Comparative and Population Genomics of Host-Pathogen Co-Evolution in Birds

Comparative and population genomics of host-pathogen co-evolution in birds

A dissertation presented

by

Allison Jane Shultz

to

The Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biology

Harvard University
Cambridge, Massachusetts

April, 2017

Dissertation Advisor: Professor Scott V. Edwards Allison Jane Shultz

Comparative and population genomics of host-pathogen co-evolution in birds

Abstract

Infectious disease produces some of the strongest selective forces on natural populations across the tree of life. The signatures of pathogen-mediated evolution on host genomes have been described for several traditional model organisms, but few studies of more diverse organisms have detected such signatures beyond candidate immune loci. In my dissertation, I combine population and comparative genomics to study the dynamics of pathogen-mediated selection at two evolutionary timescales in wild bird populations.

In Chapters 1 and 2, I integrate geographic and temporal sampling to compare the impacts on genomic variability of recent founder events and an epizootic on the House Finch (*Haeomorhous mexicanus*). House Finches, native to the western US and Mexico, were introduced to the Hawaiian Islands in 1870 and separately to the eastern US in 1940. In 1994, *Mycoplasma gallisepticum* (MG), previously associated with poultry, jumped to House Finches and caused severe declines. In Chapter 1, I genotype ~2,000 loci in samples collected before and 10 years after the epizootic. I demonstrate that demographic shifts brought about by the founder events have much more profound genomic impacts than pathogen-mediated natural selection. In Chapter 2, I use whole-genome resequencing data in six House Finch populations from three time points. I show that the expansion of the western US population following the last glacial maximum had the largest impact on

patterns of genomic variation, followed by the eastern founder event. This genome-wide dataset also enables me to detect a decrease in genetic diversity immediately following the epizootic. Additionally, the lack of temporal differentiation at individual SNPs suggests that House Finches evolved resistance through polygenic selection.

As a complement to these population-level analyses, in Chapter 3, I use over 11,000 well-annotated genes from 39 bird genomes to I show that immune genes encoding proteins that interact directly with pathogens consistently show evidence of positive selection. Using a dataset comprised of all genes regardless of function, I also show that immune system functional pathways are enriched for signatures of positive selection. Taken together, I demonstrate that host-pathogen interactions play an important role in shaping the bird genome over long evolutionary timescales.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my committee for all of the hard wok and their assistance with my dissertation and beyond. Dr. Edwards, thank you for your expertise, guidance, and advice whenever I needed it in regards to my research, the field in general, or my plans for the future. Dr. Wakeley, thank you for your expertise, input, and willingness to serve as PI for my NSF DDIG. Dr. Hoekstra, thank you for your input and advice.

Special thanks to my funding sources. This research was funded in part by the National Science Foundation both for research support (DEB-IOS 0923088, DEB-1501554) and fellowship support (GRFP 2008074713). The Harvard Museum of Comparative Zoology Putnam Fund, American Museum of Natural History, American Ornithologists' Union, Society for the Study of Evolution, and Graduate Women in Science provided funding for both field work and sequencing. Thank you to the Harvard University for supporting me while conducting research.

This project would not have been possible without the curators, collections managers, and museums that sent tissue loans for use in the project. At the Royal Ontario Museum, Alan Baker gave me permission to use the tissues. At the Burke Museum of Natural History, John Klicka gave me permission to use the tissues, and Sharon Birks subsampled and sent me the tissues. At the Field Museum of Natural History, John Bates gave me permission to use the tissues, and Ben Marks subsampled and sent me the tissues. At the Museum of Comparative Zoology, Scott Edwards gave me permission to use the tissues, and Breda Zimkus subsampled the tissues. At the Museum of Vertebrate Zoology, Rauri Bowie gave me permission to use the tissues, and Carla Cicero sent me the tissues. In

the MCZ bird collection, Jeremiah Trimble and Kate Eldridge always made sure I had what I needed, including equipment, access to specimens, and permits.

For assistance catching House Finches in Arizona I would like to thank Kevin McGraw and Pierce Hutton. In Washington, Rob Faucett helped get me the equipment I needed and send me my samples, and Lise Cicconi provided much needed assistance and a field site. In Massachusetts, Lily Lu, Simon Sin and Rabie Angadi helped me catch birds, and Scott Edwards and Michael Reed provided field sites. In California, Carla Cicero and the prep lab team saved me House Finches to prepare and helped send me tissue samples.

For help with the RADseq library preparation protocol, I thank Emily Kay and Brant Peterson for suggestions, and Flavia Termigioni for working with me to figure out the protocol. The Bauer Core staff members Christian Daly, Claire Reardon, Patrick Dennett, and Jennifer Couget were essential for training and consultation for library preparation and sequencing, and Simon Sin was a useful resource for answering lab questions. Tim Sackton provided useful advice on which analyses to run, and the FAS informatics team kept the Odyssey cluster up and running with the software that I needed. Anders Albrechtsen and Thorfinn Korneliussen helped me get their software, ANGSD up and running with my dataset.

Thank you to the OEB Department and Harvard staff, for help in administrative matters. I especially want to thank Alison Pirie, Nikki Hughes and Kirstin Pennarun, who always answered my questions and helped me get any materials or grants I needed, and without who my life would have been much more difficult.

To my fellow OEB graduate students, thank you for your support, input, and friendship. You truly made this project an enjoyable experience, and have helped me think about many different systems and questions. Mark Liu, Frank Rheindt, Clemens Kupper, Matt Fujita, Cassie Stoddard, Christie Riehl, Patrick Albers, Moos Blom, Elisa Dierickx, Cynthia Ursino, Maude Baldwin, Shane Campbell-Stanton, Simon Sin, Phil Grayson, Kathrin Napflin, Katie Stryjewski, Jonathan Schmitt, Tim Sackton, Gustavo Bravo, Sangeet Lamickhaney, Alison Cloutier, Pengchang Wang, John Burley, and Betul Kacar, I could not have asked for more intelligent or better lab mates. I will never stop cherishing your friendship.

To all of my friends and family, thank you for your support and patience. You truly made my thesis possible. To all of my parents, Amy and Mike Sewell, and Robert and Ann Shultz, I would not have been where I am without you. You have always supported me in everything I have done and supported me however I needed you to. To my sisters, Carlee and Melissa, you have always helped brighten my life, and have provided distraction when I most needed it. To all of my friends in Boston and beyond, thank you for your support and advice when I needed it. Finally, Rabie, thank you for all of your love and support you have given me through this process. I could not have done everything I needed without you.

INTRODUCTION

Pathogens are ubiquitous across the tree of life, and can have profound impacts on individual fitness by increasing host mortality and decreasing reproductive output (Haldane 1949). Infectious disease selects for hosts that can resist or tolerate the pathogens, pressuring pathogens to evolve ways to subvert host defenses. This host-pathogen arms race is a classic example of the "Red Queen" dynamic (Van Valen 1973; Salathe et al. 2008). The strength of infectious disease as a selective agent has been highlighted by recent studies of both immune gene candidates and genome-wide scans for selection. Across populations of among the best-studied organisms, humans and Drosophila, pathogens have the largest genome-wide signatures of selection (Fumagalli et al. 2011; Fumagalli and Sironi 2014; Early et al. 2017). However, the generality of these results across natural populations is currently unknown. Organisms may experience long co-evolutionary relationships with communities of pathogens, but are occasionally exposed to new threats as infectious diseases expand to new geographic areas (Daszak et al. 2000). These novel pathogens can have devastating effects on new hosts, causing severe population reductions or even extinctions (Dobson and Foufopoulos 2001). For example, the global amphibian declines in response to pathogenic chytrid fungus (Skerratt et al. 2007), the near-extinction of Tasmanian devils due to a transmissible cancer (Epstein et al. 2016), or the severe declines or extinctions of Hawaiian honeycreepers due to avian malaria (Foster et al. 2007). While there is some evidence that epizootic or epidemic events are increasing in frequency due to anthropogenic disturbances (Fey et al. 2015), they also may represent natural dynamics of host-pathogen arms races as even existing pathogens acquire new mutations to drastically increase virulence. These epizootic events provide the opportunity to study pathogen-mediated selection in real time.

Most of what we know about pathogen-mediated selection in non-model organisms comes from candidate-gene studies, with the polymorphic major histocompatibility complex (MHC) gene family as the focus for many of these studies. MHC is an important component of the vertebrate adaptive immune system for recognizing pathogens (Edwards and Hedrick 1998). Studies of MHC have shown evidence of positive selection or balancing selection in many species (e.g. Edwards and Hedrick 1998; Edwards et al. 2000; Westerdahl et al. 2005; Bonneaud et al. 2006; Fraser et al. 2010; Savage and Zamudio 2011; Eizaguirre et al. 2012). Beyond MHC genes, toll-like receptors (TLRs) have recently been a focus of studies of pathogen-mediated selection in non-model organisms. TLRs are a class of receptor molecules in the innate immune system that recognize pathogens (Kumar et al. 2011). There are approximately 10 different types of TLRs that respond to particular types of pathogens (Kumar et al. 2011; Chen et al. 2013). Most TLRs show evidence of positive selection across birds and mammals (Wlasiuk and Nachman 2010; Alcaide and Edwards 2011), and specific TLR alleles can be associated with pathogen resistance (Love et al. 2010; Netea et al. 2012; Tschirren et al. 2013).

MHC loci and TLRs are important components of the immune system and common targets of pathogen-mediated selection, but other genes may also play an important role in host-pathogen co-evolution (Fumagalli et al. 2011). In humans, 14 functional gene pathways, either directly or indirectly involved in immune responses, show evidence of positive selection (Daub et al. 2013). These signatures of positive selection observed at the pathway-level are often made up of smaller signatures at different genes, implying that pathogen-mediated selection may be polygenic (Daub et al. 2013). Similarly, both across mammal species and in human populations, Enard et al. (2016) recently found that proteins that interact with viruses, whether or not they are considered “immune” proteins, experience twice as much adaptation on average compared to proteins that have no documented interaction with viruses. These studies suggest that an unbiased, genome-wide approach may be important to comprehensively identify signatures of pathogen-mediated selection.

For my dissertation I am using birds as a system to study pathogen-mediated selection at different time scales. The first part of the dissertation will focus on a single bird species, the House Finch (*Haemorhous* mexicanus), to disentangle genomic signatures of demography from pathogen-mediated selection. House Finches are a North American finch (Family Fringillidae) well-known in studies of behavioral ecology (Hill 2002). House Finches, native to the western US and Mexico, were introduced to the Hawaiian Islands in 1870 and the eastern US in 1940 (Hill 2002). The host-pathogen relationship between the House Finch and a bacterium, *Mycoplasma gallisepticum* (MG) is one of the best-documented emerging epizootics (Dhondt et al. 1998). This poultry-associated bacterium was first documented in the House Finch in 1994 in the Washington DC area (Ley et al. 1996; Hochachka and Dhondt 2000). MG infects the respiratory tract and causes severe conjunctivitis (Hochachka and Dhondt 2000), suppresses pathogen-specific components of the immune system (Bonneaud et al. 2011) , and stimulates inflammatory responses (Gaunson et al. 2006; Mohammed et al. 2007; Adelman et al. 2013). The pathogen spread through an introduced eastern population rapidly, and by 1998 had caused severe declines across the region, as high as 60% in some areas (Dhondt et al. 1998). Infection experiments comparing gene expression responses of eastern individuals with 12 years of exposure and historically unexposed individuals suggested rapid evolution of gene expression, disease resistance (Bonneaud et al. 2011; 2012a) and disease tolerance (Adelman et al. 2013).

In Chapter 1, I broadly characterize the population-genomic impacts of House Finch introductions to eastern North America and Hawaii and the epizootic event. I use samples collected before the epizootic across the House Finch range and after the epizootic in the eastern population to separate out genomic signatures of the introduction and epizootic. I find substantial reductions of genetic diversity in the introduced eastern and Hawaii populations, population differentiation, increased linkage disequilibrium, and many allele frequency shifts, all expectations of a founder event. However, I do not find any signals of reduced genetic diversity, elevated linkage disequilibrium, or outlier loci as a result of the epizootic. In Chapter 2, I use whole genome resequencing to more fully characterize the genomic impacts of the eastern introduction and epizootic in both native and introduced populations. I extend the sampling from chapter 1 to include individuals sampled both 10 and 20 years after the epizootic in the eastern population, and to examine individuals before and 10 years after the epizootic in the western population. I find that the population expansion of the native western US population following the last glacial maximum had the largest genome-wide impact on patterns of genetic variation, followed by the eastern founder event. I also show a small decrease in genetic diversity following the MG epizootic, but no signatures of large selective sweeps or temporal population structuring. Together, these results suggest that House Finches evolved resistance through polygenic selection.

The second part of the dissertation will take a comparative genomics approach to focus on patterns of pathogen-mediated selection across all birds. Birds (Class Aves) are highly vagile animals with a worldwide distribution (Jetz et al. 2012). Many species undergo yearly migrations from hundreds to thousands of kilometers and even circumvent the globe (Gill 2007). They exhibit extensive phenotypic variation and life history strategies (Gill 2007) and are models for studies of ecology and evolution. In Chapter 3, I identify signatures of positive selection at specific sites in protein-coding gene sequences across bird lineages. I show that immune genes encoding proteins that interact directly with pathogens, genes *a priori* classified as "receptor" genes in the immune gene literature, are consistently under positive selection. Positive selection in these genes occurs at a higher proportion of sites and in a larger proportion of lineages compared to estimates from a dataset using all genes regardless of function. Using an approach including all genes (and pathways) pathways regardless of function, I also show that immune system functional pathways are enriched for signatures of positive selection. Taken together I demonstrate that host-pathogen interactions play an important role in shaping the bird genome over long evolutionary timescales.

Host-pathogen co-evolution is pervasive throughout the tree of life, but we are only now beginning to understand its genome-wide impacts. In this dissertation, I use a combination of population and comparative genomics to identify signatures of pathogen-mediated selection at two different time scales. By integrating these perspectives, we can better understand the impact of pathogen-mediated selection on the genome.

CHAPTER 1: SNPS ACROSS TIME AND SPACE: POPULATION GENOMIC SIGNATURES OF FOUNDER EVENTS AND EPIZOOTICS IN THE HOUSE FINCH (*HAEMORHOUS MEXICANUS*)

Co-authors: Allan J. Baker, Geoffrey E. Hill, Paul M. Nolan, Scott V. Edwards

INTRODUCTION

Expansions of organisms into novel ranges or habitats are ubiquitous across the tree of life. All species experience range expansions and contractions through time, but human activities are accelerating the pace of range shifts through alteration of habitats and direct movement of species (Hulme 2009). Introduced populations can have evolutionary impacts on native species in these new locations (Mooney and Cleland 2001), but the introductions also have evolutionary consequences for the introduced populations themselves as they experience new environments and an altered demographic history (Baker and Stebbins 1965; Baker and Moeed 1987; Dlugosch and Parker 2008). Introduced populations often experience founder effects and novel selection regimes (Dlugosch and Parker 2008), and if the bottlenecks resulting from founder events are sufficiently long and severe, reduced genetic diversity and heterozygosity can be sustained even after subsequent population expansion (Nei et al. 1975). Although colonization events often begin with bottlenecks, they are frequently followed by rapid population expansions, a situation that can ameliorate long-term reductions in genetic diversity and their detrimental effects (Dlugosch and Parker 2008). On the one hand, these expansions can facilitate rapid morphological or physiological change (Reznick and Ghalambor 2001) and the accumulation of deleterious mutations due to enhanced genetic drift on the edge of an expansion (Peischl et al. 2013). Although fluctuations in population sizes are characteristic of invasive species, species in native ranges can also experience such changes in demography, especially when they are impacted by environmental alteration, habitat degradation and fragmentation, and climate change (Wilcove et al. 1998; Moran and Alexander 2014).

In addition to novel demographic shifts, introduced species can also encounter novel selection regimes in newly colonized habitats. Infectious diseases constitute one of the strongest selective pressures encountered in novel habitats and can have profound impacts on a species by increasing host mortality and decreasing reproductive output (Haldane 1949; Altizer et al. 2003; Karlsson et al. 2014). Modern global connectivity increases the rate at which organisms are exposed to novel pathogens to which they do not have previously-evolved resistance (Daszak et al. 2000). Emerging pathogens can have devastating effects on novel hosts, causing extinctions or severe population reductions (Dobson and Foufopoulos 2001). Understanding the evolutionary dynamics between novel pathogens and hosts is of paramount importance to the preservation of biodiversity (Altizer et al. 2003), especially in the context of a demographic history reflecting past bottlenecks and range expansions. Additionally, the synergistic effects of population introductions and encounters with novel pathogens have rarely been studied (Longo et al. 2014).

The House Finch (*Haemorhous mexicanus*), a common bird in both urban and rural environments in North America, has become a model for the study of adaptation to novel environments following introductions and for host-pathogen co-evolution (Badyaev et al. 2002; Bonneaud et al. 2011). The native range is confined to the western United States and

Mexico, whereas established populations in the eastern United States and Hawaii are the result of human-mediated introductions in the 1940s and 1870s respectively (Hill 2002). Both populations were thought to have been introduced from a small number of founders (Grinnell 1911; Elliott and Arbib 1953), but underwent rapid population expansions and are abundant in their respective ranges (Hill 2002), suggesting classic examples of bottlenecks followed by exponential population growth. Even though the Hawaiian and eastern US populations were recently derived, both populations exhibit morphological and behavioral differences from the founding populations (Aldrich and Weske 1978; Vazquez-Phillips 1992; Able and Belthoff 1998; Badyaev and Hill 2000; Badyaev et al. 2002; Egbert and Belthoff 2003), and genetic divergence has been detected with both mitochondrial DNA and multilocus datasets (Benner 1991; Vazquez-Phillips 1992; Wang et al. 2003; Hawley et al. 2006; 2008).

Despite extensive study, the nature and extent of genetic change in the House Finch as a result of human-mediated introductions has been unclear, and only recently has it been suggested that different components of the House Finch genome may have responded to introductions in different ways (Vazquez-Phillips 1992; Wang et al. 2003; Hawley et al. 2006; Hess et al. 2007; Hawley et al. 2008, 2012; Backström et al. 2013; Zhang et al. 2014b). A situation complicating the analysis of genetic diversity in the House Finch has been temporal evolution of populations as a result of a novel pathogen, *Mycoplasma gallisepticum* (MG), which precipitated an epizootic event that is now well documented by demographic and genetic studies as well as statistical models of host-pathogen co-evolution (Dhondt et al. 2006; Staley and Bonneaud 2015). After its first encounter with House Finches in the mid-1990s in the mid-Atlantic region, MG rapidly spread through the introduced eastern population, causing severe declines across the region as high as 60% in some areas (Dhondt et al. 1998; Nolan et al. 1998). MG reached native populations in the west in 2002, where it spread more slowly and with a lower prevalence (Dhondt et al. 2006). Both experimental and gene expression studies have revealed mounting evidence for genetic evolution in House Finches as a result of the MG epizootic (Wang et al. 2006; Bonneaud et al. 2011; 2012; Adelman et al. 2013), with the potential for reductions in genetic diversity as a result of the epizootic itself. Some studies surveying genetic diversity have been able to directly analyze House Finch populations sampled prior to the epizootic and thereby isolate human-mediated introductions as a factor contributing to reductions in genetic diversity (Benner 1991; Vazquez-Phillips 1992; Wang et al. 2003; Hess et al. 2007; Hawley and Fleischer 2012). However, other studies with the aim of measuring changes in genetic diversity due to human-mediated introductions sampled populations after the epizootic, with the possibility that any reductions found may mistakenly be attributed to the introductions when in fact the consequences of the epizootic may have been at play (Hawley et al. 2006; 2008; Backström et al. 2013; Zhang et al. 2014b).

In the present study we compare the consequences for genetic diversity of both introductions and the epizootic by directly comparing geographically and temporally sampled populations of the House Finch. Demographic events, such as recent bottlenecks, confound the ability to study the genetic consequences of recent selection events (Thornton et al. 2007; Domingues et al. 2012). Genetic drift resulting from a bottleneck and selection event can each exhibit signatures of increased linkage disequilibrium and a reduction in effective population size, although demographic events typically have global effects on the genome whereas the effects of selection events are genomically more local

(Nielsen 2005). By sampling the same populations before and after the MG epizootic (Figure 1.1A), we have the rare opportunity to disentangle the signatures of drift and selection. Here we use double-digest restriction site associated sequencing (ddRADseq; Peterson et al. 2012) to genotype thousands of markers across the House Finch genome and quantify the evolutionary history across all the chromosomes. RADseq and its variants have proved useful in phylogeographic studies as well as studies searching for F_{ST} outliers and other signatures of natural selection in populations with complex recent histories (Hohenlohe et al. 2010; 2012; Ruegg et al. 2014, Edwards et al. 2016, Andrews et al. 2016). With our combination of temporal sampling and genome-scale genotyping we provide a comprehensive picture of the population genetic history of this emerging model system.

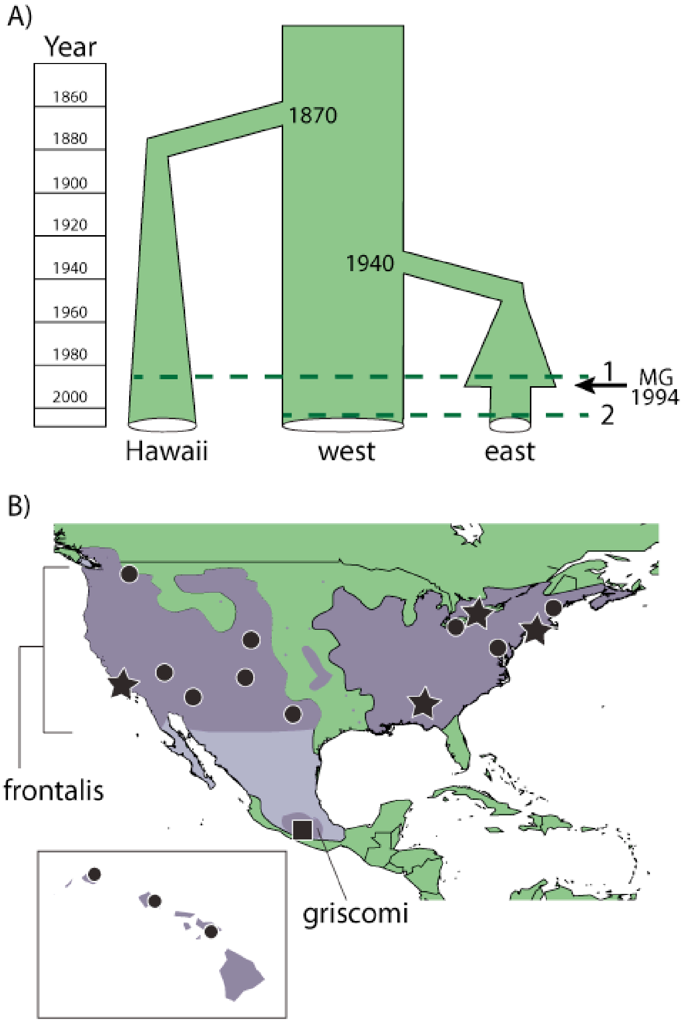


Figure 1.1. Approximate demographic history associated with population introductions in Hawaii and the eastern US, and the MG epizootic event in the eastern US. Line 1 on the Figure 1.indicates the approximate sampling time for the pre-epizootic samples, and line 2 indicates the approximate sampling time for the post-epizootic samples. B) Map of the House Finch sampling localities, including Hawaii, and the range circa 1990 (Ridgely et al. and Birdlife International 2011; NatureServe 2011). All *frontalis* subpopulations (circles and stars) were sampled before the MG epizootic (line 1 on panel A), and the subpopulations indicated by a star were sampled again in 2001 or 2003, approximately eight generations after the MG epizootic (line 2 on panel A). The *griscomi* subspecies was also sampled (square). The ranges of the *frontalis* and *griscomi* subspecies are highlighted.

MATERIALS AND METHODS

Sampling

To quantify the effects of the introductions on genetic diversity, we examined individuals collected before reports of the MG epizootic in each region. Summaries of the sampling can be found in Table 1.1 and the collection number, collection dates, and specific localities of each specimen can be found in Supplemental Table 1.1. Following a strategy similar to that of Wang et al. (2003), we sampled 16 pre-epizootic populations of the *frontalis* subspecies of House Finch in North America and the Hawaiian Islands (circles and stars, Figure 1.1B; Table 1.1), for a total of 90 individuals. To examine the effects of the epizootic, we analyzed samples from three of the eastern subpopulations in 2001 or 2003, approximately eight generations after the MG epizootic (n = 18; stars, Figure 1.1B; Table 1.1). We also analyzed samples from one western subpopulation from California in 2003, before the epizootic in this region but across the same time span as in the eastern US, allowing us to control for population genetic differences occurring in House Finches during this time span but not due to the epizootic (throughout the study we will refer to pre- and post- eastern and western populations as Pre-E, Post-E, Pre-W and Post-W respectively). To identify derived alleles within the *frontalis* subspecies we also sampled individuals from the *griscomi* (n = 6) subspecies (square, Figure 1.1B; Table 1.1) and two closely related sister species as outgroups, the Purple Finch (*H. purpurus*; n = 3) and the Cassin's Finch (*H. cassinii*; n = 3) (Smith et al. 2013; Table 1.1). Blood samples were preserved in Queen's lysis buffer (Seutin et al. 1991) and stored at -80°C. Tissue samples were frozen in liquid nitrogen following collection, and stored at -80°C, and transferred between laboratories in 100% ethanol at room temperature.

Table 1.1. Population sample information.

Regional		**Pre or Post-**	**Tissue**	**# Samples**	**# Samples**
Population	**Locality**	**epizootic**	**Type**	**Sequenced**	**Analyzed**
Western	Arizona (AZ)	Pre	tissue	6	6
	California (CA)	Pre	tissue	6	6
		Post	blood	6	6
	Colorado (CO)	Pre	tissue	6	6
	New Mexico (NM)	Pre	tissue	6	6
	Nevada (NV)	Pre	tissue	6	6
	Texas (TX)	Pre	tissue	6	6
	Washington (WA)	Pre	tissue	6	5
	western totals:	**Pre**		**42**	**41**
		Post		**6**	**6**
Eastern	Alabama (AL)	Pre	blood	6	6
		Post	blood	6	6
	Maine (ME)	Pre	tissue	6	6
	New York (NY)	Pre	tissue	6	6
		Post	blood	6	4
	Ohio (OH)	Pre	tissue	6	6
	Ontario (ON)	Pre	tissue	6	6
		Post	blood	6	5
	eastern totals:	**Pre**		**30**	**30**
		Post		**18**	**15**
Hawaiian	Kauai (HK)	*NA*	tissue	6	3
	Maui (HM)	*NA*	tissue	6	3
	Oahu (HO)	*NA*	tissue	6	6
	Hawaiian totals:	***NA***		**18**	**12**
H. m. frontalis totals:		**Pre**		**90**	**83**
		Post		**24**	**21**
Outgroup	Guerrero, Mexico (*griscomi*) (GU)	*NA*	tissue	6	5
	Haemorhous cassinii (CC)	*NA*	tissue	3	2
	Haemorhous purpeurus (CP)	*NA*	tissue	3	2

RAD Sequencing

We extracted whole genomic DNA using the DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA), using the standard blood and tissue protocols as appropriate. Each individual was barcoded and genotyped following a ddRADseq (Peterson et al. 2012) using the SphI-EcoR1 enzyme combination and isolating fragments in the range of 345-407 base pairs (bp). Details of the protocol can be found in the supplemental methods. We sequenced our library at the Bauer Core Facility of the FAS Center for Systems Biology at Harvard University (Cambridge, MA), using one lane of an Illumina HiSeq 2500 Rapid Run flow cell with 150 base pair paired-end sequencing.

Computational Analysis and Bioinformatics

Sequence data were demultiplexed using Geneious version 6.0.5 (Kearse et al. 2012) allowing for a single mismatch in the barcode. We then trimmed the four base pair restriction sites and used the *process_radtags.pl* program from STACKS version 0.99994 (Catchen et al. 2011; 2013) to filter low quality reads (on average, ~6% of reads were discarded per individual due to low quality scores). We employed a *de novo* approach to build a catalog of loci and call SNPs. Of the 126 individuals sequenced, we removed 13 that had fewer than 100,000 reads from downstream analyses (See Table 1.1 for final numbers of individuals analyzed per subpopulation). Preliminary analyses confirmed that these individuals had very low coverage and contributed very few loci to downstream analyses.

De novo assembly

We used STACKS version 1.21 to create a *de novo* catalog of loci and call SNPs (Catchen et al. 2011; 2013). We merged all reads into a single file for *de novo* locus identification and trimmed them to a length of 140 base pairs. The optimal set of parameters for library assembly varies according to study system (Catchen et al. 2011; 2013; Mastretta-Yanes et al. 2014), so with the pre-epizootic set of individuals, we tested a range of parameters with the *denovo_map.pl* pipeline for catalog construction. By creating a catalog of loci with both reads simultaneously, we were able to leverage information about which reads were derived from the same DNA fragments, which in turn allowed us to validate and test parameter performance. Briefly, we tested values of –M and –n from 1-8 and –m with 4, 10 and 20. We assessed performance based on the number of loci in the final dataset, the percentage of loci from the same DNA fragment that had more than a one-to-one match within an individual, and the percentage of loci that had more than two alleles within an individual. Final parameters used for downstream analysis were –M 4, -n 4, and –m 4.Details on parameter testing and novel python scripts are available in the supplemental material.

With the *populations* program in STACKS, we generated three different datasets for subsequent analyses. The first two datasets included all individuals (including both pre-epizootic and post-epizootic time periods), but differed in SNP filtering. Dataset 1 used a less conservative SNP filter, requiring an individual minimum locus depth of 10 to be included (-m 10), and a locus to only be included if it was present in at least 75% of individuals in half of the populations (-r 0.75 and –p 11). The second, more conservative filter (dataset 2) required an individual minimum locus depth of 30 (-m 30), and had the same inclusion parameters. Analyses of dataset 2 produced results very similar to those of dataset 1, although sometimes less resolved due to the smaller number of loci, so we only present the results of dataset 1 throughout the rest of the paper. Individuals from all time periods were used to build these STACKS catalogs. We also ran some preliminary analyses on datasets with more stringent completeness filters, but found that analyzing these data sets had very little effect on overall results. The final dataset (dataset 3) focused on maximizing high quality SNPs from the temporal samples for analyses seeking to identify MG-mediated selection. This dataset only included individuals from the diachronically sampled subpopulations (CA, AL, NY, and ON), but included both time periods. The quality filter required an individual minimum locus depth of 30 (-m 30), and a locus was to be included if it was present in at least 75% of individuals in two of the samples: Pre-E, Post-E, Pre-W and Post-W (-r 0.75 –p 2).

We further refined each dataset by removing possible problematic loci that matched any of the following three criteria: the locus contained a SNP with observed heterozygosity greater than 0.75; the locus contained a SNP that was not in Hardy-Weinberg equilibrium ($p < 0.05$) in at least two of the three populations (eastern, western, Hawaiian); or a locus mapped to the Zebra Finch (*Taeniopygia guttata*) genome version 3.2.4 (Warren et al. 2010) with BLASTN 2.2.29+ (Camacho et al. 2009) more than once with an e-value less than 10^{-40}. The Zebra Finch is the closest relative to the House Finch with a high quality reference genome. The House Finch and Zebra Finch lineages diverged approximately 50 million years ago (Brown et al. 2008), and given the conservatism of the avian genome (Ellegren 2013; Zhang et al. 2014a), the Zebra Finch genome has successfully been used to map RADseq reads from other similarly diverged bird species (Bourgeois et al. 2013). Finally, we sought to remove any related individuals from the dataset. We used the program KING (Manichaikul et al. 2010) to estimate relatedness among all individuals. We identified two individuals in the post-epizootic sample from Ontario that displayed a kinship value of 0.1534, indicative of a second-degree relationship, so we removed one of the individuals, indiv_121, from all downstream analyses. All datasets are available in the Dryad repository (http://dx.doi.org/10.5061/dryad.0h2g0).

Population Structure

To test for population structure among pre-epizootic subpopulations we used a Bayesian approach, implemented in the program STRUCTURE 2.3.4 (Pritchard et al. 2000), to determine the number of genetic groups or clusters (K) that best fit the data, and to assign individuals to these clusters. We first tested for structure among all pre-epizootic individuals, including outgroups (n = 92 individuals). Because variation among higher levels of population structure can mask substructuring (Evanno et al. 2005), we implemented a hierarchical set of analyses. We tested for structure within the *frontalis* subspecies (including the introduced populations; n = 83 individuals); within the native *frontalis* population ('western' population; n = 41 individuals); within the introduced eastern *frontalis* population ('eastern' population; n = 30); and within the introduced Hawaiian *frontalis* population ('Hawaiian' population; n = 12). To determine the optimal number of clusters (K), we considered the highest ΔK, as recommended by Evanno et al. (2005), but also considered the biological feasibility of the result. For all analyses, we used a SNP dataset that only contained a single SNP per locus, and we ran four replicates of each K value ranging from 1-8 with an admixture model, burn-in of 100,000, and 1,000,000 Markov chain Monte Carlo samples. We used STRUCTURE HARVESTER (Earl and vonHoldt 2011), CLUMPP v. 1.1.2 (Jakobsson and Rosenberg 2007), and DISTRUCT v. 1.1 (Rosenberg 2004) to visualize the combined results.

Population Genetic Analyses

We used the *populations* program in STACKS (Catchen et al. 2011; 2013) to calculate summary statistics within each subpopulation in both time periods when possible. We also calculated summary statistics for all pre-epizootic individuals grouped by population (western, eastern, Hawaiian). We further assessed population differentiation by using the *populations* program to calculate F_{ST} for each nucleotide present in each pair of subpopulations. We calculated Tajima's *D* for each subpopulation and assessed a significant difference from the neutral expectation using the beta distribution (Tajima 1989; python script available at https://github.com/ajshultz/Rad/).

We calculated linkage disequilibrium (LD; r^2; Hill and Robertson 1968) to compare levels of non-independence of SNPs among the Pre-E, Pre-W, Hawaiian and Post-E populations using a custom python script (*Pairwise_linkage_disequilibrium.py* available at https://github.com/ajshultz/Rad/). Because sample size can affect measures of LD, for each population we randomly chose eight individuals with less than 50% missing data. First, we calculated r^2 between all pairs of SNPs located on the same locus (between 1 and 139 base pairs apart). Second, we compared the mean r^2 value between SNPs located on read 1 and read 2 loci from the same DNA fragment (paired via *catalog_read_pair.py*, described in the supplemental methods). We compared these values to the equivocal number of randomly chosen pairs from the catalog of possible loci. By comparing levels of LD from loci on the same DNA fragment to levels from randomly chosen loci, we could assess whether the decrease in LD observed in the single locus analysis degraded to a level of LD lower than that observed in a single read. Additionally, by confirming that levels of LD were similar among populations with randomly chosen loci, we could ensure that any significant differences observed in loci in a single read we observed among populations were not an artifact of dataset structure (e.g. differences among populations in the amount of missing data).

