REVIEW

Genomic approaches to understanding population divergence and speciation in birds

David P. L. Toews,[1,2]*# Leonardo Campagna,[1,2]# Scott A. Taylor,[1,2]# Christopher N. Balakrishnan,[3] Daniel T. Baldassarre,[4] Petra E. Deane-Coe,[1,2] Michael G. Harvey,[5] Daniel M. Hooper,[6] Darren E. Irwin,[7] Caroline D. Judy,[5,8] Nicholas A. Mason,[1,2] John E. McCormack,[9] Kevin G. McCracken,[10] Carl H. Oliveros,[11] Rebecca J. Safran,[12] Elizabeth S. C. Scordato,[12] Katherine Faust Stryjewski,[13] Anna Tigano,[14] J. Albert C. Uy,[4] and Benjamin M. Winger[6,15]

[1] Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Ithaca, New York, USA
[2] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA
[3] Department of Biology, East Carolina University, Greenville, North Carolina, USA
[4] Department of Biology, University of Miami, Coral Gables, Florida, USA
[5] Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana, USA
[6] Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois, USA
[7] Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada
[8] Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
[9] Moore Laboratory of Zoology, Occidental College, Los Angeles, California, USA
[10] Department of Biology, University of Miami, Coral Gables, Florida, USA
[11] Department of Ecology and Evolutionary Biology and Biodiversity Institute, University of Kansas, Lawrence, Kansas, USA
[12] Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, Colorado, USA
[13] Department of Biology, Boston University, Boston, Massachusetts, USA
[14] Department of Biology, Queen's University, Kingston, Ontario, Canada
[15] Integrative Research Center, The Field Museum, Chicago, Illinois, USA
* Corresponding author: toews@cornell.edu
# These authors contributed equally to the paper.

ABSTRACT

The widespread application of high-throughput sequencing in studying evolutionary processes and patterns of diversification has led to many important discoveries. However, the barriers to utilizing these technologies and interpreting the resulting data can be daunting for first-time users. We provide an overview and a brief primer of relevant methods (e.g., whole-genome sequencing, reduced-representation sequencing, sequence-capture methods, and RNA sequencing), as well as important steps in the analysis pipelines (e.g., loci clustering, variant calling, whole-genome and transcriptome assembly). We also review a number of applications in which researchers have used these technologies to address questions related to avian systems. We highlight how genomic tools are advancing research by discussing their contributions to 3 important facets of avian evolutionary history. We focus on (1) general inferences about biogeography and biogeographic history, (2) patterns of gene flow and isolation upon secondary contact and hybridization, and (3) quantifying levels of genomic divergence between closely related taxa. We find that in many cases, high-throughput sequencing data confirms previous work from traditional molecular markers, although there are examples in which genome-wide genetic markers provide a different biological interpretation. We also discuss how these new data allow researchers to address entirely novel questions, and conclude by outlining a number of intellectual and methodological challenges as the genomics era moves forward.

Keywords: evolution, next-generation sequencing, population genomics, transcriptomics

Estrategias genómicas para entender la divergencia poblacional y la especiación en aves

RESUMEN

La aplicación generalizada de métodos de secuenciación de alto rendimiento para el estudio de procesos evolutivos y patrones de diversificación ha dado lugar a muchos descubrimientos importantes. Sin embargo, el uso de estas tecnologías así como la interpretación de los datos resultantes pueden ser intimidantes para aquellos investigadores sin experiencia previa. En este trabajo presentamos un resumen y una breve introducción a los métodos relevantes (e.g., secuenciación de genomas completos, secuenciación de librerías de representación reducida, métodos de captura de secuencias, y secuenciación de ARN), así como pasos importantes en los protocolos de análisis (e.g.,

agrupamiento de loci, asignación de variantes alélicas, ensamblaje de genomas y transcriptomas completos). También presentamos ejemplos de aplicaciones en las que los investigadores han usado estas tecnologías para responder preguntas relacionadas con la evolución de las aves. Resaltamos cómo las herramientas genómicas ayudan al avance de la ciencia al discutir sus contribuciones en tres aspectos importantes de la historia evolutiva de las aves. Nos enfocamos en 1) inferencias generales sobre biogeografía e historia biogeográfica, 2) patrones de flujo génico y aislamiento genético luego de contacto secundario e hibridación, y 3) la cuantificación de los niveles de divergencia genómica entre taxones cercanamente emparentados. Encontramos que en muchos casos los datos de secuenciación de alto rendimiento confirman los resultados de trabajos previos con marcadores moleculares tradicionales, aunque también existen ejemplos en los que el muestreo de marcadores a nivel genómico provee una interpretación biológica diferente. Finalmente, discutimos cómo estos nuevos datos permiten abordar preguntas completamente nuevas y concluimos delineando una serie de desafíos metodológicos e intelectuales de cara hacia el futuro en la era de la genómica.

Palabras clave: evolución, genómica poblacional, secuenciación de nueva generación, transcriptómica.

INTRODUCTION

In the 5 years since Lerner and Fleischer (2010) published their *Perspective* on the "Prospects for the use of next-generation sequencing methods in ornithology," the application of genomic methods to the study of avian systems has burgeoned (as it has with many other taxonomic groups). New technologies are making it possible to address important, long-standing questions in avian systematics and evolution in greater detail and are opening up exciting new avenues of research. Avian systems are particularly well suited for genomic studies: Unlike many other taxonomic groups, birds exhibit conserved genome size, chromosome structure, and ploidy across deeply divergent taxonomic levels (Ellegren 2010, 2013). Avian genomes also appear to have fewer repetitive elements and fewer genome rearrangements—additional characteristics that, in some cases, allow researchers to take advantage of existing genomic resources, such as high-quality annotated reference genomes, that were developed in other, sometimes distantly related, avian taxa (Zhang et al. 2014). Additionally, the relatively small size of the avian genome (Gregory et al. 2007) means that genomic studies are comparatively cheaper in birds than in many other vertebrates (i.e. a smaller genome means there is less to sequence). Here, we summarize a number of ongoing projects in which investigators are using genomic tools to study genetic variation and speciation in nonmodel avian systems. We first provide a primer for some of the important methods that the field is employing and then review a number of relevant empirical applications. Our major foci are case studies of closely related taxa, which are particularly relevant for research into the early stages of avian speciation, although we also discuss research concerning deeper levels of divergence. This synthesis builds upon a symposium held in Estes Park, Colorado, USA, on September 28, 2014, during the joint conference of the American Ornithologists' Union, Cooper Ornithological Society, and Society of Canadian Ornithologists. The participants of that symposium have helped map out the relevant tools involved in generating genomic data and the ornithological and evolutionary questions they are addressing. We have divided our review into 3 general sections: a short summary of current methods in the field of evolutionary genomics, a discussion of how these methods are being applied in particular avian systems, and a short summary of the current challenges and next steps that these tools present.

METHODS

"Next-generation" and "high-throughput" sequencing are arguably unfortunate monikers, given that the methods that fall under these umbrella terms can include a variety of different sequencing technologies, library preparation methods (i.e. the lab protocols used to prepare DNA for sequencing), and bioinformatics analysis pipelines. Genomic research in avian systems has thus far been biased toward reduced-representation sequencing strategies and the Illumina sequencing platform (Figure 1). Here, we briefly summarize several relevant methods and outline some practical considerations that researchers must keep in mind when designing a study utilizing genomic tools. Table 1 serves as a starting point for determining which techniques might be appropriate for addressing different broad biological questions, while also taking into account common methodological constraints, such as the types of samples that are available. We discuss whole-genome sequencing and resequencing, reduced-representation libraries, sequence-capture methods, and RNA sequencing. We then review some of the more important aspects of the analysis pipeline for most applications, including data filtering, read alignment, and variant calling.

Whole-genome Sequencing and Resequencing

The Domestic Chicken (*Gallus gallus*) and Zebra Finch (*Taeniopygia guttata*) were among the first vertebrates for which fully sequenced genomes became publicly available and were completed using traditional Sanger sequencing (International Chicken Genome Sequencing Consortium
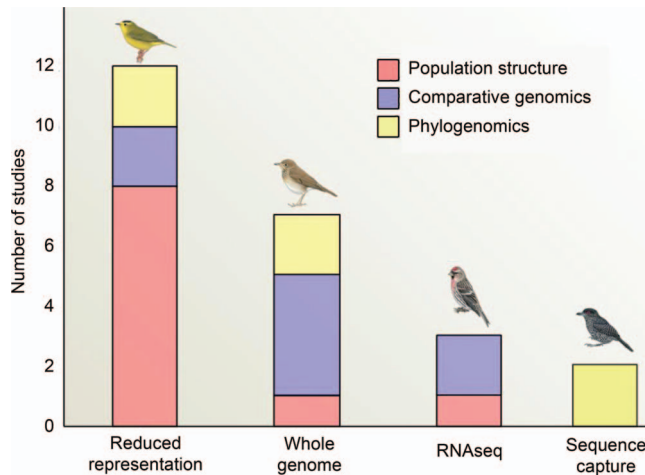
FIGURE 1. A review of the numbers of studies in avian systems in which genomic data have been applied, grouped by the main genomic technique that was used (horizontal axis), as well as a general breakdown of the theme and/or goal of the studies (colored bars). Our search criteria were designed to find all avian studies that used high-throughput, short-read sequencing data. Above each library preparation method, we include an illustration of a species that was studied using this genomic technique. From left to right: Wilson's Warbler (*Cardellina pusilla*; Ruegg et al. 2014b), Swainson's Thrush (*Catharus ustulatus*; Delmore et al. 2015), Common Redpoll (*Acanthis flammea*; Mason and Taylor 2015), and Fasciated Antshrike (*Cymbilaimus lineatus*; Smith et al. 2014). Illustrations are reproduced with permission from the *Handbook of the Birds of the World* (del Hoyo et al. 2011). We include each system only once in the figure; where multiple library strategies were used to assay a single system, we use the method indicated in Table 2.

