

## Novel insights into chromosome evolution in birds, archosaurs, and reptiles.

Marta Farré<sup>1</sup>, Jitendra Narayan<sup>2</sup>, Gancho T. Slavov<sup>2</sup>, Joana Damas<sup>1</sup>, Loretta Auvil<sup>3</sup>, Cai Li<sup>4,5</sup>, Erich D. Jarvis<sup>6</sup>, David W. Burt<sup>7</sup>, Darren K. Griffin<sup>8§</sup>, Denis M. Larkin<sup>1§\*</sup>

<sup>1</sup>Department of Comparative Biomedical Sciences, Royal Veterinary College, Royal College Street, NW1 0TU, University of London, London, UK

<sup>2</sup>Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Penglais Campus, SY23 3FL, Aberystwyth, UK

<sup>3</sup>Illinois Informatics Institute, University of Illinois, IL 61801, Urbana, Illinois, USA

<sup>4</sup>China National GeneBank, BGI-Shenzhen, 518083, Shenzhen, China

<sup>5</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark

<sup>6</sup>Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, MD 20815, Durham, USA

<sup>7</sup>Department of Genomics and Genetics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, EH25 9RG, Midlothian, UK

<sup>8</sup>School of Biosciences, University of Kent, CT2 7NY, Canterbury, UK

§Joint last authors

\*Author for Correspondence: Denis M. Larkin, Department of Comparative Biomedical Sciences, Royal Veterinary College, Royal College Street, NW1 0UT, London, UK, telephone number: +44-207-121-1906, e-mail: dlarkin@rvc.ac.uk

Data deposition: Homologous synteny block data can be viewed on the Evolution Highway comparative chromosome browser (<http://eh-demo.ncsa.uiuc.edu/birds>).

## Abstract (Max 250)

Homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in mammalian chromosomes are enriched for distinct DNA features, contributing to distinct phenotypes. To reveal HSB and EBR roles in avian evolution, we performed a sequence-based comparison of 21 avian and five outgroup species using recently sequenced genomes across the avian family tree and a newly-developed algorithm. We identified EBRs and HSBs in ancestral bird, archosaurian (bird, crocodile, dinosaur), and reptile chromosomes. Genes involved in the regulation of gene expression and biosynthetic processes were preferably located in HSBs, for example the avian-specific HSBs were enriched for genes involved in limb development. Within birds, some lineage-specific EBRs rearranged genes related to distinct phenotypes, such as forebrain development in parrots. Our findings provide novel evolutionary insights into genome evolution in birds, particularly how chromosome rearrangements likely contributed to the formation of novel phenotypes.

## INTRODUCTION

A prominent feature of animal genome evolution is the non-random rearrangement of chromosomes (Pevzner and Tesler 2003). For millions of years genomes of multiple species have maintained homologous synteny blocks (HSBs), demarcated by dynamic “evolutionary breakpoint regions” (EBRs) (Figure 1). Evidence suggests that each of them evolves by distinctly different mechanisms (Larkin, et al. 2009): HSBs maintain the order of genes related to organismal development whereas EBRs often affect chromosomal regions related to lineage-specific biology (Groenen, et al. 2012; Ullastres, et al. 2014). These data are somewhat mammal-centric and conclusions thus may not hold for other amniotes. While the availability of genetic maps and chromosome assemblies of the chicken, turkey, and zebra finch genomes provided an important insight into avian chromosome evolution (Burt, et al. 1999; Völker, et al. 2010; Warren, et al. 2010), a comprehensive study at the sequence level is lacking, making unclear if bird chromosomes follow similar patterns of evolution as their mammalian counterparts.

Birds have more compact genomes with shorter intronic and intergenic regions than mammals (ICGSC 2004; Zhang, et al. 2014). The proportion of repetitive DNA in bird genomes is ~15% (ICGSC 2004; Zhang, et al. 2014), whereas in mammals it is ~50% (Lander, et al. 2001). Birds have more gene families that lost paralogs than other amniotes (Huang, et al. 2013; Lovell, et al. 2014). Avian karyotypes have been maintained without interchromosomal changes for millions of years (Romanov, et al. 2014) and are less variable than those of mammals (Ellegren 2010; Ruiz-Herrera, et al. 2012) with a characteristic  $2n \sim 80$  in most species (Griffin, et al. 2007).

Using a new EBR-detection approach applied to 21 bird genomes assembled to whole chromosomes or large scaffolds (Zhang, et al. 2014), and four non-avian reptile genomes of similar quality, we examined the association of EBRs and multispecies HSBs (msHSBs) with

gene networks, transposable elements (TEs) and conserved non-coding sequences. We identified gene networks that: (1) were preferentially reshuffled during avian chromosome evolution, or (2) have been maintained in msHSBs for millions of years of evolution. Our results represent the first comprehensive sequence analysis of chromosome evolution in birds and reptiles, demonstrating how chromosome evolution may have acted upon the formation of various phenotypes.