Selection Scans

We used two methods to test for selection between the native and introduced populations, using Pre-W vs. Pre-E; Pre-W vs. Hawaiian with dataset 1; Pre-E vs. Post-E (combining AL, ON, and NY) samples with datasets 1 and 3; and Pre-W vs. Post-W (CA) samples with datasets 1 and 3. First, we used the *populations* program in STACKS to calculate allele frequency differences for each SNP found in both populations. We assessed the significance for these differences with a Fisher's exact test, corrected for multiple testing to a 5% false discovery rate using the Benjamini-Hochberg approach (Benjamini and Hochberg 1995). This approach for detecting allele-frequency shifts has advantages compared to using Fst outliers, which can be influenced by levels of within-population heterozygosity (Cruickshank and Hahn 2014). Second, we used BayeScan version 2.1 (Foll and Gaggiotti 2008) for each of the same population comparisons separately with a burn-in of 50,000 iterations and 100,000 generations of data collection. Because BayeScan can be sensitive to loci with low minor allele frequencies, we first filtered each dataset to include only loci with a minor allele frequency greater than 0.10.

We simulated datasets to generate expectations for the proportion of outlier loci under the neutral model for an introduction event. We used the program ms (Hudson 2002) to generate datasets that approximated historical records, and varied the size of the introduced founding population (Supplemental Figure 1.1; supplemental methods) to explore the effect the bottleneck size had on the proportion of outliers. We ran the model using a mutation rate estimated for Zebra Finch of 2.21 x 10^{-9} per site per year (Nam et al. 2010), and a generation time of 1 year. We simulated 1000 datasets for each founding N_e (20, 200, 2,000, 100,000, no bottleneck) twice, once with sampling modeled after the Pre-

W vs Pre-E comparisons (82 and 60 chromosomes respectively) and once with sampling modeled after the Pre-W vs Hawaiian comparisons (82 and 24 chromosomes respectively). For all simulations, we calculated the significance of allele frequency differences between populations with a Fisher's exact test, corrected for multiple testing as described above.

RESULTS

Sequencing and de novo library construction

We obtained a total of 40,151,299 paired-end 150 base pair reads that passed Illumina's quality filter for 126 individuals (mean 637,322 total reads per individual; Supplemental Table 1.1). The *de novo* assembly generated an average of 12,700 unique loci per individual, of which an average of 3,217 were polymorphic (Supplemental Table 1.1). Across the three datasets, when loci were mapped to the Zebra Finch genome with a minimum e-value of 10^{-40}, 12% of loci consistently mapped more than once and were subsequently dropped from the analysis. Approximately 65% of loci mapped just once to the Zebra Finch genome, and 24% did not map at all. Loci fell evenly across the entire genome (Supplemental Figure 1.2), and the number of loci per chromosome was significantly correlated with Zebra Finch chromosome length (dataset 1: R^2 = 94%, p <0.0001; Supplemental Figure 1.3). After filtering for multiple hits to the Zebra Finch genome and deviations from Hardy-Weinberg equilibrium, dataset 1 (with all individuals and a minimum locus depth of 10) contained 2,283 loci and 18,096 SNPs, was 73% complete, and had a mean depth per locus of ~60. Dataset 2 (all individuals and a minimum locus depth of 30) contained 889 loci and 6,877 SNPs, was 67% complete, and had a mean depth of ~68. Dataset 3 (only diachronically sampled populations and a minimum stack depth of 30) contained 2,150 loci and 8,561

SNPs, was 55% complete and had a mean depth of ~66. Across all datasets, there was a high amount of variability in the amount of data missing in any particular individual (Supplemental Figure 1.4), most likely caused by sensitivities of the ddRADseq protocol to tissue degradation and long-term storage.

Population Structure

When comparing all individuals across all three species sampled, the Evanno method (Evanno et al. 2005) identified two clusters as optimal, separating the outgroup species from the House Finch (Figure 1.2A; Supplemental Table 1.2). Among only the *frontalis* House Finch birds, K = 3 was optimal (Supplemental Table 1.2), separating out the native western, introduced Hawaiian, and introduced eastern populations (Figure 1.2B). Among the Hawaiian birds, K = 2 was optimal (Supplemental Table 1.2), which separated the Kauai individuals from those sampled on Oahu and Maui (Figure 1.2C). For both the eastern and western populations considered alone, K = 3 was optimal (Supplemental Table 1.2), but this result did not appear to contain any biologically useful information, with three clusters equally likely in all individuals (Figure 1.2C). The Evanno method cannot identify K = 1 as the optimal strategy (Evanno et al. 2005), but that is likely the true model in this situation.

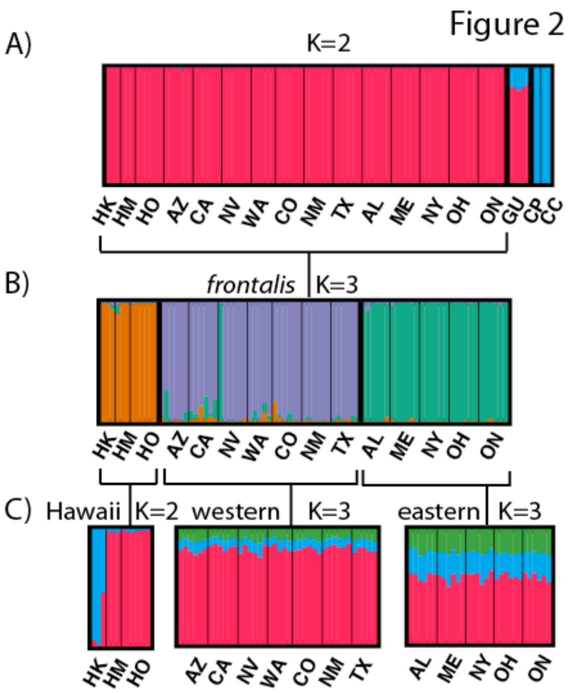


Figure 1.2. A) STRUCTURE plot with K = 2 for all pre-epizootic individuals using dataset 1. Abbreviations for populations indicated in STRUCTURE plots are: HK = Kauai, HM = Maui, HO = Oahu, AZ = Arizona, CA = California, NV = Nevada, WA = Washington, CO = Colorado, NM = New Mexico, TX = Texas, AL = Alabama, ME = Maine, NY = New York, OH = Ohio, ON = Ontario, GU = Guerrero, CP = Purple Finch, CC = Cassin's Finch. B) STRUCTURE plot with K = 3 for all *frontalis* individuals. C) STRUCTURE plot with the K = 2 for all Hawaiian individuals, K=3 for western individuals, and K = 3 for eastern individuals.

Population Genetics

Using all 18,096 SNPs in dataset 1, the Pre-W population had consistently higher levels of polymorphic sites, private sites, observed heterozygosity, and nucleotide diversity than the Pre-E population or Hawaiian population (Table 1.2). These results hold true when considering subpopulations individually as well. Reductions in nucleotide diversity (π) relative to the western region were more severe for Hawaiian birds (-14%) than for eastern birds (-7%; Figure 1.3), and were similar for haplotype diversity and heterozygosity (Table 1.2). Levels of diversity among Pre- and Post-E birds were similar. Although the small number of populations sampled in both time periods makes statistical comparison difficult, Post-E estimates were within two standard deviations of the Pre-E range for both π and observed heterozygosity (Table 1.2; Figure 1.3). We further confirmed that our results were qualitatively similar and not a by-product of sample size by examining the results with eight individuals chosen randomly from each population (Pre-E, Pre-W, Hawaiian; results not shown). Tajima's *D* was negative but non-significant for all populations, except for the post-epizootic New York population. We found similar results for the dataset with eight randomly chosen individuals per population to counter effects of different sample sizes (Pre-W=-0.34, Pre-E=-0.36, Post-E=-0.36, Hawaiian=-0.38; all $p>0.05$).

Average pairwise F_{ST} among pre-epizootic populations were consistent with the STRUCTURE analyses. For dataset 1, the species outgroups were the most divergent from the *frontalis* individuals (0.404-0.555, mean=0.455; Supplemental Figure 1.5), followed by the subspecies outgroup (0.128-0.213; mean=0.152). Populations within the eastern and western regions were the least differentiated (0.049-0.064, mean=0.056), whereas levels of differentiation between eastern and western subpopulations were slightly higher (0.061-0.070, mean=0.065). The Hawaiian population was the most differentiated from the other *frontalis* populations (0.071-0.102, mean=0.086); F_{ST} was even higher between Kauai and the other two islands within Hawaii (0.099-0.129).

The Pre-W population exhibited lower levels of LD than either the Pre-E or Hawaiian populations (Figure 1.4A; Pre-W vs. Pre-E Wilcoxon rank sum test $p < 0.0001$, Pre-W vs. Hawaiian Wilcoxon rank sum test $p < 0.0001$), but we found no difference in LD between Pre- and Post-E populations (Figure 1.4B; Wilcoxon rank sum test $p = 0.282$). There were 375 loci from read 1 and read 2 of the same DNA fragment as calculated by our novel python script (Supplemental methods). Levels of LD were higher for SNPs on pairs of loci from the same DNA fragment than for SNPs on randomly chosen pairs for all populations (Figure 1.4C; Wilcoxon rank sum tests: Pre-W, $p = 0.0119$, Hawaiian, $p < 0.0001$, Pre-E, $p < 0.0001$, Post-E, $p < 0.0001$). However, there were no differences among paired locus r^2 values from different populations (Wilcoxon rank sum test: Pre-W versus Pre-E, $p = 0.827$; Pre-W versus Hawaiian, $p = 0.361$; Pre-E versus Post-E, $p = 0.160$).

Table 1.2. Summary statistics for all populations calculated for dataset 1.

Regional Population	Locality	Pre or Post-epizootic	N	Private	Sites	Poly. Sites	% Poly. Loci	P	Observed Het.	Nucleotide Div. (π)	Hap. Div. (h)
Western	**Regional**	**Pre**	**34.42**	**3354**	**309312**	**11585**	**3.75**	**0.9965**	**0.0046**	**0.0054**	**0.78**
	Arizona (AZ)	Pre	5.83	551	300467	5401	1.80	0.9967	0.0047	0.0054	0.77
	California (CA)	Pre	5.60	286	242436	4024	1.66	0.9969	0.0044	0.0050	0.71
		Post	5.68	354	242420	4094	1.69	0.9968	0.0045	0.0051	0.72
	Colorado (CO)	Pre	5.61	351	236289	4063	1.72	0.9968	0.0046	0.0052	0.74
	New Mexico (NM)	Pre	5.67	251	157265	2728	1.73	0.9967	0.0047	0.0053	0.76
	Nevada (NV)	Pre	5.66	289	196967	3520	1.79	0.9966	0.0048	0.0055	0.80
	Texas (TX)	Pre	5.35	260	179373	2972	1.66	0.9968	0.0046	0.0052	0.73
	Washington (WA)	Pre	4.70	340	294586	4465	1.52	0.9969	0.0044	0.0050	0.73
Eastern	**Regional**	**Pre**	**26.11**	**384**	**309404**	**7591**	**2.45**	**0.9967**	**0.0045**	**0.0051**	**0.72**
	Alabama (AL)	Pre	5.80	79	299625	4636	1.55	0.9969	0.0044	0.0049	0.68
		Post	5.75	90	312905	4781	1.53	0.9969	0.0043	0.0049	0.68
	Maine (ME)	Pre	5.58	59	267323	3995	1.49	0.9969	0.0043	0.0048	0.70
	New York (NY)	Pre	5.48	44	239907	3614	1.51	0.9969	0.0046	0.0049	0.70
		Post	3.52	47	254607	3046	1.20	0.9971	0.0042	0.0047	0.68
	Ohio (OH)	Pre	5.67	35	190969	2955	1.55	0.9968	0.0046	0.0050	0.71
	Ontario (ON)	Pre	5.52	70	287609	4202	1.46	0.9970	0.0042	0.0047	0.67
		Post	4.80	61	284382	3959	1.39	0.9970	0.0043	0.0048	0.68
Hawaiian	**Regional**	**Pre**	**10.45**	**393**	**309429**	**5373**	**1.74**	**0.9969**	**0.0041**	**0.0047**	**0.67**
	Kauai (HK)	*NA*	3.00	103	287622	2654	0.92	0.9974	0.0039	0.0041	0.57
	Maui (HM)	*NA*	3.00	35	114494	1191	1.04	0.9971	0.0044	0.0047	0.70
	Oahu (HO)	*NA*	5.69	208	293342	3826	1.30	0.9972	0.0039	0.0044	0.62
Outgroup	Guerrero, Mexico (*griscomi*) (GU)	*NA*	4.48	959	272498	2815	1.03	0.9975	0.0035	0.0038	0.56
	Haemorhous cassinii (CC)	*NA*	2.00	1417	129397	1172	0.91	0.9971	0.0044	0.0049	0.71
	Haemorhous purpeurus (CP)	*NA*	2.00	1588	143094	1153	0.81	0.9973	0.0029	0.0045	0.65

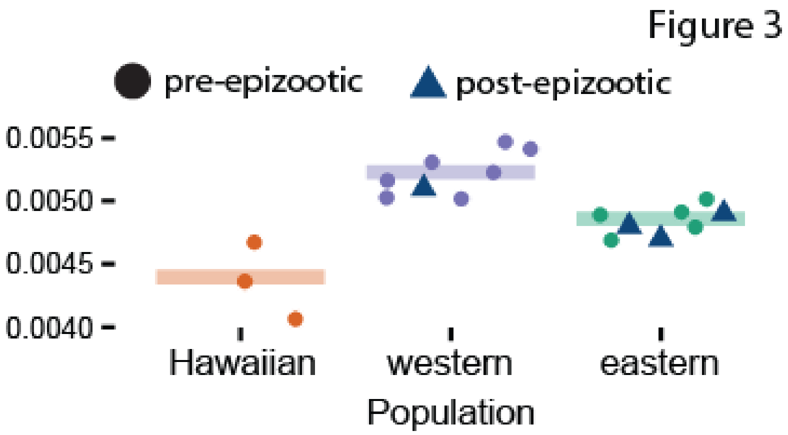


Figure 1.3. Estimates of diversity (π) for all subpopulations from dataset 1. The bars indicate the mean pre-epizootic π value. Reductions in nucleotide diversity (π) are more severe in the Hawaiian population than the eastern population, with 15.8% and 7.0% reductions in the estimated mean subpopulation π, respectively. Pre-epizootic estimates are circles and post-epizootic estimates are triangles.

Figure 4

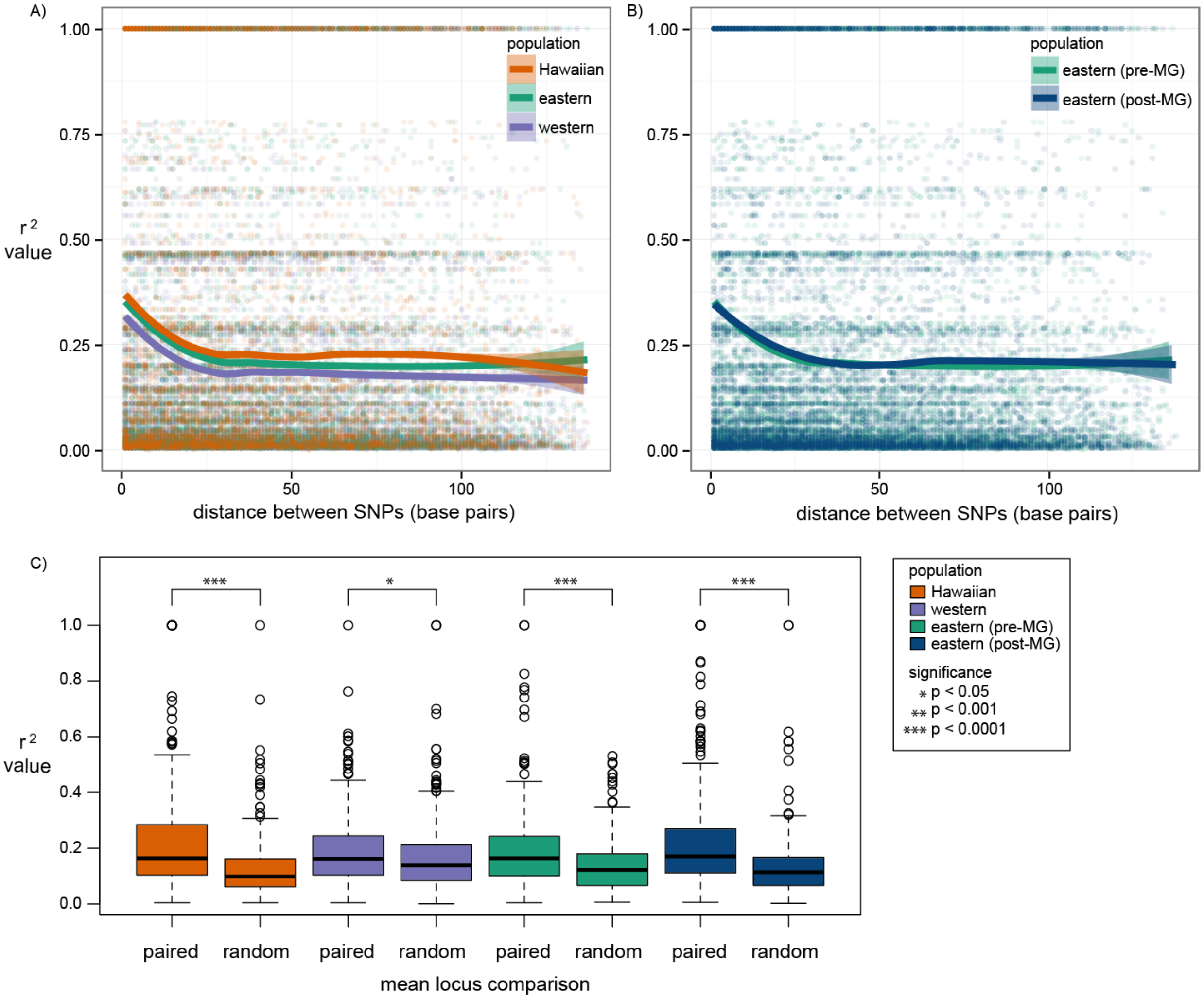


Figure 1.4. A) Pre-epizootic population linkage disequilibrium (r^2) calculations between pairs of SNPs located on the same RAD locus. From eight randomly chosen individuals with less than 50% missing data, the LOESS smoothed r^2 values (solid lines) and 95% SE confidence intervals (shaded areas) are shown for each SNP distance. B) LD (r^2) calculations between pairs of SNPs located on the same RAD locus. For eight Pre-E and Post-E individuals, the LOESS smoothed r^2 values (solid lines) and 95% SE confidence intervals (shaded areas) are shown for each SNP distance. C) Mean LD (r^2) between SNPs on read 1 and read 2 of the same DNA fragment (paired; 375 pairs of loci) compared to two randomly chosen loci (random; 375 random pairs)

Selection Scans

Using dataset 1, we detected 224 SNPs whose frequencies were significantly different between the Pre-E and Pre-W populations (out of 12,928 comparisons; 1.7% of SNPs) with a FDR of 5%. Of these, 136 could be mapped to the Zebra Finch genome (Figure 1.5A; Supplemental Table 1.3). We detected 125 SNPs whose frequencies were significantly different between the Hawaiian and Pre-W populations (out of 12,390 comparisons; 1.0% of SNPs) with a FDR of 5%. Of these, 84 could be mapped to the Zebra Finch genome (Figure 1.5B; Supplemental Table 1.4). Outlier loci from the Pre-E versus Pre-W comparison had a smaller range of F_{ST} values compared to Hawaiian and Pre-W comparisons (Figure 1.5C), with little overlap of loci identified as outliers in the two comparisons (Supplemental Figure 1.6). There were no SNPs significantly differentiated in the Pre- and Post-E comparisons, and only three SNPs were significantly differentiated in the Pre-W and Post-W comparisons, none of which mapped to the Zebra Finch genome.

BayeScan detected four outliers in the Pre-E/W comparison with a FDR of 5%, three of which could be mapped to the Zebra Finch genome (Locus 761, SNP 20; Locus 1261, SNP 43; Locus 1538, SNP 138; Supplemental Table 1.3). BayeScan detected four outliers from two loci in the Hawaiian vs. Pre-W comparison, of which one was mappable (Locus 5849, SNP 55; Supplemental Table 1.4). All loci identified by BayeScan were also detected in outlier analyses reported above.

Simulations of founding events under the neutral model had proportions of significant allele frequencies greater than or comparable to those found in our empirical data (Table 1.3). This holds true across a range of bottleneck sizes for both eastern and Hawaiian outlier loci.

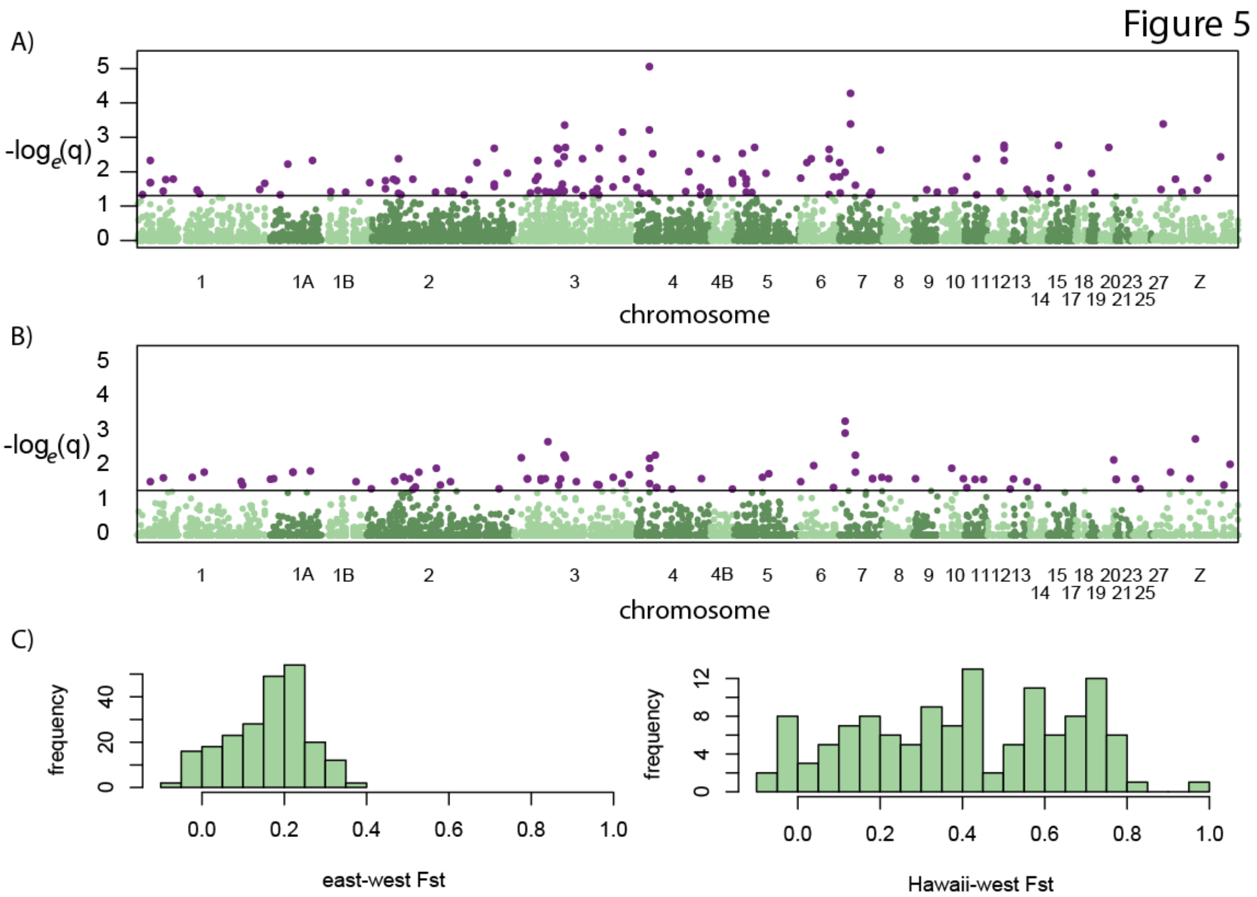


Figure 1.5. False-discovery-rate-corrected q-value from Fisher's exact test for different allele frequencies between A) Pre-E and Pre-W populations and B) Hawaiian and Pre-W populations plotted according to position on the Zebra Finch genome. A horizontal black line indicates the 5% significance threshold, and the SNPs with q-values that fall below this threshold used in downstream analyses are colored purple. C) Distribution of F_{ST} values observed in outlier loci identified in A) and B) for Pre-E and Pre-W and Hawaiian and Pre-W comparisons.

Table 1.3. Proportion of outlier loci found in founder event simulations under the neutral model. We report the proportion of loci found with significant allele frequency differences (Fisher's exact test p < 0.05 after multiple test correction) from simulations of the founder event with five different founding population sizes. We simulated data 1,000 times for each founding population size with samples sizes that matched both eastern-western comparisons and Hawaii-western comparisons. We report the mean of each set of replicates ± standard error, and the empirical proportion calculated from dataset 1.

Founder Ne	Eastern Sample Size	Hawaiian Sample Size
20	0.343±0.006	0.209±0.005
200	0.042±0.003	0.028±0.002
2,000	0.010±0.002	0.010±0.001
100,000	0.006±0.001	0.010±0.001
No Bottleneck	0.005±0.001	0.010±0.001
Empirical	**0.017**	**0.010**

DISCUSSION

Both natural selection and drift due to founder events can lead to genotypic changes in populations, and disentangling the relative contributions of these evolutionary forces can be difficult if one relies solely on contemporary samples from a single time period. Comparisons of the genotypes of populations across time and space enable better deductions of the evolutionary processes that underlie genomic change (Mathieson et al. 2015). To understand the relative importance to House Finch evolution of founder events and introduction to a novel habitat versus natural selection imposed by a novel pathogen, we employed ddRADseq to conduct a genome-wide survey of genetic variation in introduced and native populations of the House Finch both before and after an epizootic event. In addition to providing insight and clarification of conflicting results from previous studies on the genetic effects of introductions in the eastern United States and Hawaii, ours is an early study using genome-wide SNPs to quantify the effects of a major selection event with temporal sampling (see also Tin et al. 2015).

Population genetic signatures of introduced populations

Our study suggests that ddRADseq data has great power to detect even subtle changes in effective population size, and is well suited for population and conservation genomic studies even if a reference genome is not available (e.g. Backström et al. 2006; Stapley et al. 2008; Dierickx et al. 2015; Edwards et al. 2015). We demonstrate that despite the brevity of the bottlenecks prior to population expansions, introduced populations of House Finches in the eastern US and Hawaii have reduced genetic and haplotype diversity and heterozygosity. Although the number of individuals sampled in subpopulations was small compared to historical studies using one or a few loci, the large number of markers gives robust estimates of genetic diversity (Felsenstein 2006; Carling and Brumfield 2007; Mccormack et al. 2011; Therkildsen et al. 2013), and the concordance among estimates within geographic regions and subsamples (not shown) lends confidence to our results. Finally, despite biases that can be associated with ascertainment of ddRADseq data (Arnold et al. 2013), our estimates of genetic variability in the House Finch genome are similar to other species that survey non-coding genomic regions via next-generation approaches, including species with both higher and lower estimates (Supplemental Table 1.6). Although we corroborated several patterns of genetic differentiation found in previous work, including genetic differentiation of introduced House Finch populations, and no population differentiation within either western or eastern populations (Figure 1.2; Wang et al. 2003; Hawley et al. 2006; Hawley et al. 2008), for the first time, we show that genetic structure exists among birds from the Hawaiian Islands, with birds from Kauai, the most geographically isolated Hawaiian Island (Roderick and Gillespie 1998), showing differentiation from those from Oahu. One individual from the Nevada population was consistently assigned to the eastern population with a very high proportion of its genome (98%). This individual could be an example of a rare migrant from the eastern population to the western population, but without re-extracting and genotyping this individual we cannot rule out the possibility of a mislabeled or misidentified sample.

We also used LD to study the dynamics of introductions in the House Finch. We quantified r^2 both between SNPs within single 140-bp reads and between SNPs on loci from paired ends of ~300 base pair fragments. We confirm the rapid LD decay in very short segments of the genome observed in a few candidate loci sequenced by Backström et al. (2013), as well as in natural populations of other birds, at least for autosomes (Edwards and Dillon 2004; Balakrishnan and Edwards 2008; Kawakami et al. 2014). We also demonstrate elevated genome-wide levels of LD in introduced populations, a key prediction of bottleneck scenarios (Slatkin 2008). Balakrishnan and Edwards (2008) detected elevated LD in an island population of Zebra Finches, which was accompanied by a ten-fold decrease in nucleotide diversity in the island population. Our results suggest that elevated LD can persist even after a rapid expansion event and more modest decreases in genetic diversity.

Selection vs. drift during human-induced introductions

Regions of the genome exhibiting significant differentiation as a result of human introductions could be due either to adaptation to the novel environment or to the genetic effects of a bottleneck (Lee 2002; Thornton 2007; Poh et al. 2014). Despite the low LD and sparse genome sampling afforded by ddRADseq, we detected 224 SNPs with significant allele frequency differences between the pre-epizootic eastern and western populations and 125 SNPs with significant allele frequency differences between the Hawaiian and pre-epizootic western populations. Some estimates of allele divergence, such as F_{ST}, are influenced by underlying levels of polymorphism (Cruickshank and Hahn 2014). However, by focusing on loci with significant measures of allele differentiation via a Fisher's exact test, rather than by diversity-dependent measures such as F_{ST}, we avoid these biases. Indeed, the power to detect allelic differentiation in our data set appears substantial only in regions of the genome with adequate polymorphism; Supplemental Figure 1.7 suggests that only genomic regions with moderate levels of diversity yielded higher or outlier values of F_{ST}. Our simulations suggest that a large proportion of these shifts in allele frequency differences are likely a result of genetic drift during the bottleneck event (Thornton et al. 2007), or allele surfing during the population expansions (Excoffier and Ray 2008). Genetic drift can increase the variance in allele frequencies in small populations (Wright 1931, Nei and Tajima 1981). Our simulations of the founder events show that the proportions of outlier loci we detect in both introductions are comparable to neutral expectations (Table 1.3). The simulations do not account for some of the biases associated with RADseq data such as allelic dropout and missing data, but these differences are unlikely to change these conclusions. The smallest bottleneck size (N_e = 20) produced a larger proportion of allele frequency differences than our observed values, suggesting less extreme founding events in the House Finch.

Although a moderate bottleneck can explain most or all of the observed differences in allele frequency between native and introduced populations, some shifts may have been driven by selection in the novel environments experienced by the introduced populations.

The eastern population has smaller legs and feet than the western population (Aldrich and Weske 1978) and males and females of the eastern population exhibit heritable, sex-specific patterns of covariance among mensural characters (Badyaev and Hill 2000). The eastern population also exhibits significantly more short distance migration (Able and Belthoff 1998), and possibly as a consequence, has more pointed wings (Egbert and Belthoff 2003). The Hawaiian population has a smaller body size and greater morphological differentiation compared to the eastern population (Vazquez-Phillips 1992). These morphological and behavioral changes likely have at least a partial genetic basis and selection for these phenotypic traits may be responsible for some of the changes in allele frequencies. With our dataset we cannot compare an empirical distribution of outlier loci with loci known *a priori* to be evolving neutrally (Lotterhos and Whitlock 2014). But, BayeScan has been used to detect selection in populations that have undergone bottlenecks (e.g. Pilot et al. 2013). Of the 224 outliers we documented in the Pre-E population relative to Pre-W, and the 125 outliers in the Hawaiian population relative to Pre-W, we found enrichment for several gene ontology terms, but none with obvious implications for observed phenotypic differences (Supplement methods and results; Supplemental Table 1.5). Although BayeScan has low power if few populations are compared (De Mita et al. 2013), it detected 2-4 outliers in each comparison. Of the SNPs that could be mapped to the House Finch genome, all were in intergenic regions, and the closest genes had unknown functions (Supplemental Tables 3 and 4). As Domingues et al. (2012) demonstrated with a founder event 3,000 years ago in beach mice (*Peromyscus*), we find that signatures of selection are likely obscured in founder events on historical time scales due to genetic drift.

Signatures of selection as a result of the epizootic

Museum collections provide important historical snapshots and the opportunity to study changes in genetic diversity directly (e.g. Therkildsen et al. 2013; Tin et al. 2015). More importantly, these collections allow us to quantify the effects of anthropogenic change in wild organisms (Wandeler et al. 2007; Foster et al. 2007; Leonard 2008; Nielsen and Hansen 2008; Bi et al. 2013; Habel et al. 2013). Despite substantial decreases in census size after the MG epizootic (Nolan et al. 1998), we found no genome-wide signatures of a temporal bottleneck induced by the epizootic in eastern House Finches, suggesting that the effective population size remained sTable 1.despite substantial epizootic-driven decreases in census population sizes. Such a pattern is perhaps expected, given that there were still millions of individuals in the population, and genetic drift would likely have minimal effects. However, with such a large population size we expect some signatures of selection in this system between Pre- and Post-E birds. This expectation is further strengthened given the demonstrated increase in resistance (Bonneaud et al. 2011, 2012) in experimentally infected birds and decreases in bill length, tarsus length, and wing chord (Nolan et al. 1998) exhibited by post-epizootic as compared to pre-epizootic birds in the field. Our inability to detect significant temporal shifts in allele frequency in the eastern US may be expected given the low levels of LD in the population, the sparse sampling of the House Finch genome by ddRADseq (one locus approximately every 5 MB), and what was likely a mild selection event even in the face of substantial drops in census numbers. Even so, our Post-E sample size was higher than that for Hawaii, where we easily found strong evidence for allele frequency shifts among islands and relative to the Western US.

The values of Tajima's D in pre- and post-epizootic populations were similar and non-significant across space and time, suggesting that none of the populations departed significantly from a neutral model. The exception was for the post-epizootic New York population, which also had the smallest sample size possible for calculating Tajima's D (4 individuals). Small sample size has been shown to inflate values of Tajima's D (Subramanian 2016). Overall, these results suggest that there is little genome-wide deviation from a neutral model in House Finches.

To detect loci associated with the previously documented temporal and geographic differences in gene expression exhibited by House Finches in experimental infections with MG (Bonneaud et al. 2011), we likely need denser sampling of the House Finch genome (Tiffin and Ross-Ibarra 2014; Edwards et al. 2016; Andrews et al. 2016). More sensitive methods for detecting selection using both polymorphism and LD require extended tracts of sequence in genomic regions surrounding selected loci (Sabeti 2006; Barrett and Schluter 2008; Vitti et al. 2013). Studies seeking to quantify the effects of selection events on genomes in natural populations with large effective population sizes such as the House Finch likely need to employ whole genome resequencing to maximize sampling throughout the genome.

