View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Aberystwyth Research Portal

Aberystwyth University

Comparative genomics reveals insights into avian genome evolution and adaptation

Zhang, Guojie; Li, Cai; Li, Qiye; Li, Bo; Larkin, Denis M.; Lee, C.; Storz, J. F.; Antunes, Agostinho; Greenwold, Matthew J.; Meredith, Robert W.; Ödeen, Anders; Cui, Jie; Zhou, Qi; Xu, Luohao; Pan, Hailin; Wang, Zongji; Jin, Lijun; Zhang, Pei; Hu, Haofu; Yang, Wei

Published in:
Science

DOI:
10.1126/science.1251385

Publication date:
2014

Citation for published version (APA):
Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., ... Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, *346*(6215), 1311-1320. https://doi.org/10.1126/science.1251385

General rights
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Download date: 09. Jul. 2020

Comparative Genomics Across Modern Bird Species Reveal Insights into Pan-avian Genome evolution and Trait Biodiversity

Authors:

Guojie Zhang[1,2,*,#], Cai Li[1,3*], Qiye Li[1,3], Bo Li[1], Denis M. Larkin[4], Chul Lee[5], Jay F. Storz[6], Agostinho Antunes[7,8], Robert W. Meredith[9], Anders Ödeen[10], Jie Cui[11], Qi Zhou[12], Luohao Xu[1], Hailin Pan[1], Zongji Wang[1], Lijun Jin[1], Pei Zhang[1], Haofu Hu[1], Wei Yang[1], Jiang Hu[1], Jin Xiao[1], Zhikai Yang[1], Yang Liu[1], Qiaolin Xie[1], Jinmin Lian[1], Ping Wen[1], Fang Zhang[1], Hui Li[1], Yongli Zeng[1], Zijun Xiong[1], Shiping Liu[1], Long Zhou[1], Zhiyong Huang[1], Na An[1], Jie Wang[1], Qiumei Zheng[1], Yingqi Xiong[1], Guangbiao Wang[1], Bo Wang[1], Jingjing Wang[1], Yu Fang[13], Rute da Fonseca[3], Alonzo Alfaro-Núñez[3], Mikkel Schubert[3], Ludovic Orlando[3], Tobias Mourier[3], Jason Howard[14], Ganeshkumar Ganapathy[14], Julia Smith[14], Marta Farré[4], Jitendra Narayan[15], Gancho Slavov[15], Michael N Romanov[16], Rui Borges[7,8], João Paulo Machado[7,17], Imran Khan[7,8], Mark S. Springer[18], John Gatesy[18], Federico G. Hoffmann[19,20], Juan C. Opazo[21], Olle Håstad[22], Matthew J. Greenwold[23], Roger H. Sawyer[23], Heebal Kim[5], Kyu-Won Kim[5], Ning Li[24], Yinhua Huang[24,25], Michael W. Bruford[26], Xiangjiang Zhan[26], Andrew Dixon[27], Mads Bertelsen[28], Elizabeth Derryberry[29], Wesley Warren[30], Shengbin Li[31], David A. Ray[32], Richard E. Green[33], Stephen J. O'Brien[34], Darren Griffin[16],Warren E. Johnson[35], David Haussler[33], Oliver A. Ryder[36], Eske Willerslev[3], Gary Graves[37,38], Per Alström[39,40], Jon Fjeldså[38], David Mindell[41], cott V. Edwards[42], Edward L. Braun[43], Carsten Rahbek[38], David W. Burt[44], Peter Houde[45], Yong Zhang[1], Huanming Yang[1], Jian Wang[1], Erich D. Jarvis[14, #], M Thomas P Gilbert[3,46,#], Jun Wang[1,47,48,#] & Avian Genome Consortium

Consortium author list: Chen Ye[1], Shaoguang Liang[1], Zengli Yan[1], Lisandra Zepeda[3], Paula Campos[3], Amhed Missael Vargas Velazquez[3], Jose Alfredo Samaniego[3], María Avila-Arcos[3], Michael D Martin[3], Ross Barnett[3], Angela M. Ribeiro[7], Daniela Almeida[7,8], Emanuel Maldonado[7], Kartik Sunagar[7,8], Siby Philip[7,8], Maria Gloria Dominguez-Bello[49], Mike Bunce[46], David Lambert[50], Edward C. Holmes[51], Paul Gardner[52], Eric Lyons[53], Fiona McCarthy[54], Frederique Pitel[55], Douglas Rhoads[56]

Affiliations:

1 China National GeneBank, BGI-Shenzhen, Shenzen, 518083, China
2 Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark
3 Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark
4 Royal Veterinary College, University of London, London, UK
5 Laboratory of Bioinformatics and Population Genetics, Department of Food and Animal Biotechnology, Seoul National University, Seoul, Korea
6 School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA
7 CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal
8 Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo

Alegre, 4169-007 Porto, Portugal
9 Department of Biology and Molecular Biology. Montclair State University, Montclair, NJ 07043
10 Department of Animal Ecology, Uppsala University, Norbyvägen 18D, S-752 36 Uppsala, Sweden.
11 Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
12 Department of Integrative Biology University of California, Berkeley CA 94720, USA
13 Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China
14 Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC27710, USA
15 Institute of Biological, Environmental and Rural Sciences, University of Aberystwyth, Aberystwyth, UK
16 School of Biosciences, University of Kent, Canterbury, UK
17 Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Portugal.
18 Department of Biology, University of California Riverside, Riverside, CA 92521, USA
19 Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA
20 Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA
21 Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
22 Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, P.O. Box 7011, S-750 07, Uppsala, Sweden
23 Department of Biological Sciences, University of South Carolina, Columbia, South Carolina, USA
24 State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing, 100094, China
25 College of Animal Science and Technology, China Agricultural University, Beijing, 100094, China
26 Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK
27 International Wildlife Consultants Ltd., Carmarthen SA33 5YL, Wales, UK
28 Centre for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark
29 Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA
30 The Genome Institute at Washington University, St Louis, Missouri 63108, USA
31 College of Medicine and Forensics, Xi'an Jiaotong University, Xi'an, 710061, China
32 Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA
33 Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA

34 Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia
35 Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA.
36 Genetics Division, San Diego Zoo's Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, California 92027, USA
37 Department of Vertebrate Zoology , MRC-116 , National Museum of Natural History , Smithsonian
Institution, P. O. Box 37012, Washington, D.C. 20013-7012, USA
38 Center for Macroecology, Evolution and Climate, the Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O, Denmark
39 Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, P.R. China
40 Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden
41 Department of Biochemistry & Biophysics, University of California, San Francisco, CA 94158, USA
42 Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology Harvard University 26 Oxford Street Cambridge, MA 02138, USA
43 Department of Biology, University of Florida, Gainesville, FL 32611, USA
44 Dept. of Genomics and Genetics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK
45 Department of Biology, New Mexico State University, Box 30001 MSC 3AF, Las Cruces NM 88003, USA
46 Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6845, Australia
47 Department of Biology, University of Copenhagen, DK-1165 Copenhagen, Denmark
48 King Abdulaziz University, Jeddah 21589, Saudi Arabia
49 Department of Biology, University of Puerto Rico, Av Ponce de Leon, Rio Piedras Campus, JGD 224, San Juan, PR 009431-3360, USA
50 Environmental Futures Centre, Griffith University, Nathan, Queensland, Australia 4121
51 Sydney Emerging Infections and Biosecurity Institute, School of Biological Sciences and Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia
52 School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
53 Department of Plant Sciences, University of Arizona, Tucson 85721, USA
54 Department of Veterinary Science and Microbiology, The University of Arizona, 1117 E. Lowell Street, PO Box 210090-0090, Tucson AZ 85721, USA
55 Laboratoire de Génétique Cellulaire, INRA Chemin de Borde-Rouge, Auzeville, BP 52627 , 31326 CASTANET-TOLOSAN CEDEX, France
56 Department of Biological Sciences, Science and Engineering 601, Fayetteville, AR 72701, USA

* These authors contributed equally to this work.

Corresponding authors. E-mail: zhanggj@genomics.cn, jarvis@neuro.duke.edu, mtpgilbert@gmail.com and wangj@genomics.cn.

Abstract:

Genomic comparison across multiple species is a key tool for revealing macroevolutionary patterns of gene and genome evolution. Using the full genomes of 48 avian species representing all major extant clades, we present the first large scale analysis of vertebrate genome evolution outside of mammals. We found the smaller size of avian genomes was largely the result of massive erosion of repetitive elements, large segmental deletions, as well as gene loss following the split from other reptiles. Avian genomes show a remarkably high degree of evolutionary stasis from chromosomes, gene synteny, down to single nucleotides. The overall of nucleotide substitution rate in birds is slower than in mammals, and varies among avian lineages according to life history and ecological factors. We also identified many protein-coding genes that are evolving non-neutrally, as well as non-coding RNA and regulatory regions that are highly conserved. Together our analyses reveal that genomic biodiversity across birds covaries with diverse adaptations to different lifestyles.

Main Text:

With ~10,500 living species (*1*), birds are the most species-rich class of tetrapod vertebrates. Birds originated from a theropod lineage more than 150 million years ago (MYA) during the Jurassic and are the only extant dinosaurs (*2, 3*). The earliest diversification of Neornithes, the clade comprising all extant birds, occurred during the Cretaceous. However, the Neoaves, the most diverse avian clade, underwent rapid global expansion and radiation following a mass extinction event about 66 MYA near the Cretaceous-Paleogene (K-Pg) boundary ((*4*) and see also our companion paper (*5*)). The extant avian lineages possess enormous ecological diversity, having exploited all major environments except the deep oceans, and exhibit extremely diverse morphologies and rates of diversification. Given the nearly complete global inventory of avian species, and the immense amount of distributional and biological data that has been collected, birds are widely used as a model group for investigating large-scale evolutionary and ecological questions (*6, 7*). A number of avian species (e.g. chicken [*Gallus gallus*], zebra finch [*Taeniopygia guttata*], pigeon (rock dove) [*Columba livia*]) are also important model organisms in disciplines such as neuroscience and developmental biology (*8*). In addition to their scientific relevance, birds are a widely used focus group for global conservation priority setting (e.g., (*9*)) and culturally and economically important to human societies. A number of avian species have been domesticated and are farmed on extremely large and economically important scales. Birds have captured our imagination as pets and as the subject of nature study, and both farmed and wild water birds are key players in the spread of globally important pathogens, such as avian influenza virus (*10*).

Despite the need to better understand avian genomics, prior to the generation of the dataset reported here, annotated avian genomic data was publicly available for only few species: the domestic chicken, domestic turkey (*Meleagris gallopavo*) and zebra finch (*11-13*). To build an understanding of the genetic complexity of birds and to investigate links between their genomic variation and phenotypic diversity, we generated genome sequences for an additional 45 avian species, selected to represent all 32 neognath and two of the five palaeognath orders (**Fig. 1**) (*14*), thus representing nearly all of the major clades of living birds (*5*). Three of these genomes have recently been published in stand-alone studies (domestic Pekin duck [*Anas platyrhynchos*

domestica], peregrine falcon [*Falco peregrinus*], and pigeon (*15-17*)). Several of those described here are also the focus of stand-alone companion studies (crested ibis [*Nipponia nippon*] and little egret [*Egretta garzetta*], Adelie [*Pygoscelis adeliae*] and emperor penguin [*Aptenodytes forsteri*], budgerigar [*Melopsittacus undulatus*]) (*18-20*) reported in parallel with this study. We also collaborated with research groups that sequenced reptile genomes representing major non-avian lineages (American alligator [*Alligator mississippiensis*] (*21, 22*), green sea turtle [*Chelonia mydas*] (*23*) and a lizard, the green anole [*Anolis carolinensis*] (*24*), as outgroups. We note that several additional draft avian genomes of varying quality have recently been published by other parties, including the collared flycatcher (*Ficedula albcollis*.), Saker falcon (*Falco cherrug*), ground tit (*Parus humilis* [also known as *Pseudopodoces humilis*]) and Puerto Rican parrot (*Amazona vittata)* (*16, 25-27*), although they do not form part of our study.