## RESULTS AND DISCUSSION

**Lineage-specific EBRs in birds.** We developed an interactive resource for genome synteny comparison in 26 species (Evolution Highway; <http://eh-demo.ncsa.uiuc.edu/birds>; Supplementary Table 1). We aligned 20 avian and five outgroup genomes to the chicken genome to define syntenic fragments at three resolutions of rearrangement detection: 100Kbp, 300Kbp and 500Kbp (Figure 1). We developed and evaluated (Supplementary Table 2, 3 and 4) a method of detecting EBRs within scaffolds of scaffold-based assemblies that combines an algorithmic approach to identify putative EBRs (Supplementary Table 5) with independent PCR verification of these regions in several assemblies to find paired read spanning levels in scaffolds associated with confirmed EBRs in order to estimate and minimise the number of chimeric joints in the final EBR list (Supplementary Table 5 and 8). This resulted in 0-22% false positives and 33-45% false negatives in our EBR set, depending on the sequencing coverage of each assembly (Supplementary Table 7). At 100Kbp resolution 1,796 avian EBRs were assigned to phylogenetic nodes and 1,021 (56.85%) passed our chimeric scaffold detection quality controls. Out of 1,021 EBRs, 42 were specific to all Galliformes, and 16 were specific to the chicken lineage (Figure 1 and Supplementary Table 5). We detected a total of 874 lineage-specific EBRs, i.e. assigned to lineages leading to each species in our set

after the divergence from the most recent common ancestor with other included species (Supplementary Table 5).

**Lineage-specific EBRs are enriched in TEs in birds.** In mammals, lineage- and order-specific EBRs are enriched for TEs that were active at the time of lineage/order formation (Groenen, et al. 2012; Larkin, et al. 2009; Schibler, et al. 2006), and TEs can promote chromosome rearrangements by non-allelic homologous recombination (Bailey, et al. 2004). In birds, we found that one or more of four families of TEs (LINE-CR1, LTR-ERV1, LTR-ERV2, and LTR-ERV3) were significantly enriched in lineage-specific EBRs among 19 bird species (>100bp on average in the EBR- or non-EBR-containing non-overlapping 10Kbp genome intervals; FDR<10%; Figure 2). The only exceptions were ostrich and Adelie penguin lineage-specific EBRs, which had a significant negative association with the LINE-CR1 elements and LINE-CR1 and LTR-ERV1 elements, respectively, implying the presence of still unidentified lineage-specific TEs associated with EBRs in these two species. Our findings suggest that lineage-specific EBRs are associated with the presence of TE elements in birds, following the trend previously reported for mammals (Groenen, et al. 2012).

**Multispecies (ms)HSBs in avian and reptile genomes.** To evaluate if msHSBs were maintained during bird evolution, five sets of msHSBs (the regions of genomes that were not interrupted by EBRs; Supplementary Table 10 and 11) were defined: avian, archosaurian, archosaurian/testudines, sauropsid, and amniote. We detected 1,746 avian msHSBs, covering 76.29% of the chicken genome. Using the Kolmogorov-Smirnov test, the distribution of msHSB sizes was tested for goodness-of-fit to an exponential distribution, following previous publications (Larkin, et al. 2009; Pevzner and Tesler 2003). We detected 21 msHSBs longer than the maximum lengths expected from a random distribution of EBRs (Supplementary

Table 10 and 11), indicating that large msHSBs could be maintained in evolution of bird and other reptile genomes (Supplementary Table 10). Six amniote-, four sauropsid-, three archosaurian/testudines-, three archosaurian-, and five avian-msHSBs were significantly longer than would be expected from a random distribution of EBRs (Supplementary Table 10). To unravel the potential functional role of msHSBs in reptilian genomes we asked whether msHSBs were enriched in avian conserved non-coding elements (CNEs), many of which are gene regulatory sequences or miRNA (Zhang, et al. 2014), and chicken genes. All five msHSB sets were highly enriched (p-value  $<3e-12$ ) in avian CNEs, with a ratio between CNE base pairs in msHSBs and other genome intervals ranging from 1.45 for avian to 1.62 for archosaurian/testudines msHSBs (Table 1). The density of chicken genes in all msHSBs followed the opposite trend, with msHSBs having significantly fewer genes than other genome intervals (ranging from 0.58 for avian msHSBs to 0.74 for sauropsid and amniote ones; p-value  $<3e-12$ ; Table 1). To test if CNEs enrichment in msHSBs is not due to the reduction in the number of genes in msHSBs, we renamed all coding bases as additional CNE bases within the 91,947 windows in the chicken genome used to analyse the CNE density. We compared the original and obtained CNE densities in each window and found that the increment was very low with the average genome-wide ratio of the obtained to the real CNE bases of 1.02. We repeated this experiment for msHSB windows and non-msHSB windows separately and observed very similar values (1.02 for both). These values are much lower than the ratio of CNE bases in msHSBs compared to other genome intervals (Table 1), suggesting that the enrichment of CNEs in msHSBs detected is not due to the lack of genes in msHSBs. Overall, msHSBs in birds and other reptiles are gene-sparse but enriched for bird-specific non-randomly conserved DNA sequences (Table 1). Avian and reptile msHSBs lack coding genes but are enriched in CNEs, and at least the largest msHSBs are non-randomly

maintained in evolution. This likely reflects the existence of selection against chromosome rearrangements in some avian genome intervals.

**Signatures of gene-functional enrichment in msHSBs.** To identify if there are gene pathways associated with bird and/or reptile msHSBs we measured gene ontology (GO) enrichment in msHSBs. We analysed msHSBs >1.5Mbp in the chicken genome, covering from 8.03% to 18.12% of the genome in amniote and avian msHSBs, respectively and 10,830 genes with a single ortholog in human and chicken. We identified functional enrichment in all five sets of msHSBs (Figure 3 and Supplementary Table 12; FDR<10%).

