Erschienen in: Journal of Ornithology ; 156 (2015), 4. - S. 851-865
https://dx.doi.org/10.1007/s10336-015-1253-y

Avian genomics: fledging into the wild!

Robert H. S. Kraus[1,2] · Michael Wink[3]

Abstract Next generation sequencing (NGS) technologies provide great resources to study bird evolution and avian functional genomics. They also allow for the identification of suitable high-resolution markers for detailed analyses of the phylogeography of a species or the connectivity of migrating birds between breeding and wintering populations. This review discusses the application of DNA markers for the study of systematics and phylogeny, but also population genetics and phylogeography. Emphasis in this review is on the new methodology of NGS and its use to study avian genomics. The recent publication of the first phylogenomic tree of birds based on genome data of 48 bird taxa from 34 orders is presented in more detail.

Keywords Next generation sequencing · Genetics · Ornithology · Whole genome · Phylogenomics · Single nucleotide polymorphisms · Microsatellites

Communicated by J. Fjeldså.

✉ Michael Wink
wink@uni hd.de; Wink@uni heidelberg.de

Robert H. S. Kraus
robert.kraus@uni konstanz.de

[1] Department of Biology, University of Konstanz, 78457 Constance, Germany

[2] Department of Migration and Immuno Ecology, Max Planck Institute for Ornithology, Am Obstberg 1, 78315 Radolfzell, Germany

[3] Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, INF 364, 69120 Heidelberg, Germany

Zusammenfassung

Die Ornithologie ist im Zeitalter der Genomik angekommen

Neue Sequenziertechnologien (*Next Generation Sequencing*; NGS) eröffnen die Möglichkeit, Evolution und funktionelle Genomik bei Vögeln umfassend zu untersuchen. Weiterhin erlaubt die NGS-Technologie, geeignete, hochauflösende Markersysteme für Mikrosatelliten und *Single Nucleotide Polymorphisms* (SNPs) zu identifizieren, um detaillierte Analysen zur Phylogeographie einer Art oder zur Konnektivität von Zugvögeln zwischen Brut- und Winterpopulationen durchzuführen. Dieses Review widmet sich der Anwendung von DNA Markern für die Erforschung von Systematik und Phylogenie sowie Populationsgenetik und Phylogeographie. Ein Schwerpunkt liegt dabei auf der Methodik des *Next Generation Sequencing* und dessen Anwendung in der Vogelgenomik. Der 2014 in Science veröffentlichte phylogenomische Stammbaum der Vögel, der auf genomweiten Daten von 48 Vogeltaxa aus 34 Ordnungen basiert, wird dabei detailliert besprochen.

Introduction

In natural sciences, progress and new insights are often triggered by new technologies, which allow a deeper and more detailed analysis of the material world. In biology, a breakthrough came through the possibility to sequence not only few marker genes, but also complete genomes. In 2003, the human genome was sequenced for the first time (International Human Genome Sequencing Consortium 2001); this work involved the cooperation of several

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-0-302778

laboratories and the use of thousands of capillary sequencers. In the last 10 years, DNA technology has seen the advent of new sequencing technologies [next generation sequencing (NGS)], which allow the parallel sequencing of millions of DNA molecules. With NGS, complete genomes can now be sequenced in a relatively short time. Furthermore, because the costs for NGS have come down steadily, genome analyses are now possible for research groups with a normal research budget.

The new sequencing technologies can be used to analyse complete genomes, transcriptomes (i.e., all the genes that are expressed in an organism, tissue, or cell) and epigenomes (modifications of DNA, such as methylation of cytosine, relevant to differentiation and development). The availability to sequence data of full genomes brings us closer to being able to reconstruct true phylogenomic trees of life; unlike previous analyses based on few marker genes (Table 1), they will be based on complete genomes or large parts of these genomes. RNA sequencing is used to understand the phenotype and function of an organism by revealing which genes are expressed in a particular tissue or stage of development, but it can also be used for phylogenomic analyses and to develop genomic markers (due to lack of space, RNA sequencing applications in ornithology will not be reviewed here). NGS produces extremely large data sets that require the help of knowledgeable bioinformatics and powerful computers for analysis.

In this review, we give a short summary on the pre-NGS use of DNA markers in ornithology. We then focus on genomic analyses by NGS used to identify new genetic markers such as microsatellites and single nucleotide polymorphisms (SNPs) in non-model species and to reconstruct a first avian phylogenomic tree.

Table 1 DNA analysis in ornithology

Method	Applications
Sequencing of marker genes	Phylogeny and molecular systematics
	Phylogeography
	DNA barcoding
Next generation sequencing	Genomics
	Phylogenomics and evolution
	Transcriptomics (functional genomics)
	Epigenomics
	Development of STR and SNP markers
Microsatellite analysis	Paternity analysis; forensics
	Population genetics
	Phylogeography
Single nucleotide polymorphisms (SNP)	Paternity analysis; forensics
	Population genetics
	Phylogeography
	Migratory connectivity

Pre-NGS DNA analysis in ornithology

In the late nineteenth and first half of the twentieth century, the discovery of DNA by Friedrich Miescher in 1869 (Dahm 2008), the resolution of its chemical (Levene 1919) and structural (Watson and Crick 1953) characteristics, and the recognition of its role in inheritance (Avery et al. 1944; Hershey and Chase 1952) sounded the bell for the development of molecular DNA markers. Since then, methods to study directly the DNA of an organism have found their way into all fields of modern biology. It was the deciphering of a few hundred base pairs of mitochondrial DNA (mtDNA), which first became possible due to the invention of the chain termination sequencing method (Sanger 1981; Sanger et al. 1977), and the direct amplification of marker genes by polymerase chain reaction (PCR; Saiki et al. 1985, 1988).

Before PCR and DNA sequencing, analyses of mtDNA, which is available in greater copy numbers than nuclear DNA in most cells, was carried out in birds. As early as 1984, pioneering work by, for example, John Avise and colleagues involved DNA restriction enzymes to study nucleotide substitution patterns in mtDNA, and ducks were among the first study objects (Kessler and Avise 1984). Not much later, this methodology of restriction fragment length polymorphism analysis (RFLP) came to broader application in ornithology (Kessler and Avise 1985). Mack et al. (1986) recognised the great potential of studying nucleotide variation in mtDNA in ornithology as early as 1986 with the first review on the topic in an ornithological journal. The first research papers in specific ornithology journals followed quickly in the late 1980s (Avise and Zink 1988; Quinn et al. 1989). But because RFLP has limited power of resolution, it was quickly abandoned when PCR and sequencing methods became available in the 1990s. DNA sequence could then be used for taxonomic species identification in curious cases, such as a shrike for which only a single individual was known (Smith et al. 1991), and by today, DNA sequence data is available for some 80 % of all avian species, and near-complete for some groups (e.g., Furnariidae, Emberizoidea). DNA fingerprint analyses, developed by A. Jeffreys in the 1980s (Jeffreys et al. 1985a, b) and based on the fragmentation of nuclear DNA by restriction enzymes and the detection of repetitive DNA elements (“minisatellites”) by multi and single locus probes, were instrumental to understand parentage systems in many organisms, including birds (Burke and Bruford

1987; Hill 1987). This technology was also replaced by PCR approaches, which allowed the amplification of specific alleles of microsatellite loci in a more precise and quantitative way. Further advances in molecular technology enriched the toolbox for ornithologists with genetic markers such as SNPs (Schlötterer 2004; Wink 2006) (Table 1) identified in genome-wide analyses.

Ornithology in the age of genomics

So far, genomic studies of birds have not become as widespread as genetics did in the early 1990s, despite the availability of genomic technology for some years (Edwards 2007). The first genomic studies on birds were clearly motivated by agricultural interests. One of the first genomes to be sequenced among higher vertebrates was the chicken *Gallus gallus* genome over a decade ago (Hillier et al. 2004), followed a few years later by the turkey *Meleagris gallopavo* (Dalloul et al. 2010) and the domestic duck/Mallard *Anas platyrhynchos* (Huang et al. 2013). Equally likely to be among the first birds with a sequenced genome were of course those species that are maybe not model species for agriculture, but for many other main scientific disciplines; perhaps most prominently the zebra finch *Taeniopygia guttata* (Warren et al. 2010), which was believed to fuel avian genomics in the wild (Balakrishnan et al. 2010). In 2012, the genomes of pied and collared flycatchers *Ficedula hypoleuca*, *F. albicollis* were published and in 2013 the peregrine *Falco peregrinus* and saker falcon *F. cherrug* (Zhang et al. 2014b).

Very recently, a breakthrough was achieved in avian genetics on the biggest possible scale: avian genomics. A recent Science issue (vol. 346 no. 6215, 2014) not only covered a first phylogenomic tree, but also other articles, in which the genome data were used in a comparative way as to understand the evolution of birds and the genomes. Further topics included the evolution of sex chromosomes, for instance, or the development of complex traits, such as flight, loss of teeth, and vocal learning. In addition, more than 20 more articles were published concertedly in other journals (overview on http://avian.genomics.cn/en).

