Article

Drivers of avian genomic change revealed by evolutionary rate decomposition

https://doi.org/10.1038/s41586-025-08777-7

Received: 3 October 2023

Accepted: 12 February 2025

Published online: 19 March 2025

Open access

Check for updates

David A. Duchêne[1,2]✉, Al-Aabid Chowdhury[3], Jingyi Yang[4], Maider Iglesias-Carrasco[2,5], Josefin Stiller[6], Shaohong Feng[7,8,9], Samir Bhatt[1,10], M. Thomas P. Gilbert[2,11], Guojie Zhang[7,9,12], Joseph A. Tobias[4] & Simon Y. W. Ho[3]

Modern birds have diversified into a striking array of forms, behaviours and ecological roles. Analyses of molecular evolutionary rates can reveal the links between genomic and phenotypic change[1–4], but disentangling the drivers of rate variation at the whole-genome scale has been difficult. Using comprehensive estimates of traits and evolutionary rates across a family-level phylogeny of birds[5,6], we find that genome-wide mutation rates across lineages are predominantly explained by clutch size and generation length, whereas rate variation across genes is driven by the content of guanine and cytosine. Here, to find the subsets of genes and lineages that dominate evolutionary rate variation in birds, we estimated the influence of individual lineages on decomposed axes of gene-specific evolutionary rates. We find that most of the rate variation occurs along recent branches of the tree, associated with present-day families of birds. Additional tests on axes of rate variation show rapid changes in microchromosomes immediately after the Cretaceous–Palaeogene transition. These apparent pulses of evolution are consistent with major changes in the genetic machineries for meiosis, heart performance, and RNA splicing, surveillance and translation, and correlate with the ecological diversity reflected in increased tarsus length. Collectively, our analyses paint a nuanced picture of avian evolution, revealing that the ancestors of the most diverse lineages of birds underwent major genomic changes related to mutation, gene usage and niche expansion in the early Palaeogene period.

Birds are the most species-rich lineage of tetrapods and occur in all major habitat types on the planet. Their evolutionary radiation is associated with striking biological innovation in adaptive traits[1], including flight[7,8], song[3,9,10], coloration[11,12] and beak morphology[4,13]. These traits have various associations with genomic evolution[14], with notable examples including reduced genome size in birds with more aerial lifestyles[7,15], accelerated evolution of regulatory mechanisms upon the loss of flight[2], and major genomic changes associated with bone structure and song complexity[1,3,9]. However, it remains a challenge to disentangle the role of life history in driving genome evolution, and to identify any particular genes that have strong associations with the diversification of avian phenotypes[14,16]. The increased availability of whole-genome sequences for large numbers of bird species[5] offers an unprecedented opportunity to explore how life history and phenotype are linked to genomic signals across evolutionary time[17]. In particular, whole genomes have now been sequenced for almost all bird families, offering an improved resolution of phylogenetic relationships at the family level and providing a comparative framework for detailed analysis of genomic evolutionary rates[6].

Studies of the rates of genomic change across taxa have helped to explain various macroevolutionary phenomena, with the potential to improve our understanding of the factors that drive evolutionary success. For instance, the explosive diversification in birds has been linked to faster genomic evolution[18] and to changes in the probabilities of different nucleotide point mutations[19]. Absolute evolutionary rate is also expected to be influenced by fundamental life-history traits such as generation length[20], body mass[21–23], metabolic rate[23,24] and polygyny[25], supporting a close relationship between ecology, genome evolution and macroevolutionary success.

Genomic comparative research has often focused on the substantial variation in evolutionary rates across taxa[26,27], while assuming that these patterns of variation are shared across all regions of the genome[28]. However, such a focus on among-lineage variation overlooks the nuanced signals across the genome that might reveal the role of

[1]Section of Health Data Science and AI, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [2]Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen, Denmark. [3]School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia. [4]Department of Life Sciences, Imperial College London, Ascot, UK. [5]Doñana Biological Station-Spanish Research Council CSIC, Seville, Spain. [6]Centre for Biodiversity Genomics, University of Copenhagen, Copenhagen, Denmark. [7]Center for Evolutionary and Organismal Biology, Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou, China. [8]Department of General Surgery of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China. [9]Innovation Center of Yangtze River Delta, Zhejiang University, Hangzhou, China. [10]MRC Centre for Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, London, UK. [11]Department of Natural History, University Museum, Norwegian University of Science and Technology, Trondheim, Norway. [12]Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ✉e-mail: david.duchene@sund.ku.dk

particular metabolic pathways or subsets of genes as drivers of evolution. In birds, for example, sex chromosomes and macrochromosomes appear to be evolving more rapidly than the rest of the genome[1,5]. This pattern is hypothesized to reflect the lower effective population size of sex chromosomes, the reduced gene density in large chromosomes and relaxed selective constraints, all of which heighten the effect of genetic drift. Similarly, rates might be accelerated in the coding regions of songbird genomes, as complex vocalization is likely to be associated with substantial genomic changes[9]. Whole-genome data from large numbers of avian taxa now offer a wealth of information for testing such hypotheses about the molecular evolutionary process. However, these data pose a challenge for the methods that are commonly used to study evolutionary rates, which generally have limited scalability.