In addition to providing an extensive genomic resource, the generation of sequences spanning nearly all extant avian orders within a diverse and species-rich vertebrate group offers an unprecedented opportunity to study the evolution of genomes from a broad phylogenetic perspective. Avian genomes are considerably smaller and more compact than those of other vertebrates (*28*), exhibit a largely conserved karyotype consisting of few macrochromosomes and many microchromosomes (*29*), but exhibit considerable within group variation in overall pattern (*30*). This data set enables us to document the genome structure and evolutionary history of different genomes across extant birds and to provide insight into the unique avian genome characteristics and to describe genome variation that has played key roles in the evolution of specific avian lineages.

Sequencing, assembly and annotation

We employed a whole genome shotgun strategy to generate genome sequences of 45 new avian species [**Supporting Online Material (SOM) text1**], including two species representing two orders within the infraclass Paleognathae (common ostrich [*Struthio camelus*] and white-throated tinamou [*Tinamus guttatus]*), the Pekin duck representing a first genome for the order Anseriformes (clade Galloanserae), and 41 species representing 30 neoavian orders (**Table S1**). In combination with the three previously published avian genomes, the galliform chicken (*11*) and turkey (*12*), and the passerine zebra finch (*13*), the genome assemblies cover 92% (34/37) of all avian orders as presented by Howard and Moore (the three missing orders belong to the Paleognathae) (*14*). With the exception of the budgerigar, which was assembled through a multi-platform (Illumina/GS-FLX/PacBio) approach (*20*), all other new genomes were sequenced and assembled with Illumina short reads (**Fig1, SOM text 1**). For 20 species, we produced high (>50X) coverage sequences from multiple libraries with a gradient of insert sizes and built full genome assemblies. For the remaining 25 species, we generated low (~30X) coverage data from two insert-size libraries and built less complete but still sufficient assemblies for whole genome comparative analyses. The genomes were assembled *de novo* (**SOM text 1**), and despite the lower coverage for some species, the resulting assemblies typically ranged from 1.05 Gb to 1.26 Gb, consistent with independently estimated cytology-based avian genome sizes (*28*) and suggesting near complete genome coverage for all species. Scaffold N50 sizes for high coverage genomes ranged from 1.2 to 6.7 Mb, whereas those for lower coverage genomes were about 20kb. The genomes of two species, the common ostrich and budgerigar, were further assembled with input from optical mapping, doubling their N50 scaffold sizes to 17.7Mb and 13.8Mb, respectively (*20, 31*).

We annotated the protein coding sequences using a homology-based method for all genomes, aided by transcriptome sequencing for some species (**SOM text 1**). To avoid systematic biases related to the use of different methods in the annotation of the previously published avian genomes, we created a uniform reference gene set that included all genes from the previously published chicken and zebra finch genomes, and those identified in the green anole (*24*) and human genomes (*32*) as outgroups. This database was subsequently used for prediction of the protein gene models in all avian genomes. All high-coverage genomes were predicted to contain ~15,000-16,000 transposable element-free protein-coding genes (**Table S5**), a number similar to the published chicken genome (~15,000). Despite the fragmented nature of the low-depth assemblies leading to ~3,000 genes that were likely missing or partially annotated it was still possible to study 70-80% of the entire catalogue of avian genes. Orthology relationships for 8,295 genes among all birds and outgroups were assigned using the reciprocal best blast hit algorithm and information about synteny, using the chicken genome as a reference, allowing a maximum of five missing species (*5*). These genome assemblies and annotations can be found at (http://phybirds.genomics.org.cn/), and separated annotated gene sets are in public databases (NCBI and ENSEMBL).

Broad patterns of avian genome evolution

Genome size reduction

Although many fishes and some amphibians have smaller genomes than birds, among the amniotes birds have the smallest genomes. The genomes of mammals and non-avian reptiles (henceforth referred to as 'reptiles') typically range from 1.0 to 8.4 pg, whereas avian genomes are reduced, ranging from 0.91 in the black-chinned hummingbird (*Archilochus alexanderi*, not sequenced in this study) to a little over 2 pg in the common ostrich (*28*). A number of hypotheses have been proposed for the smaller avian genome size (*33-36*). Here we document key events that have contributed to this smaller genome size.

Transposable elements

The proliferation and loss of transposable elements (TEs) are thought to be a major driving factor behind vertebrate genome size evolution (*37-39*). Consistent with analyses of the previously published zebra finch and galliformes genomes (*11-13*, *40*), we found that almost all avian genomes contained lower levels of repeat elements (~4-10% of each genome, **Table S4**) than has been observed in other tetrapod vertebrate genomes (e.g. 34-52% in mammals) (*41*). The sole outlier was the downy woodpecker (*Picoides pubescens*), with TEs representing ~22% of the genome, derived mainly from species-specific expansion of LINE (Long Interspersed Elements) type CR1 (chicken repeat 1) transposons (**Fig. S1**). Likewise, we found the CR1 transposon to be the largest TE family in all avian genomes, but with widely different rates of expansion among different lineages, comprising 2% to 18% of their genomes (**Fig. S1**). Despite their abundance, most of the ancestral reptile CR1s found by mining the three non-avian reptile genomes (American alligator, green sea turtle. and green anole) have been inactivated in the birds (note that these three species are at least as divergent from each other as each are from birds). The genome alignments suggest that a deficiency of SINEs (Short Interspersed Elements) in the common ancestor of birds was the main feature of TE reduction in avian genomes (**Fig. S2**). The average copy number of SINE residues in birds has been reduced to about 1.3 million, which is significantly fewer than the number in other reptiles (12.6 million in American alligator and 34.9 million in green sea turtle).

Size reduction of avian protein coding genes

We compared the average length of different genomic elements of birds with those of available complete genome sequences of 24 mammals and the 3 non-avian reptiles. We found that for 90% of 5,087 orthologous protein coding genes, the avian orthologs were on average 50% shorter than their mammalian orthologs and 73% shorter than their reptilian orthologs (**Fig. 2A**). This reduction is largely due to the shortening of avian intronic regions. The reduction in avian genome size is represented by an increased gene density, resulting due to reduced intergenic distances in birds (**Fig. 2A**). We speculate that the overall condensed genome and reduced distances between genes may represent an adaptation tied to the need for the rapid gene regulation required during powered flight.

To further investigate whether avian genome size reduction arose due to lineage-specific reduction in the common avian ancestor of birds or expansion in other vertebrates (*42*), we identified small (<100bp) deletion events using an alignment of the 48 avian and three reptile genomes and performed ancestral state reconstructions (**SOM text 2**). We found that the avian ancestral lineage experienced the largest number of small deletion events, about twice the number in the common ancestor of birds and crocodilians (**Fig. 2B**), suggesting the largest contribution of deletion events occurred specifically in the avian ancestor. In contrast, the variance in the number of small deletion events among avian lineages is relatively small.

Large segmental deletions in avian ancestral lineage

We searched for putative large segmental deletions (*42*) in the avian genomes and their inferred common ancestor by creating a gene synteny map between the highest quality assembled basal avian genome (common ostrich) and the genome of the green anole, which allowed us to document lineage-specific deletion events in the remaining avian and reptile genomes (**SOM Text 2**). In doing so, we detected 118 syntenic blocks, spanning a total of 58 Mb, which were present in the American alligator and green sea turtle genomes but were lost in all birds (**Table S10**). By contrast, there were significantly fewer syntenic blocks missing in the American alligator (14 blocks, 9Mb) and green sea turtle (27 blocks, 8Mb) relative to the green anole, confirming the polarity of genome size reduction in birds (**Table S10**). The large segmental losses in birds were not uniform across genomes, but were skewed to losses on chr2 and chr6 of the green anole genome (**Fig. S12**). Two of the green anole's 12 pairs of microchromosomes, LGd and LGf, were completely missing in birds without any homologous genes found in avian genomes. Most of these lost segments were located at the ends of chromosome or close to the centrosomes (**Fig. S12**). A further point of interest is that the lost segments were enriched at apparent breakpoints of the avian microchromosomes (**Fig. S12**). For instance, green anole chr2, which is homologous to multiple avian macrochromosomes, including the Z and W sex chromosomes (*31, 43*), and also to microchromosomes, was enriched with 22% of the avian-specific segmental deletions at chromosomal breakpoints (**Fig. 2C**). These findings imply that the large segmental losses could have arisen as a consequence of chromosomal fragmentation events in the common ancestor of birds that also gave rise to additional microchromosomes in modern birds.

The large segmental regions deleted in birds contain at least 1,241 functional protein coding genes, with each lost segment containing at least five continuous genes. The largest region lost in all birds was represented by a 2.1 Mb segment of the green anole chr2, which contains 28 protein-coding genes (**Fig 2C**). Overall, we found that at least 7% of green anole genes on

macrochromosome were lost by segmental deletion in birds. Although gene loss is a common evolutionary process, based on what has been observed to date, this massive level of segmental deletion is rare in vertebrates. Over 77% of the 1,241 genes present in the large segmentally deleted regions have at least one additional paralog in the Green Anole Lizard genome, a level higher than the overall percentage of genes with paralogs in the green anole genome or avian genomes (both at about 70%). This suggests functional compensation by the remaining avian parologs, a process that could have reduced purifying selection against the loss of these segmental regions. Thus whole genome duplication events in early vertebrates (*44*) may have offered a buffering mechanism to tolerate such massive loss of paralogous genes later during the diversification of birds. Nevertheless, the loss of functions associated with many genes in the avian ancestor may have had a profound influence on avian-specific traits (**Table S9** and see below).

Conservative mode of genome evolution

Selective pressure for smaller genome size in birds is hypothesized to have had significant consequences on genome conservation (*45*). To investigate this hypothesis, we examined conservation in the 48 genomes at three levels: chromosome rearrangement, gene synteny, and overall mutation rate at neutrally evolving 4-fold degenerate codon (4d) positions.

Genome rearrangement

With ~30 pairs of microchromosomes for most species, the avian karyotype represents a unique pattern of genome organization; i.e. based on basic cytological analyses the “so many, so small” microchromosome configuration appears to be highly conserved in roughly two thirds of all birds and is not a general feature of any other group of organisms studied to date (*29*). Once adopted, a compact genome organization would most likely lead to the reduction in rates of chromosomal evolution and increase in re-use of approximate genomic locations at which fragmentation occurs (breakpoints), thus avoiding negative epigenetic and gene effects and, as a result, making the chromosomal organization in birds evolutionary stable. To test this idea, we analyzed chromosomal evolution of the sequenced genomes. Six of the avian genomes (chicken, turkey, zebra finch, Pekin duck, common ostrich, budgerigar) were at or near chromosomally assembled status; 15 others had large genomic scaffolds due to high coverage and large linking libraries (**SOM Text 3**). This allowed us to assess the rates of avian chromosomal evolution in several representatives of the major avian clades. Alignment to the chicken genome of the other 20 avian assemblies, plus two outgroups (green anole and the snake, *Boa constrictor* (*46*)), identified homologous synteny blocks (HSBs) and 1,746 evolutionary breakpoint regions (EBRs) in different avian lineages. We estimated the expected number of EBRs in each avian lineage by applying a correction proportional to the fraction of missed reference-specific EBRs in each genome (ranging from 1.45% to 33.33%). Based on this, we further estimated the rates of genomic rearrangements in avian lineages with our total evidence nucleotide (TENT) phylogenetic tree as a guide (*5*). We detected an unusually high fraction of small lineage-specific rearrangements in the turkey genome suggesting a high number of local misassembles, and thus it was excluded from further analyses. For the genomes built with Illumina data and large mate-pair insert libraries (18 genomes in the present study), the estimated rate of chimeric scaffolds that could lead to false EBRs in this analysis was ~6% (*47*).