2004, Warren et al. 2010). The conserved genome size and chromosomal synteny exhibited by avian taxa mean that these now well-annotated genomes provide useful structural information for researchers working in other avian systems or employing newer sequencing technologies. Thanks to advances in sequencing technology, several international collaborations have now produced many more avian genomes, including the 48 analyzed by Jarvis et al. (2014) and Zhang et al. (2014) in their comparative studies of genomes across various avian families (for a more detailed review of the specific advances resulting from this project, see Kraus and Wink 2015). Moreover, with dropping costs, sequencing and assembling full genomes does not necessarily require large research consortiums (e.g., Frankl-Vilches et al. 2015), which were common in the early avian genome projects.

Whole-genome information allows researchers to address questions regarding the conservation or divergence of genomic architecture between related groups (Ellegren et al. 2012, Delmore et al. 2015, Singhal et al. 2015), evaluate avian evolutionary relationships (Jarvis et al. 2014, Lamichhaney et al. 2015), explore patterns of adaptive molecular evolution (Nam et al. 2010), and/or link genes to

phenotypes (Shapiro et al. 2013, Poelstra et al. 2014). All of these applications make use of data from the entire genome, typically from a single individual, sequenced at high read depth (e.g., >90×). "Depth" refers to the same nucleotide sequenced multiple times (in this case, an average of 90×) and is sometimes also referred to as "coverage." In some cases it can be useful to subsequently map low-coverage (e.g., 10×) population-level genome sequences to this high-coverage reference genome (i.e. resequencing; Poelstra et al. 2014).

Genome sequencing usually requires the preparation of libraries of different fragment sizes. Once sequenced, high-quality short fragments, such as those generated by the Illumina platform, are bioinformatically aligned to create contigs and, subsequently, scaffolds. To achieve long scaffolds, contigs are linked using information from longer-fragment libraries. Scaffold length can be further improved by taking advantage of other sequencing technologies that generate longer reads. For example, the Pacific Biosciences (PacBio) platform produces read lengths of ∼20 kb. Although reads from this platform have a much higher error rate than those derived from Illumina, they can serve as a backbone for assembly of high-quality scaffolds obtained from the shorter Illumina reads (Koren et al. 2012). Deep sequencing can address this high error rate, but it comes at a significant monetary cost.

The genome-sequencing and population-resequencing approach was used to study genomic patterns of diversification between sister taxa of hybridizing Collared and Pied flycatchers by Ellegren et al. (2012) and between Hooded and Carrion crows by Poelstra et al. (2014) (scientific names of all study species are provided in Table 2). Poelstra et al. (2014) aligned low-coverage genomes from population samples to a high-coverage reference genome from 1 individual to identify variants putatively underlying phenotypic differences between these crow species. Genes potentially involved in *cis*-regulation of feather pigmentation and visual perception were found to be divergent on the basis of various metrics and, thus, are considered candidate loci potentially involved in premating reproductive isolation. For systems in which genome-wide divergence is higher, distinguishing background differentiation from that of genomic regions associated with phenotypic divergence will be more challenging. Genome-wide F_{ST} outlier analyses (Shapiro et al. 2013) and association studies across taxa with convergent traits (Zhang et al. 2014) have provided interesting insights into the genetic basis of phenotypic variation in birds. The use of captive crosses in more traditional genome-wide association studies (e.g., Colosimo et al. 2004) and association mapping in natural hybrid zones (e.g., Pallares et al. 2014) also appear promising, and the application of these methods in avian systems is on the horizon.

TABLE 1. A comparison of genomic methods, with accompanying details to consider when designing a study. The techniques can be used for the research goals outlined in Figure 1 (e.g., population structure, comparative genomics, and phylogenomics), but the most appropriate choice depends on the study's goals.

Genomic method	Possible evolutionary questions	Scale of divergence of the study taxa	Sampling of the genome	Type of samples needed
Whole-genome sequencing and resequencing	Fine-scale patterns of genomic divergence; phylogenetic relationships; linking genes to phenotypes	Both deep and shallow divergence (with resequencing particularly focused on the latter)	Sequences the entire genome	Large amounts of high-quality genomic DNA, at least for initial de novo genome assembly
Reduced-representation strategies (RADseq and GBS)	Population structure and divergence; broad-scale patterns of genomic divergence	Generally shallow divergence	Randomly samples the genome using restriction enzymes	Large amounts of high-quality genomic DNA
Sequence capture	Phylogenetic relationships and population structure	Both shallow and deep divergence	Targets areas of the genome chosen by the researcher	Can be used with degraded DNA from museum study skins
RNA sequencing	Quantifying gene expression; population structure	Shallow or intraspecific	Targets the expressed portion of the genome	Preserved RNA samples

Reduced-representation Strategies for Genomic Sampling

Reduced-representation libraries focus on a subset of the genome and thereby reduce the total amount of sequencing per individual. Rather than sequencing an entire genome, restriction-enzyme, digestion-based, reduced-representation approaches involve sequencing the regions flanking the restriction sites of the enzymes that are used to digest the genome (reviewed in Davey et al. 2011). Two popular techniques that use this strategy are genotyping-by-sequencing (GBS) and restriction-site-associated DNA sequencing (RADseq; for more detail on the differences between these 2 methods, see Davey et al. 2011). The thousands of largely unlinked markers generated by these methods have proved adequate in addressing a number of biological questions, for example in quantifying levels of differentiation among populations (Ruegg et al. 2014b) and understanding patterns of introgression in hybrid zones (Taylor et al. 2014a). The cost savings of reduced-representation methods allow researchers, depending on the sampling approach, to sequence larger sample sizes, sample a larger number of populations, or increase sequencing coverage across smaller regions of the genome. Moreover, when assessing population-level differentiation, information on individual genotypes is not necessarily required: Sampling more individuals at the expense of coverage depth (as low as 1×) can provide greater information about population-level allele frequencies and can lead to additional cost savings (Buerkle and Gompert 2013). Although in some cases the numbers of loci assayed between traditional approaches, such as amplified fragment length polymorphisms (AFLPs), and high-through-

put sequencing are similar, having short-read sequence data can be advantageous for multiple reasons. Most importantly, having information about where in the genome markers reside can speak to the architecture of divergence (Taylor et al. 2014a) and/or to whether differentiated loci are clustered in similar chromosomal locations (Parchman et al. 2013).

To obtain homologous loci in reduced-representation sequencing projects, restriction sites must be conserved across the individuals being compared. This may not be the case at deep scales of divergence, making RADseq and GBS best suited for population-level studies or studies between closely related taxa. It is also important to note that the choice of restriction enzyme for these techniques may bias the way in which the genome is sampled. DaCosta and Sorenson (2014) observed that enzymes with GC-rich recognition patterns led to subsequent overrepresentation of GC-rich genomic regions. Repetitive sequences can also affect the quality of RADseq and GBS fragment libraries by generating paralogous loci. However, precautions can be taken to eliminate paralogs using bioinformatics (see below).

The number of loci generated using RADseq or GBS can be adjusted by choosing restriction enzymes with different lengths of cleavage sites. Enzymes with longer recognition sites (e.g., *SbfI*, an 8-base cutter) are expected to find fewer matches throughout the genome and will consequently generate smaller numbers of loci than an enzyme with a shorter cut site (like *MspI*, a 4-base cutter). Alternatively, 2 enzymes can be combined, which further reduces the assayed portion of the genome to only those regions in which a restriction site from each

enzyme occurs in close proximity. This distinguishes the single-digest approach from the double-digest RAD (ddRAD) sequencing described in Peterson et al. (2012) and recently reviewed in DaCosta and Sorenson (2014). An additional reduction in complexity can be achieved by sequencing a subset of the fragment library, such as sequencing only a narrow range of fragment lengths. This practice, known as "size selection," further reduces the number of loci recovered but increases the chance that homologous loci are recovered in many individuals. Size selection is not incorporated into most GBS protocols (e.g., Elshire et al. 2011). In theory, reduced-representation techniques require high-quality DNA, which presumably makes them less suitable for working with samples derived from museum study skins from which the DNA obtained is often degraded (Besnard et al. 2015), although this assumption has not been rigorously tested.

Sequence Capture

Sequence-capture methods use primers or probes to enrich genomic DNA for regions of interest (Peñalba et al. 2014). Most approaches have used sequence capture to target highly conserved regions such as ultraconserved elements (UCEs; Faircloth et al. 2012) or conserved exons (Bi et al. 2013). This approach allows loci to be obtained from diverse species using the same probe set, and then sequenced on high-throughput platforms (Mamanova et al. 2010). For ultraconserved elements, regions flanking the conserved sequences contain variation that can be used for many analyses, from population genetics (Smith et al. 2014) to phylogenetics (e.g., McCormack et al. 2013, Keping et al. 2014). A bonus feature of this method is that it appears to be well suited for samples with low-quality or low-concentration DNA, such as those found in museum natural-history collections (Gansauge and Meyer 2013). Depending on the length of targeted loci, sequence-capture protocols can result in contigs that are generally longer (e.g., 500+ base pairs [bp]) than those obtained using RADseq and GBS. These longer sequences can be useful for some population and phylogenetic analyses. For example, building species trees and gene trees usually requires haplotype data with multiple SNPs across a locus, as well as nonvariant sites, and benefits from longer sequences. Sequence-capture approaches can also be used to examine regions of the genome that encode proteins or regulatory sequences that play an important role in controlling gene expression, although this approach has been used mostly in nonavian systems to date. Examples of such approaches may involve targets associated with physiological pathways that have been studied in humans or in traditional model organisms (reviewed in Jones and Good 2015).