Here we have identified the major axes of evolutionary rate variation across phylogenetic branches and genomic loci in birds. By using formal tests based on principal component analysis of molecular rate estimates, we are able to identify the most influential taxa and loci on each axis of evolutionary rate variation[29]. We refer to this approach as 'evolutionary rate decomposition' (Extended Data Fig. 1). We applied this method to a dataset of 218 de novo genomes produced by the Bird 10,000 Genomes (B10K) consortium[5] that sampled from almost every extant avian family (87%)[6], allowing us to identify genes that have covarying rates within subsets of lineages[29]. Using a comprehensive dataset of species traits, we identified the major correlates of evolutionary rates across families. We also leveraged estimates of rates of nonsynonymous substitutions (d_N) and synonymous substitutions (d_S) to gain insights into the contribution of mutation rate, selection and population size to evolutionary rate variation[30]. Our study provides a comprehensive description of the dominant drivers of avian evolutionary rates at the genomic scale.

Clutch size predicts mutation rates

To examine the relationship between biological traits and molecular evolution in birds, we analysed a comprehensive set of 23 life-history, morphological, ecological, geographical and environmental traits (Supplementary Data 1; Methods). We tested the association between these traits and average family-level rates across four measures of molecular evolution: d_N, d_S, d_N/d_S (ω) and rates in intergenic regions (Supplementary Data 2). These measures have the potential to provide complementary insights into the evolutionary process[30]: d_N is influenced by the mutation rate, population size and selection; d_S is influenced by the mutation rate; and their ratio (ω) is influenced by selection and population size. Genome-wide variation in ω is expected to be driven by population size. Rates in intergenic regions of the genome are expected to reflect the mutation rate. We used multiple Bayesian regression to test predictors of evolutionary rates, and supported these inferences with frequentist regression and machine learning-based random forest analyses. All regression models included body mass as a covariate, given its strong association with life-history traits.

Clutch size showed a significant positive association with mean d_N, d_S and mean rates in intergenic regions, after accounting for a diverse range of covariates (Fig. 1a,b; d_N coefficient 95% credible interval (CI): 1.19–8.57; d_S 95% CI: 4.20–10.50; intergenic 95% CI: 3.57–10.03). Generation length was further revealed by random forest analyses as the most important variable in driving molecular rate variation, receiving the maximal importance score in explaining d_N, d_S, ω and intergenic region rates, this time in a negative relationship (Fig. 1a,c). These results suggest a role for clutch size and generation length in driving mutation rates across deep timescales, which is consistent with both variables being found as predictors of instantaneous mutation rates in parent–offspring trios[31]. Our results are also consistent with findings of a link between clutch size and substitution patterns among the four nucleotide types[19]. Clutch size is likely to have an important role in adaptation, given its association with a wide range of other species characteristics, including growth rate, egg size and longevity, as well as geographical factors such as insularity and latitude[32]. Meanwhile, animals with short generations tend to copy their genomes more frequently per unit of time, whereas those with long generations are expected to place greater investment in DNA repair[20,31]. We also note that clutch size and generation length covary, such that longer-lived birds tend to have smaller clutches. The independent effects of these two variables on molecular evolutionary rates are likely to be connected via other overriding life-history strategies (for example, pace of life). Clutch size might offer greater explanatory power for molecular rate variation than other variables because of its effect on the number of viable genomic replications per generation. Larger clutch sizes are associated with greater numbers of viable copies of the genome, increasing the opportunity for mutations to be transmitted to future generations[33]. Alternatively, the expected greater parental care in small clutches could lead to reduced mutation rates, driven by reduced exposure to mutagens in the germline[34]. However, this latter hypothesis has not been described or tested in detail.

When we used species-level trait data rather than family averages, multiple regression analyses revealed additional links between ecological traits and genome-wide rates. In these analyses, taxa with shorter tarsi were found to have elevated rates of d_N (95% CI: −9.88 to −1.13) and intergenic region evolution (95% CI: −8.06 to −0.13; Extended Data Fig. 2). These links are consistent with genomically widespread adaptations to flight-intensive lifestyles associated with the shortest tarsi. Short tarsi are generally associated with arboreal and particularly aerial lifestyles, whereas long tarsi are often found in ground-dwelling birds[35]. Therefore, the links that we identified are consistent with genomically widespread adaptations for flight-intensive lifestyles associated with the shortest tarsi[36]. High intergenic rates in these taxa might reflect the mutagenic effect of the oxidative stress of intensive flight habits[24].

