RESEARCH ARTICLE

Comparative genomics reveals insights into avian genome evolution and adaptation

Guojie Zhang,[1,2*†] Cai Li,[1,3*] Qiye Li,[1,3] Bo Li,[1] Denis M. Larkin,[4] Chul Lee,[5,6] Jay F. Storz,[7] Agostinho Antunes,[8,9] Matthew J. Greenwold,[10] Robert W. Meredith,[11] Anders Ödeen,[12] Jie Cui,[13,14] Qi Zhou,[15] Luohao Xu,[1,16] Hailin Pan,[1] Zongji Wang,[1,17] 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] Hao Yu,[1] Jinmin Lian,[1] Ping Wen,[1] Fang Zhang,[1] Hui Li,[1] Yongli Zeng,[1] Zijun Xiong,[1] Shiping Liu,[1,17] Long Zhou,[1] Zhiyong Huang,[1] Na An,[1] Jie Wang,[1,18] Qiumei Zheng,[1] Yingqi Xiong,[1] Guangbiao Wang,[1] Bo Wang,[1] Jingjing Wang,[1] Yu Fan,[19] Rute R. da Fonseca,[3] Alonzo Alfaro-Núñez,[3] Mikkel Schubert,[3] Ludovic Orlando,[3] Tobias Mourier,[3] Jason T. Howard,[20] Ganeshkumar Ganapathy,[20] Andreas Pfenning,[20] Osceola Whitney,[20] Miriam V. Rivas,[20] Erina Hara,[20] Julia Smith,[20] Marta Farré,[4] Jitendra Narayan,[21] Gancho Slavov,[21] Michael N Romanov,[22] Rui Borges,[8,9] João Paulo Machado,[8,23] Imran Khan,[8,9] Mark S. Springer,[24] John Gatesy,[24] Federico G. Hoffmann,[25,26] Juan C. Opazo,[27] Olle Håstad,[28] Roger H. Sawyer,[10] Heebal Kim,[5,6,29] Kyu-Won Kim,[5] Hyeon Jeong Kim,[6] Seoae Cho,[6] Ning Li,[30] Yinhua Huang,[30,31] Michael W. Bruford,[32] Xiangjiang Zhan,[32,33] Andrew Dixon,[34] Mads F. Bertelsen,[35] Elizabeth Derryberry,[36,37] Wesley Warren,[38] Richard K Wilson,[38] Shengbin Li,[39] David A. Ray,[26‡] Richard E. Green,[40] Stephen J. O'Brien,[41,42] Darren Griffin,[22] Warren E. Johnson,[43] David Haussler,[40] Oliver A. Ryder,[44] Eske Willerslev,[3] Gary R. Graves,[45,46] Per Alström,[47,48] Jon Fjeldså,[46] David P. Mindell,[49] Scott V. Edwards,[50] Edward L. Braun,[51] Carsten Rahbek,[46,52] David W. Burt,[53] Peter Houde,[54] Yong Zhang,[1] Huanming Yang,[1,55] Jian Wang,[1] Avian Genome Consortium,§ Erich D. Jarvis,[20†] M. Thomas P. Gilbert,[3,56†] Jun Wang[1,55,57,58,59†]

Birds are the most species-rich class of tetrapod vertebrates and have wide relevance across many research fields. We explored bird macroevolution using full genomes from 48 avian species representing all major extant clades. The avian genome is principally characterized by its constrained size, which predominantly arose because of lineage-specific erosion of repetitive elements, large segmental deletions, and gene loss. Avian genomes furthermore show a remarkably high degree of evolutionary stasis at the levels of nucleotide sequence, gene synteny, and chromosomal structure. Despite this pattern of conservation, we detected many non-neutral evolutionary changes in protein-coding genes and noncoding regions. These analyses reveal that pan-avian genomic diversity covaries with adaptations to different lifestyles and convergent evolution of traits.

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 during the Jurassic and are the only extant descendants of dinosaurs (*2*, *3*). The earliest diversification of extant birds (Neornithes) occurred during the Cretaceous period. However, the Neoaves, the most diverse avian clade, later underwent a rapid global expansion and radiation after a mass extinction event ~66 million years ago near the Cretaceous-Paleogene (K-Pg) boundary (*4*, *5*). As a result, the extant avian lineages exhibit extremely diverse morphologies and rates of diversification. Given the nearly complete global inventory of avian species, and the immense collected amount of distributional and biological data, birds are widely used as models for investigating evolutionary and ecological questions (*6*, *7*). The chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and pigeon (rock dove) (*Columba livia*) are also important model organisms in disciplines such as neuroscience and developmental biology (*8*). In addition, birds are widely used for global conservation priorities (*9*) and are culturally important to human societies. A number of avian species have been domesticated and are economically important. Farmed and wild water birds are key players in the global spread of pathogens, such as avian influenza virus (*10*).

Despite the need to better understand avian genomics, annotated avian genomic data was previously available for only a few species: the domestic chicken, domestic turkey (*Meleagris gallopavo*) and zebra finch (*11*–*13*), together with a few others only published recently (*14*–*16*). To build an understanding of the genetic complexity of birds and to investigate links between their genomic variation and phenotypic diversity, we collected and compared genome sequences of these and other avian species (48 species total), representing all 32 neognath and two of the five palaeognath orders (Fig. 1) (*17*), thus representing nearly all of the major clades of living birds (*5*).

Results

Sequencing, assembly, and annotation

We used a whole-genome shotgun strategy to generate genome sequences of 45 new avian species (*18*), including two species representing two orders within the infraclass Paleognathae [common ostrich (*Struthio camelus*) and white-throated tinamou (*Tinamus guttatus*)], the other order within Galloanserae [Peking duck (*Anas platyrhynchos*)], and 41 species representing 30 neoavian orders (table S1) (*19*). In combination with the three previously published avian genomes (*11*–*13*), the genome assemblies cover 92% (34 of 37) of all avian orders (the three missing orders belong to the Paleognathae) (*17*). With the exception of the budgerigar (*Melopsittacus undulatus*), which was assembled through a multiplatform (Illumina/GS-FLX/PacBio) approach (*20*), all other new genomes were sequenced and assembled with Illumina (San Diego, CA) short reads (Fig. 1) (*18*). For 20 species, we produced high (>50×) coverage sequences from multiple libraries, with a gradient of insert sizes and built full-genome assemblies. For the remaining 25 species, we generated low (~30×) coverage data from two insert-size libraries and built less complete but still sufficient assemblies for comparative genome analyses. These de novo (*18*) genome assemblies ranged from 1.05 to 1.26 Gb, which is consistent with estimated cytology-based genome sizes (*21*), suggesting near complete genome coverage for all species. Scaffold N50 sizes for high-coverage genomes ranged from 1.2 to 6.9 Mb, whereas those for lower-coverage genomes were ~48 kb on average (table S2). The genomes of the ostrich and budgerigar were further assembled with optical maps, increasing their scaffold N50 sizes to 17.7 and 13.8 Mb, respectively (*20*, *22*).

We annotated the protein-coding sequences using a homology-based method for all genomes, aided by transcriptome sequencing for some species (*18*). To avoid systematic biases related to the use of different methods in annotations of previously published avian genomes, we created a uniform reference gene set that included all genes from the chicken, zebra finch, and human (*23*). This database was used to predict protein gene models in all avian genomes and American alligator (*Alligator mississippiensis*) (*24*). All high-coverage genomes were predicted to contain ~15,000 to 16,000 transposable element-free protein-coding genes [table S3 and annotation files in (*19*)], similar to the chicken genome (~15,000). Despite the fragmented nature of the low-coverage genomes leading to ~3000 genes likely missing or partially annotated, it was still possible to predict 70 to 80% of the entire catalog of avian genes.

Broad patterns of avian genome evolution

Although many fishes and some amphibians have smaller genomes than birds, among amniotes, birds have the smallest (*21*). The genomes of mammals and nonavian reptiles typically range from 1.0 to 8.2 Gb, whereas avian genomes range from 0.91 in the black-chinned hummingbird (*Archilochus alexanderi*) to a little over 1.3 Gb in the common ostrich (*21*). A number of hypotheses have been proposed for the smaller avian genome size (*25*–*28*). Here, we document key events that have likely contributed to this smaller genome size.

The proliferation and loss of transposable elements (TEs) may drive vertebrate genome size evolution (*29*–*31*). Consistent with the zebra finch and galliformes genomes (*11*–*13*, *32*), almost all avian genomes contained lower levels of repeat elements (~4 to 10% of each genome) (table S4) than in other tetrapod vertebrates (for example, 34 to 52% in mammals) (*33*). 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). In contrast, the average total length of SINEs (short interspersed elements) in birds has been reduced to ~1.3 Mb, which is ~10 to 27 times less than in other reptiles [12.6 Mb in alligator; 34.9 Mb in green sea turtle (*Chelonia mydas*)], suggesting that a deficiency of SINEs occurred in the common ancestor of birds.

We compared the average size of genomic elements of birds with 24 mammalian and the three nonavian reptile genomes. Avian protein-coding genes were on average 50 and 27% shorter than the mammalian and reptilian genes, respectively (Fig. 2A). This reduction is largely due to the shortening of introns and reduced intergenic distances that resulted in an increased gene density (Fig. 2A). Such genomic contraction has also evolved convergently in bats (fig. S11), the only flying mammalian group. The condensed genomes may represent an adaptation tied to rapid gene regulation required during powered flight (*34*, *35*).

To further investigate whether avian genome size reduction is due to a lineage-specific reduction in the common avian ancestor of birds or expansion in other vertebrates (*36*), we performed ancestral state reconstructions of small [<100 base pairs (bp)] deletion events across an alignment of four representative well-assembled avian and three reptile genomes (*18*) and 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 crocodiles (fig. S12). In contrast, many fewer small deletion events occurred in modern avian lineages (fig. S12).

We next created a gene synteny map between the highest-quality assembled avian genome (ostrich) and other reptile genomes to document lineage-specific events of large segmental deletions (*18*). We detected 118 syntenic blocks, spanning a total of 58 Mb, that are present in alligator and turtle genomes but lost in all birds (table S8). In contrast, ~8x and ~5x fewer syntentic blocks were missing in alligator (14 blocks, 9 Mb) and turtle (27 blocks, 8 Mb) relative to green anole, respectively, confirming the polarity of genome size reduction in birds (table S8). The large segmental losses in birds were skewed to losses from chr2 and chr6 of the green anole (fig. S13). Two of the green anole's 12 pairs of microchromosomes, LGd and LGf, were completely missing in birds, with no homologous genes found within the avian genomes. Most of these lost segments were located at the ends of chromosomes or close to the centrosomes (fig. S13). Furthermore, lost segments were enriched at apparent breakpoints of the avian microchromosomes (Fig. 2B and fig. S13). These findings imply that the large segmental losses may be a consequence of chromosomal fragmentation events in the common ancestor of birds giving rise to additional microchromosomes in modern birds.

The large segmental deletions in birds contain at least 1241 functional protein-coding genes (table S9), with each lost segment containing at least five contiguous genes. The largest region lost in birds was a 2.1-Mb segment of the green anole chr2, which contains 28 protein-coding genes (Fig. 2B). Overall, at least 7% of the green anole macrochromosomal genes were lost through segmental deletions in birds. Although gene loss is a common evolutionary process, this massive level of segmental deletion has not been previously observed in vertebrates. Over 77% of the 1241 genes present in the large segmentally deleted regions have at least one additional paralog in the green anole genome, a level higher than the overall percentage of genes with paralogs in the green anole genome or avian genomes (both at ~70%). This suggests that birds may

[1]China National GeneBank, Beijing Genomics Institute (BGI)–Shenzhen, Shenzhen, 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]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Republic of Korea. [6]Cho and Kim Genomics, Seoul National University Research Park, Seoul 151-919, Republic of Korea. [7]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA. [8]Centro de Investigación en Ciencias del Mar y Limnología (CIMAR)/Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal. [9]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal. [10]Department of Biological Sciences, University of South Carolina, Columbia, SC, USA. [11]Department of Biology and Molecular Biology, Montclair State University, Montclair, NJ 07043, USA. [12]Department of Animal Ecology, Uppsala University, Norbyvägen 18D, S-752 36 Uppsala, Sweden. [13]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia. [14]Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, Singapore 169857, Singapore. [15]Department of Integrative Biology University of California, Berkeley, CA 94720, USA. [16]College of Life Sciences, Wuhan University, Wuhan 430072, China. [17]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China. [18]BGI Education Center,University of Chinese Academy of Sciences, Shenzhen, 518083, China. [19]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. [20]Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA. [21]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK. [22]School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK. [23]Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Portugal. [24]Department of Biology, University of California Riverside, Riverside, CA 92521, USA. [25]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. [26]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. [27]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. [28]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Post Office Box 7011, S-750 07, Uppsala, Sweden. [29]Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea. [30]State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094, China. [31]College of Animal Science and Technology, China Agricultural University, Beijing 100094, China. [32]Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK. [33]Key Lab of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101 China. [34]International Wildlife Consultants, Carmarthen SA33 5YL, Wales, UK. [35]Centre for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark. [36]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA. [37]Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA. [38]The Genome Institute at Washington University, St. Louis, MO 63108, USA. [39]College of Medicine and Forensics, Xi'an Jiaotong University, Xi'an, 710061, China. [40]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [41]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia. [42]Nova Southeastern University Oceanographic Center 8000 N Ocean Drive, Dania, FL 33004, USA. [43]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA. [44]Genetics Division, San Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, CA 92027, USA. [45]Department of Vertebrate Zoology, MRC-116, National Museum of Natural History, Smithsonian Institution, Post Office Box 37012, Washington, DC 20013-7012, USA. [46]Center for Macroecology, Evolution and Climate, the Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O, Denmark. [47]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. [48]Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden. [49]Department of Biochemistry & Biophysics, University of California, San Francisco, CA 94158, USA. [50]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. [51]Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA. [52]Imperial College London, Grand Challenges in Ecosystems and the Environment Initiative, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK. [53]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. [54]Department of Biology, New Mexico State University, Box 30001 MSC 3AF, Las Cruces, NM 88003, USA. [55]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China. [56]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6102, Australia. [57]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. [58]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [59]Department of Medicine, University of Hong Kong, Hong Kong.
*These authors contributed equally to this work. **†Corresponding author. E-mail: zhanggj@genomics.cn (G.Z.); jarvis@neuro.duke.edu (E.D.J.); mtpgilbert@gmail.com (M.T.P.G.); wangj@genomics.cn (J.W.)**
‡Present address: Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA. §Avian Genome Consortium authors and affiliations are listed at the end of this paper.

have undergone functional compensation in their paralogous gene copies, reducing selection against the loss of these segmental regions. We predict that the loss of functions associated with many genes in the avian ancestor may have had a profound influence on avian-specific traits (table S11).

Conservative mode of genome evolution

With ~2/3 of avian species possessing ~30 pairs of microchromosomes, the avian karyotype appears to be distinctly conserved because this phenotype is not a general feature of any other vertebrate group studied to date (*37*). We assessed the rates of avian chromosomal evolution among the 21 more fully assembled genomes (scaffold N50 > 1 Mb) (table S2) (*18*). From the alignment of chicken with the other 20 avian genomes, plus green anole and *Boa constrictor* (*38*), we identified homologous synteny blocks (HSBs) and 1746 evolutionary breakpoint regions (EBRs) in different avian lineages and then estimated the expected number of EBRs (*18*) and the rates of genomic rearrangements, using a phylogenetic total evidence nucleotide tree (TENT) as a guide (*5*). We excluded the turkey genome after detecting an unusually high fraction of small lineage-specific rearrangements, suggesting a high number of local misassemblies. Of the 18 remaining non–Sanger-sequenced genomes (table S2), the estimated rate of chimeric scaffolds that could lead to false EBRs was ~6% (*39*).

The average rate of rearrangements in birds is ~1.25 EBRs per million years; however, bursts of genomic reorganization occurred in several avian lineages (fig. S15). For example, the origin of Neognathae was accompanied by an elevated rate of chromosome rearrangements (~2.87 EBRs per million years). Intriguingly, all vocal learning species [zebra finch, medium-ground finch (*Geospiza fortis*), American crow (*Corvus brachyrhynchos*), budgerigar, and Anna's hummingbird (*Calypte anna*)] had significantly higher rates of rearrangements than those of close vocal nonlearning relatives [golden-collared manakin (*Manacus vitellinus*), peregrine falcon (*Falco peregrinus*) and chimney swift (*Chaetura pelagica*)] [phylogenetic analysis of variance F statistic (F) = 5.78, P = 0.0499] and even higher relative to all vocal nonlearning species (F = 15.03, P = 0.004). This may be related to the larger radiations these clades experienced relative to most other bird groups. However, the golden-collared manakin, which belongs to suboscines (vocal nonlearners) that have undergone a larger radiation than parrots and hummingbirds, has a low rearrangement rate.

We next compared microsynteny (local gene arrangements), which is more robust and accurate than macrosynteny analyses for draft assemblies (*18*). We compared with eutherian mammals, which are approximately the same evolutionary age as Neoaves and whose genome assemblies are of similar quality. We examined the fraction of orthologous genes identified from each pair of two-avian/mammalian genomes, on the basis of syntenic and best reciprocal blast matches (*18*). Birds have a significantly higher percentage of synteny-defined orthologous genes than that of mammals (Fig. 2C). The fraction of genes retained in syntenic blocks in any pairwise comparison was linearly related with evolutionary time, by which the overall level of genome shuffling in birds was lower than in mammals over the past ~100 million years (Fig. 2C). This suggests a higher level of constraint on maintaining gene synteny in birds relative to mammals.

The apparent stasis in avian chromosome evolution suggests that birds may have experienced relatively low rates of gene gain and loss in multigene families. We examined the intensively studied gene families that encode the various α- and β-type subunits of hemoglobin, the tetrameric protein responsible for blood oxygen transport in jawed vertebrates (*40*). In amniotes, the α- and β-globin gene families are located on different chromosomes (*40*) and experienced high rates of gene turnover because of lineage-specific duplication and deletion events (*41*). In birds, the size and membership composition of the globin gene families have remained remarkably constant during ~100 million years of evolution, with most examined species retaining an identical complement (Fig. 2D). Estimated gene turnover rates (λ) of α- and β-globin gene families were over twofold higher in mammals than birds (λ = 0.0023 versus 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. 2D), which encodes the α-chain subunit of a

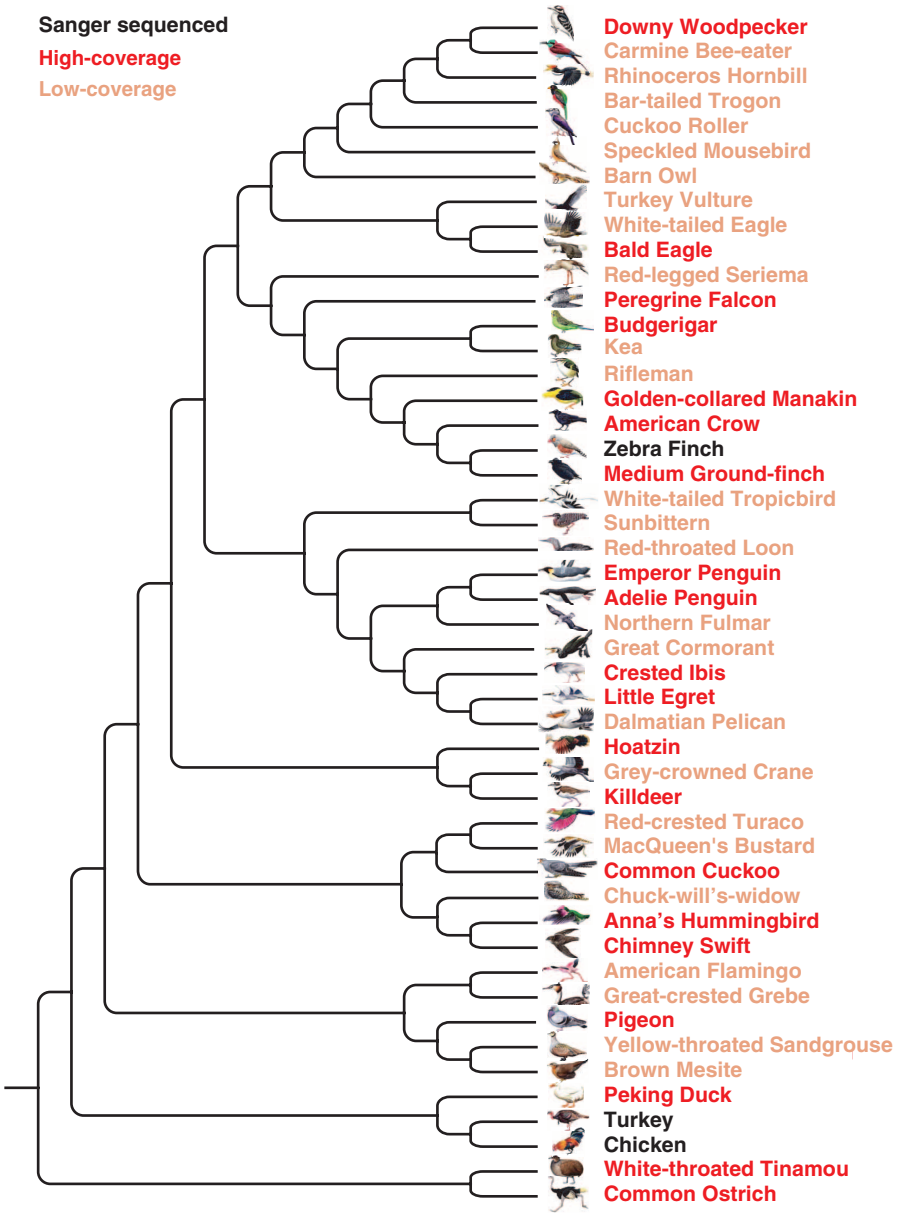

Fig. 1. Avian family tree and genomes sequenced. The phylogenomic relationships of the 48 avian genomes from (*5*), with Sanger-sequenced (black), high-coverage (dark red), and low-coverage (light red) genomes denoted.

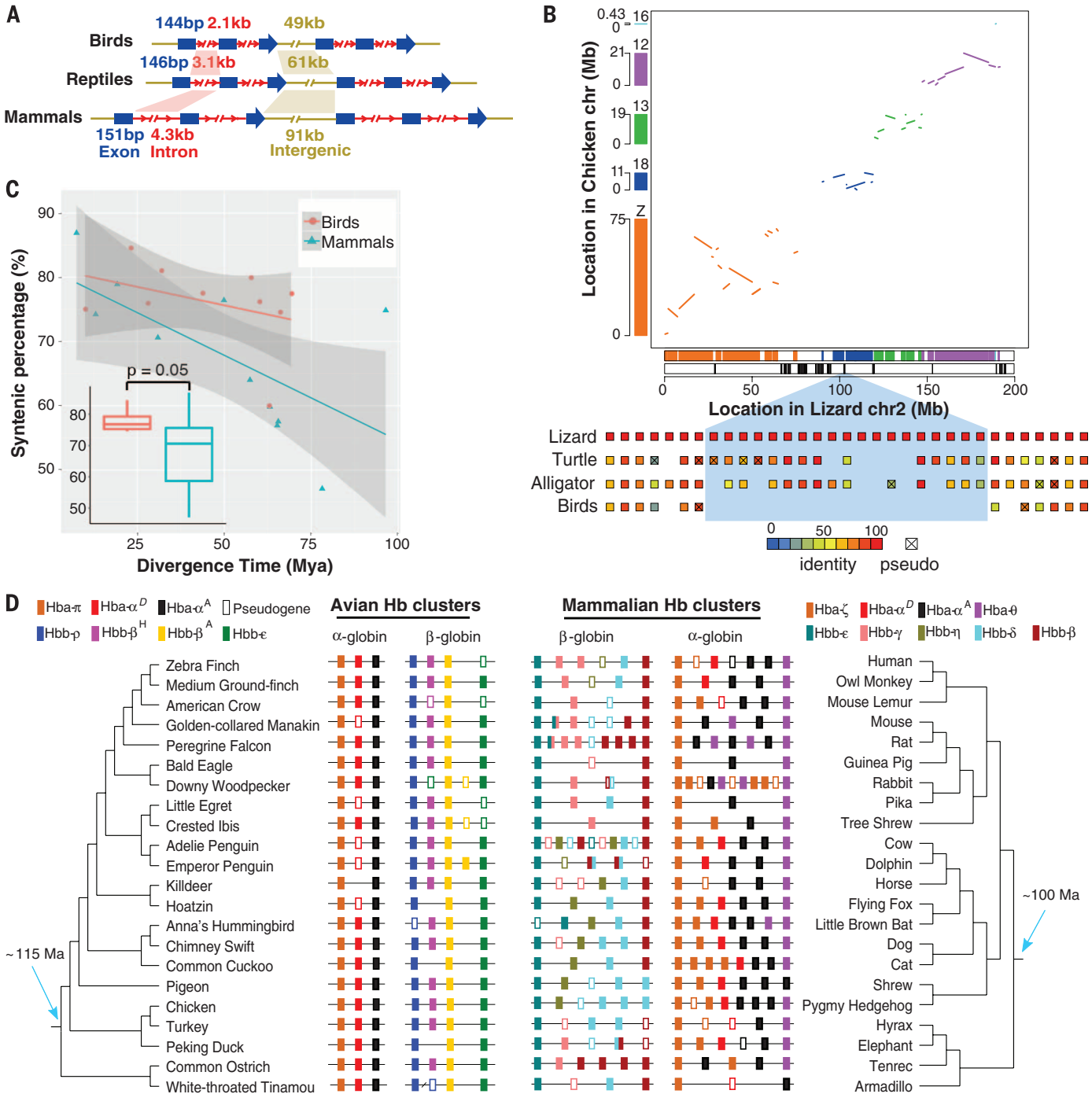

Fig. 2. Genome reduction and conservation in birds. (**A**) Comparison of average size of introns, exons, and intergenic regions within avian, reptilian, and mammalian genomes. (**B**) Synteny plot and large segmental deletions between green anole chromosome 2 and multiple chicken chromosomes. Colored bars and lines indicate homologous blocks between two species; black bars indicate location of large avian-specific segmental deletions, which are enriched at the breakpoints of interchromosome rearrangements. (Bottom) An example of a large segmental deletion in birds (represented by ostrich genes). Homologous genes annotated in each species are shown in small boxes. The color spectrum represents the percent identity of homologous genes with the green anole. (**C**) Distribution of gene synteny percentages identified for phylogenetically independent species pairs of various divergence ages. Dots indicate the percentage of genes remaining in a syntenic block in pairwise comparisons between two avian or mammalian species. Box plots indicate that the overall distributions of the synteny percentages in birds and mammals are different (*P* value was calculated by using Wilcoxon rank sum test with phylogenetically independent species pairs). (**D**) Chromosomal organization 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 (α2β2) 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 *αD*- and *αA*-globin genes. Likewise, the avian π-globin and the mammalian ζ-globin genes are 1:1 orthologs. In 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 (*90*, *91*).

hemoglobin isoform (HbD) expressed in both embryonic and definitive erythrocytes (*42*). Because of uniform and consistent differences in oxygen-binding properties between HbD and the major adult-expressed hemoglobin isoform, HbA (which incorporates products of αA-globin) (*42*), the inactivations of αD-globin likely contribute to variation in blood-oxygen affinity, which has important consequences for circulatory oxygen transport and aerobic energy metabolism. Overall, the globin gene families illustrate a general pattern of evolutionary stasis in birds relative to mammals.

Genomic nucleotide substitution rates vary across species and are determined through both neutral and adaptive evolutionary processes (*43*). We found that the overall pan-genomic background substitution rate in birds (~1.9×10^{-3} substitutions per site per million years) was lower than in mammals (~2.7×10^{-3} substitution per site per million years) (Fig. 3A). However, the substitution rate estimates also exhibited interordinal variation among birds (Fig. 3A). There was a positive correlation between the substitution rate and the number of species per order [coefficient of determination (R^2) = 0.21, P = 0.01, Pearson's test with phylogenetically independent contrasts] (Fig. 3B and fig. S19), evidencing an association with rates of macroevolution (*44*). For example, Passeriformes, the most diverse avian order, exhibited the highest evolutionary rate (~3.3×10^{-3} substitutions per site per million years), almost two times the average of Neoaves (~2×10^{-3} substitutions per site per million years, Fig. 3A). Landbirds exhibited an average higher substitution rate than that of waterbirds (landbirds, ~2.2×10^{-3} substitutions per site per million years; waterbirds, ~1.6×10^{-3} substitutions per site per million years), which is consistent with the observation that landbirds have greater net diversification rates than those of waterbirds (*7*). Among the landbirds, the predatory lineages exhibited slower rates of evolution (~1.6×10^{-3} substitutions per site per million years), similar to that of waterbirds. Moreover, the three vocal learning landbird lineages (parrots, songbirds, and hummingbirds) are evolving faster than are nonvocal learners (Fig. 3A). Overall, our analyses indicate that genome-wide variation in rates of substitution is a consequence of the avian radiation into a wide range of niches and associated phenotypic changes.

Selective constraints on functional elements

Conservation of DNA sequences across distantly related species reflects functional constraints (*45*). A direct comparison of 100-Mb orthologous genomic regions revealed more regions evolving slower than the neutral rate among birds (Fig. 3C) than mammals (*46*), which is consistent with the slower rate of avian mitochondrial sequence evolution (*47*). We predicted 3.2 million highly conserved elements (HCEs) at a resolution of 10 bp or greater spanning on average 7.5% of the avian genome, suggesting a strong functional constraint in avian genomes. Functional annotations revealed that ~12.6% of these HCEs were associated with protein-coding genes, whereas

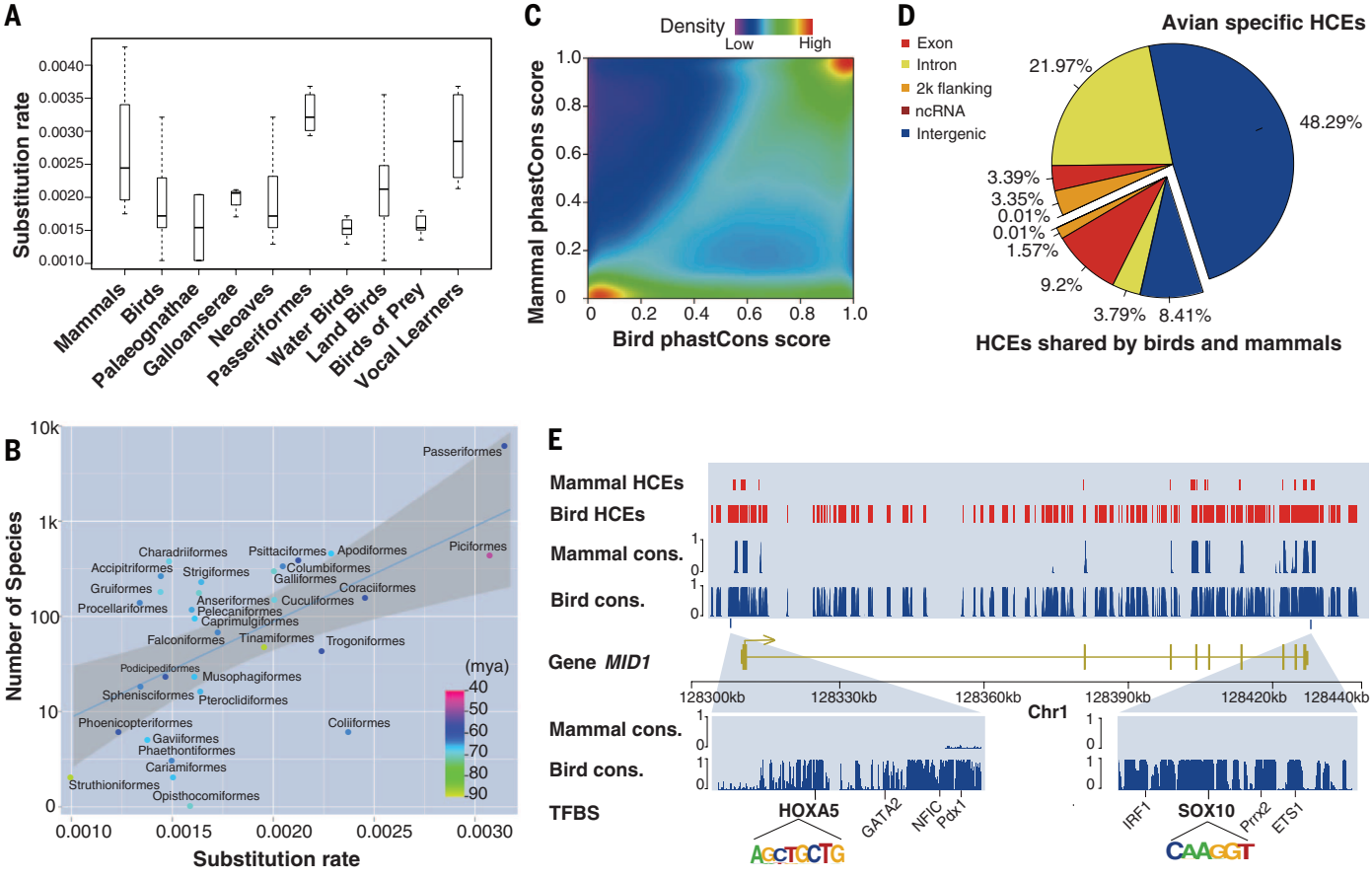

Fig. 3. Evolutionary rate and selection constraints. (**A**) Substitution rate in each lineage was estimated by the comparison of fourfold degenerate (4d) sites in coding regions, in units of substitutions per site per million years. Waterbirds and landbirds are defined in (*5*). (**B**) Correlation between average substitution rates and number of species within different avian orders. Divergence times were estimates from (*5*). The fit line was derived from least square regression analysis, and the confidence interval was estimated by "stat_smooth" in R. The units of the *x* axis are numbers of substitutions per site per million years. The correlation figure with phylogenetically independent contrasts is provided in the supplementary materials. (**C**) Density map for comparison of conservation levels between pan-avian and pan-mammalian genomes, on the basis of the homologous genomic regions between birds and mammals. Conservation levels were quantified by means of PhastCons basewise conservation scores. (**D**) HCEs found in both mammalian and avian genomes (smaller pie piece) and those that are avian-specific (larger pie piece). (**E**) MID1 contains abundant avian-specific HCEs in the upstream and downstream regulatory regions. Many regulatory motif elements are identified in these avian-specific HCEs. Cons., conservation level.

the majority of the remaining HCEs were located in intron and intergenic regions (Fig. 3, D and E). These HCEs enabled us to identify 717 new protein-coding exons and 137 new protein-coding genes, with 77% of the latter supported by the deep transcriptome data (table S17). Deep transcriptome sequencing also enabled us to annotate 5879 candidate long noncoding RNA (lncRNA) genes, of which 220 overlapped HCEs with a coverage ratio of >50% (table S18) (*18*).

Because HCEs may have different functions in different lineages, we separated the HCEs into two categories: bird-specific and amniote HCEs (shared by birds and mammals). Among the bird-specific HCEs, we identified 13 protein-coding genes that were highly conserved in birds but divergent in mammals (table S19). One of the most conserved was the sperm adhesion gene, *SPAM1*, which mediates sperm binding to the egg coat (*48*). This gene, however, was under positive selection driven by sperm competition in mammalian species (*49*). Noncoding HCEs play important roles in the regulation of gene expression (*50*); thus, we compared the transcription factor binding sites in the ENCODE project (*51*) with the HCEs and found that the avian-specific HCEs are significantly associated with transcription factors functioning in metabolism (table S20), whereas amniote core HCEs are enriched with transcription factors functioning in signal regulation, stimulus responses, and development (table S21).

To investigate evolutionary constraints on gene regions, we calculated d*N*/d*S* [the ratio of the number of nonsynonymous substitutions per nonsynonymous site (dN) to the number of synonymous substitutions per synonymous site (d*S*)] for 8295 high-quality orthologs. Consistent with the fast-Z sex chromosome hypothesis (*52*), the evolutionary rate of Z-linked genes was significantly higher than autosome genes (Fig. 4A). This is most likely driven by the reduction of effective population size (N_e) of Z-linked genes—because the N_e of Z chromosome is only 3/4 of that of autosomes—as well as by male sexual selection (*52*). Furthermore, consistent with the fast-macro hypothesis, the overall rate of macrochromosomal genic evolution is higher than that of microchromosomes (Fig. 4A), which is probably due to differences in the recombination rates and genic densities between macro- and microchromosomes in birds (*53*).

We also examined the d*N*/d*S* ratio of each avian Gene Ontology (GO) category for comparison with mammals and within birds. Those involved in development (such as spinal cord development and bone resorption) are evolving faster in birds, and those involved in the brain function (such as synapse assembly, synaptic vesicle transport, and neural crest cell migration) are evolving faster in mammals (tables S23 and S24). Genes involved in oxidoreductase activity were relatively rapidly evolving in the Palaeognathae clade that contains the flightless ratites (Fig. 4B and table S25). The fast evolving GOs in the Galloanserae participate in regulatory functions (Fig. 4B and table S26). In Neoaves, genes involved in microtubule-based processes were the fastest evolving (Fig. 4B and table S27). We speculate that these differences could 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 major modern avian lineages and their revised phylogenetic relationships (*5*), it becomes possible to conduct genome-wide association studies across species with convergent traits. 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 three times (*5*, *54*). All three groups have specialized song-learning forebrain circuits (song nuclei) not found in vocal nonlearners (Fig. 5A) (*55*).

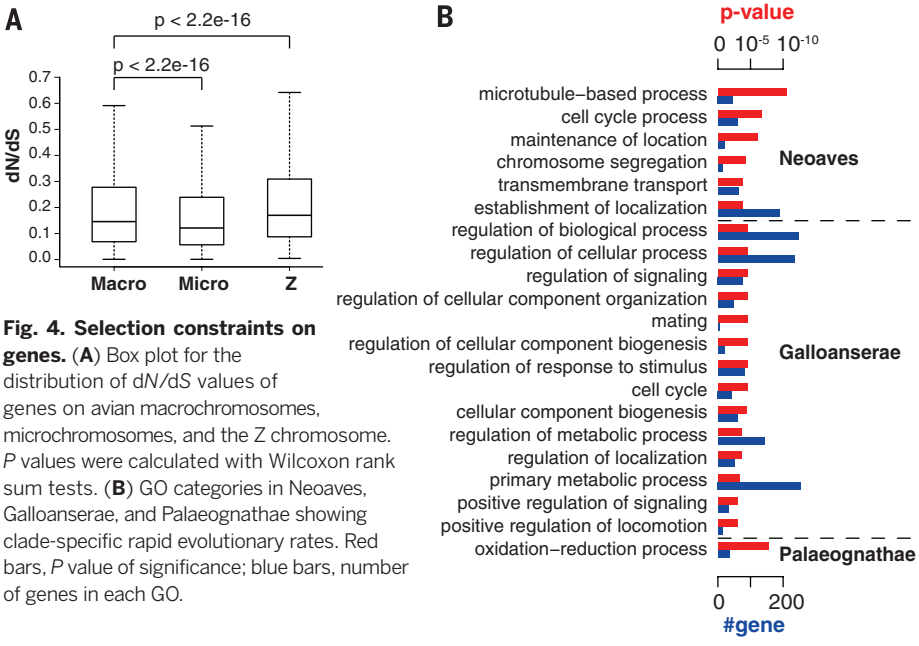

Fig. 4. Selection constraints on genes. (**A**) Box plot for the distribution of d*N*/d*S* values of genes on avian macrochromosomes, microchromosomes, and the Z chromosome. *P* values were calculated with Wilcoxon rank sum tests. (**B**) GO categories in Neoaves, Galloanserae, and Palaeognathae showing clade-specific rapid evolutionary rates. Red bars, *P* value of significance; blue bars, number of genes in each GO.

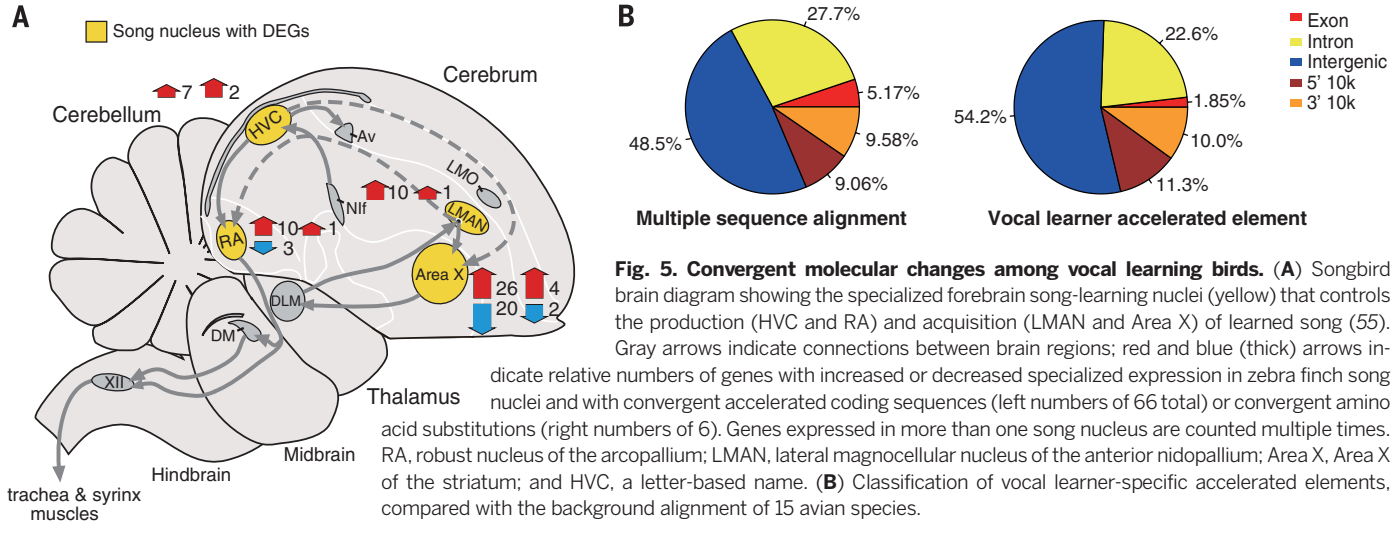

Fig. 5. Convergent molecular changes among vocal learning birds. (**A**) Songbird brain diagram showing the specialized forebrain song-learning nuclei (yellow) that controls the production (HVC and RA) and acquisition (LMAN and Area X) of learned song (*55*). Gray arrows indicate connections between brain regions; red and blue (thick) arrows indicate relative numbers of genes with increased or decreased specialized expression in zebra finch song nuclei and with convergent accelerated coding sequences (left numbers of 66 total) or convergent amino acid substitutions (right numbers of 6). Genes expressed in more than one song nucleus are counted multiple times. RA, robust nucleus of the arcopallium; LMAN, lateral magnocellular nucleus of the anterior nidopallium; Area X, Area X of the striatum; and HVC, a letter-based name. (**B**) Classification of vocal learner-specific accelerated elements, compared with the background alignment of 15 avian species.

Analyses of 7909 orthologous protein-coding genes with available amino acid sites in all three vocal-learning and control vocal nonlearning groups revealed convergent accelerated d*N*/d*S* for 227 genes in vocal learners (table S28). Of these, 73% (165) were expressed in the songbird brain (physically cloned mRNAs), and of these, 92% (151) were expressed in adult song-learning nuclei, which is much higher than the expected 60% of brain genes expressed in song nuclei (*56*). About 20% (*33*) were regulated by singing, which is twice the expected 10% (*56*). In addition, 41% of the song nuclei accelerated genes showed differential expression among song nuclei [expected 20% (*56*)], and 0.7 to 9% [0.7 to 4.3% expected (*57*)] showed specialized expression compared with the surrounding brain regions (table S28) (*58*). GO analyses of the 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. S25). For an independent measure of convergence, we developed an approach that scans for single amino acid substitutions common to species with a shared trait, controlling for phylogenetic relationships (*18*). Of the 7909 genes, 38 had one to two amino acid substitutions present only in vocal learners (table S31). At least 66% of these were expressed in the songbird brain, including in the song nuclei [58%; 20% expected (*56*)]. Two genes (*GDPD4* and *KIAA1919*) showed convergent accelerated evolution on the amino acid sites specific to vocal learners (table S31).

To identify accelerated evolution in noncoding sequences in vocal learners, we scanned the genome alignment using phyloP (*18*, *59*). We used a more limited sampling of vocal nonlearning species closely related to the vocal learners (table S32) because of the relatively faster evolutionary rate of noncoding regions. We scanned the entire genome alignment and found 822 accelerated genomic elements specifically shared by all three vocal learning groups (table S33). These convergent elements were skewed to intergenic regions in vocal learners relative to the background average accelerated elements across species (Fisher's exact test, $P < 2.2 \times 10^{-16}$) (Fig. 5B). Of these elements, 332 were associated with 278 genes (within 10 kb 5′ or 3′ of the nearest gene), of which a high proportion (76%) was expressed in the brain; almost all of those (94%, 198 genes) expressed in one or more song nuclei, 20% were regulated by singing (10% expected), 51% (20% expected) showed differential expression among song nuclei, and 2 to 15% [0.7 to 4.3% expected, based on (*56*)] had specialized expression relative to the surrounding brain regions, including the *FoxP1* gene involved in speech (table S34 and figs. S27 to S32). Overall, these analyses show a 2- to 3.5-fold enrichment of accelerated evolution in regulatory regions of genes differentially expressed in vocal learning brain regions. In contrast, there was very little overlap (2.5%) of genes with convergent accelerated noncoding changes and convergent accelerated amino acid changes, indicating two independent targets of selection for convergent evolution.

Evolution of ecologically relevant genes

We also investigated candidate genes that underlie traits relevant to avian ecological diversity. Although these analyses should be approached with caution given the phenotypic and ecological plasticity within major avian lineages, we examined genes putatively associated with major skeletal and tissue changes for the capacity for powered flight, feeding modification such as loss of teeth, the advanced visual system found in some lineages, and 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. One of the key requirements for flight is a skeleton that is both strong and lightweight. In both birds and nonavian theropods, this evolved through the fusion and elimination of some bones and the pneumatization of the remaining ones (*60*). Of 89 genes involved in ossification (table S36), 49 (~55%) showed evidence of positive selection in birds, which is almost twice as high as in mammals (31 genes, ~35%). 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 d*N*/d*S* (>0.5) were obtained for *AHSG* (*α-2-HS-glycoprotein*), which is associated with bone mineral density, and *P2RX7* (*P2X purinoceptor 7*), which is associated with bone homeostasis. The variation in the extension of pneumatization in avian post-cranial bones has been associated with the variation in body size and foraging strategies (*61*). Therefore, selection of these genes may explain variation in the levels of bone pneumatization in birds because the genes involved in the process of maintaining trabeculae within bones likely depends on the intrinsic network of genes participating in bone resorption and mineralization. These results suggest that most structural differences in bone between birds and mammals may be a result of bone remodeling and resorption (table S37).

Pulmonary structure and function: The increased metabolism associated with homeothermy and powered flight requires an efficient gas exchange process during pulmonary ventilation. Because of 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 (*62*). 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 (*63*). We found five genes that function in mammalian lung development that were lost in the avian ancestor (table S11).

Feathers: The evolution and subsequent morphological diversification of feathers have shaped avian physiology, locomotion, mate choice, and ecological niches (*64*). Feathers are composed of α- and β-keratins (*65*), the latter of which are structural proteins found only in the epidermal appendages of birds and other reptiles. The α-keratin gene family has contracted in birds relative to reptiles (except turtle) and mammals (0.7-fold change), whereas the β-keratin gene family has expanded (1.96-fold change) relative to reptiles (Fig. 6A and table S39). The avian β-keratins form six clusters, with all major avian lineages possessing members from each avian cluster (fig. S33), indicating that avian β-keratin diversity was present in the basal avian lineage. Of these, the feather β-keratin subfamily is avian-specific and comprises over 56% of the genes, whereas the remaining avian β-keratin subfamilies (claw, scale, and keratinocyte β-keratin subfamilies) are found in turtles and crocodiles (Fig. 6A and fig. S33). The mean number of keratinocyte β-keratins is similar across bird groups and their two closest living reptile relatives (turtle and alligator), suggesting copy number conservation since their common ancestor (Fig. 6A). In contrast, aquatic/semi-aquatic birds have a relatively low mean number of feather β-keratins compared with that of land birds, with land birds having more than double the number, and among them several domesticated land birds (zebra finch, chicken, pigeon, and budgerigar) having more than 8 times (Fig. 6A). Although the later observation is concordant with the hypothesis that domestication may increase the recombination rate at β-keratin loci (*66*, *67*), domestic turkey and Peking duck did not exhibit this trend. Overall, these findings indicate that feather compositional adaptations are associated with different avian lifestyles.

Evolution of genes related to diet

Edentulism: The evolution of birds also had major consequences with regard to their feeding strategies and diets, with changes at the structural, biochemical, and sensory levels (among others). 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 (*68*). However, although most phylogenetic analyses suggest that teeth were lost in the common ancestor of modern birds (*69*), several studies have recovered dentate taxa (*Hesperornis* and *Ichthyornis*) from the Mesozoic inside of crown Neornithes, suggesting that tooth loss could have occurred independently (*70*). A scan of avian genomes for molecular fossils of tooth-specific genes recovered remnants of enamel and dentin formation genes in all species examined [table 1 in (*71*)]. Frameshift mutations and whole-exon deletions were widespread in all investigated tooth genes. The vast majority of debilitating mutations were not shared, but all species shared unambiguous deletions in protein-coding exons of enamel-specific genes (*ENAM*, *AMEL*, *AMBN*, *MMP20*, and *AMTN*) and one dentin-specific gene (*DSPP*). This shared pattern of pseudogenization across living birds supports the hypothesis that the common ancestor of modern birds lacked mineralized teeth (*69*).

Diet-related enzymes: Birds have evolved an extraordinary diversity of dietary specializations. The glyoxylate detoxifying enzyme alanine/glyoxylate aminotransferase (AGT) represents a candidate for study (*72*). We recovered complete AGT genes from 22 avian genomes (table S42), of which five exhibit pseudogenized forms in their MTS region (Fig. 6B and fig. S34). MTS function was lost in three unrelated avian orders, which 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 according to human AGT; posterior probability > 99%; $P < 0.0001$).

Vitamin C (Vc) is an important nutrient cofactor in a range of essential metabolic reactions. Loss of the ability to synthesize Vc has occurred in humans, Guinea pigs, and some bats. All species that do not synthesize Vc exhibit a pseudogenized gene for L-gulonolactone oxidase (*GULO*), an enzyme essential for catalyzing the last step of Vc synthesis (*73*). Genomic mining revealed *GULO* pseudogenization in two oscines (medium ground-finch and zebra finch) and the suboscine golden-collared manakin (Fig. 6B and fig. S35). In contrast, intact *GULO* was recovered from the third oscine species, American crow, and the basal passerine rifleman (*Acanthisitta chloris*) (table S43). Similar to mammals (*74*), this pseudogenization was caused by the loss of different exons and lethal mutations. We also found purifying selection has dominated *GULO* evolution, from the ancestral amniote node ($dN/dS = 0.096$) to ancestral birds ($dN/dS = 0.133$) and mammals ($dN/dS = 0.355$), suggesting conservation of the ability to synthesize Vc both before and after avian divergence. However, both the American crow and rifleman exhibited nonsynonymous changes in *GULO* at one order of magnitude higher than the average (fig. S36), a sign of potentially harmful mutations (*75*).

Rhodopsin/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 hypothesized to play a role in reproduction and feeding (*76*). Vertebrate visual opsins are classified into five genes in two families: rhodopsin (RH1) and conopsins (RH2, OPN1sw1, OPN1sw2, and OPN1lw). In most avian genomes, we detected higher numbers of opsin genes than in mammalian genomes, which lacked *OPN4x* (*77*), *RH2*, and either *OPN1sw1* (Monotremata) or *OPN1sw2* (Theria). All avian genomes contained *RH1* and *RH2*, and most high-coverage genomes contained two to three of the remaining three conopsin genes (table S44), supporting that ancestral avian vision is tetrachromatic. Penguins were one of the exceptions, with both species exhibiting only three classes of functional opsins, and thus are trichromatic, which is in line with retinal examination (*78*). This is 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 (or pigments) (*76*).

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. S38), which corroborates that the shift from violet sensitive SWS1 cones in this clade was adaptive (*79*). Excluding these species, dN/dS values for *OPN1sw1* were lower in birds than in mammals (Fig. 6C). Optimal color discrimination requires an even distribution of spectral sensitivities (*80*), which is more easily disturbed with an increasing number of cone classes. Hence, stabilizing selection on spectral sensitivity should be stronger in birds than in mammals, and the dN/dS values are consistent with this prediction. Besides two transmembrane regions (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 acid sites important to spectral tuning of this ultraviolet-sensitive cone (fig. S40).

Sex-related and reproductive traits

Reproduction-related genes: Unlike other reptiles, almost all birds develop only a single functional ovary, on the left side (*81*), as a result of the evolutionary loss of the right ovary during the transition from nonavian theropods to birds (*82*). It has been hypothesized that this loss represents an adaptation to reduce weight during flight (*82*). We found that two genes related with ovary development (*MMP19* and *AKR1C3*) have been lost in birds. *MMP19*, a matrix metalloprotease gene, functions during the follicular growth and the ovulation process (*83*), and the enzyme AKR1C3 catalyzes the conversion of androstenedione to testosterone and has been associated with polycystic ovary syndrome (*84*).

We analyzed a range of other genes related to reproduction, under the hypothesis that some of them may have been direct targets of the morphological and behavioral adaptations related to sexual selection in birds. Reproduction genes in *Drosophila*, humans, and marine invertebrates evolve faster than do nonreproduction genes

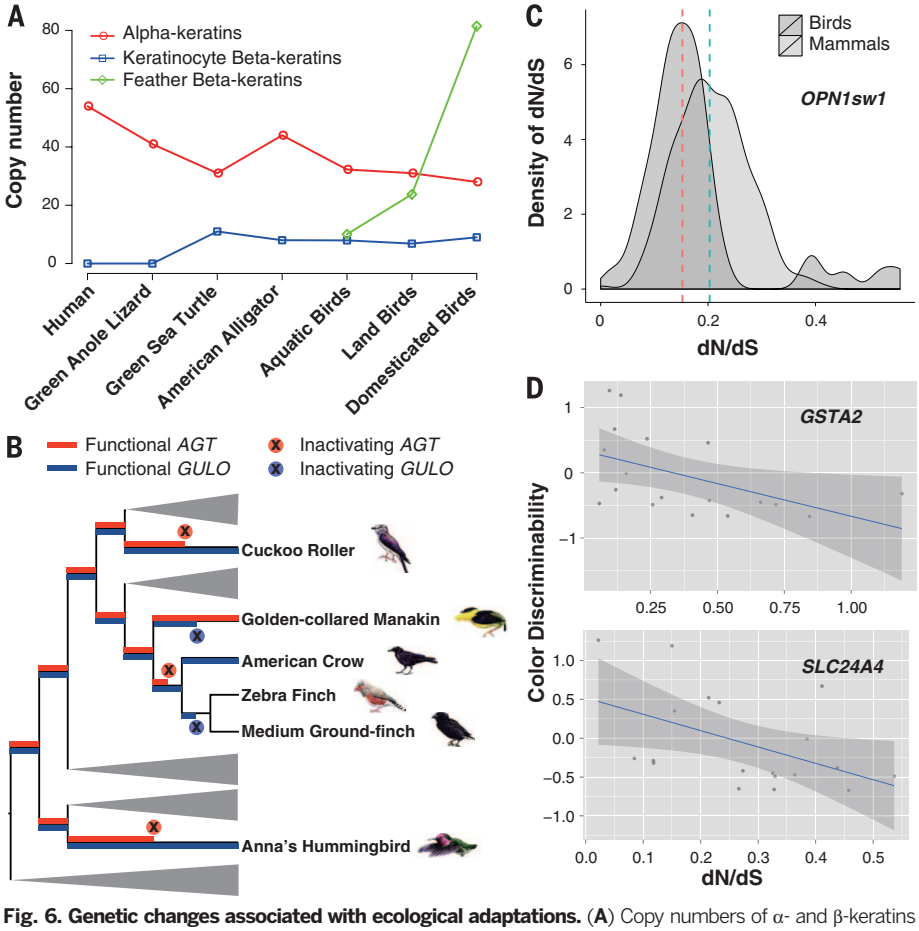

Fig. 6. Genetic changes associated with ecological adaptations. (**A**) Copy numbers of α- and β-keratins in humans, reptiles, and birds, including in aquatic birds, land birds, and domesticated birds. Definitions of aquatic and landbirds are provided in (*5*). (**B**) Pseudogenization events of the 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 color discriminability values (log transformation applied). The correlation figures with phylogenetically independent contrasts are provided in the supplementary materials.

(*49*). We chose 89 genes that may be involved in spermatogenesis (table S46) and six involved in oogenesis (table S47). We found that 19 out of 46 avian species show significantly accelerated evolution (lineage-specific d*N*/d*S* ratio) of spermatogenesis genes relative to the genomic background (table S48). In contrast, only the carmine bee-eater (*Merops nubicoides*) and Peking duck showed significantly accelerated evolution in oogenesis genes (table S49). These results suggest that male birds are the dominant targets of sexual selection, which drives rapid evolution of spermatogenesis genes via sperm competition (*85*).

Plumage color: We investigated the genomics of plumage color, a behaviorally important trait and longstanding example of sexual selection (*86*). Male birds have frequently evolved extravagant plumage color in response to both male-male competition and female choice (*87*, *88*), resulting in remarkable sexual dichromatism. Analysis of 15 genes implicated in avian plumage coloration demonstrated rapid evolutionary rates over the genomic average in 8 of 46 lineages (table S51). This pattern suggests that these genes are evolving under adaptive evolution.

Carotenoids, which are responsible for the bright yellow and red pigments that underlie some of the most conspicuous coloration patterns in vertebrates, unlike melanins can be only acquired through diet and represent trade-offs between coloration and other physiological conditions. We identified a negative correlation between color discriminability and d*N*/d*S* across birds for the gene *GSTA2*, which is involved in the binding and deposition of carotenoids and in plumage dichromatism ($R^2 = 0.24$, $P = 0.045$, Pearson's test with phylogenetically independent contrasts) (Fig. 6D and fig. S41), and similarly for *SLC24A4*, which is associated with hair color in humans ($R^2 = 0.21$, $P = 0.056$, Pearson's test with phylogenetically independent contrasts) (Fig. 6D and fig. S42), suggesting that either diversifying and stabilizing selection or the effect of different population sizes is driving the evolution of plumage color genes.

Discussion and conclusions

The small genome size of birds with fragmented microchromosomes and reduced repeat transposon activity, in contrast to other vertebrates, has been a static feature in the avian clade for >100 million years. Avian genomes consistently contain fewer genes, ~70% of the number of the human genome, and with one detected exception (downy woodpecker), an extremely reduced fraction of repeat elements. Thus, the ancestral avian lineage has distinctly lost a large number of genes by means of large segmental deletions after their divergence from other extant reptiles. These large genomic sequence deletions appear to be linked to a second defining feature of avian genomes: the putatively ancestral fission of macrochromosomes into a relatively large number of microchromosomes.