Our data demonstrate that breakpoint re-use was up to 2-fold lower in birds (~7-10%) than in mammals (~14.9%) (*48*, *49*). Birds also had a lower average chromosomal rearrangement rate

(~1.25 EBRs/MY) than mammals (~2 EBRs/MY, (*50*) and unpublished data) (p=0.04, t-test=1.79) under comparable HSB detection resolutions (**Fig. S13**). Despite this, several bursts of genomic reorganizations appeared to have occurred in the evolution of several avian lineages (**Fig. S13**). For example the origin of Neognathae was accompanied by an elevated rate of chromosome rearrangements (~2.87 EBRs/MY), and there were elevated rates in the penguins and in duck relative to chicken. Intriguingly, all vocal learning species (songbirds, parrots, and hummingbirds) also had higher rates of rearrangements compared to both their close non-vocal learning relatives (golden-collared manakin [*Manacus vitellinus*], peregrine falcon and chimney swift [*Chaetura pelagica*]) (p-value=0.03, t-test=2.39) and had some of the highest compared to all non-vocal learning species (p-value=0.0006, t-test=3.87). This difference may be related to the fact that all three groups have undergone larger radiations relative to most other bird groups, or alternatively went through bottlenecks affecting their ancestral populations resulting in fixation of a larger number of genomic rearrangements than in other bird species; the Golden-collard Manakin representing suboscine songbirds (vocal non-learners) also have undergone a larger radiation relative to parrots and hummingbirds, and have a low rate. Both scenarios favor for the selection of adaptive features associated with a large number of genomic rearrangements.

Conserved gene synteny

To better quantify and compare synteny in birds, we compared local gene arrangements (micro-synteny), which are also more robust and accurate given the current draft assemblies (**SOM Text 3**). We used avian genome assemblies with long scaffold N50 sizes (>1Mb) and focused on the eutherian mammals, which are approximately the same evolutionary age, and whose genome assemblies are similar in quality. Examining the fraction of orthologous genes that can be identified from each pair of two-avian/mammalian genomes based on syntenic and best reciprocal blast matches (**SOM Text 3**), we find that birds have a significantly higher percentage of orthologous genes at any pairwise comparison than do mammals (**Fig. 2D**). The variance of the percentage of synteny-defined orthologous genes in pairwise comparisons of birds was also less than in mammals. The fraction of genes retained in synteny blocks in any pairwise comparison was linearly related over evolutionary time, whereby the overall level of genome shuffling in birds was lower than in mammals over the past ca.100 million years (**Fig. 2D**). For example, the percentage of human genes within syntenic orthologous block with other eutherian mammals ranged from 87% with chimpanzee to 54% with the tree shrew, with an average ratio of 60%. In contrast, the percentage of syntenic genes in zebra finch with other birds was very stable at around 68%. Extrapolating this trend strongly suggests a substantially higher level of constraint on maintaining gene synteny in avian evolution than that in mammals.

Taken together, a picture emerges of a conserved and stable pattern of avian genome organization, where the smaller genome size limits the number of permissive EBR sites therefore decreasing the chromosome rearrangement rate and increasing EBR reuse compared to larger mammalian genomes. Nonetheless, punctuated examples of increased levels of chromosomal rearrangements occurred concomitantly with diversification events. One consequence of this apparently avian-unique pattern with rare inter- and intra-chromosomal evolutionary changes would be relative stasis in gene family synteny in comparison to other vertebrates, or at least mammals, an idea we tested next for a specific gene family.

Evolutionary stasis in gene families

Apparent stasis in the evolution of avian chromosomes would suggest that birds may have experienced relatively low rates of gene gain and loss in multigene families. To investigate this possibility, we examined the families of genes that encode the various α- and β-type subunits of hemoglobin, the tetrameric protein responsible for blood oxygen transport in all vertebrates. The globin gene families are among the most intensively studied vertebrate-specific gene families from a functional perspective (*51-53*). In amniote vertebrates, the α- and β-globin gene families are located on different chromosomes (*52*). In birds, we found that the size and membership composition of the globin gene families have remained remarkably constant during ~115 million years of evolution (**Fig. 2E**). Most examined species have retained an identical complement of three orthologous α-like globin genes and 4 orthologous β-like globin genes (**Fig. 2E**). In eutherian mammals, by contrast, the α- and β-globin gene families have experienced high rates of gene turnover due to lineage-specific duplication and deletion events (**Fig. 2E**), resulting in functionally significant variation in globin gene repertoires among contemporary species (*54*, *55*). Estimated rates of gene turnover (λ) were over two-fold higher for the mammalian gene clusters relative to the avian gene clusters ($\lambda = 0.0023$ vs. 0.0011, respectively).

Much of the variation in the avian α-globin gene family was attributable to multiple independent inactivations of the α^D-globin gene (**Fig. 2E**) (*56*), which encodes the α-chain subunit of a functionally distinct hemoglobin isoform (HbD) that is expressed in both embryonic and definitive erythrocytes. In the sample of 22 species with high-coverage genome assemblies, parsimony suggests that there were at least five independent pseudogenized inactivations of α^D-globin, assuming no reversals (i.e., reactivations): in the golden-collared manakin, killdeer, penguins, little egret, and hoatzin (**Fig. 2E**). Due to uniform and consistent differences in oxygen-binding properties between HbD and the major adult-expressed hemoglobin isoform, HbA (which incorporates products of α^A-globin) (*56*), these inactivations of α^D-globin likely contribute to among-species variation in blood-oxygen affinity, which has important consequences for circulatory oxygen transport and aerobic energy metabolism. Because of the pronounced degree of functional differentiation between co-expressed Hb isoforms in avian red cells (*56*), the observed variation in globin gene content among avian lineages may contribute to significant physiological variation in blood-oxygen transport. Overall, the globin gene families illustrate what appears to be a general pattern of evolutionary stasis in birds, in contrast to a higher rate of gene turnover in mammals.

Reduced genomic substitution rates

Genomic substitution rates vary across species, and are determined by both neutral and adaptive evolutionary processes (*57*). We found that the overall pan-genomic background substitution rate in birds was significantly lower than in mammals ($p= 0.00037$, Wilcoxon rank sum test), as determined by analysis of four-fold degenerate (4d) sites in coding regions (ca. 0.002 substitution per site per million years). However, the substitution rate estimates also exhibited notable inter-ordinal variation among birds (**Figs. 3A**). Intriguingly, there was a positive correlation between the substitution rate and the number of species per order (**Fig. 3B**, $p=0.001$, Pearson's test with phylogenetically independent contrasts), providing evidence of an association with rates of macroevolution. For example, Passeriformes, the most diverse bird order and one of the most diverse vertebrate orders, consisting of nearly 60% of living avian species, exhibited the highest evolutionary rate (ca. 0.0033 substitution per site per million years), almost two times the average rate of Neoaves (ca. 0.0020 substitution per site per million years). Origins for many neoavian orders are relatively close to each other in time (*5*), suggesting this finding is not

simply an artifact of taxonomy. We furthermore observed a significant negative correlation between avian body mass and substitution rate (p=0.009, Pearson's test with phylogenetically independent contrasts), supporting previous hypotheses on the effect of body size on rate of molecular evolution (*58*). Moreover, we found the substitution rates of birds also varied according to certain life-style and biological traits. In general, core landbirds (**Fig. 1;** as defined in (*5*)) exhibited on average higher substitution rates than their sister group of core waterbirds (landbirds: ca. 0.0022 substitution per site per million years; waterbirds: ca. 0.0016 substitution per site per million years), which is consistent with the observation that landbirds have greater net diversification rate than waterbirds (*7*). Among the core landbirds, the two predatory lineages (white-tailed and bald eagles [*Haliaeetus albicilla* and *H. leucocephalus*], red-legged seriema [*Cariama cristata*], peregrine falcon, turkey vulture [*Cathartes aura*], and barn owl [*Tyto alba*]), exhibited significantly slower rates of evolution (ca. 0.0016 substitution per site per million years), similar to that of core waterbirds. Given the basal placements of birds-of-prey within the core landbirds, the subsequent evolution of other core landbirds towards smaller body sizes, shorter generation times, and more fully developed altricial behavior, appears to have been associated with an overall elevated rate of substitution. Moreover, the substitution data indicates that the genomes of birds with vocal learning behavior within core landbirds (parrots and songbirds) and outside of this group (hummingbirds) are evolving faster than non-vocal learners. Overall, our analyses strongly indicate that genome-wide variation in rates of substitution are a consequence of their radiation into a wide range of niches and associated phenotypic changes, including body size, life-style, and other behavioral and ecological factors.

Selective constraints on functional elements

Comparative analyses across multiple genomes also offer the potential to provide new insight into the processes of how natural selection drives genome sequence evolution, the functional divergence of protein-coding genes, and lineage-specific evolutionary dynamics. Previous comparative studies of mammalian genomes revealed evolutionary constraints on ~5% of the genome, encoded by a variety of functionally important elements (*59*). Our dataset enables similar analyses to identify the essential core component of sequence elements within birds, and also those that vary between lineages.

DNA sequence conservation due to purifying selection

The conservation of DNA sequences across distantly related species is considered to reflect strong functional constraints (*60*). To investigate the overall conservation and selection patterns of avian genomes compared to other vertebrate groups, we performed multiple alignments using MultiZ (*61*) and quantified constraint across the genomes using a phylogenetic hidden Markov model, PhastCons (*60*) (**SOM Text 4**). We found that the overall level of conservation at the single nucleotide level among birds was higher than that in mammals (**Fig. 3C**), consistent with the slower rate of sequence evolution found in mitochondrial genes in birds (*62*). A direct comparison of 100Mb of orthologous genomic regions among birds and mammals also revealed genomic regions evolving significantly slower than the neutral rate among birds than mammals. This included 3.2 million highly conserved elements (HCEs) at a resolution of 10bp or greater within the 48 avian species, spanning on average 7.5% of the avian genome, suggesting a stronger functional constraint in avian genomes. Functional annotations revealed about 12.6% of these HCEs were associated with protein-coding genes, while the majority of the remaining HCEs were located in introns or intergenic regions (25.8% and 56.7%, respectively) (**Fig. 3D**). These HCEs enabled us to identify 717 new protein-coding exons and 137 new protein-coding

genes that were otherwise unannotated in previous pipelines (**Table S16**). Approximately 77% of the new coding genes are supported by the deep transcriptome data (**Table S16**).

Deep transcriptome sequencing also enabled us to annotate 5,879 candidate long non-coding RNA (lncRNA) genes. We found several well-characterized lncRNA families of known function to be conserved between mammals and birds, although they showed different patterns of conservation. One such family is *cyrano*, an ancestral vertebrate lncRNA that is known to play crucial roles in maintaining normal embryonic development (*63*); experimental deletion in developing zebrafish (*Danio rerio*) leads to abnormally small head and eye development and neural tube defects (*63*). In contrast to mammals, in which only a small region of this gene is conserved, the whole gene is highly conserved across all avian genomes (**Fig. S18**).

As HCEs may have unique functions in different lineages, we separated the HCEs into two categories: bird-specific and a core set of amniote HCEs shared by birds and mammals. Alignments against annotated genes identified 13 protein-coding genes that were specifically conserved in avian lineages (**Table S17**). One of the most conserved was the sperm adhesion gene, *SPAM1*, which mediates sperm binding to the egg coat (*64*), and is involved in meiotic drive and genomic conflict (*65*). This gene has been reported to be under positive selection driven by sperm competition in mammalian species (*66*). Surprisingly, >87% of the entire coding region of *SPAM1* has been highly conserved among birds, as opposed to only 23% in mammals, suggesting that this gene has been under strong negative selection in birds. We also identified three genes that are conserved in birds, but that are located in mammals on the pseudo-autosomal regions of the sex chromosome, in which they suffer the high mutagenic effects of recombination (*67*). Nevertheless, their functions have been well maintained within birds. These findings suggest that different selection constraints on these genes may be associated with divergent mating systems or sex chromosome systems between the birds and mammals (*68*).