Population size provides a further possible link between genome-wide d_N and variables such as tarsus length, wing length, brain mass, beak size and developmental mode, all found as moderately important in random forest analyses (Fig. 1a). For example, taxa with longer tarsi such as cariamids and rails are often ground-dwelling, with a tendency towards reduced dispersal or flight ability, potentially reducing connectivity among populations[37]. Nonetheless, abundance and traits associated with flight ability, such as the hand-wing index, were not significant predictors of molecular rates in our data. Unlike clutch size, the effect of tarsus length on mutation rate varies with the comparative analysis framework. Many such traits with inconsistent effects are unlikely to represent dominant drivers of genomic rates across avian history and, instead, might affect subsets of genes and taxa.

Another trait of interest is body mass, which we found to be a strong predictor of avian molecular rates in single-variable regression. This association might arise via a link between body mass and the number of cells under pressure to reduce somatic mutation and damage, particularly when body mass is associated with longevity[20]. However, body mass is correlated with several life-history traits, and once we accounted for these in multivariate models, we found no evidence that body mass predicts variation in molecular rates (Supplementary Data 3). Therefore, our data suggest that changes in body mass have not directly driven genome-wide evolutionary rates, but that these are indirectly linked to molecular evolution through more dominant mechanisms mediated by clutch size, generation length and other factors.

In contrast to other molecular rate metrics, genome-wide values of ω did not show an association with any of the sampled traits, pointing to a limited effect of fluctuations in selection or population sizes on avian molecular evolution. Population sizes are believed to have increased rapidly immediately after the Cretaceous–Palaeogene (K–Pg) transition, when a wide array of ecological niches became available following the extinction of many larger species of birds and other animals, as proposed by the 'Lilliput effect' hypothesis[23,38,39]. The large populations that came as a consequence would have experienced reduced