Genome conservation in birds—along with regard to sequence, synteny, and chromosomal structure—is remarkable in light of their rapid historical radiation. This is considerably different from the evolution of mammalian genomes, which although are experiencing a rapid radiation at a similar time, today display richer genome shuffling and variation (*89*). By comparing the genomes of 48 birds that are constrained within a largely resolved phylogeny, we discovered 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 not possible in previous genomic studies restricted to fewer taxa.

The analyses of genome sequences for taxa distributed across the avian phylogeny also explains the rich biodiversity of the avian clade because we identified selective constraints on certain categories of genes in different avian lineages. Convergent evolution also appears to be shaping the evolution of protein-coding genes and their regulatory elements, establishing similar morphological or behavioral features in distantly related bird species, as well as variation in specific gene families that correspond to avian traits and environmental adaptation. We believe that the data and analyses presented here open a new window into the evolution, diversification, and ecological adaptation of tetrapod vertebrates and offers a phylogenomic perspective that helps bridge the chasm between micro- and macroevolution.

REFERENCES AND NOTES

1. F. Gill, D. Donsker, IOC World Bird List (version 3.5) (2013).
2. L. M. Chiappe, L. M. Witmer, *Mesozoic Birds: Above the Heads of Dinosaurs* (Univ. California Press, Berkeley, CA, 2002).
3. G. Dyke, G. W. Kaiser, *Living Dinosaurs the Evolutionary History of Modern Birds* (Wiley-Blackwell, Hoboken, NJ, 2011).
4. A. Feduccia, *Trends Ecol. Evol.* **18**, 172–176 (2003).
5. E. D. Jarvis *et al*., *Science* **346**, 1320–1331 (2014).
6. B. G. Holt *et al*., *Science* **339**, 74–78 (2013).
7. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, *Nature* **491**, 444–448 (2012).
8. H. Zeigler, P. E. Marler, in Behavioral Neurobiology of Birdsong, December 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, Cambridge, UK, 1998).
10. D. J. Alexander, *Vet. Microbiol.* **74**, 3–13 (2000).
11. L. W. Hillier *et al*., *Nature* **432**, 695–716 (2004).
12. R. A. Dalloul *et al*., *PLOS Biol.* **8**, e1000475 (2010).
13. W. C. Warren *et al*., *Nature* **464**, 757–762 (2010).
14. Y. Huang *et al*., *Nat. Genet.* **45**, 776–783 (2013).
15. X. Zhan *et al*., *Nat. Genet.* **45**, 563–566 (2013).
16. M. D. Shapiro *et al*., *Science* **339**, 1063–1067 (2013).
17. E. Dickinson, J. Remsen, *The Howard and Moore Complete Checklist of the Birds of the World* (Aves Press, Eastbourne, UK, 2013).
18. Materials and methods are available as supplementary materials on *Science* Online.
19. G. Zhang *et al*., *GigaScience* **3**, 26 (2014).
20. G. Ganapathy *et al*., *GigaScience* **3,** 11 (2013).
21. T. R. Gregory, in The Animal Genome Size Database (2005); available at www.genomesize.com.
22. Q. Zhou *et al*., *Science* **346**, 1246338 (2014).
23. E. S. Lander *et al*., *Nature* **409**, 860–921 (2001).
24. R. E. Green *et al*., *Science* **346**, 1254449 (2014).
25. A. L. Hughes, M. K. Hughes, *Nature* **377**, 391 (1995).
26. M. Rho *et al*., *Genome Biol. Evol.* **1**, 2–12 (2009).
27. S. Morand, R. E. Ricklefs, *Trends Genet.* **17**, 567–568 (2001).
28. E. Waltari, S. V. Edwards, *Am. Nat.* **160**, 539–552 (2002).
29. M. G. Kidwell, *Genetica* **115**, 49–63 (2002).
30. C. Feschotte, E. J. Pritham, *Annu. Rev. Genet.* **41**, 331–368 (2007).
31. M. Lynch, J. S. Conery, *J. Struct. Funct. Genomics* **3**, 35–44 (2003).
32. N. Sela, E. Kim, G. Ast, *Genome Biol.* **11**, R59 (2010).
33. A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, *Chromosome Res.* **16**, 203–215 (2008).
34. Q. Zhang, S. V. Edwards, *Genome Biol. Evol.* **4**, 1033–1043 (2012).
35. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, *Nature* **446**, 180–184 (2007).
36. A. L. Hughes, R. Friedman, *Mol. Biol. Evol.* **25**, 2681–2688 (2008).
37. D. K. Griffin, L. B. W. Robertson, H. G. Tempest, B. M. Skinner, *Cytogenet. Genome Res.* **117**, 64–77 (2007).
38. K. R. Bradnam *et al*., *GigaScience* **2**, 10 (2013).
39. J. Kim *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1785–1790 (2013).
40. F. G. Hoffmann, J. C. Opazo, J. F. Storz, *Mol. Biol. Evol.* **29**, 303–312 (2012).
41. F. G. Hoffmann, J. C. Opazo, J. F. Storz, *Mol. Biol. Evol.* **25**, 591–602 (2008).
42. M. T. Grispo *et al*., *J. Biol. Chem.* **287**, 37647–37658 (2012).
43. C. F. Baer, M. M. Miyamoto, D. R. Denver, *Nat. Rev. Genet.* **8**, 619–631 (2007).
44. C. Venditti, M. Pagel, *Trends Ecol. Evol.* **25**, 14–20 (2010).
45. A. Siepel *et al*., *Genome Res.* **15**, 1034–1050 (2005).
46. K. Lindblad-Toh *et al*., *Nature* **478**, 476–482 (2011).
47. B. Nabholz, S. Glémin, N. Galtier, *BMC Evol. Biol.* **9**, 54 (2009).
48. W. F. Lathrop, E. P. Carmichael, D. G. Myles, P. Primakoff, *J. Cell Biol.* **111**, 2939–2949 (1990).
49. W. J. Swanson, V. D. Vacquier, *Nat. Rev. Genet.* **3**, 137–144 (2002).
50. L. A. Pennacchio *et al*., *Nature* **444**, 499–502 (2006).
51. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
52. J. E. Mank, E. Axelsson, H. Ellegren, *Genome Res.* **17**, 618–624 (2007).
53. E. Axelsson, M. T. Webster, N. G. C. Smith, D. W. Burt, H. Ellegren, *Genome Res.* **15**, 120–125 (2005).
54. A. Suh *et al*., *Nat. Commun.* **2**, 443 (2011).
55. E. D. Jarvis, *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004).
56. O. Whitney *et al*., *Science* **346**, 1256780 (2014).
57. A. R. Pfenning *et al*., *Science* **346**, 1256846 (2014).
58. H. J. Karten *et al*., *J. Comp. Neurol.* **521**, 3702–3715 (2013).
59. M. J. Hubisz, K. S. Pollard, A. Siepel, *Brief. Bioinform.* **12**, 41–51 (2011).
60. E. R. Dumont, *Proc. Biol. Sci.* **277**, 2193–2198 (2010).
61. S. C. Gutzwiller, A. Su, P. M. O'Connor, *Anat. Rec.* **296**, 867–876 (2013).
62. H. R. Duncker, *Respir. Physiol. Neurobiol.* **144**, 111–124 (2004).
63. J. B. West, R. R. Watson, Z. Fu, *Respir. Physiol. Neurobiol.* **157**, 382–390 (2007).
64. F. Gill, *Ornithology* (W.H. Freeman and Company, New York, 1995).
65. A. R. Haake, G. König, R. H. Sawyer, *Dev. Biol.* **106**, 406–413 (1984).
66. J. Ross-Ibarra, *Am. Nat.* **163**, 105–112 (2004).
67. A. Blirt, G. Bell, *Nature* **326**, 803–805 (1987).
68. A. Louchart, L. Viriot, *Trends Ecol. Evol.* **26**, 663–673 (2011).
69. B. C. Livezey, R. L. Zusi, *Zool. J. Linn. Soc.* **149**, 1–95 (2007).
70. J. Cracraft, *Syst. Biol.* **31**, 35–56 (1982).
71. R. W. Meredith, G. Zhang, M. T. P. Gilbert, E. D. Jarvis, M. S. Springer, *Science* **346**, 1254390 (2014).
72. G. M. Birdsey, J. Lewin, A. A. Cunningham, M. W. Bruford, C. J. Danpure, *Mol. Biol. Evol.* **21**, 632–646 (2004).
73. I. B. Chatterjee, *Science* **182**, 1271–1272 (1973).
74. J. Cui, X. Yuan, L. Wang, G. Jones, S. Zhang, *PLOS ONE* **6**, e27114 (2011).
75. J. Cui, Y.-H. Pan, Y. Zhang, G. Jones, S. Zhang, *Mol. Biol. Evol.* **28**, 1025–1031 (2011).
76. W. I. L. Davies, S. P. Collin, D. M. Hunt, *Mol. Ecol.* **21**, 3121–3158 (2012).
77. S. S. Pires *et al*., *Proc. Biol. Sci.* **274**, 2791–2799 (2007).
78. J. K. Bowmaker, G. R. Martin, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **156**, 71–77 (1985).
79. A. Ödeen, O. Håstad, P. Alström, *BMC Evol. Biol.* **11**, 313 (2011).
80. L. Chittka, *Naturwissenschaften* **83**, 136–138 (1996).
81. S. S. Guraya, *Ovarian Follicles in Reptiles and Birds* (Springer-Verlag, Berlin, Germany, 1989).
82. X. Zheng *et al*., *Nature* **495**, 507–511 (2013).
83. M. Jo, T. E. Curry Jr., *Biol. Reprod.* **71**, 1796–1806 (2004).
84. M. O. Goodarzi, H. J. Antoine, R. Azziz, *J. Clin. Endocrinol. Metab.* **92**, 2659–2664 (2007).
85. T. R. Birkhead, A. P. Møller, Sperm competition and sexual selection. (Academic Press, 1998).
86. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (J. Murray, London, 1871).
87. M. Zuk, J. D. Ligon, R. Thornhill, *Anim. Behav.* **44**, 999–1006 (1992).
88. C. Mateos, J. Carranza, *Anim. Behav.* **54**, 1205–1214 (1997).
89. A. R. Quinlan, I. M. Hall, *Trends Genet.* **28**, 43–53 (2012).

90. J. F. Storz, J. C. Opazo, F. G. Hoffmann, *Mol. Phylogenet. Evol.* **66**, 469–478 (2013).
91. F. G. Hoffmann, J. F. Storz, T. A. Gorr, J. C. Opazo, *Mol. Biol. Evol.* **27**, 1126–1138 (2010).

ACKNOWLEDGMENTS

Genome assemblies and annotations of avian genomes in this study are available on the avian phylogenomics website (http://phybirds.genomics.org.cn), GigaDB (http://dx.doi.org/10.5524/101000), National Center for Biotechnology Information (NCBI), and ENSEMBL (NCBI and Ensembl accession numbers are provided in table S2). The majority of this study was supported by an internal funding from BGI. In addition, 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); C.L. and Q.L. were partially supported by a Danish Council for Independent Research Grant (10-081390); and E.D.J. was supported by the Howard Hughes Medical Institute and NIH Directors Pioneer Award DP1OD000448.

The Avian Genome Consortium

Chen Ye,[1] Shaoguang Liang,[1] Zengli Yan,[1] M. Lisandra Zepeda,[2] Paula F. Campos,[2] Amhed Missael Vargas Velazquez,[2] José Alfredo Samaniego,[2] María Avila-Arcos,[2] Michael D. Martin,[2] Ross Barnett,[2] Angela M. Ribeiro,[3] Claudio V. Mello,[4] Peter V. Lovell,[4] Daniela Almeida,[3,5] Emanuel Maldonado,[3] Joana Pereira,[3] Kartik Sunagar,[3,5] Siby Philip,[3,5] Maria Gloria Dominguez-Bello,[6] Michael Bunce,[7] David Lambert,[8] Robb T. Brumfield,[9] Frederick H. Sheldon,[9] Edward C. Holmes,[10] Paul P. Gardner,[11] Tammy E. Steeves,[11] Peter F. Stadler,[12] Sarah W. Burge,[13] Eric Lyons,[14] Jacqueline Smith,[15] Fiona McCarthy,[16] Frederique Pitel,[17] Douglas Rhoads,[18] David P. Froman[19]

[1]China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China. [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [3]CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal. [4]Department of Behavioral Neuroscience Oregon Health & Science University Portland, OR 97239, USA. [5]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal. [6]Department of Biology, University of Puerto Rico, Av Ponce de Leon, Rio Piedras Campus, JGD 224, San Juan, PR 009431-3360, USA. [7]Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [8]Environmental Futures Research Institute, Griffith University, Nathan, Queensland 4121, Australia. [9]Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA. [10]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney NSW 2006, Australia. [11]School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand. [12]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Hrtelstrasse 16-18, D-04107 Leipzig, Germany. [13]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. [14]School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA. [15]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. [16]Department of Veterinary Science and Microbiology, University of Arizona, 1117 E Lowell Street, Post Office Box 210090-0090, Tucson, AZ 85721, USA. [17]Laboratoire de Génétique Cellulaire, INRA Chemin de Borde-Rouge, Auzeville, BP 52627 , 31326 CASTANET-TOLOSAN CEDEX, France. [18]Department of Biological Sciences, Science and Engineering 601, University of Arkansas, Fayetteville, AR 72701, USA. [19]Department of Animal Sciences, Oregon State University, Corvallis, OR 97331, USA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1311/suppl/DC1
Supplementary Text
Figs. S1 to S42
Tables S1 to S51
References (*92*–*192*)

27 January 2014; accepted 6 November 2014
10.1126/science.1251385

RESEARCH ARTICLE

Whole-genome analyses resolve early branches in the tree of life of modern birds

Erich D. Jarvis,[1]*† Siavash Mirarab,[2]* Andre J. Aberer,[3] Bo Li,[4,5,6] Peter Houde,[7] Cai Li,[4,6] Simon Y. W. Ho,[8] Brant C. Faircloth,[9,10] Benoit Nabholz,[11] Jason T. Howard,[1] Alexander Suh,[12] Claudia C. Weber,[12] Rute R. da Fonseca,[6] Jianwen Li,[4] Fang Zhang,[4] Hui Li,[4] Long Zhou,[4] Nitish Narula,[7,13] Liang Liu,[14] Ganesh Ganapathy,[1] Bastien Boussau,[15] Md. Shamsuzzoha Bayzid,[2] Volodymyr Zavidovych,[1] Sankar Subramanian,[16] Toni Gabaldón,[17,18,19] Salvador Capella-Gutiérrez,[17,18] Jaime Huerta-Cepas,[17,18] Bhanu Rekepalli,[20] Kasper Munch,[21] Mikkel Schierup,[21] Bent Lindow,[6] Wesley C. Warren,[22] David Ray,[23,24,25] Richard E. Green,[26] Michael W. Bruford,[27] Xiangjiang Zhan,[27,28] Andrew Dixon,[29] Shengbin Li,[30] Ning Li,[31] Yinhua Huang,[31] Elizabeth P. Derryberry,[32,33] Mads Frost Bertelsen,[34] Frederick H. Sheldon,[33] Robb T. Brumfield,[33] Claudio V. Mello,[35,36] Peter V. Lovell,[35] Morgan Wirthlin,[35] Maria Paula Cruz Schneider,[36,37] Francisco Prosdocimi,[36,38] José Alfredo Samaniego,[6] Amhed Missael Vargas Velazquez,[6] Alonzo Alfaro-Núñez,[6] Paula F. Campos,[6] Bent Petersen,[39] Thomas Sicheritz-Ponten,[39] An Pas,[40] Tom Bailey,[41] Paul Scofield,[42] Michael Bunce,[43] David M. Lambert,[16] Qi Zhou,[44] Polina Perelman,[45,46] Amy C. Driskell,[47] Beth Shapiro,[26] Zijun Xiong,[4] Yongli Zeng,[4] Shiping Liu,[4] Zhenyu Li,[4] Binghang Liu,[4] Kui Wu,[4] Jin Xiao,[4] Xiong Yinqi,[4] Qiuemei Zheng,[4] Yong Zhang,[4] Huanming Yang,[48] Jian Wang,[48] Linnea Smeds,[12] Frank E. Rheindt,[49] Michael Braun,[50] Jon Fjeldsa,[51] Ludovic Orlando,[6] F. Keith Barker,[52] Knud Andreas Jønsson,[51,53,54] Warren Johnson,[55] Klaus-Peter Koepfli,[56] Stephen O'Brien,[57,58] David Haussler,[59] Oliver A. Ryder,[60] Carsten Rahbek,[51,54] Eske Willerslev,[6] Gary R. Graves,[51,61] Travis C. Glenn,[62] John McCormack,[63] Dave Burt,[64] Hans Ellegren,[12] Per Alström,[65,66] Scott V. Edwards,[67] Alexandros Stamatakis,[3,68] David P. Mindell,[69] Joel Cracraft,[70] Edward L. Braun,[71] Tandy Warnow,[2,72]† Wang Jun,[48,73,74,75,76]† M. Thomas P. Gilbert,[6,43]† Guojie Zhang[4,77]†

To better determine the history of modern birds, we performed a genome-scale phylogenetic analysis of 48 species representing all orders of Neoaves using phylogenomic methods created to handle genome-scale data. We recovered a highly resolved tree that confirms previously controversial sister or close relationships. We identified the first divergence in Neoaves, two groups we named Passerea and Columbea, representing independent lineages of diverse and convergently evolved land and water bird species. Among Passerea, we infer the common ancestor of core landbirds to have been an apex predator and confirm independent gains of vocal learning. Among Columbea, we identify pigeons and flamingoes as belonging to sister clades. Even with whole genomes, some of the earliest branches in Neoaves proved challenging to resolve, which was best explained by massive protein-coding sequence convergence and high levels of incomplete lineage sorting that occurred during a rapid radiation after the Cretaceous-Paleogene mass extinction event about 66 million years ago.

The diversification of species is not always gradual but can occur in rapid radiations, especially after major environmental changes (*1*, *2*). Paleobiological (*3*–*7*) and molecular (*8*) evidence suggests that such "big bang" radiations occurred for neoavian birds (e.g., songbirds, parrots, pigeons, and others) and placental mammals, representing 95% of extant avian and mammalian species, after the Cretaceous to Paleogene (K-Pg) mass extinction event about 66 million years ago (Ma). However, other nuclear (*9*–*12*) and mitochondrial (*13*, *14*) DNA studies propose an earlier, more gradual diversification, beginning within the Cretaceous 80 to 125 Ma. This debate is confounded by findings that different data sets (*15*–*19*) and analytical methods (*20*, *21*) often yield contrasting species trees. Resolving such timing and phylogenetic relationships is important for comparative genomics, which can inform about human traits and diseases (*22*).

Recent avian studies based on fragments of 5 [~5000 base pairs (bp) (*8*)] and 19 [31,000 bp (*17*)] genes recovered some relationships inferred from morphological data (*15*, *23*) and DNA-DNA hybridization (*24*), postulated new relationships, and contradicted many others. Consistent with most previous molecular and contemporary morphological studies (*15*), they divided modern birds (Neornithes) into Palaeognathae (tinamous and flightless ratites), Galloanseres [Galliformes (landfowl) and Anseriformes (waterfowl)], and Neoaves (all other extant birds). Within Neoaves,

Comparative genomics reveals insights into avian genome evolution and adaptation

Guojie Zhang *et al.*

Science **346**, 1311 (2014);
DOI: 10.1126/science.1251385

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Does DNA act as a telephone line? p. 1234

Mutations enhancing leukemia development p. 1293 & 1272

A stable gold support gives sharper resolution p. 1377

Science

$10
12 DECEMBER 2014
sciencemag.org

AAAS

SPECIAL ISSUE

Avian genomes

Sequencing across the bird species tree

p. 1301

www.sciencemag.org/content/346/6215/1311/suppl/DC1

Supplementary Material for

Comparative genomics reveals insights into avian genome evolution and adaptation

Guojie Zhang,* Cai Li, Qiye Li, Bo Li, Denis M. Larkin, Chul Lee, Jay F. Storz, Agostinho Antunes, Matthew J. Greenwold, Robert W. Meredith, Anders Ödeen, Jie Cui, Qi Zhou, Luohao Xu, Hailin Pan, Zongji Wang, Lijun Jin, Pei Zhang, Haofu Hu, Wei Yang, Jiang Hu, Jin Xiao, Zhikai Yang, Yang Liu, Qiaolin Xie, Hao Yu, Jinmin Lian, Ping Wen, Fang Zhang, Hui Li, Yongli Zeng, Zijun Xiong, Shiping Liu, Long Zhou, Zhiyong Huang, Na An, Jie Wang, Qiumei Zheng, Yingqi Xiong, Guangbiao Wang, Bo Wang, Jingjing Wang, Yu Fan, Rute R. da Fonseca, Alonzo Alfaro-Núñez, Mikkel Schubert, Ludovic Orlando, Tobias Mourier, Jason T. Howard, Ganeshkumar Ganapathy, Andreas Pfenning, Osceola Whitney, Miriam V. Rivas, Erina Hara, Julia Smith, Marta Farré, Jitendra Narayan, Gancho Slavov, Michael N Romanov, Rui Borges, João Paulo Machado, Imran Khan, Mark S. Springer, John Gatesy, Federico G. Hoffmann, Juan C. Opazo, Olle Håstad, Roger H. Sawyer, Heebal Kim, Kyu-Won Kim, Hyeon Jeong Kim, Seoae Cho, Ning Li, Yinhua Huang, Michael W. Bruford, Xiangjiang Zhan, Andrew Dixon, Mads F. Bertelsen, Elizabeth Derryberry, Wesley Warren, Richard K Wilson, Shengbin Li, David A. Ray, Richard E. Green, Stephen J. O'Brien, Darren Griffin, Warren E. Johnson, David Haussler, Oliver A. Ryder, Eske Willerslev, Gary R. Graves, Per Alström, Jon Fjeldså, David P. Mindell, Scott V. Edwards, Edward L. Braun, Carsten Rahbek, David W. Burt, Peter Houde, Yong Zhang, Huanming Yang, Jian Wang, Avian Genome Consortium, Erich D. Jarvis,* M. Thomas P. Gilbert,* Jun Wang,*

*Corresponding author. E-mail: zhanggj@genomics.cn (G.Z.); jarvis@neuro.duke.edu (E.D.J.); mtpgilbert@gmail.com (M.T.P.G.); wangj@genomics.cn (J.W.)

Published 12 December 2014, *Science* **346**, 1311 (2014)
DOI: 10.1126/science.1251385

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/346/6215/1311/suppl/DC1)

Tables S9 to S11, S28 to S31, and S34 as separate text files

Authors list:

Guojie Zhang[1,2,*,#], Cai Li[1,3*], Qiye Li[1,3], Bo Li[1], Denis M. Larkin[4], Chul Lee[5,6], Jay F. Storz[7], Agostinho Antunes[8,9], Matthew J. Greenwold[10], Robert W. Meredith[11], Anders Ödeen[12], Jie Cui[13,14], Qi Zhou[15], Luohao Xu[1,16], Hailin Pan[1], Zongji Wang[1,17], 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], Hao Yu[1], Jinmin Lian[1], Ping Wen[1], Fang Zhang[1], Hui Li[1], Yongli Zeng[1], Zijun Xiong[1], Shiping Liu[1, 17], Long Zhou[1], Zhiyong Huang[1], Na An[1], Jie Wang[1, 18], Qiumei Zheng[1], Yingqi Xiong[1], Guangbiao Wang[1], Bo Wang[1], Jingjing Wang[1], Yu Fan[19], Rute R. da Fonseca[3], Alonzo Alfaro-Núñez[3], Mikkel Schubert[3], Ludovic Orlando[3], Tobias Mourier[3], Jason T. Howard[20], Ganeshkumar Ganapathy[20], Andreas Pfenning[20], Osceola Whitney[20], Miriam V. Rivas[20], Erina Hara[20],Julia Smith[20], Marta Farré[4], Jitendra Narayan[21], Gancho Slavov[21], Michael N Romanov[22], Rui Borges[8,9], João Paulo Machado[8,23], Imran Khan[8,9], Mark S. Springer[24], John Gatesy[24], Federico G. Hoffmann[25,26], Juan C. Opazo[27], Olle Håstad[28], Roger H. Sawyer[10], Heebal Kim[5,6,29], Kyu-Won Kim[5], Hyeon Jeong Kim[6], Seoae Cho[6], Ning Li[30], Yinhua Huang[30,31], Michael W. Bruford[32], Xiangjiang Zhan[32,33], Andrew Dixon[34], Mads F. Bertelsen[35], Elizabeth Derryberry[36,37], Wesley Warren[38], Richard K Wilson[38], Shengbin Li[39], David A. Ray[26, §,], Richard E. Green[40], Stephen J. O'Brien[41,42], Darren Griffin[22],Warren E. Johnson[43], David Haussler[40], Oliver A. Ryder[44], Eske Willerslev[3], Gary R. Graves[45,46], Per Alström[47,48], Jon Fjeldså[46], David P. Mindell[49], Scott V. Edwards[50], Edward L. Braun[51], Carsten Rahbek[46,52], David W. Burt[53], Peter Houde[54], Yong Zhang[1], Huanming Yang[1, 57], Jian Wang[1], Erich D. Jarvis[20, #], M Thomas P Gilbert[3,55,#], Jun Wang[1, 56, 57, 58, 59,#] & Avian Genome Consortium

Consortium author list:

Chen Ye[1], Shaoguang Liang[1], Zengli Yan[1], M. Lisandra Zepeda[3], Paula F. Campos[3], Amhed Missael Vargas Velazquez[3], José Alfredo Samaniego[3], María Avila-Arcos[3], Michael D. Martin[3], Ross Barnett[3], Angela M. Ribeiro[8], Claudio V. Mello[60], Peter V. Lovell[60], Daniela Almeida[8,9], Emanuel Maldonado[8], Joana Pereira[8], Kartik Sunagar[8,9], Siby Philip[8,9], Maria Gloria Dominguez-Bello[61], Michael Bunce[55], David Lambert[62], Robb T. Brumfield[37], Frederick H. Sheldon[37], Edward C. Holmes[13], Paul P. Gardner[63], Tammy E. Steeves[63], Peter F. Stadler[64], Sarah W. Burge[65], Eric Lyons[66], Jacqueline Smith[53], Fiona McCarthy[67], Frederique Pitel[68], Douglas Rhoads[69], David P. Froman[70]

Affiliations:

1 China National GeneBank, BGI-Shenzhen, Shenzhen, 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 Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Republic of Korea
6 CHO & KIM Genomics, Seoul National University Research Park, Seoul 151-919, Republic of Korea
7 School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA
8 CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal
9 Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
10 Department of Biological Sciences, University of South Carolina, Columbia, South Carolina, USA
11 Department of Biology and Molecular Biology. Montclair State University, Montclair, NJ 07043
12 Department of Animal Ecology, Uppsala University, Norbyvägen 18D, S-752 36 Uppsala, Sweden.
13 Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
14 Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, Singapore 169857, Singapore
15 Department of Integrative Biology University of California, Berkeley CA 94720, USA
16 College of Life Sciences, Wuhan University, Wuhan 430072, China
17 School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China
18 BGI Education Center，University of Chinese Academy of Sciences，Shenzhen, 518083, China
19 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
20 Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC27710, USA
21 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK
22 School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK
23 Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Portugal.

24 Department of Biology, University of California Riverside, Riverside, CA 92521, USA
25 Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA
26 Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA
27 Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
28 Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, P.O. Box 7011, S-750 07, Uppsala, Sweden
29 Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea
30 State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing, 100094, China
31 College of Animal Science and Technology, China Agricultural University, Beijing, 100094, China
32 Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK
33 Key Lab of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101 China
34 International Wildlife Consultants Ltd., Carmarthen SA33 5YL, Wales, UK
35 Centre for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark
36 Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA
37 Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana 70803, USA
38 The Genome Institute at Washington University, St Louis, Missouri 63108, USA
39 College of Medicine and Forensics, Xi'an Jiaotong University, Xi'an, 710061, China
40 Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA
41 Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia
42 Nova Southeastern University Oceanographic Center 8000 N. Ocean Drive, Dania, FL 33004, USA
43 Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA.
44 Genetics Division, San Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, California 92027, USA
45 Department of Vertebrate Zoology , MRC-116 , National Museum of Natural History , Smithsonian Institution, P. O. Box 37012, Washington, D.C. 20013-7012, USA
46 Center for Macroecology, Evolution and Climate, the Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O, Denmark

47 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
48 Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden
49 Department of Biochemistry & Biophysics, University of California, San Francisco, CA 94158, USA
50 Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology Harvard University 26 Oxford Street Cambridge, MA 02138, USA
51 Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA
52 Imperial College London, Grand Challenges in Ecosystems & the Environment Initiative, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK
53 Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK
54 Department of Biology, New Mexico State University, Box 30001 MSC 3AF, Las Cruces NM 88003, USA
55 Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6102, Australia
56 Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark
57 Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia
58 Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China
59 Department of Medicine, University of Hong Kong, Hong Kong
60 Department of Behavioral Neuroscience Oregon Health & Science University Portland, OR 97239, USA
61 Department of Biology, University of Puerto Rico, Av Ponce de Leon, Rio Piedras Campus, JGD 224, San Juan, PR 009431-3360, USA
62 Environmental Futures Research Institute, Griffith University, Nathan, Queensland 4121, Australia
63 School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
64 Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig, H¨rtelstrasse 16-18, D-04107 Leipzig, Germany
65 European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK
66 School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson 85721, USA
67 Department of Veterinary Science and Microbiology, The University of Arizona, 1117 E. Lowell Street, PO Box 210090-0090, Tucson AZ 85721, USA
68 Laboratoire de Génétique Cellulaire, INRA Chemin de Borde-Rouge, Auzeville, BP 52627 , 31326 CASTANET-TOLOSAN CEDEX, France

69 Department of Biological Sciences, Science and Engineering 601, University of Arkansas, Fayetteville, AR 72701, USA
70 Department of Animal Sciences, Oregon State University, Corvallis, OR 97331, USA
§ Current Address: Department of Biological Sciences, Texas Tech University, Lubbock TX 79409, 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.

Author contribution

The leaders of each working group are indicated by '*'.

Project leaders:

Guojie Zhang, Erich D. Jarvis, M Thomas P Gilbert, Jun Wang

Sample collection and DNA preparation:

M Thomas P Gilbert*, Erich D. Jarvis, Jason T. Howard*, Alonzo Alfaro-Núñez, Paula F. Campos, Ross Barnett, Mads F. Bertelsen, Elizabeth Derryberry, Maria Gloria Dominguez-Bello, Michael Bunce, David M. Lambert, Robb T. Brumfield, Frederick H. Sheldon

Genome sequencing and assembly:

Bo Li*, Erich D. Jarvis, Zijun Xiong, Qiumei Zheng, Yingqi Xiong, Guangbiao Wang, Bo Wang, Jingjing Wang, Guojie Zhang*

Genome annotation:

Bo Li*, Erich D. Jarvis, Fang Zhang, Hui Li, Yongli Zeng, Shiping Liu, Long Zhou, Tobias Mourier, Eske Willerslev, Paul P. Gardner, Tammy E. Steeves, Peter F. Stadler, Sarah W. Burge, Guojie Zhang*

Genome size reduction and gene loss analysis:

Qiye Li *, Zongji Wang, Lijun Jin, Jinmin Lian, Ping Wen, Cai Li*, Luohao Xu, Hailin Pan, Mikkel Schubert, Ludovic Orlando, Rute R. da Fonseca, M. Lisandra Zepeda, Amhed Missael Vargas Velazquez, José Alfredo Samaniego, María Avila-Arcos, Michael D. Martin, Angela M. Ribeiro, Guojie Zhang*

Chromosome rearrangement and syntenic analysis:

Denis M. Larkin*, Darren Griffin*, Marta Farré, Jitendra Narayan, Gancho Slavov, Michael N Romanov, Qiye Li*, Zongji Wang, Lijun Jin, Pei Zhang, Guojie Zhang*

Evolutionary stasis in gene families:

Jay F. Storz*, Federico G. Hoffmann, Juan C. Opazo

Substitution rates and genomic conservation analysis:

Cai Li*, Luohao Xu, Hailin Pan, Haofu Hu, Jiang Hu, Jin Xiao, Zhikai Yang, Yang Liu, Qiaolin Xie, Hao Yu, Guojie Zhang*

Global *dN/dS* analysis:

Cai Li*, Luohao Xu, Hailin Pan, Guojie Zhang*

Vocal learning related analyses:

Erich D. Jarvis*, Ganeshkumar Ganapathy, David W. Burt*, Chul Lee, Heebal Kim*, Kyu-Won Kim, Hyeon Jeong Kim, Seoae Cho, Andreas Pfenning, Osceola Whitney,

Miriam V. Rivas, Erina Hara, Julia Smith, Claudio V. Mello, Peter V. Lovell, Cai Li*, Luohao Xu, Hailin Pan, Guojie Zhang*

Analyses of ossification genes:

Agostinho Antunes*, Rui Borges, João Paulo Machado, Imran Khan, Daniela Almeida, Emanuel Maldonado, Joana Pereira, Kartik Sunagar, Siby Philip

Alpha-keratins and beta-keratins:

Matthew J. Greenwold*, Roger H. Sawyer, Cai Li, Haofu Hu

Edentulism:

Robert W. Meredith*, Mark S. Springer, John Gatesy

Diet related enzymes:

Jie Cui*, Edward C. Holmes

Visual opsins:

Anders Ödeen, Olle Håstad, Per Alström, Zhiyong Huang, Yu Fan. Note: Anders Ödeen, Olle Håstad and Per Alström contributed equally to this analysis.

Sex-related and reproductive traits:

Qi Zhou*, Cai Li, Wei Yang, Na An, Jie Wang

Database management:

Chen Ye, Shaoguang Liang, Zengli Yan, Ganeshkumar Ganapathy, Eric Lyons, Qiumei Zheng

Sharing genomic or transcriptomic data:

Erich D. Jarvis, Ning Li, Yinhua Huang, Michael W. Bruford, Xiangjiang Zhan, Andrew Dixon, Wesley Warren, Richard K Wilson, Shengbin Li, David A. Ray, Richard E. Green, Jacqueline Smith, Fiona McCarthy, Frederique Pitel, Douglas Rhoads, David P. Froman

Advisors:

Stephen J. O'Brien, Warren E. Johnson, David Haussler, Oliver A. Ryder, Gary R. Graves, Jon Fjeldså, David P. Mindell, Scott V. Edwards, Edward L. Braun, Carsten Rahbek, Peter Houde, Yong Zhang, Huanming Yang, Jian Wang

Additional Acknowledgments:
C.R., J.F. and G.R.G. thank the Danish National Research Foundation for its support of the Center for Macroecology, Evolution and Climate. H.K. was supported by a grant from the Next-Generation BioGreen 21 Program (No.PJ009019), Rural Development Administration, Republic of Korea. P.A. was supported by the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists (No. 2011T2S04). S. J. O. was supported as Principal Investigator by Russian Ministry of Science Mega-grant no.11.G34.31.0068. De.M.L. was supported in part by BBSRC awards BB/K008226/1, BB/J010170/1, and by PL-Grid Infrastructure. M.N.R. is supported by BBSRC grant BB/K008161 awarded to D.G. and De.M.L. Da.M.L. was supported by the Australian Research Council. S.V.E. was supported in part by NSF grant DEB 0743616. P. H. was supported by NSF (DBI-0821806), NMSU Manasse Award and Smithsonian Scholarly Studies Award. E. L. B was supported by Smithsonian Scholarly Studies Award. We also thank A. Suh, C Weber, and H. Ellegren for contributing to the early discussion of this study.

Content

SM Text 1 Genome sequencing, assembly and annotation

DNA sample preparation

DNA was extracted from blood or tissue samples (see **Table S1**), using either the Genomic Maxi-tip protocol (Qiagen, Germantown, MD) for DNA isolation, or a phenol-chloroform based method, as described below.

a) Qiagen Genomic Maxi-tip protocol for DNA isolation from blood

Buffers C1, G2, QBT, QC, and QF were purchased from Qiagen as a Genomic DNA Buffer Set (cat# 19060).

50µl of whole blood was brought up to 7.5 ml with 1X PBS and mixed gently. 1 volume (7.5 ml) of ice-cold Buffer C1 and 3 volumes of ice-cold distilled water (22.5 ml) was added. The sample was mixed by inverting several times and then incubated for 10 minutes on ice. The sample was then centrifuged for 15 minutes at 1300 x g, after which the supernatant was poured out. 2 ml of ice-cold Buffer C1 and 6 ml of ice-cold distilled water were added. The sample was vortexed to resuspend the pelleted nuclei and then centrifuged again at 4°C for 15 minutes at 1300 x g. The supernatant was again discarded. 10 ml of Buffer G2 was added. The sample was resuspended by vortexing for 30 seconds at maximum speed. 200 µl of QIAGEN Proteinase K stock solution (20mg/ml) was added. The sample was then incubated at 50°C for 1 hour.

The Qiagen genomic-tip 500/G column (cat# 10262) was equilibrated with 10 ml of QBT Buffer, allowing the buffer to empty by gravity flow. The sample was vortexed for 10 seconds at maximum speed and applied to the column. The sample flowed through the column by gravity flow. The column was washed with 2 x 15 ml of QC Buffer. For samples with low flow rate, positive pressure was applied with an air tight syringe (sealed with parafilm). After washing, the sample was eluted with 15 ml of QF Buffer and transferred to a new tube. To the eluted sample 10.5 ml (0.7 volumes) of isoproponal at room temperature was added, and the percipitating DNA spooled with a glass rod. The DNA was then added to a tube with 4 ml of cold 70% ethanol. DNA was vortexed for 5 seconds and then centrifuged at >5000 x g for 10 minutes at 4°C. The supernatant was removed without disturbing the pellet. The sample was left to air-dry for 30 minutes. DNA was resuspended in 250 µl of 10 mM Tris·Cl, pH 8.5, then dissolved at 55°C for 2 hours with agitation. If the sample was not completely dissolved following this time, it was given extra agitation overnight at room temperature. The DNA was quantified using either a Nanodrop (Thermo Fisher Scientific: Waltham, MA), if processed at Duke, or Qbit (Invitrogen, Carlsbad, CA), if processed in Copenhagen. If the DNA concentration was over 500ng/µl, then additional 10 mM Tris·Cl, pH 8.5 was added to bring it to a concentration of 500 ng/ul or lower.

b) Phenol Chloroform Extraction of Genomic DNA from tissue samples

150mg of tissue was minced with a sterile razor and put in 50ml tube. 7.5ml of tail buffer (0.5% SDS, 0.1M NaCl, 0.1M EDTA, 0.05M Tris (pH8)) and 200µl of 20mg/ml Protease K solution (Bioline, Tautan, MA, cat# BIO-37084) was added. The tissue sample was lysed by incubating at 55°C overnight with agitation on a shaker. The lysed tissue sample was transfer to a phenol resistant tube, after which 7ml of phenol (equilibrated with Tris pH8) was added. The sample was inverted about 20 times then centrifuged for 10 minutes at 10,000 x g. 6ml of the (top) aqueous layer was transferred to a new tube containing 2ml of tail buffer. 7ml of phenol/chloroform was added and

centrifuged for 10 minutes at 10,000 x g. 6.3ml from aqueous phase was transferred into a new tube containing 700µl of 3M NaAc (pH 6.5). 7ml of 100% ethanol was added at room temperature and mixed by inversion to mix thoroughly. The sample was centrifuge for 10 minutes at 10,000 x g. All but ~200µl of the ethanol was then removed in order to not loose the DNA pellet. 10ml of 70% ethanol was added to the sample. The sample was agitated for 1 to 5 minutes on a shaker to further wash the DNA, then centrifuged for 10 minutes at 10,000 x g. The supernatant was subsequently removed after which the tube was left open to allow the ethanol to evaporate for 10 minutes. DNA was re-suspended in 250 µl of 10 mM Tris·Cl, pH 8.5. DNA was dissolved on a shaker at 55°C for 2 hours. If the DNA was not completely dissolved after this time, it was left to continue to dissolve at room temperature overnight. The DNA was quantified using the Nanodrop, and if the sample was over 500ng/l then additional 10 mM Tris·Cl, pH 8.5. was added to bring it to a concentration of 500 ng/ul or lower.

c) DNA sex determination by PCR and restriction digest

Sex determination was done using PCR as described in (92). In brief, 500ng of genomic DNA was amplified using the P8 (5'-TCTGCATCGCTAAATCCTTT-3') and P2 (5'-CTCCCAAGGATGAGRAAYTG-3') primers. The P2 and P8 primers amplify both the CHD-W gene, which is located on the W chromosome, and the CHD-Z gene, which is located on the Z chromosome. The PCR amplification was done in 30ul total volume. The PCR reaction consisted of 0.2ul of P8 (100uM), 0.2ul P2 (100uM), 15ul 2X Premix E ((Epicenter, Madison, WI) catalog# FSP995E), 0.4ul Taq Polymerase ((NEB, Ipswich, MA) catalog# M0273L), 500ng of genomic DNA, and filled to 30ul with water. The PCR conditions were as follows: 94°C hotstart for 4min, 94°C denature for 1 min, 49°C anneling for 1 min, 72°C extension, repeat cycles 32 times, and extend 5 min at 72°C. This amplification was followed by a HaeIII (Thermo Fisher Scientific), catalog# FD0154, digest. Digested PCR products were run on a gel. A single band indicates the specimen was male and a double band indicates the specimen was female.

d) DNA barcoding

DNA was barcoded using the mitochondrial cytochrome c oxidase I (COI) gene and using the method described in (93).

Genome sequencing

All new avian genomes (except budgerigar [*Melopsittacus undulatus*] and bald eagle [*Haliaeetus leucocephalu*s]) were sequenced at BGI using the Illumina HiSeq 2000 platforms. The budgerigar genome was sequenced at Duke University and Roch/454 with a combination of Roche 454 and Illumina HiSeq reads. The sequencing of the bald eagle was done at Washington University in St. Louis (WUSTL) using Illumina HiSeq technology. The budgerigar as well as common ostrich assemblies were enhanced further with optical maps, and these enhancements are reported separately in our companion studies (*20, 22*).

Our dataset consists of two coverage levels: high-coverage species (>50X), and low-coverage species (<50X) (see **Table S2**). For high-coverage avian genomes, pair-end libraries of 7 insert sizes (170bp, 500bp, 800bp, 2kb, 5kb, 10kb and 20kb) were constructed. For four avian species (Peking duck [*Anas platyrhynchos*, also reported in (*14*), downy woodpecker [*Picoides pubescens*], hoatzin [*Ophisthocomus hoazin*] and

white-throated tinamou [*Tinamus guttatus*]), libraries of some insert sizes were not constructed due to limited sample or their specific sequencing strategy. For low-coverage avian genomes, libraries of two insert sizes (500bp and 800bp) were constructed. The read length for small insert libraries (insert size < 2kb) was 100bp, and 49bp for large insert libraries. The sequencing depths for high-coverage avian genomes range from 50X to 160X, while the sequencing depths for low-coverage avian genomes range from 24X to 39X. Small variation in the sequencing and assembly methods exist for 9 species, 8 of them (except bald eagle) have details of sequencing in their relevant companion papers (pigeon [*Columba livia*] (*16*), Peking duck (*14*), peregrine falcon [*Falco peregrinus*] (*15*), Adelie penguin [*Pygoscelis adeliae*] (*94*), emperor penguin [*Aptenodytes forsteri*] (*94*), crested ibis [*Nipponia nippon*] (*95*), little egret [*Egretta garzetta*] (*95*), budgerigar (*20*)). For bald eagle, two small insert libraries (300bp, 400bp) and two large insert libraries (3kb and 8kb) were sequenced at WUSTL.

Genome assembly

For those species not detailed in companion papers, a closely similar genome assembly strategy was used. Prior to assembly we performed several quality control steps on the raw reads: (1) Removal of reads with more than 10% of ambiguous bases (represented by the letter N). (2) For the short insert size libraries (170bp to 800bp), removal of reads with more than 65% low quality (Phred score <= 7) bases / For the large insert size libraries (2kb to 20kb), Removal of reads with more than 80% low quality (Phred score <=7) bases). (3) Filtration of PCR duplications (reads are considered duplications when read1 and read2 of the same paired end reads are identical). (4) Reads with low frequency k-mers (the cut-offs were determined based on the 17mer-frequency) were also removed in order to minimize the influence of sequencing errors.

Cleaned reads of each species were passed to SOAPdenovo v1.05 (*96*) for *de novo* assembly. Firstly, SOAPdenovo constructed *de Bruijn* graph by splitting reads of short insert sizes into k-mers and then merging k-mers clipping tips, merging bubbles and removing low coverage links. Then the contigs, which exhibited unambiguous connections in *de Bruijn* graphs, were collected. We tried different k-mers (from 23-mer to 33-mer) to construct contigs and chose the k-mer with the largest N50 contig length. All usable reads were mapped back to contig sequences to construct scaffolds step by step, with at least 3 read pairs required to form a connection. We tried different cut-offs of read pairs for different libraries to join contigs into scaffolds. The assembly with largest N50 length was used to close gaps. The gap filling step was done by the intrinsic gap filling function of SOAPdenovo as well as Gapcloser v1.10 (a companion program released with SOAPdenovo) (*97*). The assembled sequences were aligned against NCBI bacteria database (ncbi_bacteria_20110803) (*98*), to filter out the contaminated sequences. For bald eagle, the sequence reads of 4 libraries (300bp, 400bp, 3kb and 8kb) were assembled using the SOAPdenovo v1.0.3 (*96*) at WUSTL, and contaminated sequences were removed by blasting against the NCBI nr/nt sequences.

The basic statistics of all avian assemblies are listed in **Table S2**. All the assemblies have similar genome sizes, ranging from 1.05Gb to 1.26Gb. All the high-coverage avian genomes have a N50 scaffold length of >1Mb, except for the assemblies of the white-throated tinamou (scaffold N50 of 242Kb, due to lack of data of 10kb and 20kb libraries)

and bald eagle (assembled at WUSTL). The scaffold N50 lengths for low-coverage avian genomes are from 30kb to 64kb. The N50 contig lengths for high-coverage avian genomes are from 19kb to 55kb (except bald eagle, 10kb), and 12kb to 20kb for low-coverage avian genomes.

Protein-coding gene annotation

Protein coding gene annotations of avian species

We principally used a homology-based method to annotate the protein-coding genes in the avian genomes by using Ensembl (*99*) gene sets (release 60) of chicken (*Gallus gallus*) (*11*), zebra finch (*Taeniopygia guttata*) (*13*) and human (*23*).

Because the quality of homology-based prediction strongly depends on the quality of the reference gene set, we carefully chose the reference genes for the annotation pipeline. This was required due to incorrectly annotated paralogs we noted between the publically available chicken and zebra finch genome annotations.

Firstly, we generated 12,484 orthologous gene pairs of chicken and zebra finch based on synteny in the chicken-zebra finch whole genome alignment from UCSC (*100*). Subsequently the protein sequences of the orthologous genes in chicken and zebra finch were blasted against human protein sequences, after which the aligned-rate was calculated between human genes and chicken/zebra finch genes (aligned rate = aligned length / human protein length). For each orthologous gene pair of chicken and zebra finch, if the chicken's aligned rate to the human gene was higher than the zebra finch's aligned rate, the chicken gene was chosen as the reference gene set for downstream annotation. Otherwise, the zebra finch gene was chosen. The zebra finch's gene was preferred if the aligned rates were equal or there was no human homolog, since most of the new avian species sequenced were Neoaves and so is zebra finch.

From this analysis, of the 12,484 genes, 7,832 zebra finch and 4,652 chicken genes from the orthologous gene pair dataset were chosen as the reference gene set. The remaining chicken and zebra finch genes (those not represented among the 12,484 orthologous gene pairs) were further added into the reference gene set. Furthermore, given the high-quality of the human genome gene annotations, the complete human gene set was also added into the reference gene set. In summary therefore, the reference gene set consists of 5 parts: 1) 7,832 zebra finch orthologous genes (named finch_ORTH); 2); a non-overlapping set of 4,652 chicken orthologous genes (named chicken_ORTH); 3) 4,987 zebra finch non-orthologous genes (named finch_NON-ORTH); 4) 4,032 chicken non-orthologous genes (named chicken_NON-ORTH) and 5) the human genes (named HUMAN).

The protein sequences of the reference gene set compiled above were used as templates for homology-based gene prediction for all of the newly assembled avian genomes. To get more comparable data, we also used this method to re-annotate the published avian genomes except chicken and zebra finch (these two birds were actually partially re-annotated, see below). The major steps of the annotation pipeline were as follows:

a) Rough alignment

We aligned protein sequences of the reference gene set to each genome by TBLASTN (*101*) with a E-value cut-off of 1e-5, and linked the blast hits into candidate

gene loci with genBlastA (*102*). We filtered out those candidate loci with homologous block length shorter than 30% of length of query protein.

b) Precise alignment

We extracted genomic sequences of candidate gene loci, including the intronic regions and 2000bp upstream/downstream sequences. The extracted sequences were used by GeneWise (v2.2.0) (*103*) to perform more precise alignment. The outcomes of GeneWise included the predicted gene models in the genome. Then we translated the predicted coding regions into protein sequences, and ran MUSCLE (v3.8.31) (*104*) for each pair of predicted protein and reference protein. We filtered out the predicted proteins with length of <30aa or percent identity of <25%, as well as the pseudogenes (genes containing frameshifts or >2 pre-mature stop codons).

c) Building a non-redundant gene set

The results of GeneWise could have redundant gene models at a single gene locus. To build a non-redundant gene set, we set a priority order to the 5 parts of the reference gene set (the priority order is: finch_ORTH > chicken_ORTH > finch_NON-ORTH > chicken_NON-ORTH > HUMAN). If a gene model predicted based on a higher priority reference gene overlapped with a gene model based on a lower priority reference gene (overlapping length > 100 bp), the latter would be removed from the final gene set. If two overlapping genes were predicted based on two same-priority reference genes, the one with higher percent identity to the reference gene would be retained.

d) Improving gene annotations of chicken and zebra finch

To obtain an improved annotation for downstream analyses, we also re-annotated some genes of chicken and zebra finch.

For chicken, 7,832 Zebra finch_ORTH proteins were used as reference genes to run Genewise to re-annotate the corresponding orthologs in the chicken genome, with the same procedure as that described in a) and b) of the previous section ”protein coding gene annotations of avian species”. For each orthologous gene pair, we used the re-predicted gene model to replace the chicken Ensembl gene model.

We also used the 4,652 chicken_ORTH genes as reference genes to re-annotate the corresponding orthologs in zebra finch, with the same procedure as that applied to chicken.

The statistics of gene annotations of all the avian genomes are listed in **Table S3**.

Protein coding gene annotations of outgroup species

We performed gene prediction for the green sea turtle (*Chelonia mydas*) (*105*) and American alligator (*Alligator mississippiensis*) (*24, 106*) with the same method as that for avian genomes. The gene sets of human (*23*) and green anole lizard(*Anolis carolinensis*) (*107*) used were downloaded from Ensembl. In some analyses performed during this study other non-avian genomes were used for comparative analyses. For the details of which genomes were used in a specific analysis, please refer to the method part of the corresponding analysis in this document.

Repeat annotation

The homology-based repeat annotation of all avian species were done by RepeatMasker-3.3.0 (*108*) (http://www.repeatmasker.org) based on the Repbase library (version: 20110419, with parameters “-nolow -no_is -norna -engine wublast -parallel 1”).

American alligator [*Alligator mississippiensis*] and green sea turtle [*Chelonia mydas*] were also re-annotated using the same method. The repeat annotations of other outgroup genomes were not re-annotated, and we used the available annotation files from Ensembl/NCBI/BGI.

We also used RepeatModeler-1.0.5 (*109*) (http://www.repeatmasker.org) to perform *de novo* repeat annotations for the avian genomes. Firstly, we ran RepeatModeler with default parameters to construct a *de novo* repeat library for each genome assembly. The *de novo* repeat library was then used by RepeatMasker-3.3.0 (with parameters "-nolow -no_is -norna -engine wublast -parallel 1") to predict repeats in each genome.

The Repbase-based annotations and the *de novo* annotations were merged into a union set. The overall annotated content of transposable elements (LINEs, SINEs, LTR elements and DNA transposons) ranges within 4-10% of the genome, except for the woodpecker which is higher at 22% (**Table S4**, **Fig. S1** to **Fig. S4**). The numbers for chicken and zebra finch we obtained were almost identical to previously published figures (*11, 13*), suggesting a high level of similarity between the RepeatMasker libraries used for annotation. The inflated transposable element content (22%) in woodpecker reflects a large number of LINE CR1 elements (18%, **Fig. S1**) , constituting a fraction of the genome almost comparable to human (LINEs~20%) (*23*).

The divergence between any TE sequence and the consensus sequence can be taken as a rough measure of age. **Fig. S5** to **Fig. S8** show the distribution of divergence for LINEs, SINEs, LTR elements, and DNA transposons, respectively. Divergences are shown for each avian genome, with the fraction of the genome shown on the y-axis, and percent divergence on the x-axis. Color codes for different TE families are shown on selected plots.

Annotation of long non-coding RNAs (lncRNA) with RNA-seq data

lncRNAs have been revealed to play important roles in many biological processes and have become a research hot spot in recent years. We made use of the chicken RNA-seq data published in (*110*) and several RNA-seq datasets under NCBI accession PRJNA247673 to annotate the lncRNAs in the chicken genome and investigate the conservation of the lncRNA candidates among the 48 avian species.

The pipeline of annotating the lncRNAs included the following steps: 1) We applied Tophat (v1.3.3) (*111*) to map RNA-seq reads to the chicken genome with default parameters and used Cufflinks (v2.0.2) (*112*) to assemble the mapped reads into transcripts with '-I 50000'. 2) Filtered the transcripts that had overlaps with chicken protein coding genes. 3) Assessed the protein-coding potential of transcripts by using CPC (0.9-r2) (*113*) and RNAcode (v0.3) (*114*). We ran CPC with default parameters on the transcripts, and removed the transcripts with SVM score larger than -1. And we ran RNAcode with default parameters on the 48-avian species multiple sequence alignments of transcripts' genomic locations, and removed the transcripts that overlapped with the potential coding exons (RNAcode p-value < 0.01) And 4) The remaining transcripts with length of >=200bp were considered as lncRNAs. In total, we identified 5,879 lncRNAs with length of >=200bp. The statistics for lncRNAs are summarized in **Table S5**.

SM Text 2 Genome size reduction

Comparison of gene features

All the avian genomes were found to have similar mean gene, CDS, exon, and intron lengths (**Fig. S9**). However, the mean values in high-coverage avian genomes are slightly bigger than those in low-coverage avian genomes. The mean intergenic lengths in low-coverage avian genomes are significantly smaller than those of high-coverage avian genomes. These findings are not unexpected, as the intergenic lengths are strongly associated with the scaffold lengths.

In addition to the human genome, we used 23 other mammalian genomes in this analysis (**Table S6**, **Table S7**). The genomic sequences and annotations of most mammalian genomes were downloaded from Ensembl (*99*), except 4 genomes from BGI (Chinese tree shrew [*Tupaia belangeri*] (*115*) , naked mole rat [*Heterocephalus glaber*] (*116*), David's myotis [*Myotis davidii*] and black flying fox [*Pteropus alecto*] (*117*)) and two genomes from NCBI (domestic Cat [*Felis catus*] (*118*) and domestic Pig [*Sus scrofa*] (*119*)). In addition to using whole gene sets for comparison, we also compared the gene features (gene length, CDS length and intron length) of each bird-mammal orthologous group (**Fig. S10**). The bird-mammal orthologs were built by using the human-chicken orthologs as the proxy. We also compared the intergenic lengths of avian genomes and mammalian genomes (**Fig. S10**, panel A). As the low-coverage avian genomes tend to have shorter intergenic regions, when comparing intergenic lengths we only considered the high-coverage avian genomes and published bird genomes and previously published Sanger method sequenced bird genomes.

Since bats are the only flying mammalian group, we would like to know whether gene sizes of bats are also shorter than other non-flying mammals. We compared gene sets of three bats (David's myotis and black flying fox, and microbat [*Myotis lucifugus*] (Ensembl release-70) (*99*) with that of 22 non-flying mammals (**Table S6**), and found similar reduction of intronic and intergenic regions in bats' genomes (**Fig. S11**).

Detecting deletion events with whole-genome alignments

To further investigate the genome size reduction in birds, we built whole-genome alignments for 4 birds (common ostrich, chicken, zebra finch and emperor penguin) and 3 non-avian reptiles (American alligator, green sea turtle and green anole lizard) with LASTZ (v1.01.50), chainNet (v2), MULTIZ (v11.2) (*120-122*), using the green anole lizard assembly as the reference. The four selected bird species represent three major clades, zebra finch and emperor penguin for the Neoaves lineage, chicken for the Galloanserae lineage, and common ostrich for the Palaeognathae lineage. Based on the 163Mb lizard-referenced alignments (including gaps within blocks), we identified the deletions within the alignment blocks using custom scripts. To reduce the alignment errors, we excluded the alignment blocks with aligned lengths of < 50bp. We defined the following criteria to assign a genomic segment as a deletion of a specific intermediate branch in the phylogenetic tree: 1) all the child branches of the investigated branch have lost this segment and 2) the closest sister branch of the investigated branch has this segment, and 3) at least one of the more basal branches has this segment. The results of identified deletions are shown in **Fig. S12.** We also calculated deletion rates for 1Mb

sliding windows (under the green anole lizard genome coordinate system) and then estimated the variation of the deletion rates across the genome.

Identification of large segmental deletions (lost syntenic blocks)

Four species, the common ostrich (representing the modern birds), the American alligator [*Alligator mississippiensis*], the green sea turtle [*Chelonia mydas*] and the green anole lizard [*Anolis carolinensis*] were used to identify large segmental deletions (lost syntenic blocks) in the bird ancestors, and the green anole lizard was used as the reference species. The green anole lizard was chosen as reference, because it is the only reptile species with chromosome-level genome assembly and phylogenetically close to birds. We first mapped the genes of the common ostrich, the American alligator, the green sea turtle to the green anole lizard gene sets, and identified one-to-one ortholog pairs between the green anole lizard and each of the other three species separately (producing three ortholog tables). All the genes were numbered based on their positions on a chromosome or a scaffold in order to obtain the gene syntenic information between two species. Then the ortholog tables were sorted according to the gene orders of the green anole lizard. We then calculated the ratio of missing orthologs (i.e. genes present in green anole lizard but absent in other species) using a sliding window of 5 genes with a step size of 1 gene, and screened the windows with ≥ 4 orthologs missed in ostrich but $\leq$ 2 orthologs missed in at least one of the two non-avian reptiles (i.e. American alligator and green sea turtle) relative to the green anole lizard as lost syntenic blocks. Finally, overlapping windows or windows separated by ≤ 10 genes were merged into larger windows to represent the final lost syntenic blocks in the common ostrich. In total, we identified 118 lost syntenic blocks in the common ostrich, encompassing a total of 1,241 green anole lizard genes. Similar analyses were also performed to identify lost syntenic blocks in the American alligator and the green alligator, which showed that much less syntenic blocks were lost in these two non-avain reptiles. The statistics of lost syntenic blocks are listed in **Table S8**, and more details about lost syntenic blocks can be found in **Table S9**. We mapped the positions of the lost blocks in the common ostrich to the lizard chromosomes, and found many of these blocks are enriched in the chromosome ends (**Fig. S13**).