RNA Sequencing (RNAseq)

RNAseq targets the genes that are being expressed in a given tissue at the time of sampling and is used primarily to study differential patterns of gene expression between intraspecific treatment groups in an experimental context (e.g., Stager et al. 2015). However, because differentiation of gene expression between populations or species may be particularly relevant during the early stages of divergence (e.g., Poelstra et al. 2014, Mason and Taylor 2015), the use of RNAseq data is becoming increasingly common in the avian evolution literature. Special care must be taken to collect and store samples in a way that preserves RNA (see Cheviron et al. 2011), a procedure that is not routine in most field or avian-museum-collection protocols (but see Balakrishnan et al. 2014). Careful attention to experimental design is also important for distinguishing differences in gene expression due to environmental variation from those generated by population- or species-level differentiation. For example, circadian effects and tissue-specific expression patterns have the potential to overwhelm more subtle between-population or between-species variation (e.g., Storch et al. 2002). Additionally, the correct tissue and ontogenetic stage must be selected to make the appropriate comparisons (for a review on the use of RNAseq for gene-expression studies, see Wang et al. 2009).

Coding DNA, which is targeted by RNAseq, makes up only a small portion of the overall genome. As such, sequencing messenger RNA is an additional way to reduce genomic complexity and allows users to generate markers for population genetic studies. When used to generate loci for population genetic studies, the typical RNAseq workflow involves mRNA extraction, conversion of mRNA to complementary DNA (cDNA) libraries, and sequencing of multiplexed cDNA libraries at high depth of coverage. After appropriate quality filtering, sequence data can be used to assemble a reference transcriptome. Aligning individual transcript libraries to the reference transcriptome can then be used to identify variants, and the loci of interest can be genotyped for large numbers of individuals using a different platform (e.g., Sequenome, SNPchip) that requires only genomic DNA. Although these panels are very useful for the same species in which they where developed, they can be less appropriate for more distantly related species, or even populations, because of ascertainment biases. When RNAseq is used for marker development, it is generally at the population level or for taxa with shallow divergence; thus far, it has been used primarily as a way to develop markers for paternity analyses (e.g., Weinman et al. 2015).

Because gene expression varies between, and often within, individuals (e.g., allele-specific expression), the use of SNP data derived from RNAseq has the potential to bias genotype calls and result in inaccurate downstream

TABLE 2. A summary of avian studies that applied high-throughput, short-read sequencing data. The library preparation methods and a coarse summary of the general research questions of each study are quantified in Figure 1.

Study taxa	Focus	Library preparation	Reference
Studies within a species complex or between species pairs			
Afrocanarian Blue Tit (*Cyanistes teneriffae*)	Population structure: resolving relationships within the complex	RADseq	Gohli et al. 2015
Black-capped Chickadee (*Poecile atricapillus*) and Carolina Chickadee (*P. carolinensis*)	Population structure: hybrid-zone dynamics and influence of climate change	RADseq	Taylor et al. 2014a, 2014b
Capuchino seedeaters (*Sporophila* spp.)	Population structure: resolving relationships within the complex and understanding patterns of genomic divergence	RADseq	Campagna et al. 2015
Collared Flycatcher (*Ficedula albicollis*) and Pied Flycatcher (*F. hypoleuca*)	Comparative genomics: resolving patterns of divergence and gene flow between species	Whole-genome resequencing	Ellegren et al. 2012
Darwin's finches (*Geospiza* spp.)	Phylogenomics: resolving relationships among the ecologically distinct groups	Whole-genome resequencing	Lamichhaney et al. 2015
Great Tit (*Parus major*)	Comparative genomics: characterizing chromosomal variation mapping to the Zebra Finch genome	Whole-genome sequencing [a] and RNAseq	van Bers et al. 2010, Santure et al. 2011
Greenish Warbler (*Phylloscopus trochiloides*)	Population structure: quantifying divergence and extent of gene flow	RADseq	Alcaide et al. 2014
Hawaiian Amakihi (*Hemignathus virens*)	Phylogenomics: resolving relationships among Hawaiian honeycreepers	Whole genome sequencing [a] and RADseq	Callicrate et al. 2014
Gunnison Sage-Grouse (*Centrocercus minimus*) and Greater Sage-Grouse (*C. urophasianus*)	Population structure: identifying species limits and units for conservation	RADseq	Oyler-McCance et al. 2015
Hoary Redpoll (*Acanthis hornemanni*) and Common Redpoll (*A. flammea*)	Population structure: quantifying divergence between species	RADseq and RNAseq [a]	Mason and Taylor 2015
Hooded Crow (*Corvus [corone] cornix*) and Carrion Crow (*C. [c.] corone*)	Comparative genomics: estimating patterns of genomic divergence and linking genetic variants with plumage characteristics	Whole-genome resequencing	Poelstra et al. 2014
Mallard (*Anas platyrhynchos*)	Comparative genomics: comparison of wild vs. domesticated taxa	Whole-genome sequencing	Kraus et al. 2011
New Guinean crowned pigeons (*Goura* spp.)	Phylogenomics: resolving relationships among 3 endemic pigeons	Other [b]	Besnard et al. 2015
Peruvian Tyrannulet (*Zimmerius viridiflavus*) and Golden-faced Tyrannulet (*Z. chrysops*)	Phylogenomics: estimating patterns of divergence and relationships within the group	RADseq	Rheindt et al. 2014
Plain Xenops (*Xenops minutus*)	Population structure: resolving relationships within the species and testing the role of geographic barriers	RADseq	Harvey and Brumfield 2015
Red-backed Fairy-wren (*Malurus melanocephalus*)	Population structure: testing for evidence of concordance between genomic and plumage patterns across a hybrid zone	RADseq	Baldassarre et al. 2014
Reunion Island White-eye (*Zosterops borbonicus*)	Comparative genomics: SNP generation and comparison with Zebra Finch	RADseq	Bourgeois et al. 2013
Swainson's Thrush (*Catharus ustulatus*)	Comparative genomics: quantifying patterns of differentiation and testing whether migration genes were associated with islands of divergence	RADseq; Pooled whole genome sequencing [a]	Ruegg et al. 2014a, Delmore et al. 2015
White-collared Manakin (*Manacus candei*) and Golden-collared Manakin (*M. vitellinus*)	Comparative genomics: estimating patterns of divergence and gene flow between the species	RADseq	Parchman et al. 2013
Wilson's Warbler (*Cardellina pusilla*)	Population structure: estimating levels of divergence and patterns of connectivity between and within eastern and western groups	RADseq	Ruegg et al. 2014b

The Auk: Ornithological Advances 133:13–30, © 2016 American Ornithologists' Union

TABLE 2. Continued.

Study taxa	Focus	Library preparation	Reference
	Studies of multiple species or deep scales of divergence		
Representatives of distinct avian families	Comparative genomics: functional divergence and deep-level relationships within birds	Whole-genome sequencing	Jarvis et al. 2014, Zhang et al. 2014
Multiple species from Galloanserae	Comparative genomics: sex-biased expression of genes on sex chromosomes	RNAseq	Wright et al. 2014
Representatives of distinct avian families within Neoaves	Phylogenomics: understanding deep-level relationships within Neoaves	UCEs	McCormack et al. 2013
Widespread Neotropical species	Phylogenomics: resolving population structure and biogeographic patterns	UCEs	Smith et al. 2014
Multiple species pairs within the Amazonia basin	Population structure: quantifying levels of divergence and gene flow across river barriers	RADseq	Weir et al. 2015

[a] Study used multiple methods; we indicate the method that is represented in Figure 1.
[b] Degraded and fragmented DNA from museum skins was used as template for sequencing of mitochondrial genomes and a subset of nuclear genes, using a whole-genome sequencing protocol.

estimates of various population genetic parameters. This problem should be carefully considered by investigators who choose RNAseq over digestion-based reduced-representation approaches to develop SNP markers. SNP identification by RNAseq may, however, be useful when investigators are interested in developing SNP genotyping assays in the absence of a reference genome, or if a study is investigating potentially adaptive differences in coding regions or gene networks between populations or species. RNAseq provides ample flanking sequence for the development of these assays, which require primers to be designed around a region with a SNP of interest. In the absence of a reference genome, digestion-based reduced-representation approaches generally do not provide sufficiently large flanking regions for primer design, although this will change as the read lengths produced by sequencing platforms continue to increase. Fortunately, the growing number of available avian genomes, combined with their generally conserved architecture (see above), may provide investigators with the resources necessary to generate genotyping panels from reduced-representation data and avoid the potential pitfalls of using RNAseq derived SNPs for population genetic investigations.