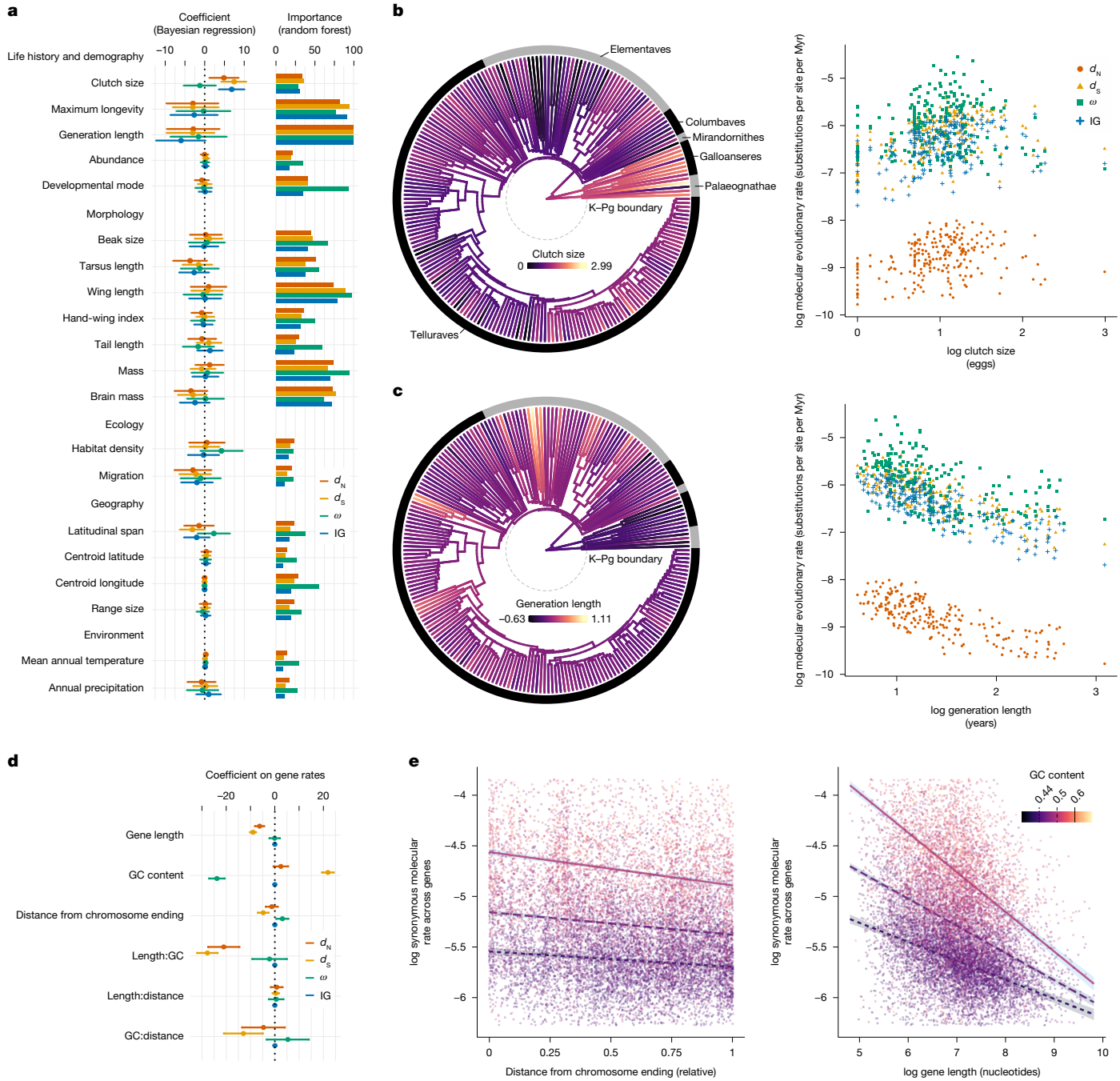


Fig. 1 | Association between molecular evolutionary rates and traits among avian families and genes. a, Clutch size significantly predicted the largest number of metrics of genome-wide molecular rates in Bayesian multiple regression, whereas generation length was also found to have the greatest importance scores in driving molecular rate variation. IG, intergenic. Error bars indicate the 95% highest posterior density interval (HPDI) in coefficients. **b**,**c**, Avian family lineages ($n = 218$) with larger clutches show greater molecular evolutionary rates (**b**), whereas lineages with longer generations show reduced molecular rates (**c**). Branch colours show uncorrected log mean clutch size per family, whereas those coloured by generation length show the residuals of the variable regressed against body mass. Myr, million years. **d**, Rates of d_S at genes ($n = 14{,}368$), with mean rate taken across all lineages at each locus, were explained in Bayesian regression by terms involving the content of guanine and cytosine (GC content). **e**, Interaction terms showed accelerated d_S rates at genes with high GC content that occur close to chromosome endings and that have short sequences.

genetic drift, leading to more efficient removal of slightly deleterious non-synonymous mutations and thus lower ω in smaller-sized avian taxa[30]. One explanation for the lack of this signal in our data is that the effects of population-size fluctuations are indistinguishable from signatures of selection, and have been eroded by other genomic processes across ancient timescales. For instance, inferences of genome-wide ω at deep timescales can be negatively affected by multiple substitutions (saturation), phylogenetic error and shifts in the evolutionary rate of subsets of genes in individual lineages. In addition, the effects of selection and population size are uneven across the genome[6], which is reflected in our finding of variation in ω across axes of variation and across chromosomes. Finally, incomplete lineage sorting and reticulate evolution can also degrade the signal of ancestral population size[40]. Future examinations of these questions at finer taxonomic levels will be able to produce a more nuanced picture, for instance, by extending estimates of population-size dynamics[41] to deeper timescales.

Evolutionary rates at individual gene loci, averaged across lineages, offer a complementary insight into overall drivers of genomic evolution. We tested whether gene-specific evolutionary rates, taken as the unweighted average rates across lineages per gene, could be explained by three key variables: the proportional content of guanine and cytosine in sequences (GC content), gene length and gene location within chromosomes. Multiple regressions of rates of d_S revealed faster evolution in genes with higher GC content (95% CI: 19.42–24.29), shorter sequences (95% CI: −10.23 to −7.74) and smaller distance to chromosome endings (95% CI: −7.01 to −2.56; Fig. 1d). This is in agreement with previous work on the effect of GC content on synonymous evolutionary rates in birds[22]. Increasing values of ω were observed with decreasing GC content (95% CI: −27.09 to −20.64) and greater distance from chromosome endings (95% CI: 0.60–5.60), probably reflecting the effect of these factors on d_S but not d_N. Interaction terms in regression revealed that the highest rates of d_N and d_S occur in short genes with high GC content (d_N 95% CI: −27.47 to −14.58; d_S 95% CI: −32.01 to −23.56), and highest d_S also occurring in genes with high GC content located near chromosome ends (95% CI: −20.97 to −4.92). However, we did not find evidence for an interaction where accelerated evolution occurs in short genes found near chromosome endings. Our results point to a dominant role of GC content in mediating mutation rate across genes given that d_S (primarily associated with mutation rate) is linked to the factors examined, whereas changes in ω appear concomitant with changes in d_S but not d_N. High GC content has previously been associated with more frequent breaks in DNA strands, greater recombination[42] and a unique evolutionary dynamic in avian genes[22]. Similarly, chromosome endings tend to have complex evolutionary signatures consistent with accelerated rates[6,35].

Major changes in rates across lineages

Rate decomposition presents a powerful approach for the analysis of molecular rates that allows a dissection of the subsets of lineages and genes that have had an unusually large effect on evolutionary rate variation[29] (Extended Data Fig. 1). By using principal component analysis of the full set of branches in the phylogenetic tree, the method uses unsupervised learning to model the main axes of variation in rates. Permutation tests can reveal the contribution of each axis, and of each taxon on each axis, to variation in rates. Out of 433 principal components in each avian molecular dataset, our analyses of rate decomposition consistently showed that only a small portion of components (less than 5%; $P < 0.01$) explained significant amounts of variation in rates. This included 8 principal components in d_N, 10 in d_S, 7 in ω and 16 in intergenic regions.