Bird genomes encompass between 0.9 and 1.3 billion bp and are thus 70 % smaller than mammalian genomes, which usually have more than three billion bp. The number of genes, however, is thought to be up to 30,000 in both mammals and birds. Bird genomes have apparently lost quite a number of repetitive DNA elements and underwent segmental deletions. Instead of 34 52 % of repetitive DNA (LINE, SINE, transposons) in the genome of mammals (Storch et al. 2013), bird genomes harbour only 4 10 % repetitive DNA (Zhang et al. 2014b). Compared to other vertebrates, birds have fewer introns and shorter intergeneric sequences (Zhang et al. 2014b). Bird genomes are highly conserved with regard to their sequences and their arrangement on chromosomes (synteny). Nucleotide substitution rates are lower in birds than in mammals (1.9×10^{-3} substitutions per nucleotide position and million years in birds vs. 2.7×10^{-3} substitutions in mammals). Within birds, Passeriformes (the largest and most divers of bird orders) show the highest substitution rates 3.3×10^{-3} substitutions per position and million years. High substitution rates were also found in birds that are able to learn songs such as parrots, humming birds and song birds (Zhang et al. 2014b).

Considering that whole genome data exist for bird references for over a decade, and that NGS technology has been mature for at least half of this time (Metzker 2010), it is interesting to observe that genomic technology has not yet become a regular tool in specific ornithological studies. While introducing genomics in more detail, we adopt the strategy to highlight how the transition from genetics to genomics has taken place in the wider scope of ecology and evolution of birds as examples of organismal biology, and in which way we expect genomics to reach into specific ornithology journals in the future. Some general genetic knowledge of the reader has to be assumed in order to keep this review realistic. We thereby often omit examples from the non-avian literature although we are very aware that methods are usually independent of the taxon studied. However, we introduce relevant reviews from general biology in all paragraphs to provide the reader with access to condensed current knowledge. Here we focus on ornithology specifically, providing starting points across the broad topic of genomics for the interested ornithologist.

Genetic marker discovery in the genomics era

A genomic study does not necessarily entail the analysis of genomes in their full complexity. In fact, short-read sequencing technologies have been used successfully in a wide range of ecological and evolutionary studies of organisms where no reference genome was available. We now present examples wherein which genomic technologies greatly facilitated genome-wide marker set development often without the need of a reference genome and wherein short-reads or other reduced representations of the genome were developed to function as genetic markers.

Microsatellites: old wine in new bottles

Sequence analyses often fail to provide phylogenetic or phylogeographic information because evolutionary events that had taken place within the last ~100,000 years are often not reflected in nucleotide substitutions of common

marker genes. One of the most important genetic markers with a higher resolution power were microsatellite markers to assess genetic relationships within populations and species (Schlötterer 2004) (Table 1). Microsatellites are stretches of repetitive DNA motifs, for example, in the form of …ACACACACACACACACACAC…, here $(AC)_{10}$, and are, therefore, also called short tandem repeat (STR) or simple sequence repeat (SSR) markers (Selkoe and Toonen 2006). Because of this repeat structure, DNA replication enzymes often introduce errors (by "polymerase slippage") and microsatellite loci, therefore, exhibit a high level of allelic diversity within species and populations (Goldstein and Schlötterer 1999). PCR primers flanking such repeat motifs in rather conserved regions of the genomic sequence lead then to amplicons that exhibit a size difference based on the number of repeats in the microsatellite locus. Such size differences can be measured and scored as co-dominant alleles (Guichoux et al. 2011). Traditional methods to detect these repeat structures in the genome of a study species used to involve the preparation of libraries enriched for a certain arsenal of repeat probes and cloned into bacterial vectors to detect a specific DNA sequence (Grunstein and Hogness 1975). These clones would then be Sanger-sequenced to identify the flanking region for primer design. This is a tedious procedure and involves many labour and time intensive steps.

With NGS technology, the situation has changed considerably. There is no longer a need to enrich genomic libraries with STR probes and sequence them individually after cumbersome bacterial cloning. With pyrosequencing technology (Ronaghi 2001) implemented in Roche's 454 sequencing technology (Margulies et al. 2005), for example, sequence reads from parallel sequenced whole genome shotgun libraries are long enough to be mined bioinformatically for adequate STR motifs and their flanking sequences at the same time (Gardner et al. 2011; Malausa et al. 2011; Schoebel et al. 2013). Among many applications, such technology can benefit bird conservation projects directly. For example, Hartmann et al. (2014) were able to extract sufficient amounts of genomic DNA for 454 sequencing from four individuals of the highly endangered pale-headed brushfinch *Atlapetes pallidiceps* in Ecuador. They obtained 39,033 sequences across the whole genome from which they were able to detect bioinformatically 869 sequences containing di-, tri-, or tetrameric microsatellite repeats. For 24 of these sequences, primers were designed and tested, leading to the discovery of nine completely new markers, which could be used in subsequent genetic diversity assessments. But also ecologically well-understood species in Europe have suffered a lack of genetic marker systems. In Montagu's harrier *Circus pygargus*, important questions regarding mating system structure and breeding behaviour remained unstudied until second generation 454 sequencing made it possible to address molecular ecological questions (Janowski et al. 2014). The sequencing run in this project yielded even more sequences than in the previous example: a total of 85,624 reads. Of these, 448 potential microsatellite loci could be extracted, and subsequent testing of 42 loci revealed a polymorphic and reliable set of 19 markers. Using the 454 method, it was even possible to create a genetic marker set for extinct moa species (Aves: Dinornithiformes) of New Zealand (Allentoft et al. 2009).

Four hundred and fifty-four sequencing is becoming out-of-date due to the ever increasing throughput of competitor systems such as Illumina sequencing (Bennett 2004; Bentley 2006; Fedurco et al. 2006), which is now established as the industry standard in short-read sequencing. Currently, Illumina sequencers (http://www.illumina.com/applications/sequencing/dna sequencing.html) produce 150 bp long paired-end reads (i.e., de facto up to 300 bp) on their high-throughput instruments (HiSeq 2500) and 300 bp paired-end reads on their medium-throughput solution (MiSeq; de facto 600 bp with paired-end sequencing). This leaves sufficient space to design primers in regions that flank the detected repeat motif. Examples in which Illumina technology has succeeded in particularly complicated cases in birds are Gunnison sage-grouse *Centrocercus minimus* and Clark's nutcracker *Nucifraga columbiana*, which have been shown to have considerably fewer STRs than a snake species used in the same study (Castoe et al. 2012). The authors also compared the paired-end sequencing protocols from these relatively early times (in terms of Illumina technology evolution) to 454 technology at that time. Illumina's method out competed 454 in every respect even when correcting for the 100× higher output of Illumina in the study. Without this correction, Illumina clearly is orders of magnitude cheaper, too.

However, developments do not stand still, and new technologies such as single molecule real-time (SMRT) sequencing on the Pacific Biosciences RS (Eid et al. 2009) are now coming into application. They are termed third generation sequencing because their fundamentally different technology yield data of very long sequence reads (Metzker 2010). Grohme et al. (2013) tested this long-read technology for microsatellite discovery in the greater white-fronted goose *Anser albifrons*. SMRT produces a relatively small number of several kb long sequences, in this study, 16,180 reads with a total of 43 Mb of sequence data. The raw data of SMRT sequencing suffer from low base call quality, around 85 %, but there are protocols to build circular consensus sequences (CSS) leading to accuracies of >99 % at the cost of coverage (Travers et al. 2010). The very long reads and high accuracy after building consensus sequences promise to yield substantial amounts of microsatellite markers with high quality

flanking sequence information allowing for primer design. The 16,180 longs-reads could be arranged into 281 SMRT-CCS reads, and 316 microsatellites were identified of which 251 flanking PCR primer pairs could be designed. Eventually, a comparison of costs showed that microsatellite development based on SMRT was cheaper than methods based on 454, but more expensive than those based on Illumina (Grohme et al. 2013). However, this proof-of-concept paper illustrates that technology is progressing constantly.

SNPs: high frequency, genome-wide marker sets

The pragmatic definition of genomics, namely that it entails studies in which a large number of genetic markers are used in ways as to represent the majority of a study species' genome (see "Introduction"; Black et al. 2001; Luikart et al. 2003), signifies how genomic techniques have been used for some years. Such genome-wide marker sets can be extensive in model species such as the chicken (Groenen et al. 2000; Wong et al. 2004). In non-model species, it has usually not been possible to conduct studies in which microsatellite markers were used in sufficiently high densities across chromosomes as to refer to them as genomic studies. However, for some time, a new type of molecular marker has found its way into ecology and evolution studies: the SNP system.