Important Aspects of the Bioinformatics Analysis Pipeline

One important issue with assembly and variant calling is how best to align similar sequences, particularly in cases where a suitable reference genome is not available. In order to assemble the thousands to millions of reads obtained from a sequencing run, sequences must be clustered by sequence identity. The identity thresholds used for clustering, however, are subjective and may result

in the assembly of nonhomologous reads if they are too liberal, or may separate reads coming from the same locus if they are too conservative (Harvey et al. 2015). Conservative thresholds may remove variable haplotypes by oversplitting them into separate loci (Harvey et al. 2015). This has the tendency to decrease the variation present in resulting datasets and can bias downstream phylogeographic and population genetic analyses (Harvey et al. 2015). More liberal thresholds prevent the loss of divergent haplotypes but allow the assembly of some nonorthologous reads into the same locus (Harvey et al. 2015). Fortunately, loci containing nonorthologous reads are often identifiable by the presence of >2 alleles within an individual (at least in diploid species such as birds). A best practice is to assemble loci under a series of different similarity thresholds in order to identify the point where the loss of alleles diminishes (see also Ilut et al. 2014). This approach both reduces the presence of artifacts in datasets and also will permit comparisons across datasets from different species and genomic regions.

Variant calling is another significant step in the analysis of high-throughput sequence data. The data generated from high-throughput sequencing technologies, while voluminous, also contain errors that can be incorrectly interpreted as true sequence variation. Reducing PCR amplification cycles and employing high-fidelity polymerases can be used to minimize errors during PCR (Brelsford et al. 2012), but all sequencing platforms introduce error. DNA sequences generated using traditional Sanger sequencing are relatively long, are associated with conserved primers with known genomic positions, and tend to contain few errors. By contrast, sequences from many new platforms are short, may be derived from virtually anywhere in the genome, and tend to contain more errors

(Shendure and Ji 2008). It is therefore important to carefully filter low-quality base calls, SNPs, or insertions/ deletions. While a thorough discussion of this topic is beyond the scope of our review, popular programs with robust variant calling, filtering, and flexible clustering capabilities are the Stacks pipeline (Catchen et al. 2011), PyRAD (Eaton 2014), and the Genome Analysis Toolkit (GATK; McKenna et al. 2010).

APPLICATIONS

The molecular tools used to study genetic patterns across a wide variety of nonmodel taxa are rapidly moving away from methods that focus on a limited set of targeted nuclear and mitochondrial loci, and toward high-throughput methods that target broad portions of the genome. Moreover, depending on the application, the costs associated with new methods have decreased such that they are comparable to, if not cheaper than, traditional methods. These tools are now being applied to avian systems to address complex questions about ecology, evolution, and patterns of diversification. To illustrate the trends in the field and to understand what general questions these data are being used to address, we searched extensively for avian studies that have utilized high-throughput, short-read sequencing data (e.g., excluding microsatellite, AFLP, micro- and SNP-array studies) by looking manually through citations. In each case, we noted which library preparation method was used and the general goal of the study. Figure 1 and Table 2 summarize the various avian studies in which this kind of sequencing has been applied. Thus far, reduced-representation approaches have been the most popular genomic methods used in nonmodel avian systems and are generally used to understand population structure. It is also notable that most, if not all, genomic studies have been focused on passerines, a taxonomic bias that hopefully will be overcome in the future. Here, we discuss the contributions of genomic tools to 3 important large-scale facets of avian evolutionary history by focusing on general inferences about biogeographic history, patterns of gene flow and isolation upon secondary contact and hybridization, and quantifying levels of genomic divergence between closely related taxa.

Biogeographic Patterns of Differentiation
A simple yet important question in the application of high-throughput sequencing data is whether the genetic and biogeographic patterns are similar between molecular datasets that differ in size by orders of magnitude. Not surprisingly, in phenotypically variable taxa in which very little genetic differentiation was resolved using a low number of markers, the larger number of markers afforded by next-generation sequencing data are usually able to

resolve some structure. For example, drawing upon a similar sampling design, Ruegg et al. (2014b) used RADseq to revisit population structure within Wilson's Warblers, which Irwin et al. (2011) originally assayed with amplified-fragment-length-polymorphism data (AFLPs; Table 2). Wilson's Warblers show subtle plumage differences between subspecies groups found in western and eastern North America, which are concordant with differences in mtDNA and AFLPs (Kimura et al. 2002, Irwin et al. 2011). The broad patterns between the AFLP and RADseq datasets were similar, splitting the species into well-resolved western and eastern groups and adding support for possible species-level differences. However, Ruegg et al. (2014b) also found evidence of structuring within the western subspecies, a pattern that AFLPs did not identify, although some structuring had been observed in earlier studies of mitochondrial DNA (Kimura et al. 2002, Paxton et al. 2013). Although a robust analysis of possible contact areas will be necessary to determine the extent of interbreeding between these groups, the ability to use thousands of genetic markers to detect low levels of differentiation is important for researchers interested in migratory connectivity and for conservation practitioners more generally.

With more power and the sensitivity to detect subtle genetic differentiation, it is also possible for genome-wide datasets to reveal different biogeographic interpretations. For example, Alcaide et al. (2014) used a GBS approach to study the Greenish Warbler ring species complex in Asia (Table 2). This group is well known in the avian speciation literature and was considered one of the best examples of speciation-with-gene-flow within a ring species complex (Coyne and Orr 2004, Price 2008). Centered on the Tibetan plateau, evidence suggested that the species complex formed when, over long periods, populations from the south slowly expanded their range northward, in parallel on either side of the plateau (Irwin et al. 2001a). After finally coming into contact after many generations in central Siberia, these populations appeared to be reproductively isolated. Moreover, the current distribution of the taxa was thought to trace the historical evolution of various phenotypic traits that could play a key role in reproductive isolation, such as song and seasonal migratory behavior (Irwin and Irwin 2005). Previous research using AFLPs found a gradual change around the ring, with the terminal forms in the north showing the strongest genetic differences (Irwin et al. 2005).

Revisiting the same samples that were analyzed by Irwin et al. (2005), the genomic data of Alcaide et al. (2014) revealed a more nuanced biogeographic history of this group. For instance, across the southern portion of the complex, there are 2 deeply divergent mtDNA clades that overlap in northern India. Alcaide et al. (2014) showed that, although there is nuclear gene flow between the 2

mtDNA groups, there is also evidence of historical isolation, and there is strong divergence in the nuclear genome across the mtDNA divide. The narrow transition between them suggests either recent contact or some level of current assortative mating and/or selection against hybrid offspring. This finding led Alcaide et al. (2014) to conclude that while the Greenish Warbler should still be considered a ring species according to the concept originally conceived by Mayr (1942) and others (reviewed by Irwin et al. 2001b), it should no longer be considered a clear example of speciation-by-distance (Alcaide et al. 2014).

There are also some instances in which genome-wide markers do not provide additional resolution compared with traditional methods. Mason and Taylor (2015) discuss the case of the redpoll finches as an example. Traditional sequencing and genotyping methods have not revealed any significant genetic differences between Hoary Redpolls and Common Redpolls, despite phenotypic variation in plumage, bill size, and shape (Table 2). Examining thousands of loci using a RADSeq dataset revealed that this lack of differentiation is pervasive among the genomes of the currently recognized species. Mason and Taylor (2015) suggest that this is consistent with redpolls comprising a single, global metapopulation with no apparent population structure. However, bringing RNAseq data to bear on the situation suggests that overall, gene expression is correlated with continuous phenotypic variation among redpolls. Mason and Taylor (2015) indicate that this pattern of differential gene expression despite negligible differentiation among anonymous loci may be due to high levels of ongoing gene flow between polymorphic populations, incomplete lineage sorting, polymorphism in *cis*-regulatory elements, or phenotypic plasticity. More generally, however, these findings suggest that the 2 species may be better treated as populations at the ends of a phenotypic continuum rather than distinct, isolated species.

Secondary Contact, Hybridization, and Introgression

Areas where taxa that have experienced periods of isolation come back into secondary contact and interbreed (i.e. hybrid zones) have provided unique insights into the process of speciation and adaptation (Harrison 1993). Applying genomic tools to these systems, which provides many more genetic markers than traditional approaches, allows for a number of advances over traditional tools, including (1) the ability to generate more robust and high-resolution hybrid indices, (2) the power to identify loci that are under selection and/or prone or resistant to introgression (e.g., Parchman et al. 2013, Taylor et al. 2014b), (3) the resolution necessary for admixture-mapping of genetic variants to phenotypic traits (e.g., Poelstra et al. 2014), and (4) the resolution necessary for understanding the genetic architecture of reproductive isolation at a fine scale across the genome.

Recent hybrid-zone studies have combined genomic data with traditional geographic transect analyses. In this case, the geographic distribution of allele frequencies across a hybrid zone is related to several parameters, including the age of the zone of contact and the extent of reproductive isolation in the zone (Szymura and Barton 1986). For example, loci that exhibit narrow cline shifts across an established hybrid zone are potentially the targets of strong selection, whereas broad transitions in allele frequencies across such hybrid zones are consistent with introgression and weak isolating barriers. In a study by Baldassarre et al. (2014) of the Red-backed Fairy-wren, an Australian endemic passerine, there is striking geographic discordance between clines in plumage and genetic characters (Table 2). The taxon comprises 2 distinct subspecies that differ primarily in a sexual signal: red versus orange male plumage color. Preliminary biogeographic analyses suggested that there was a hybrid zone between subspecies and that alleles for red plumage color introgressed asymmetrically into the genomic background of the orange subspecies (Lee and Edwards 2008, Baldassarre et al. 2013). However, data on the extent of hybridization were lacking and it was unclear whether the apparent pattern of asymmetrical introgression was real or an artifact of low spatial resolution during sampling.