The *development of primary sexual characteristics* term-related genes were significantly enriched in avian, archosaurian and archosaurian/testudines msHSB sets. Out of these 17 genes distributed across 12 chicken chromosomes, only one (*BMPRI1B*) was found in an avian-specific msHSBs but absent from the remaining msHSB sets. *BMPRI1B* plays a role in ovulation (Onagbesan, et al. 2003), and in formation of the bird three-digit limb (Welten, et al. 2005). A bird-specific CNE found 100bp upstream to *BMPRI1B* contains two transcription factor binding sites (TFBSs) for AP-1 (known as cJun) and NF-E4. The AP-1 transcription factor superfamily plays a role in the regulation of apoptosis during limb development in chickens (Suda, et al. 2014), and could account for the reported differences in expression of *BMPRI1B* in birds compared to other vertebrates (Brawand, et al. 2011). Therefore, the presence of this CNE containing a relevant TFBS could contribute to formation and stability of this msHSB in avian evolution.

*Appendage and limb development* genes (19 genes in 12 avian msHSBs on eight chicken chromosomes) were significantly enriched in the avian msHSB set only. Five genes were in avian-specific msHSBs (*SHOX*, *DLX5*, *DLX6*, *HOXA11*, and *BMPRI1B*). *DLX5* is under positive selection in birds (Zhang, et al. 2014) and mis-expression in chicken embryos leads to feather fusions and loss (Rouzankina, et al. 2004). In line with a previous study



(Lowe, et al. 2015) reporting CNEs near feather-related genes controlling the expression of these genes, we found a bird-specific CNE 1.9Kbp, containing a TFBS for TGGCA-binding proteins upstream of *DLX5*. The *HOXA11* gene is expressed during the proximodistal limb bud development leading to the formation of ulna and radius bones (Zeller, et al. 2009), and is under positive selection in birds (Zhang, et al. 2014). Overall, msHSBs are enriched for genes related to clade-specific phenotypes, suggesting a link between the formation of these genomic regions and clade-specific traits.

**Functional categories of genes in lineage-specific EBRs.** To evaluate potential associations between gene functional groups and lineage-specific EBRs, we performed GO enrichment analysis in EBRs from the 21 bird genomes. Only EBRs from genomes assembled with the aid of maps and those that passed our chimeric scaffold quality control were included in this analysis (Supplementary Table 5). We considered enriched GO terms those with genes in at least four EBRs per species to detect the terms affected by multiple chromosome rearrangements. Twenty-three categories were significantly enriched in EBRs in lineages leading to eight bird species (Table 2 and Supplementary Table 13).

The EBRs leading to budgerigar after the divergence from the ancestor of Passeriformes/parrots tended to reshuffle genes involved in *forebrain development*. Remarkably, the same term was also enriched in avian and archosaurian msHSBs, however, the gene pathways affected by EBRs and msHSBs were different (Figure 3 and 4). The msHSBs contained genes related to three of the five conserved canonical signalling pathways involved in forebrain development in vertebrates (Bertrand and Dahmane 2006; Rhinn, et al. 2006): the Hedgehog pathway (*SHH*, *Gli2* and *Gli3*), the WNT pathway (*WNT3A*, *beta-catenin* and *Lef-1*) and the FGF pathway (*FGF8* and *SOX2*) (Harrison-Uy and Pleasure 2012; Quinlan, et al. 2009) (Figure 4). Several studies demonstrated that *WNT3A* is expressed in

mouse dorsal telencephalon, but not in chicken (Hollyday, et al. 1995), possibly explaining the anatomical differences between the forebrain in these species (Robertshaw and Kiecker 2012; Shimogori, et al. 2004). In contrast, the budgerigar lineage-specific EBRs contained genes related to the *NOTCH1-NUMB* pathway (Figure 4) as well as *DRAXIN*. All three genes are involved in differentiation of neurones (Islam, et al. 2009; Wakamatsu, et al. 1999).

Although all vocal-learner bird species (songbirds, parrots and hummingbirds) have a ‘vocal brain nuclei’ in the forebrain, parrots, in addition, have a unique song-system compared to other vocal-learners (Chakraborty, et al. 2015; Jarvis 2004). To the best of our knowledge, this is the first report of distinct components of the same developmental network being found in the evolutionary stable and dynamic parts of animal genomes.

In summary, we demonstrated that genome synteny comparison represents a powerful tool to detect ancestral and lineage-specific genome-rearrangements, as well as evolutionary stable chromosomal intervals. Consistent with previous studies in mammals (Larkin, et al. 2009; Murphy, et al. 2005), chromosome breakage in reptiles and birds is not random but associated with genomic features including TEs and CNEs. We identified functional categories of genes enriched in conserved regions maintained from ancestral chromosomes or in some lineage-specific EBRs with genes related to ancestral- or lineage-specific biology. The most interesting result of EBR contribution to avian evolution (budgerigar) in our set was associated with the highest quality genome supported by additional mapping information. Therefore, the availability of more genomes supported by maps or assembled to a chromosome level will allow us to identify further genomic changes that contributed to the formation of existing species and clades.

## Methods

**Identification of SFs.** Alignments of 20 bird genomes and five outgroup genomes were performed against chicken genome using SatsumaSynteny (Grabherr, et al. 2010) (Supplementary Table 1). Syntenic fragments (SFs) were defined using three sets of parameters to detect genome rearrangements that are  $\geq 500\text{Kbp}$ ,  $\geq 300\text{Kbp}$  and  $\geq 100\text{Kbp}$  in the chicken genome with SyntenyTracker (Donthu, et al. 2009).