Basically, the genomes of all organisms are spiked with SNPs in densities of sometimes more than 1/100 bp in outbreed species. The definition of an SNP includes the criterion to be segregating at a minimum of 1 % of the minor allele in population or species (Brookes 1999) to discriminate them from rare mutations, which technically are SNPs as well, but might have very specific phenotypes and, therefore, specific applications for example, SNPs that are the causative mutations of genetic diseases (Burton et al. 2007). In the genetics of model systems, SNPs have become the common-place marker (Morin et al. 2004; Vignal et al. 2002), and also the ever increasing supply of genotyping methods spurs their application in small, ecology, and evolution oriented laboratories. This includes the genotyping of difficult material like bones (Morin and McCarthy 2007) and fish scales (Smith et al. 2011), or faeces and urine traces in snow (Kraus et al. 2015).

The initial problem of obtaining SNPs as genetic markers lies in the fact that one cannot identify an SNP by its motif, as can be done in the case of a microsatellite. Therefore, a panel of multiple individuals needs to be sequenced for the same genomic regions, their sequences aligned, and polymorphic positions extracted and validated. Traditional methods involve sequencing of many independent genome fragments (Aitken et al. 2004; Seddon et al. 2005). Studies following similar strategies were indeed carried out in birds (Backström et al. 2008; Cramer et al. 2008; Kerr et al. 2014), but this approach is impractical to yield a sufficient number of SNPs to compete with the statistical power of multi-allelic microsatellites: two to ten SNPs are usually needed to reach the power of one microsatellite, depending on the research question (Gärke et al. 2012; Morin et al. 2004; Schlötterer 2004; Schopen et al. 2008).

However, the highly parallel nature of NGS has changed the field considerably. When NGS was still relatively expensive for sequencing whole genomes, methods for complexity reduction were developed for SNP discovery. So-called reduced representation libraries (RRLs) were used to select a fraction of the genome, say 5 %, to make it possible to sequence this portion for many individuals (van Tassell et al. 2008). The generated sequence reads were aligned to a reference genome and SNP information extracted with a dedicated bioinformatics pipeline (e.g., Kraus et al. 2011). In the latter case study, 5 % of the mallard genome was extracted from genomic DNA using digestion with two DNA restriction enzymes: *Alu*I (yielding 4 %) and *Hha*I (yielding 1 %). Both libraries were created from a pool of genomic DNA that stemmed from multiple individuals across a large geographical range to avoid ascertainment bias (Boursot and Belkhir 2006; Bradbury et al. 2011; Rosenblum and Novembre 2007). The raw sequencing data from Illumina technology had a size of 35 million 76 bp reads, which were filtered aggressively for several quality indicators, including per base quality score and excessive read over-representation pointing potentially to multi-copy genes or other repetitive regions. This led to a quality trimmed data set of 16.6 million reads of 62 bp length. After finding potential SNP positions and applying several further quality filters, 122,413 SNPs were identified with great confidence. The strict quality filtering proved to be useful as it was found in a validation step of genotyping 364 of those SNPs that the SNP-to-assay conversion rate was 99.7 %, a high value among comparable studies (Davey et al. 2011). Further, work-arounds were developed for species where no reference genome was available, and that included a two-step procedure of assembly of RRLs of longer fragments and subsequent SNP information extraction from de novo assembled long RRLs (Kerstens et al. 2009; van Bers et al. 2010). Alternatively, for SNP discovery in barnacle geese *Branta leucopsis*, it has been shown to be highly efficient to use the reference genome of a close relative (here, the mallard) for the extraction of chimeric flanking sequence and subsequent marker development (Jonker et al. 2012). In birds, this is a particularly promising strategy because of the highly conserved synteny of their genomes (Jarvis et al. 2014).

SNPs can as well be mined by cross-amplification of large collections of SNPs developed for other species

(Ogden et al. 2012), sometimes as distant as 44 million years (Hoffman et al. 2013). This is because there seems to be always some general background retention of shared SNPs of around 1 5 % between species (Kraus et al. 2012). However, one needs to bear in mind that these SNPs are mostly conserved for a reason and are likely enriched for functional SNPs or are biased in other ways that are not immediately evident. These unknown sources of bias are particularly hard to account for when performing certain types of analyses that require markers to behave as neutral as possible. For birds, such large-scale SNP chips are available, of course, for the agriculturally important chicken (e.g., Groenen et al. 2011). But also models in ecology and evolution have received increased attention lately, with 10 k SNP chips for the great tit *Parus major* (van Bers et al. 2012) and the house sparrow *Passer domesticus* (Hagen et al. 2013), and a 50 k SNP chip for *Ficedula* flycatchers (Kawakami et al. 2014). Cross-amplification studies at least within the Passeriformes should therefore be straight-forward with these new resources.

As costs for NGS are coming down steadily, also high coverage shotgun sequence analyses covering complete genomes become feasible and will certainly be used in future to develop new STR and SNP markers.

Direct use of reduced genome representations

RRLs cannot only be used for marker discovery. When each individual in a sequencing run is marked individually by a DNA barcode, its sequences can be extracted from the output of large sequencing runs and directly used in subsequent population genomic analyses. This idea requires that the preparation of an RRL from each individual is carried out with greatest care as to maximise the overlap of all potential fragments recovered by the size selection of the RRL construction. Greminger et al. (2014) advanced the RRL approach by meticulously standardising restriction enzyme digestion, gel extraction, and DNA purification for a population genomic study of orangutans *Pongo* sp., calling their strategy the iRRL protocol.

Not all RRL strategies involve size selection and gel purification, however. In fact, it has not taken long for more streamlined protocols to become available that eliminate the variation in fragment size libraries when cut from gel manually. A technology called "restriction site associated DNA (RAD) sequencing" (Miller et al. 2007) was introduced and refined for ecology and evolution studies by Baird et al. (2008). For RAD sequencing, short sequence reads are generated by NGS at thousands of regions adjacent to restriction endonuclease recognition sites across the genomes of multiple individuals at once. First used for SNP discovery (Hohenlohe et al. 2011; Scaglione et al. 2012; Senn et al. 2013), it has also been a method for direct genotyping of samples (Hohenlohe et al. 2010). For example, RAD sequencing was applied to solve questions regarding the divergence between two subspecies in Swainson's thrush *Catharus ustulatus*. Twenty-five blood and tissue samples from five populations across the breeding range in North America were collected and studied for divergence on a chromosomal level, whereby 360,000 SNPs (after quality filtering) could be analysed, or 154,000 SNPs when only SNPs with a known physical location on the zebra finch genome were needed for analysis (Ruegg et al. 2014). It was found that especially genes associated with migratory traits were involved in subspecific differentiation and that the Z chromosome showed particularly accelerated divergence. Another example where SNPs from RAD sequencing were powerful markers for the study of population and species divergence is a study of the Neotropical flycatcher genus *Zimmerius*. Rheindt et al. (2014) used vocal recordings, biometrics, and spectrophotometric measurements in conjunction with genomics. RAD sequencing genotyped 37,361 SNPs from two individuals of *Z. viridiflavus*, three *Z. chrysops*, five birds from mosaic populations where species boundaries are transient, and two outgroup individuals (*Z. acer*, *Z. gracilipes*). Of those, 2710 SNPs were found across all these lineages and could be used for tests of incomplete lineage sorting versus true introgression. Allelic variation due to introgression pointed at a candidate set of genes that need to be studied in further projects.

A direct genotyping approach closely related to RAD, called "genotyping by sequencing" (GBS; Elshire et al. 2011), involves an even easier work flow and promises cheaper applications. Here, a very recent example in birds is again a study of speciation and divergence of a species complex. The Neotropical ovenbird *Xenops minutus* currently has 11 recognised subspecies, of which a large number of 72 individuals from seven subspecies could be subjected to the RAD sequencing derivative technology called GBS (Harvey and Brumfield 2015). The number of SNPs that could be extracted for analyses (3379) was lower than that for the RAD examples above, which is both due to a higher number of individuals in the latter study and to the inherent properties of GBS to yield less data, but a much more efficient work flow and thus reduced project complexity and cost. With the less complex work flow of GBS, it was possible to include museum samples (62-year-old skins) in a study to resolve the evolutionary history of Afro-Canarian blue tits *Cyanistes teneriffae*, with samples from Eurasian blue tit *Cyanistes caeruleus* and the azure tit *Cyanistes cyanus* as phylogenetic outgroups (Gohli et al. 2015). From a total set of 17,000 SNPs, subsets needed to be created to assess population structure (7500) or phylogenetics (6400, of which 2000 loci were randomly picked because of limitations of the computer programme in use).

Finally, the historical museum skin samples yielded 166 SNPs to complement analyses with individuals from the current time.

