










ARTICLE

DOI: 10.1038/s41467-018-04267-9

OPEN

Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs

Rebecca E. O'Connor ¹, Michael N. Romanov ¹, Lucas G. Kiazim ¹, Paul M. Barrett ², Marta Farré ³, Joana Damas ³, Malcolm Ferguson-Smith⁴, Nicole Valenzuela ⁵, Denis M. Larkin ³ & Darren K. Griffin ¹

Genomic organisation of extinct lineages can be inferred from extant chromosome-level genome assemblies. Here, we apply bioinformatic and molecular cytogenetic approaches to determine the genomic structure of the diapsid common ancestor. We then infer the events that likely occurred along this lineage from theropod dinosaurs through to modern birds. Our results suggest that most elements of a typical 'avian-like' karyotype (40 chromosome pairs, including 30 microchromosomes) were in place before the divergence of turtles from birds ~255 mya. This genome organisation therefore predates the emergence of early dinosaurs and pterosaurs and the evolution of flight. Remaining largely unchanged interchromosomally through the dinosaur-theropod route that led to modern birds, intrachromosomal changes nonetheless reveal evolutionary breakpoint regions enriched for genes with ontology terms related to chromatin organisation and transcription. This genomic structure therefore appears highly stable yet contributes to a large degree of phenotypic diversity, as well as underpinning adaptive responses to major environmental disruptions via intrachromosomal repatterning.

¹School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK. ²Department of Earth Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. ³Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London NW1 0TU, UK. ⁴Department of Veterinary Medicine, Cambridge University, Cambridge CB3 0ES, UK. ⁵Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Iowa, IA 50011, USA. These authors contributed equally: Denis M. Larkin, Darren K. Griffin. These authors jointly supervised this work: Denis M. Larkin, Darren K. Griffin. Correspondence and requests for materials should be addressed to D.K.G. (email: d.k.griffin@kent.ac.uk)

In the absence of cellular material and DNA from biological samples of long-extinct, early diverging lineages, data from genome sequence assemblies of extant species can nonetheless facilitate the reconstruction of gross genome structures (karyotypes). This can be achieved provided those assemblies are at, or close to, chromosome level, i.e. one scaffold per chromosome¹. In a previous study, we analysed (close to) chromosome-level assemblies from six extant birds (and a lizard outgroup) to determine the most likely karyotype of the neornithine ancestor for the macrochromosomes and the neognathe ancestor for the microchromosomes². Recreating the most parsimonious sequence of events that might have led to contemporary genome structures (karyotypes), we determined that chicken (*Gallus gallus*) was the closest karyotypically to the reconstructed ancestral pattern, with zebra finch (*Taeniopygia guttata*) and budgerigar (*Melopsittacus undulatus*) undergoing the most intra- and interchromosomal rearrangements, respectively. In the current study, to reconstruct the most likely karyotype of the diapsid common ancestor (DCA >255 mya), we applied similar approaches, i.e. the multiple-genome rearrangement and analysis (MGRA2) tool. We focussed on the best-quality chromosome-level assemblies of avian and reptilian genomes and a mammalian outgroup. Supplementing bioinformatic data with novel molecular cytogenetic approaches on turtle metaphases, we tested the hypothesis that the typical karyotype seen in neornithine birds underwent few interchromosomal rearrangements since the divergence of turtles from archosaurs (birds and crocodylians) <255 mya. Combining both sets of data, we thence inferred the most parsimonious sequence of events that occurred from the diapsid ancestor, to the archosaur ancestor³, and thence via non-avian theropod dinosaurs to extant birds (see Supplementary Note 1 for divergence times).

Studies of the best-assembled genomes (including chromosome level) also indicate that evolutionary breakpoint regions (EBRs) lie in gene-dense loci, enriched with genes related to lineage-specific biology, transposable elements and other repetitive sequences^{4–7}, while sequences that stay together during evolution (homologous synteny blocks, HSBs) are enriched in developmental genes and regulatory elements⁷. While a contribution of random breakage in chromosome evolution⁹ cannot be excluded for all (especially smaller) rearrangements that might have neutral effects on phenotypes, multiple evidence has accumulated to suggest that at least the largest HSBs and some EBRs in animal genomes are maintained non-randomly^{4,5,10,11}. Differences in the composition of their DNA features suggest that, although chromosome aberrations in germ cells may indeed occur in regions more prone to breakage (e.g. recombination hotspots or open chromatin areas), those breaks not disturbing essential genes or providing a selective advantage will more likely be fixed in populations, becoming EBRs⁴. In the current study, we investigated gene content of those EBRs and HSBs identified as being involved in the karyotypic changes from the diapsid ancestor to modern birds and identified genes that may indicate adaptive (EBRs) or conserved (HSBs) phenotypic features or those likely to be involved in gross genomic rearrangement.

Here, we analysed data from the genome assemblies of Carolina anole lizard (*Anolis carolinensis*)⁶, chicken (*G. gallus*)^{7,12}, mallard (*Anas platyrhynchos*)⁸, zebra finch (*T. guttata*)¹³, and grey short-tailed opossum (*Monodelphis domestica*)¹⁴. All of these species have robust chromosome-level assembled genomes. Moreover, *M. domestica* has a karyotypic structure thought to resemble the mammalian ancestor most closely¹⁴. Among other genome assemblies that might have proved useful in our analyses, those generated from alligators and turtles^{15,16} were discounted as they are too fragmented, i.e. not close to chromosome level.

Also, near-chromosome-level assemblies for turkey, budgerigar and ostrich were ultimately excluded because our cytogenetic studies in these species (not shown) revealed that the level of fragmentation or misassemblies in these genomes had the potential to introduce false breakpoint regions (Supplementary Note 2). Finally, we discounted crocodylians from our molecular analysis, partly because of a relative lack of fluorescence in situ hybridisation (FISH) success of multiple attempts, and partly because all crocodylian species studied have an atypical archosaur karyotype with no microchromosomes, mostly brought about by fusion¹⁷.

Our results suggest that most features of a typical ‘avian-like’ karyotype were in place before the divergence of birds and turtles, that the predominant mechanism of change thereafter was intrachromosomal rearrangement and that EBRs were enriched for GO terms associated with chromatin modification and chromosome organisation.