**Identification and classification of EBRs.** Breakpoint regions (BRs) were defined as the intervals delimited by two adjacent SF boundaries on the same reference chromosome. We developed a new multi-step approach to detect and classify EBRs from chromosome-level and fragmented assemblies. Briefly, we identified all potential BRs for every target genome pairwise comparison with the reference at each resolution in the reference genome coordinates. Then BRs from all pair-wise genome comparisons were cross-compared for reference genome coordinate overlaps. If a target genome was not assembled to chromosomal level, only BRs found within the scaffolds of the target assembly were classified as EBRs. We performed a phylogenetic classification of BRs using an *ad hoc* likelihood ratio approach, by calculating likelihoods for all possible classifications for each BR. The ratios of likelihoods were calculated for the first and second most likely classifications and were used as a quantitative basis for assigning BRs to phylogenetic branches, thereby qualifying them as EBRs, and distinguishing EBRs from so called *uncertain* BRs that could not be unambiguously assigned to a specific phylogenetic branch (see Supplementary data for more details).

To test the accuracy of our EBR classification approach we: a) compared the EBRs detected by our algorithm in the cattle genome to the previously published manually-defined cattle EBRs (Supplementary Table 2) and b) simulated a set of rearranged genomes with predefined phylogeny of EBRs (Supplementary Figure 2). We compared these EBRs and their

classification to the EBRs detected and classified by our algorithm from the same set of genomes (Supplementary Table 3 and 4). Since many of the assemblies used in this study were sequenced and assembled at scaffold level using NGS technologies, we developed a methodology to distinguish between putative assembly errors and lineage-specific EBR in NGS assemblies. First, we tested the EBR intervals by PCR using primers from the EBR-flanking DNA regions for three genomes with different sequencing coverage (63x, 85x and 105x). We calculated a minimum paired-read spanning coverage from the read libraries in all potential EBR intervals in the same genomes and correlated the levels of coverage to the rates of positive and negative PCR results to estimate the paired-read spanning level for each sequencing coverage that resulted in the minimum number of false positive and false negative EBRs (Supplementary Table 7 and 8). We applied these thresholds to other genomes with similar sequencing coverage (Supplementary Table 8).

To avoid possible underestimation of EBR numbers that would lead to detection of false regions of multispecies synteny we chose the highest (100Kbp) resolution to define msHSBs. The 500Kbp set was selected for gene enrichment analysis in EBRs to further minimize the effects of potential assembly errors in EBRs.

**Identification of msHSBs.** Multispecies HSBs were defined as the regions of reference chromosomes with no EBRs or *uncertain* BRs detected in our set of species. Five sets of msHSBs were defined: (i) avian msHSBs, including all birds, (ii) archosaurian msHSBs, including birds and crocodiles, (iii) archosaurian/testudines msHSBs, in birds, crocodiles, and turtles, (iv) sauropsida msHSBs, including all reptiles, and (v) amniote msHSBs, identified in all species studied. The distribution of msHSB sizes was tested for goodness-of-fit to an exponential distribution using the Kolmogorov-Smirnov test following previous publications (Larkin, et al. 2009; Pevzner and Tesler 2003) (Supplementary Table 9 and 10).

**Functional analysis of genes in EBRs and msHSBs.** Coordinates of all genes with a single known ortholog in the chicken and human genomes were downloaded from Ensembl (v.74). We focused on this set of genes because the follow-up analyses used functional annotation of genes generated mostly for mammalian genomes. To avoid genes that could be located in mis-assembled parts of both genomes or have erroneous definitions of orthology in Ensembl, we used the gene list to build chicken-human pairwise HSBs with SyntenyTracker using the gene coordinates. This allowed the detection of “singleton” and “out-of-place” genes located in unexpected positions within or between HSBs. These genes were removed from further analyses. We assigned the genes to EBRs or msHSBs following the previously published procedures (Larkin, et al. 2009). For the identification of GO terms overrepresented in msHSBs, we considered msHSBs >1.5Mbp in the chicken genome to avoid genes that could be located in proximity to EBRs. To evaluate gene functional enrichment in EBRs, we considered genes that were located within or  $\pm 300$ Kbp from EBR boundaries. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang, et al. 2008) to detect overrepresented GO terms in our datasets. We considered as significantly enriched terms with >2 fold-enrichment and false discovery rate (FDR) <10% in EBRs or msHSBs relative to all other regions on chicken chromosomes.

**Comparing densities of TEs in EBRs and other parts of bird genomes.** Lineage-specific EBRs identified in chicken genome coordinates were translated into the coordinates of target bird genomes using the correspondence between SF boundary coordinates in the chicken and target genomes. In the resulting EBR sets and chicken-specific EBRs we calculated the densities of TEs from major families and compared to those in other intervals of each target genome (RepeatMasker, RepBase v.18), as previously described (Elsik, et al. 2009; Groenen, et al. 2012; Larkin, et al. 2009).

**Density of bird-specific CNEs and genes in msHSBs.** Bird-specific conserved elements (Zhang, et al. 2014) defined in galGal3 coordinates were filtered to remove elements present in coding parts of chicken genes and all mRNA sequences mapped to the chicken genome, leaving only putative conserved non-coding elements (CNEs). Then, we used LiftOver (Kent, et al. 2003) to translate the CNE coordinates to galGal4 assembly to make the data compatible with our HSBs sets. We repeated filtering steps for the new genome coordinates obtained. The set of elements that was not overlapping with coding sequences after two filtering steps represented the bird CNEs in the chicken genome. Densities of CNEs and chicken genes (UCSC; all known gene set) were calculated in all msHSBs sets, and were compared to the rest of the reference genome using the previously published pipeline (Larkin, et al. 2009). After the GO enrichment analysis was performed, we screened the avian-specific CNEs nearby genes in the enriched categories for TFBSs using PROMO (Messeguer, et al. 2002) with a dissimilarity margin  $\leq 10\%$  with TFBSs found in chicken.