Detecting gene losses in avian genomes

We performed an independent analyses of gene loss in birds, not based on synteny, but on BLAST analyses with orthologs.

a) Data sets used (human genes, 5 non-avian reptiles and 48 avian genomes).

To determine what genes are lost during the origin of modern birds, we mapped all the human genes to 5 non-avian reptile genomes and 48 avian genomes and identified genes that were present in non-avian reptiles (i.e. present in the ancestor of modern birds) but absent in all the 48 birds' genomes. We chose the genes of human as reference, as the human genes were annotated much better than other species, with more than 90% annotated human genes having intact ORF (contain start and stop codons) and most human genes having functional annotations, in contrast to less than 30% annotated genes having intact ORF in the genome assembly of the green anole lizard. The human gene set was downloaded from Ensembl (*99*) (release-60). After collapsing redundant genes and

removing mitochondrial genes, we finally obtained 20,683 human genes for further analysis.

The five non-avian reptile genomes, which are used for determining if a human gene is present in the ancestor of modern birds, were three crocodile species and two turtles (American alligator [*Alligator mississippiensis*], saltwater crocodile [*Crocodylus porosus*], gharial [*Gavialis gangeticus*], green sea turtle [*Chelonia mydas*], and Chinese softshell turtle [*Pelodiscus sinensis*]) (*105, 106*).

b) Identification of genes present in the common ancestor of modern birds.

Some human genes are lineage-specific, which may originate after the divergence between mammals and reptiles and might contribute to false positives in this analysis. Thus we first used the 5 non-avian reptile genomes to confirm which genes were present in the ancestor of modern birds. We mapped all the 20,683 human genes to the 5 non-avian reptile genomes with TBLASTN (blast-2.2.23) (E-value <= 1e-2), and determined gene structure of each potential reptile locus by GeneWise (v2.2.0) (*103*). Reptile loci that mapped by multiple human genes were collapsed by keeping the query with the highest GeneWise score. We then defined a gene was present in the ancestor of modern birds if it was present in at least one of the 5 non-avian reptile genomes, with an aligning rate > 50% relative to the query human protein and without any frameshifts or premature stop codons. With these criteria, we identified 14,607 human genes that were present in the common ancestor of modern birds.

c) Identification of genes lost in the modern birds.

To identify what genes were lost during the origin of modern birds, we mapped all the 20,683 human genes to the 48 avian genomes with TBLASTN (E-value <= 1e-2), and determined gene structure of each potential avian locus by GeneWise. Avian loci that mapped by multiple human genes were collapsed by keeping the query with the highest GeneWise score. Then we defined genes as lost during the origin of modern birds if this gene was absent in all the 48 avian genomes but present in at least one of the five non-avian reptile genomes. A gene was considered to be lost in a given bird if it fit one of the following criteria: a) had no predicted record; b) with predicted record but having an aligning rate < 30% relative to the query human gene; c) with frameshifts or premature stop codons. With these criteria, we finally identified 640 human genes that were present in non-avain reptiles but lost in the modern birds (**Table S10**).

d) Functional analysis for the lost genes.

Some biological characteristic of birds, such as oxygen affinity of avian blood, lightweight but strong skeleton, edentulism (toothless phenotype), immune and nervous systems, olfaction and vision of birds have previously attracted scientific attention. As about 90% of the human genes have been assigned GO annotation (*123*), we were able to link a gene to specific functions that we are interested in by searching GO items.

GO annotations of human genes were downloaded from Ensembl (release-66). Blood, bone/cartilage, immunity, nerve, olfaction, tooth and vision related GO items were manually collected from the Gene Ontology website (,http://www.geneontology.org/) by searching for related key words, and then lost genes with GO items of interest were classified into different categories.

The numbers of lost genes related to different biological characteristic of interest are blood (51 genes), bone/cartilage (10 genes), immunity (64 genes), nerve (117 genes),

olfaction (35 genes), tooth (3 genes) and vision (5 genes), muscle (14 genes), digest (6 genes), lung (5 genes), reproduction (36 genes) and ovary (2 genes) (**Table S11**).

SM Text 3 Conservative mode of genome evolution

Identification of homologous synteny blocks and evolutionary breakpoint regions

From the 45 newly sequenced bird species in this project, we used the genomes with scaffold N50 ≥2Mbp. We aligned 16 *de novo* sequenced and assembled genomes (common cuckoo [*Cuculus canorus*], peregrine falcon, American crow [*Corvus brachyrhynchos*], little egret, crested ibis, pigeon, hoatzin, golden-collared manakin [*Manacus vitellinus*], medium ground-finch [*Geospiza fortis*], downy woodpecker, Adelie penguin, emperor penguin, Anna's hummingbird [*Calypte anna*], Peking duck, budgerigar and common ostrich) and two previously sequenced and fully assembled bird genomes (turkey [*Meleagris gallopavo*], Turkey_2.01 and zebra finch, WUGSC 3.2.4) to the chicken genome (Gallus_gallus-4.0). We aligned the green anole lizard (AnoCar2.0) and the boa constrictor snake (snake_5C, (*38*)) to the chicken genome and used them as outgroups. The alignments were generated using the Satsuma Synteny program (*124*), and were cleaned from overlapping and non-syntenic matches and finally the homologous synteny blocks (HSBs) were defined using the SyntenyTracker (*125*). 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. Visualizations were made using the Evolution Highway Comparative Chromosome Browser (http://evolutionhighway.ncsa.uiuc.edu) (**Fig. S14**). The steps for the EBR analyses are as follows:

a) Classification of evolutionary breakpoint regions

EBRs were assigned to different phylogenetic lineages using the ExaML version of the total evidence nucleotide (TENT) tree (*5*) containing only the branches leading to species used in the EBR analysis (**Fig. S15**). Briefly, our automated algorithm uses the three sets of HSBs (detected at different resolution of genomic rearrangements) and tests all possible classifications given an established phylogeny. It first parses the phylogenetic relationships of all the genomes and estimates the probability of missing an EBR in each target genome for all resolutions (β). Then, using a Poisson-based simulation, the algorithm estimates the probability of EBRs from different target genomes to overlap randomly in the reference genome (R). Later, the algorithm tests the hypothesis for each EBR to belong to each phylogenetic node using appropriate β and R values. Finally, the positions of overlapping EBRs from different resolutions are compared and the narrowest intervals are estimated for each EBR. The algorithm re-estimates the probability of "merged" EBRs to belong to each phylogenetic node. The first and second most likely classification probabilities are compared to get a ratio of probabilities between the most likely hypotheses. Therefore, the higher the ratio is, the most likely the top classification is correct. The algorithm also detects reuse EBRs (i.e., these EBRs occurring in two

different lineages without a recent common ancestor) by comparing most significant classification probabilities with the rest of classification hypotheses for each EBR. The algorithm for the automated EBR identification, classification, and reuse discovery was implemented as a custom Perl script.

We ran the automated EBR classification for three resolutions (300Kbp, 100Kbp and 50Kbp) of HSB detection. A final merged dataset was obtained by using the 100Kbp resolution as a reference but incorporating the information of EBRs in all species at the positions present in the 100Kbp dataset (from the 50Kbp and 300Kbp resolutions). This step allows for the correction of some classification errors that could result from missing EBRs in some species in the 100Kbp set.

The reuse EBRs were classified, as the EBRs with the ratio between the first and the second classification <20 but >1, and a new ratio was calculated. Finally, we filtered out those EBRs where the ratio between the first and the second classification was <45 and <50% of the 21 species were used to classify the EBR (**Table S12**).

In the merged dataset we obtained a total of 2,384 EBRs and 1,919 EBRs were unambiguously assigned to phylogenetic nodes. From these, 23 were chicken-specific and 50 galliformes-specific (shared by chicken and turkey). We found 266 reuse EBRs in the studied genomes, representing ~13.5% of the total number of EBRs (**Table S12**). This is 1.7 times higher than previously reported in mammals (~8.0%, (*126, 127*)). However, in mammalian studies fewer species were included so far leading to potentially lower resolution of reuse EBRs detection. Taking into account that we used fragmented genomes (lacking a chromosome assembly) we estimated how the fragmentation could affect EBR detection. To do so, we calculated what fraction of the reference-specific EBRs (i.e. chicken- and galliformes-specific EBRs) could be detected in each target genome (except turkey and chicken because these genomes were used to as a basis to define the chicken- and gallifomes-specific EBRs). More than 80% of these EBRs were recovered in all the species, except for hoatzin (65.75%) (**Fig. S16**). A similar fraction of the reference-specific EBRs was detected in the genomes that had chromosome assemblies (zebra finch - 86.95% and Peking duck - 84.06%) demonstrating that genome fragmentation does not significantly affect EBR detection when the N50 of fragmented genomes is relatively high (> 3Mbp, **Fig. S15**). Since we knew what fraction of the reference-specific EBRs was not detected in each genome (except turkey) we calculated the expected number of EBRs in each lineage using the recovery rate of reference-specific EBRs. Thus, the expected number of EBRs is expected to compensate for the fragmented nature or alignment issues in some genomes (**Table S12**).

b) Estimation of the rate of rearrangements in avian genomes

Rates of chromosomal rearrangements (RR) were estimated based on the TENT avian tree (*5*). The number of detected and expected EBRs in each species was normalized by the branch length in million years (MY) leading to each lineage (**Fig. S15**).

Previous cytogenetic studies suggested that birds have a stable karyotype, since the majority of the avian species have a diploid chromosome number between 76 and 80 (*37*). Using the classified EBR set, we calculated the number of EBRs in all branches of the phylogenetic tree. We excluded the turkey genome from the analysis because of the potential assembly problems as shown in **Fig. S17**. We found that most of the branches have low RR (0.00 – 1.01 EBRs/MY) or medium RR (1.01-1.82 EBRs/MY). However,

some speciation events are accompanied with higher rates of rearrangements (> 1.82 EBRs/MY) (**Fig. S15**). Contrary to what has been shown in cytogenetic studies (*37*), the divergence between Paleognathae and Neognathae species ~98 MY ago was concurrent with high rate of rearrangements (~2.87 EBRs/MY). This is also the case of the diversification of Passeriformes, where RRs vary from ~1.27 to ~5.5 EBRs/MY. We also found high RR in penguins (~2.17 EBRs/MY and 2.53 EBRs/MY in emperor penguin and Adelie penguin, respectively), which could be related to their adaptation to extreme environment. In Psittacidae (parrots, represented in this analysis by budgerigar) we observed a high RR, possibly due to the high number of chromosomal fusions occurred in this lineage (2n = 60-70) (*37*). To compare the rearrangement rate between vocal-learners and non-vocal learners, we used the phylogenetic ANOVA function implemented in R "phytools" package (*128*), which controlled for phylogenetic relatedness. We used the time TENT tree from the avian phylogenomics paper (*5*) and set the number of simulations to 10,000. We performed post-hoc mean comparisons adjusting the p-value by the Holm method. Interestingly, vocal-learners have significantly more RR than non-vocal learners ($F=15.03$, $p=0.004$).

Using the classified EBR set, we calculated the number of EBRs in all branches of the phylogenetic tree. We excluded the turkey genome from the analysis because of the potential assembly problems as shown in **Fig. S17**. We found that most of the branches have low RR (0.00 – 1.01 EBRs/MY) or medium RR (1.01-1.82 EBRs/MY). However, some speciation events are accompanied with higher rates of rearrangements (> 1.82 EBRs/MY) (**Fig. S15**). Contrary to what has been shown in cytogenetic studies (*37*), the divergence between Paleognathae and Neognathae species ~98 MY ago was concurrent with high rate of rearrangements (~2.87 EBRs/MY). This is also the case of the diversification of Passeriformes, where RR varies from ~1.27 to ~5.5 EBRs/MY. We also found high RR in penguins (~2.17 EBRs/MY and 2.53 EBRs/MY in emperor penguin and Adelie penguin, respectively) which could be related to their adaptation to extreme environment. In Psittacidae (parrots, represented in this analysis by budgerigar) we observed a high RR, possibly due to the high number of chromosomal fusions occurred in this lineage (2n = 60-70) (*37*).

c) Diversification and rearrangement rates

In order to compare the rearrangement rates to previously published diversification rates (*7*), we recalculated the RR using the same TENT phylogenetic tree. The total number of EBRs occurring in each order or family was divided by the divergence time of the node and then, by the number of species analyzed (**Table S13**) providing an average number of EBRs per million year per genome. We could not recalculate some of the RR since the tree from Jetz et al. (*7*) has different relationships for some of the species and/or because we only had one species representing the whole order. Chromosomal rearrangement rates correlate positively with the diversification rates when the diversification rates are high. For example, the Passeroidea group (including the zebra finch, medium ground-finch and American crow) has a global diversification rate of 0.2 MY^{-1} and a RR of ~1.35 EBRs/MY, as well as Anseriformes (with Peking duck) with a diversification rate of 0.28 MY^{-1} and a RR of ~9.9 EBRs/MY (**Table S13**).

Our data shows that even though the overall rate of chromosomal rearrangements in birds is low, bursts of chromosomal rearrangements had occurred in avian genomes

during their evolutionary history, which might be related to bursts of speciation and/or adaptation.

Conserved gene synteny

To compare synteny of birds and mammals, we chose those avian genomes with long scaffold N50 sizes (>600kb) in this analysis and focused on the eutherian mammals, which are approximately the same evolutionary age, and whose genome assemblies are similar in quality. Gene sets of 22 birds and 22 mammals were used to identify syntenic blocks of any two birds and mammals, respectively.

In the beginning, we used an approach combining with synteny and reciprocal best hits (RBH) for ortholog detection. To maximize flexibility, this analysis was done in a pairwise fashion, based on pairwise alignments of any two birds/mammals using BLAST (v2.2.23) (*129*). The pairwise orthologous syntenic blocks (OSBs) are defined as collections of contiguous orthologous genes located on the same scaffold/ chromosome in each pair of avian/mammalian genomes. The first and last orthologous genes of the reference genome in each OSB were used to define the OSB boundaries. The following criteria were applied to get reliable OSBs: 1) The minimum OSB should contain at least contain 5 contiguous orthologous genes for each pair of genomes; 2) The number of inserted and deleted genes between adjacent orthologs should be <=2; 3) The localized scrambling resulting from micro-inversions/translocation or assembly defects also need to be taken into account. We set a maximum of 5 genes which were due to localized scrambling for a single locus.

When comparing syntenic percentages between birds and mammals, to obtain independent data points for each group, we used the software TARGETING (*130*) to select the independent species pairs.

Analysis of globin gene family evolution

We identified contigs containing α- and β-globin genes in the genome assembly of each species by using BLAST (*101*) to make comparisons with the α- and β-globin genes of chicken. We then annotated globin genes within the identified gene clusters of each species by using GENSCAN (*131*) in combination with BLAST2 (*132*) to make comparisons with known exon sequences, following (*91*). Due to incomplete sequence coverage in the genome assemblies of several species, there were some cases where it was not possible to ascertain the full extent of conserved synteny.

For purposes of comparison, we used a stochastic birth-death model of gene family evolution to estimate rates of gene turnover in the globin gene families of birds and mammals (*133*). As input for the CAFE 3 program (*134, 135*), we used ultrametric trees based on the well-resolved phylogenies of birds (the TENT tree in (*5*)) and eutherian mammals (*136*). The analysis of globin gene family evolution was based on complete α- and β-globin gene clusters for 22 species representing each of the major avian lineages, and 22 mammal species representing each of the major eutherian lineages.

Reduced genomic substitution rates

Whole-genome alignments were created from all 48 bird genomes, 3 reptile genomes (American alligator, green sea turtle and green anole lizard) and 18 mammal genomes (see **Table S14** for details of species used), as well as *Xenopus tropicalis* (*137*) using MULTIZ (v11.2) (*122*), yielding 603Mb (including gaps) chicken-referenced alignments. The 18 mammal species MULTIZ alignments were downloaded from UCSC (*100*) (hg18 data sets).

4-fold degenerate sites (4D sites) are generally considered to be neutrally evolving (*45*). With the whole-genome alignments, we used msa_view tool in the PHAST package (v1.2.1) (*59*) to extract 4D site alignments, based on the chicken gene annotations. The phyloFit program in the PHAST package (v1.2.1) (*59*) was used to estimate the phylogenetic tree, with a known tree topology as an input parameter (the avian tree topology was based on the TENT tree from (*5*)). The estimated phylogenetic tree is shown in **Fig. S18**, and the branch lengths are in units of substitutions per site. We calculated the root-to-tip subsitution rates from the latest common ancestor of all 48 birds to each bird lineage, and then divided the root-to-tip subsitution rates by the divergence time of latest common ancestor of all birds (101.64Mya, based on the time tree in avian phylogenomics paper (*5*)). We also calculated the root-to-tip subsitution rates from the latest common ancestor of all 18 placental mammals to each mammalian lineage, and then divided the root-to-tip subsitution rates by the divergence time of latest common ancestor of all mammals (93Mya, from the estimate in (*138*)). Thus the substitution rates used for subsequent analysis were in units of substitutions per site per million years. To compare the substitution rates between different groups, we used the phylogenetic ANOVA function implemented in R "geiger" package (v2.0.1, "phy.anova()" function in "geiger") (*139*), which controlled for phylogenetic relatedness when doing comparisons. We used the time tree from avian phylogenomics paper (*5*) as input tree for phylogenetic ANOVA analysis, and set number of simulations as 10,000.

We further did correlation analysis between substitution rates and trait data. Before doing the correlation analysis, we calculated the average substitution rates for the orders with more than one species. We performed Pearson's correlation tests for the original data points, as well as the phylogenetically independent contrast data points (**Fig. S19**). After controlling for the phylogenetic relationships between species using phylogenetically independent contrast with the 'ape' package (v3.0-5) in R (v2.10.0) (*140*), we found a positive correlation between the average substitution rate and the numbers of species per order (data from IOC world checklist (*1*), R=0.46, p=0.01, Pearson's correlation test with phylogenetically independent contrasts, **Fig. S19C**). To see if the high substitution rate of Passeriformes could bias the correlation analysis, we also did Pearson's correlation tests using the data without passerines (**Fig. S19E-H**). After excluding Passeriformes, we got a p-value larger than 0.05 but still close to 0.05 (R=0.33, p=0.08, **Fig. S19G**). In addition, we used the "plot.lm()" in R (v2.10.0) (*141*) to generate the "Residuals vs Leverage" plots to see if there was any highly influential point which could bias the analysis. Our results revealed that all the data points used for correlation analyses had a Cook's distance of less than 0.5 (**Fig. S19B, D, F and H**), suggesting there was no highly influential point in these tests. We also used 'outlier.test()' in R (v2.10.0) "car" package (*141, 142*) to check if there was any outlier in the data points used for correlation analysis, but did not

find any outlier data point. Overall, we think the positive correlation between average substitution rate and the numbers of species of each order is reliable.

SM Text 4 Selective constraints on functional elements

Detecting highly conserved elements (HCEs)

PhastCons (*45*) is a widely-used program to identify highly conserved elements (HCEs) based on the multiple-genome sequence alignments. Firstly, we extracted the 4-fold degenerate sites from the 48-avian genome alignments and used phyloFit (in v0.9.9.10b PHAST package) (*59*) to estimate a neutral phylogenetic model (also considered as the nonconserved model in PhastCons), using a temporary topology (not the TENT tree) from avian phylogenomic project (*5*) as the guide tree. Because we focused on finding the conserved elements in all species, not within a substree of these species, the unresolved topology had little impact on the downstream analyses, as predicted based on (*143*). Then we ran PhastCons (in v0.9.9.10b PHAST package) to estimate conserved and nonconserved models. Then we predicted conserved elements and conservation scores based on the conserved and nonconserved models. Finally we identified 7,097,338 avian HCEs, covering 112Mb in the chicken genome (**Table S15**).

To compare the genomic conservation between birds and mammals, we also downloaded the 18-way placental mammal whole genome alignments from UCSC (*100*). (The same 18 mammalian species listed in **Table S14).** To keep consistency and make the results more comparable, we also ran PhastCons to predict the HCEs in the 18 placental mammal genome alignment, instead of using the existing HCEs provided on the UCSC FTP. 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). Finally, we identified 1,912,729 mammalian HCEs, covering 89Mb in the human genome (**Table S15**). Using the human-chicken alignment, we identified ~100Mb genomic regions that have conservation scores for both birds and mammals, and then generated the density map for comparison of conservation between birds and mammals.

Non-coding conserved regions usually contain regulatory elements, i.e. the transcription factor binding sites (TFBSes). We downloaded the human TFBS data from the UCSC data set generated by the ENCODE project (*51*) (ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeRegTfbsClustered/wgEncodeRegTfbsClustered.bed.gz) to compare the conservation degrees of these functional elements in mammals and birds. We found that of 1,582,526 human TFBSes used in analysis, 617,287 (39%) are conserved in mammals, 46,717 (3%) conserved in birds, and 31,075 (2%) conserved in both birds and mammals, and 15,642 conserved only in birds (**Table S16**).

To determine if there are some unannotated exons in the HCEs, we used Augustus (v2.5.5) (*144*) to do *de novo* gene prediction on the chicken genome, and identified the new exons that are overlapping the HCEs and not in previous annotations. We made use of the available transcriptome data to see how many of these exons are supported by expression data. The statistics of the results are shown in **Table S17**.

Of the annotated 5,879 lncRNAs with length of >=200bp, 220 lncRNAs have more than 50% overlapping sequences with avian HCEs (**Table S18**). We aligned the lncRNA candidates against the lncRNAdb database (*145*) using BLAST (v2.2.23) (*101*) (E-value < 1e-5), and found only one mapped well to the database (224 lncRNA sequences in total). The matched lncRNA is called 'cyrano' (mapped to cyrano_Human with E-value of 1e-37). The chicken cyrano lncRNA is very intriguing, as it is highly conserved among birds (**Fig. S20**) and it is known to be required for normal embryonic development in zebrafish [*Danio rerio*] (*146*). We also used the RNAfold tool in Vienna RNA server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi, (*147*)) to predict the secondary structure of the chicken cyrano gene (**Fig. S21**).

We also sought to find some genes that are very conserved in birds but divergent in mammals in terms of the coverage by HCEs in the coding regions. Based on the human-chicken ortholog, we identified 13 genes whose coding regions are >75% covered by avian HCEs and <25% covered by mammalian HCEs (**Table S19**).

The enriched GO categories for the genes near shared TFBSes and genes near avian-specific conserved TFBSes are listed in **Table S20** and **Table S21**.

Natural selection on genes

a) Comparison of the evolutionary rates (*dN/dS* ratios) between birds and mammals

To perform the *dN/dS* analysis, we obtained 8295 orthologs from the 48 avian genomes, and the corresponding CDS alignments from the avian phylogenomics project (*5*). With the ortholog alignments, we ran one-ratio branch model with CODEML in PAML (v4.4) (*148*) to estimate the overall *dN/dS* ratios for each orthologous group.

We furthermore generated 12,815 orthologs for 24 mammals species (see **Table S22** for details of species used) and generated their corresponding alignments using the same pipeline that was used for birds. Then we ran PAML one-ratio branch model to obtain the *dN/dS* estimate of each mammalian orthologous group. Using human-chicken orthologs as the bridge, we obtained 5087 orthologous groups that are shared by birds and mammals. By using the GO annotation of chicken from Ensembl, we performed a Wilcoxon signed-rank test to test whether the *dN/dS* ratios of the genes of a certain GO are significantly larger (faster evolving) in birds or significantly larger in mammals. We found 174 GOs that have significantly larger *dN/dS* ratios in birds ($p<0.05$, **Table S23**), and 150 GOs that have significantly larger *dN/dS* ratios in mammals ($p<0.05$, **Table S24**).

b) Fast evolving genes in each of the three major avian clades (Palaeognathae, Galloanserae and Neoaves)

To identify the fast evolving genes in each of the three major avian clades, we used the one-ratio branch model, which estimated one identical *dN/dS* for all branches, and three-ratio branch model, which estimated three different *dN/dS* for Palaeognathae, Galloanserae and Neoaves.

We compared the one-ratio branch model and the three-ratio branch model using a likelihood ratio test (LRT). For a given othologous group, if the three-ratio model has the significantly higher likelihood than the one-ratio model, it indicates that that gene has different *dN/dS* in the three avian clades. The LRT p-values were adjusted by FDR correction for multiple testing. Finally, we found 2093 genes with FDR < 0.05, which can be divided into three categories: 1063 genes had largest *dN/dS* in Neoaves (considered as

fast evolving genes in Neoaves), 580 in Galloanserae (considered as fast evolving genes in Galloanserae) and 450 in Palaeognathae (considered as fast evolving genes in Palaeognathae). The GO enrichment results for the three categories of genes are shown in **Table S25**, **Table S26** and **Table S27**.

Genotype-phenotype associations in birds

a) Detecting convergent accelerated evolution of protein coding genes

To find genes with convergent accelerated evolution between the three avian vocal learner lineages and vocal non-learner avian lineages, *dN* (the rate of non-synonymous substitution), the *dS* (the rate of synonymous substitution) and $\omega = dN/dS$ was estimated along each branch of the phylogenetic tree. We obtained maximum likelihood estimates of ω under the branch model implemented in codeml within PAML (v4.6) with F3X4 codon frequencies (*148*) on a platform with 100 servers equipped with quad CPUs (2.6GHz Intel Xeon) and 16 GB RAM provided by KT GenomeCloud (www.genome-cloud.com). Log likelihood ratio test (LRT) was performed to compare the two models: null hypothesis with a fixed ω among all branches (model=0) and an alternative hypothesis, which allows for varied ω among branches (model=2) (**Fig. S22**). Orthologs with the following criteria were retained: $\omega \leq 5$, $dS \leq 3$, and $\omega_{foreground} > \omega_{background}$ (branches tested for positive selection is referred to as "foreground" branches and all other are referred to as "background" branches). FDR adjustments for multiple testing correction was applied with an adjusted p-value of 0.05. For the alternative hypothesis, two different parsimonious hypotheses (two independent gains and three independent gains) were tested to determine the ancestral branches that have undergone episodic adaptive evolution leading to convergence in the vocal-learning trait.

We applied this analyses to the 8,295 exons dataset of the Avian phylogenomics consortium (*5*) and were filtered using Gblocks (v0.91b) (*149*) to discard sites with poor alignment in order to eliminate amino acid substitution sites arising from misalignment of the data. Then, from the remaining 8112 orthologs, orthologs which did not include at least one species of the songbird and one species of the parrot were further filtered out (songbirds: medium ground finch, zebra finch, American crow; parrots: budgerigar, kea; and, hummingbird: Anna's hummingbird). Rifleman was excluded from the analysis because it is unknown whether rifleman is a vocal learner or vocal non-learner. After these filtering steps, a final set of 7909 orthologs sets was retained and this set was used for all further analyses.

b) Expression data

The 7909 orthologs sets were compared with expression data from four different sources, including companion studies of this paper:

1) Convergent specialized differential expression of genes in the songbird, parrot, hummingbird RA analog of vocal all three vocal learning lineages; over 130 genes (*57*)
2) Convergent expression list from songbird, parrot, hummingbird RA and human laryngeal motor cortex; 58 genes (*57*)
3) Specialized expression of genes in Area X of zebra finches, 498 genes (*57*)
4) Specialized expression of genes in HVC in zebra finches, 866 genes(*150*)
5) Differential expression of genes between song nuclei, 5167 transcripts (*56*)
6) Genes regulated in song nuclei by the act of singing, over 2700 transcripts (*56*)

7) Genes with expression in the songbird brain from multiple studies (*56, 57, 151*) and 241 genes from ZEBrA database (http://www.zebrafinchatlas.org; Mello and Lovell)
8) Genes expressed in Area X during rapid vocal learning of songbirds, 6 genes (*152*)

We used these gene sets and the values in the associated studies to also calculate the expected values of genes with expression in the songbird brain (all genes on the brain microarray in (*56*)), expressed in song nuclei at baseline (*56*), specialized in song nuclei for RA, Area X and HVC (*57*), and regulated by singing (*57*).

c) Gene Ontology analyses

Gene Ontology analyses were conducted for accelerated genes expressed in the songbird brain (165 genes, **Table S29**, **Fig. S24**), and the subset expressed in the song nuclei (151 genes, **Table S30**, **Fig. S25**) (*153, 154*).

d) Target species specific amino acid substitution (TAAS) analysis

Amino acid substitution sites, which are mutually exclusive between avian vocal-learner and avian non-vocal learners, were examined to investigate convergent evolution of vocal learning at the molecular level. These sites were termed 'target species specific amino acid substitution' (TAAS) and the concept is shown in detail in **Fig. S23**. The 7909 orthologs sets were converted into amino acid sequences and TAAS analysis was performed with the six avian vocal learners designated as the target species.

e) Detecting accelerated genomic regions in vocal learner genomes

phyloP in the PHAST package (v1.2.1) (*59*) can be used to detect non-neutrally evolving genomic fragments in a given set of branches based on the multiple sequence alignments. To identify the accelerated genomic regions in vocal learner genomes we initially estimated a neutral model based on the 4-fold degenerate sites in the multiple genome alignment. Subsequently, the neutral model was used as input of phyloP to detect the accelerated elements (--mode ACC in phyloP) in the specified lineage compared to the outgroups. Our pipeline has two steps: (1) at first we compared each of the 6 vocal learners with 9 non-vocal learner species (**Table S32**) to identify the accelerated elements in each vocal learner species; (2) based on the results of step (1), we identified the accelerated elements shared by 6 vocal learner species which were used for downstream analyses. We ran phyloP on all the 50bp and 100bp overlapping windows (a step size of 10bp) of the alignments. By comparing the neutral model with the non-neutral model, phyloP produced a p-value for each window, which represents the significance of non-neutrality. A p-value threshold of 0.05 was used in our analysis. 360 elements are found shared by all 6 vocal learners when using 100bp windows, while 462 elements are found shared by all 6 vocal learners when using 50bp windows (**Table S33**). We used phyloFit in the PHAST package (v1.2.1) to re-estimate the phylogenetic trees with alignments of all 100bp/50bp shared accelerated elements (**Fig. S26**). From **Fig. S26** we can see the accelerated element trees have increased substitution rates in the vocal learning lineages as expected. We classified the elements into 5 non-overlapping categories by annotation types: exonic, intronic, 5'/3' 10k flanking regions of genes and intergenic region. The accelerated elements show a similar composition pattern with that of all 15 species sequence alignments, though a higher fraction in intergenic regions (**Fig. 4B** in main text).

We identified the nearest gene for each element and considered them as the accelerated-element associated genes. Some elements were very distant from its nearest

gene (i.e. > 100kb) and the association relationships cannot be considered reliable, so we focused on the genes that are within 5'/3' 10kb range of the accelerated elements. Under this criterion, 278 genes were found associated with the accelerated elements. We did GO enrichment analysis for these genes but did not find any enriched GO terms. By looking into the expression data described above, we found 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**).

Sytl2 is one candidate gene with an accelerated element located in its 5' flanking region (**Fig. S28**). A previous study in mouse revealed that *Sytl2* has a slicing isoform predominantly expressed in brain (*155*). The conservation scores predicted by phastCons also indicate this accelerated element is divergent in vocal learners but conserved in non-vocal learners. Another case is the gene *Vwc2l* with the accelerated element located within the first intron (**Fig. S29**). *Vwc2l* is predominantly expressed in the adult brain and embryonic neural tissues and the inhibition of *Vwc2l* functions in zebrafish results in the impairment of neural development (*156*). We hypothesize that accelerated evolution in the non-coding regions of these genes may lead to changes in their gene expression in vocal learners.

We also wanted to see whether some accelerated elements are associated with the differentially expressed genes (39 genes, from (*57*)) which were identified by comparing gene expression in RA region to the surround region in vocal learners (zebra finch, budgerigar, and hummingbird). We found three differentially expressed genes (*TSHZ3*, *JAZF1* and *SH3RF1*) are associated with accelerated elements (**Table S35**, **Fig. S30** to **Fig. S32**).

SM Text 5 Evolution of ecologically relevant genes

Analyses of ossification genes

The gene list involved in the ossification was retrieved from quickGO (*157*). The retrieval of genes associated with ossification was obtained using the term ("bone") and limited to non-redundant term. Subsequently the sequences for each of these were obtained from the annotated 48 avian genomes. The sequences were translated to amino acids, aligned using the MUSCLE (v3.8) (*104*) implemented in SEAVIEW (v4.4.2) (*158*) then back translated to nucleotides. The aberrant sequences and/or sequences containing stop codons were removed from the alignment. For the mammalian dataset the phylogenetic trees were retrieved when available from Ensembl (*99*). Gene sequences of 39 mammalian species were used: human [*Homo sapiens*], gorilla [*Gorilla gorilla*], chimpanzee [*Pan troglodytes*], orangutan [*Pongo abelii*], Gibbon [*Nomascus leucogenys*], macaque [*Macaca mulatta*], marmoset [*Callithrix jacchus*], tarsier [*Tarsius syrichta*], mouse lemur [*Microcebus murinus*], bushbaby [*Otolemur garnettii*], mouse [*Mus musculus*], rat [*Rattus norvegicus*], kangaroo rat [*Dipodomys ordii*], squirrel [*Ictidomys tridecemlineatus*], Guinea pig [*Cavia porcellus*], pika [*Ochotona princeps*], rabbit [*Oryctolagus cuniculus*], tree shrew [*Tupaia belangeri*], sloth [*Choloepus hoffmanni*],

armadillo [*Dasypus novemcinctus*], lesser hedgehog tenrec [*Echinops telfairi*], elephant [*Loxodonta africana*], hyrax [*Procavia capensis*], hedgehog [*Erinaceus europaeus*], shrew [*Sorex araneus*], microbat [*Myotis lucifugus*], megabat [*Pteropus vampyrus*], horse [*Equus caballus*], cat [*Felis catus*], dog [*Canis familiaris*], panda [*Ailuropoda melanoleuca*], dolphin [*Tursiops truncatus*], pig [*Sus scrofa*], cow [*Bos taurus*], alpaca [*Vicugna pacos*], Tasmanian devil [*Sarcophilus harrisii*], wallaby [*Macropus eugenii*], opossum [*Monodelphis domestica*], platypus [*Ornithorhynchus anatinus*]. Multiple sequence alignments were built using the same strategy used in the avian genes.

For each gene we used CODEML implemented in PAML v4.7 to test the selection signatures in the avian and mammalian ossification genes, using 3 models (Model 0, 1 and 2) under the species tree phylogenetic assumption. Model 0 was used to test the global selective pattern observed in those genes. The nested comparison of the models M1a (nearly neutral) vs M2a (positive selection) using the LRT obtained in each model was used to assess the statistical significance of the comparison. The sites showing significant signatures of selection were retrieved after the post-hoc analysis (Bayesian Empirical Bayes) BEB, this post-hoc analysis accommodate error and therefore is better than Naive Empirical Bayes, which is a less reliable analysis particularly in smaller datasets. The results of PAML analysis are shown in **Table S36**.

The positively and negatively selected genes were submitted to DAVID (*154, 159*) for functional clustering and restricted to the top 10 results of each gene list (**Table S37**).

Alpha-keratins and beta-keratins

Alpha-keratins were downloaded via NCBI for the chicken, green anole lizard and human. The copy number and position of these sequences coincides with the results reported in (*160*). Genome searches for alpha-keratins in the avian genomic dataset and 4 reptiles (green anole lizard, green sea turtle, American alligator, saltwater crocodile [*Crocodylus porosus*]) were conducted using standalone BLAT v35 fast sequence search command line tool (*161*) with chicken and green anole lizard α-keratin sequences downloaded from NCBI (**Table S38**). Alpha-keratins of 4 mammals (human, opossum, house mouse and platypus) were obtained from (*160*). The copy number results are listed in **Table S39**.

Avian genome searches for beta-keratins were conducted using 18 avian β-keratin sequences (**Table S40**) from Swiss-prot database as query sequences with TBLASTN in BLAST+ and GeneWise (*103*). First, avian β-keratins were aligned to the genome of each bird by TBLASTN. For each aligned region, the most similar homolog was selected having a length of not less than 50% of the query protein. Selected aligned regions were extracted from the genome and gene model structures were predicted by GeneWise. For the American alligator and green sea turtle genome searches, we used the dataset from (*162*) as queries for BLAST+ searches. For the green anole lizard β-keratins, we employed the dataset from (*163*). The summary of these results are listed in **Table S41**.

The phylogenetic analysis of the green anole lizard, green sea turtle and American alligator β-keratins and the complete avian β-keratins (**Fig. S33**) was constructed using the RAxML rapid bootstrapping algorithm, WAG substitution matrix with a gamma model of rate heterogeneity and 1000 bootstrap replicates (*164*). Alignment was

performed using ClustalW2 followed by visual inspection (*165*). Annotation of the avian sequences and phylogenetic clades was performed using BLAST+ and the (*162*) dataset as the database (*129*).

Edentulism

Tooth specific genes in the bird genomes were identified by blasting (blastn; (*101*)) with crocodylian mRNA sequences taken from Genbank: Cuvier's dwarf caiman (*Paleosuchus palpebrosus*); *AMEL* (AF095568), spectacled caiman (*Caiman crocodylus*); *AMBN* (AY043290), Nile crocodile (*Crocodylus niloticus*); *ENAM* (GU344683). Complete *DSPP*, *MMP20*, and *AMTN* mRNA sequences for crocodylians are unknown. As a result, the American alligator genome was blasted (blastn; (*101*)) with Ensembl predicted green anole lizard sequences (*AMNT* LOC100554538; *MMP20*, ENSACAG00000021026; *DSPP*, ENSACAG00000012488). The annotated Ensembl green anole lizard *DSPP* sequence is not complete. House mouse (*Mus musculus*; NM_010080) and human (NM_014208) *DSPP* sequences were first used to blast the green anole lizard genome prior to blasting the American alligator genome in order to identify all of the un-annotated green anole lizard *DSPP* exons. *DSPP* exons were further verified with mRNA sequence data isolated from American alligator teeth (provided by the Crocodilian Genome Working Group).

Genes flanking the tooth-specific genes *MMP20* (*MMP27* and *MMP7*) and *DSPP* (*SPARCL1* and *DMP1*) were identified in the birds, green sea turtle, and American alligator genome by blasting (blastn) crocodyliform and bird Genbank (experimentally obtained)/Ensembl (predicted) mRNA sequences: *DMP1* Cuvier's dwarf caiman (AB185286), chicken [*MMP27* (NM_205000); *MMP7* (NM_001006278); *SPARCL1* (ENSGALT00000017778)].

All scaffolds with significant blastn hits (E<10) for the 10-targeted genes were subsequently visually inspected to verify the hit and then manually annotated using Geneious (Versions 5.6.5–R6-1) (*166*). The annotated American alligator scaffolds were then aligned to all corresponding bird scaffolds containing the same genes one at a time using the Mauve genome aligner (mauveAligner algorithm) (*167*) in Geneious (Versions 5.6.5–R6-1). Mauve was set to automatically calculate the seed weight and minimum Local Collinear Blocks. Furthermore, we assumed synteny of overlapping genes between genomes, which were subsequently verified by the scaffold alignments. Fine-tuning of the Mauve aligned annotated regions was then executed using Muscle (*104*) and/or MAFFT (*168*) (Geneious Versions 5.6.5–R6-1 default settings). In many cases the annotated scaffold Mauve alignments identified exons that blast searches failed to find, particularly for the small exons. Subsequently, each of the homologous annotated bird scaffolds were aligned to one another using the Mauve genome aligner followed by Muscle/MAFT refinements (same settings as stated above). Each of the aligned exonic regions was then visually inspected and all newly identified exons were annotated.

Identified exonic regions for each of the tooth-specific genes were then incorporated in a multi-species alignment in Geneious (Versions 5.6.5-6) using Muscle/MAFT. Inactivating mutations were visually identified.

Diet related enzymes

We employed tBLASTn (v2.2.23) to screen the avian genomes using four public *AGT* protein sequences: human (ENSP00000302620), chicken (XP_003641783), turkey (XP_003214151), and zebra finch (XP_002192415). We then extracted the matching sequences with high-identity (*e*-value < $1e^{-5}$) in each bird genome. The putative *AGT* gene structure was then inferred using GeneWise (2.2.0) (*103*). In all, 22 avian genomes were found to have the complete *AGT* genes, thus were appropriate for gene/evolutionary analysis (**Table S42**). The MTS (mitochondrial targeting sequence) regions of all 22 genes were extracted and aligned in MEGA5 (*169*). To support evolutionary changes regarding to diet, we performed positive selection test of the 22 *AGT* genes. Several pairs of models, M1 versus M2, M7 versus M8 and M8a versus M8, implemented in PAML4.7 (*148*) were used to detect the positively selected codons. The TENT tree from avian phylogenomics project (*5*) was used as the reference for the analysis.

We employed the tBLASTn (v2.2.23) to screen the avian genomes using the following *GULO* protein sequences available in GenBank: 1) the fish, white sturgeon [*Acipenser transmontanus*] (EF397521), 2) the amphibian, African clawed frog [*Xenopus laevis*] (NM_001095065), 3) the reptile, Chinese soft turtle [*Pelodiscus sinensis*] (HQ619721), 4) the mammals, gray short-tailed opossum [*Monodelphis domestica*] (XM_001380006), platypus [*Ornithorhynchus anatinus*] (XM_001521551), Tasmanian devil [*Sarcophilus harrisii*] (XM_003758870), Norway rat [*Rattus norvegicus*] (NM_022220), house mouse [*Mus musculus*] (NM_178747), Chinese hamster [*Cricetulus griseus*] (XM_003505165), European rabbit [*Oryctolagus cuniculus*] (XM_002709304), small-eared galago [*Otolemur garnettii*] (XM_003794019), African savanna elephant [*Loxodonta africana*] (XM_003412362), domestic dog [*Canis lupus familiaris*] (XM_543226), Leschenault's rousette [*Rousettus leschenaultii*] (HQ415789), great roundleaf bat [*Hipposideros armiger*] (HQ415790), giant panda [*Ailuropoda melanoleuca*] (XM_002914414), horse [*Equus caballus*] (XM_001492727), domestic pig [*Sus scrofa*] (NM_001129948), and cow [*Bos taurus*] (NM_001034043). We then extracted the matching sequences with high-identity (*e*-value < $1e^{-5}$) in each avian genome. The putative *GULO* gene structure was inferred using GeneWise. In all, 38 avian species have conserved complete *GULO* genes in their genomes (**Table S43**). All avian genes and non-avian genes were aligned using MEGA5, in a guidance of amino acid sequences. A codon-based ML method was used to estimate the ratio of nonsynonymous (*dN*) and synonymous (*dS*) substitutions using PAML4.7, the selection forces, shaped *GULO* genes. A free-ratio model, which allows the *dN/dS* ratio varying for different branch, was employed to generate an overall selection pattern shaping avian *GULO* evolution. The avian phylogenomic species tree (*5*) and a mammal species tree (*170*) tree was used to guide the analysis.

Visual opsins

Tblastn searches with protein sequences of the *Gallus gallus* opsins were performed on the CDS and genome databases of the Avian Phylogenomics Project (http://phybirds.genomics.org.cn/blast.jsp). *RH1* and *RH2* opsins were found in all of the avian genomes. However, for the *OPN1sw1*, *OPN1sw2* and *OPN1lw* opsins, the complete

coding sequence could not be detected in all species, with cases where only a small fragment of the gene were found (**Table S44**).

Opsins were aligned using Seaview 4.3.4 (*158*) and UGENE 1.11.5 (*171*). Published sequences of the same genes were used as references. The amino acid sequences were aligned manually against Bovine rhodopsin (Uniprot ID: P02699) (*172*) and various translated mRNA sequences retrieved from GenBank (**Table S45**). Sequences that were apparently not of opsin origin were removed (e.g. sequences with insertions and/or deletions that would shift the reading frame).

To identify species with significant levels of positive selection on the gene, the sequences were analyzed using the online Ka/Ks calculation tool at BCCS (http://services.cbu.uib.no/tools/kaks) using default parameters. Ka/Ks is the ratio of non-synonymous mutations to synonymous mutations, also called *dN/dS* or ω, that can be used as an indicator of the selection pressure on a protein coding gene. Ratios below one indicate stabilizing selection, while ratios greater than one are interpreted as positive selection. The tree in **Fig. S37** was used for the hierarchical comparison.

The opsin genes showed evidence of stabilizing selection with mean *dN/dS* values below 0.25. This applied to both mammals and birds for genes whose sequences were available from enough species to make the distinction (*OPN1sw1*, *RH2*, *OPN3*, *OPN4* and *OPN5*). The average *dN/dS* against all other species in the group was calculated for each species using the "kaks" function and subsampling from the larger group to get matched sample sizes for a Wilcoxon test. The procedure was repeated 100 times to calculate an average p-value for the gene. For three genes, *OPN3*, *OPN4* and *OPN1sw1*, there was a significant difference in selection pressure between birds and mammals. For *OPN3* the median was 0.09 for birds and 0.15 for mammals ($p<0.001$), whereas it was 0.2403 and 0.1339 ($p<0.001$), respectively, for *OPN4*. For *OPN1sw1* the result was non-significant due to the bimodal distribution of *dN/dS* values in birds. As discussed later, we identified a strong positive selection on the gene in the branch leading to the Passerida passeriforms. When the two Passerida species were removed from the analysis the difference between birds and mammals became significant (birds 0.14, mammals 0.21, $p<0.05$).

Significant positive selection was found in two taxa on different genes: (1) in *OPN1sw1* on the branch leading to Passerida (here represented by zebra finch and medium ground-finch), and (2) in *OPN1Rh2* on the branches within Sphenisciformes (the emperor and Adelie penguins)) (**Fig. S38**).

RH2 in the two penguins has a 12 bp deletion where the chromophore binding aa Lys296 is encoded in all known functional opsins. Further indication that the gene is non-functional in penguins can be seen when comparing the *dN/dS* profile, generated by sliding a 15 aa frame over the gene, with *dN/dS* profiles of the little egret, crested ibis and dalmatian pelican [*Pelecanus crispus*] (**Fig. S39**). The profile was generated using custom scripts, calling the "kaks" function from the seqinR package (*173*) on all species pairs within the penguin and northern fulmar [*Fulmaru glacialis*] group and little egret/crested ibis/dalmatian pelican group, respectively. Not only are there peaks of positive selection, but the median level is also elevated (emperor-Adelie penguin, 0.48; little egret-dalmatian pelican, 0.28), indicating relaxed selection among penguins compared to the others. Furthermore, the *dN/dS* values are as large in comparisons

between the penguins (red dots in **Fig. S39A**) as in comparisons between these and their sister,the northern fulmar. Accordingly, it appears that the relaxed selection in the penguins is the result mainly of substitutions accumulated independently in the two species rather than a common selection gradient driving towards a new functional optimum. By contrast, there is very little difference between the little egret and dalmatian pelican in the reference group (red dots in **Fig. S39B**).

Using the same method on the passeriform *OPN1sw1* gene, we found strong stabilizing selection interspersed by three short regions with positive selection (**Fig. S40**). Two of these positively selected regions include known spectral tuning sites that have been shown to be important for adjusting the wavelength of maximum sensitivity of the visual pigment (transmembrane helices II and VII; (*174*), (*175*)). As the Passerida passeriforms have radically shifted their spectral tuning from being a violet sensitive pigment to one mainly sensitive in the ultraviolet region of the spectrum (*79*), it is not surprising that the gene shows signs of positive selection around the tuning sites. However, we also found evidence of strong positive selection in transmembrane region IV, around amino acid-position 165–173 (Bovine rhodopsin numbering; aa:s LGVALPPWF in pigeon), suggesting that there is one or more unidentified, but apparently important, spectral tuning amino acid site located there.

Sex-related and reproductive traits

Based on Gene Ontology annotation, we chose 89 genes (**Table S46**) that may be involved in spermatogenesis, and 6 genes (**Table S47**) involved in oogenesis for analysis. The *dN/dS* ratios used in this section were generated using free-ratio branch models of PAML (v4.4) (*148*) on the orthologs of the 48 bird genomes generated. The two eagles of genus *Haliaeetus* (bald eagle and white-tailed eagle) whose *dN/dS* ratios tended to have bias (because they are very close to each other), were excluded from the analysis. Based on the estimates of free-ratio branch models, the root-to-tip *dN/dS* ratios were calculated for each species in each ortholog. The abnormal *dN/dS* estimates (*dN/dS* <0.001 or *dN/dS* >5) were excluded from analysis. Wilcoxon rank sum test was used to compare the *dN/dS* ratios of these sex-related genes and that of macro-chomosome genes (as genome background). The results are listed in **Table S48** and **Table S49**.

We obtained 15 genes that have previously been implicated in influencing avian plumage colors (*176-180*) (**Table S50**). For some plumage-related genes which were not included in the ortholog list used in the phylogenomics project, we obtained their orthologs with reciprocal best blast hits. The median *dN/dS* ratios of plumage genes and that of macro-chomosome genes are listed in **Table S51**. The color discrimination data of birds from (*181*) was used to do the correlation analysis with the *dN/dS* ratios. The color discriminability focuses on the signal received by other members of the same specie of birds comparing to the traditional methods of measuring sexual dichromatism. It measures transmission spectrum of the oil droplets contained within the color cones of birds with limited receptor noise. We performed Pearson's correlation tests for the original data points and also for the phylogenetically independent contrasts (controlling for phylogenetic relatedness). We found two genes (*GSTA2* and *SLC24A4*) showing significant negative correlations between color discriminability and *dN/dS* (see **Fig. S41** and **Fig. S42**).

Supplementary figures

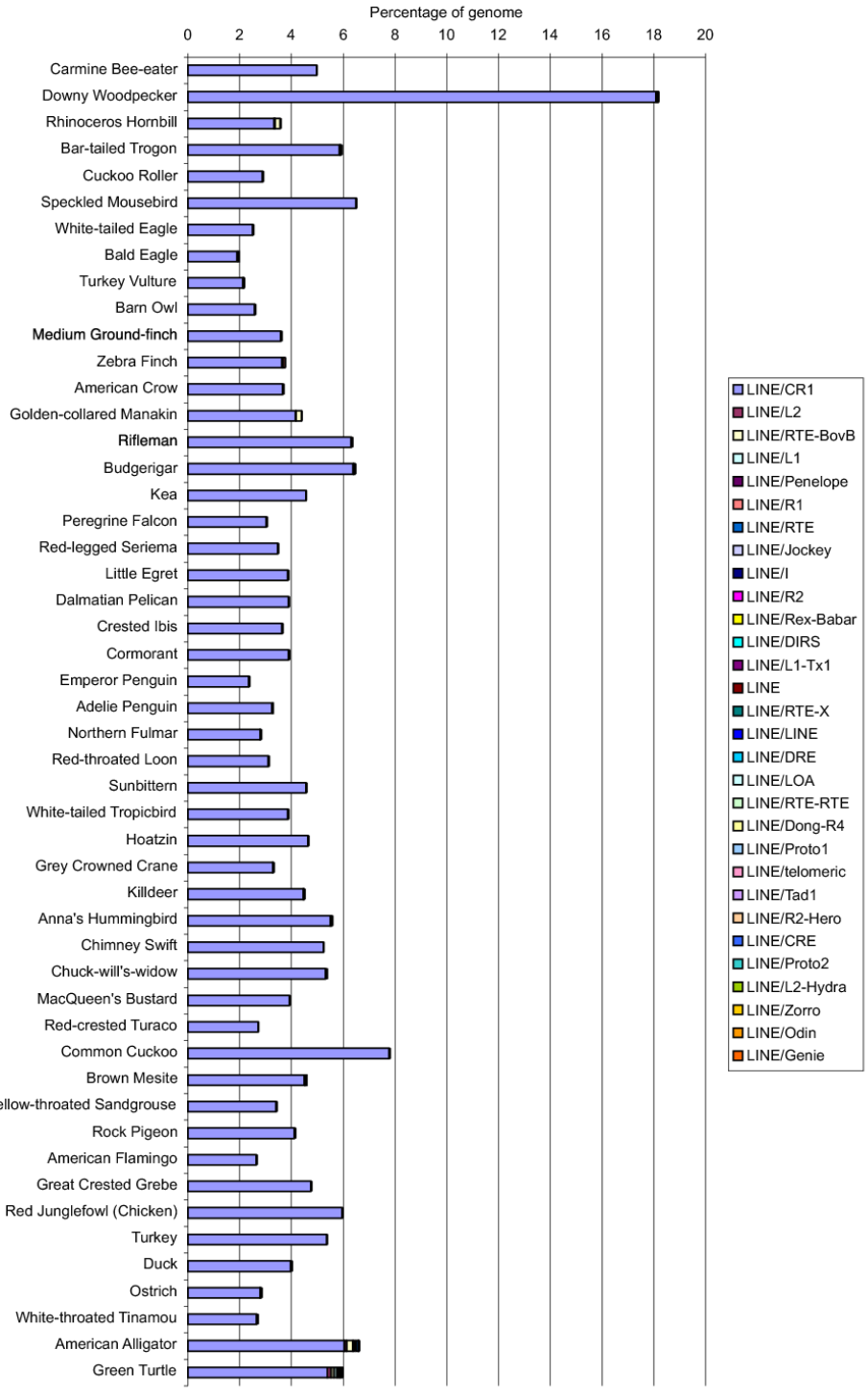

Fig. S1. The proportion of LINEs in the genomes of birds and outgroups. Species are ordered according to phylogenetic relationship in the avian TENT tree.

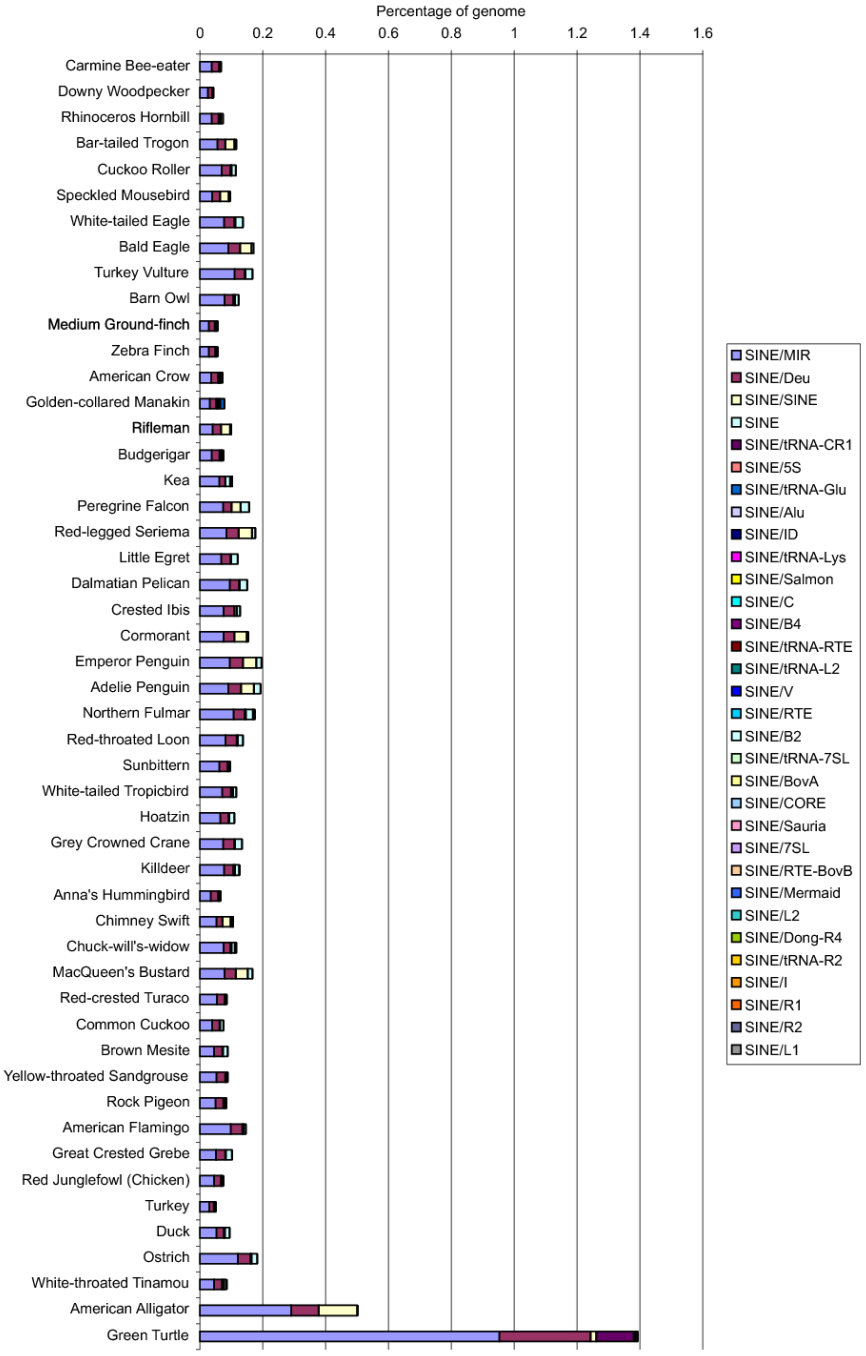

Fig. S2. The proportion of SINEs in the genomes of birds and outgroups. Species are ordered according to phylogenetic relationships in the avian TENT tree.