Baldassarre et al. (2014) used a GBS approach to generate a dataset of SNPs from populations along a previously unsampled transect through the hybrid zone. Their analyses confirmed the presence of a tension zone but found that introgression was variable across loci. While the majority of loci exhibited coincident and concordant clines that were narrow in relation to a model of neutral diffusion, several loci had cline centers and/or widths that deviated significantly from this pattern. Moreover, the plumage cline was shifted significantly into the genomic background of the orange subspecies (i.e. genetically "orange" individuals with red plumage patterns). Most notably, several SNPs had similar cline-shape parameters to the plumage clines, which suggests that asymmetrical introgression could explain the displaced plumage cline. In this case, complementary experimental studies suggest that introgression of plumage color is likely driven by sexual selection through extrapair mating (Baldassarre and Webster 2013). Although the SNP data used in this analysis were sufficient to quantify differential introgression across the genome, future work would presumably focus on targeting specific genomic regions that may confer plumage-color variation (e.g., carotenoid synthesis pathways), to identify genes that may have introgressed asymmetrically from red to orange populations.

Studies analyzing hybrid-zone dynamics often investigate a single sampling period and draw conclusions from that temporal snapshot. However, stochasticity can result in loci with spurious outlier patterns, which can be exacerbated by limited temporal or geographic sampling. In genetic data, outlier loci are usually defined as those loci below or above a given F_{ST} divergence threshold, which can indicate loci putatively under balancing or positive selection, respectively, and can be identified using various methods (e.g., BayeScan; Foll and Gaggiotti 2008). Comparing admixed populations from different geographic regions may help distinguish stochastic patterns from signals of selection, allowing the detection of repeatedly divergent genomic regions. This approach has been successfully undertaken in some nonavian studies (e.g., Teeter et al. 2010, Larson et al. 2014). Temporal comparisons may also help identify consistent outlier loci, as demonstrated by a recent study of North American chickadee hybridization (Taylor et al. 2014a; Table 2). Black-capped and Carolina chickadees hybridize in a contact zone that extends from New Jersey to Kansas. The hybrid zone is likely maintained by strong intrinsic selection against hybrids, and evidence suggests that it is moving north, possibly in response to climate change (Taylor et al. 2014b).

Taylor et al. (2014a) used 2 groups of sampled individuals collected a decade apart across a transect through the chickadee hybrid zone in southeastern Pennsylvania and compared patterns of genomic introgression, within a Bayesian framework, in the two periods. They hypothesized that the most divergent loci would be more likely to be involved in reproductive isolation and would therefore be affected by temporally consistent selection. Using sequence data generated with GBS, they compared locus-specific F_{ST} values to estimates of introgression in both periods, taking a genomic cline approach (Gompert and Buerkle 2011a, 2011b). Taylor et al. (2014a) found consistently low levels of introgression for the most highly divergent loci between the species, which is compatible with a role in reproductive isolation. Moreover, they found that these loci were significantly more likely to be located on the Z-chromosome than on the autosomes. Although they suggested that population history likely contributed to consistent patterns between decades, this remains strong evidence that the loci they identified may be linked to genomic regions involved in reproductive isolation between chickadees. This type of analysis highlights (1) the benefits of examining hybrid zones at multiple time points and (2) the added power that genomic approaches can provide when studying natural hybrid zones.

Genomic Patterns and the Architecture of Divergence

There has been considerable recent effort to document the architecture of genomic divergence between closely related taxa in terms of both variation in sequence differentiation and structural rearrangements (Feder et al. 2012, Pala et al. 2012, Yeaman 2013). These studies investigate both the evolutionary processes that contributed to patterns of divergence (e.g., selection and/or drift) and the interplay between these processes and the basic genomic architecture of a species (e.g., recombination rate and/or chromosomal inversions). At least in avian systems, most comparisons involve taxa in which reproductive barriers are incomplete and there is either current or historical evidence of hybridization. For instance, the earliest and most prominent example of this kind of dataset in birds was provided by Ellegren et al. (2012), who compared the patterns of divergence between the genomes of Collared and Pied flycatchers (Table 2). These sister taxa have a broad Eurasian distribution and hybridize at low frequency in narrow regions of overlap, including the well-studied populations on Gotland in the Baltic Sea. Ellegren et al. (2012) found a strong pattern of heterogeneous genomic differentiation between these 2 species, with large peaks of differentiation separated by regions of low divergence. In this and in other, similar studies, when authors identify genomic regions with varied levels of differentiation between populations (as measured by F_{ST}), the patterns are sometimes attributed to variation in levels of gene flow at these loci. The verbal metaphor of "divergence islands" surrounded by a sea of gene flow is commonly used to describe these patterns (reviewed in Cruickshank and Hahn 2014). More simply, these highly divergent regions are sometimes interpreted as containing genes that resist introgression, and regions of low divergence are interpreted as experiencing high levels of gene flow. These explanations follow from many years of hybrid-zone studies and a general increase in the appreciation of the "porous" nature of the genome (e.g., Wu 2001).

However, it is becoming increasingly recognized that such studies should consider the possible effects of gene flow as well as processes operating within taxa that can generate analogous patterns of variable levels of differentiation (Yeaman 2013, Cruickshank and Hahn 2014). In particular, scenarios involving historical gene flow need to be considered alongside alternatives that include differential selection, drift, and variation in mutation or recombination rates, which can generate similar patterns of increased divergence between populations (e.g., Renaut et al. 2013). For example, allopatric populations that share no gene flow but experience different levels of positive or negative selection can have reduced diversity and show similar heterogeneous divergence in discrete genomic regions (Cruickshank and Hahn 2014). Reduced recombination is a particularly important process that is thought to contribute to differences in levels of differentiation: Reduced recombination rates are evident in gene-rich regions (Stapley et al. 2010) as well as near chromosomal

centromeres (Backström et al. 2010). There is also some evidence that these "deserts of recombination" are associated with elevated levels of divergence in some avian systems, including between Collared and Pied flycatchers (Ellegren et al. 2012) and between subspecies of Swainson's Thrush (Delmore et al. 2015), although the generality of this pattern remains unknown.

One prediction that follows from the assumption that highly divergent regions are involved in reproductive isolation is that they also contain the genes that may generate reproductive barriers. The analysis of genomic patterns of divergence between coastal and inland subspecies of Swainson's Thrush by Ruegg et al. (2014a) attempted to address this (Table 2). These 2 thrush subspecies differ markedly in their migratory behavior, with the inland form migrating in the fall toward the southeast while the coastal form's route parallels the Pacific coast (Ruegg and Smith 2002, Delmore et al. 2012, Delmore and Irwin 2014). This difference in migration may be an important component of reproductive isolation between the 2 taxa, given that hybrids display intermediate and/or mixed migratory behaviors that are possibly inferior to those of the parental groups (Delmore and Irwin 2014).

One of the main motivations of Ruegg et al.'s (2014a) study was to ask (1) whether previously identified genes for migration occurred in large islands of genomic divergence and (2) whether these genes were more differentiated than one would expect by chance. They found that while patterns of divergence were heterogeneous across the genome, markers on the Z-chromosome were, on average, more differentiated than those mapped to the autosomes (as measured by F_{ST}). This fits with a variety of models, including faster Z-chromosome evolution in ZW taxa possibly due to lower effective population size, the immediate exposure of recessive favorable mutations to selection in the heterogametic sex, and/or suggestions that genes involved in speciation may be more likely to occur on the Z-chromosome (Ellegren 2011, Harrison et al. 2015). In terms of the migration genes identified a priori, Ruegg et al. (2014a) found that although they were, on a locus-by-locus basis, more differentiated than expected, they were not more likely to occur in large, differentiated regions of the genome than by chance. However, in a more recent analysis of Swainson's Thrush that used whole-genome sequencing and pooled population samples, Delmore et al. (2015) found more of these genes within divergence peaks, leaving open the debate about the role of these genes in generating reproductive barriers between these taxa. More generally, their results highlight some of the limitations of reduced-representation approaches: While they sample orders of magnitude more markers than traditional methods, they still are only sampling a small fraction of the genome (usually <1%), and this is a

limitation for studies that aim to find genes associated with phenotypes (see Campagna et al. 2015).

In the future, to directly test the role of divergent genomic regions in reproductive isolation, it will be useful to draw upon additional estimates of introgression between taxa. For example, there are statistical tests that can be used to infer levels of introgression across markers (e.g., Rheindt and Edwards 2011), in addition to more direct methods, by assaying levels of gene flow across hybrid zones for markers with varying degrees of divergence (e.g., Parchman et al. 2013, Taylor et al. 2014a). For instance, in a study of hybridizing manakins, Parchman et al. (2013) found a weak positive relationship between a marker's level of divergence and reduced introgression across the hybrid zone (Table 2). More generally, however, the genetic basis for isolating barriers in this and other avian systems is unclear. Therefore, testing for a genomic connection between these characteristics (islands of divergence, loci showing evidence of reduced introgression, and the genetic basis of speciation phenotypes) will be an important focus of future work.

When interpreting variation in levels of genomic differentiation between taxa, one must consider that divergence between species involves not only changes in DNA sequence, but also rearrangements of genome architecture. Chromosomal rearrangements, namely translocations and inversions, are often found as fixed differences between closely related species or segregating within species in most taxonomic groups (Coyne and Orr 2004, Hoffmann and Rieseberg 2008, Faria and Navarro 2010). Once established, theory suggests that inversions can foster divergence between populations (Lowry and Willis 2010, Jones et al. 2012) and aid in the speciation process (Noor et al. 2001, Rieseberg 2001, Fishman et al. 2013, Poelstra et al. 2014) because they can constitute barriers to gene flow through the suppression of recombination and/or induction of structural underdominance in heterokaryotypes.