**Key words:** Chromosome rearrangements, birds, reptiles, genome evolution, comparative genomics.

## References

- Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE 2004. Hotspots of mammalian chromosomal evolution. *Genome Biology* 5: R23. doi: 10.1186/gb-2004-5-4-r23
- Bertrand N, Dahmane N 2006. Sonic hedgehog signaling in forebrain development and its interactions with pathways that modify its effects. *Trends Cell Biol* 16: 597-605. doi: 10.1016/j.tcb.2006.09.007
- Brawand D, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478: 343-348. doi: 10.1038/nature10532
- Burt DW, et al. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* 402: 411-413. doi: 10.1038/46555
- Chakraborty M, et al. 2015. Core and Shell Song Systems Unique to the Parrot Brain. *PLoS ONE* 10: e0118496. doi: 10.1371/journal.pone.0118496
- Donthu R, Lewin HA, Larkin DM 2009. SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC Res Notes* 2: 148. doi: 10.1186/1756-0500-2-148
- Ellegren H 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol Evol* 25: 283-291. doi: 10.1016/j.tree.2009.12.004



- Elsik CG, Tellam RL, Worley KC 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 324: 522-528. doi: 10.1126/science.1169588
- Grabherr MG, et al. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26: 1145-1151. doi: 10.1093/bioinformatics/btq102
- Griffin DK, Robertson LB, Tempest HG, Skinner BM 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* 117: 64-77. doi: 10.1159/000103166
- Groenen MA, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393-398. doi: 10.1038/nature11622
- Harrison-Uy SJ, Pleasure SJ 2012. Wnt signaling and forebrain development. *Cold Spring Harb Perspect Biol* 4: a008094. doi: 10.1101/cshperspect.a008094
- Hollyday M, McMahon JA, McMahon AP 1995. Wnt expression patterns in chick embryo nervous system. *Mech Dev* 52: 9-25.
- Huang DW, Sherman BT, Lempicki RA 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4: 44-57. doi: 10.1038/nprot.2008.211
- Huang Y, et al. 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet* 45: 776-783. doi: 10.1038/ng.2657
- ICGSC 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695-716. doi: 10.1038/nature03154
- Islam SM, et al. 2009. Draxin, a repulsive guidance protein for spinal cord and forebrain commissures. *Science* 323: 388-393. doi: 10.1126/science.1165187
- Jarvis ED 2004. Learned Birdsong and the Neurobiology of Human Language. *Annals of the New York Academy of Sciences* 1016: 749-777. doi: 10.1196/annals.1298.038
- Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346: 1320-1331. doi: Doi 10.1126/Science.1253451
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America* 100: 11484-11489. doi: 10.1073/pnas.1932072100
- Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921. doi: 10.1038/35057062
- Larkin DM, et al. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research* 19: 770-777. doi: 10.1101/gr.086546.108
- Lovell PV, et al. 2014. Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biology* 15: 565. doi: 10.1186/s13059-014-0565-1
- Lowe CB, Clarke JA, Baker AJ, Haussler D, Edwards SV 2015. Feather development genes and associated regulatory innovation predate the origin of Dinosauria. *Molecular Biology and Evolution* 32: 23-28. doi: 10.1093/molbev/msu309
- Messeguer X, et al. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18: 333-334.
- Murphy WJ, et al. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science (New York, N.Y.)* 309: 613-617. doi: 10.1126/science.1111387
- Onagbesan OM, et al. 2003. BMPs and BMPRs in chicken ovary and effects of BMP-4 and -7 on granulosa cell proliferation and progesterone production in vitro. *American Journal of Physiology - Endocrinology and Metabolism* 285: E973-E983. doi: 10.1152/ajpendo.00104.2003
- Pevzner P, Tesler G 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100: 7672-7677. doi: 10.1073/pnas.1330369100
- Quinlan R, Graf M, Mason I, Lumsden A, Kiecker C 2009. Complex and dynamic patterns of Wnt pathway gene expression in the developing chick forebrain. *Neural Dev* 4: 35. doi: 10.1186/1749-8104-4-35
- Rhinn M, Picker A, Brand M 2006. Global and local mechanisms of forebrain and midbrain patterning. *Curr Opin Neurobiol* 16: 5-12. doi: 10.1016/j.conb.2006.01.005
- Robertshaw E, Kiecker C 2012. Phylogenetic origins of brain organisers. *Scientifica (Cairo)* 2012: 475017. doi: 10.6064/2012/475017