Non-coding HCEs are thought to play important functions in regulation of gene expression (*69*). We combined experimental data on gene expression from the ENCODE project (*70*) with bioinformatic prediction of transcription factor binding sites in gene regulatory regions, and found that the avian-specific HCEs are significantly associated with transcription factors that show precise functions for metabolism, while vertebrate core HCEs are also enriched with transcription factors functioning in signal regulation, stimulus responses, and development (**Table S18, S19**). For example, although birds and mammals have different facial structure, their facial development appears to utilize the same genetic toolkit, such as MID1, which is associated with avian beak development and the human cleft palate (*71*). Patterns of evolutionary conservation enabled us to identify several conserved elements unique to promoter regions of MID1 in birds (**Fig. 3E**), a feature that may regulate the spatial and/or temporal regulation in gene expression and lead to the evolution of lineage-specific cranio-facial morphologies.

Natural selection on genes

To investigate evolutionary constraints on gene regions, we calculated the ratio of nonsynonymous versus synonymous substitution rates (*dN/dS*) for the 8,295 high-quality orthologs identified among birds annotated from syntenic information. We found that the level of selective constraint on genes varied across species and chromosomes (**Fig. 4A**). Consistent with the “fast-Z” hypothesis, we observed that the evolutionary rate of Z-linked genes was significantly higher than that of the autosomes (**Fig. 4A**). This is most likely driven by the reductio of effective population size (Ne) of Z-linked genes, as the Ne of Z chromosome is only

3/4 of that of autosome, and male sexual selection would reduce it further (72). Furthermore, consistent with the "fast-macro" hypothesis, the overall rate of macrochromosomal genic evolution is higher than that of microchromosomes, which is probably caused by differences in the recombination rates and genic densities between macro- and microchromosomes in birds (73).

We also examined the median *dN/dS* ratio of each GO functional category of avian genes and compared it with that of mammals. The data indicated clear differences in categories, with those involved in development (spinal cord development, and bone resorption) evolving faster in birds, and those involved in the brain function (synapse assembly, synaptic vesicle transport and neural crest cell migration) evolving faster in mammals (**Table S21, S22**). We also observed heterogeneity in *dN/dS* ratios among avian lineages. Genes involved in oxidoreductase activity were relatively rapid evolving in the basal Palaeognathae clade that contains the flightless ratites. The fastest evolving genes within the Galloanserae participate in microtubule-based processes. In Neoaves, transmembrane transporters were the fastest evolving genes. These differences could perhaps be caused by relaxed selective constraints or positive selection in different lineages.

Genotype-phenotype convergent associations: evolution of vocal learning

With the availability of genomes representing all the major lineages of modern birds and their revised phylogenetic relationships (5), it becomes possible to conduct genome-wide association studies across species with convergent traits. Here we focused on vocal learning, which given our phylogenetic analyses is inferred as having evolved independently, either twice, in hummingbirds and the common ancestor of songbirds and parrots, or independently in the three groups (5, 74). All three groups have specialized song learning forebrain circuits not found in non-vocal learners (**Fig. 5A)** (75). To assess whether this behavioral and anatomical convergence is associated with molecular convergence, we first looked for convergent accelerated evolution of genes in avian vocal leaners, and then ascertained their expression profiles in the brains of different species from experimental data we generated in our companion studies (76, 77) and from the zebra finch brain expression database (78). Taking a conservative approach, we only examined well-aligned genes and amino acid sites available for at least one species of all three vocal learning groups, which reduced the 8295 orthologous protein-coding gene set to 7909 genes. We found convergent accelerated evolution (increased dN/dS above background) for 227 genes (~2.9%) in vocal learners (120 for the two and 107 for the three independent gain hypotheses; **Table S28**). Of these 227, at least 73% (165) were expressed in the songbird brain (physically cloned mRNAs), and of these 92% (151) were expressed in adult song learning nuclei (much higher than the expected 60%; (77)), including ~20% (33) regulated by singing (much higher than the expected 10%). In addition, 41% (20% expected) of the song nuclei accelerated genes showed differential expression among song nuclei, and 0.7 to 9% (0.7 to 4.3% expected), depending on song nucleus, showed specialized expression compared to the surrounding brain regions (**Table S28**). Of the candidate genes, cadherin 4 (*CDH4*) showed convergent specialized regulation in robust nucleus of the archistriatum (RA) of all vocal learning birds and the human laryngeal motor cortex, convergent specialized expression in songbird Area X and the human speech activated striatum, differential regulation among all song nuclei, and it controls development of neural connections (**Table S28**) (76). More generally, gene ontology analyses of the subset of accelerated differentially-expressed song nuclei genes revealed 30 significant functionally enriched gene sets, which clustered into four major categories, including neural connectivity, brain development and neural metabolism (**Fig. S23**).

We also developed an approach that scans for single amino acid substitutions common to species that have a shared trait, controlling for phylogenetic relationships (see methods; similar to Wang and Jarvis, in preparation). Of the 7909 genes, we found 38 that had one to two amino acid substitutions uniquely present in vocal learners (**Table S31**). At least two thirds (66%) of the genes were expressed in the songbird brain, including in the song nuclei (58%); the latter were higher than the expected (20%). Four genes (*CCDC135, CHGB, COL6A3,* and *TSEN2*) were regulated by singing (*77*), and two (*GDPD4* and *KIAA1919*) showed convergent accelerated evolution in vocal learners from the above analyses above, including on the specific amino acid sites unique to vocal learners (**Table S31**). Several of these genes (*KCNS3, PTPRB, SACS*) had been previously shown, when mutated, to be associated with diseases that include speech-language disorders (*79*). There were several genes whose amino acid sites were completely conserved and thus invariant in all bird species examined (only 1 amino acid) except vocal learners. These included *B3GNT2*, a polylactosamine synthase linked to brain autoimmune diseases and schizophrenia (*80*), which was differentially down-regulated in the RA and HVC song nuclei (**Fig. 5E-G**) and synaptojanin 2 (*SYJN2*), a synaptic vesicle coating protein that regulates neural connectivity, and has SNPs in humans associated with longevity and cognitive functions (*81*).

To identify accelerated evolution in non-coding sequences in vocal learners, we scanned the genome alignment using phyloP, a method that tests conservation and acceleration based on a neutral evolution model (*82*) (**SOM text 4**). Here we used a more limited sampling of vocal non-learning species closely related to the vocal learners (**Table S32**), due to the relatively faster evolutionary rate of non-coding regions. We scanned the entire genome alignment and found 822 accelerated genomic elements uniquely shared by all three vocal learning groups (**Table S33**). These convergent elements were more skewed to intergenic regions in vocal learners relative to the background average accelerated elements across species (Fisher's exact test, $p<2.2e-16$) (**Fig. 5B**). 332 of these elements were associated with 278 genes (within 5'/3' 10kb of the nearest genes), of which a high proportion (76%) was expressed in the brain and almost all of those (94%, 198 genes) expressed in one or more song nuclei (**Table S34**). Even more so than the genes identified above with accelerated coding nucleotide changes, 51% (20% expected) of the genes with accelerated non-coding region changes showed differential expression among song nuclei, 20% were regulated by singing (10% expected) and 2 to 15% (0.7 to 4.3% expected) had specialized expression relative to the surrounding brain regions (**Table S34**). These candidates included (*i*) convergent acceleration in the 5' region non-coding region of the synaptotagmin-like 2 (*SYTL2*) gene (**Fig. 5C**), a brain specific synaptic protein, (*ii*) an intron of *GRIN2B*, a glutamate neurotransmitter receptor that shows convergent specialized expression in avian song learning nuclei (*83*) and possibly human speech regions (*84*), and is required for learning and memory, including for song learning in songbirds (*85*); (*iii*) the 5' flanking region of *FOXP1*, a gene which shows convergent specialized expression in the HVC of all three vocal learning bird groups (**Fig 5H-J**) (*86*) and functions with *FOXP2* in its requirement for birdsong and speech acquisition (*87*); and (*iv*) three additional genes (*TSHZ3*, *JAZF1* and *SH3RF1*) that we independently found (*76*) to show convergent differential regulation in the RA song nucleus across the three vocal learning groups (**Table S35, Fig. S26-S28**). Overall, these analyses show a 2- to 20-fold enrichment of accelerated evolution in regulatory regions of genes differentially regulated in vocal learning brain regions. In contrast, there was very little overlap (2.5%) of genes with convergent accelerated non-coding changes and convergent accelerated amino acid changes, indicating two independent targets of selections for convergent evolution. Further

analyses on the convergent vocal learner associated molecular changes are described in our companion studies (*76*, *77*, *88*).

In summary, this convergent analysis across 48 avian species identifies candidate genes, their specific amino coding acid sites, and their regulatory regions for further tests on their role in the mechanisms and evolution of the complex trait of vocal learning.

Evolution of ecologically relevant genes

The availability of genomes representing almost all major lineages of extant birds permits us to investigate candidate genes that underlie traits relevant to the diversity of avian ecologies. Although these analyses should be approached with caution given the phenotypic and ecological plasticity within major avian lineages, this set of genomes is sufficient to draw a number of important conclusions. Some of the most obvious of these traits include major skeletal and tissue changes associated with the capacity for powered flight, feeding modification such as loss of teeth, an extremely advanced visual system found in some lineages, and features related to their sexual and reproductive systems.

Evolution of the capacity for flight

Skeletal systems

The evolution of flight involved a series of adaptive changes at the morphological and molecular levels (*89*). A number of these modifications were clearly apparent within the genomic data. One of the key requirements for flight is a skeleton that is both strong and lightweight. In both birds and non-avian theropods (*90*) this evolved by the fusion and elimination of some bones and the pneumatization of the remaining ones (*91-93*). We investigated genes involved in bone function and found that of 89 genes involved in ossification (**Table S36**), 50 (~56%) showed evidence of positive selection in birds, almost twice as high as in mammals (29 genes, ~32%). For birds, most of these are involved in the regulation of remodeling and ossification associated processes, or bone development in general, and those with the highest values for global omega (> 0.5) were obtained for *AHSG*, Alpha-2-HS-glycoprotein, associated with bone mineral density (*94*) and *P2RX7*, P2X purinoceptor 7, associated with bone homeostasis (*95*). The variation in the extension of pneumatization in avian post-cranial bones has been associated with the variation in body size and foraging strategies (*96*). Therefore, positive selection observed in those genes may explain the variation in the levels of bone pneumatization in birds, since the genes involved in the process of maintaining trabeculae within bones likely depends on the intrinsic network of genes participating in bone reabsorption and mineralization. These results suggest that most structural differences in bone between birds and mammals may not occur during ontogenesis but rather as a result of bone remodeling and reabsorption (**Table S37**).

Pulmonary structure and function

A principal demand of the increased metabolism associated with homeothermy and powered flight is the capacity for frequent high levels of oxygen consumption, requiring an efficient gas exchange process during pulmonary ventilation. Due to functional integration of ventilation and locomotion, birds evolved a volume constant lung and a rigid trunk region, whereas mammals evolved a changing volume lung, often coupled to locomotory flexion of the lumbar region (*97*). In contrast to the pulmonary alveola of the mammalian lung, the avian lung has a honeycomb-like structure incorporating a flow-through system with small air capillaries (*98*). This difference in lung structure and function could be due to variation in the regulatory network of genes related

to lung development. Intriguingly, we found that at least five genes that function in mammalian lung development were completely lost in the avian ancestor. One of these genes is the latent TGF-β binding protein 3 (*LTBP3*), which plays an important role in the extra cellular matrix and modulates the secretion and activity of TGF-β. *In-vivo* functional experiments in mice revealed that the null mutation of *LTBP3* caused lung alveolar septation defects (*99*), suggesting that the absence of this gene in birds may have limited their potential to evolve a broncho-alveolar-like structure, as found uniquely in mammals (*97*). Moreover, this gene is also associated with ossification, and its null mutation in mice causes an osteopetrosis-like phenotype (*100*). Thus the loss of this gene in birds may have multiple roles in association with the evolution of the avian lung and postcranial skeletal pneumaticity.