Birds have long been used as models of speciation, but less attention has been given to the possible role of chromosome-rearrangement evolution in bird speciation (Price 2008, Ellegren 2010, 2013, Poelstra et al. 2014). The most easily observed types of macro-rearrangements (fusions, fissions, translocations, etc.), which often distinguish the karyotypes of species in other taxonomic groups, are relatively rare in birds (Ellegren 2010, 2013, Zhang et al. 2014). But one class of rearrangements, chromosome inversions, appear to occur frequently. Inversions have a long history of cytological study in birds (reviewed in Shields 1982, Christidis 1990, Hooper and Price 2015). In passerines, a review of cytological work found that chromosome inversions are a pervasive feature of genome evolution and often involve the sex

chromosomes (Hooper and Price 2015). The high frequency of inversions detected in avian genomes using cytological approaches has recently been complemented by studies that utilize genomic sequence data to find and map the breakpoints of rearrangements (Stapley et al. 2008, Backström et al. 2010, Skinner and Griffin 2011, Kawakami et al. 2014, Zhang et al. 2014). The Zebra Finch genome itself is based on a male bird heterozygous for a large inversion of the Z-chromosome, and this polymorphism segregates at different frequencies among populations (Itoh et al. 2011). One recent comparative genomic analysis reported 140 chromosome inversion differences between the genomes of the Collared Flycatcher and the Zebra Finch (Kawakami et al. 2014). The estimated mean fixation rate of 1.5–2.0 inversions Ma^{-1} between the lineages leading to these 2 species suggests again that inversions evolve rapidly enough to be an important substrate for speciation; however, the strength of the connection between reproductive-barrier loci and inversions among avian lineages is unclear. One example where this may be important is in the Hooded and Carrion crow system. In this case, initial evidence suggests that 81 of 82 fixed differences between them are located within a ~2 Mb inversion between the 2 forms and are also linked to candidate genes potentially involved in plumage differences that may be under sexual selection (Poelstra et al. 2014).

CHALLENGES AND NEXT STEPS

We have outlined and summarized exciting methodological advances that avian biologists can add to their ever-growing analytical toolbox. We have also illustrated how these methods have been applied to an important subset of avian systems and the insights into the biology of these systems that they have generated. In closing, we will highlight what we see as some important intellectual and methodological challenges and next steps in the coming years.

Intellectual Challenges

(1) *Interpreting patterns of divergence.* As noted above, the role that "islands of divergence" between closely related species play in the evolution of reproductive isolation is currently unclear. Determining whether these regions are linked to species-defining traits and to understanding the processes that contributed to the formation of these regions will be important future work. In addition, most studies have quantified patterns of genomic divergence through examination of hybridizing taxa. Comparing genomic patterns of divergence in hybridizing taxa with those of purely allopatric taxa, which have no history of postdivergence gene flow (Feder et al. 2012,

Cruickshank and Hahn 2014), will help us understand the processes that shape the genomes of closely related species.

(2) *Null models of genome-wide processes.* While patterns of genomic divergence can lead to interpretations of outliers as the product of non-neutral processes, we have a poor understanding of how genomes might evolve and become differentiated when driven purely by neutral processes. Moreover, although demographic processes are assumed to contribute to genome-wide patterns, it is possible that these effects can be more localized in the genome than previously appreciated (Yeaman and Whitlock 2011, Flaxman et al. 2013). For this, it will be important to develop and implement more efficient analytical methods, beyond simple summary statistics, that can be applied across multiple taxa and incorporate complex demographic models (e.g., Gronau et al. 2011, Campagna et al. 2015).

(3) *Conservation implications and taxonomy.* Given the power of high-throughput sequencing technologies to detect subtle genetic structure, an important challenge will be to understand how these data translate into biologically meaningful levels of gene flow (Shafer et al. 2015). This can have important conservation implications, because genetic diagnosability is a central criterion in identifying management and evolutionarily significant units (McCormack and Maley 2015, Oyler-McCance et al. 2015). These data also have taxonomic implications, such as helping us understand how heterogeneous patterns of divergence across the genome might translate into levels of reproductive isolation. Finally, most genomic studies published to date have been focused on passerines, which indicates that more studies on other avian groups are needed.

(4) *Understanding the importance of microchromosomes in avian evolution.* A key feature of avian genomes is the large number of microchromosomes. In general, these small chromosomes are not well represented in genome assemblies, even in those of model avian species like the Domestic Chicken (Ellegren 2013). For example, chromosome 16, which harbors the major histocompatibility complex (MHC), is assembled to be only ~10,000 bp in the Zebra Finch genome (Warren et al. 2010). Microchromosomes have an extremely high recombination rate, a consequence of the obligate crossing-over events in small chromosomes, a higher substitution rate, and higher gene density compared with macrochromosomes (Smith et al. 2000, Axelsson et al. 2005). Although microchromosomes are not unique to birds (Ellegren 2013), understanding how these interesting features contribute to avian evolution remains a puzzle.

Methodological Challenges

(1) *Genome annotation.* Many of the studies we have discussed in this review took advantage of annotation data associated with the Domestic Chicken and Zebra Finch genomes. Clearly, if our goal is to have insights into the putative functional basis of genetic variants in nonmodel systems, more annotation data are needed, as are data from additional species. High-quality annotation will require abundant supporting data (e.g., RNAseq) and careful curation to identify and describe the full range of transcript structures and their features (e.g., regulatory regions, splice sites). Although this is laborious, it is necessary to be strategic about the avian genomes in which the ornithological community invests, in both high-quality sequence data and associated annotations. Finally, if specific pathways are found to be important in generating phenotypes that are relevant to speciation (e.g., the melanogenesis pathway; Poelstra et al. 2014), sequence-capture methods can be used to examine protein-coding or regulatory regions across multiple species.

(2) *Bioinformatics.* Generating, understanding, and manipulating genomic data can be a challenge for new researchers. It will be important to provide adequate tools for teaching and learning bioinformatics skills. Moreover, given how library preparation, sequencing technologies, and analysis tools are rapidly changing, it will be useful to work toward generating a common set of best practices and analysis pipelines, which will allow data collected using different methods and from different systems to be compared and shared. In particular, we recommend carefully and critically evaluating the strengths and limitations of certain methods as they are applied to avian systems. For example, benchmarks are now available to assess the efficacy of several genome-assembly methods (Bradnam et al. 2013). Generating similar benchmarks for other methodologies will be very useful for experimental design and analysis using genomic methods.

(3) *Improving genome assemblies.* Next-generation sequencing has greatly facilitated the sequencing of entire avian genomes at relatively low cost. However, the product of these assemblies is a number of large scaffolds of unknown chromosomal location. Traditional linkage maps remain a valuable tool for anchoring scaffolds on chromosomes and providing both the relative position of scaffolds and their orientation. Obtaining linkage maps involves laboratory crosses or known pedigrees across multiple generations, both of which are relatively uncommon in avian systems. Chromosome-level genome assemblies can also be obtained through optical mapping. This new technology uses intact DNA molecules that are linearized on a surface to produce a map of endonuclease recognition sites across the genome. A restriction enzyme generates nicks on a molecule that are used to incorporate a fluorescent dye. Various independent molecules are then imaged and the patterns of fluorescence are overlaid to produce a consensus map of the relative positions of the initial sites in the genome. This map can then be used in combination with large sequenced scaffolds to produce chromosome-level genome assemblies and has already been applied to at least 2 assemblies of avian genomes: the Budgerigar (Ganapathy et al. 2014) and the Ostrich (Zhang et al. 2015).

(4) *Data sharing.* Data sharing is not only becoming more efficient but is also now mandated by many journals. However, sharing large amounts of genomic data, while necessary, is challenging. It will also be important to decide what stage of the data-analysis pipeline to provide. For example, it is unclear what data products will be useful for future studies: raw reads, demultiplexed reads for each individual in a study, or simply final variant calls.

GLOSSARY

assembly. Linking the large number of short reads into a full representation of an organism's genome, divided into chromosomes or scaffolds (see below). Can also be used in the context of assembling a transcriptome—the collection of messenger RNA molecules being expressed in a tissue at a given time—from short-read sequence data.

contig. A set of bioinformatically aligned, continuous sequencing reads that yield a consensus sequence.

genomic architecture. The number, distribution, and size of various genomic elements (e.g., regions of divergence, genes, or inversions).

genotyping-by-sequencing (GBS). See *reduced-representation library*.

high-throughput sequencing. A collection of sequencing platforms that generate DNA fragments in high volume and that do not rely on chain termination methods (i.e. Sanger sequencing), with the most popular current technologies generating short reads (i.e. ~100 bp).

RADseq. See *reduced-representation library*.

read depth. The number of times a base is sequenced. Also referred to as "coverage" or "depth of coverage."

reduced-representation library. A collection of genome-sequencing methods that target only specific regions of the genome. The most popular methods, GBS and RADseq, target the regions surrounding restriction-enzyme cut sites. The main differences between them are whether a size-selection step is included (RAD vs. GBS) and the number and kind of enzymes used.

RNAseq. Sequencing of genes that are being expressed in a given tissue at the time of sampling (i.e. mRNA).

scaffold. Bioinformatic alignment of contigs, defining their order and orientation. The sequence between contigs in a scaffold is often unknown.

synteny. Conserved localization and the order of genes on chromosomes of different taxa.

ultraconserved elements (UCEs). Extremely conserved genomic elements that are used to design probes for sequence capture methods. Regions flanking the conserved sequences contain variation that can be used for a number of different analyses.

variant calling. Identifying polymorphisms between samples.