Results

Summary of results. Pairwise sequence alignments permitted visualisation of 397 multispecies HSBs, their orientation in each genome and EBRs between them (listed in Supplementary Data 2). Using MGRA2, we generated 19 contiguous ancestral regions (CARs), roughly correlating to chromosomes, in the most likely ancestral karyotype of the DCA. CAR sizes are given in Supplementary Data 1 and the analysis pipeline is described in Methods. Our analysis of chromosomes from the turtle genome with one of the largest chelonian diploid numbers ($2n$) (spiny soft-shelled turtle (*Apalone spinifera*) ($2n = 66$)) revealed little or no evidence of interchromosomal differences between this species and chicken. That is, chicken-derived fluorescent probes highly selected to hybridise across large evolutionary distances⁵ plus chromosome painting data provided evidence that most chicken chromosomes 1–28 + Z are each represented by a single-turtle counterpart (see Methods and Supplementary Data 3). Successful hybridisation to the chromosomes of red-eared slider (*Trachemys scripta*) ($2n = 50$) revealed a karyotype with microchromosomal homeologues either having fused to macrochromosomes or, more likely, having retained the ancestral state of the DCA. Indeed, one of the main technical advances made in this study was the isolation of a probe set that would hybridise directly across species that diverged ~255 mya. More limited success in hybridising to *A. carolinensis* metaphases ($2n = 36$) (Fig. 1) revealed some broad similarities to the DCA established by the bioinformatic approach (see Supplementary Data 3).

FISH analysis suggests avian and turtle chromosomes are precise counterparts. Cross-species hybridisation (zoo-FISH) of chicken bacterial artificial chromosome (BAC) probes previously designed to work in multiple avian species and located subtelomerically⁵, as well as chicken chromosome paints were successfully hybridised to the chromosomes of *A. carolinensis* (Carolina anole lizard), as well as two turtles *T. scripta* and *A. spinifera*. Results, for the most part, provide evidence that, for avian chromosomes 1–28 + Z (with rare exceptions, e.g. 16 for which we could not generate data), each chicken chromosome is syntenic to the turtle with the largest diploid number, *A. spinifera* ($2n = 66$). That is, we found little or no evidence of interchromosomal rearrangement, with the exception of rare events, e.g. fusions of chromosome 4 in chicken and chromosome 22 in turtles. For the macrochromosomes and some pools of microchromosomes, chromosome paints¹⁸ produced signals cross-species and, for the microchromosomes, selected BAC probes⁵ provided strong BAC signals (note, macrochromosomal BACs,

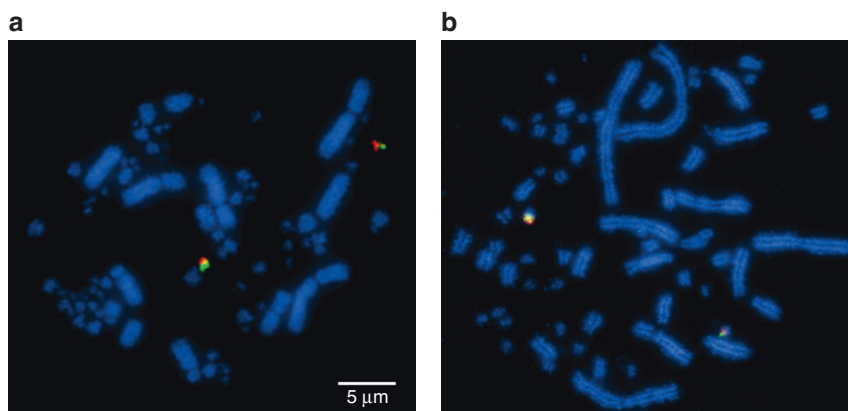


Fig. 1 Cross-species hybridisation (zoo-FISH) results. Hybridisation of chicken chromosome 23 sub-telomeric BACs to **a** chicken ($2n = 78$) and **b** turtle (*Apalone spinifera* $2n = 66$) metaphases. This is an example of how most chromosomes studied in birds and turtles (species examined with the highest diploid number) are precise counterparts of one another. All syntenic chromosomes were of similar sizes and morphologies. All chromosomes are labelled in blue (DAPI) with BAC probes labelled in red (Texas Red) and green (FITC), respectively (where signals overlap a little, a yellow/orange colour is seen). Successful strong hybridisation across large evolutionary distances was one of the technical advances of this project. Scale bar applies to both images

despite working well on other avian species, did not produce many successful hybridisations on non-avian reptiles). Specifically, of the microchromosomal BACs, 29 of the original 36 (81%) worked successfully (chromosomes 10–15 and 17–28; 16 not included due to lack of sequence coverage and chromosomes smaller than 28 do not have sequences assigned to them) in both turtle species (all but chromosome 20 had at least one BAC signal) and 17 of 36 (47%) worked in Carolina anole lizard (chromosomes 14, 18, 20, 25, 27 and 28 were not represented). Of those that worked on all species, results revealed that the orthologues of chicken chromosomes 12 and 13 were fused and chromosome 26 attached to chromosome 4 in the red-eared slider and Carolina anole but represented as separate microchromosomes in the spiny soft-shelled turtle (*A. spinifera*). The chromosome 22 orthologue appeared as a separate chromosome in the Carolina anole, but as fused to the centre of a macrochromosome in the two turtle species. Eight avian microchromosome orthologues for chromosomes 10, 11, 15, 17, 19, 21, 23 and 24, appeared to be conserved as single microchromosomes in all three reptiles studied, and in red-eared slider (*T. scripta*), chromosomes 14, 18, 25 and 28 were also represented as single microchromosomes, with chromosomes 12, 13 and 26 also as single microchromosomes for spiny soft-shelled turtle. All macrochromosomal assignments confirmed those previously reported for the lizard and for the two turtles. In order to establish that there were no rearrangements between microchromosomes of birds and reptiles, working BACs as close to the telomere as possible were used, and all microchromosomal sizes were measured in comparison to known macrochromosomal size by ImageJ. We cannot preclude the possibility that some interchromosomal events were not detected by our approach; for instance, cryptic translocations could possibly give chromosome paint signals too weak to detect microscopically or be sub-telomeric to the BAC used. However the fact that no additional rearrangements other than those already known were detected between *A. spinifera* ($2n = 66$) and avian ($2n = 80$) karyotypes supports our central assertion of identity by descent, in place ~255 mya, with ~7 fissions (presumably at least some, but possibly all) occurring before the emergence of the dinosaurs.