Feathers

The evolution and subsequent morphological diversification of feathers have shaped avian physiology, locomotion, mate choice and ecological niches (*101*). Feathers are composed of alpha and beta keratins (*102*), the latter of which are structural proteins found only in the epidermal appendages of birds (comprising 90% of mature feathers) and other reptiles. We found that birds have a significantly lower number of α-keratins than reptiles and mammals (**Table S39;** $p < 0.001$; Mann-Whitney U-test, U=26.0). Although the α-keratins genes appear to have contracted in birds, β-keratins have expanded, with a total of 1623 complete and 1084 incomplete copies across all birds (**Fig. S31**). Of these, the feather β-keratin subfamily is avian specific and comprises 893 of the complete and 637 of the incomplete genes while the remaining avian β-keratins (claw, scale and keratinocyte β-keratin subfamilies) are orthologous to β-keratins found in the crocodiles and turtles (**Fig. S31**) (*103, 104*). Interestingly, the proportion of feather β-keratins to the total number of β-keratins is significantly lower for aquatic/semi-aquatic birds than for landbirds ($p = 0.011$; Mann-Whitney U-test, U=144.5), while the proportion of keratinocyte β-keratins is significantly higher for aquatic/semi-aquatic birds than landbirds ($p=0.001$; Mann-Whitney U-test, U=116.0) indicating that feather compositional adaptations, such as an increase in hydrophobicity to aid in thermal regulation, may have occured in species with an aquatic lifestyle. We found that the avian β-keratins form six significant phylogenetic clades across the 48 species, and all major avian lineages possess members from each phylogenetic clade. This result strongly indicates that the avian β-keratin diversity was present in the basal avian lineage. However, there are four species that are statistical outliers with the highest number of β-keratins (zebra finch, chicken, rock dove and budgerigar) (*105*) all of which represent domestic birds. Assembly differences can be excluded as the possible cause of this observation, as no lineage correlation is observed between gene copy number and assembly quality. While this observation is concordant with the hypothesis that domestication may increase the recombination rate (*106, 107*) at β-keratin loci, we note that (i) two other domestic species (domestic turkey and Pekin duck) did not exhibit this trend, (ii) selection pressures on both zebra finch and budgerigar during domestication are unlikely to have been high, and (iii) three of the four above mentioned genomes (zebra finch, chicken, budgerigar) are of particularly high quality. Thus additional work is needed to explore this further.

The evolution of genes related to diet

Edentulism

The evolution of birds also had significant consequences with regards to their feeding strategies and diets, with changes happening at (among others) the structural, biochemical and sensory

levels. With regards to the former, one of the most immediately obvious avian-specific traits is edentulism, the phenotype of being toothless. Edentulism is thought to have evolved independently in multiple theropod lineages (*108*). However, although most phylogenetic analyses suggest that teeth were lost in the common ancestor of modern birds (*109, 110*), a few studies have recovered dentate taxa from the Mesozoic (*Hesperornis*, *Ichthyornis*) inside of crown Neornithes, suggesting that tooth loss could have occurred independently in different clades (*111*). To address this point, we scanned the avian genomes for molecular fossils of tooth-specific genes that function in dentin and enamel formation. We recovered remnants of these genes in all species examined (**Table S42**). Frameshift mutations and whole exon deletions were widespread in all investigated tooth genes. The vast majority of debilitating mutations were not shared, but importantly all species examined shared unambiguous deletions in protein-coding exons of enamel-specific genes (*ENAM*, *AMEL*, *AMBN, MMP20*, *AMTN*) and one dentin-specific gene (*DSPP*). The shared pattern of pseudogenization across representatives of all living birds suggests that enamel and dentin production were both knocked out in their common ancestor and supports the hypothesis that the common ancestor of modern birds lacked mineralized teeth (*109, 110*).

Diet related enzymes

At the biochemical level, avian-specific adaptation can be seen in several important pathways. One molecular adaptational feature of animal diet is the diversity in intracellular targeting of the glyoxylate detoxifying enzyme AGT that arises according to the intracellular origin of glycolate (*112*). The main source of the glycolate precursor (thus target of AGT) is the mitochondria in carnivores, the peroxisome in herbivores, and both organelles in omnivores (*113*). Variation in AGT targeting is controlled by differential expression patterns of the mitochondrial targeting sequence (MTS) that is located at the N-terminal of AGT (*113*). Because birds have evolved a wide range of dietary diversification, AGT represents an excellent candidate for study. We recovered complete AGT genes from 22 of the avian genomes (**Table S43**), of which five exhibit pseudogenized forms in their MTS region (**Fig. 6B, Fig S31**). These species included all three oscine passerine birds (the omnivorous American crow [*Corvus brachyrhynchos*] and medium ground-finch, and the herbivorous zebra finch), the nectar feeding Anna's hummingbird (*Calytpe anna*) within the order Apodiformes, and the insectivorous Cuckoo-roller (*Leptosomus discolor*) within the order Leptosomiformes. The loss of MTS function in three unrelated avian orders is consistent with multiple independent dietary transitions during avian evolution. Detection of positively selected amino acids at 137 Q (dN/dS = 2.153) and 378 R (dN/dS = 2.153) in all birds provided additional support for diet-related adaptation in AGT (positions referring to the human AGT; posterior probability > 99%; $P < 0.0001$).

A second example of dietary pathway evolution relates to vitamin C, an important nutrient cofactor functioning in a range of essential metabolic reactions in animals (*114*). Organisms that cannot synthesize vitamin C face considerable health problems (e.g. scurvy in humans) unless sufficient quantities can be obtained through diet. Loss of the ability to synthesize vitamin C has been to date observed in a limited number of species, including humans (*115*), Guinea pigs (*116*) and some bats (*117*). All species that do not synthesize vitamin C exhibit a pseudogenized gene for L-gulonolactone oxidase (*GULO*), an enzyme that is essential for catalyzing the last step of the vitamin C *de novo* synthesis pathway (*118*). Although a previous study noted *GULO* enzyme activity was lacking in some Passeriformes birds (*119*), no molecular evidence existed to support this. Genomic mining of the avian genomes revealed *GULO* pseudogenization has occurred in

two of the three oscine passerine birds (the medium ground-finch, the zebra finch) and the suboscine golden-collared manakin. In contrast, an intact form of *GULO* was recovered from all other genomes, including the third oscine species sequenced (American crow) and the basal passerine rifleman (*Acanthisitta chloris*) (**Table S44**). Similar to previous observations in mammals (*120*), this pseudogenization was caused by the loss of different exons and lethal mutations in the gene. Reconstruction of the molecular evolutionary history of *GULO* among amniotes, using a free-ratio model in PAML4.7 (*121*), shows purifying selection has dominated the evolution of *GULO*, from the ancestral node of amniotes (dN/dS = 0.096) to ancestral birds (dN/dS = 0.133) and mammals (dN/dS = 0.355), suggesting conservation of the ability to synthesize vitamin C both before and after avian divergence. However, two passerine birds with intact forms of the gene (American crow and rifleman), exhibited nonsynonymous changes at one order of magnitude higher than the average (**Fig. S34**), a sign of harmful mutations in the *GULO* protein (*117*). Together with the pseudogenization in the other three passerine birds, this suggests a relaxation in the evolutionary constraints on *GULO* in passerine birds.

Rhodosin/opsins and Vision

Birds exhibit what is possibly the most advanced vertebrate visual system, with a highly developed ability to distinguish colors over a wide range of wavelengths. In contrast to mammals, which have relatively few photoreceptor classes, almost all birds studied to date have retained an ancestral, tetrapod set of cones (*122*) that is hypothesized to play an essential role in a wide range of behaviors, such as reproduction and feeding (*123*). In most of the avian genomes, we detected higher numbers of opsin genes than in mammalian genomes; the mammalian genomes were all missing *OPN4x* (*124*), *Rh2* and either *OPN1sw1* (Monotremata) or *OPN1sw2* (Theria). Most avian genomes contained four classes of conopsin genes, supporting the hypothesis that avian vision is based on a tetrachromatic system. Penguins however were an exception, with both species sequenced exhibiting only three classes, and thus trichromatic, in line with findings based on retinal examination (*125*). This difference is most likely due to their aquatic lifestyle, and is consistent with observations of marine mammals that also appear to have lost one, or even both, cone pigment(s) (*122*). The *OPN1Rh2* gene in the two penguin species lacks amino acid K296 that binds the opsin to the chromophore to form a functional pigment. The high omega values (2.7 for emperor penguin and 2.2 for the Adelie penguin) are consistent with relaxed or positive selection. For most other species, the conopsin genes were under stronger stabilizing selection in birds than in mammals. However, signs of strong positive selection were detected in the branch leading to the passerine group Passerida (represented by the medium ground-finch and zebra finch**; Fig. S36),** which corroborates that the shift from violet sensitive SWS1 cones to ultraviolet sensitive cones in this clade (*126*) was adaptive. Optimal color discrimination is achieved when cone classes are evenly distributed over the spectrum (*127*). Therefore, as suggested by the omega dN/dS values, less variation in spectral sensitivity is tolerated in birds than in mammals due to their higher number of cone classes, which increases the risk of overlapping sensitivities. Besides two transmembrane regions (no. II and VII) encompassing previously identified spectral tuning amino acids in the SWS1 conopsin, we found markedly positive selection in region IV, strongly suggesting that there is one or more unknown amino acids sites that apparently are important to the spectral tuning of this UV-sensitive cone (**Fig. S38**).

Sex-related and reproductive traits

Reproduction related genes

Unlike other reptiles, almost all birds develop only a single functional ovary, on the left side (*128*), as a result of the evolutionary loss of the right ovary during the transition from non-avian theropods to birds (*129*). It has been hypothesized that this loss represents an adaptation to reduce weight during flight (*129*), and may be the consequence of loss of genes involved in ovary development. We found that two genes, *MMP19* and *AKR1C3*, with known association to ovary function are absent in all birds. *MMP19*, a matrix metalloprotease gene, functions during the follicular growth and the ovulation process (*130*), while the enzyme *AKR1C3* catalyzes the conversion of androstenedione to testosterone and has been associated with polycystic ovary syndrome (*131*).

We also performed analyses on a range of other genes related to reproduction, under the hypothesis that some of them may have been direct targets of the sophisticated morphological and behavioral adaptations related to sexual selection in birds. Previous analyses of these genes in *Drosophila*, humans and marine invertebrates have generally shown that they evolve faster than genes unassociated with reproduction (*66*). Specifically we chose 194 genes that may be involved in spermatogenesis, and 30 genes involved in oogenesis as defined by Gene Ontology annotation. We found that 19 out of 46 species show significantly ($p<0.05$; Wilcoxon test) accelerated evolution (measured by lineage-specific dN/dS ratio) of spermatogenesis genes relative to the genomic background. In contrast, only two species (carmine bee-eater [*Merops nubicoides*] and Pekin duck) showed significantly accelerated evolution in oogenesis genes. These results suggest that male birds are the dominant targets of sexual selection, which drives rapid evolution of spermatogenesis genes via sperm competition (*132*).