The first principal component explained large proportions of variation in d_N (35%), ω (38%) and rates in intergenic regions (24%), and the lineages with the greater loadings on this component were consistently the terminal branches in the avian phylogeny (Fig. 2 and Extended Data Figs. 3 and 4). This indicates that there is no particular subset of family lineages that drives rate variation across loci. Instead, most loci show a consistent pattern of rate variation across families, supporting the dominant effect of gene-specific rates[26]. As this result of terminal branches dominating rate heterogeneity is observed in d_N and ω, but not d_S, it is likely to reflect widespread variation in genome-wide selective constraints or population sizes within families (Extended Data Fig. 5). Denser taxon sampling will enable further disentanglement of the drivers of rate variation, such as any concomitant climate-driven fluctuations in population size across families[43]. Alternatively, the strong influence of terminal branches might indicate the difficulty in recovering signals of selection or population-size fluctuation from nucleotide data at deeper timescales[44].

The other significant principal components of d_N and ω highlighted the stem branches leading to the large neoavian clades Telluraves and Elementaves, as well as several deep branches in the passerine clade, as being hotspots of variance in molecular rates (Fig. 2). The prominent role of these early neoavian branches in explaining variation in molecular rates points towards major ecological and adaptive changes across birds occurring immediately after the K–Pg transition. This is consistent with evidence of rapid change across a range of traits such as body mass and brain mass at that time[23,38]. The major principal components of d_S variation also point to the same branches (stem lineages leading to Telluraves and Elementaves) as having the greatest explanatory power, consistent with genome-wide increases in mutation rate soon after the K–Pg transition[23,40].

Extensive variation in chromosomal rates

Evolutionary rates vary considerably across the chromosomes of avian genomes, with the largest macrochromosomes showing evidence of lower rates of mutation (d_S) than expected under permutation of loci (Fig. 3a). Similarly, microchromosomes frequently have significantly higher evolutionary rates than expected under permutation. These results are consistent with greater rates of intrachromosomal rearrangement and recombination in microchromosomes than in macrochromosomes[6,42]. Microchromosomes have also been found to contain greater amounts of simple repeats and hotspots of double-strand breakage that are GC rich. Our data also show microchromosomes to be GC rich overall (Fig. 3), consistent with the hypothesis that the increased GC content that arises from recombination is a possible source of DNA damage and a likely cause of elevated mutation rates.

Rate decomposition further revealed the subsets of lineages in which chromosomes showed extremes in evolutionary rates. The extreme values of principal components of ω and intergenic regions included an overrepresentation of loci in microchromosomes (Fig. 3b). That is, although all microchromosomes were overrepresented at the extremes of at least one principal component, chromosomes 1–3 and 5–7 were not overrepresented in any significant principal component. Furthermore, this distribution of ω highlights the role of microchromosomes in driving rates in principal component 3, a component that is substantially explained by the earliest neoavian and passerine lineages, suggesting widespread selective constraints on microchromosomes immediately after the K–Pg transition. As microchromosomes are thought to be crucial for the expression of housekeeping genes[42], the result suggests that the earliest members of Neoaves experienced major evolutionary rate changes in their housekeeping machinery. Intergenic regions show similar patterns of extreme rates occurring primarily in microchromosomes and at principal components that are explained by early passerine branches (Extended Data Fig. 3).

The overrepresentation of microchromosomes in the evolutionary rate shifts along early neoavian branches suggests a greater influence from microchromosomes than macrochromosomes on avian evolution. This result is consistent with evidence that microchromosomes are highly acetylated powerhouses of gene expression and that they form the building blocks of larger chromosomes[41]. Meanwhile, macrochromosomes are less resistant to major rearrangements, and their large size is a consequence of the fusion of microchromosomes that are less transcriptionally active[42]. Evolutionary rates are likely to have varied the most in microchromosome regions undergoing intense selection, including those fundamental for development and adaptation. Overall, our data provide evidence that severe selective constraints and conserved synteny can co-exist with high nucleotide mutation rates over large temporal scales.

Early shifts in key genomic machineries

To investigate the functional groups of genes associated with the most prominent avian families driving rate variation (defined by branches with significant loadings on principal components), we examined the loci with the most extreme values at each principal component.