Conclusions

In summary, by using both temporal and geographic sampling of House Finch populations, we extend previous findings suggesting signatures of human-induced founder events on both the eastern US and Hawaiian populations of House Finch, with signatures of reduced heterozygosity, reduced genetic diversity, reduced haplotype diversity, and increased LD at very short genomic distances (within 140 base pairs). We detected loci with shifted allele frequencies as a consequence of the human-induced founder events and suggest that these shifts are likely more often caused by genetic drift than selection. Despite a favorable scenario for detecting pathogen-mediated signatures of selection in the eastern US with temporal sampling (no bottleneck, putatively strong selection, known phenotypic differences), RADseq was unable to detect genome-wide reductions in diversity or loci with significantly different allele frequencies before and after the epizootic, results likely driven by overall low levels of LD throughout the House Finch genome, as well as the low density of markers generated by ddRADseq. Overall, our study provides a rare direct comparison of temporal and spatial events within the same species and confirms a hypothesis that is rarely tested in side-by-side comparisons: that demographic shifts, such as bottlenecks or range expansions, may have more profound and genome-wide consequences for genomic variation than will selection imposed by a novel pathogen. Therefore, our study suggests that, despite conservation concern with selective events like epizootics, if populations maintain sufficient effective population sizes to mitigate the effects of genetic drift, there may be few genomic consequences of such events in nature.

Originally published as: Shultz, A. J., A. J. Baker, G. E. Hill, and P. M. Nolan. 2016. SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorhous mexicanus*). Ecology and Evolution 6:7475-7489.

CHAPTER 2: WHOLE GENOME SEQUENCING REVEALS RECENT SIGNATURES OF DEMOGRAPHY AND POLYGENIC SELECTION IN THE HOUSE FINCH

Co-authors: Niclas Backström, Qu Zhang, Geoffrey E. Hill, Paul M. Nolan, Scott V. Edwards

INTRODUCTION

The mandate of conservation biology is to preserve as much biodiversity as possible despite many anthropogenic disturbances (Young et al. 2016). These disturbances, including direct harvest, habitat loss and modification, invasive species, novel pathogens, pollution and climate change, are increasing in scope as the human footprint expands, particularly in high biodiversity areas (Venter et al. 2016; Young et al. 2016). To mitigate the effects of these disturbances, conservation biologists use genetic tools to understand the population structure, gene flow, and genetic diversity of a species (Ouborg et al. 2010). With this information they can identify the best ways to preserve as much biodiversity as possible with limited resources. The use of genome-wide markers has been slow to integrate into a conservation biologist's toolkit due to costs, analytical demands, and complexity (Shafer et al. 2015). However, their use enables the inference of more detailed demographic histories, and can also potentially identify regions of the genome that might be important in adaptation (Ouborg et al. 2010; Shafer et al. 2015). This information can help direct management efforts. For example, it can be used to understand how to best maintain genetic diversity in small populations (e.g. Chen et al. 2016), to identify the gene pathways responsible for adaptation to pollution (e.g. Reid et al. 2016), or to infer whether or not populations are genetically distinguishable (e.g. Dierickx et al. 2015). Recently, there have been calls to go a step further, and use prescriptive evolution as a management technique, or to manage populations based on evolutionary processes (Smith et al. 2014).

Before it is possible to manage populations for evolutionary processes, it is essential to understand how these processes operate in natural populations. For example, we are only now beginning to have an understanding of how rapid adaptation occurs, and how current population-genetic models may be insufficient to correctly approximate this process (Messer et al. 2016). Even in two of the best-studied species, *Homo sapiens* and *Drosophila melanogaster*, only recently has the importance of polygenic selection (Daub et al. 2013) or soft sweeps on standing variation (Garud et al. 2015) in adaptation been appreciated. If these modes of selection are also prevalent as the mechanism of rapid adaptation in natural populations, as has been suggested by Messer and Petrov (2013), conservation management practices like prescriptive evolution may be difficult, if not impossible to implement.

Until recently, most researchers sought to infer evolutionary processes based on present-day patterns observed across geographic localities. However, recent studies incorporating temporal samples have shown that focusing only on genetic variation present in extant taxa may obscure some of the more complex population histories (Parks et al. 2014; Leonardi et al. 2017). Most of the recent focus has been on ancient DNA from organisms that lived thousands of years ago, but museum collections offer an ideal means to study the evolutionary processes operating in natural populations on historic time scales. Populations sampled over decades or hundreds of years would be of primary interest for understanding how anthropogenic disturbances have shaped the genetics of contemporary populations (Bi et al. 2013; Habel et al. 2013). Few studies have incorporated genome-scale data to compare populations before and after anthropogenic disturbances. However, a few have offered insights into how non-model organisms might respond to drought (Franks et al. 2016), disease (Tin et al. 2015; Shultz et al. 2016; Epstein et al. 2016), introductions (Shultz et al. 2016), direct harvest (Therkildsen et al. 2013), or climate change (Bi et al. 2013). However, these studies use either reduced representation sequencing (e.g. RADseq or sequence capture), or pooled whole-genome sequencing, limiting the inferences about how different regions of a species' genome responds to anthropogenic changes (Edwards et al. 2015).

A North American songbird called the House Finch (*Haemorhous mexicanus*, family Fringillidae) provides an opportunity to quantify the genomic impacts of anthropogenic changes. Two major demographic events have impacted House Finches over the last century. The first was an introduction into a novel environment in 1940 (Elliott and Arbib 1953; Hill 2002), and the second was an epizootic associated with a novel pathogen, *Mycoplasma gallisepticum* (MG). This epizootic began in 1994 in the introduced population (Dhondt et al. 1998) and in 2002 in the native population (Dhondt et al. 2006). A set of museum samples available across the range of the House Finch, both before and after the epizootic event, enables us to study the impact of these events from both a geographic and temporal perspective. Previous work using reduced-representation double-digest RAD sequencing (ddRADseq) on populations of House Finches collected in both introduced and native populations before the epizootic, and after the epizootic in the introduced population, identified that the bottleneck associated with the introduction had larger impacts on the entire genome than the selection event (Shultz et al. 2016). Comparing allele frequencies across space and time, Shultz et al. (2016) did not identify any specific SNPs associated with the evolution of MG resistance or tolerance, and could not distinguish between SNPs with different frequencies due to the bottleneck or due to selection associated with a new environment in the introduced population. However, the lack of a House Finch reference genome, together with the sparse sampling of the genome (approximately 1 locus ever 5 MB) and low levels of linkage disequilibrium (<150 base pairs), limited the ability of Shultz et al. (2016) to detect selection. Furthermore, almost all variants were from intergenic regions, prohibiting the comparison of genetic diversity across sites from different functional classes of the genome.

In the present study, we sequenced a draft genome of the House Finch and use whole-genome resequencing to fully quantify the effects of the introduction and epizootic on the entire genome. We extend the sampling of Shultz et al. (2016) to include samples collected before and ten years after the epizootic in both the introduced and native population, and add an additional time point (20 years after the epizootic) in the introduced population. With our whole-genome data, we seek to identify fine-scale population structure and differences in genetic diversity, examine the impacts of these demographic events on different functional classes of the genome, quantify changes in autosomal and sex chromosome (Z) genetic diversity, and identify any SNPs that might have changed allele frequencies through space or time.

METHODS

Genome Sequencing, Assembly and Annotation

To construct a House Finch reference genome we sequenced a single male individual collected from Arizona in 2007, a population historically unexposed to MG. We prepared two independent 220 bp insert size fragment libraries using the PrepX ILM 8 DNA library preparation kit on the Apollo 324 NGS Library Prep system (Wafergen, Fremont, CA) following manufacturer's protocols. The two duplicate libraries were pooled in equimolar amounts and sequenced on two lanes of an Illumina 2500 Rapid Run flowcell at the Bauer Core Facility at Harvard University (Cambridge, MA). To construct larger scaffolds we also sequenced one mate-pair library with a 3kb insert size, prepared at the National Center for Genome Resources (Santa Fe, New Mexico) and sequenced on two lanes of an Illumina 2000 flow cell at the Bauer Core Facility at Harvard University. We used AllPaths-LG (Gnerre et al. 2011) to assemble the reads and generate a *de novo* draft genome.

We annotated the genome using MAKER v.2.31.8 (Cantarel et al. 2007). We used previously published House Finch spleen transcriptome data (Zhang et al. 2014c) as EST evidence, data from 14 vertebrate species, including mouse, human, zebra fish, anolis lizard, and 10 species of bird as protein evidence, and known transposable elements as TE evidence. We also used two gene finders – the SNAP gene prediction model trained on chicken (*Gallus gallus*) (Korf 2004) and chicken AUGUSTUS gene models provided with MAKER (Stanke et al. 2004).

To generate a consensus fasta file of the whole genome reads, we mapped the fragment libraries back to the House Finch genome using BWA v0.7.10 (Li and Durbin 2009) and used the resulting aligned sequences with Samtools v0.1.9 mpileup (Li et al. 2009) to call SNPs. We excluded regions with less than 1/3 of the mean coverage, or more than two times the mean coverage. We estimated the demographic history of the native

House Finch population with pairwise sequential Markovian coalescent analysis (PSMC, Li and Durbin 2011) with the parameters –p “4+25*2+2+6”. To convert coalescent time to years, we assumed a House Finch generation time of 1.5 years (Hochachka and Dhondt 2000) and used the Zebra Finch mutation rate of 2.21 x 10^{-9} mutations per site per year (Nam et al. 2010).

Genome Resequencing Sampling and Library Preparation

We selected 216 House Finches collected across eastern and western localities at three time points – 1990, 2001, and 2014 from museum collections or wild-caught individuals sampled for this study (Table 2.1). Specific localities and collection dates are listed in Supplementary Table 2.1. We chose to sample eastern populations in 1990, 2001 and 2014 to have samples collected before the epizootic, approximately 10 generations after the epizootic, and approximately 20 generations after the epizootic. We sampled western populations in 2001, before the western epizootic and 2014, approximately 10 generations after the western epizootic.

Samples consisted of muscle tissues preserved at -80°C but transported between laboratories in ethanol, or blood samples preserved in Queen’s lysis buffer (Seutin et al. 1991). We extracted whole genomic DNA with the EZNA Tissue DNA Kit (Omega Bio-tek, Norcross, GA) using the manufacturer’s protocol. We sheared DNA at a 100 ng/uL concentration using sonication with a 10% duty factor, 175 peak incident power, 200 cycles per burst and 50 seconds to obtain 300 base pair fragments (Covaris, Inc, Woburn, MA), and prepared 320 base pair libraries with the PrepX 32i ILM DNA library preparation kit on the Apollo 324 NGS Library Preparation system (Wafergen, Fremont, CA). Each library was amplified with five cycles of PCR according the protocol and reagents provided with the PrepX kit. Individuals received one of 48 unique six base pair barcodes and were pooled in equimolar amounts in groups of either 24 or 48 individuals. Each pool was quantified with the KAPA Library Quantification Kit for Illumina (Kapa Biosystems, Wilmington, MA), and was sequenced on two (24 individual pooled library) or four (48 individual pooled library) lanes of an Illumina 2500v4 high output flowcell at the Bauer Core Sequencing Facility at Harvard University (Cambridge, MA).

To compare levels of observed variation and to infer ancestral versus derived alleles, we also sequenced several outgroups. House Finches form a clade with the sister species, Purple Finch (*Haemorhous purpureus*) and Cassin's Finch (*Haemorhous cassinii*) (Smith et al. 2013), so we prepared sequencing libraries for eight individuals collected in Washington from each of these species according to the same methodologies described above. These 16 individuals were individually barcoded and sequenced on two lanes of an Illumina 2500v4 high output flowcell. We also sequenced three more divergent finch species (Fringillidae; Zuccon et al. 2012): the Long-tailed Rosefinch (*Uragus sibricus*), the Common Rosefinch (*Carpodacus erythrinus*), and the Pine Grosbeak (*Pinicola enucleator*). We collected the Long-tailed Rosefinch and Common Rosefinch specimens in Mongolia in the summer of 2012. Libraries were prepared according to the previously described methods and sequenced on three-fifths of two lanes of an Illumina 2500v4 high output flowcell.

Data Pre-processing and Read Mapping

We marked Illumina adapters with Picard v.2.7.1 (http://broadinstitute.github.io/picard) and mapped reads from each replicate lane of sequencing for each sample to the draft House Finch genome using BWA-mem v0.7.9 (Li and Durbin 2009) using default parameters. We then used Picard to merge the results from each lane of sequencing, mark both PCR and optical duplicates for each sample, and sort, index, and validate each sample's BAM file.

We used GATK v3.6 (McKenna et al. 2010) to realign House Finch, Purple Finch, and Cassin's Finch alignments around indels. First, we used all individuals to create an indel realignment target file with the "RealignerTargetCreator" program. Next, we used the resulting interval file to realign each individual BAM file with the "IndelRealigner" program.

Data Filtering and Variant Identification

Due to our low-coverage (4-6x) sequence data, we did not call variants for each sample, but instead used genotype likelihood scores for all population-genetic analyses. We generated genotype likelihoods using the program ANGSD v.0.911 (Korneliussen et al. 2014) under the SAMtools (-GL 1) model (Li et al. 2009). All analyses were conducted with filters to include only reads with unique hits, reads with a minimum mapping quality score of 20, and reads not marked as "bad" due to adapter contamination or duplication by the pre-processing pipeline.

We also excluded sites based on coverage (Singhal et al. 2015). We calculated coverage for each individual at each site in the genome using the "doCounts" program in ANGSD. We excluded all sites without sufficient coverage to be useful in downstream analyses (<0.5x mean coverage across all individuals), or with coverage greater than 1.7

times the mean coverage of all individuals sequenced to prevent the inclusion of repeat regions or paralogous genes that may have been misassembled in the draft genome. In total, we removed 6,389,321 sites as a result of these coverage filters.

To prevent any biases in results due to related individuals, we used NGSrelate (Korneliussen and Moltke 2015) to estimate pair-wise coancestry coefficients within each geographic locality and time period based on the genotype likelihood scores calculated from ANGSD. We identified seven instances where related individuals were sampled in our dataset - three individuals from the 2014 Massachusetts population (MA_14_11 and MA_14_15: 0.170; MA_14_11 and MA_14_16: 0.175; MA_14_15 and MA_14_16: 0.158), two individuals from the 2001 Maine population (ME_01_02 and ME_01_12: 0.093), two individuals from the 1990 Maine population (ME_90_12 and ME_90_13: 0.0275), two individuals from the 1990 New York population (NY_90_04 and NY_90_12: 0.081), and two individuals from the 1990 Ohio population (OH_90_02 and OH_90_15: 0.063). We removed one individual from each pair, MA_14_11, MA_14_15, ME_01_02, ME_90_12, NY_90_04, and OH_90_02 from subsequent analyses.

Identification of Z Chromosomes and Individual Sexing

Because we sequenced a male individual for our draft genome, we could not identify scaffolds from the W chromosome. This is because in birds males are the homogametic sex with a ZZ genotype and females are the heterogametic sex with a ZW genotype. We used expected coverage differences for reads mapped to the Z chromosome in males and females to identify Z chromosome scaffolds. Of our 216 House Finches, we could use morphological or phenotypic information to reliably identify the sex for 19 adult females and 45 adult males that were sexed by gonads during specimen preparation (Supplemental Table 2.1). The remaining birds were either juveniles, or had no recorded sex. We calculated average coverage for each scaffold for each individual using the coverage data as described above. From the individuals with reliable sex information, we randomly chose 10 females and 10 males, and for each scaffold conducted a t-test to test for significantly different coverage between sexes. We repeated this procedure 50 times to exclude false positives or negatives, and generated a list of scaffolds with significantly different coverage between sexes (Supplemental Figure 2.1). We identified 278 scaffolds significantly different in at least 45 of the permutations, with a female to male coverage ratio less than 0.70. The total length of these scaffolds is 73.22 MB, similar to other bird Z chromosomes (chicken: 82.31MB, Zebra Finch: 72.86 MB, Great Tit: 74.51MB).

We sexed all House Finches based on the mean coverage at these 278 scaffolds. After standardizing scaffold coverage for each individual, we performed hierarchical clustering. Individuals separated into two distinct clusters, each of which included either all males of known sex, or all females of known sex (Supplemental Table 2.1), with proportions of males and females varying across collection localities and time points (Table 2.1).

Table 2.1: Numbers of samples and breakdowns of males and females collected for each population.

region	epizootic status	year	locality	N	males	females
eastern	pre-epizootic	~1990	New York (NY)	16	11	5
			Maine (ME)	16	8	8
			Ohio (OH)	16	13	3
			Alabama (AU)	8	5	3
	post-epizootic	~2001	New York (NY)	16	14	2
			Maine (ME)	16	12	4
			Ohio (OH)	16	11	5
			Alabama (AU)	8	6	2
	post-epizootic	~2014	Massachusetts (MA)	16	11	5
			Illinois (IL)	16	8	8
			Alabama (AU)	8	4	4
western	pre-epizootic	~2001	Washington (WA)	16	12	4
			California (CA)	16	11	5
	post-epizootic	~2014	Washington (WA)	16	9	7
			California (CA)	16	13	3
outgroups			Cassin's Finch (CC)	8	-	-
			Purple Finch (CP)	8	-	-
			Common Rosefinch	1	-	-
			Pine Grosbeak	1	-	-
			Long-tailed Rosefinch	1	-	-
totals:						
eastern	pre-epizootic	1990		56	37	19
	pre-epizootic	2001		56	43	13
	post-epizootic	2014		40	23	17
western	pre-epizootic	2001		32	23	9
	post-epizootic	2014		32	22	10
all house finches				**216**	**148**	**68**
all outgroups				**19**		
total individuals				**235**		

Population Structure

We tested for population structure among both geographic and temporal populations using principle components analysis (PCA). To ensure that we identified fine-scale structuring, we used a hierarchical sampling approach. We started off using all House Finch, Cassin's Finch, and Purple Finch individuals (n = 226). With this analysis, we found that one House Finch from the California 2014 population was misidentified (CA_14_05), and clustered with the Purple Finches. We removed this individual from all subsequent analyses and reran the first PCA (n=225). We then ran analyses on all House Finches (n = 209), and regional populations [eastern House Finches (n = 146), western House Finches (n = 63)] to identify geographic or temporal structuring. Finally, we sought to identify any fine temporal structuring by analyzing every population with more than one time period sampled [California (n = 31), Washington (n = 32), Auburn (n = 24), Maine (n = 30), New York (n = 31) and Ohio (n = 31)]. All analyses were run using genotype probabilities for sites likely to be polymorphic in the populations ($p < 1e^{-6}$) with a minimum minor allele frequency of 0.05. In order to perform the PCA, we first calculated the covariance matrix between individuals with the "ngsCovar" program in NGSTools v.0.615 (Fumagalli et al. 2014), and calculated the eigenvalues and eigenvectors of the covariance matrix with the 'eigen' function in R (R Core Development Team 2008).

Site Frequency Spectra

We estimated the site frequency spectrum (SFS) for each locality and temporal population. First, we use the "doSAF" program in ANGSD (Nielsen et al. 2012; Korneliussen et al. 2014) to estimate the allele frequency for each site. To polarize sites in the unfolded site frequency spectrum, we used the sequence of the Long-tailed Finch as an ancestral sequence. We then used the "realSFS" program in ANGSD (Nielsen et al. 2012; Korneliussen et al. 2014) to compute the maximum likelihood global estimate across all sites. To compare SFS among geographic localities and years, we used dadi (Gutenkunst et al. 2009) to project 32 or 30 chromosome samples down to 28 chromosomes, the smallest sample size excluding Auburn populations. Auburn populations were excluded from these analyses due to small sample sizes. We conducted all tests for differences in the SFS using the non-parametric Kolmogorov-Smirnov test (K-S test). We first compared the SFS distributions among regions (eastern and western) and time periods. There were no significant differences among localities within each region and time period ($p > 0.05$), so we averaged SFS to compare SFS between each region and time period, and the neutral expectation SFS.

A SFS is required by ANGSD to condition the estimation of genetic diversity or neutrality statistics (Korneliussen et al. 2013). To understand how the sex ratio during the introduction may have altered the ratio of genetic diversity at autosomal sites versus sex chromosome sites, or among functional classes of sites, we computed the SFS for autosomal and Z-chromosome sites separately. For the Z-chromosome sites, we restricted our analyses to males only, because ANGSD assumes diploidy among individuals. Due to the small samples sizes in the Alabama populations, we excluded them from analyses of the Z chromosome. To understand how different functional classes of sites in the genome changed in response to the demographic and selection events, we computed the SFS for intergenic, intronic, and coding sequence sites within regional populations in each time period.

In addition to computing neutrality statistics, we directly compared SFS among functional classes of sites in the genome among regions and time periods. We projected each SFS down to 64 chromosomes, the smallest sample size (2001 western population). We first tested for differences among functional classes within each region and time period. Then, we tested for differences in functional classes among eastern time periods. Finally, we tested for differences among eastern 1990 functional classes, western 2001 functional classes, and functional classes and the neutral expectation.

Genetic Diversity and Population Genetic Statistics

We calculated two measurements of genetic diversity, pairwise nucleotide diversity (π) and Watterson's Θ (Θ_w) for each site in the genome with the "doThetas" program in ANGSD (Korneliussen et al. 2013). We conducted the analysis for each geographic locality and time period, conditioned on the SFS as calculated above. We then used the thetaStat program in ANGSD to calculate statistics of genetic diversity and neutrality (π, Θ_w, and Tajima's D) for one kilobase windows.

We compared autosomal values among populations as evidence for bottlenecks or expansions due to the introduction or epizootic, and compared the ratio of Z-chromosome π to autosomal π values to infer differences in the sex ratio at the time of introduction. Finally, we compared diversity statistics among functional genomic classes among regional populations.

Allele Frequency Change

To identify specific sites in the genome whose frequency was affected either by the introduction or the epizootic, we conducted association tests for pairwise population comparisons of allele frequency differences at each site in the genome (Kim et al. 2011). To test for allele frequency changes due to the introduction, we compared eastern and western allele frequencies in 1990 and 2001 respectively (pre-epizootic) and 2014 in both populations (post-epizootic). To test for allele frequency changes through time, we compared western populations in 2001 (pre-epizootic) and 2014 (post-epizootic), eastern populations in 1990 (pre-epizootic) and 2001 (post-epizootic), and eastern populations in 1990 (pre-epizootic) and 2014 (post-epizootic). We computed p-values for the likelihood ratio test statistic with the χ^2 distribution with one degree of freedom, and corrected p-values for multiple testing to a 5% false discovery rate with the Benjamini-Hochberg approach (Benjamini and Hochberg 1995).

RESULTS

Genome Sequencing, Assembly and Annotation

We sequenced 267,562,068 and 271,785,084 reads (150 base pairs) from each of the genome fragment libraries, of which 85.0% and 85.1% were used in the genome assembly respectively. Together, these libraries resulted in 69.5x average coverage across the genome. We obtained 541,713,586 reads (each 100 base pairs) from the 3 kb mate pair library, of which 13.5% was used in the genome assembly, resulting in 7.3x average coverage across the genome. The genome assembly consisted of 7,026 scaffolds, with a contig N50 of 71.0 kb, a scaffold N50 of 2.2 Mb, and estimated genome size of 1.12 Gb, consistent with other bird genomes (Zhang et al. 2014a). MAKER produced 17,351 protein-coding annotations.

The PSMC analysis showed a large increase in effective population size from about $1x10^5$ individuals ~400k years ago to a peak of $5x10^5$ individuals ~100k years ago, and mild decrease to about $3x10^5$ individuals 10k years ago (Figure 2.1).

Genome Resequencing

We sequenced 42,508,670 reads (each 125 base pairs) on average for each House Finch individual, of which of 96.8% on average aligned to the House Finch genome, with a 3.3% mean duplication rate. After processing, the mean depth of coverage for House Finch individuals was 4.4x (standard deviation = 0.96). We sequenced Cassin's Finches and Purple Finches to a mean 6.7x depth of coverage with an alignment rate of 96.5% and 2.9% duplication rate. We sequenced additional finch outgroups to a 20.8x mean depth of coverage with an alignment rate of 96.4% and a 5.2% duplication rate. Out of ~$896x10^6$ sites passing quality filters, we identified between ~$20.1x10^6$ and ~$27.3x10^6$ sites likely to be variant ($p<1e^{-6}$) in each eastern locality and time period, and between $10.9x10^6$ and $32.7x10^6$ in each western locality and time period (Table 2.2).

Table 2.2: Summary statistics for autosomes and Z chromosome sites for each population. Alabama populations, Cassin’s Finches, and Purple Finches were excluded from Z-chromosome calculations due to small sample sizes.

region	year	locality	autosomal sites	variant autosomal sites	autosomal Θw	autosomal π	autosomal Tajima's D
eastern	~1990	Alabama	896,114,432	21,558,094	7.30E-03	7.02E-03	-0.285
		Maine	896,339,151	25,510,507	7.25E-03	7.15E-03	-0.165
		New York	896,339,258	25,130,057	7.14E-03	7.10E-03	-0.144
		Ohio	896,336,834	25,771,629	7.32E-03	7.09E-03	-0.238
	~2001	Alabama	896,271,206	20,103,681	6.81E-03	6.77E-03	-0.153
		Maine	896,330,719	24,563,067	6.97E-03	6.96E-03	-0.123
		New York	896,338,129	24,665,264	6.89E-03	6.91E-03	-0.106
		Ohio	896,339,251	26,056,209	7.29E-03	7.11E-03	-0.223
	~2014	Alabama	896,180,065	21,279,801	7.21E-03	6.98E-03	-0.258
		Illinois	896,338,729	27,252,478	7.63E-03	7.19E-03	-0.366
		Massachusetts	896,338,132	24,245,932	7.02E-03	6.97E-03	-0.146
western	~2001	Washington	896,339,230	30,920,222	8.63E-03	7.33E-03	-0.749
		California	896,337,810	32,753,621	9.13E-03	7.47E-03	-0.880
	~2014	Washington	896,339,364	29,821,583	8.32E-03	7.27E-03	-0.645
		California	896,339,216	30,973,809	8.78E-03	7.34E-03	-0.791
cassin's finch	-	-	893,065,571	23,690,727	7.96E-03	7.29E-03	-0.525
purple finch	-	-	893,032,797	17,147,000	5.57E-03	4.83E-03	-0.775

Table 2.2 (Continued)

region	year	locality	z chr sites	z chr variant sites	z chr Θw	z chr π	z chr Tajima's D	z chr/auto Θw	z chr/auto π
eastern	~1990	Alabama	-	-	-	-	-	-	-
		Maine	69,974,820	1,433,109	6.28E-03	6.27E-03	-0.077	0.866	0.877
		New York	69,975,154	1,494,896	6.13E-03	6.18E-03	-0.059	0.858	0.871
		Ohio	69,975,122	1,636,560	6.37E-03	6.37E-03	-0.076	0.869	0.899
	~2001	Alabama	-	-	-	-	-	-	-
		Maine	69,974,893	1,589,496	6.34E-03	6.26E-03	-0.117	0.909	0.900
		New York	69,975,275	1,599,796	5.98E-03	6.19E-03	0.011	0.867	0.896
		Ohio	69,971,808	1,176,611	6.05E-03	6.05E-03	-0.064	0.830	0.852
	~2014	Alabama	-	-	-	-	-	-	-
		Illinois	69,974,748	1,454,304	6.37E-03	6.31E-03	-0.108	0.836	0.878
		Massachusetts	69,974,646	1,448,858	6.12E-03	6.22E-03	-0.041	0.872	0.892
western	~2001	Washington	69,975,355	1,875,430	7.29E-03	6.57E-03	-0.366	0.845	0.897
		California	69,975,002	1,904,671	7.59E-03	6.64E-03	-0.434	0.831	0.888
	~2014	Washington	69,975,034	1,630,896	6.88E-03	6.43E-03	-0.264	0.827	0.885
		California	69,975,378	1,977,088	7.52E-03	6.62E-03	-0.423	0.857	0.901
cassin's finch	-	-	-	-	-	-	-	-	-
purple finch	-	-	-	-	-	-	-	-	-

Bottlenecks and expansions structure House Finch populations

The House Finch, Cassin's Finch and Purple Finch PCA incorporated 70,283,925 variant sites ($p<1e^{-6}$), each with a minor allele frequency of at least 0.05. PC1, representing 19.2% of the variance, clearly separated House Finches from Cassin's Finches and Purple Finches. PC2, representing 1.33% of the variance, separated Cassin's and Purple Finches (Figure 2.2A). These patterns are consistent with a previous study that identified a House Finch divergence from the ancestor of Cassin's Finches and Purple Finches approximately 10 mya followed by a Cassin's Finch and Purple Finch split approximately 4 mya (Smith et al. 2013). PCs 3-16 primarily separated variation within and between Cassin's Finches and Purple Finches. Subsequent PC axes each represented less than 1% of the variation in the dataset, but show the patterns described below.

The House Finch PCA incorporated 39,018,583 sites. PC1, explaining 1.51% of the variation, separated eastern and western populations (Figure 2.2B). PC2, explaining 0.66% of the variation, separated the eastern populations by distance from the site of introduction. New York, the hypothesized introduction site is the most positive, followed by Massachusetts and Maine, and finally Illinois, Ohio and Auburn (Figure 2.2B).

The eastern House Finch PCA incorporated 33,900,320 sites, and PC1, explaining 0.97% of the variation, showed subdivision identical to PC2 of the House Finch analysis described above (not pictured). Subsequent PC axes did not show further subdivision of geographic or temporal populations.

The western House Finch PCA incorporated 39,809,948 sites, and PC1, explaining 2.05% of the variation, separated western House Finches into two groups, a small group of 10 House Finch and a larger group with the remaining individuals. Both groups contained individuals belonging to both geographic populations and time periods. PC2, representing 1.91% of the variation, separated California and Washington populations (Figure 2.2C).

The Alabama, Maine, New York, Ohio, California and Washington PCAs incorporated 22,141,897, 27,786,099, 27,186,402, 28,657,102, 34,453,381, and 32,848,625 sites respectively. No populations showed any temporal structuring with the exception of California, for which PC3, consisting of 3.47% of the variation in the dataset, separated the 2001 and 2014 samples (Figure 2.2D).

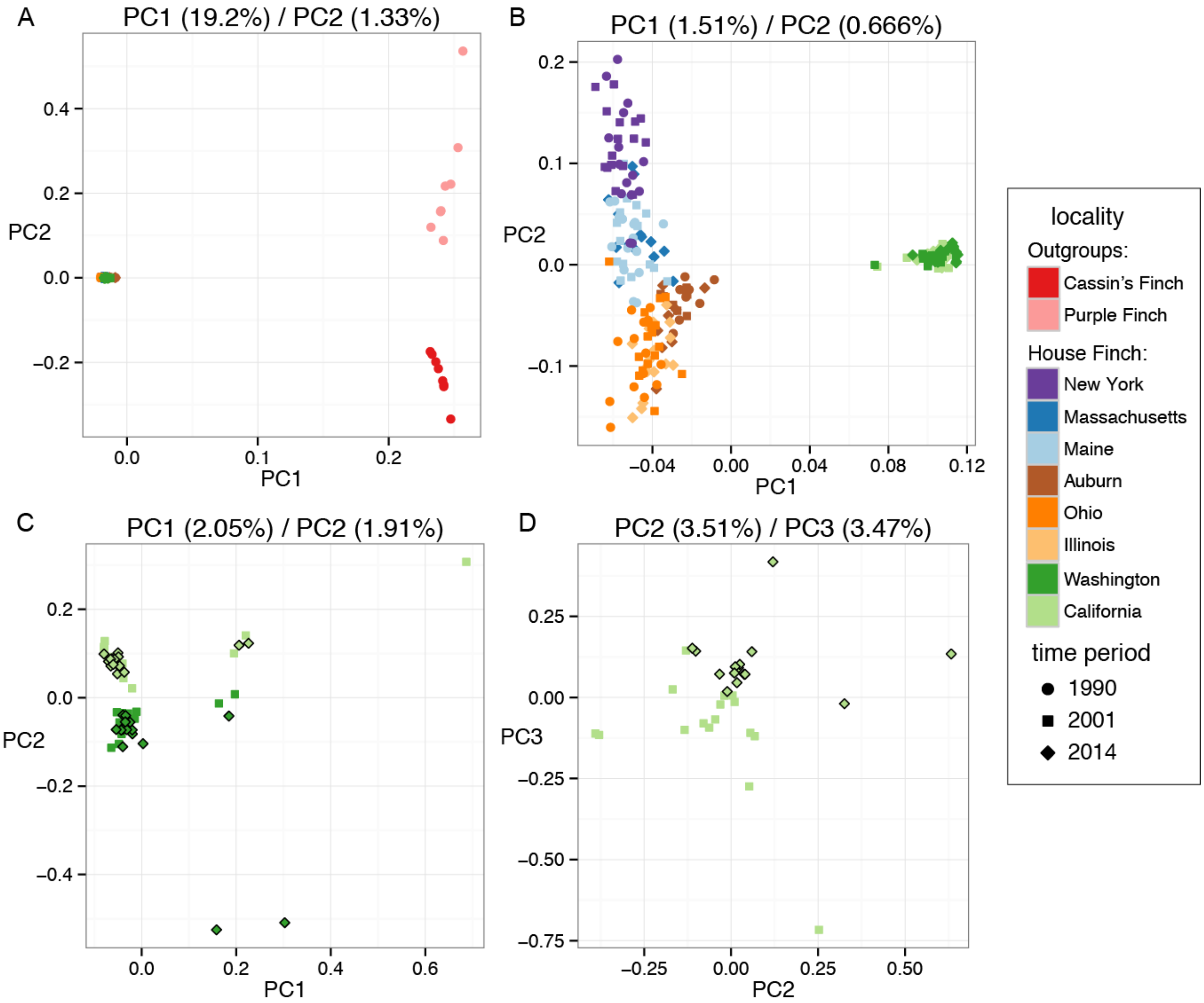


Figure 2.2: PCA of A) all House Finches, Cassin's Finches, and Purple Finches, B) all House Finches, C) western House Finches, and D) California House Finches. The percentage of variation explained by each axis displayed is indicated at the top of each figure. Point colors correspond to geographic localities and point shapes correspond to time period. Samples from the 2014 period in C) and D) are outlined in black to facilitate identification.

Deviations in the site frequency spectra

There were no differences in the autosomal SFS among localities collected in the same region (eastern or western) and time period (e.g. eastern localities collected in 1990; Figure 2.3; K-S tests, $p > 0.05$), or among average SFS from different time periods but the same region (e.g. eastern 1990 vs. 2001; Figure 2.3; K-S tests, $p > 0.05$). There were moderately significant differences among eastern and western populations - between all pairwise combinations of 1990 and 2014 eastern SFS and 2001 and 2014 western SFS, all K-S test $p = 0.0996$. There were significant differences between eastern 2001 and western 2001 or 2014 populations (Figure 2.3; K-S test: $p = 0.022$ and $p = 0.049$ respectively). All population SFS were significantly different from neutral expectations (Figure 2.3; $p < 0.001$).

We summarize the patterns described by the SFS by calculating Tajima's D for each population (Table 2.2). Western populations show more significantly negative values of Tajima's D, compared to eastern populations in all time periods (all comparisons: Mann-Whitney U-test $p < 0.0001$).