Sequence alignments, multispecies HSBs and EBRs. Pairwise bioinformatic alignments of the chicken, mallard, zebra finch, Carolina anole and grey short-tailed opossum genomes using the

Evolution Highway chromosome browser allowed for the visualisation of multispecies HSBs (msHSBs) and their orientation in each genome as well as the identification of EBRs between these msHSBs. These alignments were screened for blocks shared in the five genomes compared, and a total of 397 msHSBs were found. These were distributed across 19 of the 28 sequenced chicken chromosomes available on Evolution Highway: i.e. chromosomes 1–9, 11–13, 15, 18, 20, 24, 26, 27 and Z. The 397 msHSBs were also dispersed on 19 duck chromosomes, 21 zebra finch chromosomes, 10 anole chromosomes, and 8 opossum chromosomes. If we compare the total size of 397 msHSBs relative to the chicken 28 chromosomes available on Evolution Highway, they represented 49% of the total genome length; if we compare these msHSBs to the size of the above 19 chicken chromosomes, they represented about 53% of their combined length.

Using the MGRA2 algorithm, we produced a series of CARs representing the most likely ancestral karyotype of the diapsid common ancestor (DCA). While we cannot be entirely sure that each CAR represents a whole chromosome as MGRA2 will inherently ‘break’ the chromosome if it cannot find synteny, a total of 19 diapsid ancestral CARs were found, probably representing fewer chromosomes. The number of msHSBs per CAR varied between 2 and 59. A total of 17 chicken chromosomes were aligned to these CARs (chromosomes 1–8, 11–13, 15, 18, 24, 26, 27 and Z) meaning that some microchromosomes were not represented.

Chromosome inversions from DCA to chicken. Reconstructed CARs derived from MGRA2 were subsequently mapped to the extant genomes. The rearrangements between the DCA and chicken were then modelled using maximum parsimony. A total of 49 inversions were identified between the DCA and the chicken genome along with 10 interchromosomal changes (see Supplementary Note 3). Of the interchromosomal rearrangements found, a translocation was identified between orthologues of chicken chromosomes 5 and 20, consistent with the FISH results (Fig. 2, Supplementary Fig. 1). Comparison of the data generated for the DCA and a putative archelosaur common ancestor revealed that the majority of interchromosomal rearrangements occurred to form this basic structure with most intrachromosomal rearrangements (inversions) after (for simplicity, on Fig. 2, all intrachromosomal changes are shown after formation of the basic (archosauromorph common ancestor) pattern) although we cannot rule out the