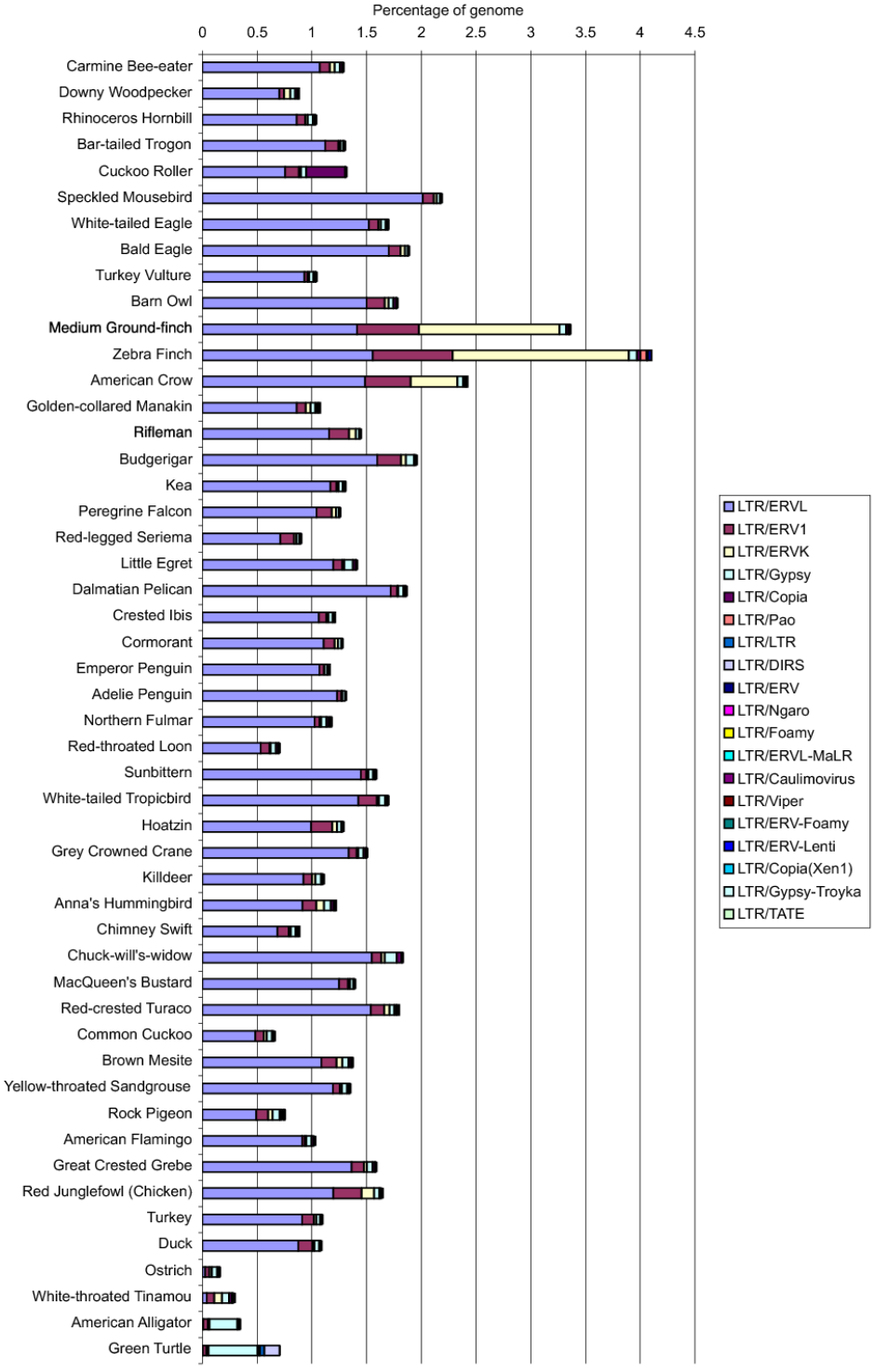

Fig. S3 The proportion of LTRs in the genomes of birds and outgroups. The species are ordered according to phylogenetic relationship of the avian TENT tree.

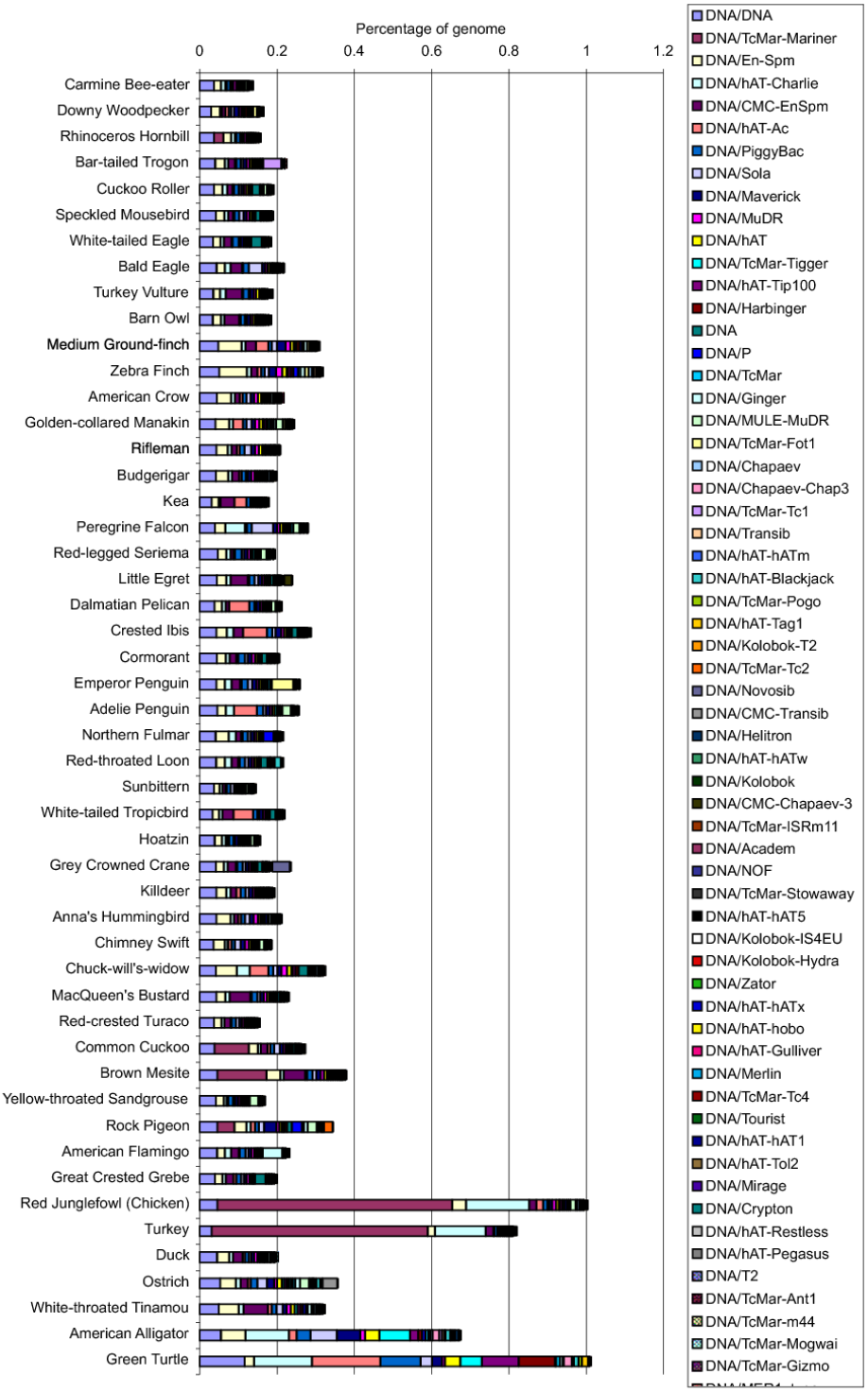

Fig. S4. The proportion of DNA transposons in the genomes of birds and outgroups. Species are ordered according to phylogenetic relationship in the avian TENT tree.

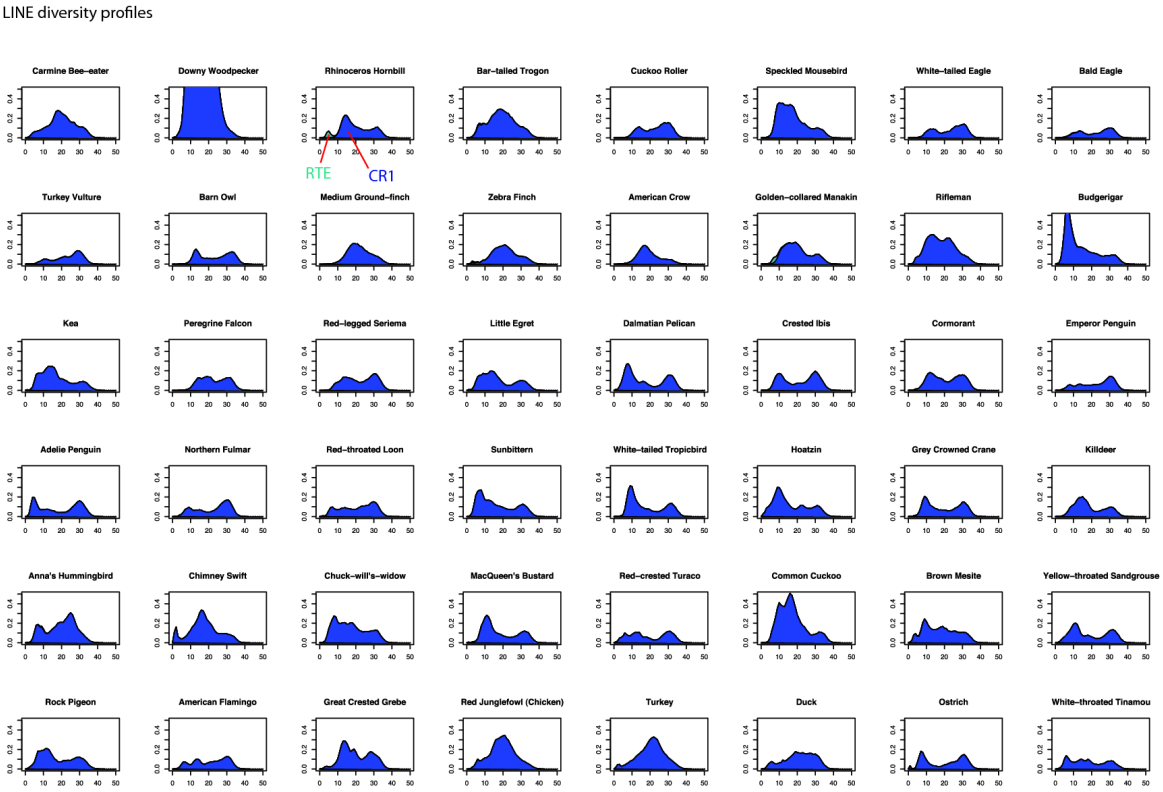

Fig. S5. The divergence profiles of LINEs in birds. Divergences are shown for each avian genome, with the fraction of the genome shown on the y-axis, and percent divergence on the x-axis.

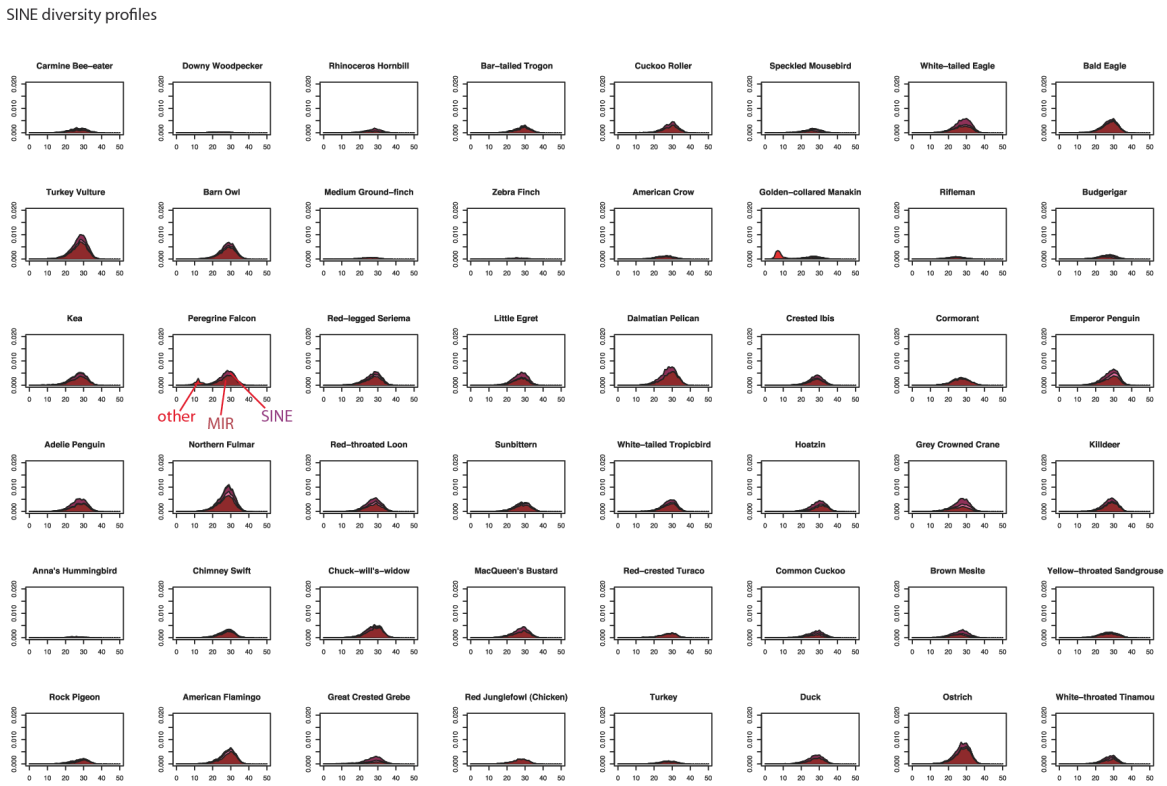

Fig. S6. Divergence profiles of SINEs in birds. Divergences are shown for each avian genome, with the fraction of the genome shown on the y-axis, and percent divergence on the x-axis.

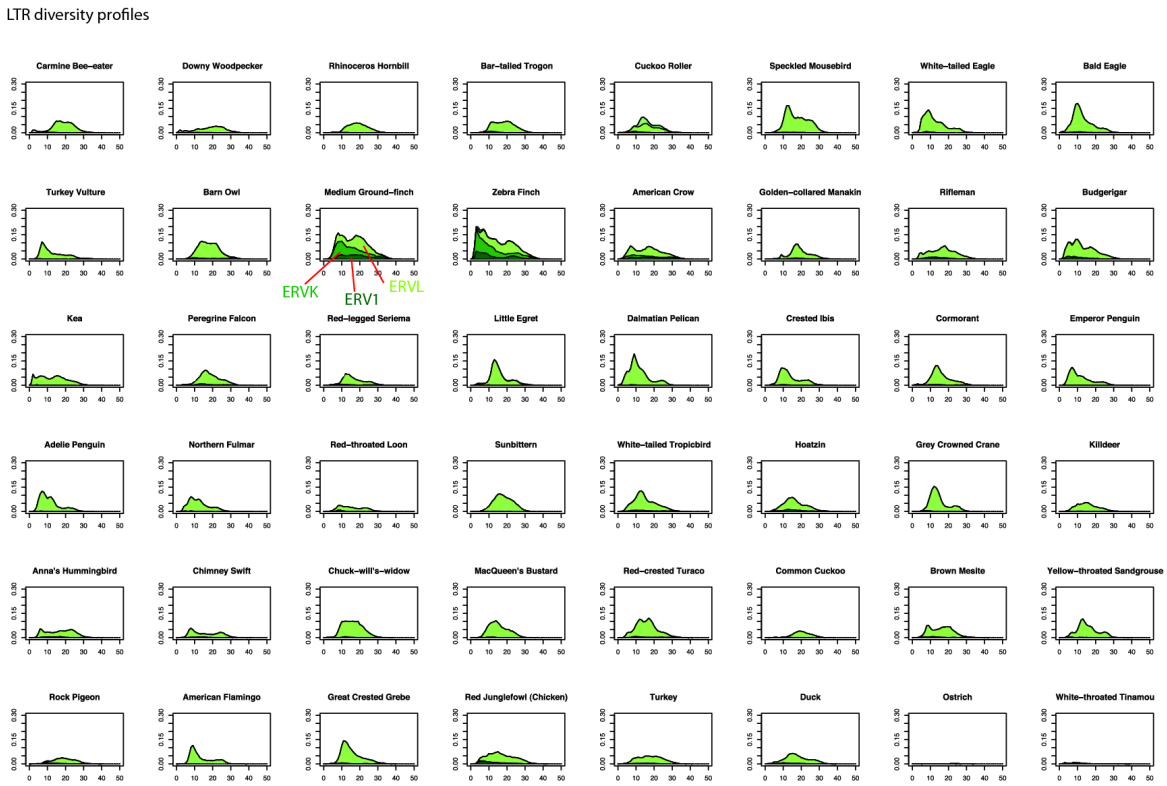

Fig. S7. Divergence profiles of LTRs in birds. Divergences are shown for each avian genome, with the fraction of the genome shown on the y-axis, and percent divergence on the x-axis.

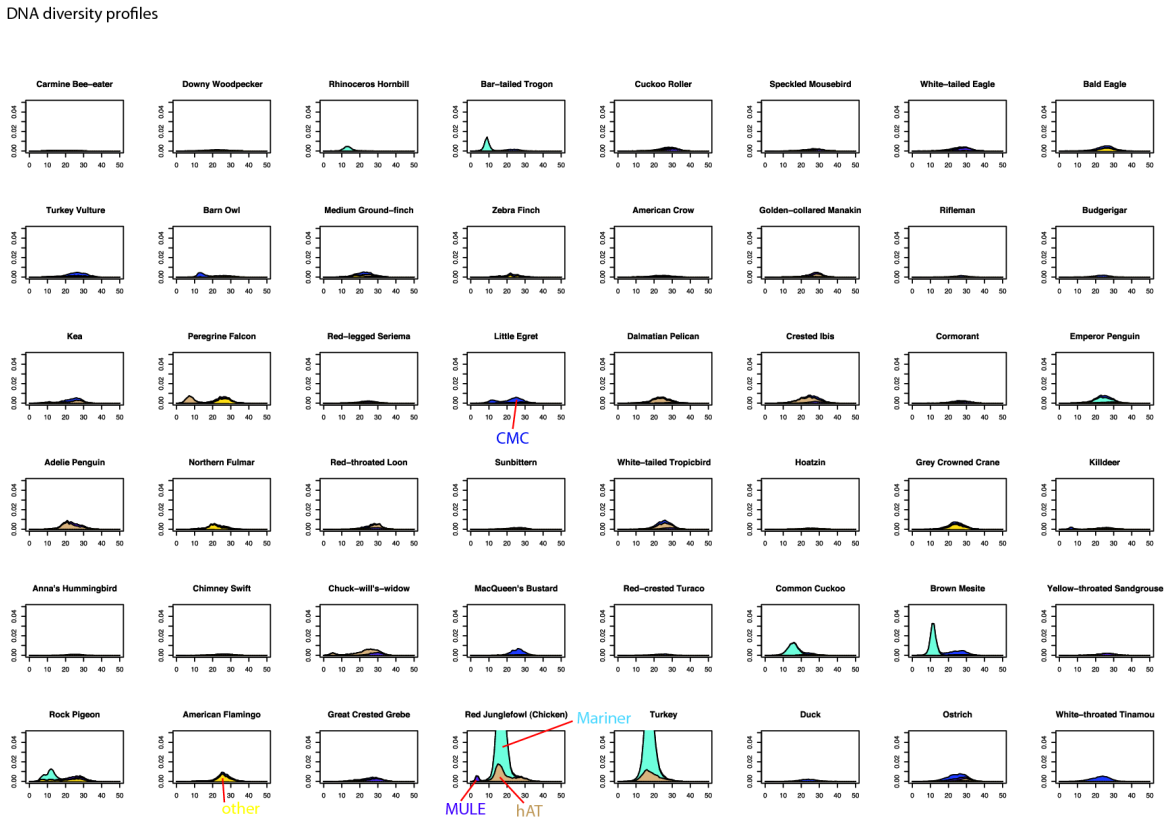

Fig. S8. Divergence profiles of DNA transposons in birds. Divergences are shown for each avian genome, with the fraction of the genome shown on the y-axis, and percent divergence on the x-axis.

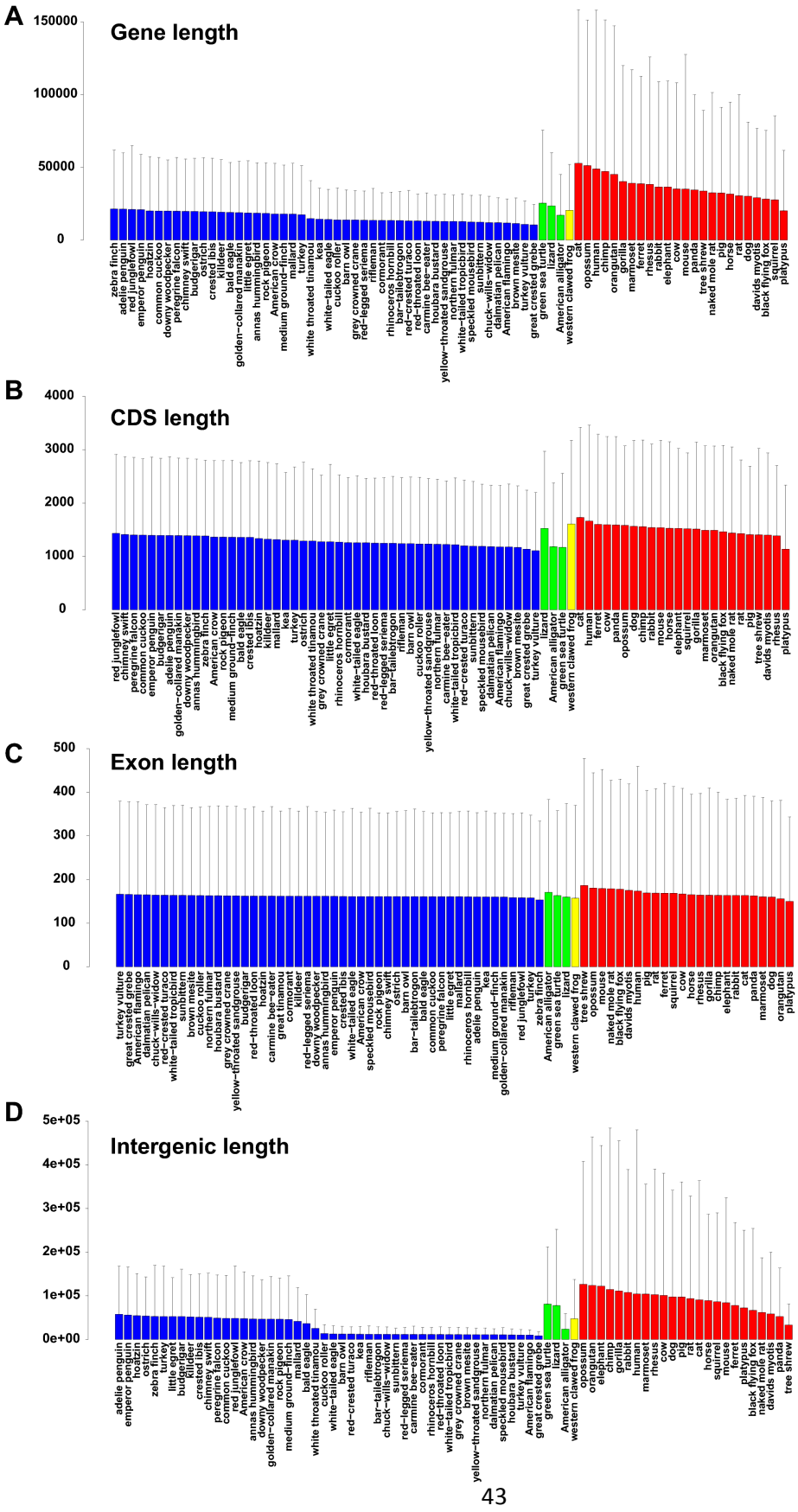
A
Gene length
150000
100000
50000
0
B
CDS length
4000
3000
2000
1000
0
C
Exon length
500
400
300
200
100
0
D
Intergenic length
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00

Fig. S9. Comparison of gene features between birds and outgroups. (A) gene length; (B) CDS length; (C) exon length; (D) intergenic length. Blue represents bird species, green for non-avian reptiles, yellow for amphibian, red for mammals. The error bars represent standard deviations.

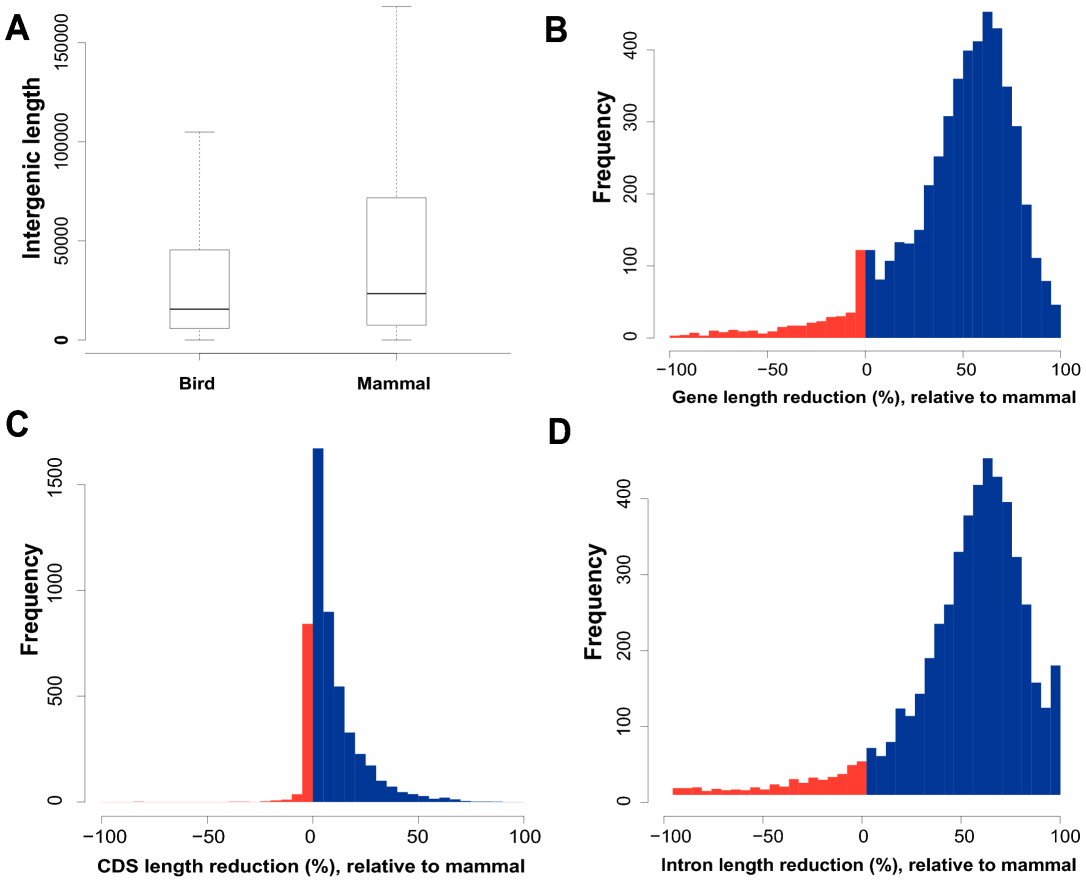

Fig. S10. Reduction of bird genomes. (A) Box plot of the lengths of intergenic regions. Median value of bird genes is about 66% of median value of mammalian genes. Degrees of reduction in length of bird genes (B), cds (C), introns (D) relative to their orthologs in mammals, expressed as a percentage: 100 (average mammalian gene length – average bird gene length) / (average mammalian gene length). The positive classes represent the shorter bird genes.

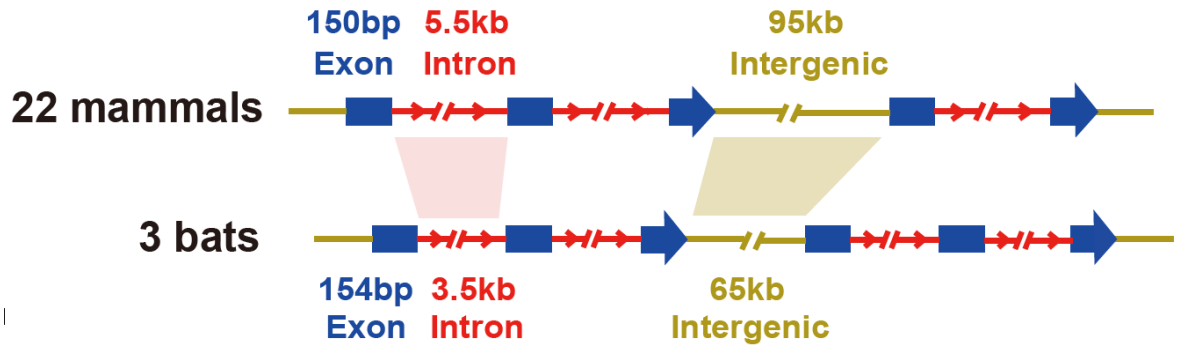

Fig. S11. Comparison of gene features between 3 bats and non-bat mammals.

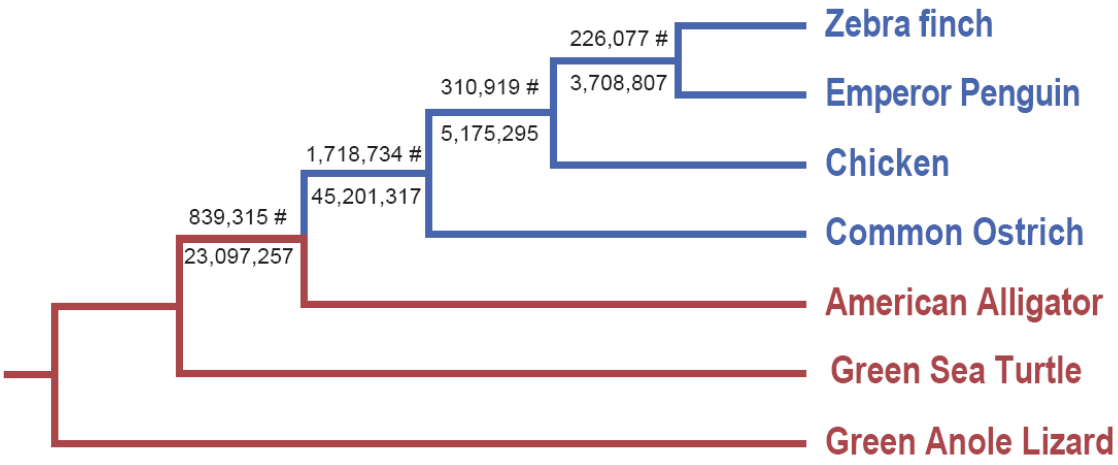

Fig. S12. Distribution of small deletion events across the phylogenetic tree from 163Mb of whole genome alignment regions. The upper values are the number of deletion events that occurred in each lineage. The lower values are the total lengths (bp) of the deletion events. Avian species are in blue, and non-avian reptiles are in red. The avian species were chosen to be representative of major clades: Paleognathae (ostrich), Galloanserae (chicken), Neoaves waterbird (penguin), and Neoaves landbird (penguin).

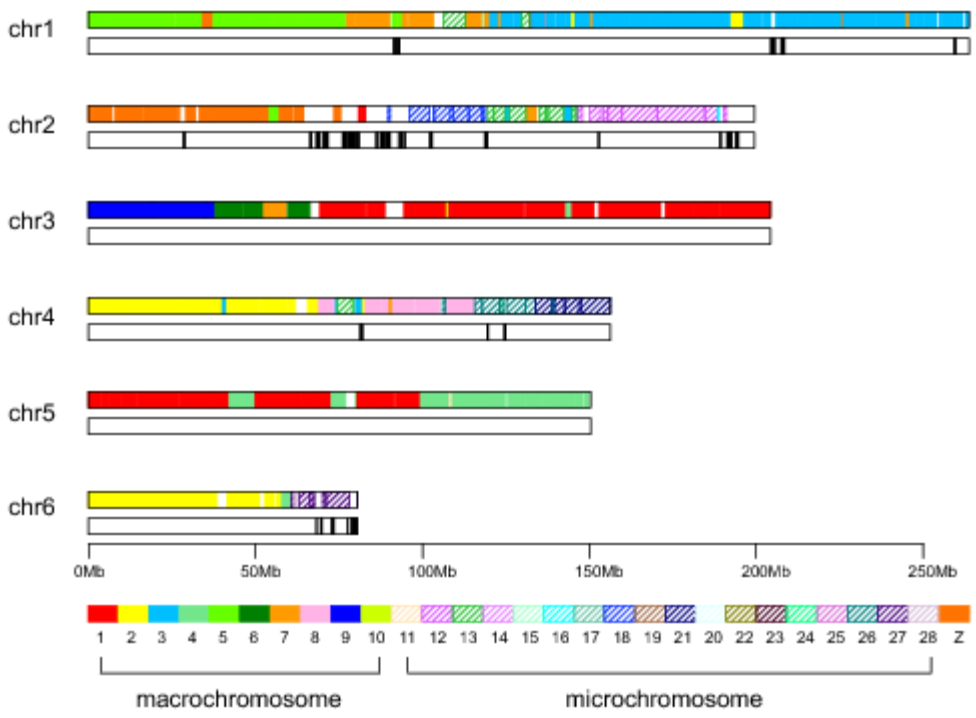

Fig. S13. Distribution of avian lost syntenic blocks across the green anole lizard genome. Chromosome numbers to the left are those of the green anole. The color bars in top rows are alignments from different chicken chromosomes (each color corresponds to a different chicken chromosome as indicated in the key). The black rectangles in bottom rows represent the syntenic blocks in lizards that are lost in birds.

GGA 11

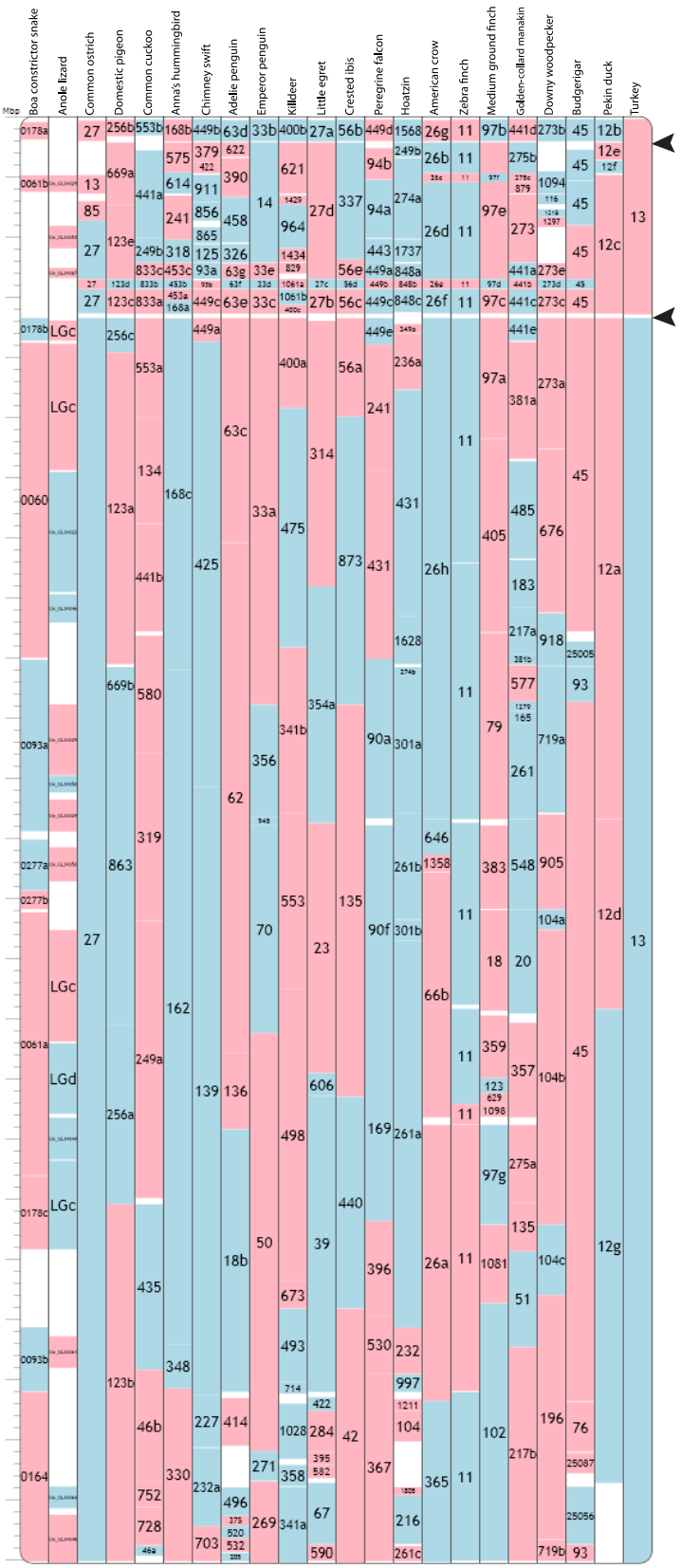
Mbp
Boa constrictor snake
Anole lizard
Common ostrich
Domestic pigeon
Common cuckoo
Anna's hummingbird
Chimney swift
Adelie penguin
Emperor penguin
Killdeer
Little egret
Crested ibis
Peregrine falcon
Hoatzin
American crow
Zebra finch
Medium ground finch
Golden-collard manakin
Downy woodpecker
Budgerigar
Pekin duck
Turkey

Fig. S14. An example showing evolutionary breakpoint regions (EBRs) and homologous synteny blocks (HSBs) in chicken chromosome 11. Blue blocks indicate HSBs and red blocks show inverted HSBs in the target genome, with the species scaffold or chromosome number inside the block. Note that the HSBs shown for the genomes with no chromosome-level assembly (common cuckoo, peregrine falcon, American crow, little egret, crested ibis, pigeon, hoatzin, golden-collared manakin, medium ground-finch, downy woodpecker, Adelie penguin, emperor penguin, Anna's hummingbird, Peking duck, budgerigar and common ostrich) may not represent complete HSBs found in species' chromosomes but be fragmented syntenic fragments found within scaffolds. Therefore not all breakpoint regions found in such genomes will represent EBRs. We applied a computational algorithm to detect EBRs in fragmented genomes (see supplementary methods). The lower case letters indicate the sequential order of HSBs in scaffolds from fragmented genomes. White areas between HSBs represent EBRs for genomes with chromosome assemblies and breakpoint regions (including EBRs) for fragmented genomes. Arrowheads show reference-specific EBRs. The EBR at 0.33 – 0.35 Mbp position is galliformes-specific and the EBR at 2.64-2.70 Mbp position is chicken-specific.

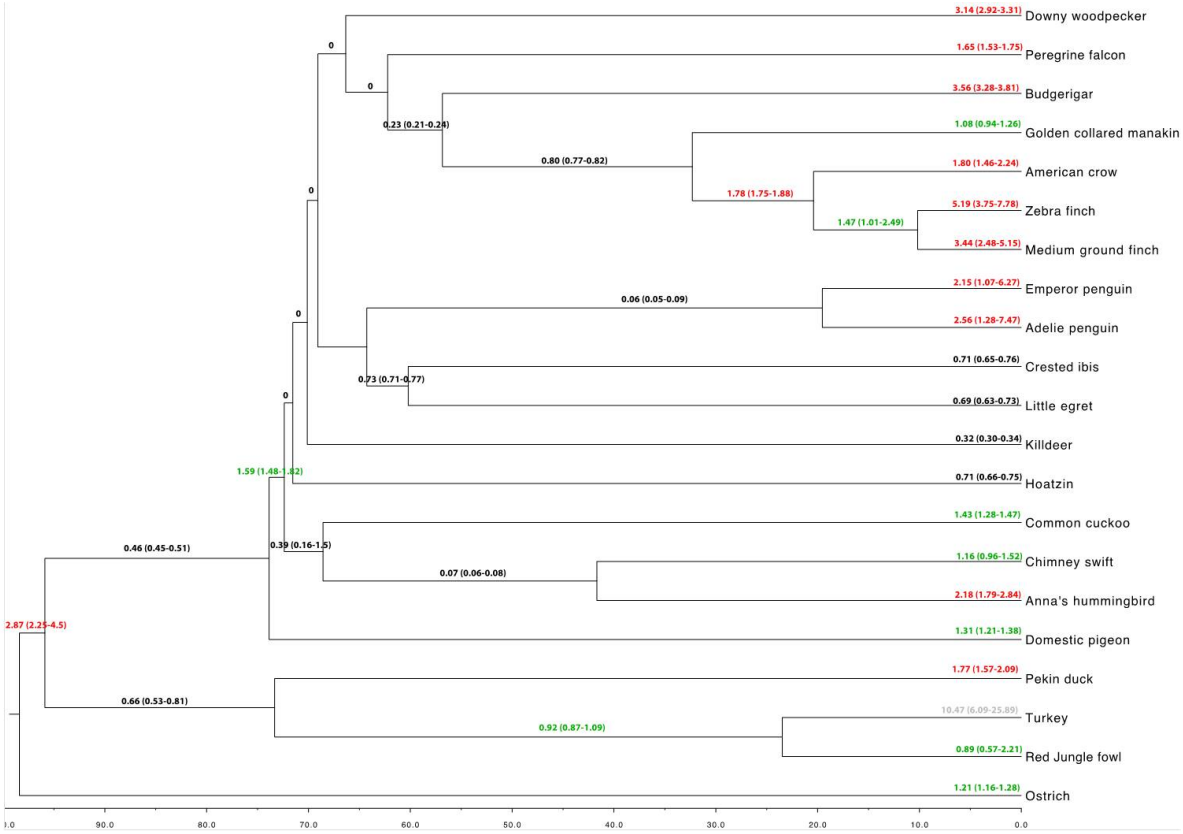

Fig. S15. Chromosomal rearrangement rates in avian lineages. The phylogenetic tree contains only the branches leading to species used in the EBR analysis. Rearrangement rates (RR) for the merged dataset are plotted on each branch, with the range of RR in parentheses. Confidence interval (95%) was calculated to classify RR. Black, low RR

(≤1.01 EBRs/MY); Green, medium RR (1.01-1.82 EBRs/MY); Red, the high RR (≥1.82 EBR/MY). Turkey RR were omitted to calculate these intervals (Pale grey) due to assembly errors.

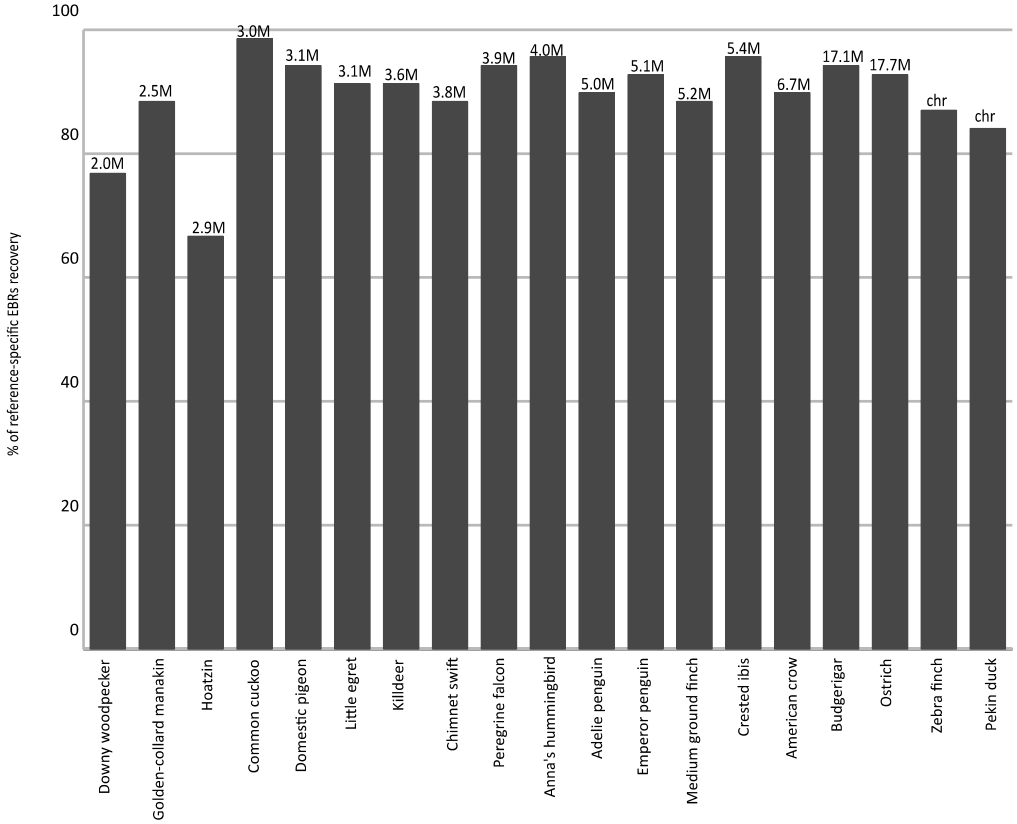

Fig. S16. Fraction of reference-specific evolutionary breakpoint regions (EBRs) detected in each genome. Black bars represent the recovery rate in the "merged" EBR dataset. Number on top of each bars indicates if the genomes are fragmented (scaffold N50 sizes in megabase-pairs (Mbp)) or assembled into chromosomes (chr).

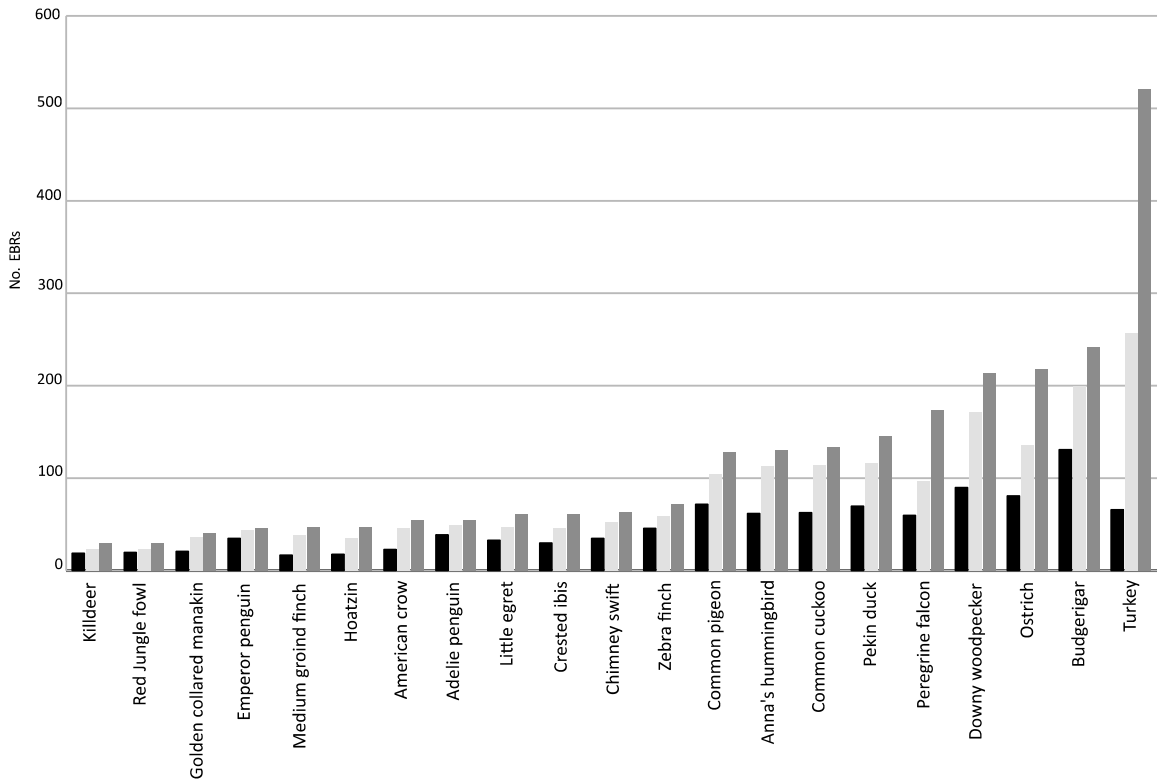

Fig. S17. Number of evolutionary breakpoint regions (EBRs) detected in each avian species using three different resolutions for HSB detection. Black bars, 300Kbp resolution; pale grey, 100Kbp resolution; and dark grey, 50Kbp resolution. Note that while all species have ~1.4x increase in number of EBRs for adjacent resolutions, turkey has ~4x more EBRs in 300Kbp than 100Kbp, and therefore, we suspect that many of the EBRs in turkey are assembly artefacts. Thus, the turkey genome was excluded from the calculations related to rates of chromosomal rearrangements.

Fig. S18. Phylogenetic tree built with 4-fold degenerate sites, using phyloFit in the PHAST package. The topology was fixed as that of TENT tree in the phylogenomics project.

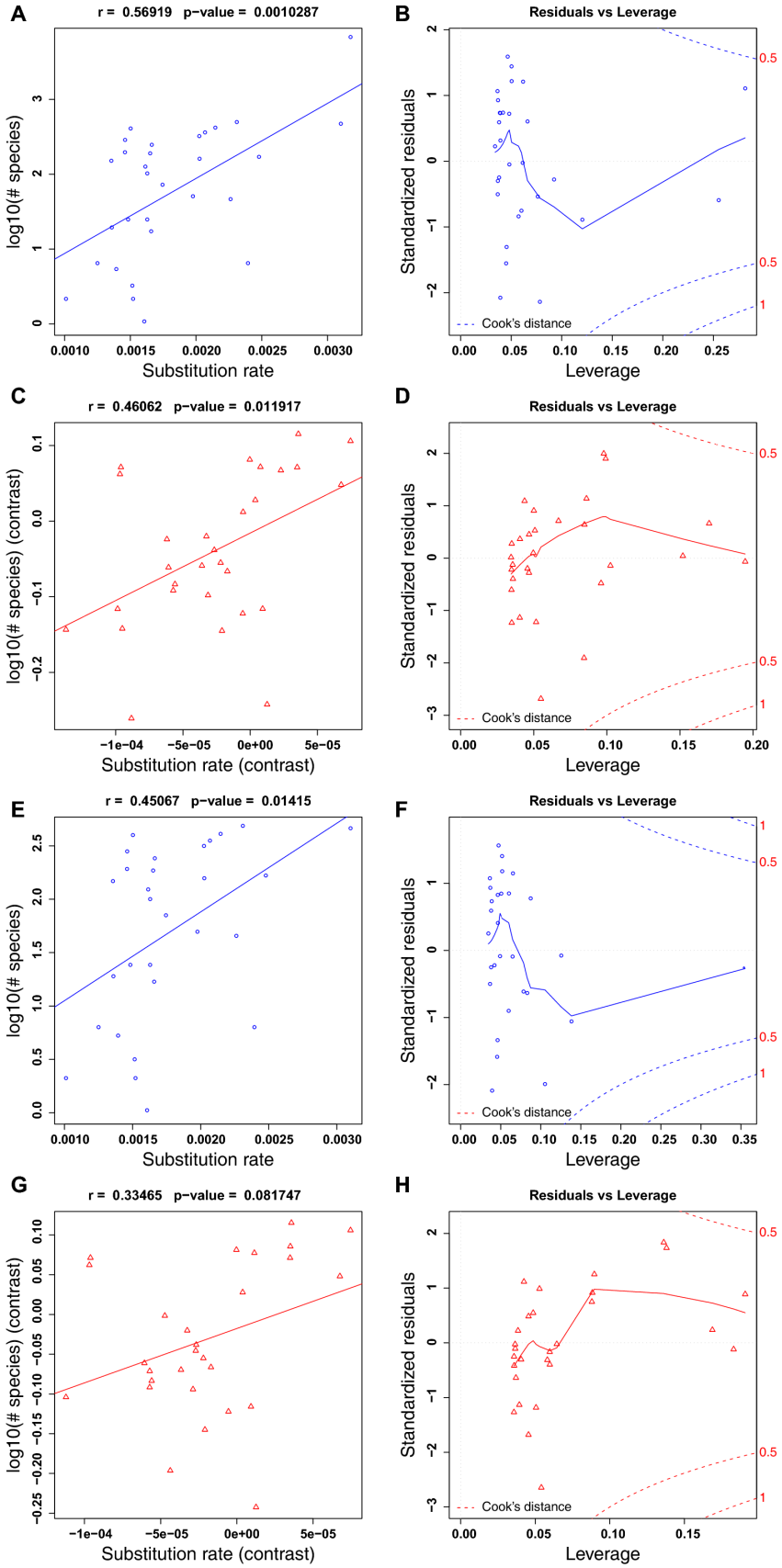
A
r = 0.56919 p−value = 0.0010287
log10(# species)
Substitution rate
B
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
C
r = 0.46062 p−value = 0.011917
log10(# species) (contrast)
Substitution rate (contrast)
D
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
E
r = 0.45067 p−value = 0.01415
log10(# species)
Substitution rate
F
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
G
r = 0.33465 p−value = 0.081747
log10(# species) (contrast)
Substitution rate (contrast)
H
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

Fig. S19. Correlation analysis between average substitution rate and the number of species per order. Panels A-D are for data of 30 orders (including Passeriformes). Panels E-H are for data of 29 orders (excluding Passeriformes). The blue symbols (A, B, E and F) were based on the original data points, and the red symbols (C, D, G and H) were based on the phylogenetic contrast data points. The "r" and "p-value" in panels (A), (C) and (E) were calculated by Pearson's correlation tests. The "Residuals vs Leverage" plots (B, D and F) were generated by "plot.lm()" in R. The fit lines were derived from least square regression analysis.

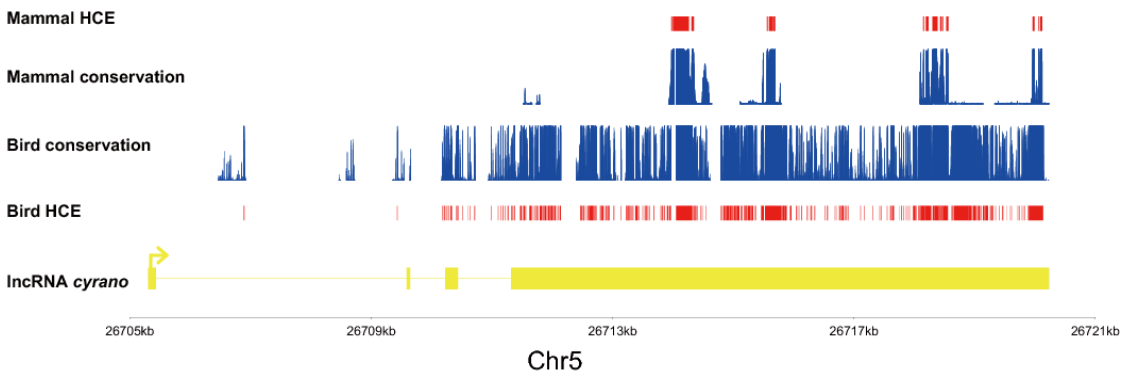

Fig. S20. Gene structure and conservation of cyrano lncRNA gene. Highly conserved elements (HCEs, red) and the conservation scores (blue) were generated by phastCons.

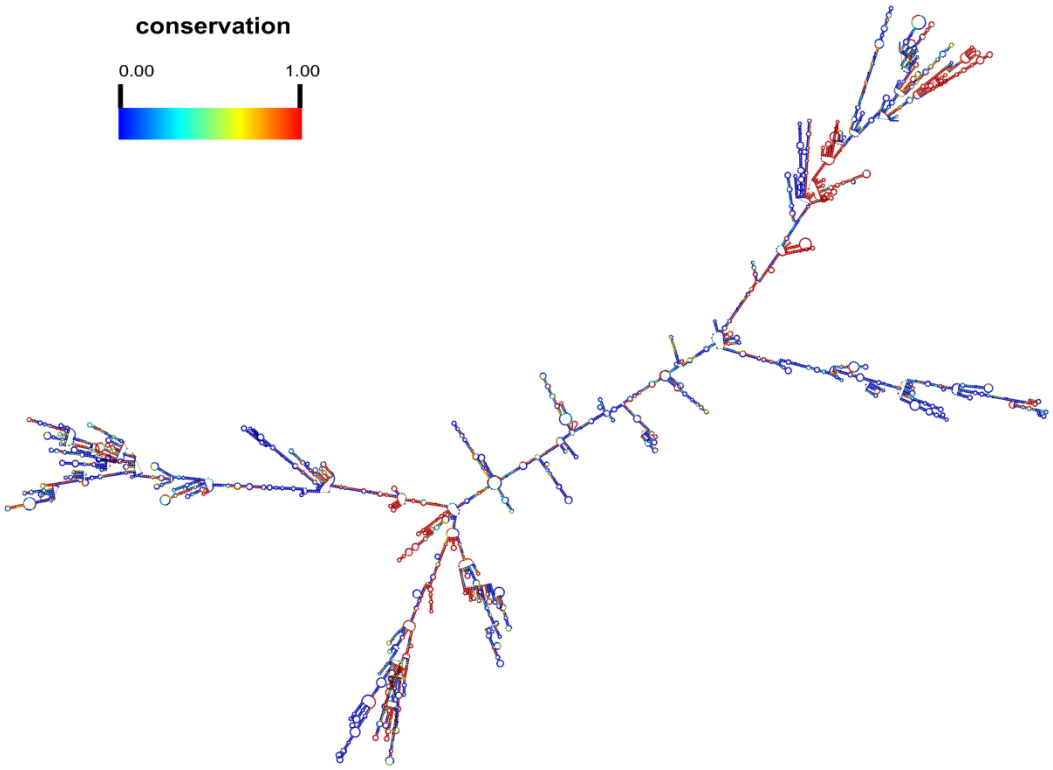

Fig. S21 Predicted secondary structure of cyrano lncRNA gene. The structure was predicted by RNAfold on Vienna RNA server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi), using minimum free energy (MFE) method. The conservation information is from the phastCons conservation scores.

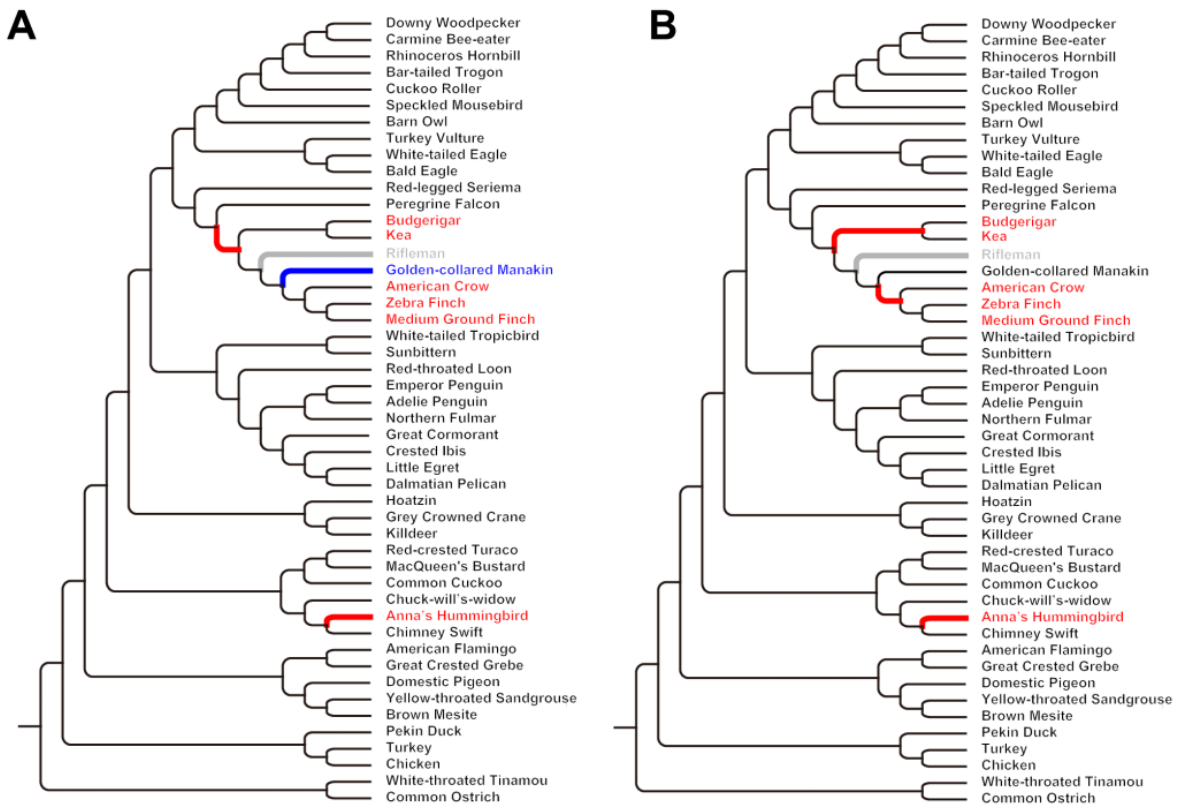

Fig. S22. Hypothesis used to test for convergent evolution of vocal learning. Phylogenetic trees showing the two different parsimonious hypotheses with (A) two independent gains and (B) three independent gains of the vocal learning trait. The species highlighted in red are vocal learners and rifleman is shown in gray as it was excluded from the analysis. The highlighted branches indicate accelerated evolution with the branch in red showing gain of the vocal learning trait and the branch in blue showing loss of the trait.