Among eastern populations, the SFS for intergenic and intronic sites were not significantly different from each other (K-S test: $p > 0.05$), but the SFS for coding sequence sites was significantly different from intergenic or intronic sites (K-S test: all $p < 0.05$) (Figure 2.4A,B,C). The functional class SFS from the 2001 western population were not significantly different from each other (Figure 2.4D; K-S test: $p > 0.05$). None of the 1990 eastern functional classes were significantly different from the neutral SFS (Figure 2.4A,B,C; K-S test: $p > 0.05$), but all of the 2001 western functional classes were significantly different from the neutral SFS (Figure 2.4D; K-S test: $p < 0.05$). The SFS for each functional class from the 1990 eastern and 2001 western populations was significantly different from the others (Figure 2.4A,D; K-S test: p < 0.001).

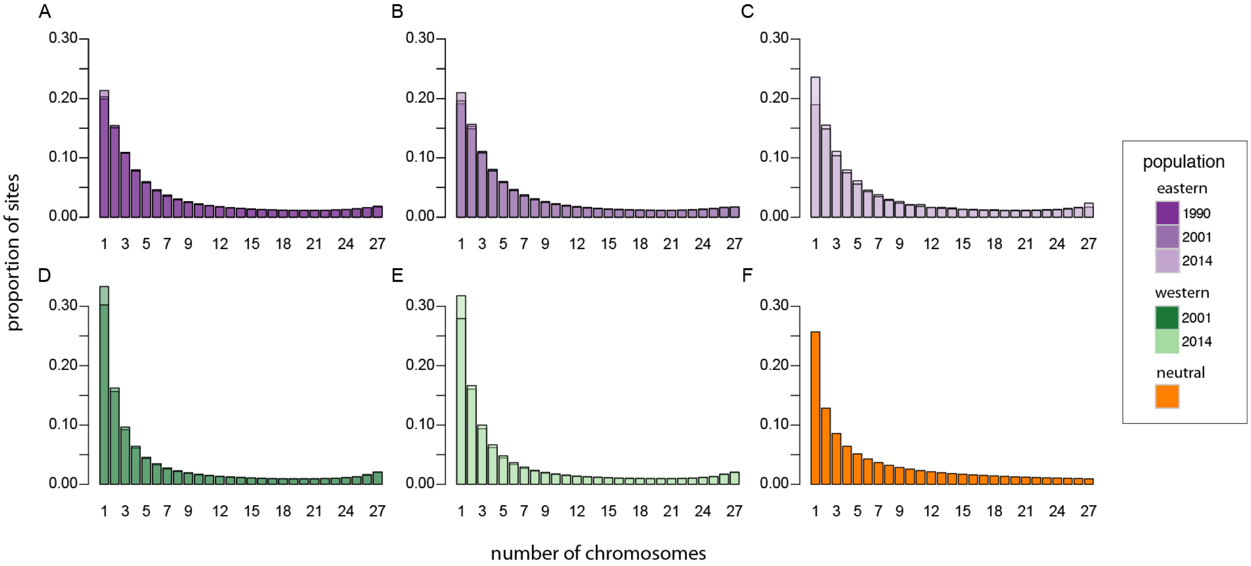


Figure 2.3: Site frequency spectra of autosomal sites from samples collected in A) eastern 1990, B) eastern 2001, C) eastern 2014, D) western 2001, E) western 2014, and F) the neutral expectation. SFS for each locality within each region and time period are overlaid to display variance within that population.

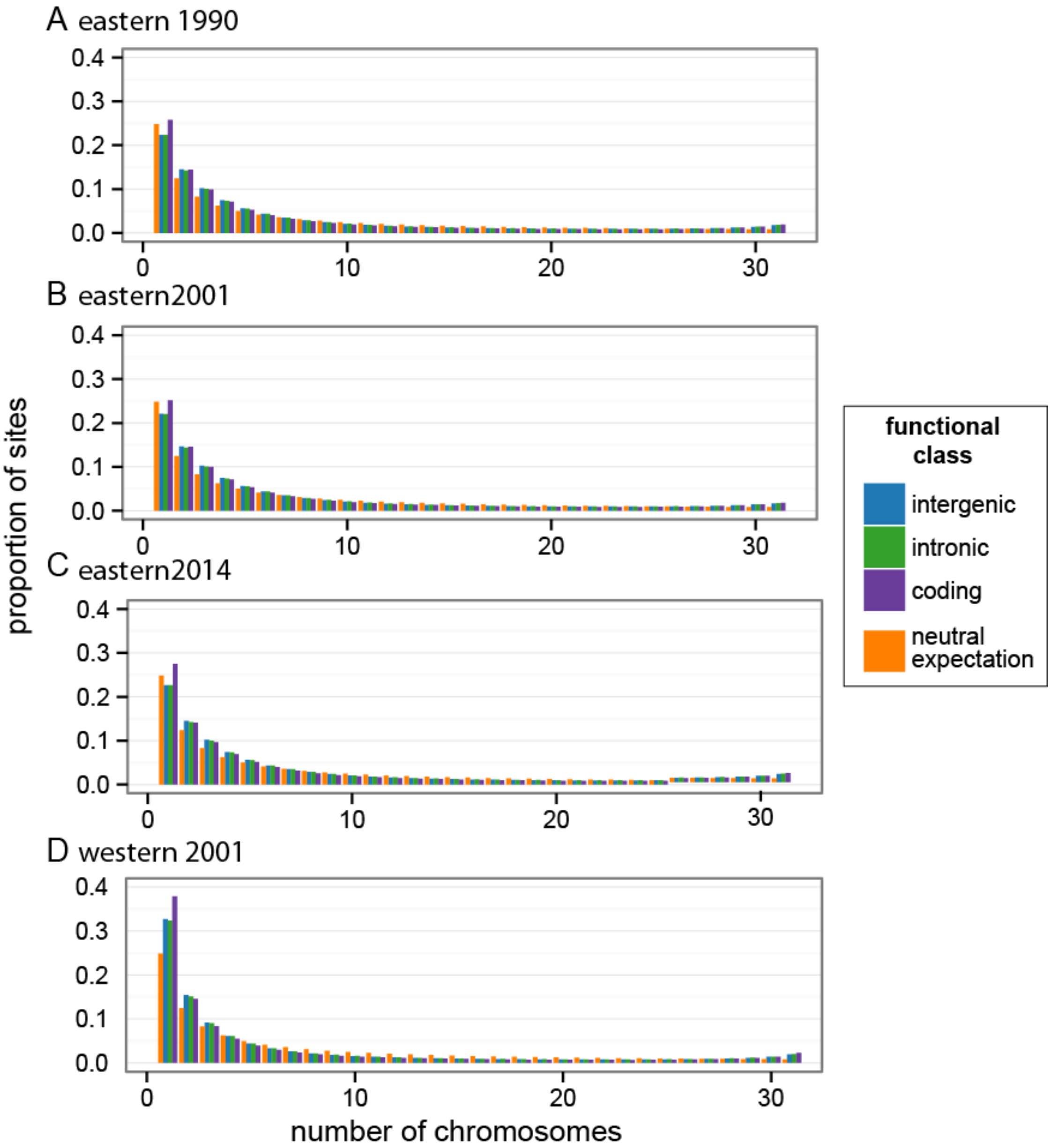


Figure 2.4: Site frequency spectra of intergenic, intronic, and coding sites within eastern populations in A) 1990, B) 2001, and C) 2014, and D) the western population in 2001. SFS have been projected down to 32 chromosomes to facilitate visualization.

Reduction in Genetic Diversity among populations but not time points

We compared patterns of genetic diversity by calculating π and Θ_w for each population (Table 2.2). Autosomal regions showed reduced levels of genetic diversity in eastern populations due to the introduction (Figure 2.5A), with an average 4.2% reduction in π for eastern 1990 populations (average π = 0.0071) compared to 2001 western populations (average π = 0.0074). Genetic diversity as measured by Θ_w showed an average 18.3% reduction in eastern 1990 populations (average Θ_w = 0.0073) compared to western 2001 populations (average Θ_w = 0.0089; Figure 2.5B. All comparisons were significant with a Mann-Whitney U-Test ($p<0.0001$).

We also detect a small decrease in genetic diversity when comparing populations collected before and after the epizootic in both eastern and western populations (Figure 2.5A and 2.5B): eastern 1990 vs eastern 2001, average π reduced by 2.2%, average Θ_w reduced by 3.7%, western 2001 vs 2014, average π reduced by 1.3%, average Θ_w reduced by 3.8%. Eastern populations collected in 2014 showed similar diversity levels compared to those observed in 1990 (Figure 2.5A and 2.5B).

Genetic diversity was more strongly reduced for Z chromosome sites compared to autosomal sites, as predicted by the difference in effective population size of sex chromosomes and autosomes (3/4 chromosomes; Table 2.2). On average, eastern populations showed slightly lower ratios of Z chromosome to autosome π values than western populations (88.3% vs 89.3% respectively). However, the variability among localities is greater than these differences (eastern SD = 1.7%, western SD = 0.8%; Figure 2.5D).

For all populations, genetic diversity (π and Θ_w) is reduced for coding sites compared to intergenic and intronic sites (Table 2.3), and for intronic sites compared to intergenic sites (Table 2.3). In eastern populations, Tajima's D from coding sites is more strongly negative than Tajima's D from intronic and intergenic sites (Table 2.3), and slightly more negative for intergenic sites than intronic sites (Table 2.3). The opposite pattern holds true for western House Finches, Cassin's Finches and Purple Finches, with more strongly negative intergenic sites, and intronic sites slightly more negative than coding sites (Table 2.3). All comparisons reported above were statistically significant with a Mann-Whitney U-test ($p < 0.0001$).

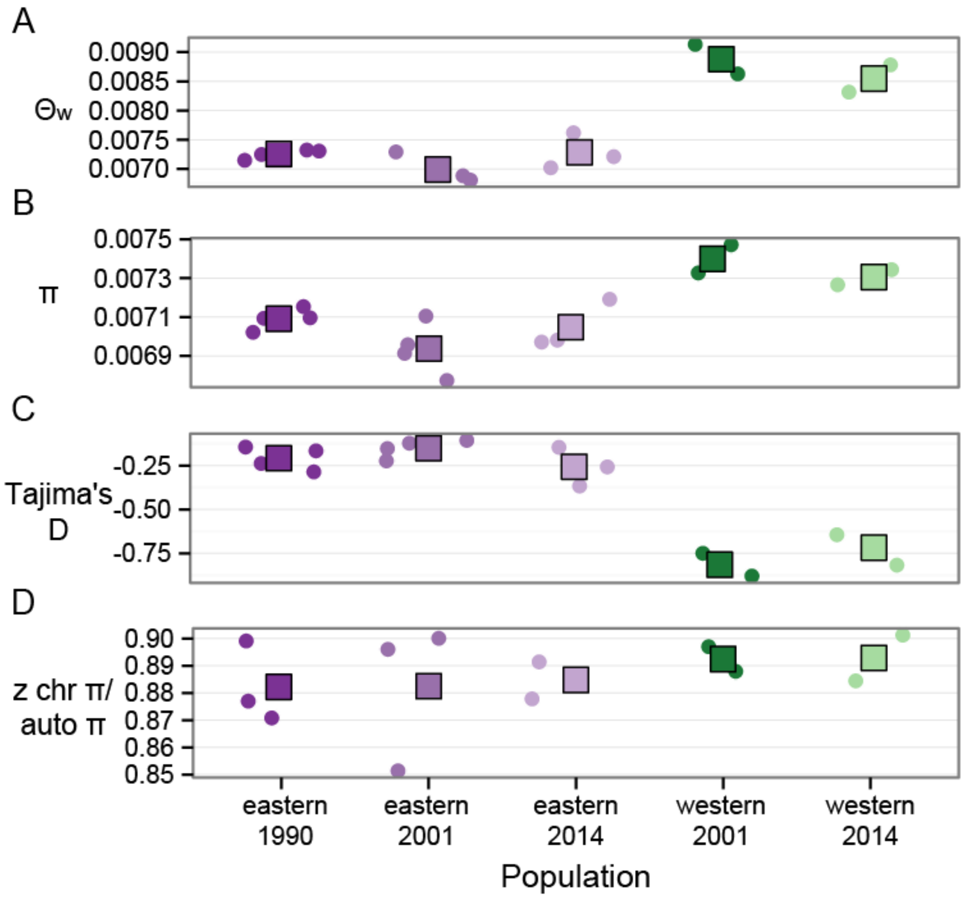


Figure 2.5: Comparisons of A) Watterson's Θ (Θ_W), B) π, C) Tajima's D, and D) autosomal π to Z chromosome π ratio among populations and time periods. Values for each locality are illustrated as circles, and means across all localities for each time period are illustrated by squares. Note that Auburn localities are excluded from Z chromosome π to Autosome π comparisons due to small sample sizes.

Table 3: Estimates of summary statistics for all regional populations, Cassin’s Finches, and Purple Finches.

		π			**Θw**			**Tajima's D**		
region	**year**	**coding**	**intronic**	**intergenic**	**coding**	**intronic**	**intergenic**	**coding**	**intronic**	**intergenic**
eastern	1990	2.90E-03	6.71E-03	7.53E-03	3.36E-03	6.78E-03	7.60E-03	-0.26	-0.09	-0.16
	2001	2.85E-03	6.61E-03	7.43E-03	3.23E-03	6.56E-03	7.38E-03	-0.23	-0.04	-0.08
	2014	2.91E-03	6.65E-03	7.47E-03	3.54E-03	6.91E-03	7.75E-03	-0.31	-0.16	-0.29
western	2001	3.05E-03	6.98E-03	7.81E-03	4.71E-03	9.06E-03	1.02E-02	-0.56	-0.66	-1.21
cassin's finch	-	2.87E-03	6.76E-03	7.79E-03	3.43E-03	7.36E-03	8.48E-03	-0.24	-0.27	-0.49
purple finch	-	2.00E-03	4.47E-03	5.15E-03	2.50E-03	5.20E-03	5.91E-03	-0.25	-0.40	-0.72

Allele frequency change across space and time

Comparisons among regions resulted in more allele frequency outliers than comparisons among time periods within a region. Between 15.3x10^6 and 16.3x10^6 comparisons were made between samples collected from different time periods within a region. Between 1990 and 2001 in the eastern population, two SNPs were significantly different after correcting for multiple tests, one on scaffold 2 ($q = 0.0021$), and one on scaffold 61 ($q = 0.0076$) (Figure 2.6A). Between 1990 and 2014 in the eastern population there were no significant differences ($q < 0.05$; not pictured), and between 2001 and 2014 in the eastern population, there was one significantly different SNP on scaffold 2 ($q = 0.0023$) (not pictured). There were no significant differences between the 2001 and 2014 western populations (not pictured).

Of the 15.7x10^6 comparisons made between the 1990 eastern population and 2001 western population, 5,165, or 0.032% of sites showed significantly different allele frequencies after correcting for multiple testing (Figure 2.6B). Of the 15.5x10^6 comparisons made between the 2014 eastern population and 2014 western population, 2,061, or 0.013% of sites showed significantly different allele frequencies (Figure 2.6C). To identify any scaffolds with higher than expected proportions of significant sites, we compared the top 1% of proportion of significant alleles per scaffold between time periods (Supplemental Figure 2.7). One scaffold had a high proportion of significant sites in two time periods – scaffolds 24. Scaffold 24 had the most significant comparisons in both tests, 219 and 182 in the eastern 1990 versus western 2001 and eastern 2014 versus western 2014 respectively (Figure 2.8).

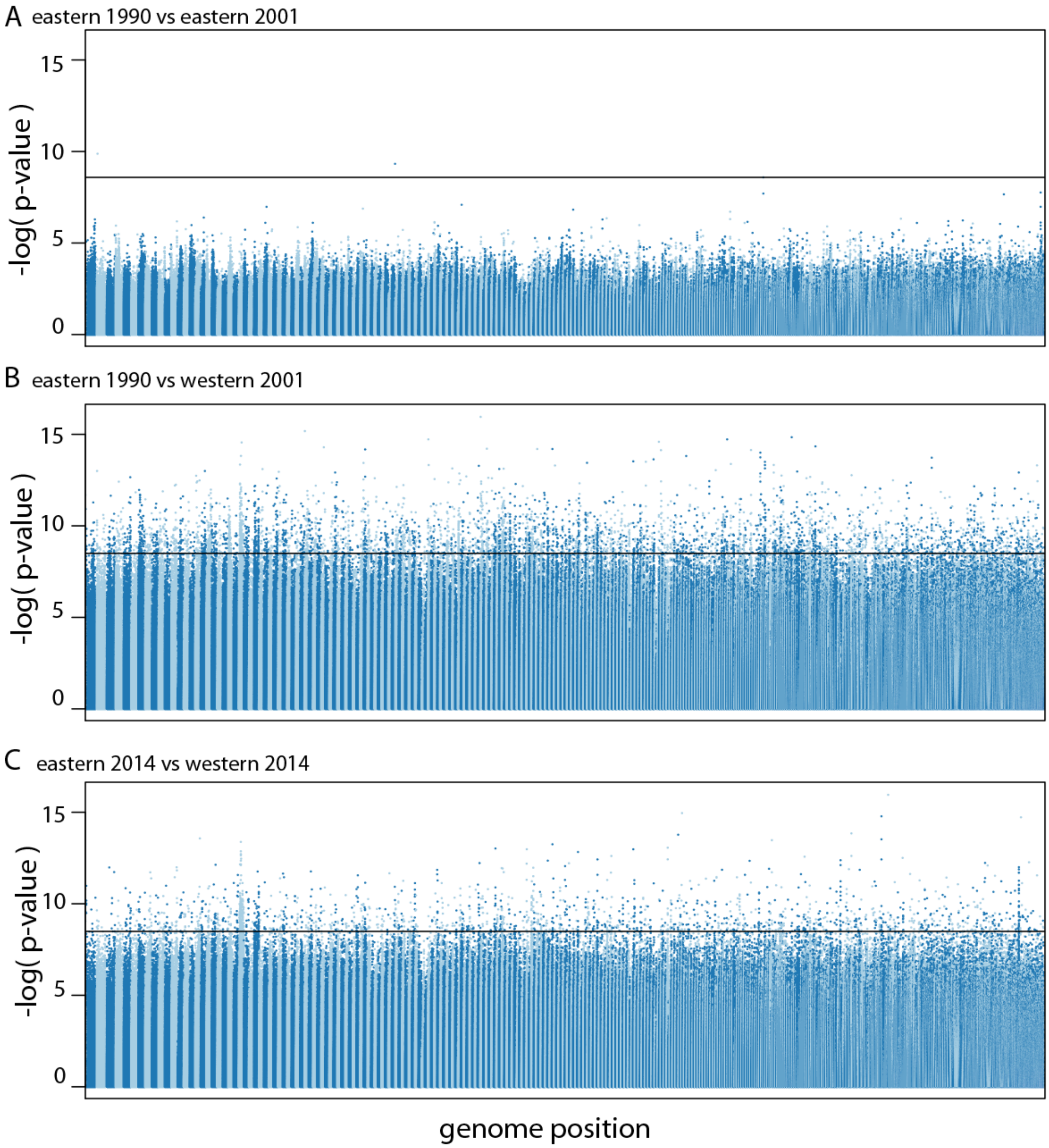


Figure 2.6: Manhattan plots of p-values for each site tested for either spatial or temporal allele frequency associations between A) eastern 1990 and eastern 2001, B) eastern 1990 and western 2001, and C) eastern 2014 and western 2014. Scaffolds are ordered by length, and alternative between light blue and dark blue. The horizontal line represents the significance cut-off for a FDR of 0.05.

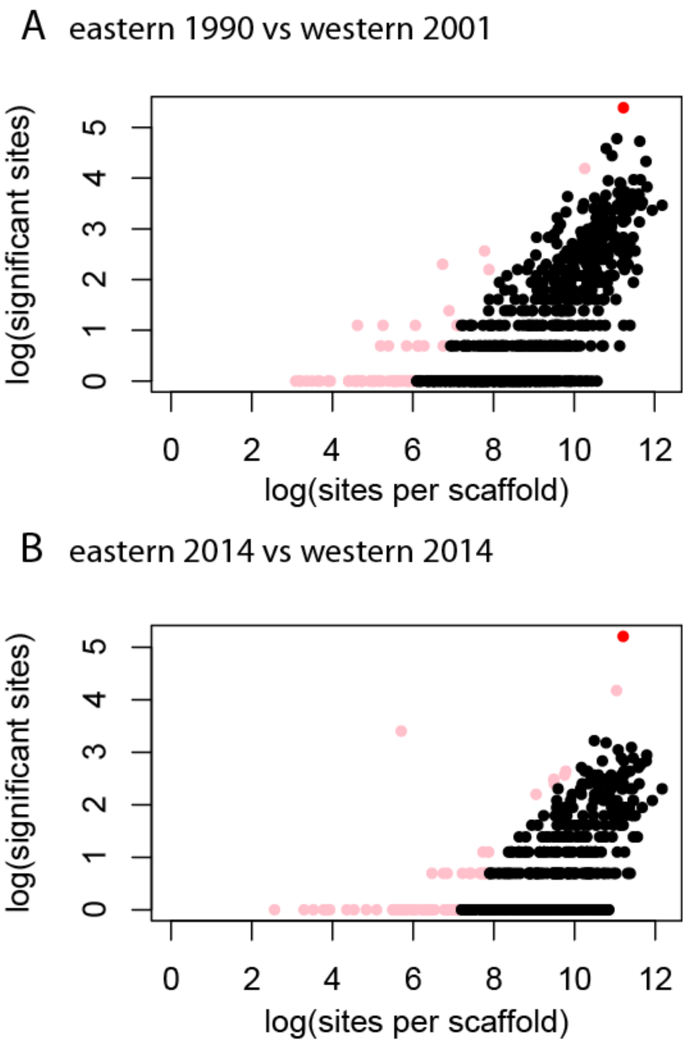


Figure 2.7: Log-scale plots of the number of comparisons conducted for each scaffold on the x-axis, and number of significant comparisons for each scaffold on the y-axis, between A) eastern 1990 and western 2001 populations and B) eastern 2014 and western 2014 populations. Scaffolds with proportions of significant sites in the top 1% of the distribution are colored pink, and the five overlapping scaffolds between the two comparisons are colored red.

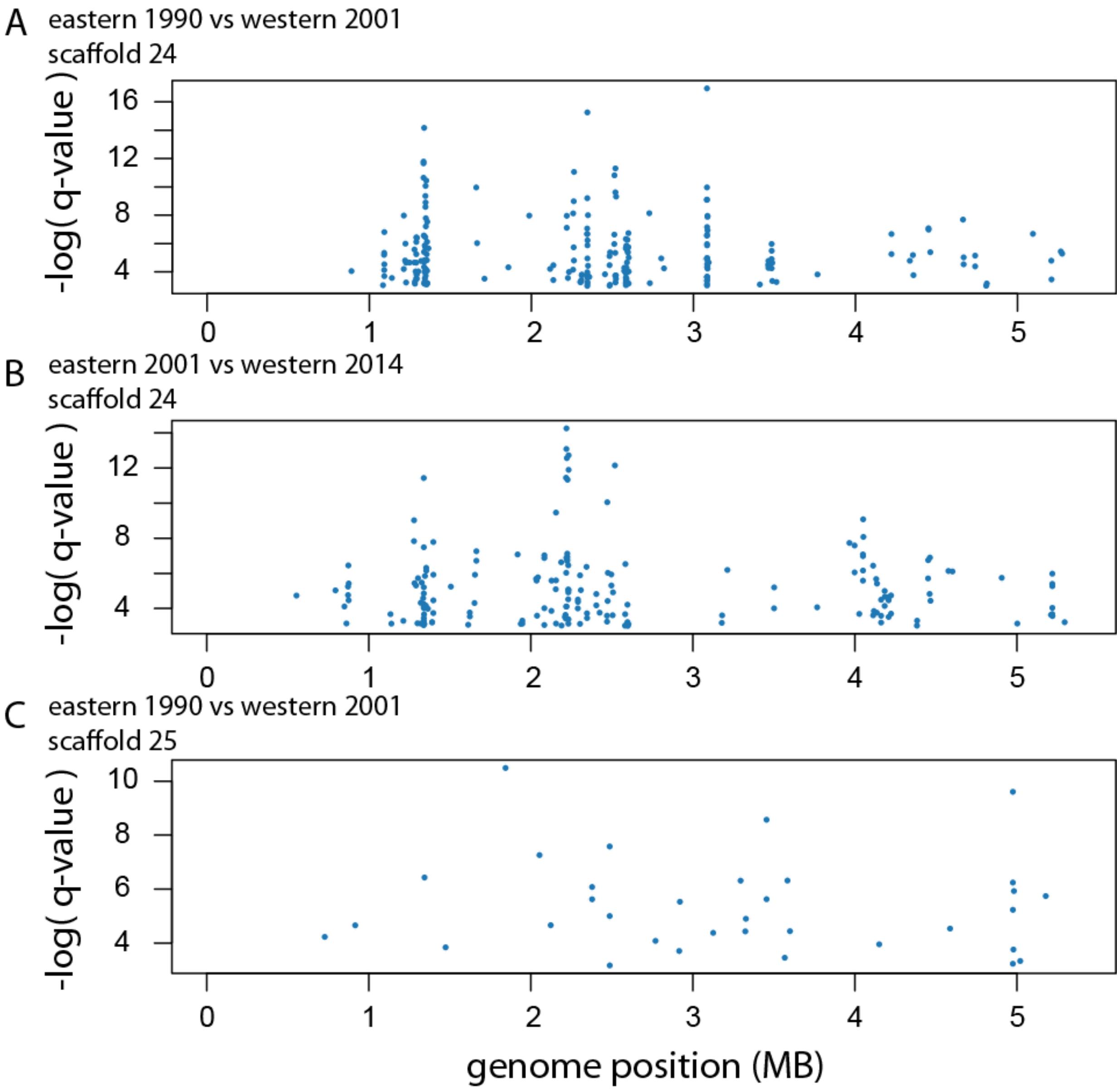


Figure 2.8: Manhattan plot of significant associations between A) eastern 1990 and western 2001 for scaffold 24, B) eastern 2014 and western 2014 for scaffold 24, and C) eastern 1990 and western 2001 for scaffold 25. Scaffold 24 has the most significant sites, including spatial clustering along the scaffold. Scaffold 25 is an example of a similarly sized scaffold not enriched for significant sites.

DISCUSSION

Across the genome, despite phenotypic and experimental evidence for recent selection, the largest patterns we observe are consistent with demographic expectations from bottlenecks and population expansions. The strongest signals in the genome of the House Finch are first for a recent population expansion that likely coincided with glacial retreat following the last glacial maximum, and second for a bottleneck in the introduced eastern population. We detected variation in the effects of the introduction on sites from different functional classes, but no effect on the z:autosome ratio of genetic diversity. We find substantial differentiation due to the introduction at specific sites throughout the genome that is likely a product of both selection and genetic drift, but which is sometimes spatially clustered. Unlike previous studies, we detect a small decrease in genetic diversity following the MG epizootic, but no signature of large selective sweeps. The lack of evidence for selective sweeps together with no temporal structuring among individuals suggests that any MG-mediated adaptation occurred by limited polygenic selection - selection at sufficient loci to reduce the power of traditional selection scans, but not pervasive enough to cause temporal structuring.

Genome-wide impacts of demographic events

Populations are structured by bottlenecks and expansions

We find strong signatures of bottlenecks and expansions across the genome with reductions in genetic diversity, shifts in the SFS patterns, and variation in Tajima's D. These signatures show that demographic processes may be the most influential processes shaping patterns of biodiversity. Over 2,000 ddRADseq loci (Shultz et al. 2016) and whole genome data both show that the strongest signal of differentiation among populations of the *frontalis* House Finch subspecies (the only subspecies investigated in this study) occurs between native (western) and introduced (eastern) House Finches (Figure 2.2B). However, unlike previous studies (Wang et al. 2003; Hawley et al. 2006; 2008; Shultz et al. 2016), our whole genome dataset enables us to observe differentiation among localities in both regional populations.

Allele surfing, or the rapid spread of rare alleles at the front of an expanding population resulting in higher allele frequencies, can result in fine-scale spatial geographic differentiation in a recently expanded population (Excoffier and Ray 2008). Following the eastern-western split, the strongest signal of differentiation occurs among eastern localities (Figure 2.2B). We observe that eastern localities are structured according to the distance from Long Island, New York (Figure 2.9), the likely origin point of the eastern House Finch population (Elliott and Arbib 1953). It is surprising that this pattern was not detectable by thousands of RADseq loci, despite using a discriminant analysis of principal components analysis, which maximizes PCA differences among populations (Jombart et al. 2010). The large amount of power required to observe this phenomenon may indicate that allele surfing is minimal due to substantial dispersal abilities (40-700 km: Veit and Lewis 1996). Allele surfing is expected to be reduced in a rapidly expanding sexual population with large dispersal distances (Goodsman et al. 2014). We also observe structuring among the western populations, although this is not the strongest signal in the data (Figure 2.2C). The western population shows genetic signatures of a population expansion following the last glacial maximum (discussed below). Without other geographic localities for comparison it is difficult to know whether or not this structure shows geographic patterns consistent with expectations from allele surfing. It is unclear why some individuals appear to be more divergent than the variation among localities, and this divergence could be a product of sequencing artifacts. However, none of the individuals show evidence of differences in coverage, duplication rate, alignment rate, were not prepared in the same library preparation batch, or sequenced together on the same lanes. An alternative biological explanation could be that these individuals represent recent immigrants to these localities. However, without broader geographic sampling in the native population we cannot address this hypothesis.

We observe very little temporal genetic differentiation in our PCAs, even when only single geographic localities were included. This indicates that neither demographic effects associated with possible bottlenecks due to the epizootic, nor pervasive polygenic selection (e.g. Bergland et al. 2014) are sufficient to differentiate populations collected at the same localities. The exception is the California population, which is temporally distinguishable via PCA (Figure 2.2D). However, the 2014 California populations were collected in Marin and Contra Costa counties, and the 2001 population was collected in Sacramento County, approximately 150 miles away, so this could be instead a signal of very fine-scale population structure.

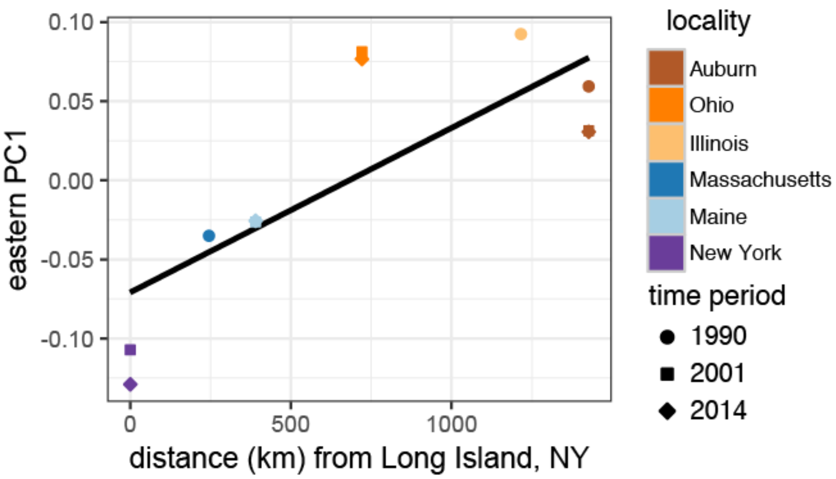


Figure 2.9: Relationship between average eastern PC1 score and distance from the original introduction site in Long Island, NY. The two variables are significantly correlated (linear model: R^2=0.61, p = 0.004).

Genetic diversity following the founding event and epizootic

The most prominent signature of population differentiation among House Finch populations was associated with the introduction, consistent with increased genetic drift during the introduction. Increased genetic drift is expected when populations have smaller effective population sizes (Nielsen 2005). So, as predicted, the largest differences in genetic diversity, as measured both by π and Θ_w, are observed between native and introduced populations. Interestingly, when comparing the 1990 eastern population and 2001 western population (both prior to the epizootic event), π showed only a 4.17% reduction in genetic diversity on average, where Θ_w showed an 18.28% reduction. This difference can be visualized by comparing the SFS between the two populations (Figure 2.3). The western population clearly exhibits a SFS with an excess of rare alleles compared to the neutral expectation, consistent with a recent population expansion (Nielsen 2005). Thus, western populations have much higher Θ_w values than π values. The eastern population exhibits a SFS shifted toward higher frequencies of intermediate and derived alleles, as expected following a bottleneck (Nielsen 2005), and producing more similar Θ_w and π estimates for each location and time period. These differences between π and Θ_w are quantified by Tajima's D, which is significantly more negative in western populations. Many studies only use a single measure to compare genetic diversity across populations. However, our results emphasize the importance of considering the possible biases in approximations of genetic diversity caused by deviations from population equilibrium when interpreting variation in these measures.

Shultz et al. (2016) found no signature of a reduction in genetic diversity following the epizootic event in eastern populations. Here, we observe a 3.66% and 2.16% reduction in average Θ_W and π respectively when comparing post-epizootic eastern 2001 populations to pre-epizootic eastern 1990 populations (Figure 2.5A,B), followed by an increase to pre-epizootic levels in 2014 in the eastern population. We also observe a 3.75% and 1.27% reduction in 2014 western populations compared to 2001 populations. While it is possible these are signatures of a minor bottleneck resulting from the epizootic event followed by a brief expansion back to pre-epizootic levels in the eastern population, one locality, Ohio, has estimates of genetic diversity slightly higher in 2001 than 1990. Tajima's D is accordingly less negative in 2001. However, due to the high power associated with genomic-scale data (Kumar et al. 2012), it is difficult to know if even significant differences in estimates are biologically meaningful. Simulations, demographic modeling, or sample permutation through time may help to distinguish between these different scenarios, or determine whether the variation we observe can be expected by chance.

While the introduction clearly had the largest effect on genetic diversity among House Finch populations, the demographic history of the native western population played an important role in shaping signatures of genome-wide genetic diversity in both populations. As discussed above, the western SFS are skewed toward rare alleles (Figure 2.3A,B), and have an associated negative Tajima's D (Figure 2.5C). The PSMC analysis also showed evidence of population growth and decline according to past glacial history (Hewitt 2004; Nadachowska-Brzyska et al. 2015), although the western population's most recent expansion in the current interglacial period would have been too recent for PSMC to detect (Li and Durbin 2011). The population history of this species has never been studied throughout its range using genetic tools, but there are 11 subspecies defined based on morphology and plumage coloration, with the most morphological and plumage diversity in southern Mexico (Moore 1939; Hill 1996; 2002). These phenotypic patterns are consistent with the hypothesis that glacial cycling plays an important role in shaping the population genetic history of many plants and animal (Avise and Walker 1998; Hewitt 2000; 2004; Zink et al. 2004). Further genetic work on House Finches collected from different subspecies and localities throughout the southern part of their range can be used to further describe the population genetic history of House Finches as a whole. However, the substantial genetic divergence between the *frontalis* and *griscomi* subspecies observed with ddRADseq data (Shultz et al. 2016), coupled with the PSMC results and signatures of expansion provide compelling evidence for the importance of glacial cycling in the history of the House Finch as a species.