Plumage color

We investigated the genomics of plumage color, a behaviorally important trait. Besides the role of camouflage, plumage color has been recognized as a classic example of sexual selection since Darwin (*133*). Male birds have frequently evolved extravagant plumage color in response to both male-male competition and female choice (*134, 135*), resulting in remarkable sexual dichromatism. Analysis of 15 genes that have previously been implicated in influencing avian plumage colors (*136-140*) demonstrated evolutionary rates that were significantly faster than the genomic average in 8 of 46 avian lineages, as measured by their branch-specific dN/dS ratios (**Table S48**). This pattern suggests these genes are evolving under adaptive evolution. *MC1R,* the melanocortin 1 receptor, whose evolutionary rate has previously been shown to be positively correlated with galliform plumage dichromatism (*137*), also showed a positive correlation (coefficient=0.93) with plumage discriminability. This relationship is, however, statistically insignificant (P=0.37), perhaps due to the small sample size (discriminability data of plumage color exists for only 6 species). Carotenoids are responsible for the bright yellow and red pigments that underlie some of the most conspicuous coloration patterns in vertebrates. Unlike melanins, carotenoids can be only acquired through diet and there are potential trade-offs between their roles in coloration and other physiological conditions. Avian carotenoids preferentially accumulate in certain parts of integument due to differential deposition, due to the action of genes such as *GSTA2* that is specifically expressed in the plumage and bill of some passerine, and is involved in the binding and deposition of carotenoids. We found that the evolutionary rate of *GSTA2* is negatively correlated with the plumage dichromatism ($p<0.05$; coefficient=-0.28, **Fig. 6D**). A second gene *SLC24A4*, which is highly associated with hair color in humans, also showed a negative correlation ($p<0.05$; coefficient=-0.11, **Fig. 6D**). This

suggests that diversifying and stabilizing selection are driving the evolution of plumage color genes.

Conclusions

The advent and rapid development of next-generation sequencing techniques has revolutionized genomic approaches to biology. These new technologies are increasingly being applied beyond the level of individuals to populations, and from model to non-model species for which little or no prior genomic data existed, and perhaps most excitingly, from single to broad taxa sampling. These developments allow us to address many longstanding questions and make new discoveries at unprecedented phylogenomic scales. As we show in this project, such studies have extraordinary advantages in the ability to produce more accurate annotations for functional genomic elements from conservation profiles across the entire avian phylogeny and reveal the evolutionary dynamics for the whole genome from a phylogenetic perspective.

The avian genome is characterized by a number of features that remain remarkably constant across the entire clade. The small genome size with fragmented micro-chromosomes and reduced repeat transposon activity, in dramatic contrast to other vertebrates, has been a static feature in the avian clade for >150 million years. Avian genomes consistently contain fewer genes, about 70% of the number observed in human genome, and with one detected exception (downy woodpecker) an extremely reduced fraction of repeat elements. Thus the ancestral avian lineage has uniquely lost a large number of genes, by means of large segmental deletions, following their divergence from other extant reptiles. As previously hypothesized, these large genomic sequence depletions appear linked to a second defining feature of avian genomes—the putatively ancestral fission of macrochromosomes into a relatively large number of microchromosomes.

Conservation at the avian genomic level, both with regards to sequence, synteny and chromosomal structure, is remarkable in light of their rapid historical radiation into what has become the rich biodiversity of extent birds. This is considerably different from the evolution of mammalian genomes, which while experiencing a rapid radiation at a similar time, today displays a richness of genome shuffling and variation (*141*). By comparing the genomes of 48 birds that are constrained within a largely resolved phylogeny, we revealed millions of highly constrained elements comprising 7.5% of avian genomes. This evolutionary profiling of genomes across >100 million years (*5*) enables their interpretation in a functional genomic context that has been missed in previous genomic studies that were restricted to fewer taxa.

The generation of genome sequences for taxa that are distributed across the full avian phylogeny also enabled investigation into some of the putative genetic bases for the rich biodiversity of the avian clade. Various selection constraints function on certain categories of genes in different avian lineages. Convergent evolution is another mechanism shaping the evolution of protein coding genes and establishing similar morphological or behavioral features in distant species over the phylogeny. For example, we revealed that several genes displaying convergent evolutionary signals among vocal-learning species that may be associated with the development of vocal learning ability. There is also considerable variation in other specific gene families that correspond to avian traits and environmental adaptation among avian lineages.

In summary, we present the largest genomic overview of a vertebrate class conducted to date. The data and analyses presented here open a new window on the evolution, diversification, and ecological adaptation of tetrapod vertebrates and offers a phylogenomic perspective that helps bridge the chasm between micro- and macro-evolution. Naturally given the wealth of potential

information available in the dataset, our analyses can represent no more than pilot investigations of the multitude of questions that could be asked. We therefore believe the dataset will be an important resource for further studies that aim to expand on avian evolutionary genomics.

Materials and Methods

Sequencing and assembly

All new avian genomes except budgerigar and bald eagle were sequenced at BGI using the Illumina HiSeq platforms. The bald eagle was sequenced at WUSTL using the Illumina HiSeq platform. The budgerigar genome was generated from a combination of Roche/454 and Illumina HiSeq reads. For 20 species, we produced high (>50X) coverage sequences from multiple libraries with a gradient of insert sizes. For the remaining 25 species we generated low (~30X) coverage data from two insert-size libraries after performing several filtering steps on the raw reads. We used SOAPdenovo (*142*) to do *de novo* assembly for each bird species.

Annotation

We used RepeatMasker (*143*) to do the homology-based repeat annotation and used RepeatModeler (*144*) to do *de novo* repeat annotations. For the annotation of the protein-coding genes, we used Genewise (*145*) to do annotation based on a carefully selected reference gene set. To obtain an improved annotation for downstream analyses we also re-annotated some genes of chicken and zebra finch.

Identification of large segmental deletions

Four species, common ostrich (representing the modern birds), American alligator, green sea turtle and green anole were used to identify lost syntenic blocks in the avian ancestors. We identified the ortholog pairs between green anole and each species separately, and merged all ortholog pairs into a large orthologous set for further analysis. Using overlapping sliding window method (5-genes windows and a step size of 1 gene), we calculated the ratio of lost ortholog pairs (LOP for short, genes present in green anole but absent in other species) for green anole and each of other species. We chose the windows with ≥0.8 LOP in ostrich and < 0.5 LOP in at least one of the two non-avian reptiles as candidate segmental deletions.

Identification of homologous synteny blocks and evolutionary breakpoint regions

The genome alignments generated using the Satsuma Synteny program (*146*) were cleaned of overlapping and non-syntenic matches. Finally the homologous synteny blocks (HSBs) were defined using the SyntenyTracker (*147*). HSBs were identified using three sets of parameters that allowed the detection of rearrangements that are ≥ 300 Kbp, ≥ 100Kbp or ≥ 50Kbp in the chicken genome. Evolutionary breakpoint regions (EBRs) were identified as the intervals delimited by two adjacent HSBs on the same chicken chromosome. For the genomes lacking complete chromosome assemblies we considered only those EBRs that were found within scaffolds.

Estimation of substitution rates

4-fold degenerate sites (4D sites) are generally considered to be neutrally evolving. We extracted the 4D site alignments from the whole-genome alignments of 48 birds, 3 reptiles, 18 mammals and one amphibian. The phyloFit program (*82*) was used to estimate the phylogenetic tree, with a known tree topology (TENT tree from (*5*)) as an input tree.

Identification of highly conserved elements

PhastCons (*60*) is a widely-used program to identify highly conserved elements (HCEs) based on

the multiple-genome sequence alignments. To compare the genomic conservation between birds and mammals, we also ran PhastCons to predict the HCEs in the 18 placental mammal genome alignment downloaded from UCSC FTP. To make the results more comparable, the 'rho' value, which is the ratio between conserved and non-conserved models in PhastCons was set to the same value as that in birds (rho=0.2506).

dN/dS analysis of birds and mammals

To compare dN/dS ratios between birds and mammals, we ran PAML one-ratio branch model to obtain the dN/dS ratios for each avian/mammalian ortholog. We used Wilcoxon signed-rank test to test whether the dN/dS ratios of the genes of a certain GO are significantly larger in birds or in mammals. To identify the fast-evolving genes in each of the three major avian clades, we ran a one-ratio branch model, which estimated one identical dN/dS for all branches, and a three-ratio branch model, which estimated three different dN/dS for Palaeognathae, Galloanserae and Neoaves. A likelihood ratio test that compared the one-ratio branch model and the three-ratio branch model was performed to identify genes with different dN/dS in the three avian clades.

Analyses on specific gene families

For analyses of a specific group of genes (i.e, globin genes, opsin genes), we usually re-annotated or manually curated the genes to obtain better gene models. PAML (*121*) was used in many sections to estimate the evolutionary rates of specific gene families, but the models used in different sections varied due to specific objectives. More details can be found in the full SOM.

References and Notes:

1. F. Gill, D. Donsker, *IOC World Bird List (v 3.5)*. (2013).
2. L. M. Chiappe, L. M. Witmer, *Mesozoic birds: above the heads of dinosaurs*. (University of California Press, Berkeley, Calif., 2002).
3. G. Dyke, G. W. Kaiser, *Living dinosaurs the evolutionary history of modern birds*. (Wiley-Blackwell, Chichester, West Sussex; Hoboken, NJ, 2011).
4. A. Feduccia, 'Big bang' for tertiary birds? *Trends in Ecology & Evolution* **18**, 172-176 (2003).
5. E. D. Jarvis *et al.*, Avian phylogenomics paper. *submitted*, (2014).
6. B. G. Holt *et al.*, An Update of Wallace's Zoogeographic Regions of the World. *Science* **339**, 74-78 (2013).
7. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, The global diversity of birds in space and time. *Nature* **491**, 444-448 (2012).
8. H. Zeigler, P. E. Marler, in *Behavioral Neurobiology of Birdsong, Dec, 2002, Hunter College, City University of New York, New York, NY, US; This volume is the result of the aforementioned conference which was one of an annual symposium series sponsored by the Hunter College Gene Center.* (2004).
9. A. J. Stattersfield, M. J. Crosby, A. J. Long, D. C. Wege, *Endemic Bird Areas of the World: Priorities for Conservation*. BirdLife Conservation (BirdLife International, 1998).
10. D. J. Alexander, A review of avian influenza in different bird species. *Veterinary Microbiology* **74**, 3-13 (2000).
11. L. W. Hillier *et al.*, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
12. R. A. Dalloul *et al.*, Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. *PLoS biology* **8**, (2010).
13. W. C. Warren *et al.*, The genome of a songbird. *Nature* **464**, 757-762 (2010).
14. E. Dickinson, J. Remsen, *The Howard and Moore Complete Checklist of the Birds of the World*. (Aves Press, Eastbourne, UK, 2013).
15. Y. Huang *et al.*, The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature genetics* **45**, 776-783 (2013).
16. X. Zhan *et al.*, Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature genetics* **45**, 563-566 (2013).
17. M. D. Shapiro *et al.*, Genomic diversity and evolution of the head crest in the rock pigeon. *Science (New York, N.Y.)* **339**, 1063-1067 (2013).
18. S. Li *et al.*, Rapid decline of genomic diversity in restored population of almost-extinct Crested Ibis. *submitted*, (2014).
19. J. Li *et al.*, The genomes of two Antarctic penguins reveal adaptations to the cold aquatic environment. *submitted*, (2014).
20. G. Ganapathy *et al.*, De novo high-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience Database*, (2013).
21. J. A. St John *et al.*, Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome biology* **13**, 415 (2012).
22. R. E. Green *et al.*, Complete genome sequences for three crocodilians reveal 'genomic stasis' for a major archosaurian lineage. *submitted*, (2014).