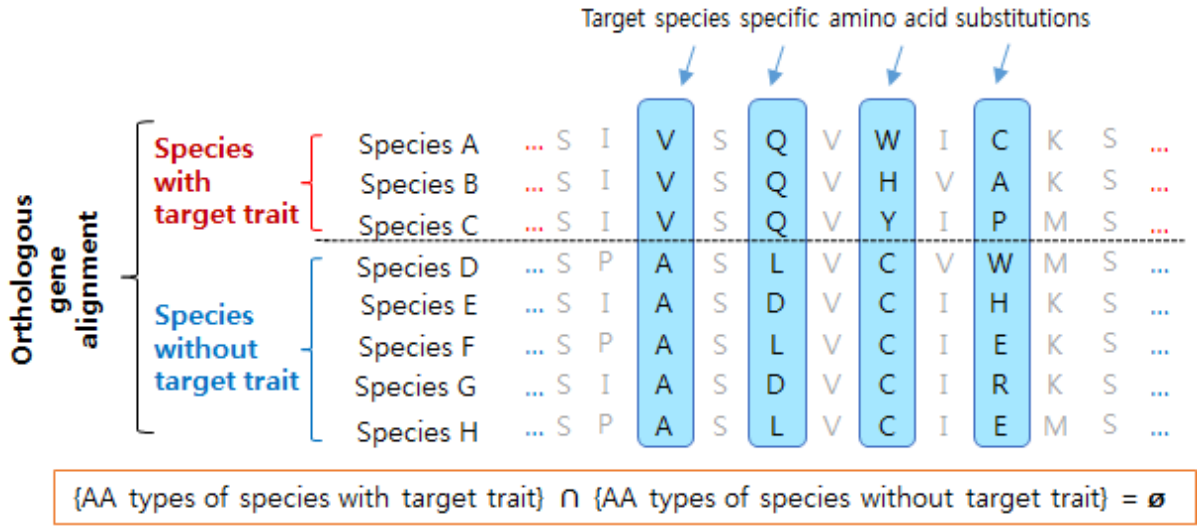

Fig. S23. Concept of target species specific amino acid substitution (TAAS). The diagram shows the mutually exclusive amino acid substitutions between species with the target trait and species without the target trait.

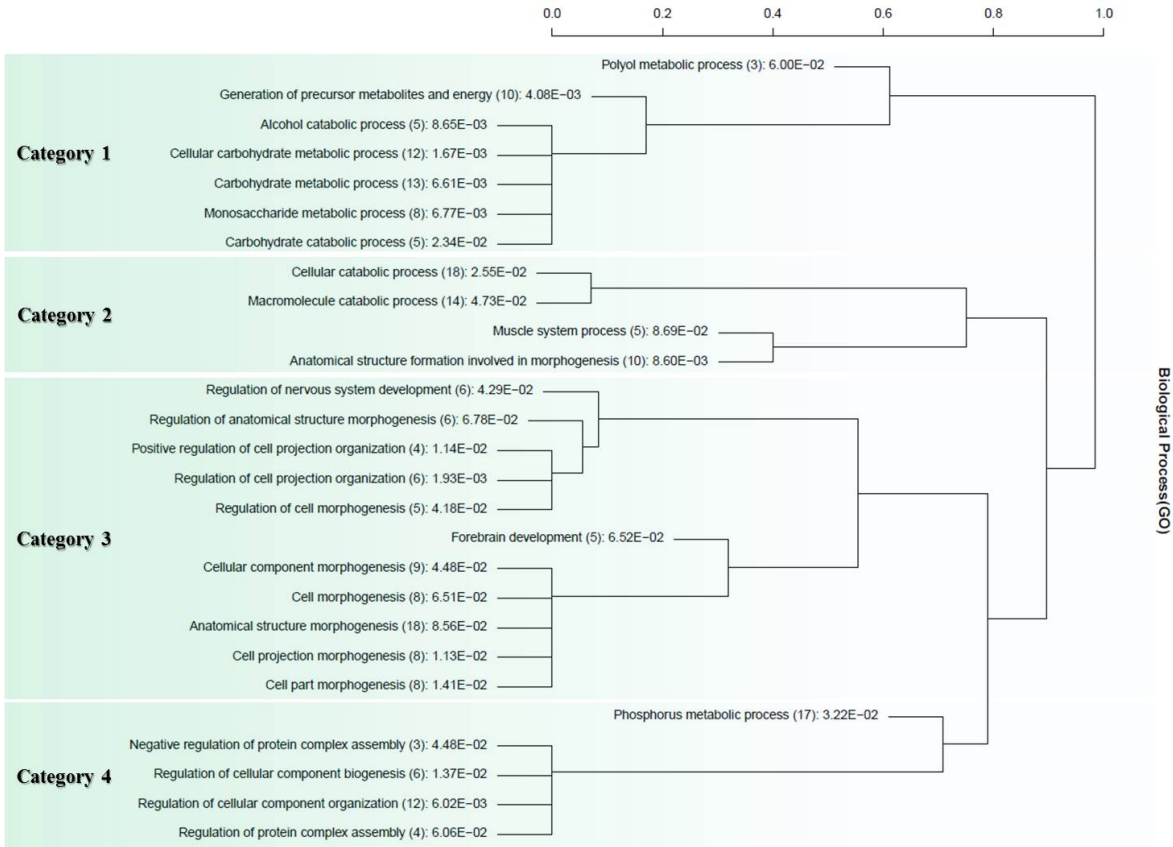

Fig. S24. Hierarchical cluster analyses of GO results (biological process). GO analysis of 165 genes expressed in the songbird brain and that are accelerated under the two evolutionary hypotheses.

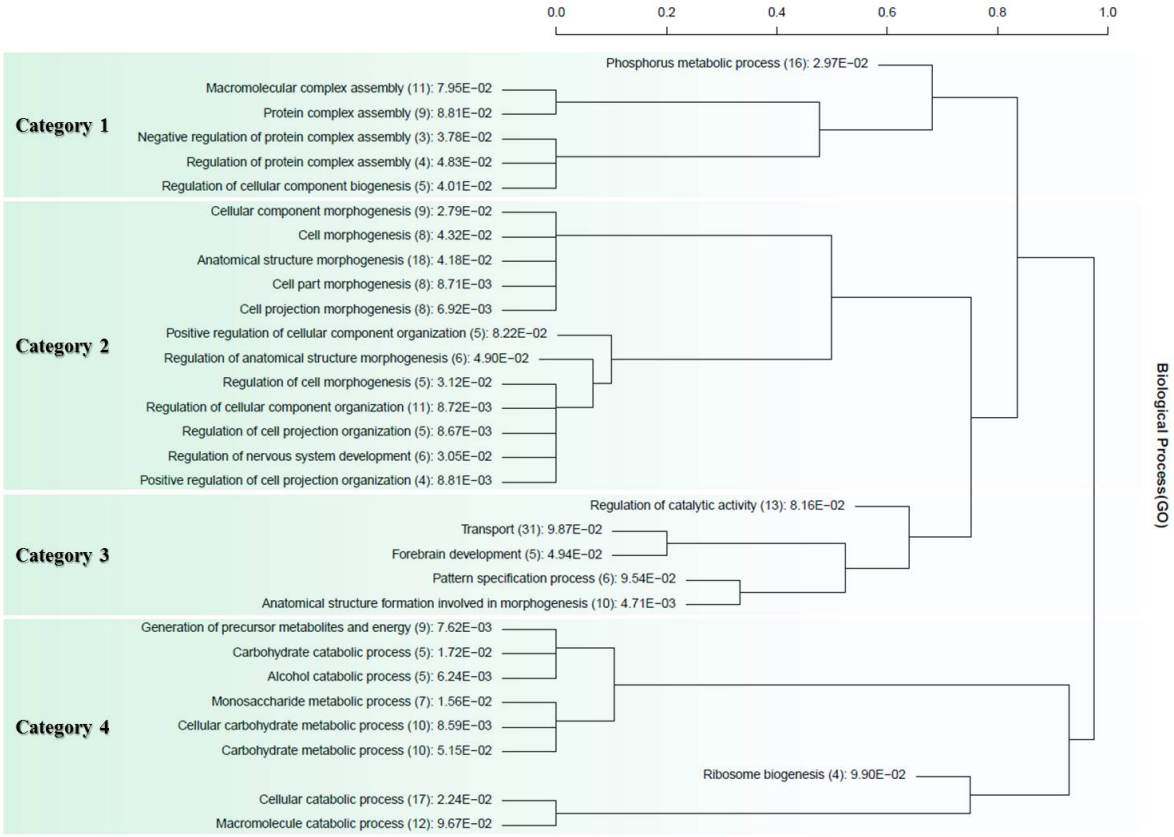

Fig. S25. Hierarchical cluster analyses of GO results (biological process). GO analysis of 151 genes expressed in the song nuclei that are accelerated under the two evolutionary hypotheses.

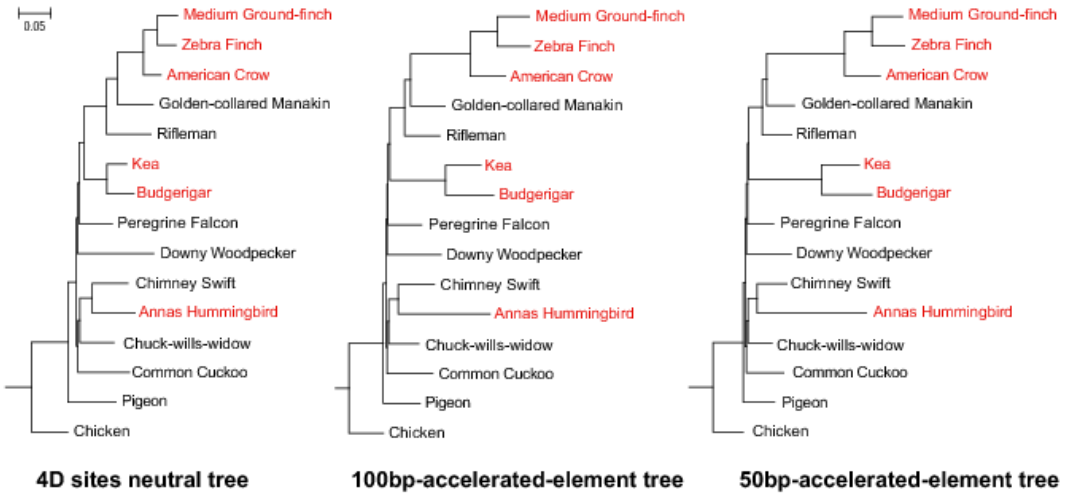

Fig. S26. Neutral tree built with 4D sites, and phylogentic trees built with accelerated elements. The species in red are vocal learners.

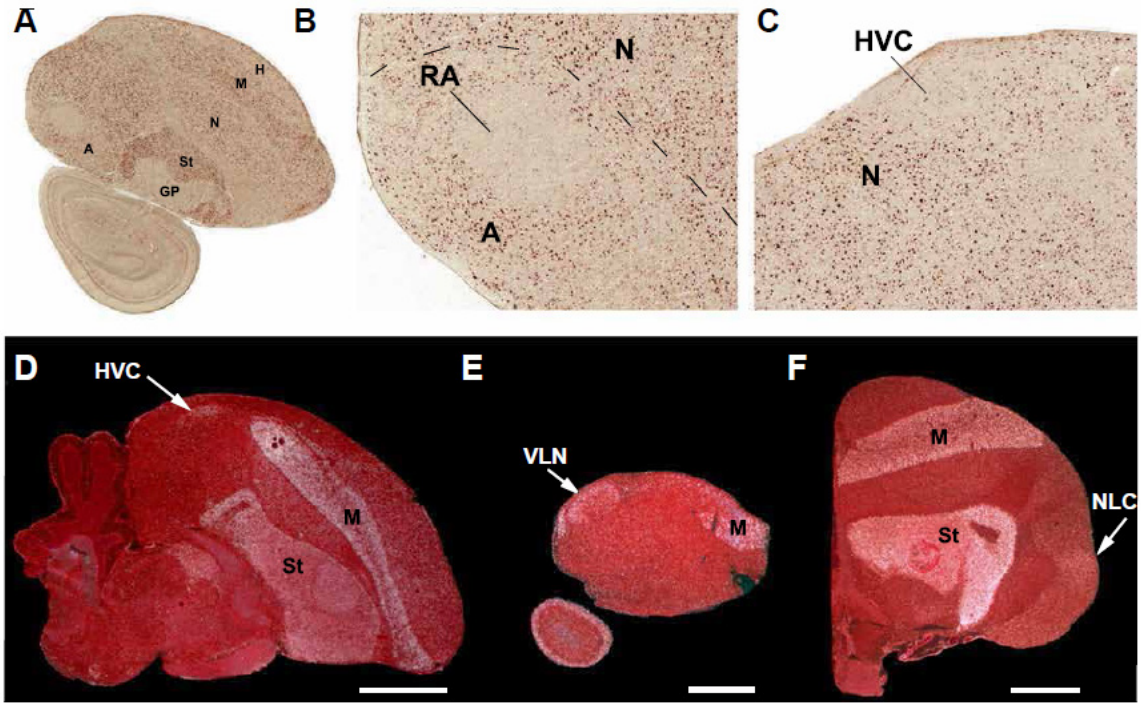

Fig. S27 Examples of differential expression in vocal learning analyses. (**A-C**) Decreased expression of *B3GNT2* mRNA (brown dioxygenin probe signal) in the (B) RA and (C) HVC song

nuclei of the zebra finch (Images derived from the ZEBrA database, http://www.zebrafinchatlas.org). (**D-F**) Increased expression of *FOXP1* mRNA (white) in the HVC analogue (arrows) in a songbird (D, zebra finch), hummingbird (E, Anna's hummingbird), and parrot (F, budgerigar). Red, Nissl stain of all cells in darkfield view; the songbird and hummingbird sections are sagital, the parrot is coronal. Images adapted from (*182*). RA, robust nucleus of the arcopallium; Area X of the striatum; HVC, a letter based name; 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.

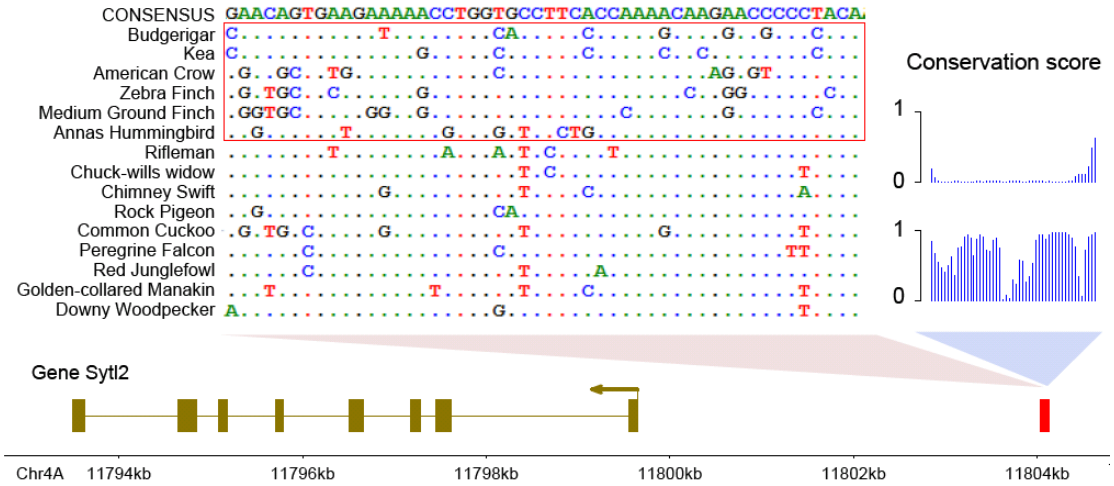

Fig. S28. Gene *Sytl2* and the accelerated element in the 5' flanking region. The levels of conservation were predicted by phastCons on the alignments of vocal learners and vocal non-learners separately.

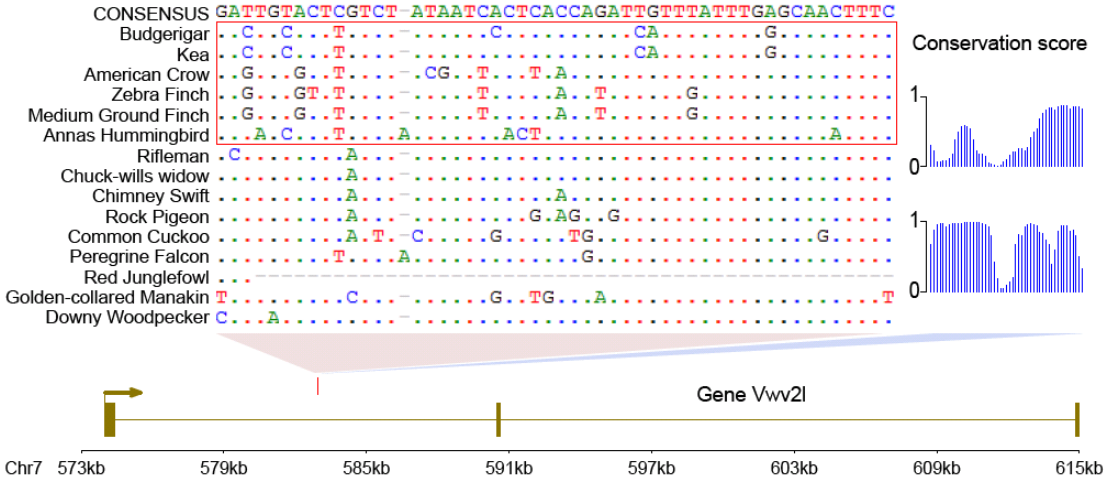

Fig. S29. Gene *Vwc2l* and the accelerated element in the intronic region. The levels of conservation were predicted by phastCons on the alignments of vocal learners and vocal non-learners separately.

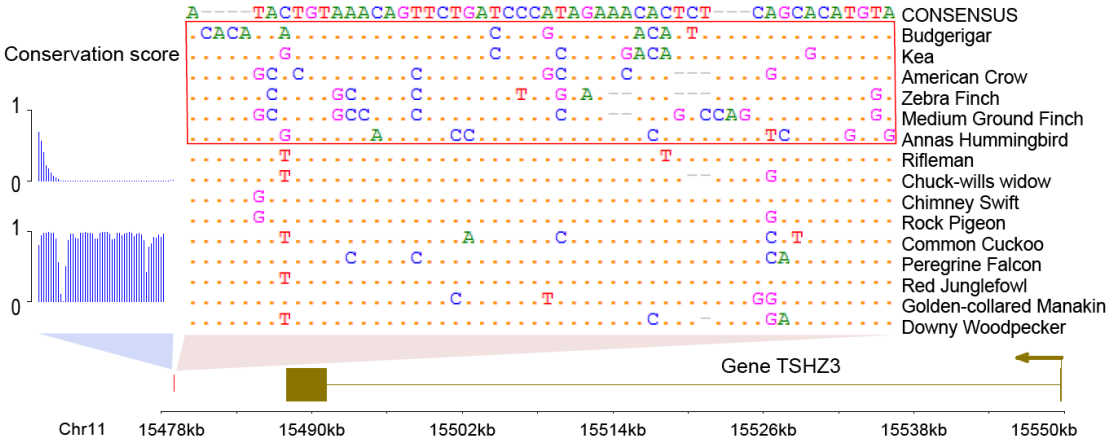

Fig. S30. Gene *TSHZ3* and the accelerated element in the 3' flanking region. The levels of conservation were predicted by phastCons on the alignments of vocal learners and vocal non-learners separately.

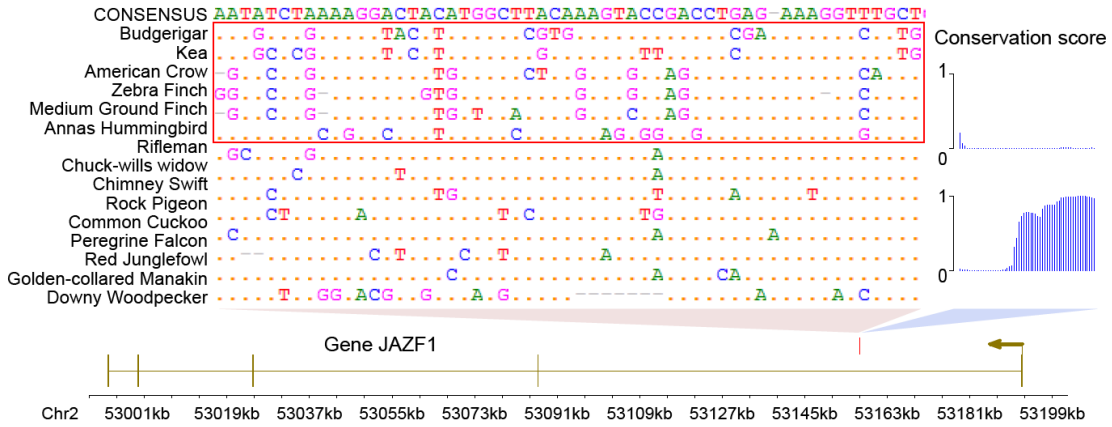

Fig. S31. Gene *JAZF1* and the accelerated element in the intronic region. The levels of conservation were predicted by phastCons on the alignments of vocal learners and vocal non-learners separately.

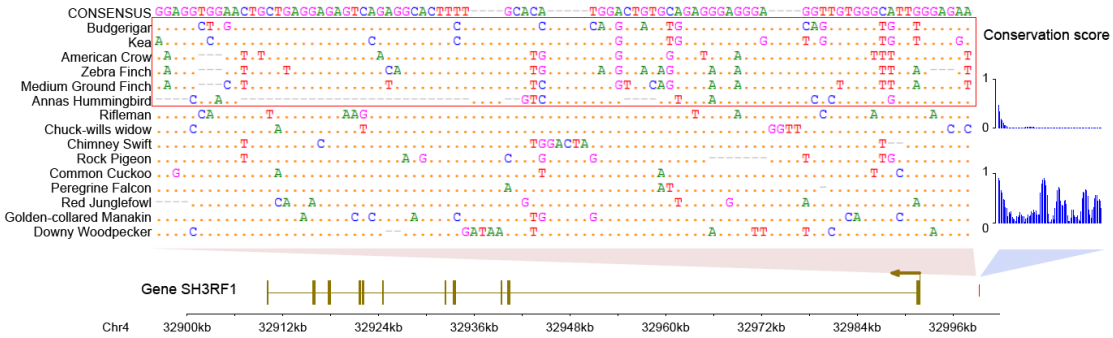

Fig. S32. Gene *SH3RF1* and the accelerated element in the 5' flanking region. The levels of conservation were predicted by phastCons on the alignments of vocal learners and vocal non-learners separately.

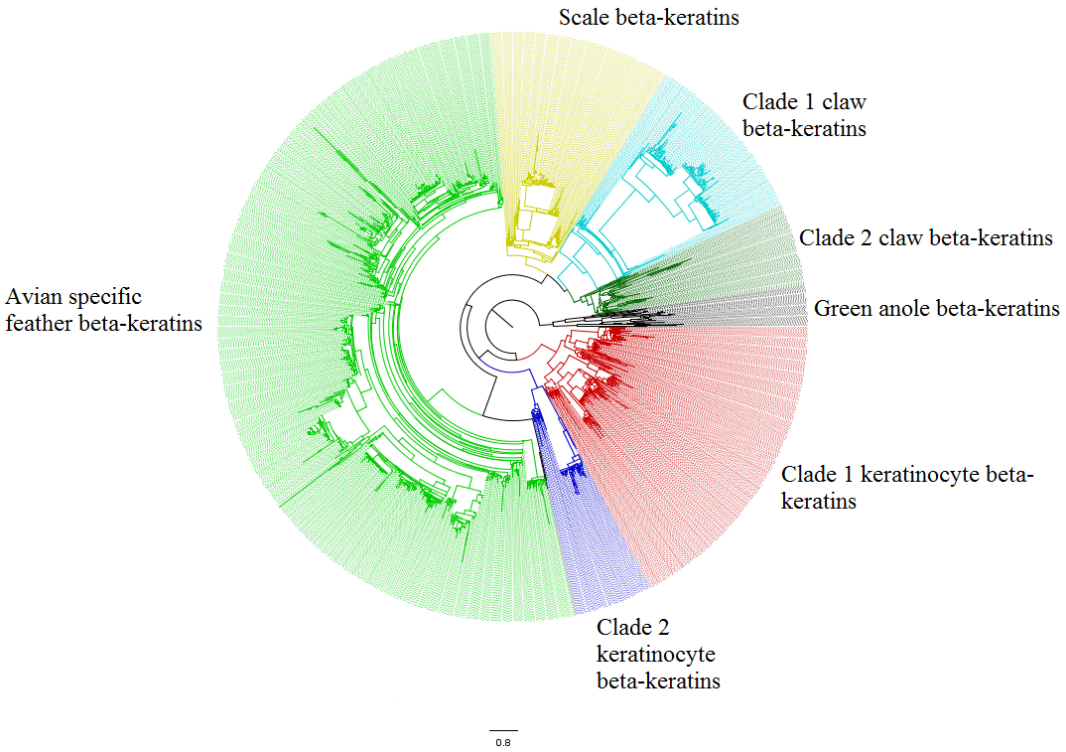

Fig. S33 Maximum likelihood analysis of the β-keratins from the green anole lizard, green sea turtle, American alligator and 48 bird species (Table S41). Annotation of clades is based upon (*162*). All colored clades are statistically significant except for the scale β-keratins, which only have a subset forming a significant clade. All clades, except for the feather β-keratins and green anole β-keratins, are composed of members from green sea turtle, American alligator and birds.

Fig. S34. Alignment of MTS sequences of the bird *AGT* gene. The boxes highlight the pseudogenized regions in the MTS of avian *AGT* genes.

```
                Medium Ground-finch exon-2 GGCCAATGCG GGCTCAGGTT CCAGAACTGG GCGGGGACGT ACGGCTGCGA GCCCGAGCTC TTCTTCCGGC CCCGCGACCT
                                           CGCGCAGATC
                Medium Ground-finch exon-4 CAGGTGGACA AGGAGAAGCT GCAGGTGAAG GTCGAGGCCG GAATTCTCCT CTCCGAGCTC AACCTGGAGC TCGATAAACA
                                           CGGAATGGCC CTG
                Medium Ground-finch exon-5 AGTTTGGGCG CCGTGTCCGA GGTGACGGCG GCCGGAGTCA TCGGCACCGG GACCCACAAC ACCGGCATCG AGCACGGCAT
                                           CCTCCCCACC CAGGTC
          Medium Ground-finch exon-7&8 GTTTCCCAGG TCCTGGACGA GCTGGAGGAG CACCTGAGGA ATTCCCGATA TTTCCGCTTC CTGTGGTTCC CGCACACCGA
                                       GAACGTGCGC GTCATCTACC AGGACCCCAC GTCCCAGCCT CCCCGCTCCT CCTCCAGCTG GTTTTGGGAT TACGCCGTGG
                                       GATATCATCT CCTGGAATTC CTGCTCTGGA TC
Medium Ground-finch exon-9 (partial) AGCCACCGCG TGCTCAACTA CGAGTGCCGC TTCCGCCAGC ACGTCCAGGA CTGGGCCATC CCCATG
           Medium Ground-finch exon-11 TACGGGCGGG AGGCGCCGCG GGGCCGCTAC TGGCGCCTCT ACGAGGGGAT CATGCGCAAG CACGGCGGGA GACCCCACTG
                                       GGCAAAG
           Medium Ground-finch exon-12 CAGGCTCACA GCTGCACGCG GAAGGATCTG GAACAAATGT TCCCCGAATT CCCCAAATTT TGTGCTCTGA GGGAAAAGCT
                                       GGACCCGGAC GGGACCTTCC TGAACCCCTA CCTGGAAAAA GTCTTGTTT
      Golden-collared Manakin exon-2 CAGGTCCACG GGCGAGCAGG AGTGAAGTTC CAGAACTGGT CCCAGACCTA TGGATGCTCC CCGGAGCTCT ACTTCCAACC
                                     CTCCTCCGTG GAGGAGCTGC GGGAGGTGAG GAGGCACCTC CGGGGGAAAT CCCGCCGG
Golden-collared Manakin exon-4&5 CAGGTGGACG CGGAGAAGCG GCAGGTGACG GTGGAAGGCG GGATCCTCCT CTCCGACCTC AACCGGGAGC TGGACAAGCA
                                 CGGGCTGGCC CTGGAAAATT TAGGGGCCGT GTCCGAGGTG TCGGCGGCCG GAGTGATCGG CACCGGGACC CACAACACCG
                                 GCATCAAACA CGGCATCCTG CCCACCCAGG TG
      Golden-collared Manakin exon-6 CAGGTCGTGG CCCTCACCCT CCTGTTGGCC TCGGGAGAGA TCCTGGAATG CTCGGAGGCG GCCAATCCCG AGGTTTTCCA
                                     GGCGGCCCGG CTGCACTTGG GATGCCTGGG AATCGTCCTG AGCGTCACCT TCCAGTGCGT CCCGGAATTC CGGCTGCGCG
                                     AGGTGGCCTT TCCGGCCACC CTGCCCCAGG TA
                      Zebra Finch exon-8 CCGCGCTCCT CGTCCAGCTG GTTCTGGGAT TACGCGGTGG GATATCACCT CCTGCAGTTC CTGCTCTGGC TC
```

Fig. S35. The exons of the pseudogenized *GULO* genes. Shown are the exon number that are psuedogenized for three passerine species.

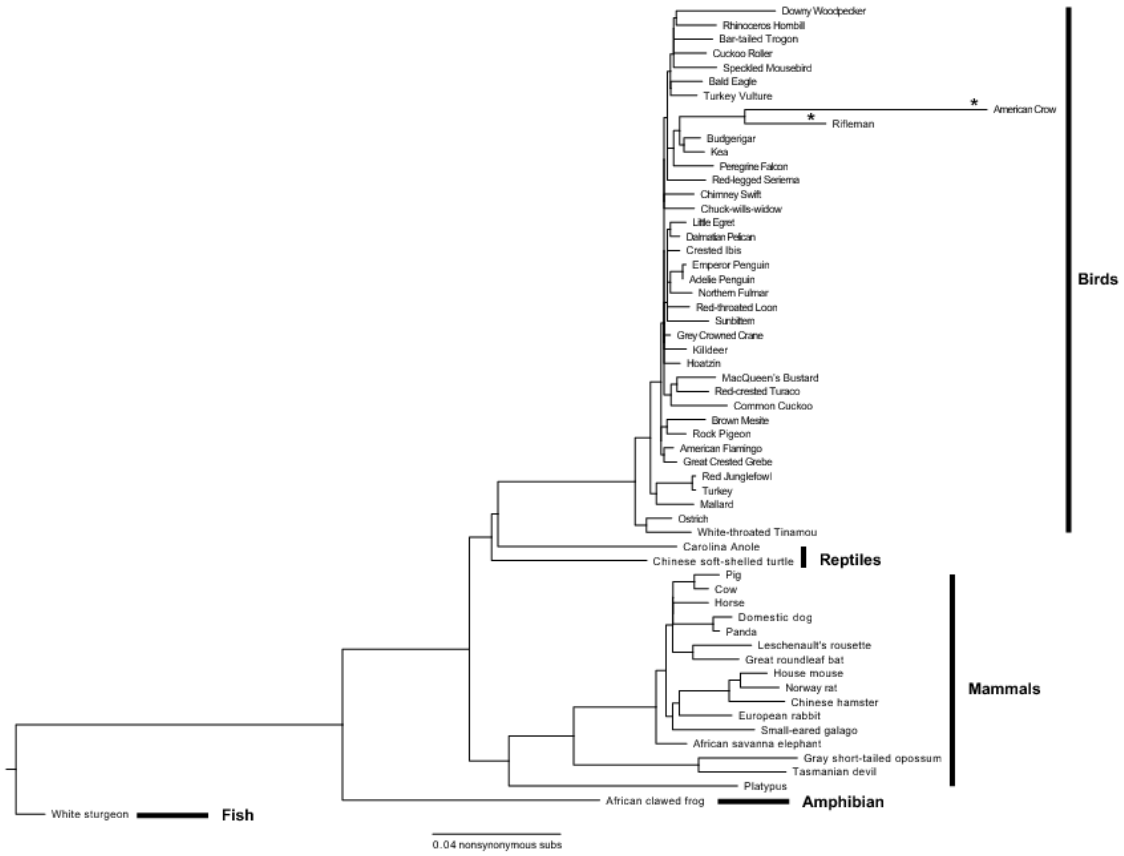

Fig. S36. The phylogenetic tree showing nonsynonymous changes of *GULO* genes among birds, mammals and reptiles. The bar represents the nonsynonymous substitutions per codon. "*" the two species that show higher non-synonymous substitutions (relaxed selection).

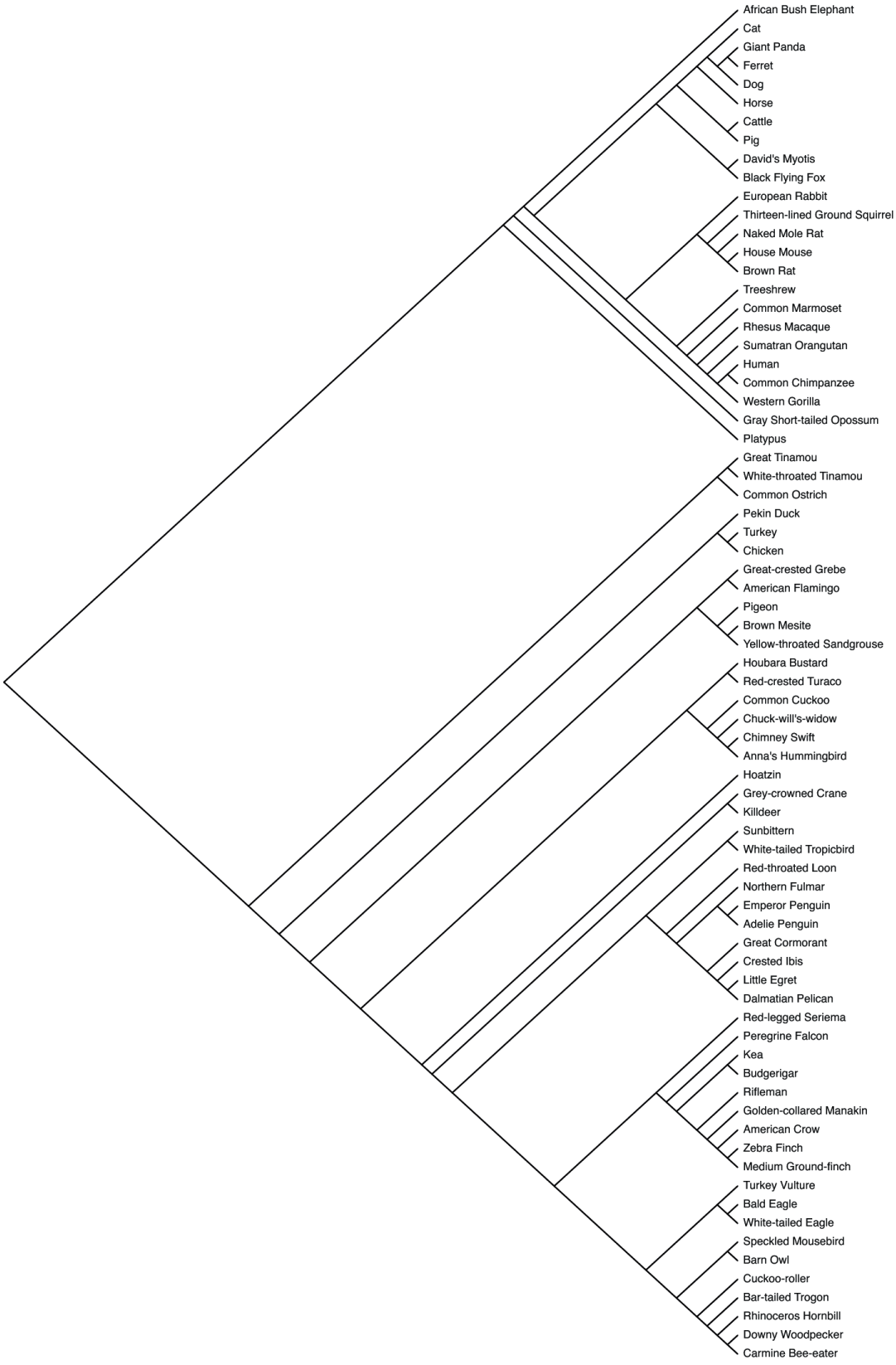
African Bush Elephant
Cat
Giant Panda
Ferret
Dog
Horse
Cattle
Pig
David's Myotis
Black Flying Fox
European Rabbit
Thirteen-lined Ground Squirrel
Naked Mole Rat
House Mouse
Brown Rat
Treeshrew
Common Marmoset
Rhesus Macaque
Sumatran Orangutan
Human
Common Chimpanzee
Western Gorilla
Gray Short-tailed Opossum
Platypus
Great Tinamou
White-throated Tinamou
Common Ostrich
Pekin Duck
Turkey
Chicken
Great-crested Grebe
American Flamingo
Pigeon
Brown Mesite
Yellow-throated Sandgrouse
Houbara Bustard
Red-crested Turaco
Common Cuckoo
Chuck-will's-widow
Chimney Swift
Anna's Hummingbird
Hoatzin
Grey-crowned Crane
Killdeer
Sunbittern
White-tailed Tropicbird
Red-throated Loon
Northern Fulmar
Emperor Penguin
Adelie Penguin
Great Cormorant
Crested Ibis
Little Egret
Dalmatian Pelican
Red-legged Seriema
Peregrine Falcon
Kea
Budgerigar
Rifleman
Golden-collared Manakin
American Crow
Zebra Finch
Medium Ground-finch
Turkey Vulture
Bald Eagle
White-tailed Eagle
Speckled Mousebird
Barn Owl
Cuckoo-roller
Bar-tailed Trogon
Rhinoceros Hornbill
Downy Woodpecker
Carmine Bee-eater

Fig. S37. Consensus tree of mammals and birds used for the hierarchial comparison in the analysis of opsins.

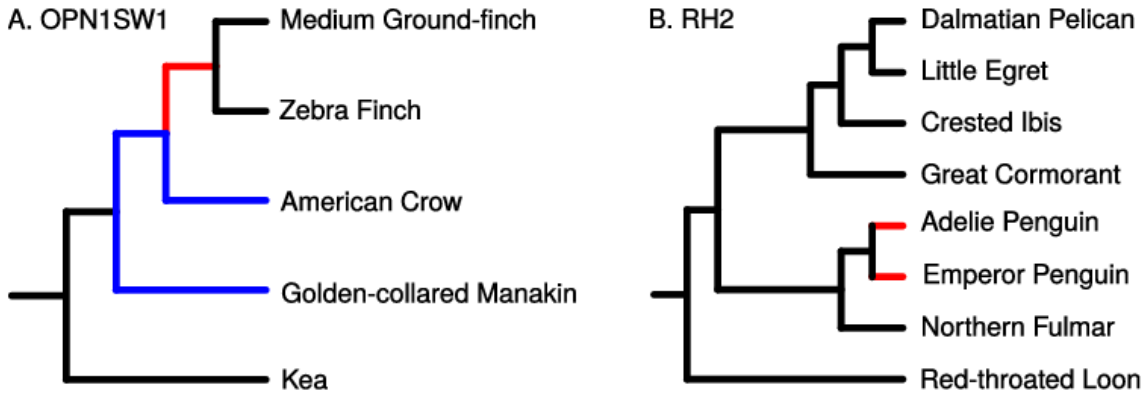

Fig. S38. Significant positive selection was found in two taxa on different genes. (A) *OPN1sw1* on the branch leading to Passerida, represented by zebra finch and medium ground-finch. (B) *RH2* on the branches within Sphenisciformes (Adelie penguin) and emperor penguin). Selection values colour coded onto trees. Black: dN/dS <0.25; blue, dN/dS <1; red: dN/dS >1.

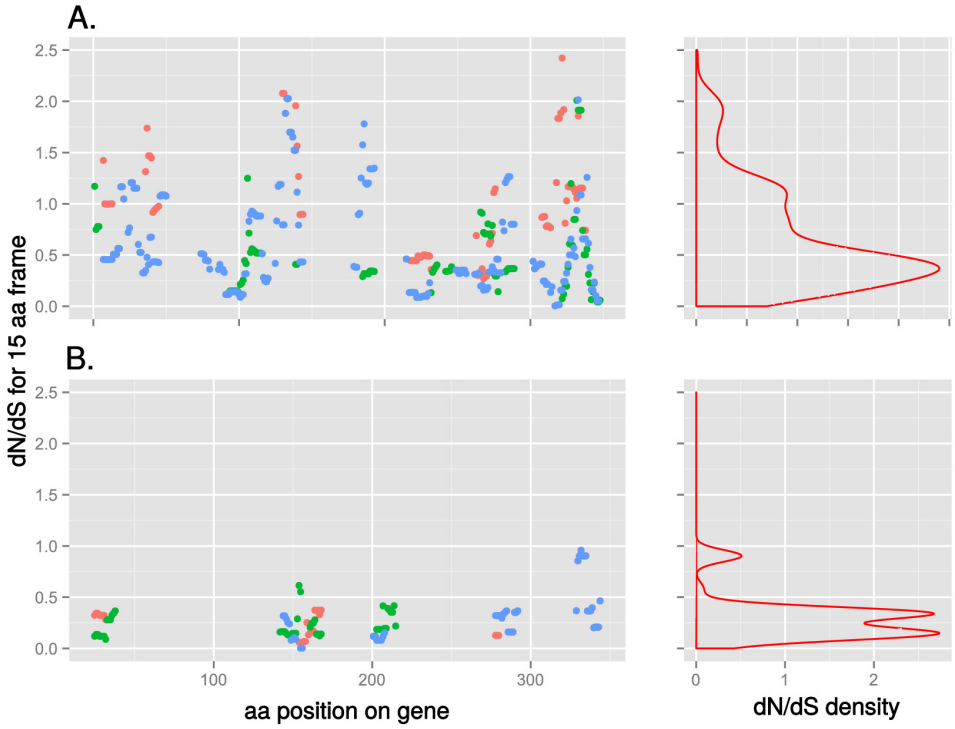

Fig. S39. Comparison of ω values of a 15aa sliding frame over the *RH2* gene. (A) Pairwise comparisons between emperor penguin, Adelie penguin and northern fulmar. (B) Pairwise comparisons between little egret, dalmatian pelican and crested ibis. Red dots, comparisons between sister species (Emperor vs Adelie penguins; Egret vs Fulmar); green and blue dots, other pairwise comparisons. Right panels for each graph show the density curves (red lines) for the distribution of *dN/dS* values.

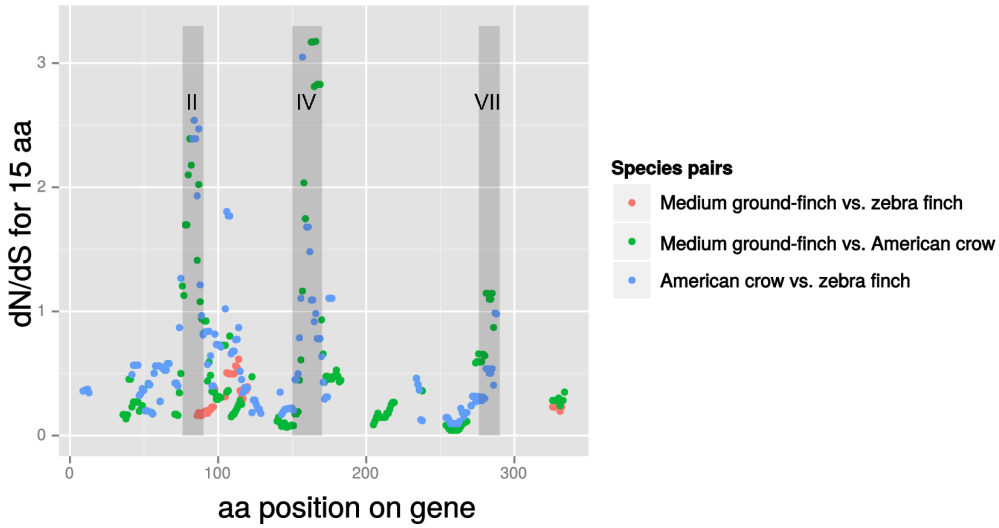

Fig. S40. Selection profile of the SWS1 gene comparing two Passerida passerines, zebra finch and medium ground-finch, and the closest relative in this study, American crow. *dN/dS* were calculated using a sliding 15aa window over the gene. Values are missing where no ω could be calculated due to low substitution levels. The difference between the two Passerida species is low, as is evident by the low *dN/dS* values (red). Three short regions show signs of positive selection and are labelled by the number (roman numerals) of the transmembrane region they are located in.

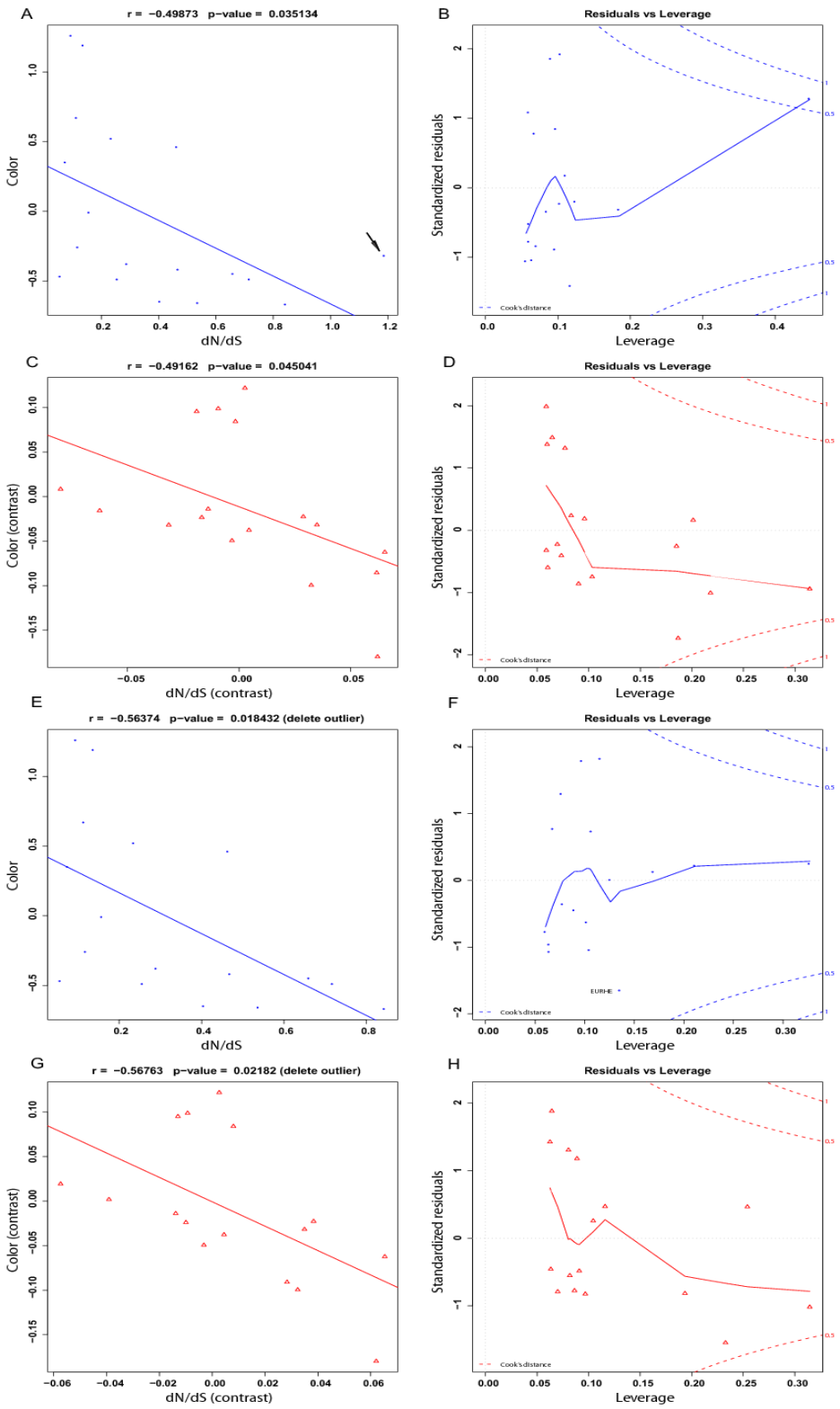

Fig. S41. Correlation analysis between *dN/dS* of *GSTA2* and color discriminability. (A) Correlation analysis based on the original data points. The "r" and "p-value" in

panels were calculated by Pearson's correlation tests. The fit line was inferred by least squares linear regression. The arrow at bottom right indicates the outlier in the original data points, predicted by 'outlier.test()' in R ($p<0.05$). (B) The "Residuals vs Leverage" plots for the data in (A). This was generated by "plot.lm()" in R, which is for spotting highly influential points. (C) Correlation analysis based on the phylogenetic contrasts of the data in (A). (D) The "Residuals vs Leverage" plots for the data in (C). (E) Correlation analysis based on the original data points (after deleting the outlier). (F) The "Residuals vs Leverage" plot for the data in (E). (G) Correlation analysis based on the phylogenetic contrasts of the data in (E). (H) The "Residuals vs Leverage" plots for the data in (G).

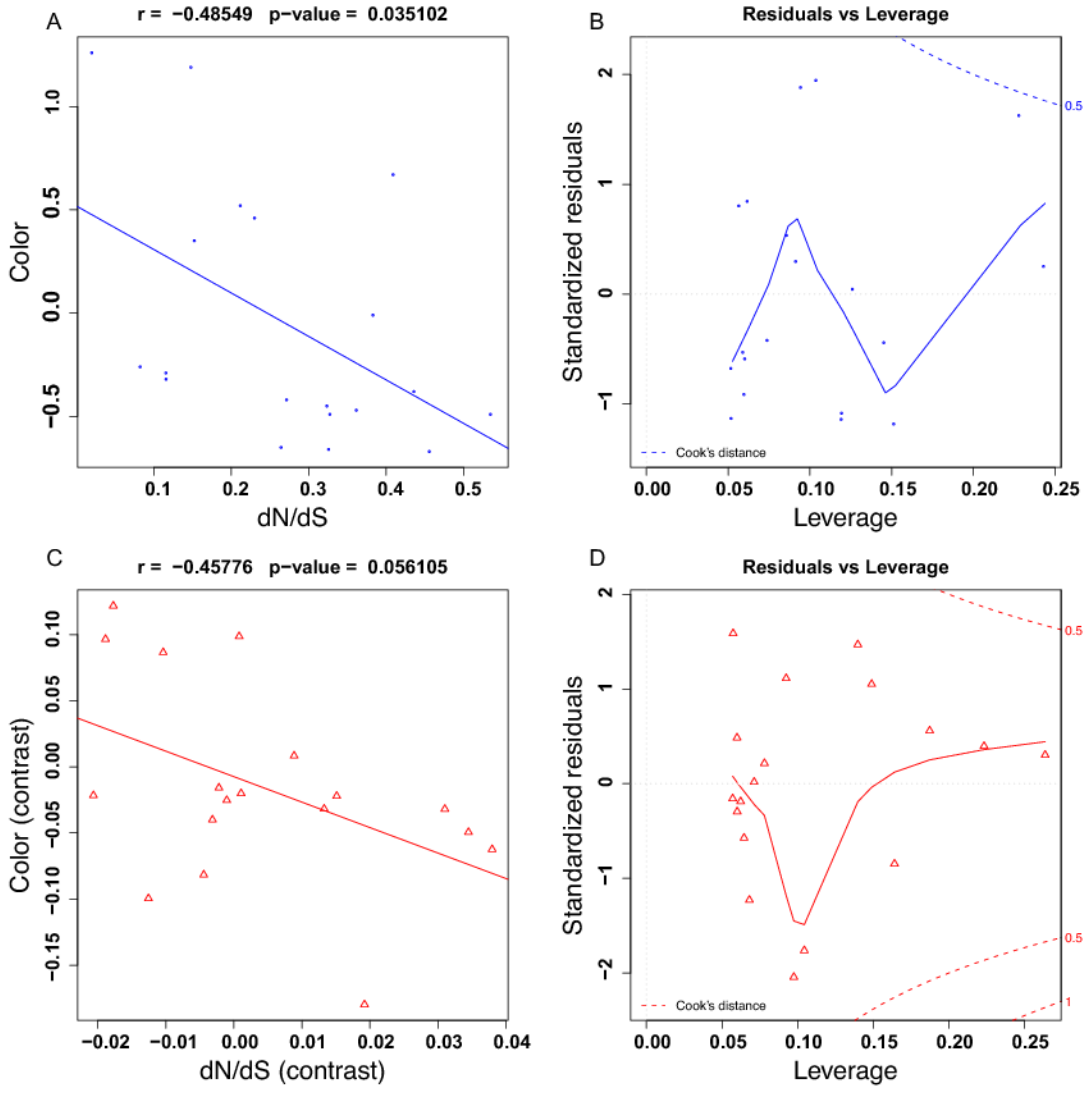

Fig. S42. Correlation analysis between *dN/dS* of *SLC24A4* and color discriminability. (A) Correlation analysis based on the original data points. The "r" and "p-value" in panels were calculated by Pearson's correlation tests. The fit line was inferred by least squares linear regression. (B) The "Residuals vs Leverage" plot for the data in (A). This

was generated by “plot.lm()” in R, which is for spotting highly influential points. (C) Correlation analysis based on the phylogenetic contrasts of the data in (A). (D) The “Residuals vs Leverage” plots for the data in (C).

Supplementary tables

Table S1. Samples of birds used for DNA isolation.

Species (sorted alphabetic)	Common name	Tissue	Source of tissue	Publication
Acanthisitta chloris	rifleman	Muscle	Rifleman 1 pectoral muscle in RNAlater - Chick; found dead 10/1/11; Kowhai bush, Kaikoura, South Island (42degrees 23' S, 173 degrees 37'E, New Zealand. Body decomposed post earthquake but will be skeletonised	This study
Anas platyrhynchos domestica	Peking duck	Blood	Gold Star Duck Production (Beijing, China)	(*14*)
Antrostomus carolinensis	chuk-will's-widow	Tissue	Louisiana, USA: Cameron Parish: East Jetty Woods, 2 mi S Cameron	This study
Apaloderma vittatum	bar-tailed trogon	Muscle	Blood in DMSO Udzyngwa Mts. Tanzania. 2007. Cat#140150	This study
Aptenodytes forsteri	emperor penguin	Blood	The Emperor Island nearby the Zhongshan Station in the South Pole	This study
Balearica regulorum gibbericeps	grey-crowned crane	Blood	Alive, Chip 8FF07, born 14-7-1992 in Givskud Zoo, Denmark	This study
Buceros rhinoceros silvestris	rhinoceros hornbill	Muscle	SNM. Chip: 208219000004613, probably born in the wild around 1989, bought from animal dealer Van der Brink in 1992. Died 04-12-2006.	This study
Calypte anna	Anna's hummingbird	Blood, Muscle, Liver	Portland, OR, USA	This study
Cariama cristata	Red-Legged Seriema	Blood	Alive, Legband: 295951 left leg, born 29-6-2000 in Noorder Dierenpark, Emmen, Holland	This study
Cathartes aura	turkey vulture	Blood	NC Raptor Center. Wild caught North Carolina, USA	This study
Chaetura pelagica	chimney swift	Tissue	Louisiana, USA: Cameron	This study

			Parish: 7 mile West Old Mouth Mermentau River	
Charadrius vociferus	killdeer	Tissue	Peru: Loreto Department: Tumbes 25km SE Zorritos, 3° 49' 1" S, 80° 29' 19" W	This study
Chlamydotis macqueenii	MacQueen's bustard	Blood	Male Houbara 12533. Private collection.	This study
Colius striatus	speckled mousebird	Blood	One of many males	This study
Columba livia	pigeon	Blood	Anders Christiansen, Danish racing pigeon association.	(*16*)
Corvus brachyrhynchos	American crow	Blood	Asheboro Zoo, NC, USA	This study
Cuculus canorus	common cuckoo	2 individuals	LSU & Wild caught in Denmark - low coverage genome done from the first one caught.	This study
Egretta garzetta	little egret	Blood	Yangxian County on the south of the Qinling Mountains	This study
Eurypyga helias	sunbittern	Liver/blood	Odense zoo - Copenhagen	This study
Falco peregrinus	peregrine falcon	Blood	United Arab Emirates	(*15*)
Fulmarus glacialis	northern fulmar	Muscle	SNM frozen tissue, Denmark 1996	This study
Gallus gallus	chicken (Red Jungle Fowl)	Blood	MSU Poultry Research & Teaching Center	(*11*)
Gavia stellata	red-throated loon	Muscle	SNM frozen tissue, Denmark, 1998	This study
Geospiza fortis	medium ground-finch	Blood	Island of Santa Cruz (Galapagos)	This study
Haliaeetus albicilla	white-tailed eagle	Muscle	Frozen tissue, Greenland, 1995	This study
Haliaeetus leucocephalus	bald eagle	Blood	Bird name Derek. NC Raptor Center, Huntersville NC, USA	This study
Leptosomus discolor	cuckoo roller	Blood	Born in captivity at Parc Botanique et Zoologique de Tsimbaz	This study

Manacus vitellinus	golden-collared manakin	Brain	Gamboa, Panama	This study
Meleagris gallopavo	turkey	Blood	Nicholas Turkey Breeding Farms	(*12*)
Melopsittacus undulatus	budgerigar (parakeet)	Blood	Duke University Aviary (Budgerigar colony)	(*20*)
Merops nubicus	carmine bee-eater	Blood	SNM Euthanised on 8-4-2011.	This study
Mesitornis unicolor	Brown ;esite	Muscle	Madagascar: Toliara; Fivondronana de Tolagnaro; Foret d'Analalava, 7 km N Manantenina	This study
Nestor notabilis	kea	Muscle	SNM Large nestling hatched April 2010, died on 27-5-2010.	This study
Nipponia nippon	crested ibis	Blood	southern Qinling Mountains, China	This study
Opisthocomus hoazin	hoatzin	Muscle-crop	Lagunas, Venezuela (a nearby forest 1998); Venezuelan Ministry of Environment # 11-001237	This study
Pelecanus crispus	dalmatian pelican	Blood	Unmarked tube	This study
Phaethon lepturus	white-tailed tropicbird	Muscle	South Atlantic, 2004, Blood in DMSO, Ascension Island, 2004, 135885	This study
Phalacrocorax carbo	great cormorant	Muscle	Frozen tissue, Gedser, Denmark , 4 Feb 2010	This study
Phoenicopterus ruber ruber	American flamingo	Blood	Legband: GWX, chip: , Born prior to 1995, we don't know where.	This study
Picoides pubescens	downy woodpecker	Tissue	Montana, USA: Cowell County: Marcum Mt.	This study
Podiceps cristatus	great-crested grebe	Muscle	Frozen Tissue Denmark, 6 Feb 2004	This study
Pterocles gutturalis	yellow-throated sandgrouse	Blood	18. From Sharjah. Wild born male, Tanzania.	This study
Pygoscelis adeliae	Adelie penguin	Blood	Inexpressible Island, Ross Sea,	This study

			Antarctica	
Struthio camelus australis	common ostrich	Blood	San Diego Zoo (originally from Botswana)	This study
Taeniopygia guttata	zebra finch	Muscle	UCLA Aviary	(*13*)
Tauraco erythrolophus	red-crested turaco	Blood	Legband: LDF 577, born July 2010 in Copenhagen Zoo. Now lives at Attica Zoo, Greece.	This study
Tinamus guttatus	white-throated tinamou	Tissue	Peru: Loreto Department: Ca 7km SW Jeberos, 5° 18' 48" S, 76° 16' 32" W	This study
Tyto alba	barn owl	Brain	Knudsen Lab Aviary, Stanford University (Breed from wild caught barn owls captured at Davis, CA USA)	This study

Table S2. Basic statistics for the assemblies of avian species.