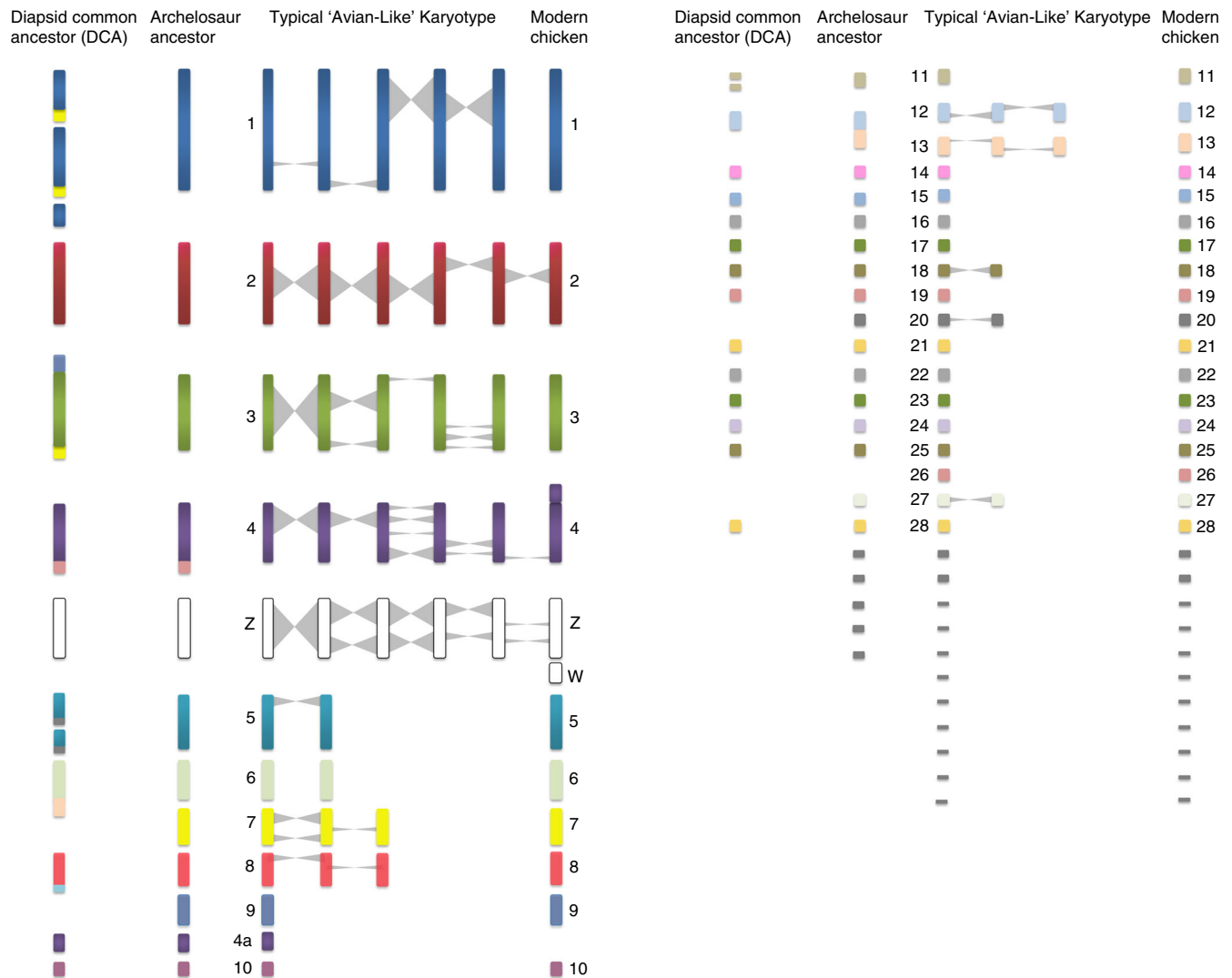


Fig. 2 Overall inferred karyotypic changes from the diapsid common ancestor (DCA), through the archosauromorph ancestor, modern chicken ($2n = 78$) arising out of theropod dinosaurs, collating all available lines of evidence, bioinformatic and molecular cytogenetic. The left side shows the 19 CARs including the ancestral microchromosomes, which gives an impression of a higher diploid number greater than 46 (total chromosome number predicted) as we cannot necessarily determine the nature of all the fusions. For simplicity, the intrachromosomal rearrangements (inversions) are all depicted after the archelosaur ancestor, however, a small proportion may have occurred before. The colour scheme is randomly assigned with each chromosome for which we have data given a different colour. Chromosomes for which we do not have FISH data are depicted in grey (e.g. chromosomes 16, 20, 22 and the very smallest microchromosomes). Diapsid common ancestor (DCA): 255 mya; likely $2n = 36\text{--}46$; 19 CARs identified (some CARs likely fused as single chromosomes, hence this diagram appears as apparently more chromosomes; some chromosomes not covered by sequence assembly; Chromosome 7 orthologue (yellow) attached to three chromosomes, chromosome 26 (dark pink) to the orthologue of chromosome 4). Archelosaur ancestor: <255 mya; likely $2n = 66$; most chromosomes syntenic to modern birds; chromosomes without direct evidence (including chromosome 16, 20, 22) depicted in grey. Typical 'Avian-Like' karyotype: likely $2n = 80$ (numbering according to chicken genome); -7 fissions; some inversions may have occurred before the pattern was established interchromosomally, but intrachromosomal changes shown separately for clarity. Modern chicken: $2n = 78$; known fusion of chromosome 4 shown (not present in most birds); sex chromosome evolution post Palaeognathae-Neognathae divergence

possibility that some occurred before. Between the DCA and the archosauromorph common ancestor, a fusion most likely occurred to form chromosome 1 and translocations/fissions occurred between avian ancestral CARs that became chromosome 7). Most rearrangements between these two ancestors were nonetheless intrachromosomal with a total of 49 inversions that appear to have occurred between the DCA and the extant chicken genome.

Enrichment of gene ontology terms in msHSBs. Within the 397 multispecies HSBs (msHSBs), significant enrichments were observed for gene ontology (GO) terms relevant to amino acid

transmembrane transport (symport; Supplementary Data 4) and signalling (group enrichment scores, 2.44–2.50 as produced using DAVID Functional Annotation Clustering tool; single GO terms as produced using DAVID Functional Annotation Chart, $p < 0.05$, false-discovery rate (FDR) < 5%). Other msHSB-specific GO term enrichments were related to synapse/neurotransmitter transport, nucleoside metabolism and use, cell morphogenesis and cytoskeleton and sensory organ development (group enrichment scores, 1.72–3.02—Supplementary Data 4). For the functional analysis of EBRs, we produced a set of 234 EBRs that were intrinsic to the archosauromorph common ancestor. Within these, we identified significant enrichments in genes and single GO terms relevant to

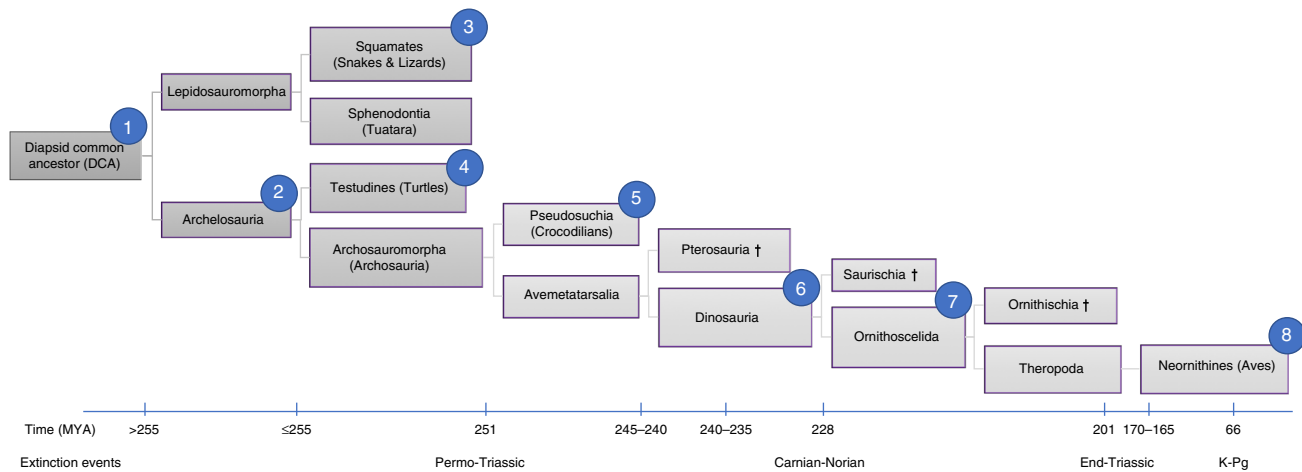


Fig. 3 Representative phylogenetic tree illustrating the lineage investigated in this analysis. For ease of reading, the timelines are not to scale. The major extinction events and the principal findings of this study are highlighted in the context of this phylogeny as follows. (1) DCA karyotype reconstructed by MGRA2 analysis. (2) Cyto-genetic analysis revealed the basic 'avian' pattern $2n = 66$ mostly in place before archelosauria divergence. (3) One lizard (*A. carolinensis*) studied genomically and cytogenetically. (4) Two turtles (*T. scripta* and *A. spinifera*) studied cytogenetically. (5) Crocodilian genomes not considered suitable for analysis because of fragmented genome assemblies and fused chromosomes. (6) Early emerging dinosaurs and pterosaurs probably had at least $2n = 66$ and up to $2n = 80$ with typical avian pattern. (7) Theropod dinosaurs and possibly other groups probably had close to $2n = 80$ with typical avian pattern. (8) Three avian species $2n = 78-80$ studied cytogenetically and using chromosomal-level genome assemblies. † extinct lineage

chromatin modification and chromosome organisation, as well as proteasome/signalosome structure ($p < 0.01$, FDR $< 5\%$, Supplementary Data 5). In particular, the first annotation cluster (enrichment score, 4.30) consisted of six genes related to proteasome/signalosome structure within EBRs located on six chicken chromosomes. The second annotation cluster (enrichment score, 1.64) included 15 genes relevant to chromatin modification within 12 EBRs on seven chicken chromosomes, these genes and their functions are listed in Supplementary Data 5.