ACKNOWLEDGMENTS

LITERATURE CITED

Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin (2014). Genomic divergence in a ring species complex. Nature 511: 83–85.

Axelsson, E., M. T. Webster, N. G. C. Smith, D. W. Burt, and H. Ellegren (2005). Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Research 15:120–125.

Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M. T. Webster, T. Öst, M. Schneider, B. Kempenaers, and H. Ellegren (2010). The recombination landscape of the Zebra Finch *Taeniopygia guttata* genome. Genome Research 20:485–495.

Balakrishnan, C. N., M. Mukai, R. A. Gonser, J. C. Wingfield, S. E. London, E. M. Tuttle, and D. F. Clayton (2014). Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. PeerJ 2:e396.

Baldassarre, D. T., H. A. Thomassen, J. Karubian, and M. S. Webster (2013). The role of ecological variation in driving divergence of sexual and non-sexual traits in the Red-backed Fairy-wren (*Malurus melanocephalus*). BMC Evolutionary Biology 13:75.

Baldassarre, D. T., and M. S. Webster (2013). Experimental evidence that extra-pair mating drives asymmetrical introgression of a sexual trait. Proceedings of the Royal Society of London, Series B 280:20132175.

Baldassarre, D. T., T. A. White, J. Karubian, and M. S. Webster (2014). Genomic and morphological analysis of a semiper-

meable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution 68:2644–2657.

Besnard, G., J. A. M. Bertrand, B. Delahaie, Y. X. C. Bourgeois, E. Lhuillier, and C. Thébaud (2015). Valuing museum specimens: High-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (*Goura*). Biological Journal of the Linnean Society 116. In press.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz (2013). Unlocking the vault: Next-generation museum population genomics. Molecular Ecology 22:6018–6032.

Bourgeois, Y. X. C., E. Lhuillier, T. Cézard, 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. Molecular Ecology Resources 13:899–907.

Bradnam, K. R., J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J. A. Chapman, G. Chapuis, R. Chikhi, H. Chitsaz, et al. (2013). Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. GigaScience 2:1–31.

Brelsford, A., H. Collin, N. Perrin, and L. Fumagalli (2012). Nonspecific PCR amplification by high-fidelity polymerases: Implications for next-generation sequencing of AFLP markers. Molecular Ecology Resources 12:123–127.

Buerkle, C. A., and Z. Gompert (2013). Population genomics based on low coverage sequencing: How low should we go? Molecular Ecology 22:3028–3035.

Callicrate, T., R. Dikow, J. W. Thomas, J. C. Mullikin, E. D. Jarvis, R. C. Fleischer, and NISC Comparative Sequencing Program (2014). Genomic resources for the endangered Hawaiian honeycreepers. BMC Genomics 15:1098.

Campagna, L., I. Gronau, L. F. Silveira, A. Siepel, and I. J. Lovette (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular Ecology 24:4238–4251.

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: Genes, Genomes, Genetics 1:171–182.

Cheviron, Z. A., M. D. Carling, and R. T. Brumfield (2011). Effects of postmortem interval and preservation method on RNA isolated from field-preserved avian tissues. The Condor 113: 483–489.

Christidis, L. (1990). Aves. In Animal Cytogenetics (B. John, Editor). Gebrüder Bornträger, Berlin, Germany. p. 116.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biology 2:e109.

Coyne, J. A., and H. A. Orr (2004). Speciation. Sinauer Associates, Sunderland, Massachusetts.

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

DaCosta, J. M., and M. D. Sorenson (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE 9:e106713.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter (2011). Genome-wide genetic

marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499–510.

del Hoyo, J., A. Elliott, J. Sargatal, and D. A. Christie (Editors) (2011). Handbook of the Birds of the World. Lynx Edicions, Barcelona, Spain.

Delmore, K. E., J. W. Fox, and D. E. Irwin (2012). Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas, revealed using light-level geolocators. Proceedings of the Royal Society of London, Series B 279: 4582–4589.

Delmore, K. E., S. Hübner, N. C. Kane, R. Schuster, R. L. Andrew, F. Câmara, R. Guigó, and D. E. Irwin (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Molecular Ecology 24:1873–1888.

Delmore, K. E., and D. E. Irwin (2014). Hybrid songbirds employ intermediate routes in a migratory divide. Ecology Letters 17: 1211–1218.

Eaton, D. A. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849.

Ellegren, H. (2010). Evolutionary stasis: The stable chromosomes of birds. Trends in Ecology & Evolution 25:283–291.

Ellegren, H. (2011). Sex-chromosome evolution: Recent progress and the influence of male and female heterogamety. Nature Reviews Genetics 12:157–166.

Ellegren, H. (2013). The evolutionary genomics of birds. Annual Review of Ecology, Evolution, and Systematics 44:239–259.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61:717–726.

Faria, R., and A. Navarro (2010). Chromosomal speciation revisited: Rearranging theory with pieces of evidence. Trends in Ecology & Evolution 25:660–669.

Feder, A. F., D. A. Petrov, and A. O. Bergland (2012). LDx: Estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS ONE 7:e48588.

Fishman, L., A. Stathos, P. M. Beardsley, C. F. Williams, and J. P. Hill (2013). Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). Evolution 67:2547–2560.

Flaxman, S. M., J. L. Feder, and P. Nosil (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. Evolution 67:2577–2591.

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

Frankl-Vilches, C., H. Kuhl, M. Werber, S. Klages, M. Kerick, A. Bakker, E. H. C. de Oliveira, C. Reusch, F. Capuano, J. Vowinckel, S. Leitner, et al. (2015). Using the canary genome

to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. Genome Biology 16:19.

Ganapathy, G., J. T. Howard, J. M. Ward, J. Li, B. Li, Y. Li, Y. Xiong, Y. Zhang, S. Zhou, D. C. Schwartz, M. Schatz, et al. (2014). High-coverage sequencing and annotated assemblies of the Budgerigar genome. GigaScience 3:1–9.

Gansauge, M. T., and M. Meyer (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8:737–748.

Gohli, J., E. H. Leder, E. Garcia-del-Rey, L. E. Johannessen, A. Johnsen, T. Laskemoen, M. Popp, and J. T. Lifjeld (2015). The evolutionary history of Afrocanarian Blue Tits inferred from genomewide SNPs. Molecular Ecology 24:180–191.

Gompert, Z., and C. A. Buerkle (2011a). A hierarchical Bayesian model for next-generation population genomics. Genetics 187:903–917.

Gompert, Z., and C. A. Buerkle (2011b). Bayesian estimation of genomic clines. Molecular Ecology 20:2111–2127.

Gregory, T. R., J. A. Nicol, H. Tamm, B. Kullman, K. Kullman, I. J. Leitch, B. G. Murray, D. F. Kapraun, J. Greilhuber, and M. D. Bennett (2007). Eukaryotic genome size databases. Nucleic Acids Research 35:D332–D338.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel (2011). Bayesian inference of ancient human demography from individual genome sequences. Nature Reviews Genetics 43:1031–1034.

Harrison, P. W., A. E. Wright, F. Zimmer, R. Dean, S. H. Montgomery, M. A. Pointer, and J. E. Mank (2015). Sexual selection drives evolution and rapid turnover of male gene expression. Proceedings of the National Academy of Sciences USA 112:4393–4398.

Harrison, R. (Editor) (1993). Hybrid Zones and the Evolutionary Process. Oxford University Press, Oxford, UK.

Harvey, M. G., and R. T. Brumfield (2015). Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. Molecular Phylogenetics and Evolution 83:305–316.

Harvey, M. G., C. D. Judy, G. F. Seeholzer, J. M. Maley, G. R. Graves, and R. T. Brumfield (2015). Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3: e895.

Hoffmann, A. A., and L. H. Rieseberg (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? Annual Review of Ecology, Evolution, and Systematics 39:21–42.

Hooper, D. M., and T. D. Price (2015). Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. Evolution 69:890–903.

Ilut, D., M. L. Nydam, and M. P. Hare (2014). Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. BioMed Research International 675158.

International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716.

Irwin, D. E., S. Bensch, J. H. Irwin, and T. D. Price (2005). Speciation by distance in a ring species. Science 307:414–416.

Irwin, D. E., S. Bensch, and T. D. Price (2001a). Speciation in a ring. Nature 409:333–337.

Irwin, D. E., and J. H. Irwin (2005). Siberian migratory divides: The role of seasonal migration in speciation. In Birds of Two Worlds: The Ecology and Evolution of Migration (R. Greenberg and P. P. Marra, Editors). Johns Hopkins University Press, Baltimore, MD, USA. pp. 27–40.

Irwin, D. E., J. H. Irwin, and T. D. Price (2001b). Ring species as bridges between microevolution and speciation. Genetica 112–113:223–243.

Irwin, D. E., J. H. Irwin, and T. B. Smith (2011). Genetic variation and seasonal migratory connectivity in Wilson's Warblers (Wilsonia pusilla): Species-level differences in nuclear DNA between western and eastern populations. Molecular Ecology 20:3102–3115.

Itoh, Y., K. Kampf, C. N. Balakrishnan, and A. P. Arnold (2011). Karyotypic polymorphism of the Zebra Finch Z chromosome. Chromosoma 120:255–264.

Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, A. Suh, et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. Nature 346:1320–1331.

Jones, M. R., and J. M. Good (2015). Targeted capture in evolutionary and ecological genomics. Molecular Ecology 24. In press.

Jones, R. T., A. P. Salazar, R. H. ffrench-Constant, C. D. Jiggins, and M. Joron (2012). Evolution of a mimicry supergene from a multilocus architecture. Proceedings of the Royal Society of London, Series B 279:316–325.