Species	Common name	Sequence Depth	Library	Assembly (contig/scaffold N50;total length)	NCBI accessions	Ensembl accessions
Published (Sanger; turkey done with illumina+454)						
Gallus gallus	chicken (*11*)	7X	-	36K/7.07M;1.05G	PRJNA13342	Gallus_gallus-4.0
Taeniopygia guttata	zebra finch (*13*)	6X	-	39K/10M;1.2G	PRJNA17289	taeGut3.2.4
Meleagris gallopavo	turkey (*12*)	17X	-	12.6K/1.5M;1.038G	PRJNA42129	UMD2
Published during this project						
Anas platyrhynchos domestica	Peking duck (*14*)	50X	200,500,2k,5k,10k	26K/1.2M;1.1G	PRJNA46621	BGI_duck_1.0
Columba livia	pigeon (*16*)	63X	200, 500, 800, 2k, 5k, 10k, 20k	22K/3.2M;1.11G	PRJNA167554	BGI_colLiv_0.0
Falco peregrinus	peregrine falcon (*15*)	105X	200,500,800,2k,5k,10k,20k	28K/3.9M;1.18G	PRJNA159791	BGI_falPer_0.0

High-coverage genomes						
Pygoscelis adeliae	Adelie penguin	60X	200,500,800,2k,5k,10k,20k	19K/5.0M; 1.23G	PRJNA235983	BGI_pygAde_0.0
Aptenodytes forsteri	emperor penguin	60X	200,500,2k,5k,10k,20k	30K/5.1M 1.26G	PRJNA235982	BGI_aptFor_0.0
Nipponia nippon	crested ibis	105X	200,500,800,2k,5k,10k,20k	22K/5.4M;1.17G	PRJNA232572	BGI_nipNip_0.0
Egretta garzetta	little egret	74X	200,500,800,2k,5k,10k,20k	24k/3.1M;1.2G	PRJNA232959	BGI_egrGar_0.0
Calypte anna	Anna's hummingbird	110X	200,500,800,2k,5k,10k,20k	23K/4M;1.1G	PRJNA212866	BGI_calAnn_0.0
Chaetura pelagica	chimney swift	103X	200,500,800,2k,5k,10k,20k	27K/3.8M;1.1G	PRJNA210808	BGI_chaPel_0.0
Charadrius vociferus	killdeer	100X	200,500,800,2k,5k,10k,20k	32K/3.6M;1.2G	PRJNA212867	BGI_chaVoc_0.0
Cuculus canorus	common cuckoo	100X	200,500,800,2k,5k,10k,20k	31K/3M;1.15G	PRJNA212870	BGI_cucCan_0.0
Ophisthocomus hoazin	hoatzin	100X	200,500,800,2k,5k,10k	24K/2.9M;1.14G	PRJNA212873	BGI_ophHoa_0.0
Geospiza fortis	medium ground-finch	115X	200,500,800,2k,5k,10k,20k	30K/5.2M;1.07G	PRJNA156703	BGI_geoFor_0.0
Manacus vitellinus	golden-collared manakin	110X	200,500,800,2k,5k,10k,20k	34K/2.5M;1.12G	PRJNA212872	BGI_manVit_0.0
Melopsittacus undulatus	budgerigar	160X	200, 500, 800, 2k, 5k, 10k	55K/10.6M;1.1G	PRJEB1588	BGI_melUnd_0.0
Picoides pubescens	downy woodpecker	105X	200,500,800,2k,5k,10k	20K/2M;1.17G	PRJNA212874	BGI_picPub_0.0
Struthio camelus	common ostrich	85X	200,500,800,2k,5k,10k,20k	29k/3.5M;1.23G	PRJNA212875	BGI_strCam_0.0
Tinamus guttatus	white-throated	100X	200,500,800,2k,5k	24K/242K;1.05G	PRJNA212876	BGI_tinGut_0.0

	tinamou					
Corvus brachyrhynchos	American crow	80X	200,500,800,2k,5k,10k,20k	24K/6.9M;1.1G	PRJNA212869	BGI_corBra_0.0
Haliaeetus leucocephalus	bald eagle	88X	300,400,3k,8k	10K/670K;1.26G	PRJNA237821	BGI_halLeu_0.0
Low-coverage genomes						
Antrostomus carolinensis	Chuck-Will's-Widow	30X	500, 800	17K/45K;1.15G	PRJNA212888	BGI_antCar_0.0
Cariama cristata	Red-Legged Seriema	24X	500, 800	17k/54k; 1.15G	PRJNA212889	BGI_carCri_0.0
Colius striatus	speckled mousebird	27X	500, 800	18K/45k ;1.08G	PRJNA212892	BGI_colStr_0.0
Merops nubicus	carmine bee-eater	37X	500, 800	20K/47K;1.06G	PRJNA212898	BGI_merNub_0.0
Gavia stellata	red-throated loon	33X	500, 800	16k/45k; 1.15G	PRJNA212895	BGI_gavSte_0.0
Balearica regulorum gibbericeps	grey-crowned crane	33X	500, 800	18K/51k; 1.14G	PRJNA212879	BGI_balReg_0.0
Apaloderma vittatum	Bar-tailed trogon	28X	500, 800	19K/56k; 1.08G	PRJNA212878	BGI_apaVit_0.0
Phalacrocorax carbo	great cormorant	24X	500, 800	15K/48K;1.15G	PRJNA212903	BGI_phaCar_0.0
Phaethon lepturus	white-tailed tropicbird	39X	500, 800	18K/47K;1.16G	PRJNA212902	BGI_phaLep_0.0
Phoenicopterus ruber ruber	American flamingo	33X	500, 800	16K/37K;1.14G	PRJNA212904	BGI_phoRub_0.0
Podiceps cristatus	great-crested grebe	30X	500, 800	13K/30K; 1.15G	PRJNA212905	BGI_podCri_0.0
Fulmarus glacialis	northern fulmar	33X	500, 800	17K/46K;1.14G	PRJNA212894	BGI_fulGla_0.0
Tyto alba	barn owl	27X	500, 800	13K/51K;1.14G	PRJNA212909	BGI_tytAlb_0.0
Tauraco erythrolophus	red-crested turaco	30X	500, 800	18K/55K;1.17G	PRJNA212908	BGI_tauEry_0.0

Cathartes aura	turkey vulture	25X	500, 800	12K/35K;1.17G	PRJNA212890	BGI_catAur_0.0
Eurypyga helias	sunbittern	33X	500, 800	16K/46K;1.1G	PRJNA212893	BGI_eurHel_0.0
Mesitornis unicolor	brown mesite	29X	500, 800	18K/46K;1.1G	PRJNA212899	BGI_mesUni_0.0
Leptosomus discolor	cuckoo roller	32X	200, 500, 800	19K/61K;1.15G	PRJNA212897	BGI_lepDis_0.0
Chlamydotis macqueenii	MacQueen's bustard	27X	500, 800	18K/45K;1.09G	PRJNA212891	BGI_chlMac_0.0
Pelecanus crispus	dalmatian pelican	34X	500, 800	18K/43K;1.17G	PRJNA212901	BGI_pelCri_0.0
Pterocles gutturalis	yellow-throated sandgrouse	25X	500, 800	17K/49K;1.07G	PRJNA212906	BGI_pteGut_0.0
Acanthisitta chloris	rifleman	29X	500, 800	18k/64k; 1.05G	PRJNA212877	BGI_acaChl_0.0
Buceros rhinoceros silvestris	Javan Rhinoceros hornbill	35X	500, 800	14K/51K;1.08G	PRJNA212887	BGI_bucRhi_0.0
Nestor notabilis	kea	32X	500, 800	16K/37K;1.14G	PRJNA212900	BGI_nesNot_0.0
Haliaeetus albicilla	white-tailed eagle	26X	500, 800	20K/56K;1.14G	PRJNA212896	BGI_halAlb_0.0

Table S3. Statistics of protein-coding gene annotations of all the birds.

Species	**Common name**	**#Gene**	**Mean gene length (kb)**	**Mean CDS length (bp)**	**Mean exon length (bp)**	**Mean intron length (bp)**	**Mean intergenic length (kb)**
Acanthisitta chloris	rifleman	14596	13.5	1242	158.6	1800	12
Anas platyrhynchos domestica	Peking duck	16521	17.8	1317	160.7	2298	42
Antrostomus carolinensis	chuck-will's-widow	14676	12.0	1177	164.1	1747	12
Apaloderma vittatum	bar-tailed trogon	13615	13.5	1247	160.8	1806	12

Aptenodytes forsteri	emperor penguin	16070	20.9	1397	161.6	2546	56
Balearica regulorum gibbericeps	grey-crowned crane	14173	13.8	1276	162.7	1828	11
Buceros rhinoceros silvestris	rhinoceros hornbill	13873	13.5	1267	160.4	1767	11
Calypte anna	Anna's hummingbird	16000	18.5	1386	161.7	2264	47
Cariama cristata	red-legged seriema	14216	13.7	1249	161.8	1849	11
Cathartes aura	turkey vulture	13534	10.8	1109	166.4	1716	10
Chaetura pelagica	chimney swift	15373	19.8	1411	161.0	2364	51
Charadrius vociferus	killdeer	16856	19.1	1324	161.8	2482	52
Chlamydotis macqueenii	MacQueen's bustard	13582	12.9	1257	162.9	1734	10
Colius striatus	speckled mousebird	13538	12.4	1190	161.1	1754	11
Columba livia	pigeon	16652	18.3	1363	161.0	2277	46
Corvus brachyrhynchos	American crow	16562	17.9	1363	161.1	2220	48
Cuculus canorus	common cuckoo	15889	20.0	1400	160.7	2413	48
Egretta garzetta	little egret	16585	18.6	1274	160.7	2496	52
Eurypyga helias	sunbittern	13974	12.3	1193	163.9	1763	11
Falco peregrinus	peregrine falcon	16242	19.9	1403	160.7	2389	49
Fulmarus glacialis	northern fulmar	14306	12.8	1230	163.0	1765	11
Gallus gallus	chicken	16516	21.1	1433	158.1	2437	48
Gavia stellata	red-throated loon	13454	13.2	1250	162.1	1776	11
Geospiza fortis	medium ground-finch	16286	17.9	1362	160.1	2198	46
Haliaeetus albicilla	white-tailed eagle	13831	14.2	1258	161.1	1903	12
Haliaeetus	bald eagle	16526	19.0	1359	160.7	2370	36

leucocephalus							
Leptosomus discolor	cuckoo roller	14831	13.9	1236	163.2	1926	14
Manacus vitellinus	golden-collared manakin	15285	18.8	1392	159.7	2262	46
Meleagris gallopavo	turkey	16051	17.4	1305	158.0	2215	52
Melopsittacus undulatus	budgerigar	15470	19.8	1395	162.2	2415	52
Merops nubicus	carmine bee-eater	13467	13.0	1224	162.1	1798	11
Mesitornis unicolor	brown mesite	15371	11.4	1169	163.6	1666	11
Nestor notabilis	kea	14074	14.4	1307	160.1	1822	12
Nipponia nippon	crested ibis	16756	19.4	1358	161.2	2434	51
Ophisthocomus hoazin	hoatzin	15702	20.0	1336	162.1	2582	55
Pelecanus crispus	dalmatian pelican	14813	11.9	1183	164.8	1740	11
Phaethon lepturus	white-tailed tropicbird	14970	12.7	1220	163.9	1781	11
Phalacrocorax carbo	great cormorant	13479	13.5	1258	162.0	1810	11
Phoenicopterus ruber ruber	American flamingo	14024	11.7	1179	165.3	1716	10
Picoides pubescens	downy woodpecker	15576	20.0	1390	161.7	2450	47
Podiceps cristatus	great-crested grebe	13913	10.4	1137	165.8	1583	8
Pterocles gutturalis	yellow-throated sandgrouse	13867	12.8	1235	162.5	1757	11
Pygoscelis adeliae	Adelie penguin	15270	21.3	1392	160.3	2589	58
Struthio camelus	common ostrich	16178	19.5	1289	161.0	2601	54
Taeniopygia guttata	zebra finch	17471	21.4	1383	153.5	2493	53
Tauraco erythrolophus	red-crested turaco	15435	13.2	1200	164.0	1894	12
Tinamus guttatus	white-throated tinamou	15773	14.7	1288	162.0	1934	25

Tyto alba	barn owl	13613	13.8	1240	160.8	1871	12

Table S4. Percentages of genome annotated as transposable elements (TEs). The species are ordered according to phylogenetic relationship of the avian TENT tree.

Species	**Common name**	**LINE**	**SINE**	**LTR**	**DNA**	**RC**	**Unknown**	**Other**	**Total (%)**
Merops nubicus	carmine bee-eater	5.01213	0.06999	1.29546	0.13936	0.00535	1.25607	0.00028	7.77865
Picoides pubescens	downy woodpecker	18.20150	0.04604	0.89066	0.16695	0.00429	2.83735	0.00046	22.14730
Buceros rhinoceros	rhinoceros hornbill	3.61909	0.07569	1.04809	0.16020	0.00592	1.08907	0.00010	5.99816
Apaloderma vittatum	bar-tailed trogon	5.96761	0.11760	1.30969	0.22583	0.00513	0.81509	0.00021	8.44116
Leptosomus discolor	cuckoo roller	2.93288	0.11669	1.32311	0.19283	0.00536	1.88262	0.00015	6.45363
Colius striatus	speckled mousebird	6.53918	0.09797	2.19283	0.19214	0.00378	0.38859	0.00051	9.41500
Haliaeetus albicilla	white-tailed eagle	2.55429	0.13944	1.70885	0.18662	0.00525	0.77347	0.00015	5.36808
Haliaeetus leucocephalus	bald eagle	2.00515	0.17320	1.89425	0.22054	0.00317	2.59326	0.00025	6.88982
Cathartes aura	turkey vulture	2.20702	0.16981	1.04973	0.19008	0.00450	0.91728	0.00011	4.53853
Tyto alba	barn owl	2.63807	0.12622	1.78848	0.18667	0.01436	0.73733	0.00019	5.49133
Geospiza fortis	medium ground-finch	3.64772	0.05835	3.37070	0.31177	0.03877	0.79969	0.00061	8.22762
Taeniopygia guttata	zebra finch	3.78953	0.05858	4.10936	0.32025	0.01772	1.38779	0.00073	9.68395
Corvus brachyrhy	American crow	3.72969	0.07357	2.42901	0.21861	0.01534	0.89940	0.00040	7.36603

nchos									
Manacus vitellinus	golden-collared manakin	4.43477	0.08049	1.08186	0.24633	0.00877	0.72441	0.00033	6.57695
Acanthisitta chloris	rifleman	6.38240	0.10215	1.45738	0.21059	0.00967	0.56122	0.00023	8.72363
Melopsittacus undulatus	budgerigar	6.48990	0.07715	1.96803	0.20069	0.00752	0.44785	0.00016	9.19131
Nestor notabilis	kea	4.59637	0.10459	1.31740	0.18054	0.00478	0.36856	0.00015	6.57240
Falco peregrinus	peregrine falcon	3.08680	0.14939	1.26518	0.28248	0.00437	0.70956	0.00031	5.49809
Cariama cristata	red-legged seriema	3.51351	0.17871	0.90914	0.19647	0.00352	0.68527	0.00018	5.48680
Egretta garzetta	little egret	3.91707	0.12260	1.41923	0.24157	0.00577	1.21957	0.00025	6.92605
Pelecanus crispus	dalmatian pelican	3.93807	0.15337	1.87329	0.21447	0.00565	1.26833	0.00019	7.45337
Nipponia nippon	crested ibis	3.68685	0.13068	1.22028	0.28937	0.00650	0.82740	0.00031	6.16138
Phalacrocorax carbo	great cormorant	3.94906	0.15573	1.29058	0.20761	0.00412	0.62450	0.00026	6.23186
Aptenodytes forsteri	emperor penguin	2.40622	0.19808	1.16966	0.26165	0.00294	1.45700	0.00024	5.49579
Pygoscelis adeliae	Adelie penguin	3.31003	0.19502	1.32086	0.25858	0.00278	0.94761	0.00025	6.03513
Fulmarus glacialis	northern fulmar	2.86062	0.17656	1.18551	0.21739	0.00646	0.87302	0.00016	5.31971
Gavia stellata	red-throated loon	3.16569	0.13876	0.71228	0.21764	0.00560	0.84803	0.00014	5.08814
Eurypyga helias	sunbittern	4.60858	0.09861	1.59535	0.14741	0.00484	0.46440	0.00017	6.91935

Phaethon lepturus	white-tailed tropicbird	3.91321	0.11787	1.70552	0.22142	0.00480	1.47534	0.00020	7.43835
Ophisthocomus hoazin	hoatzin	4.69201	0.11140	1.29602	0.15839	0.00787	1.63178	0.00022	7.89768
Balearica regulorum	grey-crowned crane	3.35137	0.13632	1.51228	0.23695	0.00587	0.83417	0.00017	6.07712
Charadrius vociferus	killdeer	4.53428	0.12823	1.11570	0.19605	0.00650	1.05240	0.00031	7.03347
Calypte anna	Anna's hummingbird	5.61961	0.06831	1.22855	0.21410	0.00990	0.90731	0.00032	8.04810
Chaetura pelagica	chimney swift	5.28185	0.10761	0.89502	0.18882	0.00352	2.56957	0.00056	9.04695
Antrostomus carolinensis	chuck-will's-widow	5.40314	0.11790	1.83793	0.32737	0.01747	0.53289	0.00032	8.23702
Chlamydotis macqueenii	MacQueen's bustard	3.97021	0.16962	1.40330	0.23258	0.00454	0.57444	0.00024	6.35493
Tauraco erythrolophus	red-crested turaco	2.75721	0.08774	1.80381	0.15721	0.00604	3.82672	0.00016	8.63889
Cuculus canorus	common cuckoo	7.83916	0.07752	0.66931	0.27460	0.00573	0.58100	0.00031	9.44763
Mesitornis unicolor	brown mesite	4.61872	0.09126	1.38263	0.38166	0.00925	1.03021	0.00041	7.51414
Pterocles gutturalis	yellow-throated sandgrouse	3.46226	0.09022	1.36026	0.17044	0.00660	0.66503	0.00013	5.75494
Columba livia	pigeon	4.18253	0.08584	0.75945	0.34731	0.00770	1.86770	0.00035	7.25089

Phoenicopterus ruber	American flamingo	2.68686	0.14818	1.03641	0.23324	0.00541	1.49416	0.00016	5.60442
Podiceps cristatus	great-crested grebe	4.80367	0.10176	1.59644	0.20172	0.00562	0.60107	0.00016	7.31045
Gallus gallus	chicken	6.01090	0.07659	1.65347	1.00501	0.00780	1.06650	0.00030	9.82058
Meleagris gallopavo	turkey	5.39961	0.05268	1.10563	0.82249	0.00478	0.51645	0.00016	7.90179
Anas platyrhynchos	Peking duck	4.05220	0.09728	1.09549	0.20375	0.00676	0.39224	0.00028	5.84800
Struthio camelus	common ostrich	2.87702	0.18499	0.16645	0.35945	0.00781	0.89673	0.00016	4.49261
Tinamus guttatus	white-throated tinamou	2.73101	0.08748	0.30251	0.32546	0.00995	0.65253	0.00034	4.10928
Alligator mississippiensis	American alligator	6.63859	0.50463	0.35148	0.67678	0.01057	0.05865	0.00017	8.24086
Chelonia mydas	green sea turtle	6.00328	1.39609	0.71170	1.01453	0.00605	0.13660	0.00011	9.26836

Table S5. lncRNA genes predicted with RNA-seq data of chicken.

Length	**#genes**
>=200bp	5879
>=500bp	3607
>=1000bp	1628
>=2000bp	548

Table S6. Basic statistics of genome assemblies of outgroup species. The repetitive content of the American alligator and green sea turtle was annotated by RepeatMasker

based on the Repbase library (version: 20110419). The repetitive content in other genomes was based on the available annotation files from Ensembl/NCBI/BGI. Note that in some analyses some specific outgroups could be used, as indicated in their own method details.

Species	**Common name**	**Source**	**Assembly result (contig/scaffold N50;total length)**	**Repeat coverage (%)**	**Publication**
Alligator mississippiensis	American alligator	St John et al. (2012)	28K/106K;2.2G	11.4	(*24, 106*)
Chelonia mydas	green sea turtle	Wang et al. (2013)	20K/3.8M;2.2G	10.1	(*105*)
Anolis carolinensis	green anole lizard	Alföldi et al. (2011)	80K/156.5M;1.8G	7.45	(*107*)
Ailuropoda melanoleuca	giant panda	Ensembl 69	40K/132M/2.3 G	38.6	(*97*)
Bos taurus	cow	Ensembl 69	96K/105.7M/2.7 G	48.7	(*183*)
Callithrix jacchus	marmoset	Ensembl 69	29K/0.1M/2.9 G	44.8	(*99*)
Canis familiaris	domestic dog	Ensembl 67	180K/67.2M/2.5 G	39.6	(*46*)
Equus caballus	horse	Ensembl 69	112K/84.7M/2.5 G	40.2	(*184*)
Felis catus	domestic cat	Ensembl 69	21K/148.5M/2.5 G	39.6	(*118*)
Gorilla gorilla	gorilla	Ensembl 69	12K/145.3M/3.0 G	48.1	(*185*)
Heterocephalus glaber	naked mole rat	BGI	19K/1.6M/2.7 G	34.5	(*116*)
Homo sapiens	human	Ensembl 69	38M146.4M/3.1 G	48.2	(*23*)
Ictidomys tridecemlineatus	squirrel	Ensembl 69	44K/8M/2.5 G	36.6	(*46*)
Loxodonta africana	elephant	Ensembl 69	69K/46M/3.2 G	46.3	(*46*)
Macaca mulatta	rhesus	Ensembl 69	26K/147.8M/3.1 G	46.2	(*186*)
Monodelphis	opossum	Ensembl 69	108K/528M/3.6G	29.1	(*187*)

domestica					
Mus musculus	mouse	Ensembl 69	28M/130.7M/2.7 G	44.6	(*188*)
Mustela putorius furo	ferret	Ensembl 69	45K/9.3M/2.4 G	38.1	(*99*)
Myotis davidii	David's myotis	BGI	13K/3.4M/2.1 G	33.2	(*117*)
Ornithorhynchus anatinus	platypus	Ensembl 69	11K/0.8M/2.1 G	42.5	(*189*)
Oryctolagus cuniculus	rabbit	Ensembl 69	65K/112M/2.7 G	42.7	(*46*)
Pan troglodytes	chimp	Ensembl 64	30K/145.1M/3.5 G	41.1	(*190*)
Pongo abelii	orangutan	Ensembl 69	15K/135.2M/3.4 G	46.4	(*191*)
Pteropus alecto	black flying fox	BGI	26K/15.8M/2.0 G	30.1	(*117*)
Rattus norvegicus	rat	Ensembl 69	33K/147.6M/2.7 G	41.2	(*192*)
Sus scrofa	pig	Ensembl 69	57K/148.5M/2.8 G	39.4	(*119*)
Tupaia belangeri	tree shrew	BGI	22K/3.7M/2.9 G	44.6	(*115*)

Table S7 Statistics of protein-coding gene annotations of reptiles and mammals.

Species	**Common name**	**#Gene**	**Mean gene length (kb)**	**Mean CDS length (bp)**	**Mean exon length (bp)**	**Mean intron length (bp)**	**Mean intergenic length (kb)**
Alligator mississippiensis	America alligator	18287	17.1	1182	170.7	2680	24
Anolis carolinensis	green anole lizard	17805	23.5	1526	159.8	2566	78

Chelonia mydas	green sea turtle	18971	25.4	1170	163.3	3928	81
Ailuropoda melanoleuca	panda	19343	34.4	1591	162.6	3738	52
Bos taurus	cow	19994	35.1	1595	166.9	3919	101
Callithrix jacchus	marmoset	20993	39.1	1492	160.5	4531	104
Canis familiaris	dog	19305	30.2	1565	159.7	3254	97
Equus caballus	horse	20436	31.6	1528	165.2	3648	89
Felis catus	cat	21718	52.8	1731	163.4	4863	91
Gorilla gorilla	gorilla	20962	40.2	1514	164.0	4700	111
Heterocephalus glaber	naked mole rat	22561	32.5	1439	178.7	4410	62
Homo sapiens	human	20424	48.9	1663	173.5	5500	104
Ictidomys tridecemlineatus	squirrel	18826	27.7	1520	168.6	3266	86
Loxodonta africana	elephant	20033	36.5	1526	163.7	4202	122
Macaca mulatta	rhesus	21905	38.3	1388	164.3	4955	103
Monodelphis domestica	opossum	21327	51.2	1584	180.3	6368	126
Mus musculus	mouse	23038	35.1	1538	179.4	4429	84
Mustela putorius furo	ferret	19910	38.7	1601	168.7	4372	78
Myotis davidii	David's myotis	21705	29.1	1400	175.0	3958	59
Ornithorhynchus anatinus	platypus	21698	20.1	1138	149.9	2878	72
Oryctolagus cuniculus	rabbit	19018	36.5	1542	163.6	4150	108
Pan troglodytes	chimpanzee	19829	47.1	1557	163.8	5356	115
Pongo abelii	orangutan	20424	45.2	1490	155.8	5102	124

Pteropus alecto	black flying fox	21392	28.2	1465	177.7	3694	67
Rattus norvegicus	rat	22938	30.5	1429	168.8	3892	94
Sus scrofa	pig	21829	32.3	1409	169.2	4218	97
Tupaia belangeri	tree shrew	22063	33.7	1405	186.2	4417	33

Table S8. Lost syntenic blocks in common ostrich, American alligator, and green sea turtle using green anole lizard as the reference genome.

Species	No. of lost blocks	Total length (Mb)	# Genes
common ostrich	118	58.18	1,241
American alligator	14	8.86	78
green sea turtle	27	7.73	190

Table S9. Detailed information of lost syntenic blocks in common ostrich using green anole lizard as the reference genome (only legend provided here, full table in a separate tab-delimited text file).

Table S10. The 640 genes absent in modern birds determined by BLAST analyses (only legend provided here, full table in a separate tab-delimited text file). Human genes served as queries, and a human gene is defined as lost in the ancestor of modern birds if it is present in at least one of the five non-avian reptile genomes but abesent (completely lost, low aligning rate (i.e. <30%) or pseudo) in all the 48 avain genomes.

Table S11. Genes involved in important functions (e.g. blood, bone, lung, muscle, etc.) that are absent in modern birds (only legend provided here, full table in a separate tab-delimited text file).

Table S12. Number of detected and expected EBRs in each lineage. * Total number of EBRs does not include the reused EBRs, because these were counted as lineage or order specific in corresponding lineages.

Phylogenetic Level	Detected no. EBRs	Expected no. EBRs
Species-specific		

Peking duck	107	126
emperor penguin	39	42
Anna's hummingbird	94	104
Chicken	23	NA
Pigeon	92	99
American crow	35	40
common cuckoo	97	103
little egret	37	40
peregrine falcon	96	105
medium ground-finch	31	37
golden-collared manakin	30	36
turkey	246	NA
budgerigar	193	210
hoatzin	40	42
crested ibis	37	56
downy woodpecker	162	207
Adelie penguin	44	49
common ostrich	112	122
zebra finch	46	56
green anole lizard	170	259
boa snake	18	42
Clade-specific		
Galliformes	50	50
Galloanserae	11	15
Cuculiformes + Trochiliformes	2	2

Ciconiiformes	3	3
Passeroidea	13	13
Passeroidea + Corvoidea	20	21
Passeriformes	18	20
Passeriformes + Psittaciformes	2	3
Sphenisciformes	3	3
Non-galloanserae	7	10
Non-galloanserae + non-columbiformes	2	2
Neognathae	6	7
Total avian EBR	1731*	
Total EBRs	1919*	--
Reused EBRs	235	--

Table S13. Rearrangement rates using Jetz *et al.* (2012) (*7*) phylogenetic tree.*Age is the estimated time of the speciation node in million years (MY) and diversification rate (r) is the mean per lineage diversification rate (in units of MY-1) from (*7*). ** Rearrangement rates are in units of EBRs/MY. *** Turkey genome was not included in this analysis.

Clade	**Age***	**Diversification rate (r)***	**Number of species included**	**Rearrangement rate ****
Passeroidea	35.6	0.2	3	1.36
Passeriformes	66.8	0.14	4	0.72
Psittaciformes + Passeriformes	77.2	0.14	5	1
Galliformes	53.9	0.13	1***	1.35
Anseriformes	10.8	0.28	1	9.91

Galloanserae	78.7	0.08	2***	1.21

Table S14. Mammalian species used in analyses of substitution rates and genomic conservation.

Species name	Common name	UCSC version	Publication
Homo sapiens	Human	hg18	(*23*)
Pan troglodytes	Chimpanzee	panTro2	(*190*)
Macaca mulatta	Rhesus	rheMac2	(*186*)
Otolemur garnettii	Bushbaby	otoGar1	(*46*)
Tupaia belangeri	tree shrew	tupBel1	(*46*)
Mus musculus	Mouse	mm8	(*188*)
Rattus norvegicus	Rat	rn4	(*192*)
Cavia porcellus	Guinea pig	cavPor2	(*46*)
Oryctolagus cuniculus	Rabbit	oryCun1	(*46*)
Sorex araneus	Shrew	sorAra1	(*46*)
Erinaceus europaeus	Hedgehog	eriEur1	(*46*)
Canis lupus	Dog	canFam2	(*46*)
Felis catus	Cat	felCat3	(*118*)
Equus caballus	Horse	equCab1	(*184*)
Bos taurus	Cow	bosTau3	(*183*)
Dasypus novemcinctus	Armadillo	dasNov1	(*46*)
Loxodonta africana	Elephant	loxAfr1	(*46*)
Echinops telfairi	Tenrec	echTel1	(*46*)

Table S15. Statistics of highly conserved elements identified in birds and mammals by phastCons.

	48 birds		18 mammals		Shared by birds and mammals	
	#segments	Total length(bp)	#segments	Total length(bp)	#segments	Total length(bp)
>=1bp	7,097,338	111,818,463	1,912,729	89,447,617	1,030,365	24,961,446
>=10bp	3,235,598	90,338,593	1,706,891	87,901,515	651,026	22,627,679
>=20bp	1,441,723	66,243,831	1,128,067	79,743,604	362,162	18,597,228
>=100bp	100,190	16,656,033	221,308	40,873,074	32,950	5,791,380
>=500bp	1,225	777,262	5,223	3,612,266	583	369,271
>=1000bp	32	38,003	405	526,482	16	18,936

Table S16. Statistics of human TFBSes found in HCEs of birds and mammals. The human TFBSes were downloaded from UCSC ENCODE website (hg18). Then human TFBSes were mapped to the chicken genome to assess the conservation in birds using the liftOver tool provided by UCSC.

Total number	Conserved in mammals	Conserved in birds	Conserved in mammals and birds	Avian-specific conserved
1,582,526	617,287	46,717	31,075	15,642

Table S17. New exons identified in the HCEs. 'Cons. ratio' means how much of the exon is covered by HCEs. The results mentioned in the main text were those exons which are 100% covered by HCEs. See the sources of the expression data in the SM text1.

Cons. ratio	#new exons		#related genes		#genes with conserved new exons and known exons		#new genes (some but not all exons are conserved)		#new genes (all exons are conserved)	
	total	expressed	total	expressed	total	expressed	total	expressed	total	expressed
>=10%	5916	3457	2675	2341	1973	1846	624	440	78	55
>=30%	4828	2831	2186	1913	1581	1474	555	403	50	36

>=50 %	4039	2362	1858	1626	1317	1223	506	377	35	26
>=70 %	3056	1772	1515	1329	1053	982	448	336	14	11
>=90 %	1818	1030	1036	911	723	678	305	227	8	6
100%	717	354	525	471	388	365	136	105	1	1

Table S18. Numbers of lncrRNAs (length >=200bp) overlapping with highly conserved elements (HCEs). See **Table S5** for statistics of annotated lncRNAs.

% of HCE length overlapping HCEs	#lncRNAs
>=10%	1120
>=30%	528
>=50%	220
>=70%	88
>=90%	29

Table S19. Genes highly conserved in birds but divergent in mammals. 'Conserved ratio' refers to the percentage of coding region of a gene covered by HCEs. "*", located in the human pseudoautosomal region (PAR).

Chicken Gene	**Human Gene**	**Bird CDS conserved ratio**	**Mammal CDS conserved ratio**	**Gene Name**
ENSGALG00000015595	ENSG00000155269	0.810089	0	*GPR78*
ENSGALG00000016717	ENSG00000178605	0.937398	0.0482315	*GTPBP6*
ENSGALG00000016702	ENSG00000167393	0.905983	0.0949074	*PPP2R3B**

ENSGALG00000016689	ENSG00000169093	0.766667	0.105061	*ASMTL**
ENSGALG00000015713	ENSG00000174137	0.854067	0.108974	*FAM53A**
ENSGALG00000016636	ENSG00000157399	0.763128	0.122599	*ARSE*
ENSGALG00000004881	ENSG00000167110	0.771144	0.153207	*GOLGA2*
ENSGALG00000016655	ENSG00000056998	0.893894	0.156042	*GYG2*
ENSGALG00000011368	ENSG00000022976	0.915954	0.181034	*ZNF839*
ENSGALG00000016623	ENSG00000130021	0.824561	0.187831	*HDHD1A*
ENSGALG00000014496	ENSG00000007062	0.946175	0.198844	*PROM1*
ENSGALG00000008761	ENSG00000106304	0.871333	0.231771	*SPAM1*
ENSGALG00000001518	ENSG00000141579	0.753603	0.241713	*ZNF750*

Table S20. Enriched GO terms in the genes near the avian specific conserved TFBSes. Statistical analyses conducted with chi-square test, FDR adjusted $p < 0.05$. Class abbreviations: BP = biological process, MF = molecular function, and CC = cellular component.

GO ID	**GO Term**	**Class**	**P-value (FDR)**	**#gene**
GO:0032993	protein-DNA complex	CC	4.91E-08	29
GO:0044237	cellular metabolic process	BP	6.34E-08	660
GO:0016043	cellular component organization	BP	1.28E-06	88
GO:0044085	cellular component biogenesis	BP	3.47E-06	51
GO:0043170	macromolecule metabolic process	BP	3.47E-06	592
GO:0044464	cell part	CC	4.85E-06	646
GO:0003700	sequence-specific DNA binding transcription factor activity	MF	5.29E-06	125
GO:0044238	primary metabolic process	BP	6.84E-06	722
GO:0006807	nitrogen compound metabolic process	BP	9.03E-06	367
GO:0019222	regulation of metabolic process	BP	9.06E-06	205
GO:0071841	cellular component organization or	BP	4.69E-05	78
GO:0006928	cellular component movement	BP	6.25E-05	15
GO:0007165	signal transduction	BP	1.26E-04	167

GO:0036094	small molecule binding	MF	4.98E-04	346
GO:0005515	protein binding	MF	5.43E-04	848
GO:0009058	biosynthetic process	BP	6.16E-04	300
GO:0005622	intracellular	CC	7.67E-04	271
GO:0003735	structural constituent of ribosome	MF	1.66E-03	38
GO:0003676	nucleic acid binding	MF	2.87E-03	330
GO:0043229	intracellular organelle	CC	4.02E-03	322
GO:0009056	catabolic process	BP	4.49E-03	93
GO:0040008	regulation of growth	BP	7.62E-03	8
GO:0043228	non-membrane-bounded organelle	CC	8.29E-03	136
GO:0044424	intracellular part	CC	9.36E-03	383
GO:0032879	regulation of localization	BP	1.19E-02	12
GO:0007275	multicellular organismal development	BP	1.35E-02	34
GO:0065009	regulation of molecular function	BP	1.35E-02	32
GO:0050789	regulation of biological process	BP	1.58E-02	520
GO:0050794	regulation of cellular process	BP	1.67E-02	515
GO:0044420	extracellular matrix part	CC	1.74E-02	11
GO:0030529	ribonucleoprotein complex	CC	2.01E-02	41
GO:0043227	membrane-bounded organelle	CC	2.15E-02	216
GO:0043167	ion binding	MF	2.76E-02	400
GO:0040012	regulation of locomotion	BP	4.33E-02	4
GO:0008047	enzyme activator activity	MF	4.80E-02	29

Table S21. Enriched GO terms in the genes near the shared conserved TFBSes between birds and mammals. Statistical analyses conducted with chi-square test, FDR adjusted $p < 0.05$. Class abbreviations: BP = biological process, MF = molecular function, and CC = cellular component.

GO ID	**GO Term**	**Class**	**P-value (FDR)**	**#gene**
GO:0005515	protein binding	MF	1.15E-19	1841
GO:0003700	sequence-specific DNA binding transcription factor activity	MF	1.76E-13	251
GO:0043167	ion binding	MF	3.20E-09	871
GO:0003676	nucleic acid binding	MF	3.26E-07	683

GO:0044237	cellular metabolic process	BP	3.26E-07	1274
GO:0007165	signal transduction	BP	6.37E-07	325
GO:0019222	regulation of metabolic process	BP	8.89E-07	388
GO:0007275	multicellular organismal development	BP	1.46E-05	69
GO:0006807	nitrogen compound metabolic process	BP	2.19E-05	702
GO:0043170	macromolecule metabolic process	BP	2.96E-05	1146
GO:0036094	small molecule binding	MF	2.96E-05	685
GO:0044238	primary metabolic process	BP	4.61E-05	1417
GO:0005622	intracellular	CC	5.91E-04	526
GO:0048583	regulation of response to stimulus	BP	8.18E-04	123
GO:0060589	nucleoside-triphosphatase regulator activity	MF	8.29E-04	115
GO:0023051	regulation of signaling	BP	1.32E-03	122
GO:0044464	cell part	CC	1.91E-03	1237
GO:0032993	protein-DNA complex	CC	2.33E-03	35
GO:0009719	response to endogenous stimulus	BP	1.32E-02	31
GO:0030312	external encapsulating structure	CC	1.38E-02	8
GO:0042597	periplasmic space	CC	1.38E-02	8
GO:0044462	external encapsulating structure part	CC	1.38E-02	8
GO:0009058	biosynthetic process	BP	2.07E-02	565
GO:0042221	response to chemical stimulus	BP	2.07E-02	41
GO:0048518	positive regulation of biological process	BP	2.27E-02	13
GO:0016740	transferase activity	MF	2.39E-02	536
GO:0048522	positive regulation of cellular process	BP	3.66E-02	12
GO:0016043	cellular component organization	BP	4.53E-02	130
GO:0031012	extracellular matrix	CC	4.82E-02	45

Table S22. Mammalian species used for global *dN/dS* analysis. The citation information is in **Table S6**.

Species	**Common name**	**Source**
Ailuropoda melanoleuca	giant panda	Ensembl 69
Bos taurus	cow	Ensembl 69

Callithrix jacchus	marmoset	Ensembl 69
Canis familiaris	domestic dog	Ensembl 67
Equus caballus	horse	Ensembl 69
Felis catus	domestic cat	Ensembl 69
Gorilla gorilla	gorilla	Ensembl 69
Heterocephalus glaber	naked mole rat	BGI
Homo sapiens	human	Ensembl 69
Ictidomys tridecemlineatus	squirrel	Ensembl 69
Loxodonta africana	elephant	Ensembl 69
Macaca mulatta	rhesus	Ensembl 69
Monodelphis domestica	opossum	Ensembl 69
Mus musculus	mouse	Ensembl 69
Mustela putorius furo	ferret	Ensembl 69
Myotis davidii	David's myotis	BGI
Ornithorhynchus anatinus	platypus	Ensembl 69
Oryctolagus cuniculus	rabbit	Ensembl 69
Pan troglodytes	chimp	Ensembl 64
Pongo abelii	orangutan	Ensembl 69
Pteropus alecto	black flying fox	BGI
Rattus norvegicus	rat	Ensembl 69
Sus scrofa	pig	Ensembl 69
Tupaia belangeri	tree shrew	BGI

Table S23. GOs of genes evolving faster in birds compared to mammals, based on the *dN/dS* ratios of orthologs of birds and mammals in each GO. The p-values were computed using Wilcoxon signed-rank test, with a cutoff of 0.05. Only the GOs with >=10 genes were considered.

GO ID	GO term	#genes	P-value
GO:0007017	microtubule-based process	155	6.59E-06
GO:0006259	DNA metabolic process	264	1.43E-05
GO:0000226	microtubule cytoskeleton organization	111	7.11E-05
GO:0051297	centrosome organization	22	8.85E-05
GO:0022403	cell cycle phase	194	9.93E-05
GO:0022402	cell cycle process	275	0.000168239
GO:0034660	ncRNA metabolic process	98	0.000275159
GO:0016072	rRNA metabolic process	32	0.000386481
GO:0010927	cellular component assembly involved in morphogenesis	57	0.000388553
GO:0007098	centrosome cycle	15	0.000427246
GO:0031023	microtubule organizing center organization	25	0.000455946
GO:0060271	cilium morphogenesis	47	0.00076669
GO:0007049	cell cycle	349	0.000819201
GO:0000279	M phase	134	0.000850537
GO:0000278	mitotic cell cycle	165	0.000870408
GO:0034470	ncRNA processing	63	0.00109334
GO:0006364	rRNA processing	28	0.001219895
GO:0051928	positive regulation of calcium ion transport	17	0.001289368
GO:0042742	defense response to bacterium	18	0.001682281
GO:0051298	centrosome duplication	12	0.001708984
GO:0042384	cilium assembly	31	0.001798591

GO:0007018	microtubule-based movement	54	0.002186084
GO:0034644	cellular response to UV	12	0.002441406
GO:0071482	cellular response to light stimulus	14	0.002624512
GO:0050830	defense response to Gram-positive bacterium	10	0.002929688
GO:0030098	lymphocyte differentiation	76	0.003156567
GO:0045619	regulation of lymphocyte differentiation	26	0.003344595
GO:0009411	response to UV	39	0.003422249
GO:0071103	DNA conformation change	43	0.003425121
GO:0000723	telomere maintenance	18	0.003845215
GO:0032200	telomere organization	18	0.003845215
GO:0051168	nuclear export	32	0.00389808
GO:0007059	chromosome segregation	51	0.004278577
GO:0071214	cellular response to abiotic stimulus	49	0.004564808
GO:2001251	negative regulation of chromosome organization	12	0.004638672
GO:0002263	cell activation involved in immune response	38	0.004781957
GO:0002366	leukocyte activation involved in immune response	38	0.004781957
GO:0051320	S phase	15	0.005126953
GO:0030001	metal ion transport	162	0.005254574
GO:0007051	spindle organization	29	0.005289147
GO:0071345	cellular response to cytokine stimulus	78	0.005461314
GO:0015849	organic acid transport	53	0.006591739
GO:0046942	carboxylic acid transport	53	0.006591739
GO:0002285	lymphocyte activation involved in immune response	32	0.006615164
GO:0007126	meiosis	45	0.007590112

GO:0051327	M phase of meiotic cell cycle	45	0.007590112
GO:0042254	ribosome biogenesis	41	0.00829211
GO:0051179	localization	1324	0.009027785
GO:0032846	positive regulation of homeostatic process	20	0.009616852
GO:0030183	B cell differentiation	26	0.01021431
GO:0071841	cellular component organization or biogenesis at cellular level	975	0.0102695
GO:0002562	somatic diversification of immune receptors via germline recombination within a single locus	18	0.01183701
GO:0016444	somatic cell DNA recombination	18	0.01183701
GO:0006811	ion transport	280	0.01200017
GO:0042113	B cell activation	55	0.0122359
GO:0051321	meiotic cell cycle	48	0.01237452
GO:0090304	nucleic acid metabolic process	1093	0.01264127
GO:0051325	interphase	71	0.01319882
GO:0034097	response to cytokine stimulus	93	0.01326063
GO:0033261	regulation of S phase	12	0.01342773
GO:0032760	positive regulation of tumor necrosis factor production	10	0.01367188
GO:0071705	nitrogen compound transport	52	0.01392757
GO:0015837	amine transport	46	0.01396126
GO:0071840	cellular component organization or biogenesis	1229	0.01437322
GO:0002200	somatic diversification of immune receptors	19	0.01446533
GO:0030217	T cell differentiation	55	0.01484345
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	38	0.01535643

GO:0032387	negative regulation of intracellular transport	25	0.01590392
GO:0042991	transcription factor import into nucleus	21	0.01595974
GO:0045453	bone resorption	11	0.01611328
GO:0002286	T cell activation involved in immune response	13	0.01635742
GO:0007010	cytoskeleton organization	249	0.01701597
GO:0006260	DNA replication	90	0.01706746
GO:0006289	nucleotide-excision repair	18	0.01711655
GO:0030509	BMP signaling pathway	34	0.01713361
GO:0009416	response to light stimulus	66	0.01723822
GO:0031109	microtubule polymerization or depolymerization	15	0.01766968
GO:0046823	negative regulation of nucleocytoplasmic transport	19	0.01803398
GO:0051924	regulation of calcium ion transport	30	0.0182175
GO:0002521	leukocyte differentiation	107	0.01830266
GO:0031400	negative regulation of protein modification process	57	0.01980424
GO:0002443	leukocyte mediated immunity	41	0.01985359
GO:0006812	cation transport	199	0.02018459
GO:0006281	DNA repair	134	0.02024611
GO:0003341	cilium movement	14	0.02093506
GO:0015844	monoamine transport	14	0.02093506
GO:0048538	thymus development	11	0.02099609
GO:0097191	extrinsic apoptotic signaling pathway	11	0.02099609
GO:0007283	spermatogenesis	90	0.02107871
GO:0022406	membrane docking	12	0.02124023
GO:0009451	RNA modification	31	0.02172683

GO:0070838	divalent metal ion transport	92	0.02256657
GO:0072511	divalent inorganic cation transport	92	0.02256657
GO:0060249	anatomical structure homeostasis	56	0.02305379
GO:0015908	fatty acid transport	18	0.0241394
GO:0042990	regulation of transcription factor import into nucleus	20	0.02422047
GO:0061337	cardiac conduction	10	0.02441406
GO:0006323	DNA packaging	26	0.02467343
GO:0032392	DNA geometric change	14	0.02471924
GO:0032508	DNA duplex unwinding	14	0.02471924
GO:0007127	meiosis I	21	0.02509594
GO:0031397	negative regulation of protein ubiquitination	16	0.02532959
GO:0048232	male gamete generation	91	0.02563167
GO:0006305	DNA alkylation	12	0.02612305
GO:0006306	DNA methylation	12	0.02612305
GO:0021510	spinal cord development	12	0.02612305
GO:0015909	long-chain fatty acid transport	11	0.02685547
GO:0009987	cellular process	3996	0.02735627
GO:0002449	lymphocyte mediated immunity	34	0.02738452
GO:0006405	RNA export from nucleus	15	0.02767944
GO:0007229	integrin-mediated signaling pathway	15	0.02767944
GO:0051235	maintenance of location	62	0.02780861
GO:0030307	positive regulation of cell growth	23	0.02812636
GO:0001775	cell activation	178	0.02834047
GO:0071842	cellular component organization at cellular level	942	0.02866391

GO:0060401	cytosolic calcium ion transport	22	0.02934504
GO:0060402	calcium ion transport into cytosol	22	0.02934504
GO:0071356	cellular response to tumor necrosis factor	22	0.02934504
GO:0033044	regulation of chromosome organization	37	0.02936513
GO:0043270	positive regulation of ion transport	29	0.02959816
GO:0000070	mitotic sister chromatid segregation	18	0.02996826
GO:0008033	tRNA processing	27	0.03094941
GO:0051329	interphase of mitotic cell cycle	68	0.03140016
GO:0032501	multicellular organismal process	1324	0.03159376
GO:0006974	response to DNA damage stimulus	197	0.03182363
GO:0001510	RNA methylation	12	0.03198242
GO:0035710	CD4-positive, alpha-beta T cell activation	12	0.03198242
GO:2001236	regulation of extrinsic apoptotic signaling pathway	10	0.03222656
GO:0030510	regulation of BMP signaling pathway	21	0.03232479
GO:0016043	cellular component organization	1195	0.03255278
GO:0002761	regulation of myeloid leukocyte differentiation	19	0.033144
GO:0030514	negative regulation of BMP signaling pathway	13	0.03405762
GO:0045321	leukocyte activation	155	0.03414413
GO:0006810	transport	1048	0.03487162
GO:0032844	regulation of homeostatic process	78	0.03491638
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	21	0.03506851
GO:0051234	establishment of localization	1065	0.03545903
GO:0006310	DNA recombination	69	0.03595695

GO:0034612	response to tumor necrosis factor	27	0.03657009
GO:0007052	mitotic spindle organization	16	0.03695679
GO:0032012	regulation of ARF protein signal transduction	16	0.03695679
GO:0015718	monocarboxylic acid transport	20	0.03792572
GO:0055001	muscle cell development	49	0.03825071
GO:0006399	tRNA metabolic process	51	0.03862261
GO:2000377	regulation of reactive oxygen species metabolic process	14	0.03924561
GO:0055002	striated muscle cell development	43	0.03942186
GO:0000087	M phase of mitotic cell cycle	79	0.0394673
GO:0006928	cellular component movement	303	0.03952884
GO:0051592	response to calcium ion	22	0.03971362
GO:0006952	defense response	158	0.03977041
GO:0019637	organophosphate metabolic process	108	0.04017459
GO:0071844	cellular component assembly at cellular level	334	0.0404805
GO:0045621	positive regulation of lymphocyte differentiation	18	0.04071426
GO:0006865	amino acid transport	28	0.04072697
GO:0019221	cytokine-mediated signaling pathway	53	0.04135784
GO:0045670	regulation of osteoclast differentiation	11	0.04150391
GO:0016447	somatic recombination of immunoglobulin gene segments	15	0.04162598
GO:0090317	negative regulation of intracellular protein transport	22	0.04270458
GO:0019724	B cell mediated immunity	17	0.04432678
GO:0006644	phospholipid metabolic process	97	0.04438099
GO:0034453	microtubule anchoring	20	0.04484749
GO:0006139	nucleobase-containing compound metabolic process	1393	0.0449874

GO:0042439	ethanolamine-containing compound metabolic process	14	0.04528809
GO:0007062	sister chromatid cohesion	12	0.04614258
GO:0012502	induction of programmed cell death	70	0.0470924
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	51	0.04714752
GO:0007140	male meiosis	15	0.04730225
GO:0000819	sister chromatid segregation	19	0.0477562
GO:0045580	regulation of T cell differentiation	19	0.0477562
GO:0030811	regulation of nucleotide catabolic process	92	0.04916961
GO:0033121	regulation of purine nucleotide catabolic process	92	0.04916961
GO:0032640	tumor necrosis factor production	17	0.04918671
GO:0032680	regulation of tumor necrosis factor production	17	0.04918671
GO:0071706	tumor necrosis factor superfamily cytokine production	17	0.04918671

Table S24. GOs of genes evolving faster in mammals compared to birds, based on the *dN/dS* ratios of orthologs of birds and mammals in each GO. The p-values were computed using Wilcoxon signed-rank test, with a cutoff of 0.05. Only the GOs with >=10 genes were considered.

GO ID	**GO term**	**#genes**	**P-value**
GO:0045454	cell redox homeostasis	32	0.000617646
GO:0001525	angiogenesis	97	0.001238581
GO:0048732	gland development	85	0.001305073
GO:0018130	heterocycle biosynthetic process	104	0.001609091
GO:0010799	regulation of peptidyl-threonine phosphorylation	10	0.001953125
GO:0005976	polysaccharide metabolic process	68	0.00235638
GO:0030850	prostate gland development	21	0.00314045

GO:0061383	trabecula morphogenesis	12	0.003417969
GO:0043405	regulation of MAP kinase activity	69	0.004335536
GO:0030218	erythrocyte differentiation	33	0.00439418
GO:0034637	cellular carbohydrate biosynthetic process	57	0.004474799
GO:0015758	glucose transport	23	0.00557828
GO:0019438	aromatic compound biosynthetic process	12	0.006103516
GO:0009126	purine nucleoside monophosphate metabolic process	17	0.00643158
GO:0048666	neuron development	163	0.006823796
GO:0043491	protein kinase B signaling cascade	30	0.006831621
GO:0005978	glycogen biosynthetic process	11	0.006835937
GO:0007416	synapse assembly	11	0.006835937
GO:0009250	glucan biosynthetic process	11	0.006835937
GO:0031346	positive regulation of cell projection organization	34	0.00727676
GO:0018205	peptidyl-lysine modification	62	0.007636684
GO:0018210	peptidyl-threonine modification	21	0.007890224
GO:0042542	response to hydrogen peroxide	18	0.007965088
GO:0009127	purine nucleoside monophosphate biosynthetic process	12	0.00805664
GO:0009168	purine ribonucleoside monophosphate biosynthetic process	12	0.00805664
GO:0010464	regulation of mesenchymal cell proliferation	14	0.008300781
GO:0030879	mammary gland development	37	0.008665251
GO:0009165	nucleotide biosynthetic process	83	0.008840776
GO:0050931	pigment cell differentiation	15	0.009033203
GO:0018107	peptidyl-threonine phosphorylation	19	0.009040833
GO:0043406	positive regulation of MAP kinase activity	52	0.009057847

GO:0009123	nucleoside monophosphate metabolic process	64	0.009203059
GO:0033673	negative regulation of kinase activity	50	0.009251388
GO:0001825	blastocyst formation	11	0.009277344
GO:0043583	ear development	52	0.009512747
GO:0001754	eye photoreceptor cell differentiation	10	0.009765625
GO:0007162	negative regulation of cell adhesion	27	0.01000088
GO:0009167	purine ribonucleoside monophosphate metabolic process	16	0.01069641
GO:0015749	monosaccharide transport	25	0.01182497
GO:0048608	reproductive structure development	75	0.01182652
GO:0033692	cellular polysaccharide biosynthetic process	36	0.01194548
GO:0034654	nucleobase-containing compound biosynthetic process	88	0.01199579
GO:0051897	positive regulation of protein kinase B signaling cascade	15	0.01278687
GO:0009124	nucleoside monophosphate biosynthetic process	51	0.01300045
GO:0009108	coenzyme biosynthetic process	31	0.01300467
GO:0044087	regulation of cellular component biogenesis	102	0.01321309
GO:0060740	prostate gland epithelium morphogenesis	10	0.01367188
GO:0042471	ear morphogenesis	31	0.0137178
GO:0048839	inner ear development	46	0.01396126
GO:0007565	female pregnancy	16	0.01449585
GO:0030318	melanocyte differentiation	14	0.01477051
GO:0009156	ribonucleoside monophosphate biosynthetic process	15	0.01507568
GO:0010827	regulation of glucose transport	11	0.01611328
GO:0002053	positive regulation of mesenchymal cell proliferation	13	0.01635742
GO:0006413	translational initiation	30	0.01636045

GO:0005975	carbohydrate metabolic process	238	0.01646011
GO:0044271	cellular nitrogen compound biosynthetic process	158	0.01694457
GO:0006367	transcription initiation from RNA polymerase II promoter	17	0.01739502
GO:0009060	aerobic respiration	15	0.01766968
GO:0006469	negative regulation of protein kinase activity	47	0.01788606
GO:0009161	ribonucleoside monophosphate metabolic process	19	0.01803398
GO:0006352	transcription initiation, DNA-dependent	27	0.01810306
GO:0008653	lipopolysaccharide metabolic process	20	0.01811695
GO:0034101	erythrocyte homeostasis	36	0.01813471
GO:0009063	cellular amino acid catabolic process	30	0.0182175
GO:0008645	hexose transport	24	0.01832461
GO:0090102	cochlea development	10	0.01855469
GO:0006164	purine nucleotide biosynthetic process	64	0.01923635
GO:0022612	gland morphogenesis	35	0.01936676
GO:0051896	regulation of protein kinase B signaling cascade	22	0.01953673
GO:0042472	inner ear morphogenesis	25	0.01966935
GO:0010463	mesenchymal cell proliferation	17	0.01976776
GO:0016568	chromatin modification	140	0.01998429
GO:0016569	covalent chromatin modification	114	0.02018897
GO:0030182	neuron differentiation	210	0.02056385
GO:0021675	nerve development	15	0.02062988
GO:0060485	mesenchyme development	46	0.02067924
GO:0051188	cofactor biosynthetic process	49	0.02085193
GO:0009112	nucleobase metabolic process	11	0.02099609

GO:0035588	G-protein coupled purinergic receptor signaling pathway	11	0.02099609
GO:0060512	prostate gland morphogenesis	11	0.02099609
GO:0021761	limbic system development	18	0.02158356
GO:0014032	neural crest cell development	19	0.02227974
GO:0010906	regulation of glucose metabolic process	23	0.02243793
GO:0044264	cellular polysaccharide metabolic process	44	0.02293751
GO:0031344	regulation of cell projection organization	67	0.0233199
GO:0015980	energy derivation by oxidation of organic compounds	51	0.02371231
GO:0048514	blood vessel morphogenesis	125	0.02379558
GO:0032272	negative regulation of protein polymerization	13	0.02392578
GO:0016570	histone modification	113	0.02469073
GO:0001568	blood vessel development	154	0.02484703
GO:0014031	mesenchymal cell development	39	0.02541628
GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	39	0.02541628
GO:0048592	eye morphogenesis	39	0.0271742
GO:0001944	vasculature development	160	0.02769822
GO:2000241	regulation of reproductive process	38	0.02841929
GO:0030521	androgen receptor signaling pathway	17	0.02844238
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	139	0.02846834
GO:0008643	carbohydrate transport	33	0.02876396
GO:0043112	receptor metabolic process	41	0.02897501
GO:0009103	lipopolysaccharide biosynthetic process	19	0.03010368
GO:0042490	mechanoreceptor differentiation	19	0.03010368

GO:0007265	Ras protein signal transduction	65	0.03036016
GO:0018193	peptidyl-amino acid modification	214	0.03042651
GO:0016051	carbohydrate biosynthetic process	85	0.03128398
GO:0072522	purine-containing compound biosynthetic process	68	0.03140016
GO:0010506	regulation of autophagy	22	0.03170967
GO:0042461	photoreceptor cell development	12	0.03198242
GO:0006725	cellular aromatic compound metabolic process	47	0.03222404
GO:0032007	negative regulation of TOR signaling cascade	10	0.03222656
GO:0043467	regulation of generation of precursor metabolites and energy	16	0.03269958
GO:0007423	sensory organ development	145	0.03284376
GO:0051893	regulation of focal adhesion assembly	13	0.03405762
GO:0090109	regulation of cell-substrate junction assembly	13	0.03405762
GO:0000302	response to reactive oxygen species	24	0.03455085
GO:0009152	purine ribonucleotide biosynthetic process	27	0.03461348
GO:0019538	protein metabolic process	1313	0.03529642
GO:0048762	mesenchymal cell differentiation	42	0.03576341
GO:0006091	generation of precursor metabolites and energy	73	0.03611222
GO:0001655	urogenital system development	80	0.0362306
GO:0009058	biosynthetic process	1317	0.03636386
GO:0046128	purine ribonucleoside metabolic process	15	0.03649902
GO:0042698	ovulation cycle	28	0.03676834
GO:0031333	negative regulation of protein complex assembly	18	0.03684235
GO:0044267	cellular protein metabolic process	1177	0.03692873
GO:0000271	polysaccharide biosynthetic process	47	0.03808772

GO:0030155	regulation of cell adhesion	79	0.03823337
GO:0001667	ameboidal cell migration	36	0.03850027
GO:0009310	amine catabolic process	35	0.03907546
GO:0051252	regulation of RNA metabolic process	695	0.0393637
GO:0021766	hippocampus development	17	0.0398407
GO:0051781	positive regulation of cell division	13	0.04016113
GO:0044249	cellular biosynthetic process	1268	0.04061494
GO:0045833	negative regulation of lipid metabolic process	18	0.04071426
GO:0045765	regulation of angiogenesis	39	0.04115667
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	135	0.04146336
GO:0001738	morphogenesis of a polarized epithelium	11	0.04150391
GO:0001755	neural crest cell migration	10	0.04199219
GO:0019318	hexose metabolic process	72	0.04221008
GO:0040013	negative regulation of locomotion	39	0.0424024
GO:0034330	cell junction organization	43	0.0426782
GO:0043484	regulation of RNA splicing	23	0.04291391
GO:0018394	peptidyl-lysine acetylation	45	0.04313977
GO:0009260	ribonucleotide biosynthetic process	30	0.04396928
GO:0014033	neural crest cell differentiation	21	0.04439926
GO:0006022	aminoglycan metabolic process	25	0.04515806
GO:0046530	photoreceptor cell differentiation	14	0.04528809
GO:0010628	positive regulation of gene expression	344	0.04556376
GO:0033555	multicellular organismal response to stress	16	0.04672241
GO:0060113	inner ear receptor cell differentiation	18	0.04936981

Table S25 Enriched GO terms for fast evolving genes in Palaeognathae. The level 3 GO terms were used for analysis.