Discussion

Combining all lines of available evidence (bioinformatic and molecular cytogenetic from this study along with previous findings), a picture emerges of an inferred DCA karyotype with a chromosome number of $2n = 36-46$, consistent with that previously proposed¹⁹ and with most other non-avian, non-testudinate reptiles. Roughly half would have been macro- and half microchromosomes^{6,19}. Alföldi et al.⁶ found direct synteny at the sequence level between the microchromosomes of *A. carolinensis* and *G. gallus*, with all but one lizard microchromosome corresponding to a single chicken microchromosome. Given that the *A. carolinensis* genome has 12 microchromosomal pairs (that are mostly syntenic to chicken chromosomes⁶) compared to the 28–30 seen in most birds, the most likely explanation is that these were present in the DCA, the remainder evolving thereafter by fission⁶ to at least $2n = 66$. Our current data suggest that interchromosomal rearrangement largely ceased thereafter, with the exception of ~ 7 fissions that explain the difference between the pattern in *A. spinifera* ($2n = 66$) and most modern birds ($2n \sim 80$). Indeed, even the karyotype of *T. scripta* ($2n = 50$) has broad similarities to the avian pattern, but with more microchromosomal homeologues attached to larger chromosomes. This would either indicate that *T. scripta* has a karyotype that represents an earlier stage of differentiation to the "bird-like" turtle pattern, or that it subsequently underwent a series of fusions (such as in the crocodilians); the former being the most likely

since it requires fewer events. The considerable range of diploid numbers in turtles ($2n = 26$ to $2n = 68$), with most being more 'avian-like' than their other reptilian counterparts, suggests that further study of this group will provide greater insight into the sequence of events that led to the establishment of a highly successful mode of genome organisation. We cannot preclude the possibility that some interchromosomal rearrangements were beyond the sensitivity of our detection, e.g. the weak chromosome painting signals may not have detected subtle changes, or cryptic chromosome translocations may have occurred that were telomeric to our fluorescent probes. This does not, however, detract from our assertions that a broad overall pattern of genome organisation was in place at least 255 mya in the archelosaur ancestor and changed little in the majority of living species. Moreover, our findings are consistent with previous studies using chicken macrochromosome paints on Chinese soft-shelled turtle (*Pelodiscus sinensis*) chromosomes ($2n = 66$)²⁰, *T. scripta*²¹ and painted turtle (*Chrysemys picta*) chromosomes (both $2n = 50$)²². Thus, similar studies on more turtles may reveal species with greater diploid numbers and a pattern resembling that of birds even more (Figs. 2 and 3).

Determining the precise timing of events that led from the archelosaur ancestor ($2n > 66$) to modern birds ($2n \sim 80$) is beyond the resolution of this study (Fig. 3). A similar rate of fission continuing beyond 255 mya, would have established a near-complete neornithine-like karyotype by ~ 240 mya, roughly coinciding with the emergence of the earliest dinosaurs and pterosaurs²³. Equally, a dramatic slowdown or halt in fission events at ~ 255 mya would suggest that the forebears of the earliest dinosaurs and pterosaurs had a pattern more similar to that of *A. spinifera*. In both scenarios (or an intermediate), however, a pattern very similar to that of most birds would have been present. Burt²⁴ proposed that most avian microchromosomes were present in the avian common ancestor > 80 mya²⁵, arguing that it probably had the small genome size characteristic of birds and a karyotype of around $2n = 60$. Our results, however, suggest a much earlier emergence of the typical avian pattern with genome size reduction occurring later in

archosaur or theropod evolution²⁴. Indeed Uno et al.²⁶ proposed that the archosauromorph ancestor may have had microchromosomes similar to the turtle but did not specify their nature.

Our results suggest that, aside from ~7 fissions, the primary mechanism for chromosomal rearrangement in the avian stem lineage after 255 mya was intrachromosomal (e.g. inversions). Reconstructed CARs, when compared to extant genomes, resulted in the identification of rearrangements between the DCA and chicken (*G. gallus*) genomes, modelling 49 inversion events. This, however, is almost certainly an underestimate due to the paucity of sequence coverage in some areas, including the smallest avian microchromosomes. Rates of change are not easy to determine, but there is some evidence of intrachromosomal change speeding up in modern birds, even in the chicken, which is thought to be very similar chromosomally to the avian common ancestor²⁷. Increased intrachromosomal change has been reported in specific groups, with several studies suggesting that the greatest rates would be found in passerines (compared to other birds)^{1,4,28}, the group containing most avian species. It is perhaps reasonable to speculate therefore that bursts of speciation may have also been accompanied by increased rates of intrachromosomal rearrangement in other non-avian dinosaur lineages.

Within the multispecies HSBs, we identified significant enrichments for GO terms relevant to transmembrane transport (symptom) and signalling, synapse/neurotransmitter transport, nucleoside metabolism and use, cell morphogenesis and cytoskeleton and sensory organ development (Supplementary Data 4). HSBs are often enriched for GO terms related to phenotypic features that remain constant²⁹ and the results presented here are consistent with this hypothesis. Sankoff³⁰ however stated that EBRs are where the 'action' in genome evolution lies and, previously, we found that GO terms in avian EBRs associated with specific adaptive features, e.g. enrichment for forebrain development in the budgerigar EBRs (consistent with vocal-learning)⁴. In the current study, we identified significant enrichments in genes and single GO terms relevant to chromatin modification, chromosome organisation and proteasome/signalosome structure in EBRs (Supplementary Data 5). This illustrates some parallels to recent findings in rodents where chromosomal changes were associated with open chromatin²⁷. Most of the 15 genes found in this GO term were related to control of gene expression. Transcription factors modify chromatin by making it accessible during transcription and this is noteworthy because EBRs rearrange transcription factor genes. This might affect expression of other genes of the same pathway. Interestingly, two of these genes showed a different expression pattern between birds and mammals, HDAC8 involved in early embryo development³¹ and PRMT8 expressed in brain³². These results suggest a correlative link between chromosomal and morphological changes among species, mediated by rearranging genes controlling the expression of developmental pathways.

Estimates of non-avian dinosaur genome size, based on osteocyte sizes inferred from bone histology, identified a distinction between small, characteristically avian, genomes in theropods and sauropodomorphs vs. much larger ornithischian genomes³³. These results were interpreted as supporting the hypothesis that small genome size and low repeat content was a genomic exaptation that preceded and facilitated the endothermic metabolic demands of birds, e.g. for flight³⁴. It was further hypothesised that the avian karyotype evolved in response to a reduction in genome size in birds³⁴. This theory was subsequently challenged by a study that suggested that a decline in overall genome size occurred in non-volant dinosaurs³³. Here, we propose further that an avian-like karyotype not only predated the origin of flight but evolved well before, and independent of any,

purported genome size reduction. Nonetheless, we note that there may be an association between genomes with fewer chromosomes (and no microchromosomes) and larger genome sizes around 2.5–3 Gb, as in mammals³⁵ and crocodylians¹⁵. More repetitive elements could provide substrates for interchromosomal rearrangement, which is commonplace in mammals but rare in birds, and it has been suggested that an avian karyotype provides fewer opportunities for interchromosomal rearrangement due to the existence of fewer recombination hotspots (despite an overall higher recombination rate)^{36,37}, repeat structures^{12,38,39} and endogenous retroviruses^{2,4,40}. Therefore, although flight evolution might be correlated with smaller genome size (consider pterosaurs vs. other avemetatarsalians⁴¹; bats vs. other mammals³⁴ and strong vs. weak flying/ flightless birds⁴²), other mechanisms are clearly involved. Specifically, the formation of an avian-like karyotype long before the evolution of flight suggests neither a causative, nor a correlative link between the two.