Details of RAD sequencing methods (sensu lato) are covered by many methodology-related reviews, to which we refer the interested reader at this point (e.g., Baxter et al. 2011; Davey et al. 2011, 2013; Narum et al. 2013). In general, RAD is implemented nowadays in many service laboratories and is, therefore, available also to groups that do not have the equipment and experience to carry out this method. GBS is even cheaper, and more samples can be pooled for the same amount of money, but at the cost of the number of genotypes (see GBS example above). A methodological drawback of RAD and GBS are increasing levels of missing data when more individuals are added to a project. This means that by adding more individuals, often already when studying more than 50 specimens, the amount of scorable SNPs drops quickly. It is further not cost-effective to add a few individuals (say, 10) to an existing data set because a whole new run for only these few samples would need to be arranged at the same cost, normally, as for the larger project. Last, but not least, a number of biases exist, which will need to be taken into account when designing and analysing RAD or GBS projects (Davey et al. 2013).

A first phylogenomic avian tree of life

Like the advent of the first generation DNA sequencing using the Sanger method, the advent of second (or next) generation sequencing technologies has been a big bang for the natural and life sciences as a whole. The new sequencing technologies are so-called second generation because they rely on fundamentally different principles. Whereas the chain-termination sequencing method of Sanger produces relatively long stretches of DNA sequence (around 1000 bp, or 1 kbp) with error rates usually as low as 1/10,000 per nucleotide, the new technologies produce millions of so-called short-reads (i.e., shorter stretches of DNA sequence), but in a highly parallel manner. This, however, comes at the cost of sequence quality, which is in turn ameliorated by a high redundancy in sequencing the same DNA fragments multiple times, whereby errors can be amended by so-called sequencing depth. These technologies were reviewed in depth a few years ago in general (Metzker 2010) and with a focus on ornithology (Lerner and Fleischer 2010), and we thus forgo to delve into details here.

In the article "Whole-genome analyses resolve early branches in the tree of life of modern birds" (Jarvis et al. 2014), a phylogenomic avian tree of life was presented (Fig. 1) that was based on partial genomes of 48 bird taxa from 34 orders (including 30 orders of Neognathae). Some of the genome data used by Jarvis et al. (2014) had derived from earlier genome analyses. The dataset of 42 million nucleotides comprised sequences of 8351 exons of protein-coding genes, of 2516 introns, and of 3769 ultraconserved elements (UCEs). Jarvis et al. demonstrated that UCE and intron sequences were more informative than those of exons. This tree (Fig. 1) confirmed quite a number of phylogenetic relationships that had been discovered by Ericson et al. (2006) and Hackett et al. (2008) based on the analysis of sequences of 19 nuclear genes (see Wink 2011, 2013, 2015). However, this information was hardly mentioned in many of the press articles, which featured the results presented in Science. In addition, Burleigh et al. (2015) have reconstructed a more comprehensive avian phylogeny from 6714 species based on a sparse supermatrix comprising 22 nuclear loci and seven mitochondrial regions. We are aware that the tree shown in Fig. 1 reflects one of several other possible hypotheses (for discussion see Jarvis et al. 2014).

The new analysis confirmed the basal position of Palaeognathae (with ostrich and tinamus) and the monophyly of Neognathae as a sister group. Within Neognathae, the Galloanserae (comprising all pheasants, grouse, quails, partridges, ducks, geese, and swans) were corroborated as a sister to the rest of the Neoaves. Differences were found, however, in the structure of Neoaves. The division of Neoaves into "Metaves" and "Coronaves" (after Hackett et al. 2008, and other authors) could not be confirmed (Fig. 1). In Neognathae aquatic and terrestrial bird taxa evolved independently several times, suggesting that this trait was strongly influenced by convergent evolution.

Jarvis et al. (2014) postulate a monophyletic clade "Columbea", clustering at the base of Neoaves, which comprises both land and water birds. This monophylum includes pigeons (Columbiformes), sandgrouse (Pterocliformes), and mesites from Madagascar (Mesitornithoformes) as land birds, and flamingos (Phoenicopteriformes) and grebes (Podicipediformes) as water birds. This monophyletic cluster of water birds was described by Hackett et al. (2008). The tropical birds (Phaethontiformes), which were then part of the flamingo/grebe cluster, are now together with Sunbittern and Kagu (Eurypygiformes) at the base of a larger water bird monophylum termed "Aequornithia". All the other bird orders belong to "Passerea" according to Jarvis et al. (2014).

The known clade of nightjars (Caprimulgiformes), swifts, and humming birds (considered as Apodiformes) was also recovered by Jarvis et al. (2014), but is now summarised as Caprimulgiformes. Cuckoos (Cuculiformes), turacos (Musophagiformes), and bustards (Otidiformes) cluster as a sister to Caprimulgiformes (similar to the position in Hackett et al. 2008). The phylogenetic position of the South American hoatzin (*Opisthocomus*), which resembles cuckoos to some degree, had been

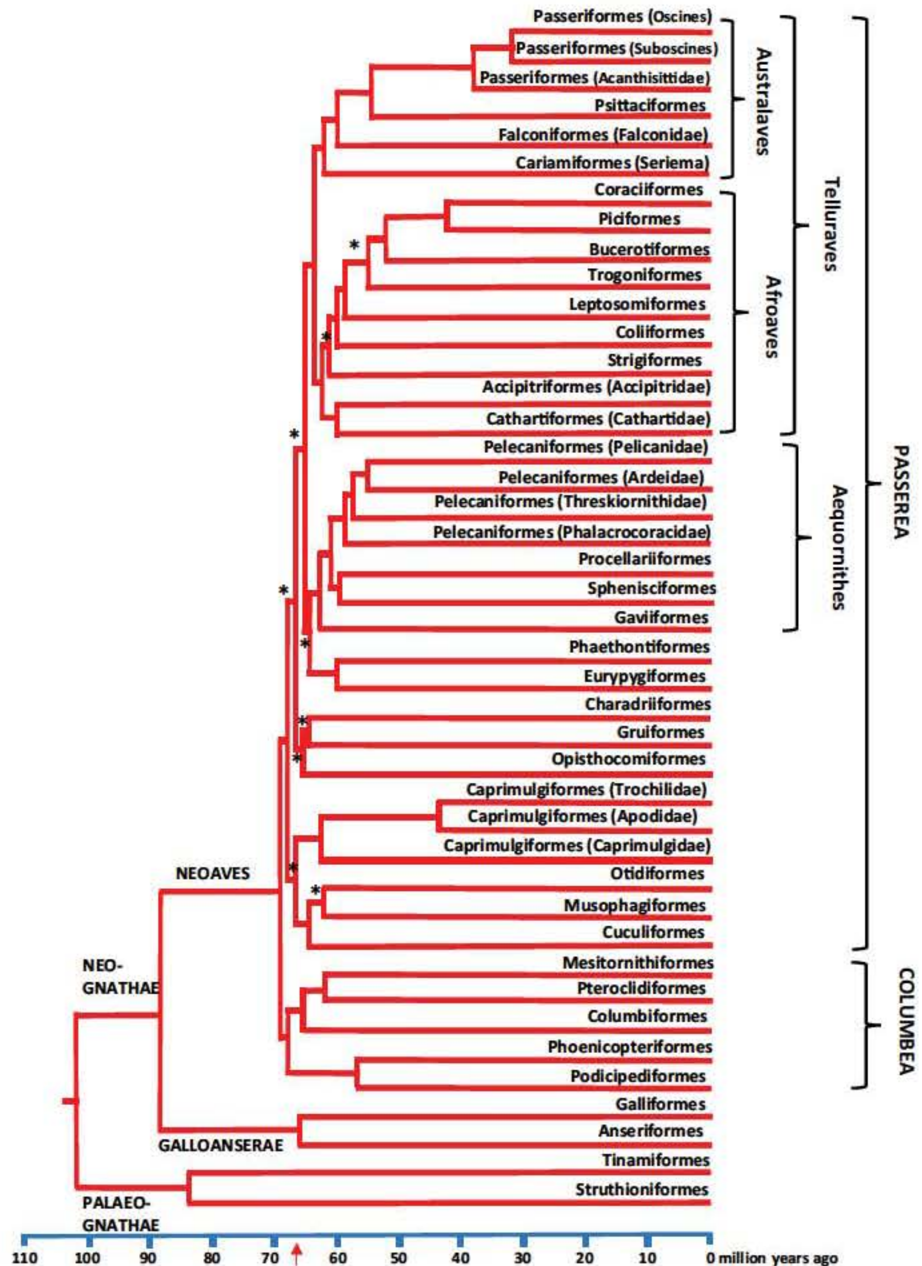

Fig. 1 Phylogenomic avian tree of life (after Jarvis et al. 2014). The phylogeny represents a "total evidence tree" reconstructed from 42 million nucleotides (from exons, introns, and UCEs) of almost every bird order. Branch lengths correspond to the evolutionary age of the taxa. Nodes without a bootstrap support of 100 % are marked with an *asterisk*. The *arrow* on the time axis indicates the border of the Cretaceous/Tertiary about 66 million years ago

enigmatic so far. The new results indicate that the Hoatzin is not a cuckoo, but clusters together with cranes (Gruiformes) and waders (Charadriiformes).