- Romanov MN, et al. 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* 15: 1060. doi: 10.1186/1471-2164-15-1060
- Rouzankina I, Abate-Shen C, Niswander L 2004. *Dlx* genes integrate positive and negative signals during feather bud development. *Dev Biol* 265: 219-233.
- Ruiz-Herrera A, Farré M, Robinson TJ 2012. Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity* 108: 28-36. doi: 10.1038/hdy.2011.102; 10.1038/hdy.2011.102
- Schibler L, et al. 2006. High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics* 7. doi: 10.1186/1471-2164-7-194
- Shimogori T, Banuchi V, Ng HY, Strauss JB, Grove EA 2004. Embryonic signaling centers expressing BMP, WNT and FGF proteins interact to pattern the cerebral cortex. *Development* 131: 5639-5647. doi: 10.1242/dev.01428
- Suda N, et al. 2014. Dimeric combinations of MafB, cFos and cJun control the apoptosis-survival balance in limb morphogenesis. *Development* 141: 2885-2894. doi: 10.1242/dev.099150
- Ullastres A, Farré M, Capilla L, Ruiz-Herrera A 2014. Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. *BMC Genomics* 15: 530. doi: 10.1186/1471-2164-15-530
- Völker M, et al. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Research* 20: 503-511. doi: 10.1101/gr.103663.109
- Wakamatsu Y, Maynard TM, Jones SU, Weston JA 1999. NUMB localizes in the basal cortex of mitotic avian neuroepithelial cells and modulates neuronal differentiation by binding to NOTCH-1. *Neuron* 23: 71-81.
- Warren WC, et al. 2010. The genome of a songbird. *Nature* 464: 757-762. doi: 10.1038/nature08819
- Welten MCM, Verbeek FJ, Meijer AH, Richardson MK 2005. Gene expression and digit homology in the chicken embryo wing. *Evol Dev* 7: 18-28. doi: 10.1111/j.1525-142X.2005.05003.x
- Zeller R, López-Ríos J, Zuniga A 2009. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat Rev Genet* 10: 845-858. doi: 10.1038/nrg2681
- Zhang G, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346: 1311-1320. doi: 10.1126/science.1251385

## Acknowledgments

We thank T. Faraut and A. Vignal for an early access to the duck chromosome assembly. This work was supported in part by the Biotechnology and Biological Sciences Research Council [BB/K008226/1 and BB/J010170/1 to D.M.L, and BB/K008161/1 to D.K.G] and by PL-Grid Infrastructure [D.M.L].

## Author contributions

DML and MF designed and performed research; JN, GTS, DML, and MF designed and tested new algorithms; JD performed PCR experiments; MF, DML, GTS, DWB, EDJ, and DKG wrote the paper; LA, EDJ, and CL contributed new reagents/analytic tools. All authors read



and approved the manuscript for publication.

## Competing financial interests

The author(s) declare no competing financial interests.

## Figure legends

**Fig. 1. Evolutionary breakpoint regions (EBRs), syntenic fragments (SFs) and homologous syteny blocks (HSBs).** Blue and red blocks define SFs in target genomes in “+” and “-“ orientation, respectively compared to the chicken chromosome 5 defined at 100Kbp resolution, with target species scaffold or chromosome numbers indicated inside the blocks. Only the columns with genomes assembled to chromosomes (turkey, duck, zebra finch, Anole lizard, and opossum) contain complete HSBs while blocks in the remaining columns represent either HSBs or SFs. EBRs are defined as white intervals in between either two adjacent SFs originating from the same scaffold in a target genome or two adjacent HSBs. Reference-specific EBRs are represented by the white intervals that overlap in all species. The arrowheads point to a chicken-specific and a Galloanserae-specific EBRs. Pale grey boxes demarcate avian msHSBs that are > 1.5Mbp in the chicken genome. Asterisks demark genomes with modified scaffold IDs for better visibility. All reference chromosome and target genome alignments are available from the avian Evolution Highway website: <http://eh-demo.ncsa.uiuc.edu/birds>.

**Fig. 2. Relationship between lineage-specific evolutionary breakpoint regions (EBRs) and transposable elements (TEs) in avian species.** The phylogenetic tree is based on (Jarvis, et al. 2014). Red bars indicate a significant enrichment of TEs from one or more abundant avian TE families (LINE-CR1, LTR-ERVL, LTR-ERVK and LTR-ERV1) in lineage-specific

EBRs ( $p\text{-value} < 0.05$ ;  $FDR < 10\%$ ); green bars show significant negative associations of TEs with lineage-specific EBRs ( $p\text{-value} < 0.05$ ;  $FDR < 10\%$ ); and grey bars indicate elevated numbers of the TE families in lineage EBRs (higher number of TEs in EBRs compared to the rest of the genome but not reaching a significance level of  $p\text{-value} < 0.05$  and  $FDR < 10\%$  likely due to a low number of lineage-specific EBRs resulting in a low power of the statistical test).

**Fig. 3. Gene Ontology (GO) terms enriched in four sets of msHSBs.** Green boxes show a fold enrichment  $> 1.3$  while red boxes depict a fold enrichment  $> 2$ . White crosses inside boxes show categories with  $FDR < 10\%$ . Underlying data could be found in Supplementary Table 12.

**Fig. 4. Gene pathways related to forebrain development in budgerigar lineage-specific EBRs and avian and archosaurian msHSBs.** Budgerigar lineage-specific EBRs (top box) are enriched for genes related to the NOTCH1-NUMB pathway, while avian and archosaurian msHSBs (bottom box) for genes related to three conserved canonical pathways (SHH pathway in blue, WNT3 pathway in pink and FGF8 pathway in purple). The function of each protein is indicated in the legend by different shapes and colours. Red lines connecting two proteins indicate inhibition, while blue lines show activation. The green circular shade represents the cell membrane, while the orange circular shade demarcates the nuclear envelope. The image was modified from Metacore version 6.22 build 67265 and Bertrand et al. 2006 (Bertrand and Dahmane 2006).