Stasis of this karyotypic structure for >255 million years nonetheless suggests a highly successful mode of genome organisation that might have provided a blueprint for evolutionary success. The reasons for its persistence are speculative but might be due to its ability, facilitated by many chromosomes, including microchromosomes with high recombination rates, to generate variation, the substrate of natural selection. A larger number of small chromosomes inherently generate variation through increased genetic recombination and increased random chromosome segregation. Burt²⁴ suggested that a higher recombination rate has also contributed to the unique genomic features seen in microchromosomes such as high GC-content, low repeat content and high gene-density, which subsequently led to its maintenance. Variation, in turn, facilitates rapid adaptation and may therefore have contributed to wide phenotypic variation, in extant animals represented by >10,000 species of birds ($2n \sim 80$), >300 species of turtles ($2n < 68$) and, quite possibly, a large number of non-avian dinosaurs also. Of course, a karyotype with many, tiny chromosomes is not the only means by which variation can be generated (genic, epigenetic and interchromosomal variation all are mechanisms reported in other groups): indeed, amphibians display enormous phenotypic variation but possess relatively few chromosomes. Nonetheless, the above may explain the apparent paradox of a group with very little interchromosomal change, but incredible phenotypic diversity.

In conclusion, any 'dinosaur genomics' effort of this type is limited to reconstructing common ancestors (e.g. of birds and crocodylians), along with other nodes that have extant descendants (e.g. Archelosauria, Diapsida etc.) and inferring the most parsimonious set of events that led to extant animals. With this in mind, few studies have attempted to infer the nature of the gross structural genomic changes that occurred from the DCA, to the archelosaur ancestor, to birds. Given our data, it is perhaps not an unreasonable speculation that, if we had the opportunity to make metaphase chromosomes from tissue of non-avian theropods, both karyotypic and molecular cytogenetic analysis (genome size aside) would reveal little difference from a modern chicken, duck or ostrich (or at least a spiny soft-shelled turtle), i.e. $2n = 66-80$ in the majority of species. Of course, we cannot preclude the possibility that certain groups of non-avian dinosaurs underwent significant interchromosomal change, as these are known to occur among extant avian dinosaurs (kingfishers⁴³ (fissions), parrots⁴⁴ and falcons⁵ (fusions) are modern examples). Rather than being simply interesting descriptions of inferred karyotypes, therefore, we propose that the overall genome organisation and evolution of dinosaur chromosomes (inclusive of the avian radiation) might have been a major contributing factor to their morphological

disparity, physiology, high rates of morphological change⁴⁵ and ultimate survival. In other words, we have an apparent paradox of a highly stable karyotype that is rarely changing, nonetheless contributing to great morphological diversity. We already believed this to be the case for the great phenotypic diversity we see in birds; the current results however suggest that the karyotype may have contributed to species diversity in non-avian dinosaurs also. Moreover, the evidence that the karyotype had deeper origins than previously appreciated is congruent with other recent discoveries about dinosaur morphology, demonstrating that features previously thought to be characteristic of crown-group birds only (e.g. feathers and pneumatised skeletons) arose first among more ancient dinosaur or archosaurian ancestors^{23,46}. Dinosaurs have pervaded popular culture and the creative arts, perpetuated, in part, through film, television, press and literature. Their dominance for many millions of years, their radiations following two mass extinction events and, despite being almost wiped out by a third (the K–Pg meteor impact), their persistence as a highly diverse and speciose clade (extant birds)⁴⁷ has fascinated scientists since the very earliest discoveries. Of course, many of the evolutionary changes were in response to a rapidly changing environment. Whether a disproportionate advantage was offered by having an ‘avian-like’ karyotype will be the subject of future studies and speculation.

Methods

Cell culture and chromosome preparation. Chromosome preparations were established from fibroblast cell lines of the Caroline anole (*A. carolinensis*) ($2n = 36$), red-eared slider (*T. scripta*) ($2n = 50$) and spiny soft-shelled turtle (*A. spinifera*) ($2n = 66$). Cells were cultured at 30 °C and 5% CO₂ in Alpha MEM (Fisher), supplemented with 10% fetal bovine serum (Gibco) and 1% Pen-Strep-L-gutamine (Sigma). Chromosome suspension preparation followed standard protocols, briefly, mitostatic treatment with colcemid at a final concentration of 5.0 µg/ml for 1 h at 40 °C was followed by hypotonic treatment with 75 mM KCl for 15 min at 37 °C and fixation with 3:1 methanol:acetic acid.

Selection of BACs. Chicken and zebra finch BACs were chosen for interspecies FISH experiments according to a range of criteria, including the proportion of conserved elements shared across multiple avian species. Due to the high degree of apparent genome conservation observed between avian and reptilian species, this set of BACs was applied to chromosome suspensions of the birds in this study and from *A. carolinensis*, *T. scripta* and *A. spinifera*.

Preparation of BAC clones for FISH. BAC clone DNA was isolated using the Qiagen Miniprep Kit (Qiagen) prior to amplification and direct labelling by nick translation. Probes were labelled with Texas Red-12-dUTP (Invitrogen) and FITC-Fluorescein-12-UTP (Roche) prior to purification using the Qiagen Nucleotide Removal Kit (Qiagen).

Fluorescence in situ hybridisation. Metaphase preparations were fixed to slides and dehydrated through an ethanol series (2 min each in 2×SSC, 70%, 85% and 100% ethanol at room temperature). Probes were diluted in a formamide buffer (Cytocell) with Chicken Hybloc (Insight Biotech) and applied to the metaphase preparations on a 37 °C hotplate before sealing with rubber cement. Probe and target DNA were simultaneously denatured on a 75 °C hotplate prior to hybridisation in a humidified chamber at 37 °C for 72 h. Slides were washed post hybridisation for 30 s in 2 × SSC/0.05% Tween 20 at room temperature, and then counterstained using VECTASHIELD anti-fade medium with DAPI (Vector Labs). Images were captured using an Olympus BX61 epifluorescence microscope with cooled CCD camera and SmartCapture (Digital Scientific UK) system. Fissions and/or translocations were detected if cross-species signals appeared on two different chromosomes on the species of interest, and fusions were identified where the signals appeared on a noticeable larger chromosome than the (usually chicken) chromosome from which it was derived. This was achieved by visual inspection aided by ImageJ analysis for the smaller chromosome.