Wang et al. (2003), Shultz et al. (2016) and the present study utilize very similar sampling strategies but different types of genome-wide markers, allowing us to compare the results of amplified fragment length polymorphisms (AFLPs), double-digest RADseq (ddRADseq), and low-coverage whole genome resequencing on estimations of genome-wide genetic diversity as measured by π. For the pre-epizootic eastern population, we observe average estimates of 0.0076, 0.0048, and 0.0071 from AFLPs, ddRADseq, and whole genome resequencing data respectively, and for the pre-epizootic western population, we observe average estimates of 0.0075, 0.0053, and 0.0074 respectively. AFLPs and whole genome resequencing produced overall similar estimates of π, although AFLPs may not be sensitive enough to detect the minor reductions in genetic diversity we observe as a result of the recent founder event. However, ddRADseq produced only 68% and 72% of whole genome resequencing π estimates for eastern and western populations respectively. This is an empirical demonstration that estimates of genetic diversity can be even more extreme with ddRADseq data than the ~80% reduction allele dropout would predict (Arnold et al. 2013). We attribute a possible cause of the additional reduction of genetic diversity to the use of a *de novo* library assembly approach with overly conservative mismatch parameters (Shultz et al. 2016). Without a reference genome for sequence mapping, researchers must identify loci based on distance. One of the most popular programs for this, Stacks (Catchen et al. 2011; 2013), requires users to *a priori* specific the number of mismatches allowed between sequences representing the same location in the genome. If overly conservative mismatch parameters are chosen, sequences from the same region of the genome may be split into two separate loci, reducing estimates of polymorphism. However, additional work would be necessary to test this hypothesis. It is also possible that the whole-genome resequencing overestimates genetic diversity due to sequencing errors. It is also worth noting that the nonequilibrium histories of both the native and introduced populations likely increased the inferred reduction in π in the eastern 1990 versus western 2001 population in ddRADseq data (7% using ddRADseq data vs. 4.2% using whole genome data). This is consistent with the idea that allelic dropout may be more or less prevalent depending on demographic history. Arnold et al. (2013) show that expanding populations are less affected by allelic dropout compared to neutral or bottlenecked populations. Thus, even in closely-related populations of the same species allele dropout can have differential effects on estimations of genetic diversity. Authors of future RADseq studies, especially ddRADseq due to additional dropout with two RAD cut sites, should be aware of these biases when interpreting their results. Simulations of several different technologies would be useful to fully characterize biases in estimates of genetic diversity with different genotyping strategies.

Genome heterogeneity and demographic events

Previous work has assessed the genome-wide impacts of the House Finch introduction and epizootic on the genome. However, these demographic events may have different impacts on sites in various functional classes due to genome heterogeneity caused by variation in effective population sizes, recombination rate variation, mutation rate variation, or linked selection (Lohmueller et al. 2011; Burri et al. 2015; Singhal et al. 2015). We investigate this possibility by comparing SFS and genetic diversity in sites from three functional classes, by comparing the ratio of genetic diversity on autosomes versus the Z chromosome, and by comparing allele frequency changes at all individual sites through time and space.

Genetic diversity among functional classes

Throughout our draft genome assembly, coding, intronic, and intergenic sites were annotated using Maker. Within all populations, the largest difference in both π and Θ_W was between coding sites and both intronic and intergenic sites (Table 2.3). Coding sites also showed the largest excess of rare variants (Figure 2.4). These patterns are consistent with purifying selection on coding regions of the genome (Lohmueller et al. 2011; Koch and Novembre 2017). Genetic diversity was also slightly reduced for intronic sites relative to intergenic sites, likely reflecting the increased proportion of intronic sites with regulatory importance (Torgerson et al. 2009).

Few studies have examined patterns of genetic diversity for functional classes of sites in non-model organisms following recent anthropogenic disturbances. We expected coding sites to have a smaller reduction in genetic diversity than intronic or intergenic sites due to heterozygosity maintained by the mutation-selection balance at sites under strong purifying selection (Simons et al. 2014; Zhang et al. 2014c; Koch and Novembre 2017). However, between eastern 1990 and western 2001 populations we observed a 4.9% reduction of π in coding sites compared to a 3.7% and 3.5% reduction in intronic and intergenic sites respectively. Coding sites not being uniformly under strong selection may explain this pattern; further categorizing these into 4-fold and 0-fold degenerate sites could result in the pattern predicted by mutation-selection balance. An intriguing possibility is that the bottleneck during the founder event may have purged some of the deleterious alleles present accumulated during the expansion of the western population due to allele surfing (Peischl et al. 2013). However, further genome annotation will be necessary to test this hypothesis.

Tajima's D values also show different patterns between western and eastern populations. All time points in the eastern population have the most negative Tajima's D values in coding sites, as would be predicted by purifying selection. Western populations, however, have significantly more negative intergenic Tajima's D values, most likely a consequence of the recent population expansion. This pattern is also observed in Cassin's Finches and Purple Finches, both of which also likely underwent an expansion similar to that in the House Finch. This pattern is consistent with the hypothesis that the bottleneck had a larger effect than the population expansion on genomic variation in the eastern House Finches, but the much longer population expansion resulting from glacial cycling produced strong signatures in the genome of the western population.

Comparisons between autosomes and the Z chromosome

Sexual selection, sex-biased dispersal, and sex ratios have been implicated as reasons birds can succeed or fail during an invasion (Sorci et al. 1998; Donald 2007; Shaw and Kokko 2014), with birds that experience strong sexual selection more likely to fail. The red plumage of male House Finches is one of the quintessential examples of a phenotypic character shaped by sexual selection (Hill 2002). When House Finches were brought to the eastern United States as part of the pet trade, it was reported that "only about 100 females were shipped to every 1,000 males, the males being the colored ones" (Elliott and Arbib 1953). However, we show that if there was an excess of males introduced relative to females, and therefore more z chromosomes in the population, this short-term change in sex ratio did not result in an increase in the z chromosome to autosome diversity ratio (Figure 2.5D). This could be a consequence of the Allee affect in House Finches (Veit and Lewis 1996), where even if excess males were introduced, only those that were able to successfully mate with females would have contributed to the founding of the eastern population. In fact, the average z:autosome π ratio was slightly decreased in the eastern population, although this difference is likely not significant as the variance among localities is larger than this decrease.

The z:autosome π ratio of ~0.89 is at first surprising, given that the expected ratio is 0.75. When deviations in the z:autosome π ratio occur, they usually trend below 0.75 and are attributed to a polygynous mating system and sexual selection (Oyler-McCance et al. 2015; Huang and Rabosky 2015). However, the increased ratio observed here is more consistent with the pattern (described above) of population expansion in the western population than with a pattern induced by mating system. Sex chromosomes reach the new equilibrium following population size changes faster than autosomal chromosomes due to their smaller effective population sizes (Pool and Nielsen 2007). Our empirical results highlight the importance of recent demography for modulating z: autosome diversity ratios as compared to life-history characteristics like sexual selection (Huang and Rabosky 2015).

Signatures of recent selection

The two major demographic events in the recent history of continental populations of the House Finch, the introduction of the eastern population and the MG epizootic in both native and eastern populations, produced phenotypic differences among populations. Introduced eastern House Finches exhibit morphological differences (Aldrich and Weske 1978; Egbert and Belthoff 2003), short distance migration (Able and Belthoff 1998), and different sex-biased hatching order (Badyaev et al. 2002) compared to their native western counterparts. Likewise, post-epizootic eastern populations show increased resistance and tolerance to MG compared to historically unexposed populations (Bonneaud et al. 2011; Adelman et al. 2013). The finding of substantial PCA differentiation between eastern and western populations and lack of PCA differentiation among temporal population samples is driven by many more significant allele frequency differences between eastern and western populations, regardless of the time period, than between temporal samples (Figure 2.6). These patterns match those observed in Shultz et al.'s (2016) ddRADseq study, and are consistent with the idea that genetic drift during population bottlenecks have much more profound genome-wide consequences than selection events.

Site-specific signatures of an introduction

Disentangling allele frequency differences occurring as a result of genetic drift and selection is challenging in a recently bottlenecked population (Thornton et al. 2007; Poh et al. 2014; Lotterhos and Whitlock 2014). We detected over 5,000 sites with significantly different allele frequencies between the pre-epizootic native and introduced populations, or 0.032% of sites. This is substantially fewer than the 1.7% of sites detected using ddRADseq (Shultz et al. 2016). This discrepancy is likely a result of the much more stringent false discovery rate necessary to correct for ~15 million comparisons made with whole genome data to the ~12,000 comparisons made with the ddRADseq data. The ddRADseq data also may have been more prone to false positives due to missing data (Arnold et al. 2013). In the whole genome data set, genetic drift is likely responsible for many significant differences, but some may be due to selection for the new environment in the introduced population or the relaxation of selection present in the native population. To obtain an expectation for the proportion of significant sites due to genetic drift, we will need to correctly model the demographic history of introduction event, incorporating both the history of the native population as well of that of the introduced population (Jensen et al. 2016).

Sites with significant allele frequency shifts are not distributed randomly throughout the genome, but are spatially clustered, with one scaffold having a large number of sites in both spatial comparisons (Figure 2.7). A further look at scaffold 24, shows that significant sites are further clustered within scaffolds and may represent linked regions (Figure 2.8A,B). This pattern is not observed in all scaffolds (e.g. Figure 2.8C). Analyses of linkage disequilibrium will allow us to identify how many independent regions of the genome are significantly different, and to help distinguish genetic drift from selection. Analyses incorporating linkage are more powerful to detect selection during a bottleneck than SFS-based methods (Poh et al. 2014). This is true even though soft sweeps are predicted to be more prevalent than hard sweeps during a founder event (Kim and Gulisija 2010), since soft sweeps may "harden" during a founder event (Wilson et al. 2014). Although we have not yet measured it with this whole-genome dataset, previous estimates of House Finch linkage using ddRADseq data (Shultz et al. 2016) or candidate loci (Backström et al. 2013a) show very short LD blocks, with LD falling off between 100 and 500 bases. Some of the clustered sites showing significant allele frequency changes appear to be several kilo bases long, and could be extended haplotype blocks, but will need to be formally quantified in the future.

Signatures of selection as a result of the epizootic

While soft sweeps may have contributed to the evolution of phenotypic traits in the eastern population, our results suggest that soft sweeps coupled with polygenic selection are responsible for the evolution of resistance and tolerance to MG. If hard sweeps, or strong selection at a small number of loci caused the phenotypic changes, we would have expected to observe a few regions in the genome along with linked variants changing through time (e.g. 12 regions observed in temporal comparisons of human populations; Mathieson et al. 2015). We identify two single significant SNPs in the eastern 1990 and 2001 comparison, but these are not located in coding regions, and the lack of any linked sites suggest that they may be false positives. This result is consistent with the hypothesis that adaptation occurs primarily by polygenic selection (Pritchard and Di Rienzo 2010; Wellenreuther and Hansson 2016), especially in cases of rapid adaptation (Messer and Petrov 2013). However, if polygenic selection was pervasive throughout the genome, as has been suggested for Drosophila in response to seasonal selection (Bergland et al. 2014), we would have expected to have observed some temporal structuring. Polygenic selection is difficult to detect with traditional outlier tests (Pritchard and Di Rienzo 2010), but modeling techniques, including Random Forest algorithms, hold promise for identifying loci under polygenic selection (Wellenreuther and Hansson 2016). This method has been used to demonstrate polygenic adaptation to freshwater and brackish habitats in the American eel (Pavey et al. 2015) and in spawning strategies of the Atlantic salmon (Bourret et al. 2014). Methods consolidating SNP changes over functional gene pathways has also been useful for detecting polygenic selection in humans (Daub et al. 2013). Using these methods with our temporal sampling data may help us to detect epizootic-mediated selection.

In addition to polygenic selection, there are other possibilities to explain the inability to detect regions of the genome under pathogen-mediated selection. First, it is possible that the reason we could not detect single SNPs under selection in the entire eastern population is that selection occurs in similar regions of the genome, but at different SNPs in different localities. To account for this possibility, temporal studies of *Brassica rapa* populations before and after a drought (Franks et al. 2016), or Tasmanian devils before and after an epizootic (Epstein et al. 2016), identified a number of genomic regions differentiated through time in parallel windows across localities. Thus, a sliding window approach using localities rather than regional populations could help detect allele frequencies that have shifted through time. A second possibility is that House Finches evolved resistance and tolerance at a MHC locus, as has been observed in temporal comparisons of human populations before and after a smallpox epidemic (Lindo et al.

2016) as well as associated to other pathogens in human populations (Hill et al. 1991; Hill 1998). One study of House Finch MHC class II loci did detect a change in heterozygosity through time (Hawley and Fleischer 2012), but signatures of selection were not found by a different study of the same locus (Hess et al. 2007). Complicated gene families like MHC are likely misassembled in our reference genome, and would have been filtered out of our resequencing data due to high coverage. Additional sequencing of MHC loci to much higher coverage would be necessary for their correct characterization and the ability to detect any temporal changes (Dilthey et al. 2015). Third, it is possible that eastern populations may have preserved genetic diversity through the epizootic by immigration, which may have obscured signatures of selection (e.g. Sackett et al. 2013). However, a recent study in honeybees identified hundreds of signatures of selection throughout the genome before and after exposure to a new pathogen, and inferred that the mechanism for this change was largely immigration (Tin et al. 2015). The lack of temporal differentiation in our dataset suggests that this would not be the case in House Finches, or that movements were not extensive enough to obscure regional differentiation. Finally, our association approach to detect allele frequency change through time may be too simplistic or incorrectly modeled. Numerous methods to detect selection from time serial data, both from experimental evolution experiments and from ancient DNA have been recently developed (Malaspinas 2015). These include methods to take into account multiple loci simultaneously in a temporal framework (Nishino 2013), and methods to take into account both spatial and temporal fluctuations in selection (Gompert et al. 2016), and may prove useful for the analysis of this dataset.

Conservation implications and future prospects

Anthropogenic disturbances can have marked ecological impacts on biodiversity, but less is known about the genomic impacts these disturbances might have on plants and animals. Here, we have used whole genome resequencing to characterize the genomic impacts of two anthropogenic disturbances within and among wild bird populations. By incorporating robust sampling across space and time, we show that demographic events, such as bottlenecks, have much stronger signatures on the genome than selection events, even those that are likely strong. These findings have important implications for conservation practices. Whether considering population translocations or identifying which populations to actively manage, we show that populations in danger of experiencing a bottleneck, even a short bottleneck if sufficiently severe, may warrant management over species that are experiencing new selection regimes, including climate change and novel pathogens. An exception would be when the selection event itself results in a bottleneck without subsequent population expansion, which may result in “genomic meltdown” (Rogers and Slatkin 2017). Although it may be alarming to observe many individuals of a species with a large effective population size succumb to disease, the most pervasive patterns of selection evident throughout the bird tree of life are associated with pathogen-mediated selection (Chapter 3), suggesting that this may be an occurrence more common than previously appreciated.

We also show that the demographic history of source population may have large impacts on the outcome of demographic events. It has long been assumed that the overall genetic diversity of a species is correlated with fitness, and thus is an important predictor of the health of that species or population (Reed and Frankham 2003). If we had looked only at the introduced eastern population, it may have been difficult to detect the bottleneck that occurred during the founder event. The signature of population expansion in the western population was so prevalent that the bottleneck produced a SFS in the eastern population that looked similar to the neutral expectation (e.g. Figure 2.3F). Future work testing demographic models that incorporate the source population as well as models that use the bottlenecked population alone will help to identify if demographic changes due to anthropogenic disturbances can be easily detected in single populations without a historical context. Future work would also be useful to explore how the demographic history of a source population might affect the chance of a successful introduction. For example, a source population with a demographic history of population expansion may improve the potential success of an introduced population if the bottleneck during the founding allows the introduced population to purge some of the deleterious alleles. Alternatively, the introduced population may have decreased success because there could be a greater chance that deleterious mutations from recent growth could drift to higher frequencies during the founding event. We may be able to untangle these possibilities in the future for this species by assessing deleterious mutational load in both populations, and possibly by including a Hawaiian population that had a more extreme bottleneck (Shultz et al. 2016).

Our whole genome dataset and temporal sampling is ideally suited to detect signatures of selection as a result of the MG epizootic (Jensen et al. 2016). Here we present a fairly simplistic selection scan, but the characterization of LD in the genome and incorporation of sophisticated polygenic and temporal models may enable us to detect the MG-mediated selection that resulted in the increased resistance and tolerance observed with experimental infections. It is interesting to note that it may be easier to detect selection in experimental populations than natural populations. By experimentally manipulating the variables of interest, it may eliminate the considerable noise associated with genome scans. However, many experimental studies, including those with the House Finch (Bonneaud et al. 2011; Adelman et al. 2013), do not actually measure selection, but only measure gene expression or phenotypic differences. Thus, combining data from both natural and experimental studies, as we have the opportunity to do with this system, could yield insights out of reach for either study alone.

CHAPTER 3: EVOLUTION OF THE INNATE IMMUNE SYSTEM ACROSS THE BIRD TREE OF LIFE

Co-authors: Julia Yu, Timothy B. Sackton

INTRODUCTION

Rapid evolution in hosts and counter-adaptation by pathogens, when played out over evolutionary time scales, is a classic example of the "Red Queen" dynamic (Salathe et al. 2008) and can have profound impacts on the genomes of both interactors. Infectious disease has been recognized as one of the most important factors shaping genetic variation in human populations (Fumagalli et al. 2011; Karlsson et al. 2014) and across comparative studies on primates, mammals, bees, ants, *Drosophila* and other organisms (Schlenke and Begun 2003; Sackton et al. 2007; Barreiro and Quintana-Murci 2009; Roux et al. 2014). For example, in mammals, proteins that interact with viruses experience about twice as many amino acid changes compared to proteins that do not (Enard et al. 2016).

The innate immune system provides the first line of defense against pathogens, including genetically encoded pathogen recognition receptors, and a wide variety of general defensive cell types like dendritic cells, macrophages and neutrophils (Iwasaki and Medzhitov 2010). The adaptive immune system, built primarily on the activities of T and B lymphocytes, includes responses specific to individual pathogens, but several of the antigen receptors that mediate such responses are not genetically encoded; rather, in the case of T-cell receptors and immunoglobulins, sequence specificity is generated de novo following exposure to a pathogen (Iwasaki and Medzhitov 2010). The innate immune system evolved in early metazoans and many pathways are conserved in both arthropods and vertebrates (Buchmann 2014; Palmer and Jiggins 2015). The adaptive immune system, including MHC genes and T-cells, arose in early vertebrates, and is maintained exclusively throughout that clade (Buchmann 2014).

Although innate immune system pathways are conserved across animals, there is a lack of evolutionary studies across entire pathways outside of some selected invertebrate groups. From these studies, receptor genes, or the genes interacting directly with pathogens and are hypothesized to exhibit adaptive evolution, are most often the target of positive selection (Sackton et al. 2007; Waterhouse et al. 2007; Ellis et al. 2012). Because of potential redundancies in the action of the innate and adaptive immune system in vertebrates, an adaptive immune system may alter the evolutionary dynamics of the host-pathogen arms race and lead to different patterns of selection across the innate immune system (Wlasiuk and Nachman 2010). For example, the specific adaptive immune responses possible with the adaptive immune system may potentially reducing the selective pressure for innate immune receptor genes to broadly respond to diverse pathogens. Studies of toll-like receptors (TLRs) in vertebrates found evidence for positive selection (Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Grueber et al. 2014), but the generality of this pattern across other innate immune receptor genes is unclear. Effectors include proteins including antimicrobial peptides that inhibit pathogen survival and reproduction (Medzhitov 2007). These genes typically have low levels of positive selection in *Drosophila* (Sackton et al. 2007) and other insects (Viljakainen 2015). In birds, β-defensins display signals of balancing or purifying selection (Chapman et al. 2016).

Signaling genes mediate and coordinate receptor and effector genes in immune pathways (Buchmann 2014). In insects, there is evidence of positive selection in some signaling genes (Viljakainen 2015), particularly in genes that module signal rather than just transduce it (Sackton et al. 2007). In birds, a population genetic study of chickens found evidence for positive selection on TLRs (receptors) and frequency-dependent selection on cytokines (signaling genes) (Downing et al. 2010), but in humans some signaling genes show evidence of purifying selection, while other show evidence of positive selection (Fornarino et al. 2011; Quintana-Murci and Clark 2013).

To assess the generality of these patterns, it is necessary to perform a genome-wide scan for positive selection in a uniform set of species. Few comparative genomic studies search for signals of positive selection at the same set of codons across all lineages in a clade; most often they focus on detecting positive selection on a few species or populations. Pathogens could potentially drive evolution in different molecular systems in different lineages, and therefore could weaken signals of positive selection across many lineages in a group (Sironi et al. 2015; Viljakainen 2015). For example, Webb et al. (2015) detected species-specific positive selection in mice and humans in different classes of immune genes, suggesting that selection pressures by pathogens were idiosyncratic across these two mammals. On the other hand, pathogens could target homologous molecular pathways in different lineages, a scenario that could strengthen signals of pathogen-mediated positive selection across whole clades (Sironi et al. 2015). Recent work has highlighted the importance of polygenic selection in adaptation (Pritchard and Di Rienzo 2010; Daub et al. 2013), emphasizing the need to incorporate functional gene pathway analyses into comparative genomic scans for selection.

Birds (class Aves) are an evolutionary and ecologically important group of animals. A radiation of approximately 10,000 species (Clements et al. 2016), they possess diverse morphologies and behaviors (Gill 2007). They have a global distribution and diverse range of habitats (Jetz et al. 2012), and many species migrate thousands of miles annually (Gill 2007), making them excellent models for studies of disease ecology. From a genomic perspective, they have small genomes, generally stable chromosomes, little repeat content, and low rates of gene loss and gain (Organ et al. 2007; Organ and Edwards 2011; Zhang and Edwards 2012; Ellegren 2013; Zhang et al. 2014b). Because of the economic importance of the poultry industry and birds as reservoirs for potentially infectious disease, we know a substantial amount about the avian immune system (Kaiser 2010). Birds have the same general blueprint of immune pathways as mammals, but with a slimmed down gene repertoire and some small differences in the functions of specific genes (Kaiser 2010; Chen et al. 2013; Juul-Madsen et al. 2014). Studies of the evolutionary dynamics of avian immune genes have almost exclusively focused on the major histocompatibility complex genes (MHC) or TLRs, with evidence of positive selection across species in MHC class I genes (Alcaide et al. 2013), MHC class II genes (Edwards et al. 1995; 2000; Hess and Edwards 2002; Burri et al. 2008; 2010) and TLRs (Alcaide and Edwards 2011; Grueber et al. 2014).

The recent influx of genomes from species representative of the bird radiation (Zhang et al. 2014b; Kapusta et al. 2017) provides an opportunity to study innate immune system dynamics using comparative genomics. First, we use existing immune gene annotations in birds to study the positive selection in three immune gene categories – receptor genes, signaling genes, and effector genes. We hypothesize that the patterns of positive selection in these three categories are similar to those observed in invertebrates given the positive selection observed for TLRs, and purifying selection observed for β-defensins. Second, we study positive selection without any *a priori* assumptions as to gene function. We take all signatures of positive selection we detect in the genome and use pathway enrichment to test for the tendency for immune gene pathways to exhibit signatures of positive selection across bird lineages. Overall, we find strong signatures of positive selection in receptor genes, at a higher proportion of sites and lineages, and enrichment in immune functional pathways for positively selected genes. Together, these results suggest that host-pathogen interactions have had strong effects in shaping bird genomes.

METHODS

Identification, alignment and filtering of avian orthologs

We used the program OMA (Roth et al. 2008; Altenhoff et al. 2013) to infer patterns of homology among protein-coding genes across sequenced bird and non-avian reptile genomes. We selected 29 existing bird annotations from NCBI in addition to 10 paleognath gene sets produced by Sackton et al. (in prep) (Figure 3.1). For each gene set, we selected the longest transcript to represent that protein in our homology search.

We then ran OMA v.1.0.0, with default parameters. Once OMA had completed, we checked and improved the annotated homology groups using HMMs. We started by building alignments for each homologous group defined by OMA using MAFFT v.7.221 (Katoh and Standley 2013), and then built HMMs for each protein alignment using HMMER (Johnson et al. 2010). We then searched each hmm against the full set of input proteins to OMA, to first verify that proteins assigned to a homology group are recovered by searching with the HMM built from that group, and secondly to assign unassigned proteins as best as possible. We parsed the results of the HMM search to add graph edges between HOG(a) and HOG(b) if multiple proteins in HOG(a) had best hmmsearch hits to HOG(b), and vice versa (but allowing connections with only a single hit for HOGs with < 20 proteins). We then parsed the resulting connectivity graph to extract weakly connected components using a custom Python script. These subgraphs are combined into new HOGs (code: https://github.com/tsackton/ratite-genomics/blob/master/homology/process_hmm_output.py). This produced a new set of homologous groups, which we use in the following analyses.

These 45,151 hierarchical orthologous groups, or HOGs, were filtered to retain 16,151 HOGs sequenced in at least four species. Protein sequences were aligned with MAFFT v. 7.245 (Katoh and Standley 2013), and filtered in three steps. First, we excluded entire columns with excess gaps, second we masked poorly aligning regions according to Jarvis et al. (Jarvis et al. 2014), and third we again removed alignment columns after masking. Entire sequences were then removed from each alignment if they were over 50% shorter than their pre-filtered length or contained excess gaps. Finally, HOGs were removed if they contained more than three sequences for any species and did not have more than 1.5x sequences for the given number of species present in the alignment. Finally, nucleotide sequences for all remaining HOGs were aligned with the codon model in Prank v. 150803 (Loytynoja and Goldman 2008). In total, 11,271 HOGs remained after all alignment and filtering steps.

Guide trees for use in the tests of selection were constructed for each alignment with RAxML v. 8.1.4 (Stamatakis 2014) under a GTR+GAMMA substitution model, partitioned into codon positions 1+2 and 3, with 200 rapid bootstrap replicates and a maximum likelihood tree search. In cases where species had more than one sequence in the alignment, we included all copies to produce a gene tree for that HOG.

Full methodological detail for the identification, alignment and filtering of avian orthologs can be found in Sackton et al. (in prep).

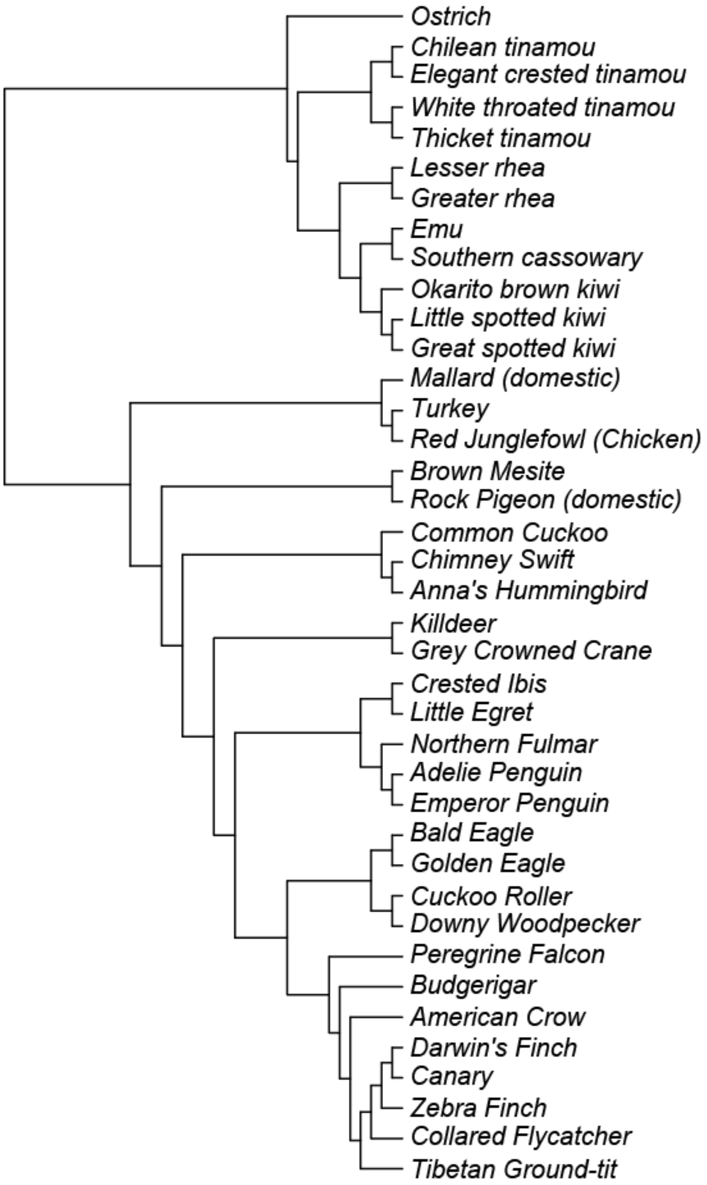


Figure 3.1: Species tree of all birds included in the study. The topology of the tree was constructed using the phylogeny of Sackton et al. (in prep) for the Paleognathae, with the Neognathae placed according to Jarvis et al. (2014).

Tests of selection

To identify positively selected HOGs, we compared models of nearly neutral evolution to those that included signatures of positive selection at specific sites across lineages in the avian phylogeny. To do this, we identified sites with elevated nonsynonymous/synonymous substitution ratios (ω = d_N/d_S), as the expectation under neutral evolution is ω = 1. We used two different programs to identify HOGs with evidence for elevated ω values at specific sites across avian lineages. We used the site models (Nielsen and Yang 1998; Yang et al. 2000) implemented in the program Phylogenetic Analysis by Maximum Likelihood v4.8 (PAML; Yang 2007) to calculate likelihood scores and parameter estimates for seven models of evolution (Table 3.1). Because some HOGs contained gene duplicates, we used the species tree from Sackton et al. (in prep) as the phylogenetic hypothesis if no duplications were present, but the gene tree if any duplication were present in any lineages as described above. First, we fit the M0 model, which estimates a single ω for all sites in the alignment. We used the branch lengths estimated with the M0 model as fixed branch lengths for subsequent models to decrease computational time. To identify HOGs with evidence of positive selection, we conducted likelihood ratio tests between neutral models and selection models (models with ω > 1). We compared likelihood scores from the M1a vs. M2a, M2a vs. M2a_fixed, M7 vs. M8, and M8 vs. M8a models (Table 3.1) (Nielsen and Yang 1998; Yang et al. 2000; Wong et al. 2004). We computed p-values according to a χ^2 distribution with two, one, two, and one degree of freedom respectively. In addition to likelihood scores, we extracted the ω parameter estimate and proportion of selected sites from the M2a and M8a models for use in downstream analyses.

In addition to the site tests implemented in PAML, we used BUSTED (Murrell et al. 2015), a modeling framework implemented in the program HyPhy (Pond et al. 2005), to identify HOGs with evidence of positive selection at a fraction of sites. BUSTED uses a model that allows branch-to-branch variation across the entire tree (Murrell et al. 2015). Similar to the PAML models, BUSTED uses a likelihood ratio test to compare a model including selection ($\omega > 1$ at a proportion of sites) with one that does not. For both sets of tests, to correct for multiple testing, we applied a false discovery rate of 0.05 with the Benjamini-Hochberg approach (Benjamini and Hochberg 1995) with the p.adjust function in the stats packages in R v.3.3.3 (R Core Development Team 2008). We considered a corrected p-value less than 0.05 as evidence for positive selection in that HOG for a given model comparison.

Finally, in addition to testing for selection at particular sites across bird lineages, we used the BS-REL method in HyPhy with default parameters (Kosakovsky Pond et al. 2011) to detect which specific lineages showed evidence of selection for each HOG. For each lineage, including both tip species and internal branches, BS-REL estimates a p-value for the presence of positive selection. We considered both the raw p-value as well as a p-value corrected for multiple testing within each HOG. Fewer lineages showed evidence of selection with a FDR corrected p-value, but overall results presented below were qualitatively consistent with both sets of tests. For simplicity and because the stringent correction may remove biologically-interesting lineages with weak to moderate selection, we present the results using the number of lineages considered nominally significant without multiple-test correction.

Table 3.1: PAML Model descriptions.

model	model description	parameters
M0	one ratio	ω
M1a	neutral	p0 (p1 = 1 - p0) ω0 < 1, ω1 = 1
M2a_fixed	neutral	p0,p1 (p1 = 1 - p0 - p1) ω0 < 1, ω1 = 1, ω2 = 1
M2a	selection	p0,p1 (p1 = 1 - p0 - p1) ω0 < 1, ω1 = 1, ω2 > 1
M7	neutral (beta distribution)	p, q
M8a	neutral (beta distribution)	p0 (p1 = 1 - p0), p, q, ωs = 1
M8	selection (beta distribution	p0 (p1 = 1 - p0), p, q, ωs > 1

HOG annotation

We annotated HOGs for downstream enrichment analyses using chicken (*Gallus gallus*) NCBI gene ids. Of the 11,271 HOGs, 10911 had a chicken sequence present in the alignment, and could be assigned to a gene id.

To compare signatures of positive selection among categories of innate immune genes, we conducted a literature search to identify avian innate immune genes. First, we compiled a list of 1,026 genes with chicken ortholog Ensembl IDs from the mammal-focused Innate Immune Database (InnateDB; Breuer et al. 2012). While many genes in this database have been experimentally validated for immune function in human, mouse or cow, there are some differences between the avian and mammalian immune systems that may cause differences in interpretation or analysis (Kaiser 2010). Thus, we identified 234 innate immune genes from this list that have been specifically examined in birds (Wigley and Kaiser 2003; Kaiser 2010; Schat et al. 2012; Sokol and Luster 2015). These genes were then classified into three different subclasses, 84 ‘receptor genes’, 71 ‘signaling genes’, and 79 ‘effector genes’ (Buchmann 2014). Receptor genes, or ‘pathogen recognition receptors’ (PRRs) detect pathogen-associated molecular patterns (PAMPs) and include proteins such as toll-like receptors (TLRs), NOD-like receptors (NLRs), and RIG-like receptors (RLRs) (Juul-Madsen et al. 2014). Signaling genes encode proteins that mediate interactions between elements of a pathway, including cytokines and chemokines (Sokol and Luster 2015). Effector genes encode molecules that regulate, kill or otherwise clear pathogens from host tissues, and include antimicrobial peptides (AMPs) and components of the complement cascade (Juul-Madsen et al. 2014).

We converted the chicken ENSEMBL gene IDs to chicken NCBI gene IDs for innate immune genes with the R biomaRt package version 2.30.0 (Durinck et al. 2005; 2009) with the Gallus_gallus-5.0 dataset (Warren et al. 2017). We filled in gaps for genes that could not be mapped with this approach by identifying synonymous gene IDs based on genome position in the chicken genome with bedtools (Quinlan and Hall 2010). In total, we were able to identify the NCBI gene IDs for 231 of the 234 avian innate immune genes.