The large clade of water birds ("Aequornithia") is largely in agreement with the phylogenetic hypotheses of Hackett et al. (2008). It comprises loons (Gaviiformes), a sister clade of Sphenisciformes/Procellariiformes, and the other water birds, now placed together as Pelecaniformes. Unfortunately, Jarvis et al. (2014) have not sequenced a member of storks, which leaves the position of the Ciconiiformes unresolved. After Hackett et al. (2008), the Ciconiiformes (with a single family Ciconiidae) cluster basal to Pelecaniformes.

The other bird orders were combined as "core land birds; Telluraves", which are subdivided into two clades, Afroaves and Australaves (Fig. 1) (Ericson 2012). In the Afroaves, birds of prey (Accipitriformes) are basal, which agrees with the tree of Hackett et al. (2008), confirming that New World vultures (Cathartidae) do not belong to the Ciconiiformes as was postulated by Sibley and Monroe (1990). The finding of Hackett et al. (2008) that falcons (Falconidae) are not part of the birds of prey clade (Accipitriformes), but belong to a Falconidae/Psittaciformes/Passeriformes, was corroborated by the genome data (Fig. 1). Therefore, the order Falconiformes now is confirmed to consist of falcons only and no longer of other

raptors (Wink 2011, 2013). The relationship of owls (Strigiformes) with mousebirds (Coliiformes) was confirmed as well. The other Afroaves consist (as in Hackett et al. 2008) of cuckoo roller (Leptosomiformes), trogons (Trogoniformes), hornbills (Bucerotiformes), woodpeckers (Piciformes), and kingfishers, bee-eaters, rollers, motmots, and toadies (Coraciiformes).

The clade comprising the Australaves is identical to the one in Hackett et al. (2008). Basal are seriemas (Cariamiformes), followed by Falconiformes, Psittaciformes, and Passeriformes. This finding was highlighted as a novelty in many press articles although the group of falcons, parrots, and songbirds had already been recognised as a monophylum by Hackett et al. (2008) and termed "Eufalconimorphae" by Suh et al. (2011), which might have originated in the austral continent of Australia Antarctica South America about 65 million years ago.

The phylogenomic data were analysed under a molecular clock model and fossils used as minimum age constraints. As can be seen in Fig. 1, it is assumed that the Palaeognathae and Neognathae shared a common ancestor before more than 100 million years ago. About 88 million years ago, the lineages leading to Galloanserae and Neoaves became separated. Within the Neoaves, a rapid radiation into a larger number of bird orders took place within a few million years at the border of the Cretaceous and Tertiary (around 67 69 million years ago) when planet Earth was apparently hit by a large meteorite leading to the extinction of dinosaurs (Alvarez et al. 1980; Vellekoop et al. 2014). This rapid radiation after the mass extinction can explain the difficulty to find statistically significant support for many of the nodes leading to the branches of extant bird orders. In the present tree, six of 46 nodes had bootstrap support of <100 % indicating incomplete lineage sorting. When the dinosaurs had disappeared, a number of ecological niches became vacant, which were soon inhabited by new groups of birds and mammals (Storch et al. 2013). As early as 50 million years ago, almost all bird orders had evolved. New bird orders evolving in the Eocene were the Apodiformes, Piciformes, and Coraciiformes. The Passeriformes evolved relatively late, around 39 million years ago in Australia, and today comprise about 60 % of all extant birds. Because we do not have fossil evidence for the time point of divergence of Passeriformes, this calibration should be treated with some caution as the molecular clock apparently runs at a different speed in this order.

The relationships discovered by Hackett et al. (2008) and Jarvis et al. (2014) differ on many points from the results of DNA DNA hybridisation analyses carried out by Sibley and co-workers about 25 years ago when they were used to reconstruct a new systematics of birds (Sibley and Monroe 1990). As we know today, DNA DNA hybridisation does not have sufficient phylogenetic resolution power as compared to sequence data. Therefore, it is not surprising that Sibley and Monroe (1990) arrived at apparently wrong systematic groupings in many instances.

Although the new avian tree of life looks quite robust, we can nevertheless assume that it is not correct in all its bifurcations. New genome sequences are in the sequencing pipeline, and the next step will be an analysis of >200 genomes covering most of the bird families (Erich Jarvis, Guojie Zhang, Tom Gilbert, personal communication). In a long-term perspective, an avian tree of life comprising all its 10,500 taxa (or more; del Hoyo and Collar 2014) will probably be available in the future. In addition, the present phylogenomic analysis was based only on those partial genome sequences that were available after quality filtering and alignment in today's computational sequence alignment frameworks. A challenge would be to produce complete genome data with sufficient quality throughout all parts of the genome for a comparative analysis. However, even with improved genome-wide bioinformatics technology it might not be possible to compare every single base in the genomes of a group of species because there will be regions that do not share a common evolutionary history. The more distant two species are in terms of their evolution, the more genomic regions are no longer comparable, and therefore, a "full phylogenomic tree of life" will always contain some gaps.

Using whole genomic data to study bird evolution

Whole genome sequencing of non-model bird genomes started only a few years ago, starting with the above mentioned zebra finch genome (Warren et al. 2010), arguably a model species as well, but the borders between model and non-model can be hard to define in ecology and evolution. Several additional projects have attempted to decipher bird genomes, involving a broad scope of biological questions. Ellegren et al. (2012) screened the genomes of *Ficedula* flycatchers for islands of speciation (Kelleher and Barbash 2010), and Zhan et al. (2013) sequenced the genomes of two falcon species to study natural adaptation to predatory lifestyle. The black grouse genome was investigated for regions with a particularly accelerated evolutionary speed, for instance, the major histocompatibility (MHC) locus (Wang et al. 2014a).

In the framework of the B10 K project (in the special issue of Science and also in other papers, which appeared on the same day), general avian genome evolution was scrutinised in Zhang et al. (2014b) in a comparative genomics manner and complemented by dedicated sex chromosome evolution studies (Wang et al. 2014c; Zhou et al. 2014). The evolution of sex chromosomes, which

have derived from normal autosomes, was analysed by Zhou et al. (2014). Different from mammals with an XX- and XY-system, female birds have a WZ- and male birds ZZ-sex chromosomes. Y- and W-chromosomes have lost a number of functional genes because they cannot recombine with their former homologues, the Z- or X-chromosomes. The DMRT1-gene, which governs male sex determinations in birds, has been lost on the W-chromosome.

It is common knowledge that birds have lost their teeth, which were present in their ancestors, the theropod dinosaurs, during evolution. Using the phylogenomic backbone and the information about all genes that are present in a genome assembly and annotation, Meredith et al. (2014) found that mineralised teeth were lost only a single time in the common avian ancestor. In all of the 48 bird taxa of the genome project, the genes that are responsible for the formation of dentin and enamel were inactivated by mutations, whereas these genes are still active in crocodiles. The authors assume that genes for tooth formation have started degenerating in an ancestor of the present birds for more than 116 million years ago. It has been postulated that the reduction of jaws and loss of teeth is an adaptation to flight (Wink 2013) because the horn structures that replace bones and teeth are much lighter.

Adding three crocodilian genomes to the analyses of avian genomes enabled substantial gain of knowledge of deeper archosaur evolution and to infer an ancestral genome of all birds (Green et al. 2014). Interestingly, it seems that the chicken lineage most closely resembles the dinosaur avian ancestor (Romanov et al. 2014). Looking deeper at specific genes of the avian genomes revealed several patterns of adaptation and further connections between ecology, biology, and genetics. The study of two Antarctic penguin species shed light on adaption to extreme climates (Li et al. 2014a), while the comparison of an immune gene complex, the MHC, between birds and a crocodilian species (Jaratlerdsiri et al. 2014) helped resolve some important open questions in avian immune gene evolution.

Birds are a special group of vertebrates in that they have conquered the sky in an explosive adaptive radiation. A study of keratin genes, which are important for the evolution of feathers early or just before the avian radiation, is an example of how to use whole genome information to gain knowledge about flight evolution in birds in general (Greenwold et al. 2014). Comparative gene family evolution across birds, and also mammals, has been shown to become highly informative by two exemplary studies: the adaptive gene complex of avian haemoglobin isoform expression (Opazo et al. 2014) and a general vertebrate-wide analysis of hedgehog genes, which are key switches during embryonal development (Pereira et al. 2014).