Reconstruction of the DCA karyotype. In order to reconstruct the hypothetical DCA, we selected the following sequenced extant amniote genomes: four diapsids (three birds, one lizard) and one basal mammalian representative as an outgroup (grey short-tailed opossum, *M. domestica*; assembly MonDom5). The three avian genomes, chicken (*G. gallus*; assembly galGal4), mallard (*A. platyrhynchos*; assembly BGI_duck_1.0; chromosome-level assembly, Faraut et al., personal

communication) and zebra finch (*T. guttata*; assembly taеGut1), were also used to reconstruct the avian ancestor, in this case using the Carolina anole genome (*A. carolinensis*; assembly anoCar2) as the outgroup. In order to compare reptilian and avian genomes cytogenetically, we selected chicken BAC probes designed to work in FISH experiments on all avian and reptilian chromosomes⁵. By combining novel bioinformatic and FISH data produced for the current study with that of previous studies, we established, by inference, the most parsimonious explanation of the available data regarding the nature of the dinosaurian genome. Initial experiments designed to establish whether the sequenced crocodile genomes⁴⁸ were suitable outgroup species met with only limited success, given the fragmented nature of these assemblies. This, and the fact that the microchromosomes are fused in extant crocodylians ($2n = 30$) meant that this genome assembly was not a suitable reference for our work; there was also no sufficiently well-assembled chromosome-level turtle genome available for analysis at the time of writing.

Alignment of multiple genomes and identification of HSBs and EBRs. Results from this study were generated from the alignment of the three best avian genomes assembled at a chromosomal level (chicken, mallard and zebra finch) along with the best-assembled reptile genome available assembled to a partial chromosomal level (Carolina anole) and one mammalian outgroup, grey short-tailed opossum (all genomes were aligned against chicken). The whole-genome sequences of the species of interest were aligned using LastZ and visualised using the interactive genome browser Evolution Highway^{4,49}. Pairwise blocks of synteny were identified relative to chromosomes of the chicken, which served as a reference genome (galGal4). Genome alignments for the five species as inferred from sequence orthology maps were mapped against chicken chromosomes. The start and end coordinates of the contiguously aligned orthologous regions observed in all the species compared were used to define msHSBs at the 300-kb resolution. These msHSBs were assigned to and subsequently sorted in individual chromosomes in each species according to their location, orientation and sequential order.

Arrangement of ancestral diapsid and avian karyotypes. To reconstruct a putative ancestral DCA karyotype, the Multiple Genomes Rearrangements and Ancestors tool version 2 (MGRA2⁵⁰; <http://mgra.bioinf.spbau.ru/>), was used as follows: based on pairwise alignments for mallard, zebra finch, Carolina anole, and grey short-tailed opossum visualised relative to the chicken, a set of respective msHSBs was generated as referred to above. In this case, the orthology map of the opossum was used as an input for the MGRA programme and included in the analysis as an outgroup. The five species-specific msHSB sets served as MGRA2 inputs for individual genomes which then produced a series of CARs representing the most likely ancestral configuration for the species identified in both hypothetical diapsid and avian ancestors.

Genome rearrangement analysis. To reconstruct the chromosomal changes that occurred between the groups, we used the MGR and Genome Rearrangements In Man and Mouse (GRIMM tools⁵¹; <http://grimm.ucsd.edu/>). MGRA2 outputs served as MGR/GRIMM inputs to trace the most parsimonious scenarios for evolutionary changes in two scenarios: first, the intrachromosomal and inter-chromosomal rearrangements that might have occurred from the hypothetical diapsid ancestor to the avian one and second, those rearrangements that may have occurred between the avian ancestor and the extant species.

Identification of gene ontology enrichment terms in HSBs and EBRs. Gene lists for msHSBs and archosaur EBRs were extracted from the Ensembl BioMart data system using chicken as the reference. With the chromosome-level assemblies available on Evolution Highway, definitions of EBRs are only possible for the archosaur ancestor. Since human genes are best annotated, the gene lists derived from chicken were matched to orthologous human genes and filtered for homology type and orthology confidence, leaving only those genes that were one-to-one orthologues. Background gene lists were also generated using all chicken–human orthologues with the maximum orthology confidence. The first two background gene lists tested covered all assembled chicken chromosomes and the second two lists only included results for 19 of the chicken chromosomes where the msHSBs and EBRs were found. In addition, in order to test whether genes with low gene identity matches affected the GO analysis, thresholds of 70, 60 and 50% homology at nucleotide level for the orthologue gene lists were set and the resulting GO outputs were compared. Based on these tests, the 70% gene identity threshold was selected for generating the msHSB/EBR gene lists, and the 19-chromosome list with all orthologous genes was used for the background GO analysis list. Gene lists were used as inputs for the web-based functional annotation tool DAVID⁵² using human Ensembl Gene ID as the list identifier and subsequently analysed using the Functional Annotation Clustering tool. Cluster data from each gene list output were downloaded into Microsoft Excel and filtered using an enrichment score of ≥ 1.3 and a p -value < 0.05 to edit the list for clusters considered to be significant. In addition, Functional Annotation Chart reports containing single GO terms and their associated genes were generated using the same gene lists. The latter information was also taken into account to identify significant GO terms for the tested gene lists, especially in situations when the Functional Annotation Clustering

analysis did not result in any significant gene–GO term enrichment groups. In order to correct for multiple sampling error, a FDR threshold of 5% was used. Finally, individual lists of genes that fit the GO criteria were manually curated and their function established, initially from Ensembl and thereafter from the original publications that described their isolation and analysis.

Data availability. The authors declare that the data supporting the findings of this study are available within the paper (and its supplementary information files).