23. Z. Wang *et al.*, The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet* **45**, 701-706 (2013).
24. J. Alföldi *et al.*, The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587-591 (2011).
25. H. Ellegren *et al.*, The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756-760 (2012).
26. Y. Qu *et al.*, Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nature communications* **4**, 2071 (2013).
27. T. K. Oleksyk *et al.*, A locally funded Puerto Rican parrot (Amazona vittata) genome sequencing project increases avian data and advances young researcher education. *Gigascience* **1**, 14 (2012).
28. T. R. Gregory, in *The Animal Genome Size Database*. (2005).
29. D. K. Griffin, L. B. W. Robertson, H. G. Tempest, B. M. Skinner, The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and genome research* **117**, 64-77 (2007).
30. C. L. Organ, S. V. Edwards, in *Living Dinosaurs: The Evolutionary History of Modern Birds*. (Wiley-Blackwell, 2011), pp. 325-337.
31. Q. Zhou *et al.*, Atlas of Avian Sex Chromosome Evolution. *submitted*, (2014).
32. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
33. A. L. Hughes, M. K. Hughes, Small genomes for better flyers. *Nature* **377**, 391 (1995).
34. M. Rho *et al.*, Independent mammalian genome contractions following the KT boundary. *Genome biology and evolution* **1**, 2-12 (2009).
35. S. Morand, R. E. Ricklefs, Genome size, longevity and development time in birds. *Trends in genetics: TIG* **17**, 567-568 (2001).
36. E. Waltari, S. V. Edwards, Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *The American naturalist* **160**, 539-552 (2002).
37. M. G. Kidwell, Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49-63 (2002).
38. C. Feschotte, E. J. Pritham, DNA transposons and the evolution of eukaryotic genomes. *Annual review of genetics* **41**, 331-368 (2007).
39. M. Lynch, J. S. Conery, The evolutionary demography of duplicate genes. *Journal of structural and functional genomics* **3**, 35-44 (2003).
40. N. Sela, E. Kim, G. Ast, The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome biology* **11**, R59 (2010).
41. A. Bohne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **16**, 203-215 (2008).
42. A. L. Hughes, R. Friedman, Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Molecular biology and evolution* **25**, 2681-2688 (2008).
43. M. K. Fujita, S. V. Edwards, C. P. Ponting, The Anolis lizard genome: an amniote genome without isochores. *Genome Biol Evol* **3**, 974-984 (2011).
44. S. Ohno, *Evolution by gene duplication*. (George Alien & Unwin Ltd., London, 1970).

45. T. R. Gregory, A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class aves. *Evolution; international journal of organic evolution* **56**, 121-130 (2002).
46. K. R. Bradnam *et al.*, Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2**, 10 (2013).
47. J. Kim *et al.*, Reference-assisted chromosome assembly. *Proc Natl Acad Sci U S A* **110**, 1785-1790 (2013).
48. P. Pevzner, G. Tesler, Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7672-7677 (2003).
49. B. M. Skinner, D. K. Griffin, Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**, 37-41 (2012).
50. M. A. Groenen *et al.*, Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398 (2012).
51. F. G. Hoffmann, J. C. Opazo, J. F. Storz, Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. *Proceedings of the National Academy of Sciences* **107**, 14274-14279 (2010).
52. F. G. Hoffmann, J. C. Opazo, J. F. Storz, Whole-Genome Duplications Spurred the Functional Diversification of the Globin Gene Superfamily in Vertebrates. *Molecular biology and evolution* **29**, 303-312 (2012).
53. J. F. Storz, J. C. Opazo, F. G. Hoffmann, Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Molecular phylogenetics and evolution* **66**, 469-478 (2013).
54. F. G. Hoffmann, J. C. Opazo, J. F. Storz, Rapid rates of lineage-specific gene duplication and deletion in the alpha-globin gene family. *Molecular biology and evolution* **25**, 591-602 (2008).
55. J. C. Opazo, F. G. Hoffmann, J. F. Storz, Differential loss of embryonic globin genes during the radiation of placental mammals. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12950-12955 (2008).
56. M. T. Grispo *et al.*, Gene duplication and the evolution of hemoglobin isoform differentiation in birds. *The Journal of biological chemistry* **287**, 37647-37658 (2012).
57. C. F. Baer, M. M. Miyamoto, D. R. Denver, Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics* **8**, 619-631 (2007).
58. A. P. Martin, S. R. Palumbi, Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences* **90**, 4087-4091 (1993).
59. K. Lindblad-Toh *et al.*, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-482 (2011).
60. A. Siepel *et al.*, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* **15**, 1034-1050 (2005).
61. M. Blanchette *et al.*, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-715 (2004).
62. B. Nabholz, S. Glémin, N. Galtier, The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology* **9**, 54 (2009).

63. I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, D. P. Bartel, Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537-1550 (2011).
64. W. F. Lathrop, E. P. Carmichael, D. G. Myles, P. Primakoff, cDNA cloning reveals the molecular structure of a sperm surface protein, PH-20, involved in sperm-egg adhesion and the wide distribution of its gene among mammals. *The Journal of cell biology* **111**, 2939-2949 (1990).
65. Y. Zheng, X. Deng, P. A. Martin-DeLeon, Lack of Sharing of Spam1 (Ph-20) among Mouse Spermatids and Transmission Ratio Distortion. *Biology of Reproduction* **64**, 1730-1738 (2001).
66. W. J. Swanson, V. D. Vacquier, The rapid evolution of reproductive proteins. *Nature reviews. Genetics* **3**, 137-144 (2002).
67. G. A. Rappold, The pseudoautosomal regions of the human sex chromosomes. *Human Genetics* **92**, 315-324 (1993).
68. J. A. Marshall Graves, S. Shetty, Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *The Journal of experimental zoology* **290**, 449-462 (2001).
69. L. A. Pennacchio *et al.*, In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499-502 (2006).
70. E. P. Consortium *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
71. M. T. Cobourne, The complex genetics of cleft lip and palate. *The European Journal of Orthodontics* **26**, 7-16 (2004).
72. J. E. Mank, E. Axelsson, H. Ellegren, Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Research* **17**, 618-624 (2007).
73. E. Axelsson, M. T. Webster, N. G. C. Smith, D. W. Burt, H. Ellegren, Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Research* **15**, 120-125 (2005).
74. A. Suh *et al.*, Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nature communications* **2**, 443 (2011).
75. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* **1016**, 749-777 (2004).
76. A. Pfenning *et al.*, Convergent gene expression specializations in song and speech brain regions of song learning birds and humans. *submitted*, (2014).
77. O. Whitney *et al.*, The singing genome: Core and region enriched gene expression define behaviorally regulated gene networks. *submitted*, (2014).
78. H. J. Karten *et al.*, Digital atlas of the zebra finch (Taeniopygia guttata) brain: a high-resolution photo atlas. *The Journal of comparative neurology* **521**, 3702-3715 (2013).
79. P. Pinel *et al.*, Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 817-825 (2012).
80. A. R. Sanders *et al.*, Transcriptome study of differential expression in schizophrenia. *Human molecular genetics* **22**, 5001-5014 (2013).
81. M. Luciano *et al.*, Longevity candidate genes and their association with personality traits in the elderly. *American journal of medical genetics. Part B, Neuropsychiatric genetics :*

the official publication of the International Society of Psychiatric Genetics **159B**, 192-200 (2012).

82. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in Bioinformatics* **12**, 41-51 (2011).
83. K. Wada, H. Sakaguchi, E. D. Jarvis, M. Hagiwara, Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *The Journal of comparative neurology* **476**, 44-64 (2004).
84. K. Amunts *et al.*, Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol* **8**, (2010).
85. M. E. Basham, E. J. Nordeen, K. W. Nordeen, Blockade of NMDA receptors in the anterior forebrain impairs sensory acquisition in the zebra finch (Poephila guttata). *Neurobiology of learning and memory* **66**, 295-304 (1996).
86. E. D. Jarvis *et al.*, Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns. *The Journal of comparative neurology* **521**, 3614-3665 (2013).
87. S. E. Fisher, C. Scharff, FOXP2 as a molecular window into speech and language. *Trends in genetics : TIG* **25**, 166-177 (2009).
88. C. Lee *et al.*, A comparative genetics approach of avian genomes gives insight to bird vocalization. *submitted*, (2014).
89. A. Hedenström, Aerodynamics, evolution and ecology of avian flight. *Trends in Ecology & Evolution* **17**, 415-422 (2002).
90. J. R. Codd, P. L. Manning, M. A. Norell, S. F. Perry, Avian-like breathing mechanics in maniraptoran dinosaurs. *Proceedings. Biological sciences / The Royal Society* **275**, 157-161 (2008).
91. J. Cubo, A. Casinos, Incidence and mechanical significance of pneumatization in the long bones of birds. *Zoological Journal of the Linnean Society* **130**, 499–510 (2000).
92. T. A. Dececchi, H. C. E. Larsson, Patristic evolutionary rates suggest a punctuated pattern in forelimb evolution before and after the origin of birds. *Paleobiology* **35**, 1-12 (2009).
93. E. R. Dumont, Bone density and the lightweight skeletons of birds. *Proceedings. Biological sciences / The Royal Society* **277**, 2193-2198 (2010).
94. Y.-J. Yang *et al.*, AHSG gene polymorphisms are associated with bone mineral density in Caucasian nuclear families. *European Journal of Epidemiology* **22**, 527-532 (2007).
95. R. Sluyter, L. Stokes, Significance of P2X7 receptor variants to human health and disease. *Recent patents on DNA & gene sequences* **5**, 41-54 (2011).
96. S. C. Gutzwiller, A. Su, P. M. O'Connor, Postcranial Pneumaticity and Bone Structure in Two Clades of Neognath Birds. *The Anatomical Record* **296**, 867–876 (2013).
97. H. R. Duncker, Vertebrate lungs: structure, topography and mechanics. A comparative perspective of the progressive integration of respiratory system, locomotor apparatus and ontogenetic development. *Respiratory physiology & neurobiology* **144**, 111-124 (2004).
98. J. B. West, R. R. Watson, Z. Fu, Major differences in the pulmonary circulation between birds and mammals. *Respiratory physiology & neurobiology* **157**, 382-390 (2007).
99. C. Colarossi *et al.*, Lung alveolar septation defects in Ltbp-3-null mice. *The American journal of pathology* **167**, 419-428 (2005).
100. B. Dabovic *et al.*, Osteopetrosis-like phenotype in latent TGF-β binding protein 3 deficient mice. *Bone* **37**, 25-31 (2005).

101. F. Gill, *Ornithology*. (W.H. Freeman and Company, New York, 1995).
102. A. R. Haake, G. König, R. H. Sawyer, Avian feather development: relationships between morphogenesis and keratinization. *Developmental biology* **106**, 406-413 (1984).
103. M. J. Greenwold, R. H. Sawyer, Molecular evolution and expression of archosaurian beta-keratins: diversification and expansion of archosaurian beta-keratins and the origin of feather beta-keratins. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **320**, 393-405 (2013).
104. Y. I. Li, L. Kong, C. P. Ponting, W. Haerty, Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biol Evol* **5**, 923-933 (2013).
105. M. J. Greenwold *et al.*, The dynamic evolution of the alpha (α) and beta (β) keratin multigene families has accompanied the lineage-specific evolution of avian scales and feathers. *submitted*, (2014).
106. J. Ross-Ibarra, The evolution of recombination under domestication: a test of two hypotheses. *The American naturalist* **163**, 105-112 (2004).
107. A. Burt, G. Bell, Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326**, 803-805 (1987).
108. A. Louchart, L. Viriot, From snout to beak: the loss of teeth in birds. *Trends in Ecology & Evolution* **26**, 663-673 (2011).
109. B. C. Livezey, R. L. Zusi, Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological journal of the Linnean Society* **149**, 1-95 (2007).
110. J. Cracraft, Origin and Evolution of Continental Biotas: Speciation and Historical Congruence within the Australian Avifauna. *Evolution* **40**, 977 (1986).
111. J. Cracraft, Phylogenetic relationships and monophyly of loons, grebes, and hesperornithiform birds, with comments on the early history of birds. *Systematic Biology* **31**, 35–56 (1982).
112. G. M. Birdsey, J. Lewin, A. A. Cunningham, M. W. Bruford, C. J. Danpure, Differential enzyme targeting as an evolutionary adaptation to herbivory in carnivora. *Molecular biology and evolution* **21**, 632-646 (2004).
113. C. J. Danpure, Variable peroxisomal and mitochondrial targeting of alanine: glyoxylate aminotransferase in mammalian evolution and disease. *BioEssays: news and reviews in molecular, cellular and developmental biology* **19**, 317-326 (1997).
114. C. L. Linster, E. Van Schaftingen, Vitamin C. Biosynthesis, recycling and degradation in mammals. *The FEBS journal* **274**, 1-22 (2007).
115. M. Nishikimi, R. Fukuyama, S. Minoshima, N. Shimizu, K. Yagi, Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *The Journal of biological chemistry* **269**, 13685-13688 (1994).
116. M. Nishikimi, T. Kawai, K. Yagi, Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for L-ascorbic acid biosynthesis missing in this species. *The Journal of biological chemistry* **267**, 21967-21972 (1992).
117. J. Cui, Y.-H. Pan, Y. Zhang, G. Jones, S. Zhang, Progressive pseudogenization: vitamin C synthesis and its loss in bats. *Molecular biology and evolution* **28**, 1025-1031 (2011).
118. I. B. Chatterjee, Evolution and the biosynthesis of ascorbic acid. *Science* **182**, 1271-1272 (1973).

