Bakst MR (2007) Fertilization. In: *Reproductive Biology and Phylogeny of Birds* (ed. Jamieson BGM). Science Publishers, Enfield, NH.

Stewart SG, Bausek N, Wohlrab F, *et al.* (2004) Species specificity in avian sperm:perivitelline interaction. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **137**, 657-663.

Stutchbury BJ, Morton ES (2001) *Behavioral Ecology of Tropical Birds.* Academic Press, San Diego ; London.

Sulonen A-M, Ellonen P, Almusa H, *et al.* (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biology* **12**, R94.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7375-7379.

Swanson WJ, Nielsen R, Yang QF (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution* **20**, 18-20.

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137-144.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources* **8**, 299-301.

Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Traill LW, Bradshaw CJA, Brook BW (2007) Minimum viable population size: A meta-analysis of 30 years of published estimates. *Biological Conservation* **139**, 159-166.

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: Invited review. *Molecular Ecology* **9**, 1013-1027.

Trivers RL (1972) Parental investment and sexual selection. In: *Sexual Selection and the Descent of Man, 1871-1971* (ed. Campbell B), pp. 136-179. Aldine, Chicago, IL.

Turner LM, Chuong EB, Hoekstra HE (2008) Comparative analysis of testis protein evolution in rodents. *Genetics* **179**, 2075-2089.

USFWS (1976) Determination of endangered status for yellow-shouldered blackbird and determination of critical habitat, pp. 51019-51022. Federal Register.

USFWS (1996) Yellow-shouldered Blackbird (Agelaius xanthomus) Revised Recovery Plan, p. 77. U.S. Fish and Wildlife Service, Atlanta, GA.

USFWS (2011) Mariquita or yellow-shouldered blackbird (Agelaius xanthomus) 5-Year Review: Summary and Evaluation, p. 34. U.S. Fish and Wildlife Service, Boquerón, PR.

Utleg AG, Yi EC, Xie T, *et al.* (2003) Proteomic analysis of human prostasomes. *The Prostate* **56**, 150-161.

Valera F, Hoi H, Kristin A (2003) Male shrikes punish unfaithful females. *Behavioral Ecology* **14**, 403-408.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**, 227-232.

Walters JR, Harrison RG (2010) Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in Heliconius butterflies. *Molecular Biology and Evolution* **27**, 2000-2013.

Walters JR, Harrison RG (2011) Decoupling of rapid and adaptive evolution among seminal fluid proteins in Heliconius butterflies with divergent mating systems. *Evolution* **65**, 2855-2871.

Wang JL (2009) A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology* **18**, 2148-2164.

Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Molecular Ecology Resources* **10**, 785-796.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources* **8**, 753-756.

Weatherhead PJ (1997) Breeding synchrony and extra-pair mating in red-winged blackbirds. *Behavioral Ecology and Sociobiology* **40**, 151-158.

Weatherhead PJ, Boag PT (1995) Pair and extra-pair mating success relative to male quality in red-winged blackbirds. *Behavioral Ecology and Sociobiology* **37**, 81-91.

Weatherhead PJ, Montgomerie R, Gibbs HL, Boag PT (1994) The cost of extra-pair fertilizations to female red-winged blackbirds. *Proceedings of the Royal Society B-Biological Sciences* **258**, 315-320.

Weber JN, Peterson BK, Hoekstra HE (2013) Discrete genetic modules are responsible for complex burrow evolution in Peromyscus mice. *Nature* **493**, 402-405.

Webster MS, Pruett-Jones S, Westneat DF, Arnold SJ (1995) Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. *Evolution* **49**, 1147-1157.

Weibel AC, Moore WS (2002) A test of a mitochondrial gene-based phylogeny of woodpeckers (genus Picoides) using an independent nuclear gene, beta-fibrinogen intron 7. *Molecular Phylogenetics and Evolution* **22**, 247-257.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321-2328.

Westneat DF (1992) Do female red-winged blackbirds engage in a mixed mating strategy? *Ethology* **92**, 7-28.

Westneat DF (1993) Polygyny and extrapair fertilizations in eastern red-winged blackbirds (Agelaius phoeniceus). *Behavioral Ecology* **4**, 49-60.

Westneat DF, Mays HL (2005) Tests of spatial and temporal factors influencing extra-pair paternity in red-winged blackbirds. *Molecular Ecology* **14**, 2155-2167.

Westneat DF, McGraw LA, Fraterrigo JM, Birkhead TR, Fletcher F (1998) Patterns of courtship behavior and ejaculate characteristics in male red-winged blackbirds. *Behavioral Ecology and Sociobiology* **43**, 161-171.