Comparisons of avian innate immune gene categories

We compared signatures of positive selection among innate immune receptors, signaling genes, and effectors. We classified the 8,406 genes analyzed by both PAML and BUSTED as ‘selected’ or ‘not selected’. To examine genes with signatures of selection across bird lineages, we classified genes with signatures of selection in all PAML and BUSTED analyses as ‘selected’. This cutoff is likely conservative, as some of the genes identified by either PAML or BUSTED alone likely contain true signatures of positive selection. However, we felt this list likely identified the genes with strongest signatures and reduced false positives, and preliminary analyses with different sets of selected genes did not qualitatively differ in overall results. For all tests of selection in PAML and BUSTED, we used Fisher's exact tests to compare ratios of 'selected' and 'not selected' (selected = FDR p-values < 0.05) between immune gene categories and all genes (e.g., receptors vs. all genes). We also compared proportions of positively selected genes in the avian orthologs of innateDB genes, and literature-searched avian immune genes. P-values were corrected for multiple testing with a FDR of 0.05 for the five comparisons within each test of selection (e.g. M1 vs. M2). We further investigated whether the observed patterns changed with more stringent requirements as evidence for positive selection. For the M1 vs. M2 tests and M7 vs. M8 tests, we required genes to both have significant p-values, at least 5% of sites with signatures of positive selection, and ω values of at least two. For BUSTED tests, we compared genes with significant LRTs and those with signatures of selection in at least 20% of BS-REL lineages.

Finally we investigated whether immune gene categories differed in ω values, proportions of selected sites, or numbers of selected lineages. We used the Kolmogorov-Smirnov test to compare the distribution of PAML M2 ω values and proportions of significant sites from significant genes between immune categories (receptors vs. signaling, receptors vs. effectors and signaling vs. effectors), and between each immune category and genome-wide estimates. Estimates from the M8 model and BUSTED produced qualitatively similar results and are not presented. We also compared the proportion of significant lineages estimated by BS-REL for those genes identified as under positive selection by BUSTED, among the same categories of genes as the PAML M2 model.

Functional gene pathway enrichment

We looked for patterns of positive selection among groups genes with similar functions using KEGG pathway enrichment tests (Kanehisa and Goto 2000; Kanehisa et al. 2011). We also investigated functional categories of genes with a high proportion of lineages under selection as detected by BS-REL. We divided these into two categories – those that were under selection at the same sites across avian lineages, and those with different sites under selection across avian lineages. To accomplish this, we created two additional classes of genes under selection: those selected in at least 20% of lineages with significant BUSTED results and those selected in at least 20% of lineages with no significant BUSTED results. The 20% cutoff allows us to detect enough genes in each class to run enrichment tests, but preliminary sensitivity tests (not shown) with cutoffs of 10%, 15%, 20%, and 25% showed qualitatively similar results.

For the three sets of analyses, we used the 'enrichKEGG' command from clusterProfiler v. 3.2.14 (Yu et al. 2012) from Bioconductor v. 3.4 (Gentleman et al. 2004) with chicken as the organism. We used p-value and q-value cutoffs of 0.05, and the genes included in both PAML and HyPhy analyses with NCBI gene IDs (N = 8,273) as the gene universe for enrichment tests with genes selected in all tests. For the two enrichment tests with genes selected in particular lineages with BS-REL, we used and the genes included in both HyPHy analyses with NCBI gene IDs (N = 8,456) as the gene universe. We visualized the results used the 'dotplot' and 'cnetplot' commands in clusterProfiler, and the 'pathview' command in the pathvew Bioconductor package v. 1.14.0 (Luo and Brouwer 2013).

RESULTS

Strong signatures of positive selection throughout the avian genome

In total, we obtained results from all four PAML tests of selection for 10,852 genes out of the 10,911 that could be assigned to NCBI gene IDs. With HyPhy, we obtained results from 8,605 genes with both BUSTED and BS-REL. The difference in the number of genes included in tests of selection is due to the inability of BUSTED to process certain sequence names, and future work will fix this discrepancy. The combined dataset for both PAML and HyPhy results consisted of 8,273 genes. The mean ω value for the 10,852 genes analyzed by PAML (from the M0 model) was 0.147, the median was 0.101 and standard deviation was 0.141 (Figure 3.2).

For all individual tests, between 19.3% and 72.5% of genes were positively selected (Table 3.2). About 20% of genes were positively selected with the M1a vs. M2a test, M2a vs M2a_fixed test, or both, with large overlaps among the genes identified. The less conservative M7 vs. M8 tests showed much higher levels of positive selection (~70%), although this was reduced to about 35% with the M8 vs. M8a test or combination of the two, indicating that the M8 model may often improve fit by adding a class of sites with ω very close to 1. Most of the genes identified by the various M1 vs. M2 tests were also identified by the M7 vs. M8 tests (Table 3.2). BUSTED identified 32.6% of genes as positively selected. Fewer than half of these were also identified as being positively selected by all PAML tests, resulting in 864, or 10.4% of analyzed genes found to be under positive selection in all tests. Genes identified by BUSTED as having sites with evidence of positive selection across avian lineages implicated significantly more lineages under selection as compared to BS-REL (K-S test: D = 0.341, $p<10^{-16}$, Figure 3.3).

Previous work has found that alignment errors can result in substantial false positives (Markova-Raina and Petrov 2011). However, our strict alignment filtering strategy and use of the evolution-aware PRANK aligner would have minimized the possibility that our results are solely false positives (Markova-Raina and Petrov 2011). Recombination also can elevate ω estimates, but the M7 vs M8 model has been shown to be robust to recombination (Anisimova et al. 2003), and these results give us the highest proportions of positive selection we observe in our dataset (Table 3.2). Finally, despite observing high proportions of selected genes, the overall trend of gene-wide estimates of ω << 1 are consistent with patterns of purifying selection on coding regions of the genome (Figure 3.2). Furthermore the similarity in estimated ω values between this study and previous studies in birds with different sets of genome sequences or the use of pairwise estimates between chicken and zebra finch (Nam et al. 2010; Zhang et al. 2014a) give us confidence that our results are robust.

Table 3.2: Counts and proportions of genes with evidence of selection for single tests and combinations of tests.

dataset	# genes	m1a vs m2a	m2a vs m2a_fixed	m1a vs m2a and m2a vs m2a_fixed	m7 vs m8	m8 vs m8a	m7 vs m8 and m8 vs m8a	m2a vs m2a_fixed and m8 vs m8a	all PAML	BUSTED	all PAML + BUSTED
PAML only	10852	2094	2382	2086	7683	3965	3875	2337	2062	-	-
HyPHY + PAML	8273	1651	1878	1644	5994	3168	3097	1840	1622	2694	864
% PAML only	-	19.3%	21.9%	19.2%	70.8%	36.5%	35.7%	21.5%	19.0%	-	-
% HyPHY + PAML	-	20.0%	22.7%	19.9%	72.5%	38.3%	37.4%	22.2%	19.6%	32.6%	10.4%

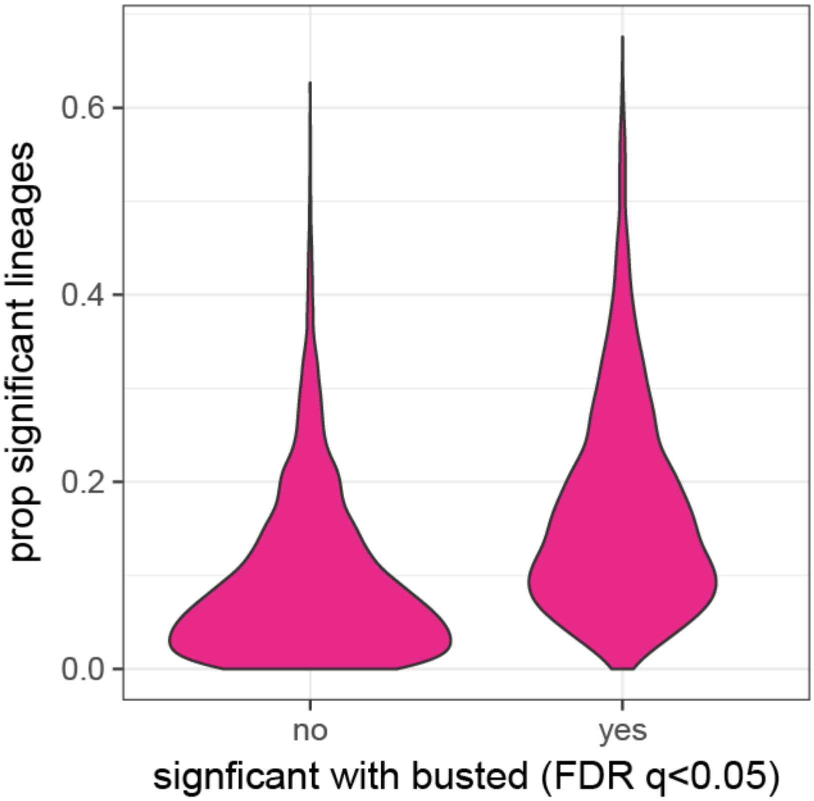


Figure 3.3: Proportion of bird lineages with significant tests of positive selection for HOGs identified as not significant (FDR p-value >= 0.05), or significant (FDR p-values < 0.05) for positive selection across all birds as tested with BUSTED.

Innate immune receptor genes show strong signals of positive selection

The strong signals of positive selection we observed in avian genes were not uniformly distributed among innate immune gene categories. Among innate immune gene categories, receptor genes showed significantly higher proportions of selected genes across almost all tests (Table 3.3), although the proportions in signaling and effector genes under selection were higher than the results for all genes, if not significantly so for most tests. These weak signatures of selection for the signaling and effector genes combined with the strong signatures for the receptor genes mirrors the distribution of selected avian innate immune genes identified by our literature search (described above). Innate immune genes from the InnateDB did not show patterns of positive selection above genome-wide averages. This pattern might arise because the InnateDB may include genes with secondary immune functions. Additionally, genes without direct host-pathogen interactions (most non-receptor genes) may be less likely to be under selection in specific sites in many bird lineages. However, the InnateDB is also based entirely on mammalian evidence, and the lack of a selection signature of innateDB genes could indicate that more work is required to elucidate the broader functions of avian genes. Only the M7 vs. M8 test did not show any evidence of positive selection in any category of immune genes, most likely due to the very high proportions of genes with significant likelihoods of positive selection as discussed above.

When only considering genes with evidence of strong positive selection and with proportions of selected sites > 0.05 and ω > 2, we observe patterns similar to those observed without these constraints. However, it is worth noting that the genome-wide average ω drops from ~20% with the M1 vs. M2 test and ~80% with the M7 vs. M8 test to

~3% and ~4% respectively, under these more stringent criteria. As a result, the proportions of positively selected genes were much more similar among immune categories with these stringent criteria. The proportion of selected effector genes also drops below genome-wide averages, suggesting that the higher proportions observed with less stringent criteria were comprised primarily of genes with very weak selection or with selection at a very small number of sites. Among tests requiring selection in at least 20% of lineages, only receptor genes show a significantly higher proportion of selected genes than the genome-wide proportion.

Under the stringent criteria, there was no difference in ω values among immune genes categories or between immune genes categories and genome-wide values (K-S test; all $p > 0.05$; Figure 3.4A). Proportions of selected sites were similar in receptor genes as compared to signaling genes (K-S test: $D = 0.27$, $p = 0.58$), significantly higher in receptor genes than effector genes (K-S test: $D = 0.53$, $p = 0.026$), and marginally significantly higher in signaling genes than effector genes (K-S test: $D = 0.053$; $p = 0.058$) (Figure 3.4B). Compared to genome-wide estimates, receptor genes have significantly higher proportions of selected sites (K-S test: $D = 0.35$, $p = 0.0073$), signaling genes have significantly higher proportions of selected sites (K-S test: $D = 0.039$, $p = 0.036$), and effector genes have similar proportions of selected sites (K-S test: $D = 0.23$, $p = 0.54$) (Figure 3.4B). Receptor genes had significantly higher proportions of selected lineages compared to signaling genes (K-S test: $D = 0.49$, $p = 0.042$) and effector genes (K-S test: $D = 0.50$, $p = 0.011$), and signaling and effector genes had similar proportions of selected lineages (K-S test: $D = 0.021$, $p =0.89$). Similarly, only receptor genes had significantly higher proportions of lineages than all genes (K-S Test: Receptor: $D=0.38$, $p = 0.0053$, Signaling $D = 0.18$, $p = 0.80$, Effector: $D =$

0.16, p = 0.68; Figure 3.4C). We observe qualitatively similar results with the Mann-Whitney U-test for all comparisons, suggesting that the observed differences are not just in the shape of the distribution, but also in the median values.

Table 3.3: Proportions of significant genes in each category of immune genes. Counts of total genes in each category are shown by parentheses, and significant comparisons are bolded.

Category	m1a vs m2a	m2a vs m2a_fixed	m7 vs m8	m8 vs m8a	BUSTED	All tests	m1 vs m2 ω>2 & >5% sites	m7 vs m8 ω>2 & >5% sites	BUSTED >20% BS-REL lineages
Receptor	**0.535 (43)*** **	**0.581 (43)*** **	0.884 (43)	**0.791 (43)*** **	**0.618 (34)* **	**0.433 (30)*** **	**0.233 (43)*** **	**0.186 (43)* **	**0.294 (34)* **
Signaling	**0.382 (34)* **	0.382 (34)	0.735 (34)	0.559 (34)	0.481 (27)	0.261 (23)	**0.176 (34)* **	0.118 (34)	0.111 (27)
Effector	0.316 (38)	0.316 (38)	0.632 (38)	0.447 (38)	0.556 (36)	**0.276 (29)* **	0.026 (38)	0.026 (38)	0.111 (36)
Immune	**0.407 (118)*** **	**0.424 (118)*** **	0.763 (118)	**0.593 (118)*** **	**0.55 (100)*** **	**0.318 (85)*** **	**0.144 (118)*** **	**0.11 (118)* **	0.17 (100)
InnateDB	0.211 (560)	0.238 (560)	0.693 (560)	0.361 (560)	0.371 (426)	0.115 (418)	**0.052 (560)* **	0.055 (560)	0.108 (426)
All	0.193 (10852)	0.219 (10852)	0.708 (10852)	0.365 (10852)	0.328 (8605)	0.104 (8273)	0.032 (10852)	0.039 (10852)	0.100 (8605)

* $p < 0.05$ ** $p < 0.001$ *** $p < 0.0001$

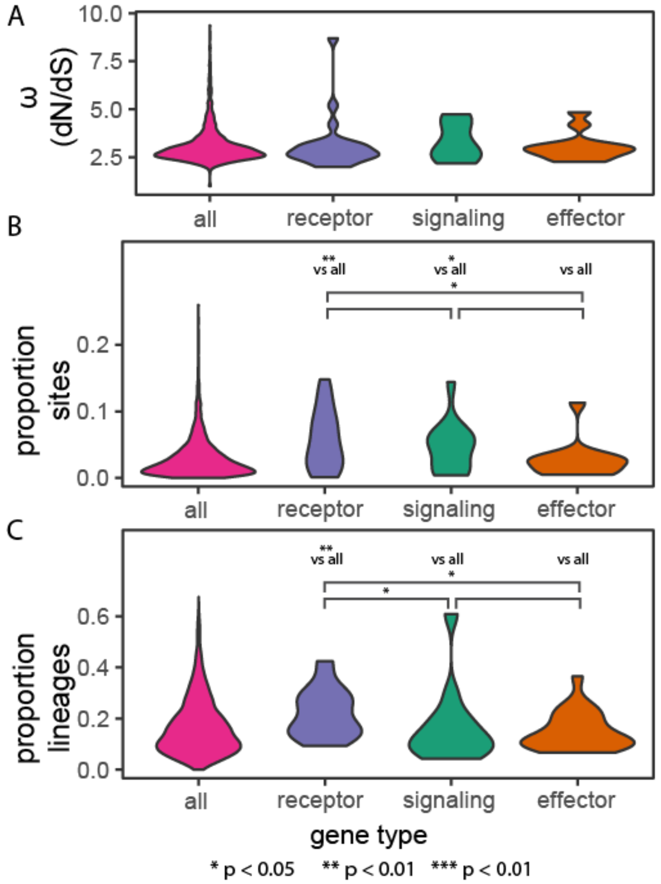


Figure 3.4: Violin plots of all genes, receptor genes, signaling genes, and effector genes for A) ω values for all positively selected genes from the PAML M2 model, restricted to values between and 10, to enhance visualization (there were a few outlier loci in all genes), B) the proportion of significant sites for all positively selected genes estimated by the PAML M2 model, and C), the proportion of significant branches calculated by BS-REL for hogs with significant sites in BUSTED. The significance of pairwise K-S tests is indicated by *, and if the p >= 0.05, no asterisk is present. No pairwise tests were significant for A), so we did not include any brackets.

Immune and recombination pathways enriched for positive selection

For an unbiased perspective on whether or not positively selected genes are concentrated in particular functional pathways, we performed a pathway enrichment test of positively selected genes. Of the 864 genes identified as evolving by positive selection with all tests, 208 genes could be mapped to KEGG pathways. For use as the gene universe, 2,499 of the possible 8,273 genes could be mapped to KEGG pathways. Out of the 166 KEGG pathways available for chicken, or any other bird species, 95 had at least 1 gene with evidence of positive selection (Supplemental Table 3.1) in all tests. Of these 95 pathways, 10 were significantly enriched with positively selected genes (Table 3.4; Figure 3.5A). These 10 pathways belonged to five classes of KEGG pathways: infectious disease: viral, infectious disease: bacterial, immune system, replication and repair, and signaling molecules and interaction. Some genes are found in multiple enriched pathways, particularly among those with immune functions or those with functions in recombination, but many genes are uniquely enriched in a single pathway as well (Figure 3.5B).

We observed 836 genes identified as positively selected by BUSTED at specific sites and selected in at least 20% of lineages by BS-REL, of which 199 could be mapped to KEGG pathways. Of the 8,456 genes with BUSTED results, BS-REL results, and NCBI gene IDs, 2,536 could be mapped to KEGG pathways for use in the gene universe. We observed 94 KEGG pathways with at least one gene with evidence of positive selection, and three pathways significant enriched for positively selected genes of the 10 identified by the previous analysis (Table 3.4). These three pathways were also detected by the all gene selection criterion.

We observed 643 genes not identified as positively selected by BUSTED at specific sites, but selected in at least 20% of lineages by BS-REL, of which 143 could be mapped to KEGG pathways. 89 KEGG pathways contained at least one selected gene, but no pathways were significantly enriched.

Table 3.4: Results of enrichment analyses for genes significant in all tests

Gene Set	ID	Description	KEGG Classification	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
sig all tests	gga05164	Influenza A	Infectious disease: viral	17/208	59/2499	2.8E-06	2.6E-04	2.3E-04	17
sig all tests	gga03440	Homologous recombination	Replication and repair	11/208	31/2499	2.0E-05	8.8E-04	7.8E-04	11
sig all tests	gga03460	Fanconi anemia pathway	Replication and repair	11/208	32/2499	2.8E-05	8.8E-04	7.8E-04	11
sig all tests	gga04620	Toll-like receptor signaling pathway	Immune system	11/208	36/2499	9.5E-05	2.1E-03	1.9E-03	11
sig all tests	gga04060	Cytokine-cytokine receptor interaction	Signaling molecules and interaction	19/208	91/2499	1.1E-04	2.1E-03	1.9E-03	19
sig all tests	gga04512	ECM-receptor interaction	Signaling molecules and interaction	9/208	32/2499	8.3E-04	1.3E-02	1.2E-02	9
sig all tests	gga05168	Herpes simplex infection	Infectious disease: viral	13/208	60/2499	9.7E-04	1.3E-02	1.2E-02	13
sig all tests	gga04514	Cell adhesion molecules (CAMs)	Signaling molecules and interaction	11/208	51/2499	2.5E-03	3.0E-02	2.6E-02	11
sig all tests	gga03410	Base excision repair	Replication and repair	6/208	19/2499	3.3E-03	3.6E-02	3.1E-02	6
sig all tests	gga05132	Salmonella infection	Infectious disease: bacterial	8/208	33/2499	4.5E-03	4.3E-02	3.8E-02	8
sig busted bs-rel >0.25	gga05164	Influenza A	Infectious disease: viral	14/199	59/2536	1.2E-04	1.1E-02	1.1E-02	14
sig busted bs-rel >0.25	gga03410	Base excision repair	Replication and repair	7/199	19/2536	3.7E-04	1.7E-02	1.6E-02	7
sig busted bs-rel >0.25	gga04060	Cytokine-cytokine receptor interaction	Signaling molecules and interaction	17/199	92/2536	5.8E-04	1.8E-02	1.7E-02	17

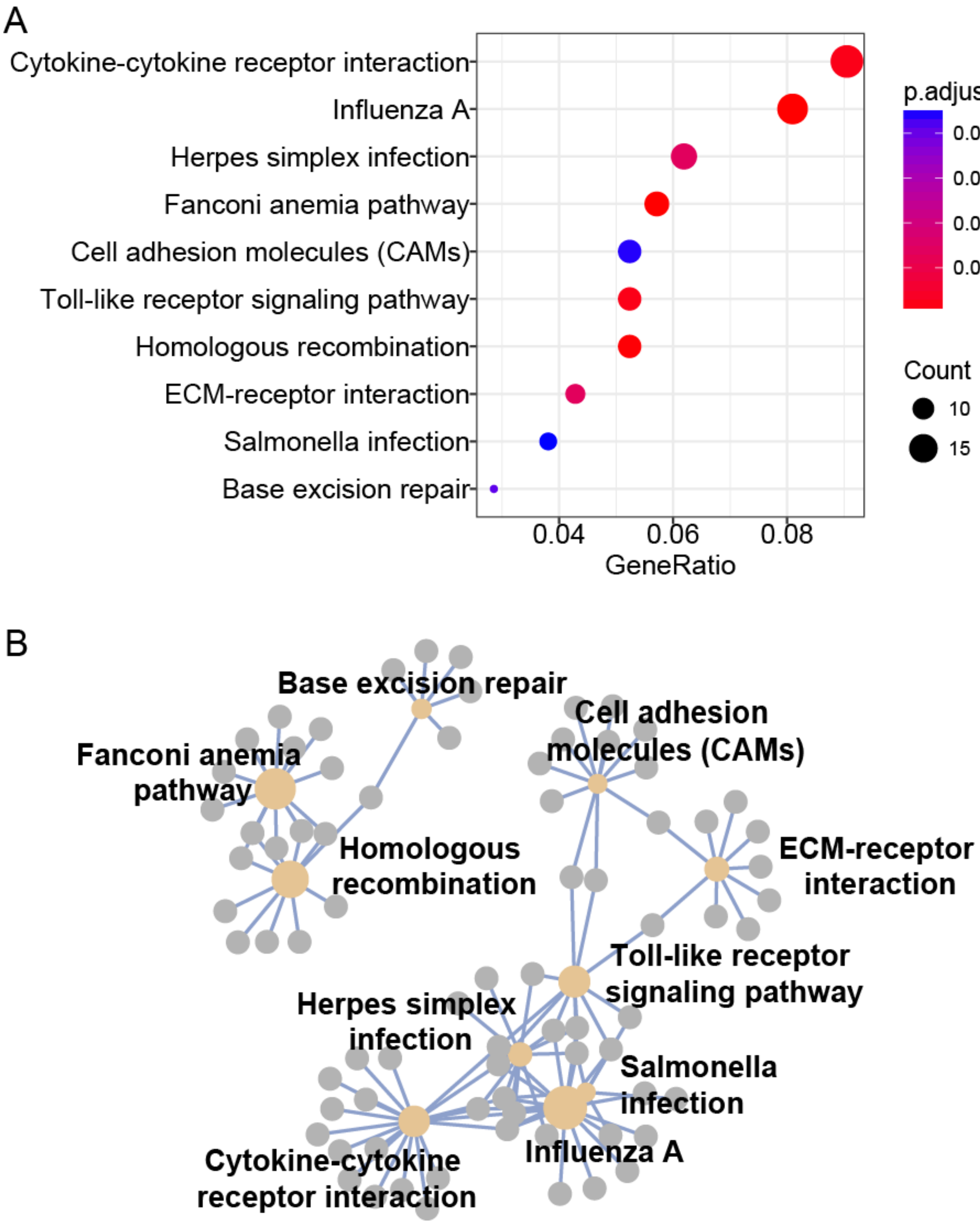


Figure 3.5: A) Plot depicting the significance, ratio of significant genes (gene ratio) and number of significant genes (count) for 10 functional pathways out of 166 possible pathways enriched for 864 genes positively selected with all tests. B) Map showing how each gene (grey circles) is connected to single or multiple pathways (brown circles).

DISCUSSION

Our focus on a wide array of tests for positively selected genes allowed us to use two different approaches to investigate the dynamics of immune gene evolution in birds. First, using *a priori* classifications of bird innate immune genes, we show that receptor genes are consistently under positive selection at a higher proportion of sites and a larger proportion of lineages than genome-wide averages. Second, with an approach blind to *a priori* annotations, we find that immune system functional pathways are enriched for signatures of positive selection. Taken together, our results demonstrate that host-pathogen interactions play an important role in shaping the long-term evolution of the avian genome.

Evolutionary dynamics of functional immune classes

Despite the presence of an adaptive immune system in addition to the innate immune system in vertebrates, we find that innate immune receptor genes are enriched for positive selection compared to genome-wide estimates (Table 3.3), similar to patterns observed in invertebrates (Sackton et al. 2007; Waterhouse et al. 2007). Receptor genes also exhibited a significantly higher proportion of selected sites and proportion of selected lineages compared to genome-wide estimates. The high proportion of lineages under selection at the same sites suggests that pathogens may continuously target the same pathogen recognition receptors throughout the avian tree. However, we will need to examine the distribution of selected lineages to determine whether selection occurs in lineages dispersed throughout the tree, or if particular clades are targeted over and over again.

Despite higher proportions of selected sites in recognition genes compared to genome-wide estimates, the mean proportion per gene is still only 5%. This small percentage means that host-pathogen co-evolution occurs at specific codons, likely where pathogen ligand binding occurs. An analysis of the specific codons and their locations on protein crystal-structures are beyond the scope of this study, but previous studies of avian TLRs have demonstrated that at least half of the codons identified as being under positive selection are in these ligand-binding regions (Alcaide and Edwards 2011). Five of the 10 avian TLRs are present in our dataset, with TLRs 1A, 1B, 2A, and 2B likely filtered out due to their recent duplication. TLR21 may be difficult to sequence as it has the fewest available sequences in previous TLR studies (Alcaide and Edwards 2011; Grueber et al. 2014), and may have been filtered out due to missing data. For the five TLRs in our dataset we observe that the M0 ω (Table 3.3) are similar to those observed by a previous study (Table 3.1 from Alcaide and Edwards 2001 - TLR3: 0.343, TLR4: 0.517, TLR5: 0.482, TLR7: 0.386, TLR15: 0.423). In addition, Alcaide and Edwards (2011), and a follow-up study by Grueber et al. (Grueber et al. 2014) both identified as TLR5 as having the highest proportion of selected sites, a pattern consistent with our dataset. These similarities from independent studies focusing on just a few genes give us additional confidence in the result of our larger analysis. Finally, viral TLRs in mammals show more constraint due to more complex trade-offs to prevent problems with auto-immunity due to the incorrect recognition of 'self' as a pathogen (Wlasiuk and Nachman 2010). This pattern is also observed in our results. TLR3 and TLR7, the viral TLRS (Chen et al. 2013) in birds, both have the lowest overall ω values as well as the fewest selected sites (Table 3.5).

Table 3.5: Results for Toll-like receptors, including the origin of binding ligand (Chen et al. 2013), if significant in all tests (selected by all? - yes or specific tests with significance), ω from the M0 model, ω from the M2 model, proportional of sites estimated by M2 model, and proportion of branches selected by BS-REL.

Gene	hog	origin of ligand	selected by all?	M0 ω	M2 ω	M2 prop sites	bs-rel sel branches
TLR3	25497	viral	yes	0.35	3.0	0.019	0.16
TLR4	18627	unknown possibly bacterial, fungal, bacterial	yes	0.42	3.2	0.033	0.20
TLR5	3054	bacterial	yes	0.46	2.6	0.061	0.13
TLR7	1385	viral	no *not run in BUSTED	0.32	2.8	0.027	0.09

We observe that signaling genes have an excess of positive selection for some tests (Table 3.3) and have a proportion of selected sites comparable to receptor genes (Figure 3.4). Some signaling genes may also evolve under host-pathogen arms race dynamic if they are hijacked by pathogens to prevent immune signaling (Sironi et al. 2015). Additionally or alternatively, the arms race between pathogens and receptors may also spill over onto signaling genes if signaling genes interact directly with receptors, or could be under selection for other non-immune functions (Quintana-Murci and Clark 2013).

Effector genes have the least evidence of positive selection in birds, consistent with studies demonstrating that avian β-defensins are primarily under purifying selection (Chapman et al. 2016). However, the proportion of positively selected effector genes is elevated non-significantly compared to background levels for many tests (Table 3.3). These proportions drop below background levels if we apply filters to only examine strongly selected genes, implying that they might be under weak positive selection we can identify due to the substantial power we have to detect selection in 39 bird genomes. These findings corroborating the hypothesis that antimicrobial peptides may be under weak positive selection in many lineages of animals (Tennessen 2005) and recent evidence of balancing selection for some genes in *Drosophila* (Unckless et al. 2016). It is also important to remember that we are only examining selection for coding sequence with this study, and changes in gene expression could also be under selection in signaling or effector genes (Sironi et al. 2015). Additional work on the evolution of expression differences in these genes might provide important insights into the evolutionary dynamics among immune gene categories.

Here we have used only a single categorization of immune genes – receptors, signaling genes, and effectors. However, as we demonstrate with our pathway enrichment results, we anticipate that our main results would not change if slightly different categorization strategies were used. Previous studies that compared categorization strategies found that they did not change the main conclusions of the study (e.g. Sackton et al. 2007).

Functional gene pathways are enriched for immune function and recombination

When using genes under selection in all tests, we observed 10 KEGG functional pathways enriched for positively selected genes, with functions related to either immune response or recombination and DNA repair. Furthermore, although only three pathways are enriched with genes selected in a high proportion of bird lineages, these categories of genes are both represented (Table 3.4). The lack of enrichment for genes selected in a high proportion of bird lineages but not consistently at the same sites across birds (not significant with BUSTED) could have two explanations. The first could be that BS-REL is identifying to a large number of false positives. Although plausible, we feel this explanation is unlikely. Our results hold true if we only consider lineages to be positively selected with FDR-corrected p-values, or if we vary the cutoff for the proportion of selected lineages to 10%, 15% 20% or 25%. A possible biological explanation is that host-pathogen interactions are constrained to target very specific sites in genes, which is why we observe enrichment in functional categories that can be explained by host-pathogen interactions (discussed in detail below). Without these constraints, genes may be under selection in many different lineages of birds, but there may be much greater variation in biological function. Three studies that performed genome-wide scans for positive selection in specific bird lineages provide evidence for this hypothesis. First, Nam et al. (Nam et al. 2010) compared positive selection acting on three avian lineages, and while 1,751 genes were evolving more rapidly than average in one of these three lineages, only 208 were common to all. Backstrom et al. (2013b) compared signatures of positive selection in two species of galliformes and two species of passerines, and found that only the passerine lineages showed GO enrichment with terms related to fat metabolism, neurodevelopment and ion binding. Finally, Zhang et al. (2014b) found evidence for positive selection in the three vocal-learning bird lineages enriched for neural-related GO terms.

It is worth noting that only 166 functional gene pathways are available for birds, including chicken, our reference species, compared to the 318 available for human (http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=has; accessed 4/8/2017). It is possible that we are missing other pathways that would be enriched for positive selection. However, the observation that immune and recombination pathways are consistent targets of selection is robust to tests with different gene sets (results not shown). Future work to test for enrichment using orthologous mammalian genes IDs with our selected gene sets could identify additional pathways of interest.

Immune system evolution

The strongest signal of positively selected functional pathways is for pathways related to immune system evolution (Table 3.4). Even those pathways categorized under "signaling molecules and interactions" are enriched for immune system genes. The cytokine-cytokine receptor interaction pathway consists of all immune system genes. The cell adhesion molecule pathway is half immune genes and half neural genes, and of the 11 selected genes, only 1 occurs in the neural category. The ECM (extra-cellular matrix) receptor interaction pathway can be targeted by pathogens to adhere and invade a host (Singh et al. 2012). Additional immune pathways that have direct pathogen interactions, the NOD-like and RIG-1-like pathways (Kumar et al. 2011), are nearly significantly enriched (Supplemental Table 3.1). The strong enrichment of the Influenza A pathway, including for genes selected in many bird lineages, reinforces the hypothesis that viruses are important drivers of adaptation in bird genomes, as has been observed with mammals (Enard et al. 2016).

If we examine the Influenza A pathway in detail (Figure 3.6), we observe a tendency for the enriched genes to be concentrated at the host-pathogen interface, consistent with receptor gene enrichment for positively selected genes. However, we also observe genes under intense selection at other positions in the gene pathway, consistent with the above results that some signaling genes may also be under positive selection. Furthermore, we observe that the strength of selection is not evenly distributed among genes, but that some genes appear to be under intense selection (Obbard et al. 2009). We display the proportion of selected lineages for each gene in the Influenza A pathways in Figure 3.6, but similar conclusions can be drawn if we consider the proportion of selected sites or ω values. The strongest signal of positive selection in the Influenza A pathway is not for a receptor gene, but for a signaling gene, TRIF. TRIF, also known as TICAM1, is recruited by TLR3, one of only two viral sensing TLRs in birds (Kumar et al. 2011; Chen et al. 2013; Santhakumar et al. 2017). TRIF then activates a set of molecules that culminate in the activation of IRF7 or NF-κB, transcription factors that induce transcription of type I interferons and pro-inflammatory cytokines (Santhakumar et al. 2017). Although interferon pathways are only beginning to be studied in birds (Santhakumar et al. 2017), viruses are known to inhibit NF-κB (Le Negrate 2011b), and bacteria impair signaling for NF-κB at the IKK complex (Le Negrate 2011a). We observe high levels of selection of NF-κB at a small proportion of sites (M2: ω = 2.9 at 0.5% of sites), but observe the some of the highest signals of positive selection in our dataset with TRIF, which is under selection in 61% of avian lineages (M2: ω = 2.2 at 5.5% of sites). These results suggest that TRIF may be a target for viral or bacterial suppression of host immune responses, and may be an interesting candidate for future studies of host-pathogen co-evolution.

Here we have only characterized a single pathway in detail, but a formal analysis of how network connectivity might constrain or shape patterns of positive selection could provide additional insight into how functional networks evolve in birds (Casals et al. 2011).