119. C. M. d. Rio, Can Passerines Synthesize Vitamin C? *The Auk* **114**, 513-516 (1997).
120. J. Cui, X. Yuan, L. Wang, G. Jones, S. Zhang, Recent loss of vitamin C biosynthesis ability in bats. *PloS one* **6**, e27114 (2011).
121. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular biology and evolution* **24**, 1586 -1591 (2007).
122. W. I. L. Davies, S. P. Collin, D. M. Hunt, Molecular ecology and adaptation of visual photopigments in craniates. *Molecular ecology* **21**, 3121-3158 (2012).
123. C. B. V. Carvalho, R. H. F. Macedo, J. A. Graves, Reproduction of Blue-black Grassquits in central Brazil. *Brazilian journal of biology = Revista brasleira de biologia* **67**, 275-281 (2007).
124. S. S. Pires *et al.*, Isolation and characterization of melanopsin (Opn4) from the Australian marsupial Sminthopsis crassicaudata (fat-tailed dunnart). *Proceedings. Biological sciences / The Royal Society* **274**, 2791-2799 (2007).
125. J. K. Bowmaker, G. R. Martin, Visual pigments and oil droplets in the penguin,Spheniscus humboldti. *Journal of Comparative Physiology A* **156**, 71-77 (1985).
126. A. Ödeen, O. Håstad, P. Alström, Evolution of ultraviolet vision in the largest avian radiation - the passerines. *BMC Evolutionary Biology* **11**, 313 (2011).
127. L. Chittka, Does bee color vision predate the evolution of flower color? *Naturwissenschaften* **83**, 136-138 (1996).
128. S. S. Guraya, *Ovarian follicles in reptiles and birds*. (Springer-Verlag, Berlin, Germany, 1989).
129. X. Zheng *et al.*, Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour. *Nature* **495**, 507-511 (2013).
130. M. Jo, T. E. Curry, Jr., Regulation of matrix metalloproteinase-19 messenger RNA expression in the rat ovary. *Biology of reproduction* **71**, 1796-1806 (2004).
131. M. O. Goodarzi, H. J. Antoine, R. Azziz, Genes for enzymes regulating dehydroepiandrosterone sulfonation are associated with levels of dehydroepiandrosterone sulfate in polycystic ovary syndrome. *The Journal of clinical endocrinology and metabolism* **92**, 2659-2664 (2007).
132. T. R. Birkhead, A. P. Møller, *Sperm competition and sexual selection*. (Academic Press, 1998).
133. C. Darwin, *The descent of man, and selection in relation to sex*. (J. Murray, London,, 1871).
134. M. Zuk, J. D. Ligon, R. Thornhill, Effects of experimental manipulation of male secondary sex characters on female mate preference in red jungle fowl. *Animal Behaviour* **44**, 999-1006 (1992).
135. C. Mateos, J. Carranza, The role of bright plumage in male–male interactions in the ring-necked pheasant. *Animal Behaviour* **54**, 1205-1214 (1997).
136. N. Walsh, J. Dale, K. J. McGraw, M. A. Pointer, N. I. Mundy, Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. *Proceedings. Biological sciences / The Royal Society* **279**, 58-66 (2012).
137. N. J. Nadeau, T. Burke, N. I. Mundy, Evolution of an avian pigmentation gene correlates with a measure of sexual selection. *P R Soc B* **274**, 1807-1813 (2007).

138. J. K. Hubbard, J. A. Uy, M. E. Hauber, H. E. Hoekstra, R. J. Safran, Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in genetics : TIG* **26**, 231-239 (2010).
139. A. Roulin, J. Mangels, K. Wakamatsu, T. Bachmann, Sexually dimorphic melanin-based colour polymorphism, feather melanin content, and wing feather structure in the barn owl (Tyto alba). *Biological Journal of the Linnean Society* **109**, 562–573 (2013).
140. J. J. Bull, *Evolution of sex determining mechanisms*. (The Benjamin/Cummings Publishing Company, Inc, 1983).
141. A. R. Quinlan, I. M. Hall, Characterizing complex structural variation in germline and somatic genomes. *Trends in genetics : TIG* **28**, 43-53 (2012).
142. R. Li *et al.*, De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**, 265-272 (2010).
143. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0. (http://www.repeatmasker.org). (1996).
144. A. F. A. Smit, R. Hubley, RepeatModeler Open-1.0. (http://www.repeatmasker.org). (2008).
145. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Research* **14**, 988-995 (2004).
146. M. G. Grabherr *et al.*, Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145-1151 (2010).
147. R. Donthu, H. A. Lewin, D. M. Larkin, SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC research notes* **2**, 148 (2009).
148. G. Feenders *et al.*, Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS One* **3**, e1768 (2008).

Acknowledgments: Genome assemblies, and annotations of avian genomes in this study are available on the avian phylogenomics website (http://phybirds.genomics.org.cn/). This study was supported by inner funding from BGI. G.Z. was supported by a Marie Curie International Incoming Fellowship grant (300837). M.T.P.G. was supported by a Danish National Research Foundation grant (DNRF94) and a Lundbeck Foundation grant (R52-A5062). E.D.J. was supported by funding from Howard Hughes Medical Institute and NIH Directors Pioneer Award DP1OD000448.

Figure 1. Phylogenetic relationship of the 48 avian genomes analyzed (taken from (*5*)), indicating their genome status. Percentage of repeat elements within each genome given on terminal branches.

Figure 2. (A) Average size of introns (red), exons (blue) and intergenic (yellow) regions within avian, reptilian and mammalian genomes. The blue bars represent the gene models. **(B)** Cumulative curve of small deletion events inferred within Reptilia. Most small deletions occurred prior to the origin of birds. The shade along the curve depicts the standard deviation of the deletion rates, estimated by searching the alignments with 1Mb sliding windows. **(C)** Synteny plot and large segmental deletions. The synteny plot shows a syntenic comparison between Green Anole Lizard chromosome 2 and multiple Chicken chromosomes. The colored bars and lines represent homologous blocks between two species. The black bars on the horizontal axis indicate the location of large avian-specific segmental deletions, which are enriched at the breakpoints of inter-chromosome rearrangements. The lower panel displays gene synteny of a case of a large segmental deletion in birds. The homologous genes annotated in each species are shown in small boxes. The color spectrum represents the identity of homologous genes against Green Anole Lizard genes. **(D)** Distribution of gene synteny percentages identified for species pairs of various divergence ages. Each dot represents the percentage of genes remained in synteny block in pairwise comparison between two avian or mammalian species. The box plot shows the overall distributions of the synteny percentages in birds and mammals. **(E)** Comparative structure of the α- and β-globin gene clusters in representative avian and mammalian taxa. These genes encode the α- and β-type subunits of tetrameric ($\alpha_2\beta_2$) hemoglobin isoforms that are expressed at different ontogenetic stages. In the case of the α-like globin genes, birds and mammals share orthologous copies of the α^D- and α^A-globin genes. Likewise, the avian π-globin and the mammalian ζ-globin genes (which are exclusively expressed in embryonic erythroid cells) are 1:1 orthologs. By contrast, the genes in the avian and mammalian β-globin gene clusters are derived from independent duplications of one or more β-like globin genes that were inherited from the common ancestor of tetrapod vertebrates (*51, 53*). Whereas the mammalian α- and β-globin gene clusters have undergone high rates of gene turnover due to lineage-specific duplications, deletions, and inactivations, the size and membership composition of the corresponding avian gene clusters have remained remarkably constant over a similar timescale of species diversification.

Figure 3. Evolutionary rate and selection constraints. **(A)** Substitution rate in each lineage was estimated by the comparison of 4-fold degenerate sites, in units of substitutions per site per million years. **(B)** Correlation between substitution rates and number of species within different avian orders. The divergence times were estimated in the avian phylogenomics project (*5*). **(C)** The density map for comparison of conservation levels between pan-avian and pan-mammalian genomes, based on the homologous genomic regions between birds and mammals. The conservation levels were quantified by PhastCons basewise conservation scores. **(D)** Highly conserved elements (HCEs) in avian genomes are categorized into two subsets; HCEs found in both mammalian and avian genomes and avian-specific HCEs. **(E)** The *MID1* gene contains abundant avian specific HCEs in upstream and downstream regulatory regions. Many regulatory motif elements are identified in these avian specific HCEs.

Figure 4. Selection constraints on genes. **(A)** Box plot for the distribution of dN/dS values of genes on avian macro-chromosomes, micro-chromosomes, and Z chromosomes. **(B)** Faster evolving GO categories in birds (upper) or mammals (lower). The red bar shows the p-value of significance, and the blue bar shows the number of genes in each GO.

Figure 5. Convergent molecular changes among vocal learning birds. (**A**) Brain diagram showing the specialized forebrain song learning system (yellow) that controls the production (HVC and RA) and acquisition (LMAN and Area X) of learned song (adapted from (*75*)). Grey arrows represent the connections between brain regions. Thick arrows (in red and blue), relative numbers of genes with increased or decreased expressed between song nuclei genes and with convergent accelerated coding sequences (left numbers of 66) or convergent target specific amino acid substitution sites (TAAS, right numbers of 6) in the four song nuclei of the zebra finch brain. Genes that are expressed in more than one song nucleus are counted multiple times. RA, robust nucleus of the arcopallium; LMAN, lateral magnocellular nucleus of the anterior neostraitum; Area X of the striatum; HVC, a letter based name. (**B**) Distribution of the vocal learner-specific accelerated elements, compared the background levels across 15 species in the sequence alignment. (**C**) Alignment of an accelerated element in the upstream region of the *Sytl2* gene between vocal learners and non-learners. Nucleotides are color-coded. (**D**) Levels of sequence conservation across the alignment predicted by phastCons on vocal learners and non-learners separately. (**E-G**) Decreased expression of B3GNT2 mRNA (brown dioxygenin probe signal) in the RA and HVC song nuclei of the zebra finch (Images derived from the ZEBRa database). (**H-J**) Convergent increased expression of FOXP1 mRNA (white) in the HVC analogue (arrows) in a songbird (H, zebra finch), hummingbird (I, Anna's hummingbird), and parrot (J, budgerigar). Red, Nissl stain of all cells in darkfield view; the songbird and hummingbird sections are sagital, the parrot is coronal. Images based on (*148*). A, Arcopallium; H, Hyperpallium; M, Mesopallium; N, Nidopallium; P, Pallidum; St, Striatum; VLN, vocal nucleus of lateral nidopallium; NLC, central nucleus of lateral nidopallium. Scale bars, 1mm for each species.

Figure 6. (A) Contraction of alpha-keratins in birds and the dynamic evolution of the feather and keratinocyte beta-keratins in aquatic birds, land birds and domesticated birds. **(B)** Pseudogenization events of two diet-related genes (*AGT* and *GULO*) along the avian phylogeny. **(C)** Density distribution of dN/dS values of the *OPN1sw1* gene for mammals (median=0.21) and birds (median=0.16). **(D)** dN/dS values of two plumage color related genes (*GSTA2* and *SLC24A4*) show negative correlation with the plumage dichromatism values (log transformation applied).