## Tables

**Table 1. Density per 10Kbp window of conserved non-coding elements and genes in msHSBs and other genome intervals.**

msHSB set	Genes*						CNEs*					
	All msHSBs			msHSBs>1.5Mbp			All msHSBs			msHSBs>1.5Mbp		
	msHSBs	Other	Ratio	msHSBs	Other	Ratio	msHSBs	Other	Ratio	msHSBs	Other	Ratio
Avian	0.14	0.24	0.58	0.10	0.17	0.59	2.20	1.52	1.45	2.44	1.96	1.25
Archosaurian	0.14	0.20	0.70	0.10	0.17	0.59	2.33	1.47	1.58	2.58	1.96	1.32
Archosaurian/ Testudines	0.14	0.20	0.70	0.11	0.17	0.65	2.35	1.45	1.62	2.49	1.98	1.26
Sauropsid	0.14	0.19	0.74	0.12	0.17	0.71	2.45	1.55	1.58	2.60	1.99	1.31
Amniote	0.14	0.19	0.74	0.12	0.17	0.71	2.44	1.58	1.54	2.36	2.01	1.17

\* All differences are statistically significant (raw p-values<0.0000000001).

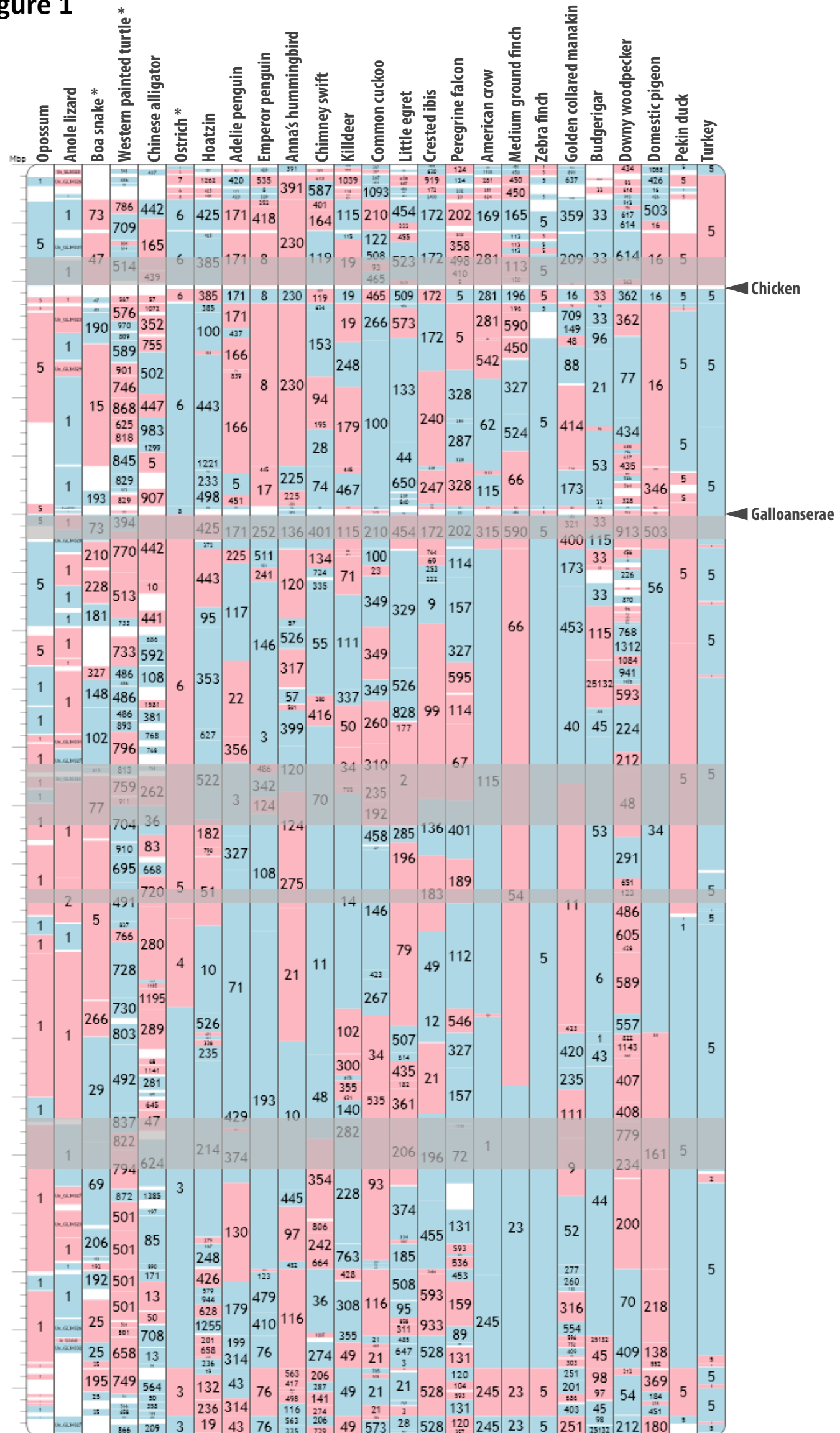
**Table 2. Gene Ontology terms enriched in lineage-specific EBRs\*.**

<b>EBR classification</b>	<b>GO term</b>	<b>No. genes</b>	<b>No. EBRs</b>	<b>Fold-enrichment</b>	<b>FDR (%)</b>
Budgerigar	Forebrain development	12	11	2.74	5.47
	Neuron differentiation	15	13	2.33	6.83
	Neuron development	12	11	2.62	8.19
	Response to wounding	11	11	2.77	8.35
Common cuckoo	Mitotic cell cycle	11	11	3.57	1.14
	Condensed chromosome	7	5	4.88	2.67
	M phase	10	9	3.25	4.50
Little egret	Passive transmembrane transport	10	5	4.15	0.59
	Cation channel activity	7	4	4.32	5.61
Anna's hummingbird	Hexose metabolic process	10	8	2.90	9.70
Peregrine falcon	RNA degradation	6	6	6.13	2.29
	Soluble fraction	5	4	6.23	8.35
Downy woodpecker	Histidine metabolism	6	5	10.30	0.16

\*An extended version of this table, including the gene names in each GO term is the Supplementary Table 13.