Received: 4 December 2017 Accepted: 12 April 2018

Published online: 21 May 2018

References

- Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
- Romanov, M. N. et al. 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 (2014).
- Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1–106 (2015).
- Farré, M. et al. Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biol. Evol.* **8**, 2442–2451 (2016).
- Damas, J. et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res.* **27**, 875–884 (2017).
- Alföldi, J. et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
- Hillier, L. W. et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Rao, M. et al. A duck RH panel and its potential for assisting NGS genome assembly. *BMC Genomics* **13**, 513 (2012).
- Nadeau, J. H. & Taylor, B. A. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl Acad. Sci. USA* **81**, 814–818 (1984).
- Pevzner, P. & Tesler, G. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc. Natl Acad. Sci. USA* **100**, 7672–7677 (2003).
- Larkin, D. M. et al. A cattle–human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res.* **13**, 1966–1972 (2003).
- Warren, W. C. et al. A new chicken genome assembly provides insight into avian genome structure. *G3* **7**, 109–117 (2017).
- Warren, W. C. et al. The genome of a songbird. *Nature* **464**, 757–762 (2010).
- Mikkelsen, T. S. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
- St John, J. A. et al. Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* **13**, 415 (2012).
- Shaffer, H. B. et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013).
- Srikulnath, K., Thapana, W. & Muangmai, N. Role of chromosome changes in *Crocodylus* evolution and diversity. *Genomics Inform.* **13**, 102 (2015).
- Griffin, D. K. et al. Micro- and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenet. Cell Genet.* **87**, 278–281 (1999).
- Beçak, W., Beçak, M. L., Nazareth, H. R. S. & Ohno, S. Close karyological kinship between the reptilian suborder Serpentes and the class Aves. *Chromosoma* **15**, 606–617 (1964).
- Matsuda, Y. et al. Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. *Chromosom. Res.* **13**, 601–615 (2005).
- Kasai, F., O'Brien, P. C. M., Martin, S. & Ferguson-Smith, M. A. Extensive homology of chicken macrochromosomes in the karyotypes of *Trachemys scripta elegans* and *Crocodylus niloticus* revealed by chromosome painting despite long divergence times. *Cytogenet. Genome Res.* **136**, 303–307 (2012).
- Badenhorst, D. et al. Physical mapping and refinement of the painted turtle genome (*Chrysemys picta*) inform amniote genome evolution and challenge turtle–bird chromosomal conservation. *Genome Biol. Evol.* **7**, 2038–2050 (2015).
- Baron, M. G., Norman, D. B. & Barrett, P. M. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature* **543**, 501–506 (2017).
- Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97–112 (2002).
- Cracraft, J. et al. Response to Comment on “Whole-genome analyses resolve early branches in the tree of life of modern birds”. *Science* **349**, 1460 (2015).
- Uno, Y. et al. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS ONE* **7**, e53027 (2012).
- Capilla, L. et al. Mammalian comparative genomics reveals genetic and epigenetic features associated with genome reshuffling in Rodentia. *Genome Biol. Evol.* **8**, 3703–3717 (2016).
- Skinner, B. M. & Griffin, D. K. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**, 37–41 (2011).
- Larkin, D. M. et al. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* **19**, 770–777 (2009).
- Sankoff, D. The where and wherefore of evolutionary breakpoints. *J. Biol.* **8**, 66 (2009).
- Murko, C. et al. Expression of class I histone deacetylases during chick and mouse development. *Int. J. Dev. Biol.* **54**, 1527–1537 (2010).
- Wang, Y.-C. et al. Identification, chromosomal arrangements and expression analyses of the evolutionarily conserved *prmt1* gene in chicken in comparison with its vertebrate paralogue *prmt8*. *PLoS ONE* **12**, e0185042 (2017).
- Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M. & Edwards, S. V. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).
- Hughes, A. L. & Hughes, M. K. Small genomes for better flyers. *Nature* **377**, 391–391 (1995).
- Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl Acad. Sci. USA* **114**, E1460–E1469 (2017).
- Kawakami, T. et al. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014).
- Smeds, L., Mugal, C. F., Qvarnström, A. & Ellegren, H. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. *PLoS Genet.* **12**, e1006044 (2016).
- Gao, B. et al. Low diversity, activity, and density of transposable elements in five avian genomes. *Funct. Integr. Genom.* **17**, 427–439 (2017).
- Mason, A. S., Fulton, J. E., Hocking, P. M. & Burt, D. W. A new look at the LTR retrotransposon content of the chicken genome. *BMC Genomics* **17**, 688 (2016).
- Cui, J. et al. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biol.* **15**, 539 (2014).
- Organ, C. L. & Shedlock, A. M. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol. Lett.* **5**, 47–50 (2009).
- Gregory, T. R. In *The Evolution of the Genome* (ed. Gregory, T. R.) 3–87 (Elsevier, New York City, NY, 2005).
- Christidis, L. *Animal Cytogenetics 4: Chordata 3 B: Aves*. (Gebrüder Borntraeger, Berlin, 1990).
- Nanda, I., Karl, E., Griffin, D. K., Scharlt, M. & Schmid, M. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenet. Genome Res.* **117**, 43–53 (2007).
- Berv, J. S. & Field, D. J. Genomic signature of an avian Lilliput Effect across the K-Pg extinction. *Syst. Biol.* **67**, 1–13 (2018).
- Zhou, Z. The origin and early evolution of birds: discoveries, disputes, and perspectives from fossil evidence. *Naturwissenschaften* **91**, 455–471 (2004).
- Barrowclough, G. F., Cracraft, J., Klicka, J. & Zink, R. M. How many kinds of birds are there and why does it matter? *PLoS ONE* **11**, e0166307 (2016).
- Green, R. E. et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).
- Murphy, W. J. et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613–617 (2005).
- Avdeyev, P., Jiang, S., Aganezov, S., Hu, F. & Alekseyev, M. A. Reconstruction of ancestral genomes in presence of gene gain and loss. *J. Comput. Biol.* **23**, 150–164 (2016).
- Bourque, G. & Pevzner, P. A. Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* **12**, 26–36 (2002).
- Dennis, G. Jr et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).

Acknowledgements

This work was supported in part by the Biotechnology and Biological Sciences Research Council (BB/K008161/1 to D.K.G. and BB/K008226/1 and BB/J010170/1 to D.M.L.).

Author contributions

R.E.O.C. was the post-doc lead on the project, providing the first draft of the manuscript and performing the bulk of the experimental work. L.G.K. assisted with the cytogenetic work, while M.N.R., J.D. and M.F. contributed to the bioinformatic analysis. P.M.B., M.F.-S. and N.V. provided much needed insight into the nature of chromosome changes, essential material (M.F.-S. and N.V.) and insight into diapsid taxonomy and biology (P.M.B.). D.M.L. and D.K.G. were the academic leads on the manuscript, D.M.L. with an

emphasis on bioinformatics and D.K.G. on molecular cytogenetics. D.K.G. was the custodian of the manuscript and first point of contact dealing with reviewer comments. All authors contributed to writing the manuscript and contributing on drafts.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04267-9>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018