Special features of avian genomes also seem to be wired in lineage specific patterns of gene conversion (Weber et al. 2014a) or selection (Weber et al. 2014b) and other factors leading to a rather compact genome (Lovell et al. 2014; Zhang et al. 2014b). A large group of papers used the bird genomes to investigate a key feature of general genome evolution: transposable and endogenised retroviral elements (Chong et al. 2014; Cui et al. 2014; Suh et al. 2014a, b). Finally, one of the most eminent traits of birds, the ability of multiple lineages to sing and especially learn songs, was studied in a multifaceted way by comparative genomics (Whitney et al. 2014; Wirthlin et al. 2014), transcriptomics (Pfenning et al. 2014; Wang et al. 2014b), and candidate gene approaches (Wang et al. 2014b).

Conclusions

Soon after becoming available, genetic techniques were readily used to study ecology and evolution in bird species. Traditionally, ornithological topics are interesting to a relatively wide readership, and thus the use of new genetic technologies in ornithology was often published in journals of a taxonomically wide scope (e.g., Nature and Science; Britten 1986; Quinn et al. 1987). Since the early 1990s, the use of genetics in ornithology is part of most working groups (sometimes by collaboration). In contrast, the costs associated with genome-wide data collection have precluded broad applications of genomics studies for a while. If we take the publication of the chicken genome (Hillier et al. 2004) as a milestone in avian genomics, we have since seen merely reviews and opinion papers in specific bird journals (Balakrishnan et al. 2010; Edwards 2007; Piertney 2006). The reasons are twofold: the financial and analytical burden remains a challenge that is hard to overcome, and avian genomics projects are at the forefront of inter- and cross-disciplinary research and, therefore, are rarely published in specific ornithology journals, which we also observe in most examples presented in this review. It is indeed still prohibitively expensive to perform molecular ecological studies with genomic techniques. Although marker sets covering hundreds of markers can be considered genome-wide, their application has remained exotic (Jonker et al. 2013; Kraus et al. 2013; Kurvers et al. 2013; Santure et al. 2010) and are usually published in journals with a broader scope. The same accounts for the examples about RAD and GBS above. For whole genome analyses of individuals to arrive at a meaningful sample size is a challenge even for today's pricing. However, individual samples of genomic portions were used for phylogenomics (Baker and Haddrath 2006; Kriegs et al. 2006). It is not only the monetary cost that makes life difficult for ornithologists. Often, relatively heavy bioinformatics is involved in data

analysis, and this usually is beyond the regular training of an ecologist. It will take some time until bioinformatics becomes part of MSc level education in ecology and evolution programmes at universities.

Avian evolutionary patterns can, however, be studied not only by genomics, and therefore, an integrative approach to understanding bird origins was highlighted as well (Xu et al. 2014), in addition to stating clearly the value of classical and modern (i.e., including biobanks) natural history collections (Kress 2014). Neither do the avian genome project data (Zhang et al. 2014a) stand isolated with regard to other major genome sequencing projects. Not long ago, an ambitious consortium formed with the aim of sequencing 10,000 vertebrate species (Haussler et al. 2009) for building an indispensable foundation for future biology studies. In this respect, the bird genomes are part of an even bigger initiative (O'Brien et al. 2014), providing a glimpse of what can actually be achieved in the study of evolution and ecology of non-model organisms. Actually, the Avian Genomics papers from December 2014 offer a first example of how true genomic information fosters more applied fields of research, long waiting for large-scale genomic information: conservation genomics (Allendorf et al. 2010; Ouborg et al. 2010; Piertney 2006; Shafer et al. 2015). Li et al. (2014b) characterised important baseline data in the genome of the crested ibis regarding processes such as near-extinction and subsequent rescue.

Last, but not least, even with the "big bang" of avian genomics in the special issue of Science in late 2014, our understanding of comparative genomic landscapes of birds is insufficient to go into more fine-scale or population-wide analyses in many taxa. When will a larger set of reference genomes be available? This is hard to predict considering the over-paced developments in genomics (Hayden 2014), but Harr and Price (2012) estimated that the majority of all birds would be sequenced by 2022. We know of one more published bird genome, which came out after the Science special issue (Frankl-Vilches et al. 2015), and we expect that in 2 3 years at the latest, hundreds of bird genomes will be publically available through dedicated online resources such as "Avianbase" (Eöry et al. 2015). Avian genomics currently builds up with great leaps forward. The outcomes of all the many bird genomes will fuel general biological research for many years, and thus the ornithological community will find itself again at the forefront of research of general importance (Konishi et al. 1989). Avian genomics has grown considerably, as we could show in this review. Maybe not immediately, but almost guaranteed, this will lead to the publishing of more genomics articles in native ornithology journals. Until then, keep your eyes open for ground-breaking ornithology, which is published in many high ranking, cross-disciplinary journals.

References

Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted gene approach. Mol Ecol 13:1423 1431. doi:10.1111/j.1365 294X.2004.02159.x

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697 709. doi:10.1038/nrg2844

Allentoft ME et al (2009) Identification of microsatellites from an extinct moa species using high throughput (454) sequence data. Biotechniques 46:195 200. doi:10.2144/000113086

Alvarez LW, Alvarez W, Asaro F, Michel HV (1980) Extraterrestrial cause for the Cretaceous Tertiary extinction. Science 208:1095 1108

Avery O, MacLeod C, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococ cal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med 79:137 158

Avise JC, Zink RM (1988) Molecular genetic divergence between avian sibling species: king and clapper rails, long billed and short billed dowitchers, boat tailed and great tailed grackles, and tufted and black crested titmice. Auk 105:516 528

Backström N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene based set of reference markers evenly spread across the avian genome. Mol Ecol 17:964 980

Baird NA et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3. doi:10.1371/journal.pone.0003376

Baker A, Haddrath O (2006) Rare genomic events as phylogenetic markers to help resolve the avian tree of life. J Ornithol 147:43 44

Balakrishnan CN, Edwards SV, Clayton DF (2010) The zebra finch genome and avian genomics in the wild. Emu 110:233 241. doi:10.1071/MU09087

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML (2011) Linkage mapping and comparative genomics using next generation RAD sequencing of a non model organism. PLoS ONE 6:e19315

Bennett S (2004) Solexa Ltd. Pharmacogenomics 5:433 438

Bentley DR (2006) Whole genome re sequencing. Curr Opin Genet Dev 16:545 552

Black WC IV, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome wide sampling of insect populations. Annu Rev Entomol 46:441 469

Boursot P, Belkhir K (2006) Mouse SNPs for evolutionary biology: beware of ascertainment biases. Genome Res 16:1191 1192

Bradbury IR et al (2011) Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. Mol Ecol Res 11:218 225

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393 1398

Brookes AJ (1999) The essence of SNPs. Gene 234:177 186

Burke T, Bruford MW (1987) DNA fingerprinting in birds. Nature 326:149 152

Burleigh JG, Kimball RT, Braun EL (2015) Building the avian tree of life using a large scale, sparse supermatrix. Mol Phylogenet Evol 84:53 63

Burton PR et al (2007) Genome wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661 678. doi:10.1038/nature05911

Castoe TA et al (2012) Rapid microsatellite identification from Illumina paired end genomic sequencing in two birds and a snake. PLoS One 7. doi:10.1371/journal.pone.0030953