Figure 1

GGA 5



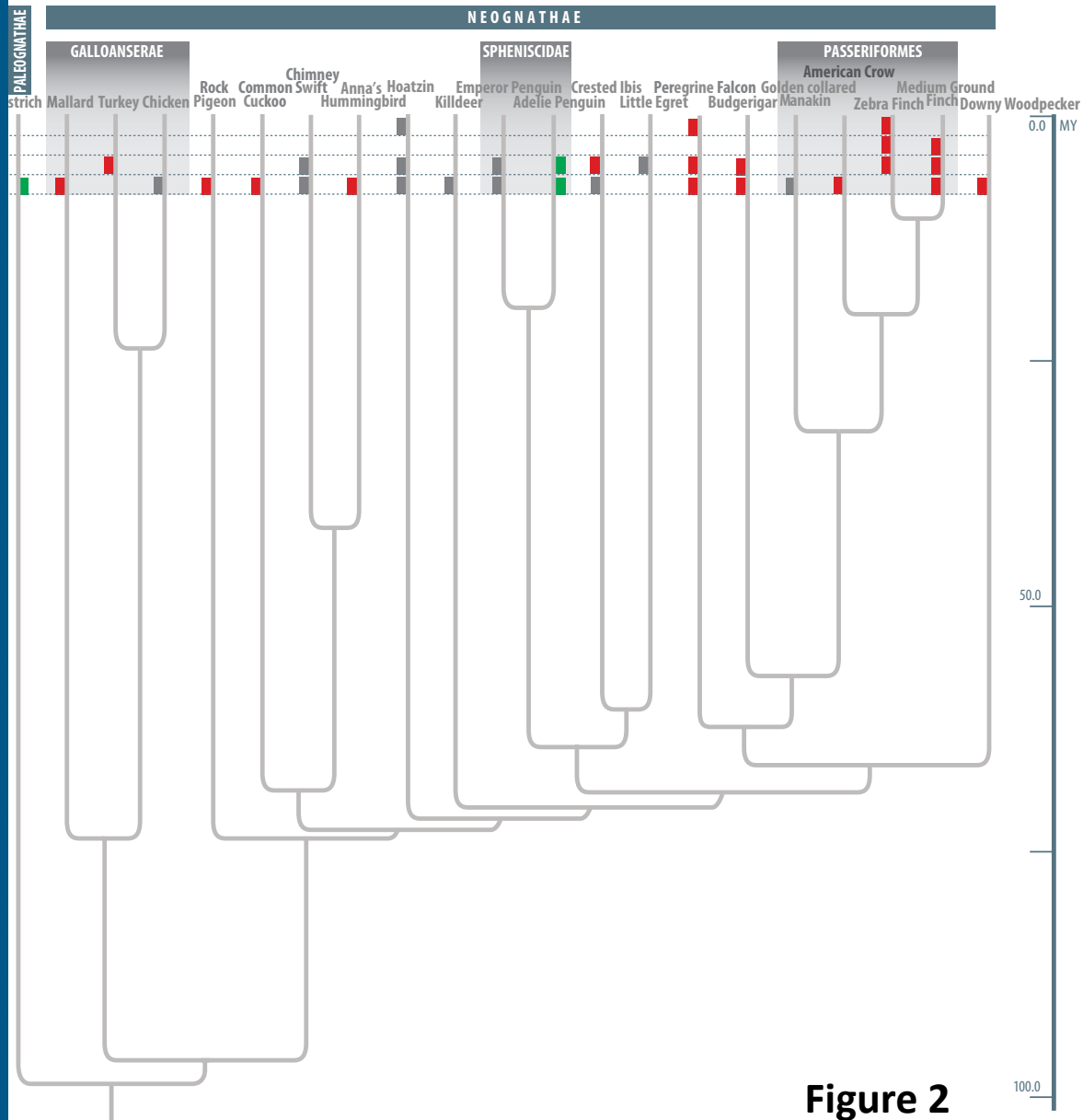


Figure 2

	AVIAN	ARCHOSAURIAN	ARCHOSAURIAN/ TESTUDINES	SAUROPSID
esis	*			
rocyte differentiation	*			
rocyte homeostasis	*		*	
g of blood vessels	*			*
ation of blood vessel size	*			
d development	*	*		
d cell differentiation	*		*	
ctive developmental process	*	*	*	
opment of primary sexual characteristics	*	*	*	*
egulation of biosynthetic process	*		*	
ive regulation of macromolecule biosynthetic process	*		*	
etal protein binding	*	*		
egulation of gene expression	*	*	*	
cription factor complex	*	*		
ive regulation of transcription	*	*	*	
ive regulation of transcription, DNA-dependent	*	*	*	
ation of transcription from RNA polymerase II promoter	*	*	*	
egulation of nitrogen compound metabolic process	*		*	
pecification process	*			
ic morphogenesis	*	*		
ization	*	*		
ge/Limb development	*			
umen	*			
ding	*			
velopment in camera-type eye	*	*		
de binding	*			
enesis		*		*
ogenesis of dentine-containing tooth		*		*

**Figure 3**

Figure 4