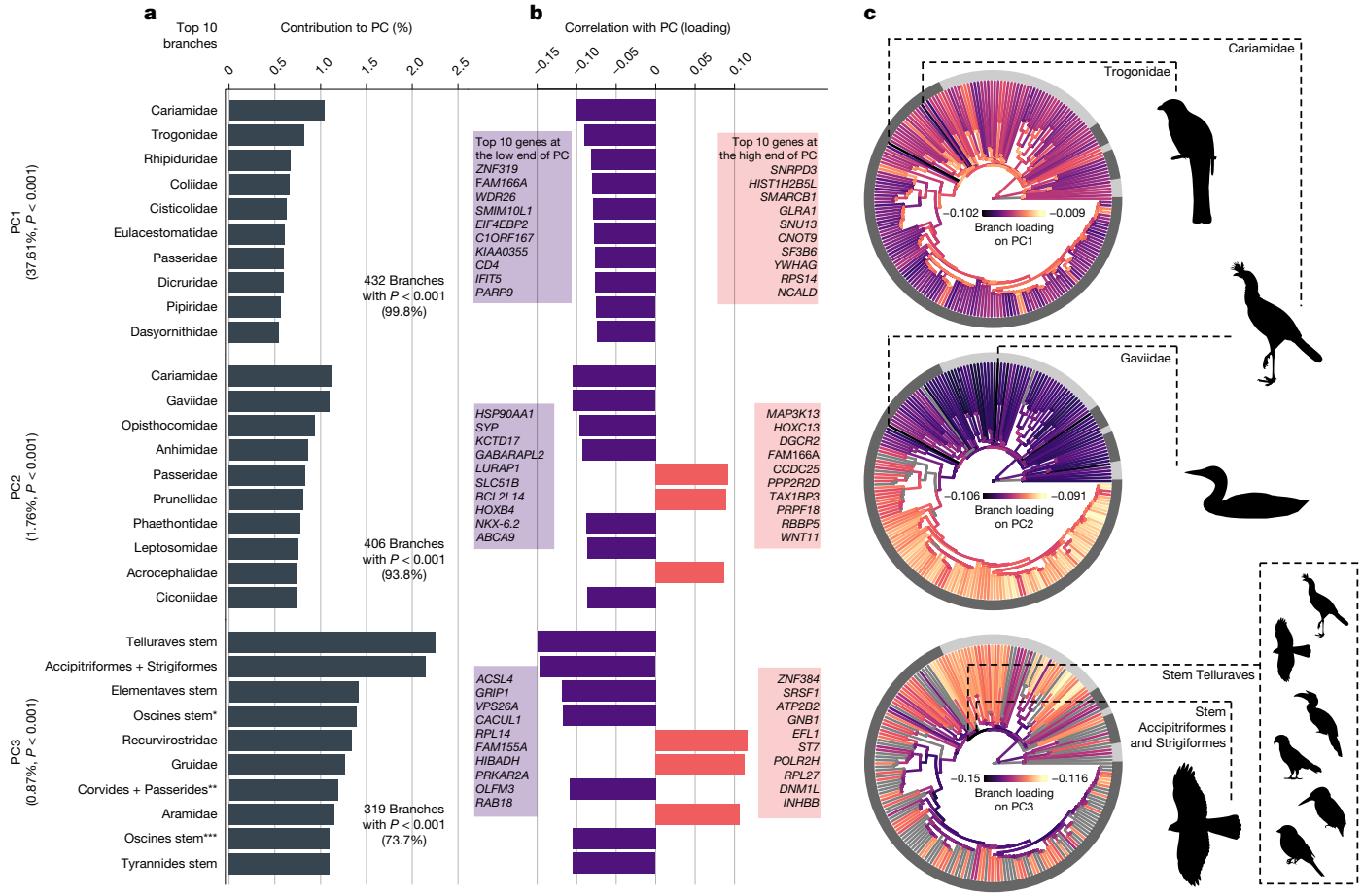


Fig. 2 | Contribution to variance and loading across taxa on principal components of rates of ω across loci. The rates matrix that undergoes decomposition includes branches as columns and loci as rows, and is built separately for each type of rate (d_N, d_S, ω and intergenic; the results of ω rates are shown here). Principal components (PCs) describe the variation of locus rates as explained by taxa with significantly high loadings (Extended Data Fig. 1). **a**, Taxon contributions for the top ten taxa in the top three principal components. The principal components and taxa shown here explain significantly more variation than expected under permutation ($P < 0.01$). According to the taxonomy of passerines[56]: * indicates the stem of the oscines minus the grouping of Menuridae and Atrichornithidae; ** denotes the stem of infraorders Corvides and Passerides, plus the grouping of families Orthonychidae and Pomatostomidae; and *** indicates the stem of the oscines minus Menuridae, Atrichornithidae, Climacteridae and Ptilonorhynchidae. **b**, Branch loadings indicate the correlation between taxa and locus rates in each principal component. **c**, Branches are coloured by loadings, in which grey branches are those that did not have significantly greater explanatory power than expected under permutation. The outer ring indicates the major taxonomic groups named in Fig. 1, whereas the dashed lines indicate the two branches with the greatest loading in each principal component. Silhouettes in panel **c** were adapted from PhyloPic (http://phylopic.org) under CC0 1.0 (Alcedinidae, *Buceros rhinoceros*, *Nestor notabilis*, *Passer domesticus*, trogon and *Accipiter nisus*) and Public Domain Mark 1.0 (*Gavia immer* and *Cariama cristata*) licences.