Chong AY, Kojima KK, Jurka J, Ray DA, Smit AFA, Isberg SR, Gongora J (2014) Evolution and gene capture in ancient endogenous retroviruses insights from the crocodilian gen omes. Retrovirology 11:71. doi:10.1186/s12977 014 0071 2
Cramer ERA, Stenzler L, Talaba AL, Makarewich CA, Vehrencamp SL, Lovette IJ (2008) Isolation and characterization of SNP variation at 90 anonymous loci in the banded wren (*Thryothorus pleurostictus*). Conserv Genet 9:1657 1660. doi:10.1007/s10592 008 9511 7
Cui J et al (2014) Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol 15:539
Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Hum Genet 122:565 581
Dalloul RA et al (2010) Multi platform next generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol 8. doi:10.1371/journal.pbio.1000475
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome wide genetic marker discovery and geno typing using next generation sequencing. Nat Rev Genet 12:499 510
Davey JW, Cezard T, Fuentes Utrilla P, Eland C, Gharbi K, Blaxter ML (2013) Special features of RAD sequencing data: implica tions for genotyping. Mol Ecol 22:3151 3164. doi:10.1111/mec.12084
del Hoyo J, Collar NJ (2014) Introduction. In: del Hoyo J, Collar NJ (eds) HBW and BirdLife international illustrated checklist of the birds of the world. Non passerines, vol 1. Lynx Edicions, Barcelona, pp 19 54
Edwards SV (2007) Genomics and ornithology. J Ornithol 148:S27 S33. doi:10.1007/s10336 007 0238 x
Eid J et al (2009) Real time DNA sequencing from single polymerase molecules. Science 323:133 138. doi:10.1126/science.1162986
Ellegren H et al (2012) The genomic landscape of species divergence in *Ficedula flycatchers*. Nature 491:756 760. doi:10.1038/nature11584
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping by sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. doi:10.1371/journal.pone.0019379
Eöry L et al (2015) Avianbase: a community resource for bird genomics. Genome Biol 16:21. doi:10.1186/s13059 015 0588 2
Ericson PG (2012) Evolution of terrestrial birds in three continents: biogeography and parallel radiations. J Biogeogr 39:813 824
Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M, Ohlson JI, Parsons TJ, Zuccon D, Mayr M (2006) Diversification of Neoaves: integration of molecular sequence data and fossils. Biol Lett 2:543 547
Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid phase amplified DNA colonies. Nucleic Acids Res 34:e22
Frankl Vilches C et al (2015) Using the canary genome to decipher the evolution of hormone sensitive gene regulation in seasonal singing birds. Genome Biol 16:19. doi:10.1186/s13059 014 0578 9
Greminger MP et al (2014) Generation of SNP datasets for orangutan population genomics using improved reduced representation sequencing and direct comparisons of SNP calling algorithms. BMC Genomics 15. doi:10.1186/1471 2164 15 16
Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines recommendations for ecologists when using next generation sequencing for microsatellite development. Mol Ecol Resour 11:1093 1101. doi:10.1111/j.1755 0998.2011.03037.x
Gärke C, Ytournel F, Bed'Hom B, Gut I, Lathrop M, Weigend S, Simianer H (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. Anim Genet 43:419 428. doi:10.1111/j.1365 2052.2011.02284.x
Gohli J et al (2015) The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. Mol Ecol 24:180 191. doi:10.1111/mec.13008
Goldstein DB, Schlötterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford
Green RE et al (2014) Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346:1254449. doi:10.1126/science.1254449
Greenwold MJ et al (2014) Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. BMC Evol Biol 14:249
Groenen MAM et al (2000) A consensus linkage map of the chicken genome. Genome Res 10:137 147
Groenen MAM et al (2011) The development and characterization of a 60 K SNP chip for chicken. BMC Genomics 12. doi:10.1186/1471 2164 12 274
Grohme MA, Frias Soler R, Wink M, Frohme M (2013) Microsatel lite marker discovery using single molecule real time circular consensus sequencing on the Pacific Biosciences RS. Biotech niques 55:253 256. doi:10.2144/000114104
Grunstein M, Hogness DS (1975) Colony hybridization a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72:3961 3965. doi:10.1073/pnas.72.10.3961
Guichoux E et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591 611. doi:10.1111/j.1755 0998.2011.03014.x
Hackett SJ et al (2008) A phylogenomic study of birds reveals their evolutionary history. Science 320:1763 1768
Hagen IJ, Billing AM, Rønning B, Pedersen SA, Pärn H, Slate J, Jensen H (2013) The easy road to genome wide medium density SNP screening in a non model species: development and application of a 10 K SNP chip for the house sparrow (*Passer domesticus*). Mol Ecol Resour 13:429 439. doi:10.1111/1755 0998.12088
Harr B, Price T (2012) Speciation: clash of the genomes. Curr Biol 22:R1044 R1046. doi:10.1016/j.cub.2012.11.005
Hartmann SA, Schaefer HM, Segelbacher G (2014) Development of 12 microsatellite loci for the endangered pale headed Brushfinch (*Atlapetes pallidiceps*) and their cross amplification in two co occurring brushfinches. J Ornithol 155:835 839
Harvey MG, Brumfield RT (2015) Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, popula tion expansion, and gene flow. Mol Phylogenet Evol 83:305 316. doi:10.1016/j.ympev.2014.10.023
Haussler D et al (2009) Genome 10 K: a proposal to obtain whole genome sequence for 10,000 vertebrate species. J Hered 100:659 674
Hayden EC (2014) The $1,000 genome. Nature 507:294 295. doi:10.1038/507294a
Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 36:39 56
Hill WG (1987) DNA fingerprints applied to animal and bird populations. Nature 327:98 99
Hillier LW et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695 777
Hoffman JI, Thorne MAS, McEwing R, Forcada J, Ogden R (2013) Cross amplification and validation of SNPs conserved over 44 million years between seals and dogs. PLoS ONE 8:e68365. doi:10.1371/journal.pone.0068365

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6:e1000862. doi:10.1371/journal.pgen.1000862

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Res 11:117 122

Huang Y et al (2013) The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet 45:776 783. doi:10.1038/ng.2657

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860 921

Jaratlerdsiri W et al (2014) Comparative genome analyses reveal distinct structure in the saltwater crocodile MHC. PLoS One 9. doi:10.1371/journal.pone.0114631

Janowski S, Grohme MA, Frohme M, Wink M (2014) Development of new microsatellite (STR) markers for Montagu's harrier (*Circus pygargus*) via 454 shot gun pyrosequencing. Open Ornithol J 7:11 18

Jarvis ED et al (2014) Whole genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320 1331. doi:10.1126/science.1253451

Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable 'minisatel lite' regions in human DNA. Nature 314:67 73

Jeffreys AJ, Wilson V, Thein SL (1985b) Individual specific 'finger prints' of human DNA. Nature 316:76 79

Jonker RM et al (2012) The development of a genome wide SNP set for the barnacle goose *Branta leucopsis*. PLoS ONE 7:e38412

Jonker RM et al (2013) Genetic consequences of breaking migratory traditions in barnacle geese *Branta leucopsis*. Mol Ecol 22:5835 5847

Kawakami T et al (2014) Estimation of linkage disequilibrium and interspecific gene flow in *Ficedula* flycatchers by a newly developed 50 k single nucleotide polymorphism array. Mol Ecol Resour 14:1248 1260. doi:10.1111/1755 0998.12270

Kelleher ES, Barbash DA (2010) Expanding islands of speciation. Nature 465:1019 1020

Kerr KCR, Cloutier A, Baker AJ (2014) One hundred new universal exonic markers for birds developed from a genomic pipeline. J Ornithol 155:561 569. doi:10.1007/s10336 014 1041 0

Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin A Woeng TFC, den Dunnen JT, Groenen MAM (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. BMC Genom 10:479

Kessler LG, Avise JC (1984) Systematic relationships among waterfowl (Anatidae) inferred from restriction endonuclease analysis of mitochondrial DNA. Syst Zool 33:370 380. doi:10.2307/2413089

Kessler LG, Avise JC (1985) A comparative description of mito chondrial DNA differentiation in selected avian and other vertebrate genera. Mol Biol Evol 2:109 125

Konishi M, Emlen ST, Ricklefs RE, Wingfield JC (1989) Contribu tions of bird studies to biology. Science 246:465 472

Kraus RHS et al (2011) Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). BMC Genom 12:150

Kraus RHS et al (2012) Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. BMC Evol Biol 12:45

Kraus RHS, Van Hooft P, Megens H J, Tsvey A, Fokin SY, Ydenberg RC, Prins HHT (2013) Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. Mol Ecol 22:41 55. doi:10.1111/mec.12098

Kraus RHS et al (2015) A single nucleotide polymorphism based approach for rapid and cost effective genetic wolf monitoring in Europe based on noninvasively collected samples. Mol Ecol Resour 15:295 305. doi:10.1111/1755 0998.12307

Kress WJ (2014) Valuing collections. Science 346:1310. doi:10.1126/science.aaa4115

Kriegs JO, Matzke A, Churakov G, Brosius J, Schmitz J (2006) Retroposons as phylogenetic markers in bird genomes. J Ornithol 147:197 198

Kurvers RHJM et al (2013) Contrasting context dependence of familiarity and kinship in animal social networks. Anim Behav 86:993 1001. doi:10.1016/j.anbehav.2013.09.001

Lerner HRL, Fleischer RC (2010) Prospects for the use of next generation sequencing methods in ornithology. Auk 127:4 15. doi:10.1525/auk.2010.127.1.4

Levene P (1919) The structure of yeast nucleic acid. J Biol Chem 40:415 424

Li C et al (2014a) Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. GigaScience 3:27

Li S et al (2014b) Genomic signatures of near extinction and rebirth of the crested ibis and other endangered bird species. Genome Biol 15:557

Lovell PV et al (2014) Conserved syntenic clusters of protein coding genes are missing in birds. Genome Biol 15:565

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981 994

Mack AL, Gill FB, Colburn R, Spolsky C (1986) Mitochondrial DNA: a source of genetic markers for studies of similar passerine bird species. Auk 103:676 681

Malausa T et al (2011) High throughput microsatellite isolation through 454 GS FLX titanium pyrosequencing of enriched DNA libraries. Mol Ecol Resour 11:638 644. doi:10.1111/j.1755 0998.2011.02992.x