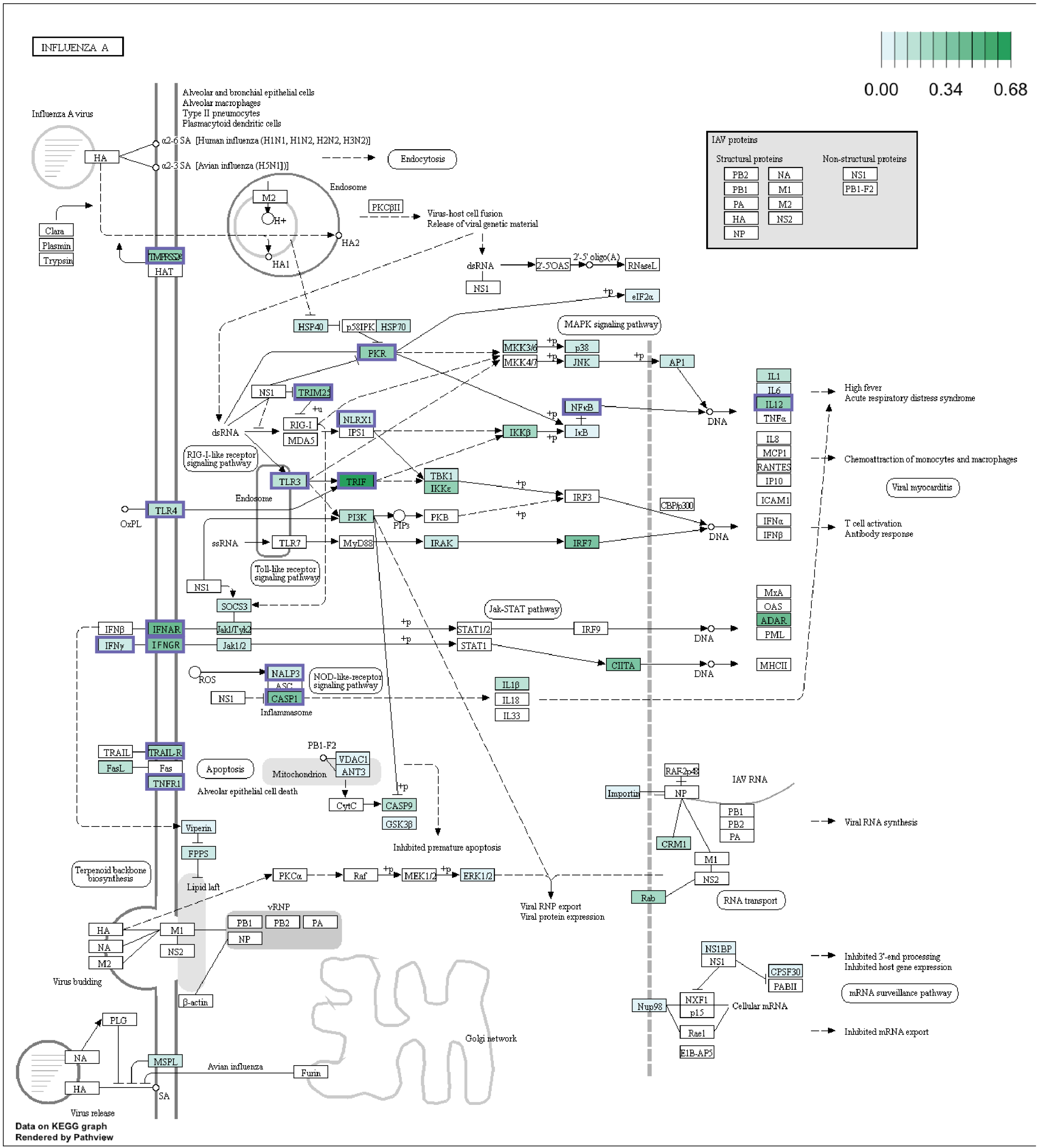


Figure 3.6: Representation of the Influenza A pathway. Each gene is colored according to the proportion of lineages under positive selection with BS-REL, with light green indicating a small proportion of lineages and dark green indicating a large proportion. Lineages under positive selection at specific sites in all tests are outlined by a purple box. Genes colored white were not included in the set of HOGs with results from all PAML and HyPhy.

Chromosomal recombination

In addition to the immune functional pathways, three pathways related to DNA replication and repair were significantly enriched for positively selected genes – homologous recombination, fanconi anemia pathway, and base excision repair. Fanconi anemia is a genetic disease that prevents the removal of DNA cross-links that can block DNA replication and transcription, and uses genes from three DNA repair pathways – homologous recombination, nucleotide excision repair, and mutagenic translation synthesis (Moldovan and D'Andrea 2009). All three pathways promote chromosomal stability and remove damaged DNA bases. Birds are known for their compact genomes that show greater than average chromosomal stability (Zhang et al. 2014b), and a surprising paucity of transposable elements (TEs) (Cui et al. 2014; Zhang et al. 2014b; Kapusta and Suh 2017). One effect on genome structure during the insertion of transposable elements is genome rearrangement due to homologous recombination (Kazazian 2004). Cui et al. (2014) hypothesized that homologous recombination may be responsible for purging transposable elements from the genome, and even observed a galliform hepadnavirus in the process of being removed via homologous recombination. Kapusta and Suh's (2017) observation that the non-recombining W chromosome and regions near centromeres had the highest TE richness also suggest that homologous recombination may prevent the insertion of TEs. Our pathway enrichment results support this hypothesis, and the similar dynamics of positive selection at specific sites as those observed with immune gene pathways suggest that birds may experience a form of host-pathogen co-evolution with TEs.

The process of double-strand break repair, potentially associated with the excision of TEs, can lead to genome-size reductions if biased toward deletions (Schubert and Vu 2016). This model, combined with the evidence for deletions in the ancestral bird lineage (Zhang et al. 2014b; Kapusta et al. 2017), the lack of TEs throughout the bird genome (Cui et al. 2014; Kapusta and Suh 2017), and our observation of strong positive selection for homologous recombination, provide strong support that the mechanism for the evolution of small genomes in birds is host-pathogen co-evolution against TEs (Kapusta et al. 2017). Powered flight and metabolic stress have long been hypothesized as the selective pressure driving this decrease in genome size (Zhang and Edwards 2012; Wright et al. 2014; Kapusta et al. 2017). The strong positive selection for base excision repair suggests that there may be continued selection in functional pathways that may help correct breakage or DNA damage that could be a result of metabolic damage (Kapusta and Suh 2017). An in depth analysis of the lineages experiencing positive selection for genes in these pathways and potential correlations with metabolic requirements would help to test this hypothesis.

Conclusions and implications

Across bird lineages, there is a clear signal of positive selection acting on immune genes, whether against viral pathogens, bacterial pathogens, or transposable elements. Our results demonstrate that innate immune receptor genes continue to be important in the host-pathogen arms race, despite the evolution of an adaptive immune system in vertebrates. Selection at receptors often occurs at a small proportion of sites in the genome, and in a large proportion of avian lineages, suggesting that the same genes and codons may be common targets of pathogens to subvert the immune response. Genes with particularly strong evidence of selection (e.g. TRIF) may be good candidates for further study from a functional and ecological perspective, and could broaden options beyond the traditional MHC and TLRs typically examined. From an applied perspective, there is a great need to understand which proteins or genes in immune gene networks are important in pathogen resistance to improve breeding strategies in economically important species (e.g. poultry; (Kaiser 2010). Our work is a good first step in this direction, and we provide a rich resource for the examination of specific genes and pathways.

Here we have only considered positive selection at the broadest level in birds – across all bird lineages. Combining these results with those from bird populations or specific clades of birds may provide new insights on similarities or differences in long and short-term selection. Pathogen load is the strongest driver of local adaptation in humans (Fumagalli et al. 2011) and viruses are important drivers of population adaptation in flies (Early et al. 2017). From a network perspective, functional gene pathways under strong selection in humans are directly or indirectly involved in immunity (Daub et al. 2013). Given the signatures of host-pathogen co-evolution we observe across birds, we expect that pathogens may be an important driver of adaptation in bird populations as well.

CONCLUSIONS

Integrating different methodologies across evolutionary timescales can provide insights into how micro-scale evolutionary changes might eventually result in observed macro-scale patterns. Only micro-scale changes that are consistent through time should result in observable macroevolutionary differences. Infectious disease is a selective pressure well suited to study this phenomenon, as host-pathogen co-evolution is ubiquitous across the tree of life. Many studies have documented host-pathogen co-evolution at specific immune loci in many diverse systems, but rarely has this been investigated in the context of an entire genome across evolutionary timescales. In my dissertation I use population and comparative genomics to study the dynamics of pathogen-mediated selection at two evolutionary timescales.

In Chapter 1, we use low-density genome-wide markers to identify population genetic consequences of exposure to a novel pathogen in the House Finch. To disentangle the impact of pathogen-mediated selection and the demographic effects of recent founder events, we use a combination of temporal and geographic sampling before the epizootic in both native and introduced populations, and after the epizootic in introduced populations. We show that two different introduced populations showed evidence of reduced genetic diversity, elevated linkage disequilibrium, population differentiation, and many allele frequency shifts. But comparing the same populations before and after the epizootic, we failed to find any population differentiation, elevated linkage disequilibrium, or outlier loci. These results demonstrate that demographic shifts like founder events have much more profound impacts on the genome than natural selection.

In Chapter 2, we use the same system as Chapter 1, but expand our dataset with whole-genome resequencing and additional temporal sampling. This extended sampling allows us to perform population-genetic comparisons before and after the epizootic in both introduced and native populations of House Finch. We show that the population expansion of the native western US population following the last glacial maximum had the largest genome-wide impact on patterns of genetic variation, followed by the eastern founder event. Spatial clustering of allele variation between native and introduced populations suggests that overall genetic differences between populations may be a product of both selection and genetic drift, but simulations and addition testing will be necessary to further support this hypothesis. The greater power of this dataset enables me to detect a decrease in genetic diversity immediately following the epizootic, but we detect no signatures of large selective sweeps. Together, our results suggest that House Finches evolved resistance through polygenic selection.

In Chapter 3, we use comparative genomics to investigate genome-wide signatures of host-pathogen co-evolution across longer temporal scales in birds. We test for positive selection at over 11,000 genes in 39 species of birds. We show that immune genes characterized as receptors, encoding proteins that interact directly with pathogens, are consistently under selection at a higher proportion of sites and in a larger proportion of lineages than estimates from all genes. Using an unbiased approach with all genes and pathways regardless of function, we also show that immune system functional pathways are enriched for signatures of positive selection.

Taken together, this dissertation research shows at short timescales, pathogen-mediated selection is likely to favor many different genes simultaneously, a process known as polygenic selection. However, at long evolutionary timescales, pathogen-mediated selection is the strongest and most-consistent signature of selection compared to other biotic or abiotic possibilities like temperature regimes or metabolic constraints. In both studies, this research shows that pathogen-mediated selection at more than just a single locus. These results at two ends of an evolutionary spectrum provide a framework to better understand how pathogen-mediated selection at short timescales might project into the future.

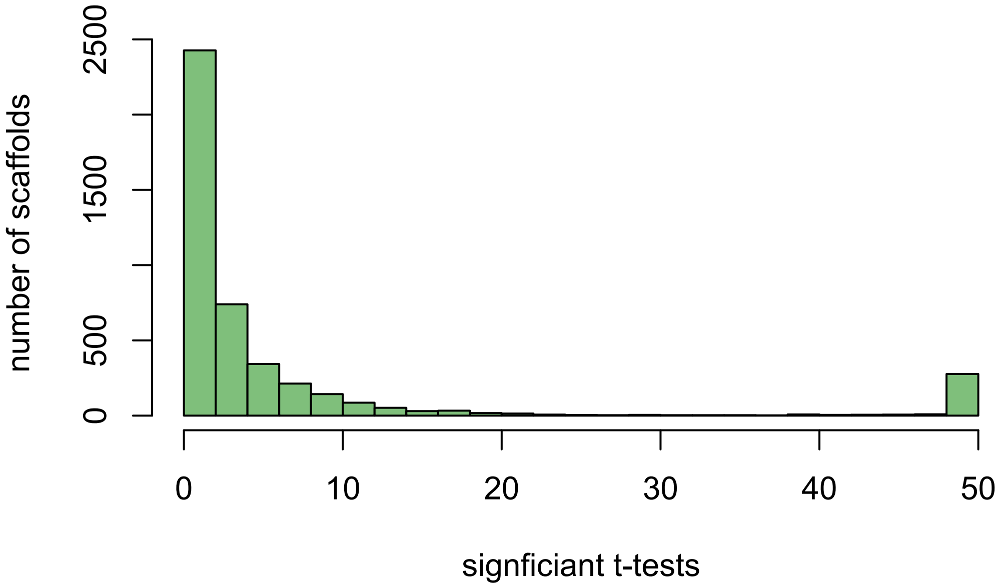
APPENDIX 1

Supplemental materials for Chapter 1 can be found at

http://onlinelibrary.wiley.com/doi/10.1002/ece3.2444/full.

APPENDIX 2

Supplemental materials for Chapter 2



Supplemental Figure 2.1: Histogram of the scaffolds with significantly different male and female coverage. For each scaffold, we performed 50 t-tests for 10 randomly chosen males and 10 randomly chosen females.

Supplemental Table 2.1. Sampling details for all resequenced individuals.

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
ROM	Carpodacus	mexicanus	MKP 1030	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1031	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1033	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1034	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1036	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1037	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1038	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1040	male	adult	male	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1043	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1044	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1046	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1048	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1049	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1050	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1051	female	adult	female	ME_90	ME	1991
ROM	Carpodacus	mexicanus	MKP 1052	female	adult	female	ME_90	ME	1991
OurLab	Carpodacus	mexicanus	ME01	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME02	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME03	-	-	female	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME04	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME05	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME06	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME07	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME08	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME09	-	-	female	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME10	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME11	-	-	female	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME12	-	-	female	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME13	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME14	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME15	-	-	male	ME_01	ME	2003
OurLab	Carpodacus	mexicanus	ME16	-	-	male	ME_01	ME	2003
ROM	Carpodacus	mexicanus	MKP 940	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 941	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 944	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 945	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 946	female	adult	female	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 947	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 948	female	adult	female	NY_90	NY	1990

Supplemental Table 2.1 (Continued).

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
ROM	Carpodacus	mexicanus	MKP 949	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 950	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 951	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 953	female	adult	female	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 954	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 955	female	adult	female	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 956	female	adult	female	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 957	male	adult	male	NY_90	NY	1990
ROM	Carpodacus	mexicanus	MKP 959	male	adult	male	NY_90	NY	1990
OurLab	Carpodacus	mexicanus	NY01	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY02	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY03	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY06	-	-	female	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY07	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY08	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY11	-	-	female	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY12	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY16	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY18	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY20	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY21	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY27	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY28	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY29	-	-	male	NY_01	NY	2003
OurLab	Carpodacus	mexicanus	NY30	-	-	male	NY_01	NY	2003
ROM	Carpodacus	mexicanus	MKP 815	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 816	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 817	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 818	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 820	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 821	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 822	female	adult	female	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 824	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 829	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 831	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 833	female	adult	female	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 834	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 838	female	adult	female	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 839	male	adult	male	OH_90	OH	1990
ROM	Carpodacus	mexicanus	MKP 840	male	adult	male	OH_90	OH	1990

Supplemental Table 2.1 (Continued).

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
ROM	Carpodacus	mexicanus	MKP 845	male	adult	male	OH_90	OH	1990
OurLab	Carpodacus	mexicanus	OH01	-	-	male	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH02	-	-	male	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH03	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH04	-	-	male	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH05	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH06	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH07	-	-	male	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH08	-	-	male	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH09	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH10	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH11	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH12	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH13	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH14	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH15	-	-	female	OH-01	OH	2003
OurLab	Carpodacus	mexicanus	OH16	-	-	female	OH-01	OH	2003
Auburn	Carpodacus	mexicanus	DV-11	-	-	female	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DV-5	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-1	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-13	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-16	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-2	-	-	female	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-20	-	-	female	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-3	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-34	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-37	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-4	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	DX-45	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	W-10	-	-	male	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	W-14	-	-	female	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	W-28	-	-	female	CA_01	CA	2001
Auburn	Carpodacus	mexicanus	W-5	-	-	male	CA_01	CA	2001
MVZ	Carpodacus	mexicanus	AJS 194	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 195	female	juv	female	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 196	female	adult	female	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 197	male	adult	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 198	female	adult	female	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 200	male	adult	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 201	male	juv	male	CA_14	CA	2014

Supplemental Table 2.1. (Continued).

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
MVZ	Carpodacus	mexicanus	AJS 202	male	adult	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 203	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 204	male	adult	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 205	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 206	male	adult	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 207	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 208	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 210	male	juv	male	CA_14	CA	2014
MVZ	Carpodacus	mexicanus	AJS 213	male	juv	male	CA_14	CA	2014
UWBM	Carpodacus	mexicanus	DRF 044	male	-	male	WA_01	WA	2001
UWBM	Carpodacus	mexicanus	DRF 045	male	-	male	WA_01	WA	2001
UWBM	Carpodacus	mexicanus	GKD 127	male	-	male	WA_01	WA	2000
UWBM	Carpodacus	mexicanus	GKD 21	male	-	male	WA_01	WA	1997
UWBM	Carpodacus	mexicanus	HEW 10	female	-	female	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	LNS 11	male	-	male	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	MGS 015	-	-	female	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	MGS 48	male	-	male	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	MNP 11	male	-	male	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	DRF 246	male	-	male	WA_01	WA	2001
UWBM	Carpodacus	mexicanus	EVL 464	male	-	male	WA_01	WA	2001
UWBM	Carpodacus	mexicanus	LRM 021	male	-	male	WA_01	WA	2000
UWBM	Carpodacus	mexicanus	LRM 022	female	-	female	WA_01	WA	1999
UWBM	Carpodacus	mexicanus	PRM 023	female	-	female	WA_01	WA	2002
UWBM	Carpodacus	mexicanus	SCD 020	male	-	male	WA_01	WA	2001
UWBM	Carpodacus	mexicanus	YEC 031	male	-	male	WA_01	WA	1998
MCZ	Carpodacus	mexicanus	363853	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363854	female	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363855	female	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363856	female	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363857	female	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363858	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363859	male	juv	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363861	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363862	female	adult	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363863	female	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363864	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363865	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363866	male	adult	male	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363867	femlae	juv	female	WA_14	WA	2014
MCZ	Carpodacus	mexicanus	363868	male	adult	male	WA_14	WA	2014

Supplemental Table 2.1 (Continued).

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
MCZ	Carpodacus	mexicanus	363869	male	adult	male	WA_14	WA	2014
FMNH	Carpodacus	mexicanus	WWH-8497	female	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8457	-	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8669	male	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8893	male	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-9066	-	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-9139	-	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-9140	-	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	S15-2917	female	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8108	-	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-9148	-	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	S15-006	-	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8328	-	-	female	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-8804	-	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-9123	-	-	male	IL_14	IL	2015
FMNH	Carpodacus	mexicanus	WWH-7881	-	-	female	IL_14	IL	2014
FMNH	Carpodacus	mexicanus	WWH-7880	-	-	male	IL_14	IL	2014
MCZ	Carpodacus	mexicanus	364300	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364301	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364302	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364303	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364304	-	-	female	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364305	-	-	female	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364306	-	-	female	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364307	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364308	-	-	female	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364309	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364310	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364311	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364312	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364313	-	-	female	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364314	-	-	male	MA_14	MA	2016
MCZ	Carpodacus	mexicanus	364315	-	-	male	MA_14	MA	2016
UWBM	Carpodacus	cassinii	GAV 299	-	-	-	CC	WA	1994
UWBM	Carpodacus	cassinii	GAV 303	-	-	-	CC	WA	1994
UWBM	Carpodacus	cassinii	JMB 1570	-	-	-	CC	WA	1994
UWBM	Carpodacus	cassinii	SAR 7007	-	-	-	CC	WA	1995
UWBM	Carpodacus	cassinii	SVD 980	-	-	-	CC	WA	1995
UWBM	Carpodacus	cassinii	GAV 302	-	-	-	CC	WA	1994
UWBM	Carpodacus	cassinii	DAB 464	-	-	-	CC	WA	1994

Supplemental Table 2.1 (Continued).

Institution	Genus	Species	Number	Specimen Sex	Age class	Genetic Sex	WGR_Pop	State	Year
UWBM	Carpodacus	cassinii	JMB 1567	-	-	-	CC	WA	1994
UWBM	Carpodacus	purpureus	LKB 23	-	-	-	CP	WA	1996
UWBM	Carpodacus	purpureus	NAM 14	-	-	-	CP	WA	1998
UWBM	Carpodacus	purpureus	SAR 6975	-	-	-	CP	WA	1995
UWBM	Carpodacus	purpureus	SMB 60	-	-	-	CP	WA	1998
UWBM	Carpodacus	purpureus	WCW 033	-	-	-	CP	WA	1999
UWBM	Carpodacus	purpureus	GKD 14	-	-	-	CP	WA	1998
UWBM	Carpodacus	purpureus	DOT 027	-	-	-	CP	WA	1999
UWBM	Carpodacus	purpureus	GKD 125	-	-	-	CP	Newfoundland	N/A
MCZ	Pinicola	enucleator	338983	-	-	-	Out	MA	2008
MCZ	Uragus	sibiricus	362760	-	-	-	Out	Ikh Gutai	2012
MCZ	Carpodacus	erythrinus	362768	-	-	-	Out	Ikh Gutai	2012

APPENDIX 3

Supplemental materials for Chapter 3.

Supplemental Table 3.1: KEGG pathway enrichment results for all KEGG pathways with genes positively selected in all tests.

ID	Description	GeneRatio	BgRatio	pvalue	qvalue
gga05164	Influenza A	17/208	59/2499	0.0000	0.0002
gga03440	Homologous recombination	11/208	31/2499	0.0000	0.0008
gga03460	Fanconi anemia pathway	11/208	32/2499	0.0000	0.0008
gga04620	Toll-like receptor signaling pathway	11/208	36/2499	0.0001	0.0019
gga04060	Cytokine-cytokine receptor interaction	19/208	91/2499	0.0001	0.0019
gga04512	ECM-receptor interaction	9/208	32/2499	0.0008	0.0116
gga05168	Herpes simplex infection	13/208	60/2499	0.0010	0.0116
gga04514	Cell adhesion molecules (CAMs)	11/208	51/2499	0.0025	0.0260
gga03410	Base excision repair	6/208	19/2499	0.0033	0.0312
gga05132	Salmonella infection	8/208	33/2499	0.0045	0.0379
gga04621	NOD-like receptor signaling pathway	11/208	63/2499	0.0131	0.1004
gga03008	Ribosome biogenesis in eukaryotes	8/208	41/2499	0.0174	0.1221
gga04145	Phagosome	9/208	50/2499	0.0199	0.1248
gga04622	RIG-I-like receptor signaling pathway	6/208	27/2499	0.0207	0.1248
gga00770	Pantothenate and CoA biosynthesis	3/208	10/2499	0.0440	0.2316
gga03450	Non-homologous end-joining	3/208	10/2499	0.0440	0.2316
gga04744	Phototransduction	3/208	11/2499	0.0569	0.2818
gga04510	Focal adhesion	11/208	81/2499	0.0690	0.3162
gga03430	Mismatch repair	3/208	12/2499	0.0713	0.3162
gga00140	Steroid hormone biosynthesis	3/208	13/2499	0.0873	0.3674
gga04142	Lysosome	8/208	61/2499	0.1297	0.4915
gga00310	Lysine degradation	4/208	24/2499	0.1334	0.4915
gga04920	Adipocytokine signaling pathway	5/208	33/2499	0.1342	0.4915
gga04210	Apoptosis	6/208	43/2499	0.1426	0.4943
gga04146	Peroxisome	7/208	53/2499	0.1467	0.4943
gga01212	Fatty acid metabolism	4/208	26/2499	0.1653	0.5355
gga00830	Retinol metabolism	2/208	10/2499	0.1997	0.5799
gga00910	Nitrogen metabolism	2/208	10/2499	0.1997	0.5799
gga00980	Metabolism of xenobiotics by cytochrome P450	2/208	10/2499	0.1997	0.5799
gga03320	PPAR signaling pathway	4/208	29/2499	0.2177	0.6091

Supplemental Table 3.1 (Continued).

ID	Description	GeneRatio	BgRatio	pvalue	qvalue
gga04623	Cytosolic DNA-sensing pathway	3/208	20/2499	0.2289	0.6091
gga00410	beta-Alanine metabolism	2/208	11/2499	0.2314	0.6091
gga04933	AGE-RAGE signaling pathway in diabetic complications	5/208	41/2499	0.2521	0.6381
gga00052	Galactose metabolism	2/208	12/2499	0.2635	0.6381
gga04270	Vascular smooth muscle contraction	4/208	32/2499	0.2738	0.6381
gga04216	Ferroptosis	3/208	22/2499	0.2752	0.6381
gga04810	Regulation of actin cytoskeleton	9/208	86/2499	0.2844	0.6381
gga00071	Fatty acid degradation	2/208	13/2499	0.2955	0.6381
gga00983	Drug metabolism - other enzymes	2/208	13/2499	0.2955	0.6381
gga03420	Nucleotide excision repair	3/208	25/2499	0.3460	0.7194
gga00240	Pyrimidine metabolism	5/208	47/2499	0.3527	0.7194
gga04672	Intestinal immune network for IgA production	2/208	15/2499	0.3588	0.7194
gga00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	2/208	16/2499	0.3897	0.7631
gga04260	Cardiac muscle contraction	2/208	17/2499	0.4199	0.8036
gga00600	Sphingolipid metabolism	3/208	29/2499	0.4391	0.8216
gga00480	Glutathione metabolism	2/208	18/2499	0.4493	0.8226
gga04114	Oocyte meiosis	4/208	44/2499	0.5051	0.8769
gga04371	Apelin signaling pathway	4/208	44/2499	0.5051	0.8769
gga00230	Purine metabolism	6/208	72/2499	0.5624	0.8769
gga00860	Porphyrin and chlorophyll metabolism	1/208	10/2499	0.5813	0.8769
gga03020	RNA polymerase	1/208	11/2499	0.6163	0.8769
gga00531	Glycosaminoglycan degradation	1/208	12/2499	0.6484	0.8769
gga00561	Glycerolipid metabolism	2/208	26/2499	0.6506	0.8769
gga00051	Fructose and mannose metabolism	1/208	13/2499	0.6778	0.8769
gga00601	Glycosphingolipid biosynthesis - lacto and neolacto series	1/208	13/2499	0.6778	0.8769
gga00190	Oxidative phosphorylation	3/208	44/2499	0.7230	0.8769
gga01040	Biosynthesis of unsaturated fatty acids	1/208	15/2499	0.7295	0.8769
gga02010	ABC transporters	1/208	15/2499	0.7295	0.8769
gga00010	Glycolysis / Gluconeogenesis	1/208	16/2499	0.7521	0.8769
gga00500	Starch and sucrose metabolism	1/208	16/2499	0.7521	0.8769
gga00590	Arachidonic acid metabolism	1/208	17/2499	0.7729	0.8769

Supplemental Table 3.1 (Continued).

ID	Description	GeneRatio	BgRatio	pvalue	qvalue
gga03010	Ribosome	3/208	49/2499	0.7886	0.8769
gga04115	p53 signaling pathway	2/208	34/2499	0.7891	0.8769
gga04130	SNARE interactions in vesicular transport	1/208	18/2499	0.7919	0.8769
gga04137	Mitophagy - animal	2/208	35/2499	0.8025	0.8769
gga04110	Cell cycle	4/208	65/2499	0.8038	0.8769
gga03030	DNA replication	1/208	20/2499	0.8254	0.8769
gga04530	Tight junction	4/208	68/2499	0.8315	0.8769
gga03050	Proteasome	1/208	21/2499	0.8400	0.8769
gga04350	TGF-beta signaling pathway	2/208	39/2499	0.8490	0.8769
gga00970	Aminoacyl-tRNA biosynthesis	1/208	22/2499	0.8534	0.8769
gga04910	Insulin signaling pathway	3/208	56/2499	0.8584	0.8769
gga00565	Ether lipid metabolism	1/208	23/2499	0.8657	0.8769
gga04261	Adrenergic signaling in cardiomyocytes	2/208	42/2499	0.8770	0.8769
gga04020	Calcium signaling pathway	4/208	77/2499	0.8959	0.8769
gga04330	Notch signaling pathway	1/208	26/2499	0.8968	0.8769
gga04520	Adherens junction	1/208	27/2499	0.9055	0.8769
gga03018	RNA degradation	2/208	46/2499	0.9070	0.8769
gga00520	Amino sugar and nucleotide sugar metabolism	1/208	28/2499	0.9135	0.8769
gga03022	Basal transcription factors	1/208	29/2499	0.9207	0.8769
gga00564	Glycerophospholipid metabolism	2/208	50/2499	0.9301	0.8769
gga00510	N-Glycan biosynthesis	1/208	32/2499	0.9391	0.8769
gga04012	ErbB signaling pathway	1/208	32/2499	0.9391	0.8769
gga04914	Progesterone-mediated oocyte maturation	1/208	34/2499	0.9490	0.8769
gga03040	Spliceosome	2/208	55/2499	0.9514	0.8769
gga03013	RNA transport	3/208	75/2499	0.9569	0.8769
gga01200	Carbon metabolism	1/208	37/2499	0.9608	0.8769
gga04070	Phosphatidylinositol signaling system	1/208	42/2499	0.9748	0.8769
gga03015	mRNA surveillance pathway	1/208	46/2499	0.9823	0.8769
gga04080	Neuroactive ligand-receptor interaction	6/208	152/2499	0.9909	0.8769
gga04144	Endocytosis	5/208	141/2499	0.9939	0.8769
gga04140	Autophagy - animal	1/208	62/2499	0.9957	0.8769
gga04120	Ubiquitin mediated proteolysis	1/208	68/2499	0.9975	0.8769
gga04010	MAPK signaling pathway	2/208	100/2499	0.9985	0.8769
gga04150	mTOR signaling pathway	1/208	76/2499	0.9988	0.8769
gga04141	Protein processing in endoplasmic reticulum	1/208	90/2499	0.9997	0.8769

REFERENCES

Able, K. P., and J. R. Belthoff. 1998. Rapid “evolution” of migratory behaviour in the introduced house finch of eastern North America. P R Soc B 265:2063–2071.

Adelman, J. S., L. Kirkpatrick, J. L. Grodio, and D. M. Hawley. 2013. House Finch populations differ in early inflammatory signaling and pathogen tolerance at the peak of *Mycoplasma gallisepticum* infection. Am Nat 181:674–689.

Alcaide, M., and S. V. Edwards. 2011. Molecular Evolution of the Toll-Like Receptor Multigene Family in Birds. MBE 28:1703–1715.

Alcaide, M., M. Liu, and S. V. Edwards. 2013. Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. PeerJ 1:e86.

Aldrich, J. W., and J. S. Weske. 1978. Origin and evolution of the eastern House Finch population. Auk 528–536.

Altenhoff, A. M., M. Gil, G. H. Gonnet, and C. Dessimoz. 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. PLoS ONE 8:e53786.

Altizer, S., D. Harvell, and E. Friedle. 2003. Rapid evolutionary dynamics and disease threats to biodiversity. Trends in Ecology and Evolution 18:589–596.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81–92.

Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236.

Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol Ecol 22:3179–3190.

Avise, J. C., and D. Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc Biol Sci B 265:457–463.

Backström, N., A. Qvarnstrom, L. Gustafsson, and H. Ellegren. 2006. Levels of linkage disequilibrium in a wild bird population. Biol Letters 2:435–438.

Backström, N., D. Shipilina, M. P. K. Blom, and S. V. Edwards. 2013a. Cis-regulatory sequence variation and association with Mycoplasma load in natural populations of the house finch (*Carpodacus mexicanus*). Ecol Evol 3:655–666.

Backström, N., Q. Zhang, and S. V. Edwards. 2013b. Evidence from a House Finch (*Haemorhous mexicanus*) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds. MBE 30:1046–1050.

Badyaev, A. V., and G. E. Hill. 2000. The evolution of sexual dimorphism in the house finch. I. Population divergence in morphological covariance structure. Evolution 54:1784–1794.

Badyaev, A. V., G. E. Hill, M. L. Beck, A. A. Dervan, R. A. Duckworth, K. J. Mcgraw, P. M. Nolan, and L. A. Whittingham. 2002. Sex-biased hatching order and adaptive population divergence in a passerine bird. Science 295:316–318.

Baker, A. J., and A. Moeed. 1987. Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). Evolution 41:525–538.

Baker, H. G., and G. L. Stebbins (eds). 1965. The Genetics of Colonizing Species. Academic Press, New York.

Balakrishnan, C. N., and S. V. Edwards. 2008. Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the Zebra Finch (*Taeniopygia guttata*). Genetics 181:645–660.

Barreiro, L. B., and L. Quintana-Murci. 2009. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17–30.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. TREE 23:38–44.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B 57:289–300.

Benner, W. L. 1991. Mitochondrial DNA variation in the House Finch (*Carpodacus mexicanus*). Cornell University.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov. 2014. Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. PLoS Genet 10:e1004775.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next-generation museum population genomics. Mol Ecol 22:6018–6032.

Bonneaud, C., J. Pérez-Tris, P. Federici, O. Chastel, and G. Sorci. 2006. Major histocompatability alleles associated with local resistance to malaria in a passerine. Evolution 60:383-389.

Bonneaud, C., S. L. Balenger, A. F. Russell, J. Zhang, G. E. Hill, and S. V. Edwards. 2011. Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. PNAS 108:7866–7871.

Bonneaud, C., S. L. Balenger, G. E. Hill, and A. F. Russell. 2012a. Experimental evidence for distinct costs of pathogenesis and immunity against a natural pathogen in a wild bird. Mol Ecol 21:4787–4796.

Bonneaud, C., S. L. Balenger, J. Zhang, S. V. Edwards, and G. E. Hill. 2012b. Innate immunity and the evolution of resistance to an emerging infectious disease in a wild bird. Mol Ecol 21:2628–2639.

Bourgeois, Y. X. C., E. Lhuillier, T. Cezard, J. A. M. Bertrand, B. Delahaie, J. Cornuault, T. Duval, O. Bouchez, B. Milá, and C. Thébaud. 2013. Mass production of SNP markers in a nonmodel passerine bird through RAD sequencing and contig mapping to the zebra finch genome. Mol Ecol Res 13:899–907.

Bourret, V., M. Dionne, and L. Bernatchez. 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. Mol Ecol 23:4444–4457.

Breuer, K., A. K. Foroushani, M. R. Laird, C. Chen, A. Sribnaia, R. Lo, G. L. Winsor, R. E. W. Hancock, F. S. L. Brinkman, and D. J. Lynn. 2012. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic Acids Research 41:D1228–D1233.

Brown, J. W., J. S. Rest, J. Garcia-Moreno, M. D. Sorenson, and D. P. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology 6:6.

Buchmann, K. 2014. Evolution of innate immunity: clues from invertebrates via fish to mammals. Front Immunol 5:459.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnström, M. Ružić, S.-A. Sæther, G.-P. Sætre, J. Török, and H. Ellegren. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedulaflycatchers. Gen Res 25:1656–1665.

Burri, R., H. N. Hirzel, N. Salamin, A. Roulin, and L. Fumagalli. 2008. Evolutionary Patterns of MHC Class II B in Owls and Their Implications for the Understanding of Avian MHC Evolution. MBE 25:1180–1191.

Burri, R., N. Salamin, R. A. Studer, A. Roulin, and L. Fumagalli. 2010. Adaptive Divergence of Ancient Gene Duplicates in the Avian MHC Class II. MBE 27:2360–2374.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado, and M. Yandell. 2007. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Gen Res 18:188–196.

Carling, M. D., and R. T. Brumfield. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). PLoS ONE 2:e160.

Casals, F., M. Sikora, H. Laayouni, L. Montanucci, A. Muntasell, R. Lazarus, F. Calafell, P. Awadalla, M. G. Netea, and J. Bertranpetit. 2011. Genetic adaptation of the antibacterial human innate immunity network. BMC Evol Biol 11:202.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. G3 1:171–182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140.