Westneat DF, Rambo TB (2000) Copulation exposes female Red-winged Blackbirds to bacteria in male semen. *Journal of Avian Biology* **31**, 1-7.

Westneat DF, Sherman PW (1997) Density and extra-pair fertilizations in birds: a comparative analysis. *Behavioral Ecology and Sociobiology* **41**, 205-215.

Westneat DF, Stewart IRK (2003) Extra-pair paternity in birds: Causes, correlates, and conflict. *Annual Review of Ecology Evolution and Systematics* **34**, 365-396.

Whittingham LA, Kirkconnel A, Ratcliffe LM (1992) Differences in song and sexual dimorphism between Cuban and North American red-winged blackbirds (Agelaius phoeniceus). *Auk* **109**, 928-933.

Whittingham LA, Kirkconnell A, Ratcliffe LM (1996) Breeding behavior, social organization and morphology of red-shouldered (Agelaius assimilis) and tawny-shouldered (A-humeralis) blackbirds. *Condor* **98**, 832-836.

Wiley JW, Post W, Cruz A (1991) Conservation of the yellow-shouldered blackbird Agelaius xanthomus, an endangered West Indian species. *Biological Conservation* **55**, 119-138.

Wilk T, Cichon M, Wolff K (2008) Lack of evidence for improved immune response of extra-pair nestlings in collared flycatcher Ficedula albicollis. *Journal of Avian Biology* **39**, 546-552.

Wolfner MF (2002) The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. *Heredity* **88**, 85-93.

Wolfner MF (2009) Battle and ballet: molecular interactions between the sexes in Drosophila. *Journal of Heredity* **100**, 399-410.

Wong A (2010) Testing the effects of mating system variation on rates of molecular evolution in primates. *Evolution* **64**, 2779-2785.

Wong WSW, Yang ZH, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041-1051.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 0097-0159.

Wright S (1938) Size of population and breeding structure in relation to evolution. *Science* **87**, 430-431.

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nature Review Genetics* **13**, 329-342.

Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586-1591.

Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* **15**, 496-503.

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**, 1107-1118.

Yasukawa K, Searcy WA (1995) Red-winged Blackbird (Agelaius phoeniceus) In: *The Birds of North America Online* (ed. Poole A). Cornell Lab of Ornithology, Ithaca.

Zhdanova OL, Pudovkin AI (2008) Nb_HetEx: a program to estimate the effective number of breeders. *Journal of Heredity* **99**, 694-695.

Zwarts L, Magwire MM, Carbone MA, *et al.* (2011) Complex genetic architecture of Drosophila aggressive behavior. *Proceedings of the National Academy of Sciences* **108**, 17070-17075.

Biography

Irene A. Liu was born on 18 May 1984 in Fargo, North Dakota, and graduated with a B.S. and B.A. in Biology, *summa cum laude,* from the University of Maryland in 2006. During her Ph.D. she has earned numerous fellowships and awards, including the James B. Duke Fellowship, a National Science Foundation (NSF) Graduate Research Fellowship, a Sigma Delta Epsilon Fellowship from Graduate Women in Science, the 2013-2014 Katherine Stern Fellowship from Duke University, and the 2014 Dean's Award for Excellence in Mentoring. Irene has also received several grants in support of her laboratory and field research, including a 2011 NSF Doctoral Dissertation Research Grant, three student research grants from Sigma Xi, and two travel grants from the Duke Center for Latin-American and Caribbean Studies. Irene was invited on separate occasions to present her research on mating systems and population genetics of *Agelaius* blackbirds by the U.S. Fish and Wildlife Service in Puerto Rico, the University of Pennsylvania, Archbold Biological Station, and the Bahamas National Trust. She has also presented her research at international scientific meetings, including the North American Ornithological Conference and the Animal Behavior Society. Throughout her time at Duke, Irene was active in Women in Science and Engineering, the University Scholars Program, the Basketball Committee (part of Duke's Graduate and Professional Student Council), and the Biology Department's Behavioral, Population, and Community Ecology seminar series. For her circus trick, the final graduation

requirement for students in the Nowicki lab, she performed a 30-minute voice and piano recital chronicling her journey through graduate school.

Selected Publications

Olsen BJ, Greenberg R, **Liu IA**, Felch JM, Walters JR (2010) Interactions between sexual and natural selection on the evolution of a plumage badge. *Evolutionary Ecology* **24**, 731-748.

Liu IA., Lohr B, Olsen BJ, Greenberg R (2008) Macrogeographic vocal variation in subspecies of swamp sparrow. *Condor* **110**, 102-109.