GO ID	Description	Class	FDR	#genes
GO:0055114	oxidation-reduction process	BP	4.10E-04	36

Table S26 Enriched GO terms for fast evolving genes in Galloanserae. The level 3 GO terms were used for analysis.

GO ID	Description	Class	FDR	#genes
GO:0050789	regulation of biological process	BP	1.11E-02	247
GO:0050794	regulation of cellular process	BP	1.11E-02	235
GO:0023051	regulation of signaling	BP	1.11E-02	77
GO:0051128	regulation of cellular component organization	BP	1.11E-02	49
GO:0007618	mating	BP	1.11E-02	5
GO:0044087	regulation of cellular component biogenesis	BP	1.11E-02	19
GO:0048583	regulation of response to stimulus	BP	1.11E-02	82
GO:0007049	cell cycle	BP	1.11E-02	43
GO:0044085	cellular component biogenesis	BP	1.18E-02	59
GO:0019222	regulation of metabolic process	BP	2.64E-02	142
GO:0032879	regulation of localization	BP	2.64E-02	51
GO:0044238	primary metabolic process	BP	3.87E-02	255
GO:0023056	positive regulation of signaling	BP	5.00E-02	33
GO:0040017	positive regulation of locomotion	BP	5.00E-02	14

Table S27 Enriched GO terms for fast evolving genes in Neoaves. The level 3 GO terms were used for analysis.

GO ID	Description	Class	FDR	#genes
GO:0007017	microtubule-based process	BP	2.68E-05	46
GO:0022402	cell cycle process	BP	1.26E-03	60
GO:0051235	maintenance of location	BP	2.38E-03	19
GO:0007059	chromosome segregation	BP	1.54E-02	13
GO:0055085	transmembrane transport	BP	2.39E-02	63
GO:0051234	establishment of localization	BP	2.40E-02	188

Table S28. Summary of the 227 genes with convergent accelerated evolution in the three vocal-learning bird groups under two evolutionary hypotheses, with their expressions profiles in the various regions and overlapping target species specific amino acid substitution site (TAAS) genes (only legend provided here, full table in a separate tab-delimited text file). In the acceleration columns, *dN/dS*(FG)= *dN/dS* of the foreground branches; *dN/dS*(BG)= *dN/dS* of the background branches; H0= null hypothesis; and, H1= alternative hypothesis. Area X = area X of the striatum; HVC = letter based name; LMAN = lateral magnocellular nucleus of the anterior neostraitum; RA = robust nucleus of the arcopallium; and, + = expression of gene. In # TAAS sites column, numbers are present of counts of TAAS sites in the gene.

Table S29. Gene Ontology (GO) analysis (biological process) of 165 genes in the 6 vocal-learners that are both accelerated under the two evolutionary hypotheses and expressed in the songbird brain. (only legend provided here, full table in a separate tab-delimited text file).

Table S30. Gene Ontology (GO) analysis (biological process) of 151 genes in the 6 vocal-learners that are both accelerated under the two evolutionary hypotheses and expressed in the song nuclei. (only legend provided here, full table in a separate tab-delimited text file).

Table S31. Genes with target species specific amino acid substitution (TAAS) sites for the three vocal learners represented by six species with their acceleration results and expression results from microarray analyses of the songbird brain (only legend provided here, full table in a separate tab-delimited text file). Abbreviations: *dN/dS*(FG)= *dN/dS* of the foreground branches; *dN/dS*(BG)= *dN/dS* of the background branches; H0= null hypothesis; and, H1= alternative hypothesis. Area X = area X of the striatum; HVC = letter based name; LMAN = lateral magnocellular nucleus of the anterior neostraitum; RA = robust nucleus of the arcopallium; + = expression of gene; M = medium ground-finch, Z = zebra finch, C= American crow, B = budgerigar, K = kea; and, H = Anna's humming bird.

Table S32. 6 vocal learning birds and 9 non-vocal learning birds used in identification of accelerated elements.

Vocal learning birds	Songbirds	zebra finch, medium ground-finch, American crow
	Hummingbirds	Anna's hummingbird
	Parrots	budgerigar, kea
Vocal non-learning birds	golden-collared manakin, rifleman, downy woodpecker, chimney swift, peregrine falcon, pigeon, common cuckoo, chuck-wills-widow, chicken	

Table S33. Number of vocal learner specific accelerated elements predicted by phyloP. This analysis was done with two search window sizes (100 and 50 bp). The combined values = 822 genomic elements. The 278 described in Table S34, associated with known genes, is a non-redunant overlap of these two window size scans.

Window size	**Anna's humming bird**	**American crow**	**medium ground-finch**	**budgerigar**	**kea**	**zebra finch**	**ALL**	**ALL (merge adjacent)**
100bp	1168949	412745	445711	706905	623194	496158	580	360
50bp	1476453	617832	659301	982339	889866	692906	550	462

Table S34. Details of accelerated elements, the associated genes, and brain expression in vocal learners (only legend provided here, full table in a separate tab-delimited text file). Area X = area X of the striatum; HVC = letter based name; RA = robust nucleus of the arcopallium; and, + = expression of gene.

Table S35. The accelerated elements associated with differentially expressed genes.

Element size	ZebraFinch geneID	Gene name	Chr	Start	End	Distance to gene	Location
50bp	ENSTGUG00000009016	*TSHZ3*	11	15479003	15479059	8942	3'flanking
	ENSTGUG00000002953	*JAZF1*	2	53157287	53157342	0	intron
100bp	ENSTGUG00000006100	*SH3RF1*	4	32999920	33000012	7464	5'flanking

Table S36. List of genes involved in ossification. The gene list was obtained from GO using terms associated with bones in mammals and birds. The crossed list of genes annotated from the BGI consortium and the lists of genes involved in ossification are the genes presented in this list. For each gene we performed a separate analysis for mammals and birds (Model 0, 1 and 2). The positively selected genes in mammals and birds are highlighted in bold. The calculated *dN/dS* was compared by subtracting the value obtained from mammalian genes to avian genes. Negative values suggest a de-acceleration of the evolutionary rate in birds, while positive values suggests acceleration (when compared with mammals).

	Birds								Mammals								
Genes	Sequences number	Alignment Length (aa)	Model 0	*dN/dS* (ω)	Model 1 (*lnL*)	Model 2 (*lnL*)	2ΔL	p-value	Sequences number	Alignment Length (aa)	Model 0	*dN/dS* (ω)	Model 1 (*lnL*)	Model 2 (*lnL*)	2ΔL	p-value	*dN/dS* difference (birds - mammals)
ACVR2A	***43***	***513***	***-11702.76***	***0.02***	***-11466.7***	***-11445.7***	***42***	***0***	37	581	-9750.79	0.052	-9539.9	-9539.9	0	1	-0.032
ACVR2B	***43***	***516***	***-8605.05***	***0.03***	***-8449.32***	***-8442.83***	***13***	***0.002***	32	493	-9584.95	0.026	-9580.9	-9580.9	0	1	0.004
ADAM8	***32***	***531***	***-18973.37***	***0.17***	***-18440.8***	***-18393***	***95.5***	***0***	***29***	***941***	***-38815.55***	***0.269***	***-37792.2***	***-37771.3***	***41.7***	***0***	-0.099
AHSG	***43***	***384***	***-14302.16***	***0.53***	***-13632.2***	***-13495.5***	***273.4***	***0***	***37***	***662***	***-23765.4***	***0.482***	***-23084.3***	***-23025.7***	***117.1***	***0***	0.048
ANKH	***44***	***509***	***-8167.2***	***0.07***	***-7820.16***	***-7796***	***48.3***	***0***	36	492	-11855.21	0.029	-11809.3	-11809.3	0	1	0.041
ASPN	***40***	***641***	***-14968.79***	***0.16***	***-14453.4***	***-14372.6***	***161.6***	***0***	37	390	-10844.52	0.109	-10722	-10722	0	1	0.051
BMP2	***45***	***367***	***-3609.81***	***0.11***	***-3541.92***	***-3525.43***	***33***	***0***	36	400	-10613.16	0.1	-10489.3	-10489.3	0	1	0.01
BMPR1A	***42***	***541***	***-9405.23***	***0.04***	***-9100.45***	***-9041.77***	***117.4***	***0***	30	544	-10887.29	0.043	-10829.7	-10829.7	0	1	-0.003
CBS	***42***	***94***	***-2649.25***	***0.31***	***-2558.12***	***-2553.57***	***9.1***	***0.011***	30	609	-18230.99	0.099	-17803.8	-17803.3	1.1	0.589	0.211
CD38	***44***	***312***	***-8767.54***	***0.28***	***-8476.3***	***-8462.75***	***27.1***	***0***	***37***	***321***	***-16217.98***	***0.536***	***-15777.4***	***-15752.7***	***49.4***	***0***	-0.256
CER1	***45***	***273***	***-9925.34***	***0.42***	***-9745.14***	***-9731.71***	***26.8***	***0***	33	293	-10914.24	0.339	-10637.8	-10637.8	0	1	0.081
CITED2	***16***	***293***	***-2556.44***	***0.14***	***-2512.43***	***-2501.12***	***22.6***	***0***	21	276	-3334.56	0.079	-3316.2	-3316.2	0	1	0.061
CREB3L1	***43***	***520***	***-10058.35***	***0.2***	***-9621.25***	***-9528.95***	***184.6***	***0***	36	547	-12925.83	0.078	-12874	-12871.2	5.7	0.059	0.122
CTHRC1	***45***	***244***	***-5164.86***	***0.05***	***-5014.78***	***-5004.2***	***21.2***	***0***	35	256	-6754	0.097	-6473.3	-6473.3	0	1	-0.047
DLX5	***16***	***316***	***-5587.19***	***0.3***	***-5095.22***	***-5026.08***	***138.3***	***0***	35	299	-5210.04	0.064	-5192.6	-5192.6	0	1	0.236
DUOX2	***43***	***1623***	***-53940.85***	***0.16***	***-51910.3***	***-51771.4***	***277.8***	***0***	***36***	***1634***	***-54148.58***	***0.159***	***-52320.6***	***-52277***	***87***	***0***	0.001
FBXL15	***44***	***297***	***-7031.28***	***0.08***	***-6874.67***	***-6862.75***	***23.9***	***0***	***29***	***393***	***-9545.87***	***0.131***	***-9243.5***	***-9212.4***	***62.3***	***0***	-0.051
FGF23	***45***	***275***	***-7909.5***	***0.17***	***-7651.95***	***-7623.73***	***56.4***	***0***	32	263	-9396.6	0.131	-9229.6	-9229.6	0	1	0.039
GAS6	***44***	***686***	***-16805.46***	***0.17***	***-16397.9***	***-16382.2***	***31.5***	***0***	***32***	***791***	***-27414.98***	***0.156***	***-26683.2***	***-26668.7***	***28.9***	***0***	0.014
GHR	***42***	***619***	***-15525.26***	***0.25***	***-15089.2***	***-15065.9***	***46.5***	***0***	***37***	***694***	***-21977.59***	***0.297***	***-21422.2***	***-21414.2***	***16***	***0***	-0.047
GPLD1	***44***	***856***	***-26406.69***	***0.3***	***-25547.9***	***-25504.3***	***87.3***	***0***	36	881	-33064.51	0.23	-32263.9	-32263.9	0	1	0.07
HOXA11	***29***	***310***	***-3932.45***	***0.24***	***-3743.63***	***-3725.66***	***35.9***	***0***	33	346	-6697.5	0.147	-6439.6	-6437.6	3.9	0.141	0.093

HOXB4	***41***	***56***	***-657.23***	***0.24***	***-623.393***	***-611.602***	***23.6***	***0***	26	294	-4825.18	0.118	-4752.4	-4752.4	0	1	0.122
HSD17B2	***40***	***388***	***-10641.72***	***0.27***	***-10299.7***	***-10289.5***	***20.4***	***0***	***35***	***403***	***-18921.52***	***0.376***	***-18363.8***	***-18244.8***	***238***	***0***	-0.106
IAPP	***45***	***136***	***-3801.58***	***0.31***	***-3743.83***	***-3738.29***	***11.1***	***0.004***	28	98	-3292.77	0.398	-3282.5	-3282.5	0	1	-0.088
IGF1	***45***	***153***	***-1467.27***	***0.07***	***-1450.58***	***-1445.7***	***9.8***	***0.008***	***28***	***241***	***-4716.72***	***0.197***	***-4514.3***	***-4494.5***	***39.6***	***0***	-0.127
IL6	***42***	***169***	***-5727.51***	***0.31***	***-5508.97***	***-5500.35***	***17.3***	***0***	***35***	***317***	***-12481.92***	***0.701***	***-12348.7***	***-12336.1***	***25.2***	***0***	-0.391
INPP5D	***23***	***1227***	***-18906.06***	***0.13***	***-18395.9***	***-18392.6***	***6.7***	***0.036***	33	1262	-40139.73	0.136	-38975.4	-38975.4	0	1	-0.006
LRRC17	***45***	***442***	***-10115.64***	***0.16***	***-9910.87***	***-9899.53***	***22.7***	***0***	37	458	-14573.29	0.141	-14390.1	-14390.1	0	1	0.019
MEF2A	***44***	***527***	***-11043.54***	***0.09***	***-10783***	***-10779.9***	***6.2***	***0.045***	***38***	***574***	***-15390.23***	***0.126***	***-14691.2***	***-14602.9***	***176.4***	***0***	-0.036
MEF2C	***43***	***478***	***-6133.89***	***0.09***	***-5998***	***-5969.52***	***57***	***0***	***34***	***494***	***-8439.71***	***0.139***	***-7873.3***	***-7594.3***	***558***	***0***	-0.049
MEPE	***35***	***82***	***-2877.16***	***0.4***	***-2730.58***	***-2697.83***	***65.5***	***0***	***35***	***637***	***-27855.67***	***0.492***	***-27564.4***	***-27559.2***	***10.3***	***0.006***	-0.092
MGP	***44***	***104***	***-3947***	***0.35***	***-3809.69***	***-3799.87***	***19.7***	***0***	35	129	-3828.38	0.202	-3732.7	-3732.7	0	1	0.148
NBR1	***41***	***1081***	***-37167.42***	***0.36***	***-35468.7***	***-34913.6***	***1110.3***	***0***	***32***	***1118***	***-26784.06***	***0.257***	***-26018***	***-25964***	***108.1***	***0***	0.103
NCDN	***44***	***772***	***-27557.63***	***0.24***	***-26571.6***	***-26537.3***	***68.7***	***0***	28	764	-15754.48	0.048	-15686	-15686	0	1	0.192
NOX4	***41***	***597***	***-11228.1***	***0.2***	***-11021.2***	***-11000.4***	***41.7***	***0***	35	661	-15140.69	0.17	-14798.4	-14796.4	4	0.135	0.03
OSR2	***27***	***316***	***-5164.5***	***0.09***	***-4936.43***	***-4925.65***	***21.5***	***0***	***32***	***312***	***-5900.05***	***0.077***	***-5569.5***	***-5561.1***	***16.9***	***0***	0.013
P2RX7	***11***	***598***	***-9939.44***	***0.71***	***-9681.71***	***-9619.37***	***124.7***	***0***	34	613	-19177.78	0.206	-18579.7	-18579.7	0	1	0.504
PKDCC	***40***	***296***	***-7580.94***	***0.19***	***-7116.84***	***-7060.6***	***112.5***	***0***	34	504	-9721.1	0.084	-9550.9	-9550.9	0	1	0.106
PLXNB1	***42***	***2250***	***-49382.17***	***0.13***	***-47722.7***	***-47666.9***	***111.6***	***0***	34	2247	-64864.6	0.143	-63307.3	-63307.3	0	1	-0.013
PTGER4	***43***	***478***	***-10433.82***	***0.09***	***-10051***	***-10047.3***	***7.4***	***0.024***	***33***	***546***	***-14372.1***	***0.098***	***-13939.1***	***-13923.3***	***31.5***	***0***	-0.008
SRD5A1	***42***	***177***	***-4345.07***	***0.23***	***-4180.84***	***-4170.7***	***20.3***	***0***	***33***	***266***	***-11004.89***	***0.284***	***-10623.1***	***-10619.5***	***7***	***0.029***	-0.054
SULF1	***40***	***892***	***-17554.42***	***0.08***	***-16891.3***	***-16869.8***	***43***	***0***	37	1169	-27848.1	0.092	-27286.5	-27284.36	4.3	0.116	-0.012
SYK	***36***	***655***	***-8616.47***	***0.1***	***-8411.41***	***-8400.41***	***22***	***0***	36	665	-19535.2	0.069	-19002.95	-19002.95	0	1	0.031
TCF7L2	***45***	***511***	***-5655***	***0.14***	***-5547.61***	***-5530.51***	***34.2***	***0***	37	697	-13560.88	0.133	-13154.84	-13154.84	0	1	0.007
TFRC	***45***	***791***	***-32110.43***	***0.48***	***-30514.3***	***-30267.6***	***493.3***	***0***	***37***	***807***	***-32317.76***	***0.332***	***-31012.26***	***-30917.3***	***189.9***	***0***	0.148
TNFAIP3	***44***	***833***	***-23443.12***	***0.19***	***-22689.9***	***-22643.8***	***92.3***	***0***	38	809	-27479.46	0.11	-26943.04	-26943.04	0	1	0.08
TPP1	***21***	***480***	***-17754.24***	***0.41***	***-15933.1***	***-15699.6***	***467***	***0***	***33***	***589***	***-15351.68***	***0.199***	***-14972.8***	***-14969.07***	***7.5***	***0.024***	0.211
VEGFA	***44***	***239***	***-1916.55***	***0.21***	***-1874.23***	***-1846.21***	***56***	***0***	32	484	-9874.3	0.34	-9745.29	-9745.29	0	1	-0.13
AQP1	44	271	-5813.3	0.06	-5703.4	-5703.4	0	1	***28***	***304***	***-6611.57***	***0.057***	***-6459.7***	***-6454.4***	***10.6***	***0.005***	0.003
BCOR	*29*	*1798*	*-32581.66*	*0.13*	*-31944*	*-31941.2*	*5.6*	*0.061*	*37*	*1859*	*-54967.63*	*0.117*	*-54297.5*	*-54297.5*	*0*	*1*	0.013
BMP7	44	425	-6531.04	0.03	-6482.96	-6482.96	0	1	35	438	-9473.43	0.027	-9451.5	-9451.5	0	1	0.003

CA2	45	271	-6611.39	0.13	-6420.57	-6420.57	0	1	***36***	***274***	***-10580.57***	***0.257***	***-10246.2***	***-10221.2***	***50***	***0***	-0.127
CARM1	19	412	-6668.17	0.23	-6552.55	-6552.55	0	1	***13***	***658***	***-7523.57***	***0.089***	***-7421.5***	***-7413.7***	***15.6***	***0***	0.141
CDX1	30	255	-4528.4	0.06	-4515.56	-4515.56	0	1	24	296	-6379.04	0.141	-6328.5	-6328.5	0	1	-0.081
COL2A1	19	268	-5208.96	0.04	-5076.04	-5076.04	0	1	***34***	***1521***	***-34176.27***	***0.115***	***-33359.5***	***-33184.1***	***350.8***	***0***	-0.075
CTSK	19	132	-3109.43	0.08	-2973.96	-2973.96	0	1	37	353	-9294.53	0.125	-9071	-9070.6	0.7	0.7	-0.045
DYM	24	680	-12203.96	0.05	-12147.7	-12147.7	0	1	36	686	-16342.33	0.08	-16209.4	-16209.4	0	1	-0.03
EIF2AK3	44	1012	-23818.04	0.1	-23421.4	-23421.4	0	1	39	1150	-36979.36	0.132	-36048.8	-36048.4	0.7	0.722	-0.032
FGF8	36	229	-3605.86	0.06	-3523.46	-3520.74	5.5	0.065	27	211	-4540.68	0.091	-4480.6	-4480.6	0	1	-0.031
GPM6B	45	328	-4105.11	0.03	-4067.28	-4067.28	0	1	***35***	***378***	***-8218.49***	***0.106***	***-7918.6***	***-7913.5***	***10.1***	***0.006***	-0.076
GREM1	44	186	-3942.12	0.06	-3838.69	-3836.11	5.2	0.076	***32***	***240***	***-3819.46***	***0.025***	***-3769.2***	***-3764.7***	***9.1***	***0.01***	0.035
HOXD11	44	284	-3775.63	0.12	-3730.09	-3730.09	0	1	18	367	-7337.44	0.182	-7153	-7153	0	1	-0.062
IFITM5	16	151	-3307.72	0.13	-3246.79	-3246.79	0	1	24	135	-3392.62	0.102	-3378.3	-3378.3	0	1	0.028
IHH	22	312	-6585.46	0.11	-6476.32	-6476.32	0	1	33	461	-10287.35	0.068	-10149.6	-10149.6	0	1	0.042
IL7	44	145	-3418.75	0.35	-3349.29	-3347.12	4.3	0.114	***14***	***185***	***-3236.87***	***0.607***	***-3157***	***-3143.3***	***27.4***	***0***	-0.257
KLF10	42	485	-18012.24	0.23	-17407.5	-17405.7	3.8	0.151	***32***	***512***	***-14207.6***	***0.154***	***-13961.3***	***-13957.8***	***6.9***	***0.031***	0.076
LRP6	45	1618	-30462.53	0.03	-30335.2	-30335.2	0	1	38	1639	-33752.84	0.043	-33466.3	-33466.3	0	1	-0.013
MC4R	45	331	-6203.29	0.06	-6084.67	-6083.7	1.9	0.379	32	341	-8135.01	0.053	-8046.3	-8046.3	0	1	0.007
MITF	45	468	-8087.2	0.03	-7997.14	-7997.14	0	1	36	533	-11836.16	0.082	-11631.7	-11631.7	0	1	-0.052
MMP2	43	672	-13067.17	0.05	-12839.4	-12839.4	0	1	36	687	-19106.74	0.074	-18818.5	-18818.2	0.5	0.761	-0.024
MSX1	31	294	-4678.14	0.03	-4552.59	-4552.59	0	1	31	320	-6840.9	0.071	-6784	-6784	0	1	-0.041
NF1	41	2834	-53801.41	0.02	-53482.5	-53481.5	2.2	0.337	36	2854	-52208	0.037	-52042.5	-52042.5	0	1	-0.017
PAPSS2	42	626	-17640.91	0.09	-17302.6	-17302.6	0	0.986	***39***	***653***	***-20583.03***	***0.098***	***-19971.2***	***-19930.5***	***81.4***	***0***	-0.008
PLA2G4A	37	772	-15650.53	0.06	-15447.4	-15447.4	0	1	38	764	-19685.71	0.071	-19416.7	-19416.7	0	1	-0.011
PTH	45	119	-3004.34	0.18	-2976.46	-2976.46	0	1	34	126	-4372.73	0.324	-4269.7	-4269.7	0	1	-0.144
PTK2B	14	1127	-16055.71	0.1	-15649.8	-15649.8	0	1	37	1036	-26169.58	0.057	-25961.6	-25961.6	0	1	0.043
PTN	43	165	-3060.44	0.1	-2982.97	-2982.42	1.1	0.577	***35***	***257***	***-6069.17***	***0.173***	***-5779.7***	***-5748.1***	***63.2***	***0***	-0.073
SBDS	45	172	-4495	0.04	-4468.29	-4468.29	0	1	21	255	-4311.49	0.042	-4279	-4279	0	1	-0.002
SFRP1	9	314	-2190.51	0.02	-2183.89	-2182.22	3.3	0.19	31	329	-5941.62	0.045	-5873.1	-5873.1	0	1	-0.025
SFRP2	22	313	-4193.65	0.08	-4046.79	-4046.79	0	1	18	304	-4128.58	0.037	-4070.9	-4070.9	0	1	0.043
SH3PXD2B	41	927	-24447.81	0.15	-23678.9	-23678.9	0	1	29	939	-25719.88	0.127	-25059.3	-25059.3	0	1	0.023

SPP2	44	193	-7264.52	0.31	-7089.98	-7089.98	0	1	33	224	-10155.77	0.36	-9993.9	-9993.9	0	1	-0.05
SRGN	44	148	-3843.52	0.2	-3731.79	-3730.79	2	0.37	33	212	-7677.94	0.411	-7571.1	-7571.1	0	1	-0.211
SULF2	45	906	-19135.59	0.06	-18846	-18844.6	2.8	0.251	31	929	-21842.44	0.052	-21433.46	-21433.46	0	1	0.008
TGFB3	45	296	-5431.84	0.02	-5395.57	-5394.42	2.3	0.316	35	490	-8801.84	0.058	-8762.83	-8762.83	0	1	-0.038
TPH1	43	444	-8989.13	0.09	-8898.23	-8898.23	0	1	***30***	***535***	***-10926.03***	***0.128***	***-10654.62***	***-10606.05***	***97.1***	***0***	-0.038
TRAF6	44	545	-14122.88	0.07	-13859	-13859	0	1	***36***	***559***	***-16310.82***	***0.128***	***-15984.32***	***-15981.19***	***6.3***	***0.043***	-0.058
TUFT1	38	394	-4587.75	0.05	-4562.3	-4562.3	0	1	35	422	-12809.61	0.213	-12493.67	-12493.67	0	1	-0.163

Table S37. Gene Set Enrichment Analysis. Comparison of the functional annotation (accessed in DAVID (*153*)) in positively and negatively selected genes in birds.

Positively Selected Genes (49)					Negatively Selected Genes (40)				
Term	Count	%	P-Value	Benjamini	Term	Count	%	P-Value	Benjamini
GO:0030278~regulation of ossification	13	2.43	1.00E-17	1.20E-14	GO:0001501~skeletal system development	18	2.44	7.89E-19	8.25E-16
GO:0030500~regulation of bone mineralization	10	1.87	2.30E-16	1.33E-13	GO:0060348~bone development	14	1.89	2.90E-18	1.51E-15
GO:0070167~regulation of biomineral formation	10	1.87	4.49E-16	1.78E-13	GO:0001503~ossification	13	1.76	8.23E-17	3.86E-14
GO:0001501~skeletal system development	15	2.81	7.52E-13	2.25E-10	GO:0046849~bone remodeling	6	0.81	1.97E-09	5.15E-07
GO:0045124~regulation of bone resorption	6	1.12	8.44E-10	2.02E-07	GO:0005615~extracellular space	14	1.89	2.26E-09	2.08E-07
GO:0046850~regulation of bone remodeling	6	1.12	8.44E-10	2.02E-07	GO:0031214~biomineral formation	6	0.81	4.10E-08	8.56E-06
GO:0034103~regulation of tissue remodeling	6	1.12	3.24E-09	6.47E-07	GO:0042981~regulation of apoptosis	14	1.89	6.75E-08	1.18E-05
GO:0045667~regulation of osteoblast differentiation	7	1.31	4.63E-09	7.93E-07	GO:0030282~bone mineralization	5	0.68	7.51E-08	1.12E-05
GO:0046851~negative regulation of bone remodeling	5	0.94	6.75E-09	1.01E-06	GO:0043067~regulation of programmed cell death	14	1.89	7.59E-08	9.91E-06
GO:0045779~negative regulation of bone resorption	5	0.94	6.75E-09	1.01E-06	GO:0010941~regulation of cell death	14	1.89	7.92E-08	9.20E-06

Table S38. NCBI identification numbers for α-keratin genes used as queries for genome searches.

Anolis carolinensis Type I	*Anolis carolinensis* Type II	*Gallus gallus* Type I	*Gallus gallus* Type II
GI: 327279477	GI: 327275737	cdsid_NP_001001311.2	GI: 45384377
GI: 327264559	GI: 327275733	cdsid_NP_990340.1	GI: 47604941
GI: 327264555	GI: 327275729	cdsid_NP_001001312.1	GI: 363745007
GI: 327264551	GI: 327275725	cdsid_XP_003642876.1	cdsid_NP_001001313.1
GI: 327264547	GI: 327275721	cdsid_XP_418162.3	cdsid_NP_001001314.1
GI: 327264543	GI: 327279491	cdsid_XP_001235182.2	cdsid_NP_001001195.1
GI: 327264539	GI: 327275735	cdsid_XP_003642867.1	cdsid_XP_003643660.1
GI: 327264444	GI: 327275731	cdsid_XP_418168.2	cdsid_XP_424505.3
GI: 327264432	GI: 327275727	cdsid_XP_001233972.1	cdsid_XP_428852.3
GI: 327264561	GI: 327275723	cdsid_XP_003642868.1	cdsid_XP_428851.3
GI: 327264553	GI: 327275719	cdsid_XP_418163.3	cdsid_XP_001232027.2
GI: 327264549	GI: 327275507	cdsid_XP_418167.3	cdsid_XP_001232152.1
GI: 327264545	GI: 327275503	cdsid_XP_425874.2	cdsid_XP_428853.2
GI: 327264541	GI: 327275499		cdsid_XP_001232221.2
GI: 327264537	GI: 327260349		cdsid_XP_003643171.1
GI: 327264442	GI: 327275505		cdsid_XP_428850.2
GI: 327264438	GI: 327275494		cdsid_XP_001231586.2
GI: 327264434	GI: 327275496		
GI: 327264430	GI: 327275501		
GI: 327260347	GI: 208972040		
GI: 208972044			

Table S39. Copy numbers for Type I and II α-keratins for each species of mammal, reptile and bird used in this study. Copy numbers for mammalian α-keratins were obtained from (*160*).

Common name	Type I α-keratins	Type II α-keratins	Total α-keratins
Mammals			
human	28	26	54
opossum	35	27	62
house mouse	27	26	53
platypus	27	20	47
Reptiles			
green anole lizard	20	21	41
green sea turtle	17	14	31
American alligator	23	21	44
saltwater crocodile	20	20	40
Birds			
Adelie penguin	15	18	33
American crow	14	17	31
red-crested turaco	14	18	32
Anna's hummingbird	14	16	30
bald eagle	13	17	30
barn owl	12	18	30
bar-tailed trogon	12	18	30
brown mesite	15	22	37
budgerigar	15	11	26

American flamingo	13	17	30
chicken	14	13	27
chimney swift	16	15	31
chuck-will's-widow	13	21	34
common cuckoo	12	16	28
crested ibis	17	17	34
grey-crowned crane	14	18	32
cuckoo roller	12	19	31
dalmatian pelican	15	18	33
pigeon	13	15	28
downy woodpecker	11	18	29
emperor penguin	14	18	32
golden-collared manakin	15	11	26
great cormorant	13	19	32
great-crested grebe	17	19	36
hoatzin	15	16	31
MacQueen's bustard	15	19	34
rhinoceros hornbill	14	18	32
kea	17	16	33
killdeer	16	17	33
little egret	18	18	36
medium ground-finch	14	15	29
carmine bee-eater	10	17	27
northern fulmar	14	19	33

common ostrich	17	14	31
Peking duck	14	16	30
peregrine falcon	18	17	35
red-legged seriema	16	15	31
red-throated loon	15	20	35
rifleman	15	14	29
speckled mousebird	15	21	36
sunbittern	13	16	29
turkey vulture	16	22	38
white-throated tinamou	15	17	32
white-tailed eagle	14	17	31
white-tailed tropicbird	14	17	31
turkey	13	14	27
yellow-throated sandgrouse	14	19	33
zebra finch	18	13	31

Table S40. Swiss-Prot ID numbers, gene and species names and protein length for β-keratin genes used as queries for avian genome searches.

Swiss-prot ID	Protein name	Organism	Length
P25692	Claw keratin	*Gallus gallus*	128
Q9PRI5	Feather keratin Cos1-1/Cos1-3/Cos2-1	*Columba livia*	101
O93499	Feather keratin Cos1-2	*Columba livia*	101
O93500	Feather keratin Cos2-2	*Columba livia*	101
Q9PSV3	Feather keratin Cos2-3	*Columba livia*	101

P20308	Feather keratin 4	*Gallus gallus*	98
Q98U06	Feather beta keratin	*Cathartes aura*	98
P04458	Feather keratin 2	*Gallus gallus*	98
Q98U05	Feather beta keratin	*Mycteria Americana*	98
P02450	Feather keratin 1	*Gallus gallus*	98
P20307	Feather keratin 3	*Gallus gallus*	98
O13152	Beta-keratin-related protein	*Gallus gallus*	109
Q92012	Beta-keratin-related protein	*Coturnix coturnix japonica*	109
P08335	Feather keratin B-4	*Anas platyrhynchos*	95
P07521	Feather keratin B-4	*Columba livia*	95
P02449	Feather keratin	*Dromaius novaehollandiae*	102
P02451	Feather keratin	*Larus novaehollandiae*	98
P04459	Scale keratin	*Gallus gallus*	155

Table S41. The total number of β-keratins for each β-keratin subfamily. Column C shows complete sequences without any obvious errors and column I shows incomplete sequences that have sequencing errors, inframe stop codons, frame shift mutations or missing data due to short reads or unknown sequence.

Common Name	**Total**		**Claw**		**Feather**		**Keratinocyte**		**Scale**	
	C	**I**	**C**	**I**	**C**	**I**	**C**	**I**	**C**	**I**
Reptiles										
green anole lizard	40	0	0	0	0	0	0	0	0	0
green sea turtle	26	0	15	0	0	0	11	0	0	0
American alligator	20	0	6	0	0	0	8	0	6	0

Birds										
rifleman	18	19	4	5	7	4	5	6	2	4
Peking duck	46	15	3	3	26	8	14	0	3	4
bar-tailed trogon	39	33	4	0	27	25	7	2	1	6
emperor penguin	38	8	8	1	12	4	14	0	4	3
grey-crowned crane	18	10	3	1	4	6	7	2	4	1
rhinoceros hornbill	16	20	3	1	9	8	3	2	1	9
Anna’s hummingbird	40	19	6	1	19	8	10	7	5	3
chuck-will's-widow	21	11	6	1	7	6	6	1	2	3
red-legged seriema	21	13	3	2	9	7	7	2	2	2
turkey vulture	17	16	4	1	7	9	5	2	1	4
chimney swift	33	17	7	1	16	5	6	8	4	3
killdeer	58	21	7	3	30	13	14	2	7	3
MacQueen’s bustard	20	18	2	2	8	10	8	1	3	5
pigeon	81	18	6	2	62	14	6	1	7	1
speckled mousebird	17	16	5	2	3	8	7	3	2	3
American crow	24	32	2	2	15	22	4	5	4	3
common cuckoo	47	39	7	6	27	22	7	3	6	8
little egret	27	24	3	2	12	12	5	6	7	4
sunbittern	19	12	4	1	8	4	6	4	1	3
peregrine falcon	35	12	5	0	11	7	10	4	6	1
northern gulmar	16	17	4	2	3	9	7	3	2	3
chicken	133	38	13	4	99	31	11	0	10	3
red-throated loon	15	32	3	2	7	22	4	4	1	4

medium ground-finch	43	38	1	5	35	22	4	9	3	2
white-tailed eagle	19	16	5	1	3	13	8	0	3	2
bald eagle	15	81	5	8	4	61	5	4	1	8
cuckoo roller	23	21	4	1	13	13	4	2	2	5
golden-collared manakin	31	24	2	4	25	12	2	2	2	6
turkey	46	42	7	3	29	28	7	2	3	9
budgerigar	71	38	13	4	38	24	9	4	11	6
carmine bee-eater	18	15	3	1	8	8	5	2	2	4
brown mesite	37	20	6	0	15	15	9	1	7	4
kea	33	17	8	0	11	11	11	2	3	3
crested ibis	29	32	3	6	15	14	7	7	5	5
hoatzin	34	33	8	2	16	22	5	5	5	4
dalmatian pelican	19	15	3	1	7	5	7	2	2	7
great cormorant	11	18	1	2	5	9	4	2	1	5
white-tailed tropicbird	15	25	2	3	6	14	5	3	2	5
American flamingo	19	15	5	0	8	6	5	5	1	4
downy woodpecker	39	22	4	4	30	9	0	5	5	4
great-crested grebe	16	11	4	1	5	5	6	2	1	3
yellow-throated sandgrouse	29	19	5	1	16	15	5	0	3	3
Adelie penguin	33	20	4	7	10	8	12	4	7	1
common ostrich	23	10	4	1	7	4	4	2	8	3
zebra finch	149	17	5	5	127	11	10	0	7	1
red-crested turaco	39	40	5	3	18	28	11	4	5	5

white-throated tinamou	27	16	4	2	13	8	5	4	5	2
barn owl	6	19	2	1	1	8	3	4	0	6
BIRD TOTALS	**1623**	**1084**								

28

29 **Table S42. Scafolds containing the complete avian *AGT* genes.**

Bird name	**Genome Location**
bar-tailed trogon	scaffold23963
emperor penguin	Scaffold647
rhinoceros hornbill	scaffold44635
Anna's hummingbird	scaffold98
chimney swift	scaffold264
killdeer	scaffold559
pigeon	scaffold80
American crow	scaffold223
little egret	scaffold448
peregrine falcon	scaffold303_1
chicken	Chr9
medium ground-finch	scaffold118
bald eagle	Scaffold2418
cuckoo roller	scaffold37318
turkey	GL428176.1
budgerigar	Adam_Phillippy_v6_sli_scf900160277061
brown mesite	scaffold31209

crested ibis	Scaffold53
dalmatian pelican	scaffold43973
yellow-throated sandgrouse	scaffold4397
Adelie penguin	Scaffold544
zebra finch	Chr9

Table S43. Genomic locations of the complete bird *GULO* genes.* Retrieved from GenBank and verified in our genomic mining: chicken, XM_003204567 and turkey, XM_003204567. ** Green anole lizard(*Anolis carolinensis*)

Species	**Genome location**
rifleman	scaffold19528
Peking duck	scaffold3910
Carolina Anole	GL343529.1, GL344527.1
bar-tailed trogon	scaffold5637
emperor penguin	scaffold192
grey-crowned crane	scaffold33311
rhinoceros hornbill	scaffold6969
chuck-wills-widow	scaffold45258, saffold16059
red-legged seriema	scaffold25271, scaffold8563
turkey vulture	scaffold8861, scaffold25477, C13110740, scaffold47142
chimney swift	scaffold107
killdeer	scaffold41
MacQueen's bustard	scaffold41185
pigeon	scaffold661
speckled mousebird	scaffold2429

American crow	scaffold948, scaffold3805
common cuckoo	scaffold436
little egret	scaffold218
sunbittern	scaffold8079, scaffold50246
peregrine falcon	scaffold312_1
northern fulmar	scaffold2872
chicken**	Chr3
red-throated loon	scaffold40674, scaffold28774
bald eagle	scaffold612
cuckoo roller	scaffold36217, scaffold40357
turkey**	Chr2
budgerigar	scf900160277026
brown mesite	scaffold14945
kea	scaffold15854
crested ibis	scaffold155
hoatzin	scaffold1147
dalmatian pelican	scaffold20629
American flamingo	scaffold24923, scaffold23433, scaffold52878
downy woodpecker	scaffold389
great-crested grebe	scaffold13824
Adelie penguin	scaffold666
common ostrich	scaffold372
red-crested turaco	scaffold33784
white-throated tinamou	scaffold8823, C15630071, C16205607, C16618990

green anole lizard**	GL343529.1, GL344527.1

35
36

Table S44. Significant matches of rhodopsin and conopsin representatives identified using tBLASTn searches. Green circles, complete sequences; Yellow circles, partial sequences; Shaded rows, species that have all four conopsin genes necessary for tetrachromatic vision. High coverage genomes (≥60X) are highlighted in bold.

Species		Rh1	Rh2	OPN1sw1	OPN1sw2	OPN1lw
Rifleman	Acanthisitta chloris	●	●	●	●	●
Peking duck	Anas platyrhynchos domestica	●	●	●		●
Chuck will's widow (Nightjar)	Antrostomus carolinensis	●	●	●		●
Bar-tailed trogon	Apaloderma vittatum	●	●	●		
Emperor penguin	Aptenodytes forsteri	●	●	●		
Crowned crane	Balearica r. gibbericeps	●	●	●	●	
Javan rhinoceros hornbill	Buceros r. silvestris	●	●		.	
Anna's hummingbird	Calypte anna	●	●		●	●
Red-legged seriema	Cariama cristata	●	●	●		
Turkey vulture	Cathartes aura	●	●	●		
Chimney swift	Chaetura pelagica	●	●		●	●
Killdeer	Charadrius vociferus	●	●	●		
Houbara Bustard	Chlamydotis undulata	●	●	●	●	
Speckled mousebird	Colius striatus	●	●	●	●	●
domestic pigeon	Columba livia	●	●	●	●	●
American Crow	Corvus brachyrhynchos	●	●	●	●	●
Common Cuckoo	Cuculus canorus	●	●	●	●	●
Little egret	Egretta garzetta	●	●	●		●
Sunbittern	Eurypyga helias	●	●	●	●	
Peregrine falcon	Falco peregrinus	●	●	●	●	●
Northern Fulmar	Fulmarus glacialis	●	●	●	●	
Chicken	Gallus gallus	●	●	●	●	●
Red throated loon	Gavia stellata	●	●			
Medium ground finch	Geospiza fortis	●	●	●	●	●
white-tail eagle	Haliaeetus albicilla	●	●			
Bald Eagle	Haliaeetus leucocephalus	●	●	●	●	●
Cuckoo roller	Leptosomus discolor	●	●	●	●	●
Golden-collared Manakin	Manacus vitellinus	●	●	●		●
Turkey	Meleagris gallopavo	●	●			
Budgerigar	Melopsittacus undulatus	●	●			
Northern Carmine bee-eater	Merops nubicus	●	●			
Brown mesite	Mesitornis unicolor	●	●	●	●	●
Kea	Nestor notabilis	●	●	●	●	●
Crested Ibis	Nipponia nippon	●	●	●	●	●
Hoatzin	Ophisthocomus hoazin	●	●	●		
Dalmatian pelican	Pelecanus crispus	●	●			
White-tailed tropicbird	Phaethon lepturus	●	●	●		
Great black cormorant	Phalacrocorax carbo	●	●			
Caribbean flamingo	Phoenicopterus ruber	●	●	●		
Downy Woodpecker	Picoides pubescens	●	●	●	●	●
Great-crested grebe	Podiceps cristatus	●	●	●		
Yellow-thoated Sandgrouse	Pterocles guturalis	●	●	●	●	●
Adelie penguin	Pygoscelis adeliae	●	●			●
Ostrich	Struthio camelus	●	●		●	
Zebra finch	Taeniopygia guttata	●	●	●	●	●
Angola turaco	Tauraco erythrolophus	●	●			
Great tinamou	Tinamus guttatus	●	●	●	●	●
Barn owl	Tyto alba	●	●			

Table S45. NCBI accession numbers for avian visual opsin genes used for sequence alignments.

Species	***RH1***	***RH2***	***OPN1sw1***	***OPN1sw2***	***OPN1lw***
pigeon				XM_005514005.1	AAD38036
chicken	NM_001030606.1	M92038	M92037	M92037	NM_205440.1
great cormorant			EF568933.1		
turkey	XM_003210211.1	XM_003212975.1			
zebra finch	AF222329	AF222330	AF222331	AF222332	NM_001076702.1

Table S46. Spermatogenesis-related genes used for *dN/dS* analysis.

Gene name	Chicken gene ID	Gene name	Chicken gene ID
MKKS	ENSGALG00000009013	*STRA8*	ENSGALG00000011722
UBR2	ENSGALG00000009906	*ADAD1*	ENSGALG00000011862
GSR	ENSGALG00000010271	*MEI1*	ENSGALG00000011919
TDRD9	ENSGALG00000011565	*MLH1*	ENSGALG00000012060
JAG2	ENSGALG00000011696	*TXNDC3*	ENSGALG00000012078
SPATA5	ENSGALG00000011833	*STYX*	ENSGALG00000012411
SREBF2	ENSGALG00000011916	*GPR64*	ENSGALG00000016511
FSHB	ENSGALG00000012140	*CEP57*	ENSGALG00000017201
HERPUD2	ENSGALG00000012149	*BBS4*	ENSGALG00000001798
SYCP3	ENSGALG00000012766	*SOD1*	ENSGALG00000015844
MAK	ENSGALG00000012770	*SPATA6*	ENSGALG00000010502
CDYL	ENSGALG00000012808	*SSTR2*	ENSGALG00000004418

TLK2	ENSGALG00000000410	*MOV10L1*	ENSGALG00000008570
TBPL1	ENSGALG00000013981	*ADRM1*	ENSGALG00000005200
SOX30	ENSGALG00000003723	*RBP4*	ENSGALG00000006629
SLC22A16	ENSGALG00000015055	*MICALCL*	ENSGALG00000005523
DAZAP1	ENSGALG00000015200	*GOLGA3*	ENSGALG00000002158
RNF151	ENSGALG00000015957	*FKBP6*	ENSGALG00000000837
ALMS1	ENSGALG00000016039	*JAM3*	ENSGALG00000001472
HSF2BP	ENSGALG00000016202	*NDRG3*	ENSGALG00000001492
PCYT1B	ENSGALG00000016318	*INPP5B*	ENSGALG00000001606
APOB	ENSGALG00000016491	*PDILT*	ENSGALG00000002006
TDRD6	ENSGALG00000016712	*ACOX1*	ENSGALG00000002159
FNDC3A	ENSGALG00000017002	*SYCP1*	ENSGALG00000002511
CCNA1	ENSGALG00000017052	*NKD1*	ENSGALG00000003767
RNF6	ENSGALG00000017105	*IFT81*	ENSGALG00000003854
MAEL	ENSGALG00000019211	*SIAH1*	ENSGALG00000003916
SPATA7	ENSGALG00000010607	*PIWIL1*	ENSGALG00000002645
KDM2B	ENSGALG00000004225	*ODF2*	ENSGALG00000004767
TBP	ENSGALG00000011171	*RAD51C*	ENSGALG00000005055
FAS	ENSGALG00000006351	*TUBD1*	ENSGALG00000005173
MYCBP	ENSGALG00000024295	*PAFAH1B1*	ENSGALG00000005834
GFER	ENSGALG00000005579	*ADAMTS2*	ENSGALG00000006000
PATZ1	ENSGALG00000006934	*RACGAP1*	ENSGALG00000006271
CIT	ENSGALG00000007354	*SPATA20*	ENSGALG00000006857
PEBP1	ENSGALG00000007403	*AFF4*	ENSGALG00000007126
MYCBPAP	ENSGALG00000007661	*SPAG9*	ENSGALG00000007352

SPAG6	ENSGALG00000007892	*SPO11*	ENSGALG00000007662
BOLL	ENSGALG00000008116	*PROK2*	ENSGALG00000007785
NPHP1	ENSGALG00000008195	*CELF1*	ENSGALG00000008097
CLDN11	ENSGALG00000009355	*PSME4*	ENSGALG00000008163
PLEKHA1	ENSGALG00000009534	*NEURL*	ENSGALG00000008281
MAST2	ENSGALG00000010313	*NR2C2*	ENSGALG00000008519
CHN2	ENSGALG00000011164	*TYRO3*	ENSGALG00000008631
DAZL	ENSGALG00000011243		

Table S47. Oogenesis-related genes used for *dN/dS* analysis.

Gene name	Chicken gene ID
HEXB	ENSGALG00000014933
GDF9	ENSGALG00000007091
RPS6KA2	ENSGALG00000011473
STRA8	ENSGALG00000011722
MLH1	ENSGALG00000012060
ADRM1	ENSGALG00000005200

Table S48. Comparison of *dN/dS* ratios of spermatogenesis genes to that of the genome background (Macrochromosome genes). "*" denotes significant differences (Wilcoxon rank sum test, $p<0.05$).

Species	Macrochromosome gene median *dN/dS* ratio	Spermatogenesis gene median *dN/dS* ratio	One-tailed Wilcoxon rank sum test p-value (Macro < Spermatogenesis)
downy woodpecker	0.1217	0.13985	0.006858353*

carmine bee-eater	0.1248	0.1599	0.04713542*
rhinoceros hornbill	0.1389	0.1505	0.317827
bar-tailed trogon	0.1277	0.1985	0.0087315*
cuckoo roller	0.1363	0.158	0.07364234
speckled mousebird	0.1301	0.1805	0.01257755*
barn owl	0.1535	0.1959	0.0041785*
turkey vulture	0.1858	0.2253	0.003208803*
red-legged seriema	0.15375	0.1929	0.1590293
peregrine falcon	0.1462	0.2194	0.01211696*
kea	0.1495	0.1778	0.1140301
budgerigar	0.1515	0.1835	0.03172751*
rifleman	0.14	0.1877	0.005718737*
golden-collared manakin	0.15085	0.1573	0.1327841
American crow	0.1642	0.17475	0.2902343
zebra finch	0.1834	0.2727	0.000405036*
medium ground-finch	0.1386	0.1725	0.07511656
sunbittern	0.1268	0.1654	0.02533376*
white-tailed tropicbird	0.1418	0.1703	0.1416681
red-throated loon	0.1527	0.1957	0.02357394*
emperor penguin	0.2251	0.2224	0.465767
Adelie penguin	0.2036	0.2199	0.4379779
northern fulmar	0.1697	0.21075	0.08830071
great cormorant	0.15185	0.15525	0.08720042
crested ibis	0.1527	0.1775	0.2049325
little egret	0.13815	0.1578	0.1382569

dalmatian pelican	0.1699	0.18105	0.02524784*
grey crowned crane	0.1383	0.1606	0.04436508*
killdeer	0.1296	0.16415	0.05403358
hoatzin	0.15225	0.1936	0.01065571*
red-crested turaco	0.1395	0.1592	0.1747277
MacQueen's bustard	0.1451	0.1764	0.1868291
common cuckoo	0.1432	0.1825	0.003375151*
chuck-wills-widow	0.1472	0.1734	0.1414638
chimney Swift	0.1408	0.1815	0.004769873*
Anna's Hummingbird	0.14085	0.1762	0.1265243
great crested grebe	0.1434	0.1676	0.1765632
American flamingo	0.1548	0.1689	0.09382822
pigeon	0.1418	0.1611	0.1105944
brown mesite	0.1375	0.1845	0.04062427*
yellow-throated sandgrouse	0.1647	0.21525	0.06766375
Peking duck	0.1164	0.1597	0.008793625*
turkey	0.1413	0.1787	0.06081294
chicken	0.1416	0.16035	0.142966
white throated tinamou	0.1163	0.1294	0.1421856
ostrich	0.14055	0.1619	0.1959332

55

56

Table S49. Comparison of *dN/dS* ratios of oogenesis genes to that of the genome background (Macrochromosome genes). "*" denotes significant differences (Wilcoxon rank sum test, $p<0.05$).

Species	**Macrochromosome gene median *dN/dS* ratio**	**Oogenesis gene median *dN/dS* ratio**	**One-tailed Wilcoxon rank sum test p-value (Macro < Oogenesis)**
downy woodpecker	0.1217	0.11935	0.5334802
carmine bee-eater	0.1248	0.26025	0.03724567*
rhinoceros hornbill	0.1389	0.13965	0.6042751
bar-tailed trogon	0.1277	0.1532	0.4226511
cuckoo roller	0.1363	0.1845	0.2373658
speckled mousebird	0.1301	0.2048	0.1443924
barn owl	0.1535	0.21735	0.2942273
turkey vulture	0.1858	0.19065	0.3450349
red-legged seriema	0.15375	0.1619	0.4526495
peregrine falcon	0.1462	0.2035	0.20318
kea	0.1495	0.1096	0.7376749
budgerigar	0.1515	0.2446	0.2127455
rifleman	0.14	0.2736	0.1979866
golden-collared manakin	0.15085	0.22625	0.3405046
American crow	0.1642	0.29365	0.1249903
zebra finch	0.1834	0.2514	0.3136557
medium ground-finch	0.1386	0.195	0.3004913
sunbittern	0.1268	0.1516	0.3544757
white-tailed tropicbird	0.1418	0.1621	0.3254336
red-throated loon	0.1527	0.0868	0.8061988
emperor penguin	0.2251	0.2998	0.07696919

Adelie penguin	0.2036	0.2121	0.6440341
northern fulmar	0.1697	0.16085	0.4191518
great cormorant	0.15185	0.2313	0.1493874
crested ibis	0.1527	0.13105	0.8043484
little egret	0.13815	0.27055	0.2171667
dalmatian pelican	0.1699	0.1618	0.3128074
grey crowned crane	0.1383	0.1855	0.350855
killdeer	0.1296	0.1386	0.4709844
hoatzin	0.15225	0.1346	0.6870383
red-crested turaco	0.1395	0.11755	0.6273355
MacQueen's bustard	0.1451	0.1431	0.597612
common cuckoo	0.1432	0.1874	0.3778765
chuck-wills-widow	0.1472	0.2047	0.1483873
chimney Swift	0.1408	0.1896	0.4724601
Anna's Hummingbird	0.14085	0.1435	0.63425
great crested grebe	0.1434	0.11145	0.6896767
American flamingo	0.1548	0.34165	0.1647673
pigeon	0.1418	0.16445	0.4556204
brown mesite	0.1375	0.237	0.1189107
yellow-throated sandgrouse	0.1647	0.1629	0.4060455
Peking duck	0.1164	0.22285	0.0352362*
turkey	0.1413	0.2233	0.1219174
chicken	0.1416	0.27285	0.06687185
white throated tinamou	0.1163	0.1296	0.675169
ostrich	0.14055	0.1359	0.7439552

Table S50. 15 genes that have previously been implicated in influencing avian plumage colors.

Gene name	Chicken gene ID
StAR	ENSGALG00000003242
Sox10	ENSGALG00000012290
DCT	ENSGALG00000016899
CD36	ENSGALG00000008439
APOD	ENSGALG00000006995
SLC24A4	ENSGALG00000010793
c-Kit	ENSGALG00000011206
GSTA2	ENSGALG00000016328
SLC45A2	ENSGALG00000003310
SLC24A5	ENSGALG00000004885
BCMO1	ENSGALG00000005408
SR-BI	ENSGALG00000003018
Mitf	ENSGALG00000007679
MC1R	ENSGALG00000023459
ASIP	ENSGALG00000021455

Table S51. Comparison of *dN/dS* ratios of plumage genes to that of the genome background (Macrochromosome genes). "*" denotes significant differences (Wilcoxon rank sum test, $p<0.05$).

Species	Macrochromosome gene median *dN/dS* ratio	Plumage gene median *dN/dS* ratio	One-tailed Wilcoxon rank sum test p-value (Macro < Plumage)
downy woodpecker	0.1219	0.1495	0.2694523
carmine bee-eater	0.1249	0.22755	0.0133170*

rhinoceros hornbill	0.1396	0.21685	0.0444251*
bar-tailed trogon	0.128	0.2124	0.2490953
cuckoo roller	0.137	0.1848	0.2177296
speckled mousebird	0.1303	0.2035	0.064638
barn owl	0.1539	0.1699	0.5554984
turkey vulture	0.1858	0.1997	0.1730825
red-legged seriema	0.1541	0.2598	0.0576704
peregrine falcon	0.1462	0.1612	0.2178247
budgerigar	0.1523	0.2139	0.1143497
kea	0.14955	0.10685	0.454076
rifleman	0.14055	0.1229	0.3264488
golden-collared manakin	0.1512	0.2034	0.1476205
American crow	0.1645	0.3278	0.0103829*
zebra finch	0.1823	0.25405	0.2126022
medium ground-finch	0.1387	0.2011	0.2244475
white-tailed tropicbird	0.142	0.1778	0.2536954
sunbittern	0.1269	0.14445	0.3167713
red-throated loon	0.1531	0.16395	0.3563454
emperor penguin	0.2267	0.11715	0.8320503
Adelie penguin	0.2039	0.20745	0.6222136
northern fulmar	0.17	0.2549	0.05850807
great cormorant	0.15255	0.1916	0.1531844
crested ibis	0.1534	0.363	0.0001536*
little egret	0.1387	0.18985	0.0334631*

dalmatian pelican	0.1708	0.16345	0.3136998
killdeer	0.1299	0.1836	0.0844354
grey-crowned crane	0.13875	0.18645	0.1979334
hoatzin	0.1523	0.1996	0.5186856
red-crested turaco	0.14015	0.2427	0.0834987
MacQueen's bustard	0.1451	0.1506	0.6904103
common cuckoo	0.1434	0.18615	0.0777690
chuck-wills-widow	0.148	0.2023	0.2203966
Anna's hummingbird	0.141	0.2132	0.0637746
chimney swift	0.14105	0.2836	0.0321046*
American flamingo	0.1548	0.1672	0.5621139
great-crested grebe	0.1436	0.2238	0.0830221
pigeon	0.1429	0.2909	0.0395470*
yellow-throated sandgrouse	0.1649	0.16135	0.598092
brown mesite	0.13775	0.1931	0.0986216
Peking duck	0.11675	0.1786	0.0363636*
turkey	0.1418	0.1545	0.0552585
chicken	0.1423	0.09285	0.7361299
white throated tinamou	0.1162	0.14115	0.2600189
common ostrich	0.1404	0.16005	0.5172032

67

References and Notes

1. F. Gill, D. Donsker, IOC World Bird List (version 3.5) (2013).

2. L. M. Chiappe, L. M. Witmer, *Mesozoic Birds: Above the Heads of Dinosaurs* (Univ. California Press, Berkeley, CA, 2002).