Chapman, J. R., O. Hellgren, A. S. Helin, R. H. S. Kraus, R. L. Cromie, and J. Waldenström. 2016. The Evolution of Innate Immune Genes: Purifying and Balancing Selection on β-Defensins in Waterfowl. MBE 33:3075–3087.

Chen, N., E. J. Cosgrove, R. Bowman, J. W. Fitzpatrick, and A. G. Clark. 2016. Genomic consequences of population decline in the endangered Florida Scrub-Jay. Curr Biol 26:2974–2979.

Chen, S., A. Cheng, and M. Wang. 2013. Innate sensing of viruses by pattern recognition receptors in birds. Vet. Res. 44:82.

Clements, J. F., T. S. Schulenberg, M. J. Iliff, T. A. Fredericks, B. L. Sullivan, and C. L. Wood. 2016. The Clements checklist of the birds of the world: v2016.

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

Cui, J., W. Zhao, Z. Huang, E. D. Jarvis, M. T. P. Gilbert, P. J. Walker, E. C. Holmes, and G. Zhang. 2014. Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol 15:539.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. Emerging infectious diseases of wildlife--threats to biodiversity and human health. Science 287:443–449.

Daub, J. T., T. Hofer, E. Cutivet, I. Dupanloup, L. Quintana-Murci, M. Robinson-Rechavi, and L. Excoffier. 2013. Evidence for polygenic adaptation to pathogens in the human genome. MBE 30:1544–1558.

De Mita, S., A.-C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux. 2013.

Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol 22:1383–1399.

Dhondt, A. A., A. V. Badyaev, A. P. Dobson, D. M. Hawley, M. J. L. Driscoll, W. M. Hochachka, and D. H. Ley. 2006. Dynamics of mycoplasmal conjunctivitis in the native and introduced range of the host. EcoHealth 3:95–102.

Dhondt, A. A., D. L. Tessaglia, and R. L. Slothower. 1998. Epidemic mycoplasmal conjunctivitis in house finches from eastern North America. J Wildl Dis 34:265–280.

Dierickx, E. G., A. J. Shultz, F. Sato, T. Hiraoka, and S. V. Edwards. 2015. Morphological and genomic comparisons of Hawaiian and Japanese Black-footed Albatrosses (*Phoebastria nigripes*) using double digest RADseq: implications for conservation. Evol Appl 8:662–678.

Dilthey, A., C. Cox, Z. Iqbal, M. R. Nelson, and G. McVean. 2015. Improved genome inference in the MHC using a population reference graph. Nat Genet 47:682-688.

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

Dobson, A. P., and J. Foufopoulos. 2001. Emerging infectious pathogens of wildlife. Phil Trans of the Roy Soc B 356:1001–1012.

Domingues, V. S., Y.-P. Poh, B. K. Peterson, P. S. Pennings, J. D. Jensen, and H. E. Hoekstra. 2012. Evidence of adaptation from ancestral variation in young populations of beach mice. Evolution 66:3209–3223.

Donald, P. F. 2007. Adult sex ratios in wild bird populations. Ibis 149:671-692.

Downing, T., A. T. Lloyd, C. O'Farrelly, and D. G. Bradley. 2010. The Differential Evolutionary Dynamics of Avian Cytokine and TLR Gene Classes. J Immunol 184:6993–7000.

Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4:1184–1191.

Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21:3439–3440.

Earl, D. A., and B. M. vonHoldt. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resour 4:359–361.

Early, A. M., J. R. Arguello, M. Cardoso-Moreira, S. Gottipati, J. K. Grenier, and A. G. Clark.

2017. Survey of Global Genetic Diversity Within the Drosophila Immune System. Genetics 205:353–366.

Edwards, S. V., A. J. Shultz, and S. C. Campbell-Staton. 2015. Next-generation sequencing and the expanding domain of phylogeography. Folia Zoologica 64:187–206.

Edwards, S. V., and M. Dillon. 2004. Hitchhiking and recombination in birds: evidence from Mhc-linked and unlinked loci in Red-winged Blackbirds (Agelaius phoeniceus). Genet. Res. 84:175–192.

Edwards, S. V., and P. W. Hedrick. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. TREE 13:305–311.

Edwards, S. V., E. K. Wakeland, and W. K. Potts. 1995. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. PNAS 92:12200–12204.

Edwards, S. V., J. Gasper, D. Garrigan, D. Martindale, and B. F. Koop. 2000. A 39-kb sequence around a blackbird Mhc class II gene: ghost of selection past and songbird genome architecture. MBE 17:1384–1395.

Egbert, J. R., and J. R. Belthoff. 2003. Wing shape in House Finches differs relative to migratory habit in eastern and western North America. Condor 105: 825-829.

Eizaguirre, C., T. L. Lenz, M. Kalbe, and M. Milinski. 2012. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. Ecol Lett 15:723–731.

Ellegren, H. 2013. The evolutionary genomics of birds. Annu. Rev. Ecol. Evol. Syst. 44:239–259.

Elliott, J. J., and R. S. Arbib Jr. 1953. Origin and status of the house finch in the eastern United States. Auk 70:31–37.

Ellis, J. S., L. M. Turner, and M. E. Knight. 2012. Patterns of selection and polymorphism of innate immunity genes in bumblebees (Hymenoptera: Apidae). Genetica 140:205–217.

Enard, D., L. Cai, C. Gwennap, and D. A. Petrov. 2016. Viruses are a dominant driver of protein adaptation in mammals. Elife 5:e12469.

Epstein, B., M. Jones, R. Hamede, S. Hendricks, H. McCallum, E. P. Murchison, B. Schonfeld, C. Wiench, P. Hohenlohe, and A. Storfer. 2016. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Comms 7:12684.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620.

Excoffier, L., and N. Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. TREE 23:347–351.

Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? MBE 23:691–700.

Fey, S. B., A. M. Siepielski, S. Nusslé, K. Cervantes-Yoshida, J. L. Hwan, E. R. Huber, M. J. Fey, A. Catenazzi, and S. M. Carlson. 2015. Recent shifts in the occurrence, cause, and magnitude of animal mass mortality events. PNAS 112:1083–1088.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993.

Fornarino, S., G. Laval, L. B. Barreiro, J. Manry, E. Vasseur, and L. Quintana-Murci. 2011. Evolution of the TIR Domain-Containing Adaptors in Humans: Swinging between Constraint and Adaptation. MBE 28:3087–3097.

Foster, J. T., B. L. Woodworth, L. E. Eggert, P. J. Hart, D. Palmer, D. C. Duffy, and R. C. Fleischer. 2007. Genetic structure and evolved malaria resistance in Hawaiian honeycreepers. Mol Ecol 16:4738–4746.

Franks, S. J., N. C. Kane, N. B. O'Hara, S. Tittes, and J. S. Rest. 2016. Rapid genome-wide evolution in *Brassica rapa* populations following drought revealed by sequencing of ancestral and descendant gene pools. Mol. Ecol. 25:3622–3631.

Fraser, B. A., I. W. Ramnarine, and B. D. Neff. 2010. Temporal variation at the MHC class IIB in wild populations of the guppy (*Poecilia reticulata*). Evolution 64:2086-2096.

Fumagalli, M., and M. Sironi. 2014. Human genome variability, natural selection and infectious diseases. Current Opinion in Immunology 30:9–16.

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

Fumagalli, M., M. Sironi, U. Pozzoli, A. Ferrer-Admettla, L. Pattini, and R. Nielsen. 2011. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLoS Genet 7:e1002355.

Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet 11:e1005004.

Gaunson, J. E., C. J. Philip, K. G. Whithear, and G. F. Browning. 2006. The cellular immune response in the tracheal mucosa to *Mycoplasma gallisepticum* in vaccinated and unvaccinated chickens in the acute and chronic stages of disease. Vaccine 24:2627–

2633.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80.

Gill, F. B. 2007. Ornithology. W H Freeman & Company.

Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, and S. Sykes. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. PNAS 108:1513–1518.

Gompert, Z., S. P. Egan, R. D. H. Barrett, J. L. Feder, and P. Nosil. 2016. Multilocus approaches for the measurement of selection on correlated genetic loci. Mol Ecol 26:365–382.

Goodsman, D. W., B. Cooke, D. W. Coltman, and M. A. Lewis. 2014. Theoretical Population Biology. Theor Popul Biol 98:1–10.

Grueber, C. E., G. P. Wallis, and I. G. Jamieson. 2014. Episodic Positive Selection in the Evolution of Avian Toll-Like Receptor Innate Immunity Genes. PLoS ONE 9:e89632.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5:e1000695.

Habel, J. C., M. Husemann, A. Finger, P. D. Danley, and F. E. Zachos. 2013. The relevance of time series in molecular ecology and conservation biology. Biological Reviews 89:484–492.

Haldane, J. 1949. Disease and evolution. Ric Sci Suppl 19:68–76.

Hawley, D. M., and R. C. Fleischer. 2012. Contrasting epidemic histories reveal pathogen-mediated balancing selection on class II MHC diversity in a wild songbird. PLoS ONE 7:e30222.

Hawley, D. M., D. Hanley, A. A. Dhondt, and I. J. Lovette. 2006. Molecular evidence for a founder effect in invasive house finch (*Carpodacus mexicanus*) populations experiencing an emergent disease epidemic. Mol Ecol 15:263–275.

Hawley, D. M., J. Briggs, A. A. Dhondt, and I. J. Lovette. 2008. Reconciling molecular signatures across markers: mitochondrial DNA confirms founder effect in invasive North American house finches (*Carpodacus mexicanus*). Conserv Genet 9:637–643.

Hess, C. M., and S. V. Edwards. 2002. The Evolution of the Major Histocompatibility Complex in Birds. BioScience 52:423–431.

Hess, C. M., Z. Wang, and S. V. Edwards. 2007. Evolutionary genetics of *Carpodacus mexicanus*, a recently colonized host of a bacterial pathogen, *Mycoplasma gallisepticum*. Genetica 129:217–225.

Hewitt, G. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philos T Roy Soc B 359:183-195.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907-913.

Hill, A. V. 1998. The immunogenetics of human infectious diseases. Annu. Rev. Immunol. 16:593–617.

Hill, A. V., C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood. 1991. Common west African HLA antigens are associated with protection from severe malaria. Nature 352:595–600.

Hill, G. E. 2002. A Red Bird in a Brown Bag: The Function and Evolution of Colorful Plumage in the House Finch. Oxford University Press, New York.

Hill, G. E. 1996. Subadult plumage in the house finch and tests of models for the evolution of delayed plumage maturation. Auk 858–874.

Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–231.

Hochachka, W. M., and A. A. Dhondt. 2000. Density-dependent decline of host abundance resulting from a new infectious disease. PNAS 97:5303–5306.

Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos T Roy Soc B 367:395–408.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genet 6:e1000862.

Huang, H., and D. L. Rabosky. 2015. Sex-linked genomic variation and its relationship to avian plumage dichromatism and sexual selection. BMC Evol Biol 15:199.

Hudson, R. R. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. J Appl Ecol 46:10–18.

Iwasaki, A., and R. Medzhitov. 2010. Regulation of Adaptive Immunity by the Innate Immune System. Science 327:291–295.

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

Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, and C. Li. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320-1331.

Jensen, J. D., M. Foll, and L. Bernatchez. 2016. The past, present and future of genomic scans for selection. Mol Ecol 25:1-4.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Johnson, L. S., S. R. Eddy, and E. Portugaly. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Juul-Madsen, H. R., B. Viertlböeck, S. Härtle, A. L. Schmidt, and T. W. Göbel. 2014. Innate Immune Responses. Pp. 121–147 *in* K. A. Schat, B. Kaspars, and P. Kaiser, eds. Avian Immunology. Elsevier, Ltd.

Kaiser, P. 2010. Advances in avian immunology—prospects for disease control: a review. Avian Pathology 39:309–324.

Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28:27–30.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2011. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research 40:D109–D114.

Kapusta, A., A. Suh, and C. Feschotte. 2017. Dynamics of genome size evolution in birds and mammals. PNAS 114:E1460–E1469.

Kapusta, A., and A. Suh. 2017. Evolution of bird genomes-a transposon's-eye view. Ann NY Acad Sci 1389:164-185.

Karlsson, E. K., D. P. Kwiatkowski, and P. C. Sabeti. 2014. Natural selection and infectious disease in human populations. Nat Rev Genet 15:379–393.

Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. MBE 30:772–780.

Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, P. Olason, and H. Ellegren. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol 23:4035-4058.

Kazazian, H. H. 2004. Mobile elements: drivers of genome evolution. Science 303:1626–1632.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kim, S. Y., K. E. Lohmueller, A. Albrechtsen, Y. Li, T. Korneliussen, G. Tian, N. Grarup, T. Jiang, G. Andersen, D. Witte, T. Jørgensen, T. Hansen, O. Pedersen, J. Wang, and R. Nielsen. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics 12:231.

Kim, Y., and D. Gulisija. 2010. Signatures of Recent Directional Selection Under Different Models of Population Expansion During Colonization of New Selective Environments. Genetics 184:571–585.

Koch, E., and J. Novembre. 2017. A temporal perspective on the interplay of demography and selection on deleterious variation in humans. G3; Genes|Genomes|Genetics 7:1027–1037.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

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

Korneliussen, T. S., and I. Moltke. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics 31:4009–4011.

Korneliussen, T. S., I. Moltke, A. Albrechtsen, and R. Nielsen. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics 14:289.

Kosakovsky Pond, S. L., B. Murrell, M. Fourment, S. D. W. Frost, W. Delport, and K. Scheffler. 2011. A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. MBE 28:3033–3043.

Kumar, H., T. Kawai, and S. Akira. 2011. Pathogen Recognition by the Innate Immune System. Int Rev Immunol 30:16–34.

Kumar, S., A. J. Filipski, F. U. Battistuzzi, S. L. Kosakovsky Pond, and K. Tamura. 2012. Statistics and truth in phylogenomics. MBE 29:457–472.

Le Negrate, G. 2011a. Subversion of innate immune responses by bacterial hindrance of NF-κB pathway. Cellular Microbiology 14:155–167.

Le Negrate, G. 2011b. Viral interference with innate immunity by preventing NF-κB activity. Cellular Microbiology 14:168–181.

Lee, C. E. 2002. Evolutionary genetics of invasive species. TREE 17:386–391.

Leonard, J. A. 2008. Ancient DNA applications for wildlife conservation. Mol Ecol 17:4186–4196.

Leonardi, M., P. Librado, C. Der Sarkissian, M. Schubert, A. H. Alfarhan, S. A. Alquraishi, K. A. S. Al-Rasheid, C. Gamba, E. Willerslev, and L. Orlando. 2017. Evolutionary Patterns and Processes: Lessons from Ancient DNA. Syst Biol 66:e1–e29.

Ley, D. H., J. E. Berkhoff, and J. M. McLaren. 1996. Mycoplasma gallisepticum isolated from house finches (*Carpodacus mexicanus*) with conjunctivitis. Avian Dis. 40:480–483.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. Nature 475:493–496.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Lindo, J., E. H.-S. A. nchez, S. Nakagome, M. Rasmussen, B. Petzelt, J. Mitchell, J. S. Cybulski, E. Willerslev, M. DeGiorgio, and R. S. Malhi. 2016. A time transect of exomes from a Native American population before and after European contact. Nat Comms 7:13175.

Lohmueller, K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, N. Vinckenbosch, G. Tian, E. Huerta-Sanchez, A. F. Feder, N. Grarup, T. Jørgensen, T. Jiang, D. R. Witte, A. Sandbæk, I. Hellmann, T. Lauritzen, T. Hansen, O. Pedersen, J. Wang, and R. Nielsen. 2011. Natural Selection Affects Multiple Aspects of Genetic Variation at Putatively Neutral Sites across the Human Genome. PLoS Genet 7:e1002326.

Longo, A. V., P. A. Burrowes, and K. R. Zamudio. 2014. Genomic studies of disease-outcome in host-pathogen dynamics. Integrative and Comparative Biology 54:427–438.

Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol 23:2178–2192.

Love, W., N. Dobbs, L. Tabor, and J. W. Simecka. 2010. Toll-Like Receptor 2 (TLR2) Plays a Major Role in Innate Resistance in the Lung against Murine Mycoplasma. PLoS ONE 5:e10739.

Loytynoja, A., and N. Goldman. 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. Science 320:1632–1635.

Luo, W., and C. Brouwer. 2013. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 29:1830–1831.

Malaspinas, A.-S. 2015. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. Mol Ecol 25:24-41.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W. M. Chen. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873.

Markova-Raina, P., and D. Petrov. 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. Genome Research 21:863–874.

Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero, and B. C. Emerson. 2014. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources 15:28–41.

Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, and D. Reich. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503.

Mccormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2011. Applications of next-generation sequencing to phylogeography and phylogenetics. MPE 66:526–538.

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

Medzhitov, R. 2007. Recognition of microorganisms and activation of the immune response. Nature 449:819–826.

Messer, P. W., and D. A. Petrov. 2013. Population genomics of rapid adaptation by soft selective sweeps. TREE 28:659–669.

Messer, P. W., S. P. Ellner, and N. G. Hairston Jr. 2016. Can population genetics adapt to rapid evolution? Trends Genet 32:408–418.

Mohammed, J., S. Frasca Jr., K. Cecchini, D. Rood, A. C. Nyaoke, S. J. Geary, and L. K. Silbart. 2007. Chemokine and cytokine gene expression profiles in chickens inoculated with *Mycoplasma gallisepticum* strains Rlow or GT5. Vaccine 25:8611–8621.

Moldovan, G.-L., and A. D. D'Andrea. 2009. How the Fanconi Anemia Pathway Guards the Genome. Annu. Rev. Genet. 43:223–249.

Mooney, H. A., and E. E. Cleland. 2001. The evolutionary impact of invasive species. PNAS 98:5446–5451.

Moore, R. T. 1939. A review of the house finches of the subgenus Burrica. Condor 41:177–205.

Moran, E. V., and J. M. Alexander. 2014. Evolutionary responses to global change: lessons from invasive species. Ecol Lett 17:637–649.

Murrell, B., S. Weaver, M. D. Smith, J. O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D. P. Martin, D. M. Smith, K. Scheffler, and S. L. Kosakovsky Pond. 2015. Gene-Wide Identification of Episodic Selection. MBE 32:1365–1371.

Nadachowska-Brzyska, K., C. Li, L. Smeds, G. Zhang, and H. Ellegren. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biol. 25:1375–1380.

Nam, K., C. Mugal, B. Nabholz, H. Schielzeth, J. B. Wolf, N. Backström, A. Kunstner, C. N. Balakrishnan, A. Heger, C. P. Ponting, D. F. Clayton, and H. Ellegren. 2010. Molecular evolution of genes in avian genomes. Genome Biol 11:R68.

Nei, M., and F. Tajima. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution 29:1–10.

Netea, M. G., C. Wijmenga, and L. A. J. O'Neill. 2012. Genetic variation in Toll-like receptors and disease susceptibility. Nat Immunol 13:535–542.

Nielsen, E. E., and M. M. Hansen. 2008. Waking the dead: the value of population genetic analyses of historical samples. Fish and Fisheries 9:450–461.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39:197–218.

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

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

Nishino, J. 2013. Detecting Selection Using Time-Series Data of Allele Frequencies with Multiple Independent Reference Loci. G3; Genes|Genomes|Genetics 3:2151–2161.

Nolan, P. M., G. E. Hill, and A. M. Stoehr. 1998. Sex, size, and plumage redness predict house finch survival in an epidemic. Proc Roy Soc B 265:961–965.

Obbard, D. J., J. J. Welch, K.-W. Kim, and F. M. Jiggins. 2009. Quantifying Adaptive Evolution in the Drosophila Immune System. PLoS Genet 5:e1000698.

Organ, C. L., A. M. Shedlock, A. Meade, M. Pagel, and S. V. Edwards. 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature 446:180–184.

Organ, C. L., and S. V. Edwards. 2011. Major Events in Avian Genome Evolution. Pp. 325–337 *in* G. Dyke and G. Kaiser, eds. Living Dinosaurs: the evolutionary history of modern birds. John Wiley & Sons, Ltd, Hoboken, NJ.

Ouborg, N. J., C. Pertoldi, V. Loeschcke, R. K. Bijlsma, and P. W. Hedrick. 2010. Conservation genetics in transition to conservation genomics. Trends Genet 26:177–187.

Oyler-McCance, S. J., R. S. Cornman, K. L. Jones, and J. A. Fike. 2015. Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. Heredity 115:452-459.

Palmer, W. J., and F. M. Jiggins. 2015. Comparative Genomics Reveals the Origins and Diversity of Arthropod Immune Systems. MBE 32:2111–2129.

Parks, M., S. Subramanian, C. Baroni, M. C. Salvatore, G. Zhang, C. D. Millar, and D. M. Lambert. 2014. Ancient population genomics and the study of evolution. Phil Trans Roy Soc B 370:20130381.

Pavey, S. A., J. Gaudin, E. Normandeau, M. Dionne, M. Castonguay, C. Audet, and L. Bernatchez. 2015. RAD Sequencing Highlights Polygenic Discrimination of Habitat Ecotypes in the Panmictic American Eel. Curr Biol 25:1666–1671.

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier. 2013. On the accumulation of deleterious mutations during range expansions. Mol Ecol 22:5972–5982.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Pilot, M., C. Greco, B. M. Vonholdt, B. J. E. drzejewska, E. Randi, W. J. E. drzejewski, V. E. Sidorovich, E. A. Ostrander, and R. K. Wayne. 2013. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity 112:428–442.

Poh, Y.-P., V. S. Domingues, H. E. Hoekstra, and J. D. Jensen. 2014. On the prospect of identifying adaptive loci in recently bottlenecked populations. PLoS ONE 9:e110579.

Pond, S. L. K., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

Pool, J. E., and R. Nielsen. 2007. Population size changes reshape genomic patterns of diversity. Evolution 61:3001–3006.

Pritchard, J. K., and A. Di Rienzo. 2010. Adaptation – not by sweeps alone. Nature Reviews Genetics 11:665–667.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Quintana-Murci, L., and A. G. Clark. 2013. Population genetic tools for dissecting innate immunity in humans. Nat Rev Immunol 13:280–293.

R Core Development Team. 2008. R: A Language and Environment for Statistical Computing.

Reed, D. H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. Conservation Biology.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, D. L. Crawford, and A. Whitehead. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science 354:1305–1308.

Reznick, D. N., and C. K. Ghalambor. 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica 112:183–198.

Roderick, G. K., and R. G. Gillespie. 1998. Speciation and phylogeography of Hawaiian terrestrial arthropods. Mol Ecol 7:519–531.

Rogers, R. L., and M. Slatkin. 2017. Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet 13:e1006601.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137–138.

Roth, A. C., G. H. Gonnet, and C. Dessimoz. 2008. Algorithm of OMA for large-scale orthology inference. BMC Bioinformatics 9:518.

Roux, J., E. Privman, S. Moretti, J. T. Daub, M. Robinson-Rechavi, and L. Keller. 2014.

Patterns of Positive Selection in Seven Ant Genomes. Molecular Biology and Evolution 31:1661–1685.

Ruegg, K., E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol 23:4757–4769.

Sabeti, P. C. 2006. Positive Natural Selection in the Human Lineage. Science 312:1614–1620.

Sackett, L. C., S. K. Collinge, and A. P. Martin. 2013. Do pathogens reduce genetic diversity of their hosts? Variable effects of sylvatic plague in black-tailed prairie dogs. Mol Ecol 22:2441–2455.

Sackton, T.B., P. Grayson, A. Cloutier, Z. Hu, J. Liu, N. Wheeler, P. Gardner, J. Clarke, A. Baker, M. Clamp, S.V. Edwards. in prep. Convergent regulatory evolution and the origin o fflightlessness in palaeognathous birds.

Sackton, T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark, and A. G. Clark. 2007. Dynamic evolution of the innate immune system in Drosophila. Nat Genet 39:1461–1468.

Salathe, M., R. D. Kouyos, and S. Bonhoeffer. 2008. The state of affairs in the kingdom of the Red Queen. TREE 23:439–445.

Santhakumar, D., D. Rubbenstroth, L. Martinez-Sobrido, and M. Munir. 2017. Avian Interferons and Their Antiviral Effectors. Front. Immunol. 8:49.

Savage, A. E., and K. R. Zamudio. 2011. MHC genotypes associate with resistance to a frog-killing fungus. PNAS 108:16705–16710.

Schat, K. A., B. Kaspers, and P. Kaiser. 2012. Avian Immunology. Academic Press.

Schlenke, T. A., and D. J. Begun. 2003. Natural selection drives Drosophila immune system evolution. Genetics 164:1471–1480.

Schubert, I., and G. T. H. Vu. 2016. Genome Stability and Evolution: Attempting a Holistic View. Trends in plant science 21:749–757.

Seutin, G., B. N. White, and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology 69:82–90.

Shafer, A. B. A., J. B. W. Wolf, P. C. Alves, L. Bergström, M. W. Bruford, I. Brännström, G. Colling, L. Dalén, L. De Meester, R. Ekblom, K. D. Fawcett, S. Fior, M. Hajibabaei, J. A. Hill, A. R. Hoezel, J. Höglund, E. L. Jensen, J. Krause, T. N. Kristensen, M. Krützen, J. K. McKAY, A. J. Norman, R. Ogden, E. M. Österling, N. J. Ouborg, J. Piccolo, D. Popović, C. R. Primmer, F. A. Reed, M. Roumet, J. Salmona, T. Schenekar, M. K. Schwartz, G.

Segelbacher, H. Senn, J. Thaulow, M. Valtonen, A. Veale, P. Vergeer, N. Vijay, C. Vilà, M. Weissensteiner, L. Wennerström, C. W. Wheat, and P. Zieliński. 2015. Genomics and the challenging translation into conservation practice. Trends in Ecology and Evolution 30:78–87.

Shaw, A. K., and H. Kokko. 2014. Mate finding, Allee effects and selection for sex-biased dispersal. Journal of Animal Ecology 83:1256–1267.

Shultz, A. J., A. J. Baker, G. E. Hill, and P. M. Nolan. 2016. SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorhous mexicanus*). Ecology and Evolution 6:7475-7489.

Simons, Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella. 2014. The deleterious mutation load is insensitive to recent population history. Nat Genet 46:220–224.

Singh, B., C. Fleury, F. Jalalvand, and K. Riesbeck. 2012. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol Rev 36:1122–1180.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, and M. Przeworski. 2015. Stable recombination hotspots in birds. Science 350:928–932.

Sironi, M., R. Cagliani, D. Forni, and M. Clerici. 2015. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nature Reviews Genetics 16:224–236.

Skerratt, L. F., L. Berger, R. Speare, S. Cashins, K. R. McDonald, A. D. Phillott, H. B. Hines, and N. Kenyon. 2007. Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. EcoHealth 4:125–134.

Slatkin, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9:477–485.

Smith, B. T., R. W. Bryson Jr, V. Chua, L. Africa, and J. Klicka. 2013. Speciational history of North American Haemorhous finches (Aves: Fringillidae) inferred from multilocus data. MPE 66:1055–1059.

Smith, T. B., M. T. Kinnison, S. Y. Strauss, T. L. Fuller, and S. P. Carroll. 2014. Prescriptive Evolution to Conserve and Manage Biodiversity. Annu. Rev. Ecol. Evol. Syst. 45:1–22.

Sokol, C. L., and A. D. Luster. 2015. The Chemokine System in Innate Immunity. Cold Spring Harb Perspect Biol 7:a016303.

Sorci, G., A. P. Møller, and J. Clobert. 1998. Plumage dichromatism of birds predicts introduction success in New Zealand. J Animal Ecol 67:263-269.

Staley, M., and C. Bonneaud. 2015. Immune responses of wild birds to emerging infectious diseases. Parasite Immunol 37:242–254.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stanke, M., R. Steinkamp, S. Waack, and B. Morgenstern. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic acids research 32:W309–W312.

Stapley, J., T. R. Birkhead, T. Burke, and J. Slate. 2008. A linkage map of the Zebra Finch Taeniopygia guttata provides new Iinsights Into avian genome evolution. Genetics 179:651–667.

Subramanian, S. 2016. The effects of sample size on population genomic analyses--implications for the tests of neutrality. BMC Genomics 17:123.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tennessen, J. A. 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. J Evol Biol 18:1387–1394.

Therkildsen, N. O., J. Hemmer-Hansen, T. D. Als, D. P. Swain, M. J. Morgan, E. A. Trippel, S. R. Palumbi, D. Meldrup, and E. E. Nielsen. 2013. Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. Mol Ecol 22:2424–2440.

Thornton, K. R., and J. D. Jensen. 2007. Controlling the False-Positive Rate in Multilocus Genome Scans for Selection. Genetics 175:737–750.

Thornton, K. R., J. D. Jensen, C. Becquet, and P. Andolfatto. 2007. Progress and prospects in mapping recent selection in the genome. Heredity 98:340–348.

Tiffin, P., and J. Ross-Ibarra. 2014. Advances and limits of using population genetics to understand local adaptation. TREE 29:673–680.

Tin, M. M. Y., J. Arora, T. D. Seeley, and A. S. Mikheyev. 2015. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. Nat Comms 6:7991.

Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu, T. J. White, J. J. Sninsky, M. Cargill, M. D. Adams, C. D. Bustamante, and A. G. Clark. 2009. Evolutionary Processes Acting on Candidate cis-Regulatory Regions in Humans Inferred from Patterns of Polymorphism and Divergence. PLoS Genet 5:e1000592.

Tschirren, B., M. Andersson, K. Scherman, H. Westerdahl, P. R. E. Mittl, and L. Raberg. 2013. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. P R Soc B 280:20130364.

Unckless, R. L., V. M. Howick, and B. P. Lazzaro. 2016. Convergent Balancing Selection on an Antimicrobial Peptide in Drosophila. Curr Biol 26:257–262.

Van Valen, L. 1973. A new evolutionary law. Evolutionary theory 1:1–30.

Vazquez-Phillips, M. A. 1992. Population differentiation of the House Finch (C*arpodacus mexicanus*) in North America and the Hawaiian Islands. University of Toronto.

Veit, R. R., and M. A. Lewis. 1996. Dispersal, population growth, and the Allee effect: dynamics of the house finch invasion of eastern North America. Am Nat 148:255–274.

Venter, O., E. W. Sanderson, A. Magrach, J. R. Allan, J. Beher, K. R. Jones, H. P. Possingham, W. F. Laurance, P. Wood, B. M. Fekete, M. A. Levy, and J. E. M. Watson. 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nat Comms 7:12558.

Viljakainen, L. 2015. Evolutionary genetics of insect innate immunity. Briefings in Functional Genomics 14:407–412.

Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting Natural Selection in Genomic Data. Annu. Rev. Genet. 47:97–120.

Wandeler, P., P. E. A. Hoeck, and L. F. Keller. 2007. Back to the future: museum specimens in population genetics. TREE 22:634–642.

Wang, Z., A. J. Baker, G. E. Hill, and S. V. Edwards. 2003. Reconciling actual and inferred population histories in the house finch (*Carpodacus mexicanus*) by AFLP analysis. Evolution 57:2852–2864.

Wang, Z., K. FARMER, G. E. Hill, and S. V. Edwards. 2006. A cDNA macroarray approach to parasite-induced gene expression changes in a songbird host: genetic response of house finches to experimental infection by Mycoplasma gallisepticum. Mol Ecol 15:1263–1273.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. WHITNEY, A. R. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, and R. K. Wilson. 2010. The genome of a songbird. Nature 464:757–762.

Warren, W. C., L. W. Hillier, C. Tomlinson, P. Minx, M. Kremitzki, T. Graves, C. Markovic, N. Bouk, K. D. Pruitt, F. Thibaud-Nissen, V. Schneider, T. A. Mansour, C. T. Brown, A. Zimin, R. Hawken, M. Abrahamsen, A. B. Pyrkosz, M. Morisson, V. Fillon, A. Vignal, W. Chow, K. Howe, J. E. Fulton, M. M. Miller, P. Lovell, C. V. Mello, M. Wirthlin, A. S. Mason, R. Kuo, D. W. Burt, J. B. Dodgson, and H. H. Cheng. 2017. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. G3: Genes|Genomes|Genetics 7:109–117.

Waterhouse, R. M., E. V. Kriventseva, S. Meister, and Z. Xi. 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316:1738-1743.

Webb, A. E., Z. N. Gerek, C. C. Morgan, T. A. Walsh, C. E. Loscher, S. V. Edwards, and M. J. O'Connell. 2015. Adaptive evolution as a predictor of species-specific innate immune response. MBE 32:1717-1729.

Wellenreuther, M., and B. Hansson. 2016. Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. Trends Genet 32:155–164.

Westerdahl, H., J. Waldenstrom, B. Hansson, D. Hasselquist, T. von Schantz, and S. Bensch. 2005. Associations between malaria and MHC genes in a migratory songbird. P R Soc B 272:1511–1518.

Wigley, P., and P. Kaiser. 2003. Avian cytokines in health and disease. Revista Brasileira de Ciência Avícola 5:1-14.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. BioScience 48:607–615.

Wilson, B. A., D. A. Petrov, and P. W. Messer. 2014. Soft Selective Sweeps in Complex Demographic Scenarios. Genetics 198:669–684.

Wlasiuk, G., and M. W. Nachman. 2010. Adaptation and Constraint at Toll-Like Receptors in Primates. Molecular Biology and Evolution 27:2172–2186.

Wong, W. S. W., Z. Yang, N. Goldman, and R. Nielsen. 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, N. A., T. R. Gregory, and C. C. Witt. 2014. Metabolic "engines" of flight drive genome size reduction in birds. P R Soc B 281:20132780–20132780.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. MBE 24:1586–1591.

Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.

Young, H. S., D. J. McCauley, M. Galetti, and R. Dirzo. 2016. Patterns, Causes, and Consequences of Anthropocene Defaunation. Annu. Rev. Ecol. Evol. Syst. 47:333–358.

Yu, G., L.-G. Wang, Y. Han, and Q.-Y. He. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16:284–287.

Zhang, G., B. Li, C. Li, M. T. P. Gilbert, E. D. Jarvis, J. Wang, Avian Genome Consortium. 2014a. Comparative genomic data of the Avian Phylogenomics Project. GigaSci 3:26.

Zhang, G., C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. WHITNEY, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K.-W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldsa, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang, Avian Genome Consortium, E. D. Jarvis, M. T. P. Gilbert, and J. Wang. 2014b. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zhang, Q., and S. V. Edwards. 2012. The Evolution of Intron Size in Amniotes: A Role for Powered Flight? Genome Biol Evol 4:1033–1043.

Zhang, Q., G. E. Hill, S. V. Edwards, and N. Backström. 2014c. A house finch (Haemorhous mexicanus) spleen transcriptome reveals intra- and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. BMC Genomics 15:305.

Zink, R. M., J. Klicka, and B. R. Barber. 2004. The tempo of avian diversification during the Quaternary. Phil Trans Roy Soc B 359:215–220.

Zuccon, D., R. Prŷs-Jones, P. C. Rasmussen, and P. G. P. Ericson. 2012. The phylogenetic relationships and generic limits of finches (Fringillidae). MPE 62:581–596.