To understand the ecological implications of these genes, we also examined whether the proportions of variance explained by each avian family (loadings) were associated with their life-history and morphological traits described above. Gene set enrichment analysis was performed for the 20% of loci at the extremes of principal components of ω (Extended Data Fig. 1), focusing on those components that explained significantly greater heterogeneity than expected under permutation ($P < 0.01$). Along the first principal component of ω, enrichment was found for the ribosome gene term at high values of the axis (Supplementary Data 2). As this axis is negatively linked with rates on terminal branches, the result indicates a marked slowdown in ribosomal evolutionary rates across many present-day avian families.

The loadings at the third principal component of ω (Supplementary Data 3) showed a significant positive relationship with tarsus length (Fig. 4a, 95% CI: 0.12–1.28), suggesting that genes with greater values on this axis had accelerated rates in taxa with long tarsi. This principal component (PC3 of ω) was one of the few in which large numbers of taxa contributed to variation significantly more than expected under permutation, and its link with tarsus length is the only significant association among any trait variable and components with large numbers of taxon representatives.

One distinctive feature of tarsus length is its very rapid increase in diversity immediately following the K–Pg transition (Fig. 4b). The branches leading to birds with long tarsi represent a diverse and ancient set of avian families including plovers (Charadriidae), cranes (Gruidae), several raptor families (Accipitridae, Pandionidae and Sagittariidae) and some passerine families (Hyliidae, Cardinalidae and Thraupidae). Given that tarsus length is strongly associated with primary modes of locomotion and lifestyle in birds[45,46], this finding suggests rapid diversification of lineages into an array of foraging niches associated with aerial, arboreal, aquatic and terrestrial lifestyles, all of which have evolved diverse tarsi that fit their ecology. The rapid evolution of this trait soon after the K–Pg transition suggests strong positive selection that can be explained by the enriched gene terms at the ends of PC3 of ω. In addition to corresponding to accelerated evolutionary rates along terminal branches of birds with long tarsi, gene terms enriched at the extreme of this principal component indicate decelerated rates in some of the branches with the greatest