3. G. Dyke, G. W. Kaiser, *Living Dinosaurs the Evolutionary History of Modern Birds* (Wiley-Blackwell, Hoboken, NJ, 2011).

4. A. Feduccia, 'Big bang' for tertiary birds? *Trends Ecol. Evol.* **18**, 172–176 (2003). doi:10.1016/S0169-5347(03)00017-X

5. E. D. Jarvis *et al*., Whole genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).

6. B. G. Holt, J. P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araújo, D. Dimitrov, P. H. Fabre, C. H. Graham, G. R. Graves, K. A. Jønsson, D. Nogués-Bravo, Z. Wang, R. J. Whittaker, J. Fjeldså, C. Rahbek, An update of Wallace's zoogeographic regions of the world. *Science* **339**, 74–78 (2013). Medline doi:10.1126/science.1228282

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). Medline doi:10.1038/nature11631

8. H. Zeigler, P. E. Marler, in Behavioral Neurobiology of Birdsong, December 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, Cambridge, UK, 1998).

10. D. J. Alexander, A review of avian influenza in different bird species. *Vet. Microbiol.* **74**, 3–13 (2000). Medline doi:10.1016/S0378-1135(00)00160-7

11. L. W. Hillier, W. Miller, E. Birney, W. Warren, R. C. Hardison, C. P. Ponting, P. Bork, D. W. Burt, M. A. M. Groenen, M. E. Delany, J. B. Dodgson, A. T. Chinwalla, P. F. Cliften, S. W. Clifton, K. D. Delehaunty, C. Fronick, R. S. Fulton, T. A. Graves, C. Kremitzki, D. Layman, V. Magrini, J. D. McPherson, T. L. Miner, P. Minx, W. E. Nash, M. N. Nhan, J. O. Nelson, L. G. Oddy, C. S. Pohl, J. Randall-Maher, S. M. Smith, J. W. Wallis, S.-P. Yang, M. N. Romanov, C. M. Rondelli, B. Paton, J. Smith, D. Morrice, L. Daniels, H. G. Tempest, L. Robertson, J. S. Masabanda, D. K. Griffin, A. Vignal, V. Fillon, L. Jacobbson, S. Kerje, L. Andersson, R. P. M. Crooijmans, J. Aerts, J. J. van der Poel, H. Ellegren, R. B. Caldwell, S. J. Hubbard, D. V. Grafham, A. M. Kierzek, S. R. McLaren, I. M. Overton, H. Arakawa, K. J. Beattie, Y. Bezzubov, P. E. Boardman, J. K. Bonfield, M. D. R. Croning, R. M. Davies, M. D. Francis, S. J. Humphray, C. E. Scott, R. G. Taylor, C. Tickle, W. R. A. Brown, J. Rogers, J.-M. Buerstedde, S. A. Wilson, L. Stubbs, I. Ovcharenko, L. Gordon, S. Lucas, M. M. Miller, H. Inoko, T. Shiina, J. Kaufman, J. Salomonsen, K. Skjoedt, G. K.-S. Wong, J. Wang, B. Liu, J. Wang, J. Yu, H. Yang, M. Nefedov, M. Koriabine, P. J. deJong, L. Goodstadt, C. Webber, N. J. Dickens, I. Letunic, M. Suyama, D. Torrents, C. von Mering, E. M. Zdobnov, K. Makova, A. Nekrutenko, L.

Elnitski, P. Eswara, D. C. King, S. Yang, S. Tyekucheva, A. Radakrishnan, R. S. Harris, F. Chiaromonte, J. Taylor, J. He, M. Rijnkels, S. Griffiths-Jones, A. Ureta-Vidal, M. M. Hoffman, J. Severin, S. M. J. Searle, A. S. Law, D. Speed, D. Waddington, Z. Cheng, E. Tuzun, E. Eichler, Z. Bao, P. Flicek, D. D. Shteynberg, M. R. Brent, J. M. Bye, E. J. Huckle, S. Chatterji, C. Dewey, L. Pachter, A. Kouranov, Z. Mourelatos, A. G. Hatzigeorgiou, A. H. Paterson, R. Ivarie, M. Brandstrom, E. Axelsson, N. Backstrom, S. Berlin, M. T. Webster, O. Pourquie, A. Reymond, C. Ucla, S. E. Antonarakis, M. Long, J. J. Emerson, E. Betrán, I. Dupanloup, H. Kaessmann, A. S. Hinrichs, G. Bejerano, T. S. Furey, R. A. Harte, B. Raney, A. Siepel, W. J. Kent, D. Haussler, E. Eyras, R. Castelo, J. F. Abril, S. Castellano, F. Camara, G. Parra, R. Guigo, G. Bourque, G. Tesler, P. A. Pevzner, A. Smit, L. A. Fulton, E. R. Mardis, R. K. Wilson, International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004). Medline doi:10.1038/nature03154

12. R. A. Dalloul, J. A. Long, A. V. Zimin, L. Aslam, K. Beal, L. A. Blomberg, P. Bouffard, D. W. Burt, O. Crasta, R. P. Crooijmans, K. Cooper, R. A. Coulombe, S. De, M. E. Delany, J. B. Dodgson, J. J. Dong, C. Evans, K. M. Frederickson, P. Flicek, L. Florea, O. Folkerts, M. A. Groenen, T. T. Harkins, J. Herrero, S. Hoffmann, H. J. Megens, A. Jiang, P. de Jong, P. Kaiser, H. Kim, K. W. Kim, S. Kim, D. Langenberger, M. K. Lee, T. Lee, S. Mane, G. Marcais, M. Marz, A. P. McElroy, T. Modise, M. Nefedov, C. Notredame, I. R. Paton, W. S. Payne, G. Pertea, D. Prickett, D. Puiu, D. Qioa, E. Raineri, M. Ruffier, S. L. Salzberg, M. C. Schatz, C. Scheuring, C. J. Schmidt, S. Schroeder, S. M. Searle, E. J. Smith, J. Smith, T. S. Sonstegard, P. F. Stadler, H. Tafer, Z. J. Tu, C. P. Van Tassell, A. J. Vilella, K. P. Williams, J. A. Yorke, L. Zhang, H. B. Zhang, X. Zhang, Y. Zhang, K. M. Reed, Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLOS Biol.* **8**, e1000475 (2010). Medline doi:10.1371/journal.pbio.1000475

13. W. C. Warren, D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y. C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S. P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson, The genome of a songbird. *Nature* **464**, 757–762 (2010). Medline doi:10.1038/nature08819

14. Y. Huang, Y. Li, D. W. Burt, H. Chen, Y. Zhang, W. Qian, H. Kim, S. Gan, Y. Zhao, J. Li, K. Yi, H. Feng, P. Zhu, B. Li, Q. Liu, S. Fairley, K. E. Magor, Z. Du, X. Hu, L. Goodman, H. Tafer, A. Vignal, T. Lee, K. W. Kim, Z. Sheng, Y. An, S. Searle, J. Herrero, M. A. Groenen, R. P. Crooijmans, T. Faraut, Q. Cai, R. G. Webster, J. R. Aldridge, W. C. Warren, S. Bartschat, S. Kehr, M. Marz, P. F. Stadler, J. Smith, R. H.

Kraus, Y. Zhao, L. Ren, J. Fei, M. Morisson, P. Kaiser, D. K. Griffin, M. Rao, F. Pitel, J. Wang, N. Li, The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat. Genet.* **45**, 776–783 (2013). Medline doi:10.1038/ng.2657

15. X. Zhan, S. Pan, J. Wang, A. Dixon, J. He, M. G. Muller, P. Ni, L. Hu, Y. Liu, H. Hou, Y. Chen, J. Xia, Q. Luo, P. Xu, Y. Chen, S. Liao, C. Cao, S. Gao, Z. Wang, Z. Yue, G. Li, Y. Yin, N. C. Fox, J. Wang, M. W. Bruford, Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat. Genet.* **45**, 563–566 (2013). Medline doi:10.1038/ng.2588

16. M. D. Shapiro, Z. Kronenberg, C. Li, E. T. Domyan, H. Pan, M. Campbell, H. Tan, C. D. Huff, H. Hu, A. I. Vickrey, S. C. Nielsen, S. A. Stringham, H. Hu, E. Willerslev, M. T. Gilbert, M. Yandell, G. Zhang, J. Wang, Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**, 1063–1067 (2013). Medline doi:10.1126/science.1230422

17. E. Dickinson, J. Remsen, *The Howard and Moore Complete Checklist of the Birds of the World* (Aves Press, Eastbourne, UK, 2013).

18. Materials and methods are available as supplementary materials on *Science* Online.

19. G. Zhang *et al*., Comparative genomic data of Avian Phylogenomics Project. *GigaScience* **3**, 26 (2014).

20. G. Ganapathy *et al*., High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience* **3**, 11 (2013).

21. T. R. Gregory, in The Animal Genome Size Database (2005); available at www.genomesize.com.

22. Q. Zhou *et al*., Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338 (2014).

23. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T.

Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). Medline doi:10.1038/35057062

24. R. E. Green *et al.*, Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).

25. A. L. Hughes, M. K. Hughes, Small genomes for better flyers. *Nature* **377**, 391 (1995). Medline doi:10.1038/377391a0

26. M. Rho, M. Zhou, X. Gao, S. Kim, H. Tang, M. Lynch, Independent mammalian genome contractions following the KT boundary. *Genome Biol. Evol.* **1**, 2–12 (2009). Medline doi:10.1093/gbe/evp007

27. S. Morand, R. E. Ricklefs, Genome size, longevity and development time in birds. *Trends Genet.* **17**, 567–568 (2001). Medline doi:10.1016/S0168-9525(01)02414-3

28. E. Waltari, S. V. Edwards, Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**, 539–552 (2002). Medline doi:10.1086/342079

29. M. G. Kidwell, Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49–63 (2002). Medline doi:10.1023/A:1016072014259

30. C. Feschotte, E. J. Pritham, DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007). Medline doi:10.1146/annurev.genet.40.110405.090448

31. M. Lynch, J. S. Conery, The evolutionary demography of duplicate genes. *J. Struct. Funct. Genomics* **3**, 35–44 (2003). Medline doi:10.1023/A:1022696612931

32. N. Sela, E. Kim, G. Ast, The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biol.* **11**, R59 (2010). Medline doi:10.1186/gb-2010-11-6-r59

33. A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* **16**, 203–215 (2008). Medline doi:10.1007/s10577-007-1202-6

34. Q. Zhang, S. V. Edwards, The evolution of intron size in amniotes: A role for powered flight? *Genome Biol. Evol.* **4**, 1033–1043 (2012). Medline doi:10.1093/gbe/evs070

35. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007). Medline doi:10.1038/nature05621

36. A. L. Hughes, R. Friedman, Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Mol. Biol. Evol.* **25**, 2681–2688 (2008). Medline doi:10.1093/molbev/msn207

37. 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. *Cytogenet. Genome Res.* **117**, 64–77 (2007). Medline doi:10.1159/000103166

38. K. R. Bradnam, J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J. A. Chapman, G. Chapuis, R. Chikhi, H. Chitsaz, W. C. Chou, J. Corbeil, C. Del Fabbro, T. R. Docking, R. Durbin, D. Earl, S. Emrich, P. Fedotov, N. A. Fonseca, G. Ganapathy, R. A. Gibbs, S. Gnerre, E. Godzaridis, S. Goldstein, M. Haimel, G. Hall, D. Haussler, J. B. Hiatt, I. Y. Ho, J. Howard, M. Hunt, S. D. Jackman, D. B. Jaffe, E. D. Jarvis, H. Jiang, S. Kazakov, P. J. Kersey, J. O. Kitzman, J. R. Knight, S. Koren, T. W. Lam, D. Lavenier, F. Laviolette, Y. Li, Z. Li, B. Liu, Y. Liu, R. Luo, I. Maccallum, M. D. Macmanes, N. Maillet, S. Melnikov, D. Naquin, Z. Ning, T. D. Otto, B. Paten, O. S. Paulo, A. M. Phillippy, F. Pina-Martins, M. Place, D. Przybylski, X. Qin, C. Qu, F. J. Ribeiro, S. Richards, D. S. Rokhsar, J. G. Ruby, S. Scalabrin, M. C. Schatz, D. C. Schwartz, A. Sergushichev, T. Sharpe, T. I. Shaw, J. Shendure, Y. Shi, J. T. Simpson, H. Song, F. Tsarev, F. Vezzi, R. Vicedomini, B. M. Vieira, J. Wang, K. C. Worley, S. Yin, S. M. Yiu, J. Yuan, G. Zhang, H. Zhang, S. Zhou, I. F. Korf, Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2**, 10 (2013). Medline doi:10.1186/2047-217X-2-10

39. J. Kim, D. M. Larkin, Q. Cai, Y. Asan, R. L. Zhang, L. Ge, B. Auvil, G. Capitanu, H. A. Zhang, J. Lewin, J. Ma, Reference-assisted chromosome assembly. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1785–1790 (2013). Medline doi:10.1073/pnas.1220349110

40. F. G. Hoffmann, J. C. Opazo, J. F. Storz, Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. *Mol. Biol. Evol.* **29**, 303–312 (2012). Medline doi:10.1093/molbev/msr207

41. F. G. Hoffmann, J. C. Opazo, J. F. Storz, Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. *Mol. Biol. Evol.* **25**, 591–602 (2008). Medline doi:10.1093/molbev/msn004

42. M. T. Grispo, C. Natarajan, J. Projecto-Garcia, H. Moriyama, R. E. Weber, J. F. Storz, Gene duplication and the evolution of hemoglobin isoform differentiation in birds. *J. Biol. Chem.* **287**, 37647–37658 (2012). Medline doi:10.1074/jbc.M112.375600

43. C. F. Baer, M. M. Miyamoto, D. R. Denver, Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat. Rev. Genet.* **8**, 619–631 (2007). Medline doi:10.1038/nrg2158

44. C. Venditti, M. Pagel, Speciation as an active force in promoting genetic evolution. *Trends Ecol. Evol.* **25**, 14–20 (2010). Medline doi:10.1016/j.tree.2009.06.010

45. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). Medline doi:10.1101/gr.3715005

46. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, Broad Institute Sequencing Platform and Whole Genome Assembly Team, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, Genome Institute at Washington University, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011). Medline doi:10.1038/nature10530

47. 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 Evol. Biol.* **9**, 54 (2009). Medline doi:10.1186/1471-2148-9-54

48. 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. *J. Cell Biol.* **111**, 2939–2949 (1990). Medline doi:10.1083/jcb.111.6.2939

49. W. J. Swanson, V. D. Vacquier, The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**, 137–144 (2002). Medline doi:10.1038/nrg733

50. L. A. Pennacchio, N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, E. M. Rubin, In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006). Medline doi:10.1038/nature05295

51. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). Medline doi:10.1038/nature11247

52. J. E. Mank, E. Axelsson, H. Ellegren, Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res.* **17**, 618–624 (2007). Medline doi:10.1101/gr.6031907

53. 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 Res.* **15**, 120–125 (2005). Medline doi:10.1101/gr.3021305

54. A. Suh, M. Paus, M. Kiefmann, G. Churakov, F. A. Franke, J. Brosius, J. O. Kriegs, J. Schmitz, Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2**, 443 (2011). Medline doi:10.1038/ncomms1448

55. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004). Medline doi:10.1196/annals.1298.038

56. O. Whitney *et al.*, Core and region enriched networks of behaviorally regulated 1 genes and the singing genome. *Science* **346**, 1256780 (2014).

57. A. R. Pfenning *et al.*, Convergent transcriptional specializations in the brains of humans and song learning birds. *Science* **346**, 1256846 (2014).

58. H. J. Karten, A. Brzozowska-Prechtl, P. V. Lovell, D. D. Tang, C. V. Mello, H. Wang, P. P. Mitra, Digital atlas of the zebra finch (*Taeniopygia guttata*) brain: A high-resolution photo atlas. *J. Comp. Neurol.* **521**, 3702–3715 (2013). Medline doi:10.1002/cne.23443

59. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011). Medline doi:10.1093/bib/bbq072

60. E. R. Dumont, Bone density and the lightweight skeletons of birds. *Proc. Biol. Sci.* **277**, 2193–2198 (2010). Medline doi:10.1098/rspb.2010.0117

61. S. C. Gutzwiller, A. Su, P. M. O'Connor, Postcranial pneumaticity and bone structure in two clades of neognath birds. *Anat. Rec.* **296**, 867–876 (2013). Medline doi:10.1002/ar.22691

62. H. R. Duncker, Vertebrate lungs: Structure, topography and mechanics. A comparative perspective of the progressive integration of respiratory system, locomotor apparatus and ontogenetic development. *Respir. Physiol. Neurobiol.* **144**, 111–124 (2004). Medline doi:10.1016/j.resp.2004.07.020

63. J. B. West, R. R. Watson, Z. Fu, Major differences in the pulmonary circulation between birds and mammals. *Respir. Physiol. Neurobiol.* **157**, 382–390 (2007). Medline doi:10.1016/j.resp.2006.12.005

64. F. Gill, *Ornithology* (W.H. Freeman and Company, New York, 1995).

65. A. R. Haake, G. König, R. H. Sawyer, Avian feather development: Relationships between morphogenesis and keratinization. *Dev. Biol.* **106**, 406–413 (1984). Medline doi:10.1016/0012-1606(84)90240-9

66. J. Ross-Ibarra, The evolution of recombination under domestication: A test of two hypotheses. *Am. Nat.* **163**, 105–112 (2004). Medline doi:10.1086/380606

67. A. Blirt, G. Bell, Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326**, 803–805 (1987). Medline doi:10.1038/326803a0

68. A. Louchart, L. Viriot, From snout to beak: The loss of teeth in birds. *Trends Ecol. Evol.* **26**, 663–673 (2011). Medline doi:10.1016/j.tree.2011.09.004

69. B. C. Livezey, R. L. Zusi, Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.* **149**, 1–95 (2007). Medline doi:10.1111/j.1096-3642.2006.00293.x

70. J. Cracraft, Phylogenetic relationships and monophyly of loons, grebes, and hesperornithiform birds, with comments on the early history of birds. *Syst. Biol.* **31**, 35–56 (1982). doi:10.1093/sysbio/31.1.35

71. R. W. Meredith, G. Zhang, M. T. P. Gilbert, E. D. Jarvis, M. S. Springer, Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**, 1254390 (2014).

72. 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. *Mol. Biol. Evol.* **21**, 632–646 (2004). Medline doi:10.1093/molbev/msh054

73. I. B. Chatterjee, Evolution and the biosynthesis of ascorbic acid. *Science* **182**, 1271–1272 (1973). Medline doi:10.1126/science.182.4118.1271

74. J. Cui, X. Yuan, L. Wang, G. Jones, S. Zhang, Recent loss of vitamin C biosynthesis ability in bats. *PLOS ONE* **6**, e27114 (2011). Medline doi:10.1371/journal.pone.0027114

75. J. Cui, Y.-H. Pan, Y. Zhang, G. Jones, S. Zhang, Progressive pseudogenization: Vitamin C synthesis and its loss in bats. *Mol. Biol. Evol.* **28**, 1025–1031 (2011). Medline doi:10.1093/molbev/msq286

76. W. I. L. Davies, S. P. Collin, D. M. Hunt, Molecular ecology and adaptation of visual photopigments in craniates. *Mol. Ecol.* **21**, 3121–3158 (2012). Medline doi:10.1111/j.1365-294X.2012.05617.x

77. S. S. Pires, J. Shand, J. Bellingham, C. Arrese, M. Turton, S. Peirson, R. G. Foster, S. Halford, Isolation and characterization of melanopsin (Opn4) from the Australian marsupial *Sminthopsis crassicaudata* (fat-tailed dunnart). *Proc. Biol. Sci.* **274**, 2791–2799 (2007). Medline doi:10.1098/rspb.2007.0976

78. J. K. Bowmaker, G. R. Martin, Visual pigments and oil droplets in the penguin, *Spheniscus humboldti*. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **156**, 71–77 (1985). doi:10.1007/BF00610668

79. A. Ödeen, O. Håstad, P. Alström, Evolution of ultraviolet vision in the largest avian radiation—The passerines. *BMC Evol. Biol.* **11**, 313 (2011). Medline doi:10.1186/1471-2148-11-313

80. L. Chittka, Does bee color vision predate the evolution of flower color? *Naturwissenschaften* **83**, 136–138 (1996). doi:10.1007/BF01142181

81. S. S. Guraya, *Ovarian Follicles in Reptiles and Birds* (Springer-Verlag, Berlin, Germany, 1989).

82. X. Zheng, J. O'Connor, F. Huchzermeyer, X. Wang, Y. Wang, M. Wang, Z. Zhou, Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour. *Nature* **495**, 507–511 (2013). Medline doi:10.1038/nature11985

83. M. Jo, T. E. Curry Jr., Regulation of matrix metalloproteinase-19 messenger RNA expression in the rat ovary. *Biol. Reprod.* **71**, 1796–1806 (2004). Medline doi:10.1095/biolreprod.104.031823

84. 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. *J. Clin. Endocrinol. Metab.* **92**, 2659–2664 (2007). Medline doi:10.1210/jc.2006-2600

85. T. R. Birkhead, A. P. Møller, Sperm competition and sexual selection. (Academic Press, 1998).

86. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (J. Murray, London, 1871).

87. M. Zuk, J. D. Ligon, R. Thornhill, Effects of experimental manipulation of male secondary sex characters on female mate preference in red jungle fowl. *Anim. Behav.* **44**, 999–1006 (1992). doi:10.1016/S0003-3472(05)80312-4

88. C. Mateos, J. Carranza, The role of bright plumage in male-male interactions in the ring-necked pheasant. *Anim. Behav.* **54**, 1205–1214 (1997). Medline doi:10.1006/anbe.1997.0516

89. A. R. Quinlan, I. M. Hall, Characterizing complex structural variation in germline and somatic genomes. *Trends Genet.* **28**, 43–53 (2012). Medline doi:10.1016/j.tig.2011.10.002

90. J. F. Storz, J. C. Opazo, F. G. Hoffmann, Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol. Phylogenet. Evol.* **66**, 469–478 (2013). Medline doi:10.1016/j.ympev.2012.07.013

91. F. G. Hoffmann, J. F. Storz, T. A. Gorr, J. C. Opazo, Lineage-specific patterns of functional diversification in the α- and β-globin gene families of tetrapod vertebrates. *Mol. Biol. Evol.* **27**, 1126–1138 (2010). Medline doi:10.1093/molbev/msp325

92. R. Griffiths, M. C. Double, K. Orr, R. J. Dawson, A DNA test to sex most birds. *Mol. Ecol.* **7**, 1071–1075 (1998). Medline doi:10.1046/j.1365-294x.1998.00389.x

93. P. D. Hebert, M. Y. Stoeckle, T. S. Zemlak, C. M. Francis, Identification of birds through DNA barcodes. *PLOS Biol.* **2**, e312 (2004). Medline doi:10.1371/journal.pbio.0020312

94. C. Li, D. Lambert, S. Subramanian, C. D. Millar, C. P. Ponting, L. Kong, C. M. Rands, M. K. Fujita, S. Meader, H. Ellegren, K. Nam, T. A. Castoe, D. D. Pollock, W. Gu, M. J. Greenwold, S. Y. W. Ho, D. W. Burt, E. D. J. Jarvis, T. P. Gilbert, Y. Zhang, J. Li, H. Hu, H. Pan, L. Xu, Y. Deng, H. Yu, Q. Li, L. Jin, Y. Chen, B. Liu, L. Yang, S. Liu, Y. Zhang, Y. Lang, J. Xia, W. He, Q. Shi, J. Wang, G. Zhang, J. Jian Wang, H. Yang, Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience* 10.1186/2047-217X-3-27 (2014).

95. S. Li et al., Rapid decline of genomic diversity in restored population of almost-extinct Crested Ibis. *Xxxxxxxxxxxxx* **XX**, XXXX (2014).

96. R. Li, H. Zhu, J. Ruan, W. Qian, X. Fang, Z. Shi, Y. Li, S. Li, G. Shan, K. Kristiansen, S. Li, H. Yang, J. Wang, J. Wang, De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010). Medline doi:10.1101/gr.097261.109

97. R. Li, W. Fan, G. Tian, H. Zhu, L. He, J. Cai, Q. Huang, Q. Cai, B. Li, Y. Bai, Z. Zhang, Y. Zhang, W. Wang, J. Li, F. Wei, H. Li, M. Jian, J. Li, Z. Zhang, R. Nielsen, D. Li, W. Gu, Z. Yang, Z. Xuan, O. A. Ryder, F. C. Leung, Y. Zhou, J. Cao, X. Sun, Y. Fu, X. Fang, X. Guo, B. Wang, R. Hou, F. Shen, B. Mu, P. Ni, R. Lin, W. Qian, G. Wang, C. Yu, W. Nie, J. Wang, Z. Wu, H. Liang, J. Min, Q. Wu, S. Cheng, J. Ruan, M. Wang, Z. Shi, M. Wen, B. Liu, X. Ren, H. Zheng, D. Dong, K. Cook, G. Shan, H. Zhang, C. Kosiol, X. Xie, Z. Lu, H. Zheng, Y. Li, C. C. Steiner, T. T. Lam, S. Lin, Q. Zhang, G. Li, J. Tian, T. Gong, H. Liu, D. Zhang, L. Fang, C. Ye, J. Zhang, W. Hu, A. Xu, Y. Ren, G. Zhang, M. W. Bruford, Q. Li, L. Ma, Y. Guo, N. An, Y. Hu, Y. Zheng, Y. Shi, Z. Li, Q. Liu, Y. Chen, J. Zhao, N. Qu, S. Zhao, F. Tian, X. Wang, H. Wang, L. Xu, X. Liu, T. Vinar, Y. Wang, T. W. Lam, S. M. Yiu, S. Liu, H. Zhang, D. Li, Y. Huang, X. Wang, G. Yang, Z. Jiang, J. Wang, N. Qin, L. Li, J. Li, L. Bolund, K. Kristiansen, G. K. Wong, M. Olson, X. Zhang, S. Li, H. Yang, J. Wang, J. Wang, The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311–317 (2010). Medline doi:10.1038/nature08696

98. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35** (Database), D61–D65 (2007). Medline doi:10.1093/nar/gkl842

99. P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. K. Kähäri, D. Keefe, S. Keenan, R. Kinsella, M. Komorowska, G. Koscielny, E. Kulesha, P. Larsson, I. Longden, W. McLaren, M. Muffato, B. Overduin, M. Pignatelli, B. Pritchard, H. S. Riat, G. R. Ritchie, M. Ruffier, M. Schuster, D. Sobral, Y. A. Tang, K. Taylor, S. Trevanion, J. Vandrovcova, S. White, M. Wilson, S. P. Wilder, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernández-Suarez, J. Harrow, J. Herrero, T. J. Hubbard, A. Parker, G. Proctor, G. Spudich, J. Vogel, A. Yates, A. Zadissa, S. M. Searle, Ensembl 2012. *Nucleic Acids Res.* **40** (D1), D84–D90 (2012). Medline doi:10.1093/nar/gkr991

100. P. A. Fujita, B. Rhead, A. S. Zweig, A. S. Hinrichs, D. Karolchik, M. S. Cline, M. Goldman, G. P. Barber, H. Clawson, A. Coelho, M. Diekhans, T. R. Dreszer, B. M. Giardine, R. A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R. M. Kuhn, K. Learned, C. H. Li, L. R. Meyer, A. Pohl, B. J. Raney, K. R. Rosenbloom, K. E. Smith, D. Haussler, W. J. Kent, The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011). Medline doi:10.1093/nar/gkq963

101. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). Medline doi:10.1093/nar/25.17.3389

102. R. She, J. S. C. Chu, K. Wang, J. Pei, N. Chen, GenBlastA: Enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009). Medline doi:10.1101/gr.082081.108

103. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004). Medline doi:10.1101/gr.1865504

104. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). Medline doi:10.1093/nar/gkh340

105. Z. Wang, J. Pascual-Anaya, A. Zadissa, W. Li, Y. Niimura, Z. Huang, C. Li, S. White, Z. Xiong, D. Fang, B. Wang, Y. Ming, Y. Chen, Y. Zheng, S. Kuraku, M. Pignatelli, J. Herrero, K. Beal, M. Nozawa, Q. Li, J. Wang, H. Zhang, L. Yu, S. Shigenobu, J. Wang, J. Liu, P. Flicek, S. Searle, J. Wang, S. Kuratani, Y. Yin, B. Aken, G. Zhang, N. Irie, 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). Medline doi:10.1038/ng.2615

106. J. A. St John, E. L. Braun, S. R. Isberg, L. G. Miles, A. Y. Chong, J. Gongora, P. Dalzell, C. Moran, B. Bed'hom, A. Abzhanov, S. C. Burgess, A. M. Cooksey, T. A. Castoe, N. G. Crawford, L. D. Densmore, J. C. Drew, S. V. Edwards, B. C. Faircloth, M. K. Fujita, M. J. Greenwold, F. G. Hoffmann, J. M. Howard, T. Iguchi, D. E. Janes, S. Y. Khan, S. Kohno, A. J. de Koning, S. L. Lance, F. M. McCarthy, J. E. McCormack, M. E. Merchant, D. G. Peterson, D. D. Pollock, N. Pourmand, B. J. Raney, K. A. Roessler, J. R. Sanford, R. H. Sawyer, C. J. Schmidt, E. W. Triplett, T. D. Tuberville, M. Venegas-Anaya, J. T. Howard, E. D. Jarvis, L. J. Guillette Jr., T. C. Glenn, R. E. Green, D. A. Ray, Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* **13**, 415 (2012). Medline doi:10.1186/gb-2012-13-1-415

107. J. Alföldi, F. Di Palma, M. Grabherr, C. Williams, L. Kong, E. Mauceli, P. Russell, C. B. Lowe, R. E. Glor, J. D. Jaffe, D. A. Ray, S. Boissinot, A. M. Shedlock, C. Botka, T. A. Castoe, J. K. Colbourne, M. K. Fujita, R. G. Moreno, B. F. ten Hallers, D. Haussler, A. Heger, D. Heiman, D. E. Janes, J. Johnson, P. J. de Jong, M. Y. Koriabine, M. Lara, P. A. Novick, C. L. Organ, S. E. Peach, S. Poe, D. D. Pollock, K. de Queiroz, T. Sanger, S. Searle, J. D. Smith, Z. Smith, R. Swofford, J. Turner-Maier, J. Wade, S. Young, A. Zadissa, S. V. Edwards, T. C. Glenn, C. J. Schneider, J. B. Losos, E. S. Lander, M. Breen, C. P. Ponting, K. Lindblad-Toh, The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011). Medline doi:10.1038/nature10390

108. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0 (1996); available at www.repeatmasker.org.

109. A. F. A. Smit, R. Hubley, RepeatModeler Open-1.0 (2008); available at www.repeatmasker.org.

110. D. Brawand, M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, F. W. Albert, U. Zeller, P. Khaitovich, F. Grützner, S. Bergmann, R. Nielsen, S. Pääbo, H. Kaessmann, The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011). Medline doi:10.1038/nature10532

111. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009). Medline doi:10.1093/bioinformatics/btp120

112. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010). Medline doi:10.1038/nbt.1621

113. L. Kong, Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao, L. Wei, G. Gao, CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007). Medline doi:10.1093/nar/gkm391

114. S. Washietl, S. Findeiss, S. A. Müller, S. Kalkhof, M. von Bergen, I. L. Hofacker, P. F. Stadler, N. Goldman, RNAcode: Robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **17**, 578–594 (2011). Medline doi:10.1261/rna.2536111

115. Y. Fan, Z. Y. Huang, C. C. Cao, C. S. Chen, Y. X. Chen, D. D. Fan, J. He, H. L. Hou, L. Hu, X. T. Hu, X. T. Jiang, R. Lai, Y. S. Lang, B. Liang, S. G. Liao, D. Mu, Y. Y. Ma, Y. Y. Niu, X. Q. Sun, J. Q. Xia, J. Xiao, Z. Q. Xiong, L. Xu, L. Yang, Y. Zhang, W. Zhao, X. D. Zhao, Y. T. Zheng, J. M. Zhou, Y. B. Zhu, G. J. Zhang, J. Wang, Y. G. Yao, Genome of the Chinese tree shrew. *Nat. Commun.* **4**, 1426 (2013). Medline doi:10.1038/ncomms2416

116. E. B. Kim, X. Fang, A. A. Fushan, Z. Huang, A. V. Lobanov, L. Han, S. M. Marino, X. Sun, A. A. Turanov, P. Yang, S. H. Yim, X. Zhao, M. V. Kasaikina, N. Stoletzki, C. Peng, P. Polak, Z. Xiong, A. Kiezun, Y. Zhu, Y. Chen, G. V. Kryukov, Q. Zhang, L. Peshkin, L. Yang, R. T. Bronson, R. Buffenstein, B. Wang, C. Han, Q. Li, L. Chen, W. Zhao, S. R. Sunyaev, T. J. Park, G. Zhang, J. Wang, V. N. Gladyshev, Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**, 223–227 (2011). Medline doi:10.1038/nature10533

117. G. Zhang, C. Cowled, Z. Shi, Z. Huang, K. A. Bishop-Lilly, X. Fang, J. W. Wynne, Z. Xiong, M. L. Baker, W. Zhao, M. Tachedjian, Y. Zhu, P. Zhou, X. Jiang, J. Ng, L. Yang, L. Wu, J. Xiao, Y. Feng, Y. Chen, X. Sun, Y. Zhang, G. A. Marsh, G. Crameri, C. C. Broder, K. G. Frey, L. F. Wang, J. Wang, Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013). Medline doi:10.1126/science.1230835

118. J. U. Pontius, J. C. Mullikin, D. R. Smith, K. Lindblad-Toh, S. Gnerre, M. Clamp, J. Chang, R. Stephens, B. Neelam, N. Volfovsky, A. A. Schäffer, R. Agarwala, K. Narfström, W. J. Murphy, U. Giger, A. L. Roca, A. Antunes, M. Menotti-Raymond, N. Yuhki, J. Pecon-Slattery, W. E. Johnson, G. Bourque, G. Tesler, S. J. O'Brien, Agencourt Sequencing Team, NISC Comparative Sequencing Program, Initial sequence and comparative analysis of the cat genome. *Genome Res.* **17**, 1675–1689 (2007). Medline doi:10.1101/gr.6380007

119. M. A. Groenen, A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H. J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S.

Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S. H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z. Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z. L. Hu, S. J. Humphray, T. Hunt, H. Hornshøj, J. T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J. H. Kim, K. W. Kim, T. H. Kim, G. Larson, K. Lee, K. T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. T. Nguyen, P. Ni, S. J. Oh, S. Onteru, F. Panitz, E. W. Park, H. S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, L. B. Schook, Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012). Medline doi:10.1038/nature11622

120. R. S. Harris, thesis, Pennsylvania State University (2007).

121. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484–11489 (2003). Medline doi:10.1073/pnas.1932072100

122. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004). Medline doi:10.1101/gr.1933104

123. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, The Gene Ontology Consortium, Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). Medline doi:10.1038/75556

124. M. G. Grabherr, P. Russell, M. Meyer, E. Mauceli, J. Alföldi, F. Di Palma, K. Lindblad-Toh, Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151 (2010). Medline doi:10.1093/bioinformatics/btq102

125. 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 Res. Notes* **2**, 148 (2009). Medline doi:10.1186/1756-0500-2-148

126. J. Ma, L. Zhang, B. B. Suh, B. J. Raney, R. C. Burhans, W. J. Kent, M. Blanchette, D. Haussler, W. Miller, Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16**, 1557–1565 (2006). Medline doi:10.1101/gr.5383506

127. D. M. Larkin, G. Pape, R. Donthu, L. Auvil, M. Welge, H. A. Lewin, Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* **19**, 770–777 (2009). Medline doi:10.1101/gr.086546.108

128. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Meth. Ecol. and Evol.* **3**, 217–223 (2012). doi:10.1111/j.2041-210X.2011.00169.x

129. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). Medline doi:10.1186/1471-2105-10-421

130. C. Arnold, P. F. Stadler, Polynomial algorithms for the maximal pairing problem: Efficient phylogenetic targeting on arbitrary trees. *Algorithms Mol. Biol.* **5**, 25 (2010). Medline doi:10.1186/1748-7188-5-25

131. C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997). Medline doi:10.1006/jmbi.1997.0951

132. T. A. Tatusova, T. L. Madden, BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247–250 (1999). Medline doi:10.1111/j.1574-6968.1999.tb13575.x

133. M. W. Hahn, T. De Bie, J. E. Stajich, C. Nguyen, N. Cristianini, Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* **15**, 1153–1160 (2005). Medline doi:10.1101/gr.3567505

134. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006). Medline doi:10.1093/bioinformatics/btl097

135. M. V. Han, G. W. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013). Medline doi:10.1093/molbev/mst100

136. R. W. Meredith, J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. Simão, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, W. J. Murphy, Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011). Medline doi:10.1126/science.1211028

137. U. Hellsten, R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, D. S. Rokhsar, The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010). Medline doi:10.1126/science.1183670

138. O. R. Bininda-Emonds, M. Cardillo, K. E. Jones, R. D. MacPhee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, A. Purvis, The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007). Medline doi:10.1038/nature05634

139. L. J. Harmon, J. T. Weir, C. D. Brock, R. E. Glor, W. Challenger, GEIGER: Investigating evolutionary radiations. *Bioinformatics* **24**, 129–131 (2008). Medline doi:10.1093/bioinformatics/btm538

140. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004). Medline doi:10.1093/bioinformatics/btg412

141. R. C. Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2012).

142. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage, Los Angeles, CA, 2011).

143. X. Fan, J. Zhu, E. E. Schadt, J. S. Liu, Statistical power of phylo-HMM for evolutionarily conserved element detection. *BMC Bioinformatics* **8**, 374 (2007). Medline doi:10.1186/1471-2105-8-374

144. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (Suppl 2), ii215–ii225 (2003). Medline doi:10.1093/bioinformatics/btg1080

145. P. P. Amaral, M. B. Clark, D. K. Gascoigne, M. E. Dinger, J. S. Mattick, lncRNAdb: A reference database for long noncoding RNAs. *Nucleic Acids Res.* **39** (Database), D146–D151 (2011). Medline doi:10.1093/nar/gkq1138

146. 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). Medline doi:10.1016/j.cell.2011.11.055

147. I. L. Hofacker, Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003). Medline doi:10.1093/nar/gkg599

148. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). Medline doi:10.1093/molbev/msm088

149. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000). Medline doi:10.1093/oxfordjournals.molbev.a026334

150. P. V. Lovell, D. F. Clayton, K. L. Replogle, C. V. Mello, Birdsong "transcriptomics": Neurochemical specializations of the oscine song system. *PLOS ONE* **3**, e3440 (2008). Medline doi:10.1371/journal.pone.0003440

151. K. Wada, H. Sakaguchi, E. D. Jarvis, M. Hagiwara, Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *J. Comp. Neurol.* **476**, 44–64 (2004). Medline doi:10.1002/cne.20201

152. D. T. Whitaker, Differential gene expression in the anterior forebrain pathway nucleus area X during rapid vocal learning: Honors and undergraduate research; available at http://hdl.handle.net/19691 (2010).

153. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009). Medline doi:10.1038/nprot.2008.211

154. B. T. Sherman, W. Huang, Q. Tan, Y. Guo, S. Bour, D. Liu, R. Stephens, M. W. Baseler, H. C. Lane, R. A. Lempicki, DAVID Knowledgebase: A gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics* **8**, 426 (2007). Medline doi:10.1186/1471-2105-8-426

155. M. Fukuda, K. Mikoshiba, Synaptotagmin-like protein 1-3: A novel family of C-terminal-type tandem C2 proteins. *Biochem. Biophys. Res. Commun.* **281**, 1226–1233 (2001). Medline doi:10.1006/bbrc.2001.4512

156. H. Miwa, A. Miyake, Y. Kouta, A. Shimada, Y. Yamashita, Y. Nakayama, H. Yamauchi, M. Konishi, N. Itoh, A novel neural-specific BMP antagonist, Brorin-like, of the Chordin family. *FEBS Lett.* **583**, 3643–3648 (2009). Medline doi:10.1016/j.febslet.2009.10.044

157. D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan, R. Apweiler, QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009). Medline doi:10.1093/bioinformatics/btp536

158. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010). Medline doi:10.1093/molbev/msp259

159. W. Huang, B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane, R. A. Lempicki, The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183 (2007). Medline doi:10.1186/gb-2007-8-9-r183

160. W. Vandebergh, F. Bossuyt, Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Mol. Biol. Evol.* **29**, 995–1004 (2012). Medline doi:10.1093/molbev/msr269

161. W. J. Kent, BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002). Medline doi:10.1101/gr.229202. Article published online before March 2002

162. M. J. Greenwold, R. H. Sawyer, Genomic organization and molecular phylogenies of the beta (β) keratin multigene family in the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*): Implications for feather evolution. *BMC Evol. Biol.* **10**, 148 (2010). Medline doi:10.1186/1471-2148-10-148

163. L. Dalla Valle, A. Nardi, G. Bonazza, C. Zucal, D. Emera, L. Alibardi, Forty keratin-associated β-proteins (β-keratins) form the hard layers of scales, claws, and adhesive pads in the green anole lizard, *Anolis carolinensis*. *J. Exp. Zoolog. B Mol. Dev. Evol.* **314**, 11–32 (2010). Medline doi:10.1002/jez.b.21306

164. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). Medline doi:10.1093/bioinformatics/btl446

165. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G.

Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007). Medline doi:10.1093/bioinformatics/btm404

166. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012). Medline doi:10.1093/bioinformatics/bts199

167. A. C. Darling, B. Mau, F. R. Blattner, N. T. Perna, Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004). Medline doi:10.1101/gr.2289704

168. K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005). Medline doi:10.1093/nar/gki198

169. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011). Medline doi:10.1093/molbev/msr121

170. W. J. Murphy, P. A. Pevzner, S. J. O'Brien, Mammalian phylogenomics comes of age. *Trends Genet.* **20**, 631–639 (2004). Medline doi:10.1016/j.tig.2004.09.005

171. K. Okonechnikov, O. Golosova, M. Fursov, UGENE team, Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012). Medline doi:10.1093/bioinformatics/bts091

172. T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, Y. Shichida, Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5982–5987 (2002). Medline doi:10.1073/pnas.082666399

173. D. Charif, J. Thioulouse, J. R. Lobry, G. Perrière, Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* **21**, 545–547 (2005). Medline doi:10.1093/bioinformatics/bti037

174. S. E. Wilkie, P. R. Robinson, T. W. Cronin, S. Poopalasundaram, J. K. Bowmaker, D. M. Hunt, Spectral tuning of avian violet- and ultraviolet-sensitive visual pigments. *Biochemistry* **39**, 7895–7901 (2000). Medline doi:10.1021/bi992776m

175. S. Yokoyama, F. B. Radlwimmer, N. S. Blow, Ultraviolet pigments in birds evolved from violet pigments by a single amino acid change. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7366–7371 (2000). Medline doi:10.1073/pnas.97.13.7366

176. N. J. Nadeau, T. Burke, N. I. Mundy, Evolution of an avian pigmentation gene correlates with a measure of sexual selection. *Proc. Biol. Sci.* **274**, 1807–1813 (2007). Medline doi:10.1098/rspb.2007.0174

177. J. K. Hubbard, J. A. Uy, M. E. Hauber, H. E. Hoekstra, R. J. Safran, Vertebrate pigmentation: From underlying genes to adaptive function. *Trends Genet.* **26**, 231–239 (2010). Medline doi:10.1016/j.tig.2010.02.002

178. 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. *Proc. Biol. Sci.* **279**, 58–66 (2012). Medline doi:10.1098/rspb.2011.0765

179. A. Roulin, A. L. Ducrest, Genetics of colouration in birds. *Semin. Cell Dev. Biol.* **24**, 594–608 (2013). Medline doi:10.1016/j.semcdb.2013.05.005

180. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings Publishing Company, San Francisco, 1983).

181. J. K. Armenta, P. O. Dunn, L. A. Whittingham, Quantifying avian sexual dichromatism: A comparison of methods. *J. Exp. Biol.* **211**, 2423–2430 (2008). Medline doi:10.1242/jeb.013094

182. G. Feenders, M. Liedvogel, M. Rivas, M. Zapka, H. Horita, E. Hara, K. Wada, H. Mouritsen, E. D. Jarvis, Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. *PLOS One* **3**, e1768 (2008). Medline doi:10.1371/journal.pone.0001768

183. A. V. Zimin, A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, S. L. Salzberg, A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* **10**, R42 (2009). Medline doi:10.1186/gb-2009-10-4-r42

184. C. M. Wade, E. Giulotto, S. Sigurdsson, M. Zoli, S. Gnerre, F. Imsland, T. L. Lear, D. L. Adelson, E. Bailey, R. R. Bellone, H. Blöcker, O. Distl, R. C. Edgar, M. Garber, T. Leeb, E. Mauceli, J. N. MacLeod, M. C. Penedo, J. M. Raison, T. Sharpe, J. Vogel, L. Andersson, D. F. Antczak, T. Biagi, M. M. Binns, B. P. Chowdhary, S. J. Coleman, G. Della Valle, S. Fryc, G. Guérin, T. Hasegawa, E. W. Hill, J. Jurka, A. Kiialainen, G. Lindgren, J. Liu, E. Magnani, J. R. Mickelson, J. Murray, S. G. Nergadze, R. Onofrio, S. Pedroni, M. F. Piras, T. Raudsepp, M. Rocchi, K. H. Røed, O. A. Ryder, S. Searle, L. Skow, J. E. Swinburne, A. C. Syvänen, T. Tozaki, S. J. Valberg, M. Vaudin, J. R. White, M. C. Zody, E. S. Lander, K. Lindblad-Toh, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865–867 (2009). Medline doi:10.1126/science.1178158

185. A. Scally, J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lappalainen, T. Mailund, T. Marques-Bonet, S. McCarthy, S. H. Montgomery, P. C. Schwalie, Y. A. Tang, M. C. Ward, Y. Xue, B. Yngvadottir, C. Alkan, L. N. Andersen, Q. Ayub, E. V. Ball, K. Beal, B. J. Bradley, Y. Chen, C. M. Clee, S. Fitzgerald, T. A. Graves, Y. Gu, P. Heath, A. Heger, E. Karakoc, A. Kolb-Kokocinski, G. K. Laird, G. Lunter, S. Meader, M. Mort, J. C. Mullikin, K. Munch, T. D. O'Connor, A. D. Phillips, J. Prado-Martinez, A. S. Rogers, S. Sajjadian, D. Schmidt, K. Shaw, J. T. Simpson, P. D. Stenson, D. J. Turner, L. Vigilant, A. J. Vilella, W. Whitener, B. Zhu, D. N. Cooper, P. de Jong, E. T. Dermitzakis, E. E. Eichler, P. Flicek, N. Goldman, N. I. Mundy, Z. Ning, D. T. Odom, C. P. Ponting, M. A. Quail, O. A. Ryder, S. M. Searle, W. C. Warren, R. K. Wilson, M. H. Schierup, J. Rogers, C. Tyler-Smith, R. Durbin, Insights into hominid

evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012). Medline doi:10.1038/nature10842

186. R. A. Gibbs, J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y. S. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S. Pohl, Y. H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S. P. Yang, S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csürös, J. Glasscock, R. A. Harris, P. Havlak, A. R. Jackson, H. Jiang, Y. Liu, D. N. Messina, Y. Shen, H. X. Song, T. Wylie, L. Zhang, E. Birney, K. Han, M. K. Konkel, J. Lee, A. F. Smit, B. Ullmer, H. Wang, J. Xing, R. Burhans, Z. Cheng, J. E. Karro, J. Ma, B. Raney, X. She, M. J. Cox, J. P. Demuth, L. J. Dumas, S. G. Han, J. Hopkins, A. Karimpour-Fard, Y. H. Kim, J. R. Pollack, T. Vinar, C. Addo-Quaye, J. Degenhardt, A. Denby, M. J. Hubisz, A. Indap, C. Kosiol, B. T. Lahn, H. A. Lawson, A. Marklein, R. Nielsen, E. J. Vallender, A. G. Clark, B. Ferguson, R. D. Hernandez, K. Hirani, H. Kehrer-Sawatzki, J. Kolb, S. Patil, L. L. Pu, Y. Ren, D. G. Smith, D. A. Wheeler, I. Schenck, E. V. Ball, R. Chen, D. N. Cooper, B. Giardine, F. Hsu, W. J. Kent, A. Lesk, D. L. Nelson, W. E. O'brien, K. Prüfer, P. D. Stenson, J. C. Wallace, H. Ke, X. M. Liu, P. Wang, A. P. Xiang, F. Yang, G. P. Barber, D. Haussler, D. Karolchik, A. D. Kern, R. M. Kuhn, K. E. Smith, A. S. Zwieg, Rhesus Macaque Genome Sequencing and Analysis Consortium, Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007). Medline doi:10.1126/science.1139247

187. T. S. Mikkelsen, M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, J. Jurka, M. Kamal, E. Mauceli, S. M. Searle, T. Sharpe, M. L. Baker, M. A. Batzer, P. V. Benos, K. Belov, M. Clamp, A. Cook, J. Cuff, R. Das, L. Davidow, J. E. Deakin, M. J. Fazzari, J. L. Glass, M. Grabherr, J. M. Greally, W. Gu, T. A. Hore, G. A. Huttley, M. Kleber, R. L. Jirtle, E. Koina, J. T. Lee, S. Mahony, M. A. Marra, R. D. Miller, R. D. Nicholls, M. Oda, A. T. Papenfuss, Z. E. Parra, D. D. Pollock, D. A. Ray, J. E. Schein, T. P. Speed, K. Thompson, J. L. VandeBerg, C. M. Wade, J. A. Walker, P. D. Waters, C. Webber, J. R. Weidman, X. Xie, M. C. Zody, J. A. Graves, C. P. Ponting, M. Breen, P. B. Samollow, E. S. Lander, K. Lindblad-Toh, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007). Medline doi:10.1038/nature05805

188. R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R.

Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigó, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, E. S. Lander, Mouse Genome Sequencing Consortium, Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002). Medline doi:10.1038/nature01262

189. W. C. Warren, L. W. Hillier, J. A. Marshall Graves, E. Birney, C. P. Ponting, F. Grützner, K. Belov, W. Miller, L. Clarke, A. T. Chinwalla, S. P. Yang, A. Heger, D. P. Locke, P. Miethke, P. D. Waters, F. Veyrunes, L. Fulton, B. Fulton, T. Graves, J. Wallis, X. S. Puente, C. López-Otín, G. R. Ordóñez, E. E. Eichler, L. Chen, Z. Cheng, J. E. Deakin, A. Alsop, K. Thompson, P. Kirby, A. T. Papenfuss, M. J. Wakefield, T. Olender, D. Lancet, G. A. Huttley, A. F. Smit, A. Pask, P. Temple-Smith, M. A. Batzer, J. A. Walker, M. K. Konkel, R. S. Harris, C. M. Whittington, E. S. Wong, N. J. Gemmell, E. Buschiazzo, I. M. Vargas Jentzsch, A. Merkel, J. Schmitz, A. Zemann, G. Churakov, J. O. Kriegs, J. Brosius, E. P. Murchison, R. Sachidanandam, C. Smith, G. J. Hannon, E. Tsend-Ayush, D. McMillan, R. Attenborough, W. Rens, M. Ferguson-Smith, C. M. Lefèvre, J. A. Sharp, K. R. Nicholas, D. A. Ray, M. Kube, R. Reinhardt, T. H. Pringle, J. Taylor, R. C. Jones, B. Nixon, J. L. Dacheux, H. Niwa, Y. Sekita, X. Huang, A. Stark, P. Kheradpour, M. Kellis, P. Flicek, Y. Chen, C. Webber, R. Hardison, J. Nelson, K. Hallsworth-Pepin, K. Delehaunty, C. Markovic, P. Minx, Y. Feng, C. Kremitzki, M. Mitreva, J. Glasscock,

T. Wylie, P. Wohldmann, P. Thiru, M. N. Nhan, C. S. Pohl, S. M. Smith, S. Hou, M. Nefedov, P. J. de Jong, M. B. Renfree, E. R. Mardis, R. K. Wilson, Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008). Medline doi:10.1038/nature06936

190. Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005). Medline doi:10.1038/nature04072

191. D. P. Locke, L. W. Hillier, W. C. Warren, K. C. Worley, L. V. Nazareth, D. M. Muzny, S. P. Yang, Z. Wang, A. T. Chinwalla, P. Minx, M. Mitreva, L. Cook, K. D. Delehaunty, C. Fronick, H. Schmidt, L. A. Fulton, R. S. Fulton, J. O. Nelson, V. Magrini, C. Pohl, T. A. Graves, C. Markovic, A. Cree, H. H. Dinh, J. Hume, C. L. Kovar, G. R. Fowler, G. Lunter, S. Meader, A. Heger, C. P. Ponting, T. Marques-Bonet, C. Alkan, L. Chen, Z. Cheng, J. M. Kidd, E. E. Eichler, S. White, S. Searle, A. J. Vilella, Y. Chen, P. Flicek, J. Ma, B. Raney, B. Suh, R. Burhans, J. Herrero, D. Haussler, R. Faria, O. Fernando, F. Darré, D. Farré, E. Gazave, M. Oliva, A. Navarro, R. Roberto, O. Capozzi, N. Archidiacono, G. Della Valle, S. Purgato, M. Rocchi, M. K. Konkel, J. A. Walker, B. Ullmer, M. A. Batzer, A. F. Smit, R. Hubley, C. Casola, D. R. Schrider, M. W. Hahn, V. Quesada, X. S. Puente, G. R. Ordoñez, C. López-Otín, T. Vinar, B. Brejova, A. Ratan, R. S. Harris, W. Miller, C. Kosiol, H. A. Lawson, V. Taliwal, A. L. Martins, A. Siepel, A. Roychoudhury, X. Ma, J. Degenhardt, C. D. Bustamante, R. N. Gutenkunst, T. Mailund, J. Y. Dutheil, A. Hobolth, M. H. Schierup, O. A. Ryder, Y. Yoshinaga, P. J. de Jong, G. M. Weinstock, J. Rogers, E. R. Mardis, R. A. Gibbs, R. K. Wilson, Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529–533 (2011). Medline doi:10.1038/nature09687

192. R. A. Gibbs, G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott, D. Steffen, K. C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, R. A. Celera, M. D. Holt, P. G. Adams, H. Amanatides, M. Baden-Tillson, S. Barnstead, C. A. Chin, S. Evans, C. Ferriera, A. Fosler, Z. Glodek, D. Gu, C. L. Jennings, T. Kraft, C. M. Nguyen, C. Pfannkoch, G. G. Sitter, J. C. Sutton, T. Venter, D. Woodage, H. M. Smith, E. Lee, P. Gustafson, A. Cahill, L. Kana, K. Doucette-Stamm, K. Weinstock, R. B. Fechtel, D. M. Weiss, E. D. Dunn, R. W. Green, G. G. Blakesley, P. J. Bouffard, K. De Jong, B. Osoegawa, M. Zhu, J. Marra, I. Schein, C. Bosdet, S. Fjell, M. Jones, C. Krzywinski, A. Mathewson, N. Siddiqui, J. Wye, S. McPherson, C. M. Zhao, J. Fraser, S. Shetty, K. Shatsman, Y. Geer, S. Chen, W. C. Abramzon, P. H. Nierman, R. Havlak, K. J. Chen, A. Durbin, Y. Egan, X. Z. Ren, B. Song, Y. Li, X. Liu, S. Qin, K. C. Cawley, A. J. Worley, L. M. Cooney, K. D'Souza, J. Q. Martin, M. L. Wu, A. R. Gonzalez-Garay, K. J. Jackson, M. P. Kalafus, A. McLeod, D. Milosavljevic, A. Virk, D. A. Volkov, Z. Wheeler, J. A. Zhang, E. E. Bailey, E. Eichler, E. Tuzun, E. Birney, A. Mongin, C. Ureta-Vidal, E. Woodwark, P. Zdobnov, M. Bork, D. Suyama, M. Torrents, B. J. Alexandersson, J. M. Trask, H. Young, H. Huang, H. Wang, S. Xing, D. Daniels, J. Gietzen, K. Schmidt, U. Stevens, J. Vitt, F. Wingrove, M. Camara, J. F. Mar Albà, R. Abril, A. Guigo, I. Smit, E. M. Dubchak, O. Rubin, A. Couronne, N. Poliakov, D. Hübner, C. Ganten, O. Goesele, T. Hummel, Y. A. Kreitler, J. Lee, H. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. J. Lehrach, S. Jacob, J. Bromberg, M. I. Gullings-

Handley, A. E. Jensen-Seaman, J. Kwitek, D. Lazar, P. J. Pasko, S. Tonellato, C. P. Twigger, J. M. Ponting, S. Duarte, L. Rice, S. A. Goodstadt, R. D. Beatson, E. E. Emes, C. Winter, P. Webber, G. Brandt, M. Nyakatura, F. Adetobi, L. Chiaromonte, P. Elnitski, R. C. Eswara, M. Hardison, D. Hou, K. Kolbe, W. Makova, A. Miller, C. Nekrutenko, S. Riemer, J. Schwartz, S. Taylor, Y. Yang, K. Zhang, T. D. Lindpaintner, M. Andrews, M. Caccamo, L. Clamp, V. Clarke, R. Curwen, E. Durbin, S. M. Eyras, G. M. Searle, S. Cooper, M. Batzoglou, A. Brudno, E. A. Sidow, J. C. Stone, B. A. Venter, G. Payseur, C. Bourque, X. S. López-Otín, K. Puente, S. Chakrabarti, C. Chatterji, L. Dewey, N. Pachter, V. B. Bray, A. Yap, G. Caspi, P. A. Tesler, D. Pevzner, K. M. Haussler, R. Roskin, H. Baertsch, T. S. Clawson, A. S. Furey, D. Hinrichs, W. J. Karolchik, K. R. Kent, H. Rosenbloom, M. Trumbower, D. N. Weirauch, P. D. Cooper, B. Stenson, M. Ma, M. Brent, D. Arumugam, R. R. Shteynberg, M. S. Copley, H. Taylor, U. Riethman, J. Mudunuri, M. Peterson, A. Guyer, S. Felsenfeld, S. Old, F. Mockrin, Collins, Rat Genome Sequencing Project Consortium, Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004). Medline
doi:10.1038/nature02426