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. Molecular Ecology 23:4035–4058.

Keping, S., K. A. Meiklejohn, B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball (2014). The evolution of peafowl and other taxa with ocelli (eyespots): A phylogenomic approach. Proceedings of the Royal Society of London, Series B 281: 20140823.

Kimura, M., S. M. Clegg, I. J. Lovette, K. R. Holder, D. J. Girman, B. Milá, P. Wade, and T. B. Smith (2002). Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic–Neotropical warbler (Wilsonia pusilla). Molecular Ecology 11:1605–1616.

Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard, G. Ganapathy, Z. Wang, D. A. Rasko, W. R. McCombie, E. D. Jarvis, and A. M. Phillippy (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature Biotechnology 30:693–700.

Kraus, R. H. S., H. H. D. Kerstens, P. Van Hooft, R. P. M. A. Crooijmans, J. J. Van Der Poel, J. Elmberg, A. Vignal, Y. Huang, N. Li, H. H. T. Prins, and M. A. M. Groenen (2011). Genome wide SNP discovery, analysis and evaluation in Mallard (Anas platyrhynchos). BMC Genomics 12:150.

Kraus, R. H. S., and M. Wink (2015). Avian genomics: Fledging into the wild! Journal of Ornithology 156. In press.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Larson, E. L., T. A. White, C. L. Ross, and R. G. Harrison (2014). Gene flow and the maintenance of species boundaries. Molecular Ecology 23:1668–1678.

Lee, J. Y., and S. V. Edwards (2008). Divergence across Australia's Carpentarian Barrier: Statistical phylogeography of the Red-backed Fairy-wren (Malurus melanocephalus). Evolution 62: 3117–3134.

Lerner, H. R. L., and R. C. Fleischer (2010). Prospects for the use of next-generation sequencing methods in ornithology. The Auk 127:4–15.

Lowry, D. B., and J. H. Willis (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biology 8:1–14.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure, and D. J. Turner (2010). Target-enrichment strategies for next-generation sequencing. Nature Methods 7:111–118.

Mason, N. A., and S. A. Taylor (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. Molecular Ecology 24:3009–3025.

Mayr, E. (1942). Systematics and the Origin of Species. Columbia University Press, New York, NY, USA.

McCormack, J. E., M. G. Harvey, B. C. Faircloth, N. G. Crawford, T. C. Glenn, and R. T. Brumfield (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8:e54848.

McCormack, J. E., and J. M. Maley (2015). Interpreting negative results with taxonomic and conservation implications: Another look at the distinctness of coastal California Gnatcatchers. The Auk 132:380–388.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, D. 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.

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

Noor, M. A. F., K. L. Grams, L. A. Bertucci, Y. Almendarez, J. Reiland, and K. R. Smith (2001). The genetics of reproductive isolation and the potential for gene exchange between Drosophila pseudoobscura and D. persimilis via backcross hybrid males. Evolution 55:512–521.

Oyler-McCance, S. J., R. S. Cornman, K. L. Jones, and J. A. Fike (2015). Genomic single-nucleotide polymorphisms confirm that Gunnison and Greater sage-grouse are genetically well differentiated and that the Bi-State population is distinct. The Condor 117:217–227.

Pala, I., S. Naurin, M. Stervander, D. Hasselquist, S. Bensch, and B. Hansson (2012). Evidence of a neo-sex chromosome in birds. Heredity 108:264–272.

Pallares, L. F., B. Harr, L. M. Turner, and D. Tautz (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. Molecular Ecology 23: 5756–5770.

Parchman, T. L., Z. Gompert, M. J. Braun, R. T. Brumfield, D. B. McDonald, J. A. C. Uy, G. Zhang, E. D. Jarvis, B. A. Schlinger,

and C. A. Buerkle (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology 22:3304–3317.

Paxton, K. L., M. Yau, F. R. Moore, and D. E. Irwin (2013). Differential migratory timing of western populations of Wilson's Warbler (*Cardellina pusilla*) revealed by mitochondrial DNA and stable isotopes. The Auk 130:689–698.

Peñalba, J. V., L. L. Smith, M. A. Tionone, C. Sass, S. M. Hykin, P. L. Skipwith, J. A. McGuire, R. C. Bowie, and C. Moritz (2014). Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. Molecular Ecology Resources 14: 1000–1010.

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.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344: 1410–1414.

Price, T. (2008). Speciation in Birds. Roberts, Boulder, CO, USA.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J. M. Burke, and L. H. Rieseberg (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nature Communications 4:1827.

Rheindt, F. E., and S. V. Edwards (2011). Genetic introgression: An integral but neglected component of speciation in birds. The Auk 128:620–632.

Rheindt, F. E., M. K. Fujita, P. R. Wilton, and S. V. Edwards (2014). Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): Population genetic and phylogenetic inferences from genome-wide SNPs. Systematic Biology 63: 134–152.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. Trends in Ecology & Evolution 16:351–358.

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

Ruegg, K. C., E. C. Anderson, K. L. Paxton, V. Apkenas, S. Lao, R. B. Siegel, D. F. DeSante, F. Moore, and T. B. Smith (2014b). Mapping migration in a songbird using high-resolution genetic markers. Molecular Ecology 23:5726–5739.

Ruegg, K. C., and T. B. Smith (2002). Not as the crow flies: A historical explanation for circuitous migration in Swainson's Thrush (*Catharus ustulatus*). Proceedings of the Royal Society of London, Series B 269:1375–1381.

Santure, A. W., J. Gratten, J. A. Mossman, B. C. Sheldon, and J. Slate (2011). Characterisation of the transcriptome of a wild Great Tit *Parus major* population by next generation sequencing. BMC Genomics 12:283.

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, et al. (2015). Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution 30:78–87.

Shapiro, M. D., Z. Kronenberg, C. Li, E. T. Domyan, H. Pan, M. Campbell, H. Tan, C. D. Huff, H. Haofu, A. I. Vickrey, S. C. A. Nielsen, et al. (2013). Genomic diversity and evolution of the head crest in the rock pigeon. Science 339:1063–1067.

Shendure, J., and H. Ji (2008). Next-generation DNA sequencing. Nature Biotechnology 26:1135–1145.

Shields, G. F. (1982). Comparative avian cytogenetics: A review. The Condor 84:45–58.

Singhal, S., E. Leffler, K. Sannareddy, I. Turner, O. Venn, D. Hooper, A. Strand, Q. Li, B. Raney, C. Balakrishnan, S. Griffith, et al. (2015). Stable recombination hotspots in birds. bioRxiv: 023101.

Skinner, B. M., and D. K. Griffin (2011). Intrachromosomal rearrangements in avian genome evolution: Evidence for regions prone to breakpoints. Heredity 108:37–41.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63:83–95.

Smith, J., C. K. Bruley, I. R. Paton, I. Dunn, C. T. Jones, D. Windsor, D. R. Morrice, A. S. Law, J. Masabanda, A. Sazanov, D. Waddingtion, et al. (2000). Differences in gene density on chicken macrochromosomes and microchromosomes. Animal Genetics 31:96–103.

Stager, M., D. L. Swanson, and Z. A. Cheviron (2015). Regulatory mechanisms of metabolic flexibility in the Dark-eyed Junco (*Junco hyemalis*). Journal of Experimental Biology 218. In press.

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

Stapley, J., T. R. Birkhead, T. Burke, and J. Slate (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the Zebra Finch genome. Genome research 20:496–502.

Storch, K.-F., O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong, and C. J. Weitz (2002). Extensive and divergent circadian gene expression in liver and heart. Nature 417:78–83.

Szymura, J. M., and N. H. Barton (1986). Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution 40:1141–1159.

Taylor, S. A., R. L. Curry, T. A. White, V. Ferretti, and I. Lovette (2014a). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. Evolution 68: 3066–3081.

Taylor, S. A., T. A. White, W. M. Hochachka, V. Ferretti, R. L. Curry, and I. Lovette (2014b). Climate mediated movement of an avian hybrid zone. Current Biology 24:671–676.

Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker (2010). The variable genomic architecture of isolation between hybridizing species of house mouse. Evolution 64:472–485.

van Bers, N. E., K. van Oers, H. H. Kerstens, B. W. Dibbits, R. P. Crooijmans, M. E. Visser, and M. A. Groenen (2010). Genome-wide SNP detection in the Great Tit *Parus major* using high throughput sequencing. Molecular Ecology 19:89–99.

Wang, Z., M. Gerstein, and M. Snyder (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57–63.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, et al. (2010). The genome of a songbird. Nature 464:757–762.

Weinman, L. R., J. W. Soloman, and D. R. Rubenstein (2014). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. Molecular Ecology Resources 15: 502–511.

Weir, J. T., M. S. Faccio, P. Pulido-Santacruz, A. O. Barrera-Guzmán, and A. Aleixo (2015). Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. Evolution 69:1823–1834.

Wright, A. E., P. W. Harrison, S. H. Montgomery, M. A. Pointer, and J. E. Mank (2014). Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68:3281–3295.

Wu, C. I. (2001). The genic view of the process of speciation. Journal of Evolutionary Biology 14:851–865.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proceedings of the National Academy of Sciences 110:E1743–E1751.

Yeaman, S., and M. C. Whitlock (2011). The genetic architecture of adaptation under migration–selection balance. Evolution 65:1897–1911.

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, et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zhang, J., C. Li, Q. Zhou, and G. Zhang (2015). Improving the ostrich genome assembly using optical mapping data. GigaScience 4:24.