Margulies M et al (2005) Genome sequencing in microfabricated high density picolitre reactors. Nature 437:376 380

Meredith RW, Zhang G, Gilbert MTP, Jarvis ED, Springer MS (2014) Evidence for a single loss of mineralized teeth in the common avian ancestor. Science 346:1254390. doi:10.1126/science.1254390

Metzker ML (2010) Sequencing technologies the next generation. Nat Rev Genet 11:31 46

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost effective polymorphism identification and geno typing using restriction site associated DNA (RAD) markers. Genome Res 17:240 248. doi:10.1101/gr.5681207

Morin PA, McCarthy M (2007) Highly accurate SNP genotyping from historical and low quality samples. Mol Ecol Notes 7:937 946

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208 216

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping by sequencing in ecological and conserva tion genomics. Mol Ecol 22:2841 2847. doi:10.1111/mec.12350

O'Brien S, Haussler D, Ryder O (2014) The birds of Genome10 K. GigaScience 3:32

Ogden R, Baird J, Senn H, McEwing R (2012) The use of cross species genome wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar horned oryx. Conserv Genet Res 4:471 473. doi:10.1007/s12686 011 9577 2

Opazo JC, Hoffman FG, Natarajan C, Witt CC, Berenbrink M, Storz JF (2014) Gene turnover in the avian globin gene families and evolutionary changes in hemoglobin isoform expression. Mol Biol Evol. doi:10.1093/molbev/msu341

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26:177 187. doi:10.1016/j.tig.2010.01.001

Pfenning AR et al (2014) Convergent transcriptional specializations in the brains of humans and song learning birds. Science 346:1256846. doi:10.1126/science.1256846

Pereira J et al (2014) Evolutionary genomics and adaptive evolution of the hedgehog gene family (Shh, Ihh and Dhh) in vertebrates. PLoS One 9. doi:10.1371/journal.pone.0074132

Piertney S (2006) Avian conservation genetics in the era of genomics. J Ornithol 147:17 18

Quinn TW, Quinn JS, Cooke F, White BN (1987) DNA marker analysis detects multiple maternity and paternity in single broods of the lesser snow goose. Nature 326:392 394

Quinn TW, Davies JC, Cooke F, White BN (1989) Genetic analysis of offspring of a female female pair in the lesser snow goose (*Chen c. caerulescens*). Auk 106:177 184

Rheindt FE, Fujita MK, Wilton PR, Edwards SV (2014) Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyran nidae): population genetic and phylogenetic inferences from genome wide SNPs. Syst Biol 63:134 152. doi:10.1093/sysbio/syt070

Romanov MN et al (2014) Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. BMC Genom 15:1060

Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. Genome Res 11:3 11

Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: a case study in the Eastern Fence Lizard. J Hered 98:331 336

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration linked genes and genomic islands in divergence of a songbird. Mol Ecol 23:4757 4769. doi:10.1111/mec.12842

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350 1354

Saiki RK et al (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487 491

Sanger F (1981) Determination of nucleotide sequences in DNA. Science 214:1205 1210

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463 5467

Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol Ecol 19:1439 1451

Scaglione D, Acquadro A, Portis E, Tirone M, Knapp SJ, Lanteri S (2012) RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. BMC Genomics 13. doi:10.1186/1471 2164 13 3

Schlötterer C (2004) The evolution of molecular markers just a matter of fashion? Nat Rev Genet 5:63 69

Schoebel CN et al (2013) Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequenc ing. J Evol Biol 26:600 611. doi:10.1111/jeb.12077

Schopen GCB, Bovenhuis H, Visker MHPW, Van Arendonk JAM (2008) Comparison of information content for microsatellites and SNPs in poultry and cattle. Anim Genet 39:451 453. doi:10.1111/j.1365 2052.2008.01736.x

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol Ecol 14:503 511

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

Senn H et al (2013) Reference free SNP discovery for the Eurasian beaver from restriction site associated DNA paired end data. Mol Ecol 22:3141 3150. doi:10.1111/mec.12242

Shafer ABA et al (2015) Genomics and the challenging translation into conservation practice. Trends Ecol Evol 30:78 87. doi:10.1016/j.tree.2014.11.009

Sibley CG, Monroe BL (1990) Distribution and taxonomy of birds of the world. Yale University Press, New Haven

Smith EFG, Arctander P, Fjeldså J, Amir OG (1991) A new species of shrike (Laniidae: Laniarius) from Somalia, verified by DNA sequence data from the only known individual. Ibis 133:227 235

Smith MJ, Pascal CE, Grauvogel Z, Habicht C, Seeb JE, Seeb LW (2011) Multiplex preamplification PCR and microsatellite val idation enables accurate single nucleotide polymorphism geno typing of historical fish scales. Mol Ecol Res 11:268 277

Storch V, Welsch U, Wink M (2013) Evolutionsbiologie (in German), vol 3. Spektrum Springer, Heidelberg

Suh A et al (2011) Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat Commun 2. doi:10.1038/ncomms1448

Suh A et al (2014a) Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. Genome Biol Evol. doi:10.1093/gbe/evu256

Suh A et al (2014b) Early Mesozoic coexistence of amniotes and Hepadnaviridae. PLoS Genet 10:e1004559. doi:10.1371/journal.pgen.1004559

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38:e159 e159. doi:10.1093/nar/gkq543

van Bers NE, van Oers K, Kerstens HH, Dibbits BW, Crooijmans RPMA, Visser ME, Groenen MAM (2010) Genome wide SNP detection in the great tit *Parus major* using high throughput sequencing. Mol Ecol 19:89 99

van Bers NEM et al (2012) The design and cross population application of a genome wide SNP chip for the great tit *Parus major*. Mol Ecol Res 12:753 770. doi:10.1111/j.1755 0998.2012.03141.x

van Tassell CP et al (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247 252

Vellekoop J, Sluijs A, Smit J, Schouten S, Weijers JWH, Sinninghe Damsté JS, Brinkhuis H (2014) Rapid short term cooling following the Chicxulub impact at the Cretaceous Paleogene boundary. Proc Natl Acad Sci USA 111:7537 7541. doi:10.1073/pnas.1319253111

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34:275 305

Wang B, Ekblom R, Bunikis I, Siitari H, Höglund J (2014a) Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster Z and MHC evolution. BMC Genomics 15. doi:10.1186/1471 2164 15 180

Wang R et al (2014b) Convergent differential regulation of SLIT ROBO axon guidance genes in the brains of vocal learners. J Comp Neurol. doi:10.1002/cne.23719

Wang Z, Zhang J, Yang W, An N, Zhang P, Zhang G, Zhou Q (2014c) Temporal genomic evolution of bird sex chromosomes. BMC Evol Biol 14:250

Warren WC et al (2010) The genome of a songbird. Nature 464:757 762. doi:10.1038/nature08819

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids a structure for deoxyribose nucleic acid. Nature 171:737 738

Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H (2014a) Evidence for GC biased gene conversion as a driver of between lineage differences in avian base composition. Genome Biol 15:549

Weber CC, Nabholz B, Romiguier J, Ellegren H (2014b) K_r/K_c but not d_N/d_S correlates positively with body mass in birds, raising implications for inferring lineage specific selection. Genome Biol 15:542

Whitney O et al (2014) Core and region enriched networks of behaviorally regulated genes and the singing genome. Science 346:1256780. doi:10.1126/science.1256780

Wink M (2006) Use of DNA markers to study bird migration. J Ornithol 147:234 244

Wink M (2011) Evolution und Phylogenie der Vögel Taxonomische Konsequenzen (in German). Vogelwarte 49:17 24

Wink M (2013) Ornithologie für Einsteiger (in German). Spektrum Springer, Heidelberg

Wink M (2015) Der erste phylogenomische Stammbaum der Vögel (in German). Vogelwarte 53:23 28

Wirthlin M, Lovell PV, Jarvis ED, Mello CV (2014) Comparative genomics reveals molecular features unique to the songbird lineage. BMC Genom 15:1082

Wong GKS et al (2004) A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms. Nature 432:717 722. doi:10.1038/nature03156

Xu X, Zhou Z, Dudley R, MacKem S, Chuong CM, Erickson GM, Varricchio DJ (2014) An integrative approach to understanding bird origins. Science 346:1253293. doi:10.1126/science.1253293

Zhan X et al (2013) Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. Nat Genet 45:563 566. doi:10.1038/ng.2588

Zhang G, Li B, Gilbert MTP, Jarvis ED, Wang J, The Avian Genome Consortium (2014a) Comparative genomic data of the Avian Phylogenomics Project. GigaScience 3:26

Zhang G et al (2014b) Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311 1320. doi:10.1126/science.1251385

Zhou Q et al (2014) Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346:1246338. doi:10.1126/science.1246338