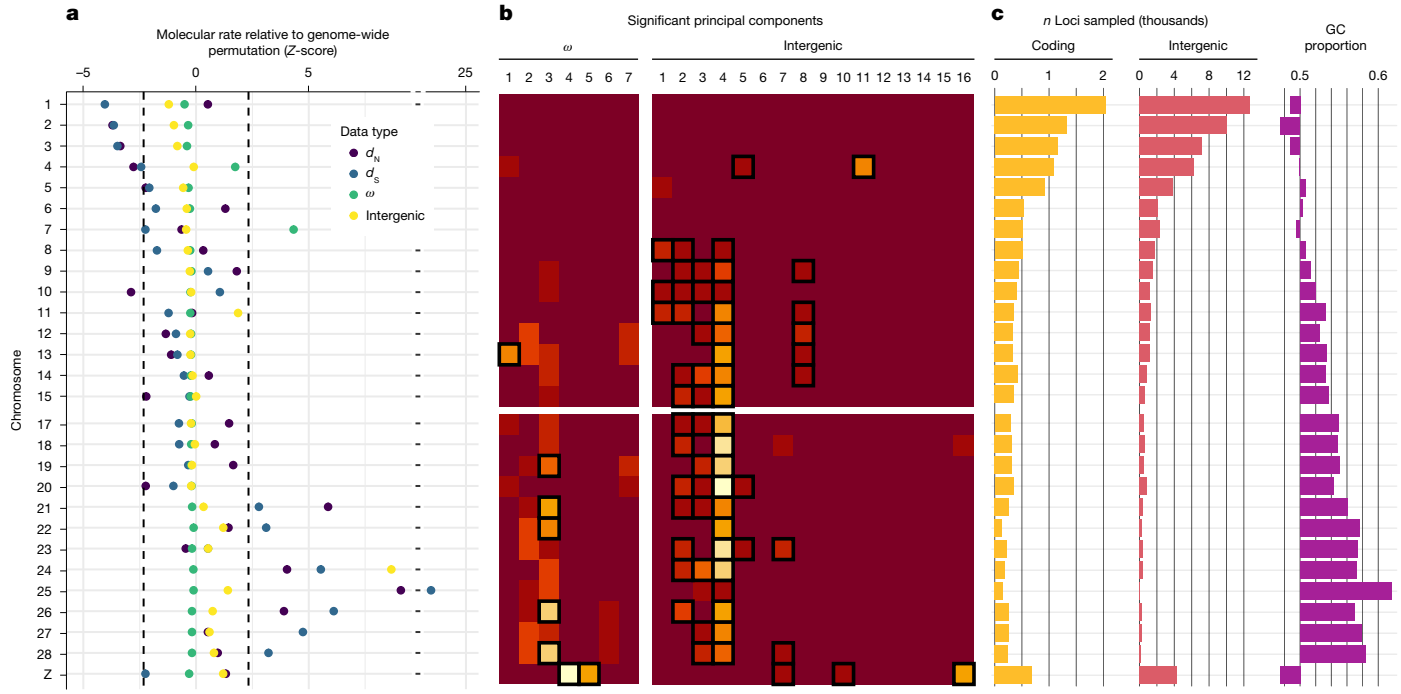


Fig. 3 | Avian chromosome rates and chromosomal contribution to principal components in rate decomposition analyses. a, Each chromosome was assessed for its mean rate being different from that expected under genome-wide permutation. **b**, Loci at the extremes of principal components were tested for having an overrepresentation of each chromosome. The light colours indicate increasing overrepresentation of a chromosome (rows) at a particular principal component (columns). The boxed cells indicate chromosomes with a significantly greater portion of genes at the extremes of a principal component ($P < 0.01$). Chromosome 16 was excluded from all analyses due to a small sample size of loci. **c**, The number of loci in our samples reflects chromosome size, whereas the mean proportion of GC content is likely to reflect the density of genes and recombination breakpoints.

loadings on this axis, including early Neoaves (stem Telluraves and stem Elementaves) and early passerine branches (Fig. 2). These findings of a nuanced effect of the ecology surrounding tarsus length on selection regimes also complements previous work that has identified gene regulation of limb development as crucial in avian evolution[2], and the tarsus as being linked to the evolution of non-coding genomic elements[47].

In addition to higher ω in taxa with long tarsi and lower ω in deep post-K–Pg branches, we found the top end of the ω PC3 to be enriched for five gene functional terms (Fig. 4c). One of these terms was oocyte meiosis (Fig. 4d), which is the primary pathway involved in the replication machinery of the germline and can be expected to have widespread effects across the genome through its effect on mutation rate, and hence estimates of d_S. The principal components of d_S did not show enrichment for any gene term, and instead consistently supported stem branches of Telluraves and of Elementaves (or related early Neoaves) as dominant drivers of evolutionary rate changes (Extended Data Fig. 3). One explanation is that, in addition to a broad range of adaptive changes, avian genomes were subject to changes in meiotic DNA repair mechanisms that led to increased mutation rates along those early branches. Support for this hypothesis includes the presence of genes that are directly involved in DNA repair among those in the enriched oocyte meiosis term, such as MAPK[48,49]. A further finding was the enrichment of genes encoding adrenergic signalling in cardiomyocytes (Fig. 4e). This pathway is associated with cardiac performance[50,51] and suggests rapid evolution in heart function in old families with long tarsi, possibly related to relaxation of the selective constraints for intense flight. Therefore, this is also likely to reflect stronger selective constraints in taxa with short tarsi, allowing early avian lineages to sustain intense flight habits.

Additional enriched pathways included the combination of the spliceosome, mRNA surveillance and the ribosome (Fig. 4f). In combination, these terms form the backbone of the ribonuclear cellular machinery that regulates the quality control and translation of gene products. The evolution of ribonuclear machineries has been linked to protection against deleterious insertions and mutation[52–54], whereas a link between gene regulation and limb development has been identified as central to avian evolution[2]. The change in ω in ribonuclear machineries in a broad range of taxa with long tarsi suggests that genomic changes contributed to increased efficiency or novelty in gene expression and regulation. Ribonuclear machinery evolution in some of the oldest families with long tarsi may have had a role in the colonization of a vast diversity of habitats, contributing to the rapid radiations of these groups after the K–Pg transition. The marked reduction in ω in stem Telluraves and stem Elementaves (or related early Neoaves) might reflect large population sizes that were probably seen after the K–Pg transition. It is likely that there was widespread competition among early avian lineages at this time, and changes in fundamental ribonuclear machineries in early Neoaves could have conferred a competitive advantage, providing some key steps towards the explosive diversifications that followed. Together, these results indicate that a dramatic increase in novel adaptations took place in stem neoavians shortly after the K–Pg transition, and that these changes were likely to have been facilitated by evolution in ribonuclear machineries.

Evolutionary rate decomposition can reveal various genomic evolutionary phenomena with smaller effects. Across avian genomes, the taxon contributions along the fifth principal component of ω are linked with brain mass (95% CI: −2.52 to −0.33; Supplementary Data 3) and range size (95% CI: −0.69 to −0.03), with the strongest contribution to variance coming from albatrosses (Diomedeidae) and all three families of birds of prey in the order Accipitriformes. Enriched terms that are accelerated in these clades include insulin signalling, critical for lipid and glucose homeostasis, Hedgehog signalling, linked with growth, and cardiomyocyte signalling, related to heart function

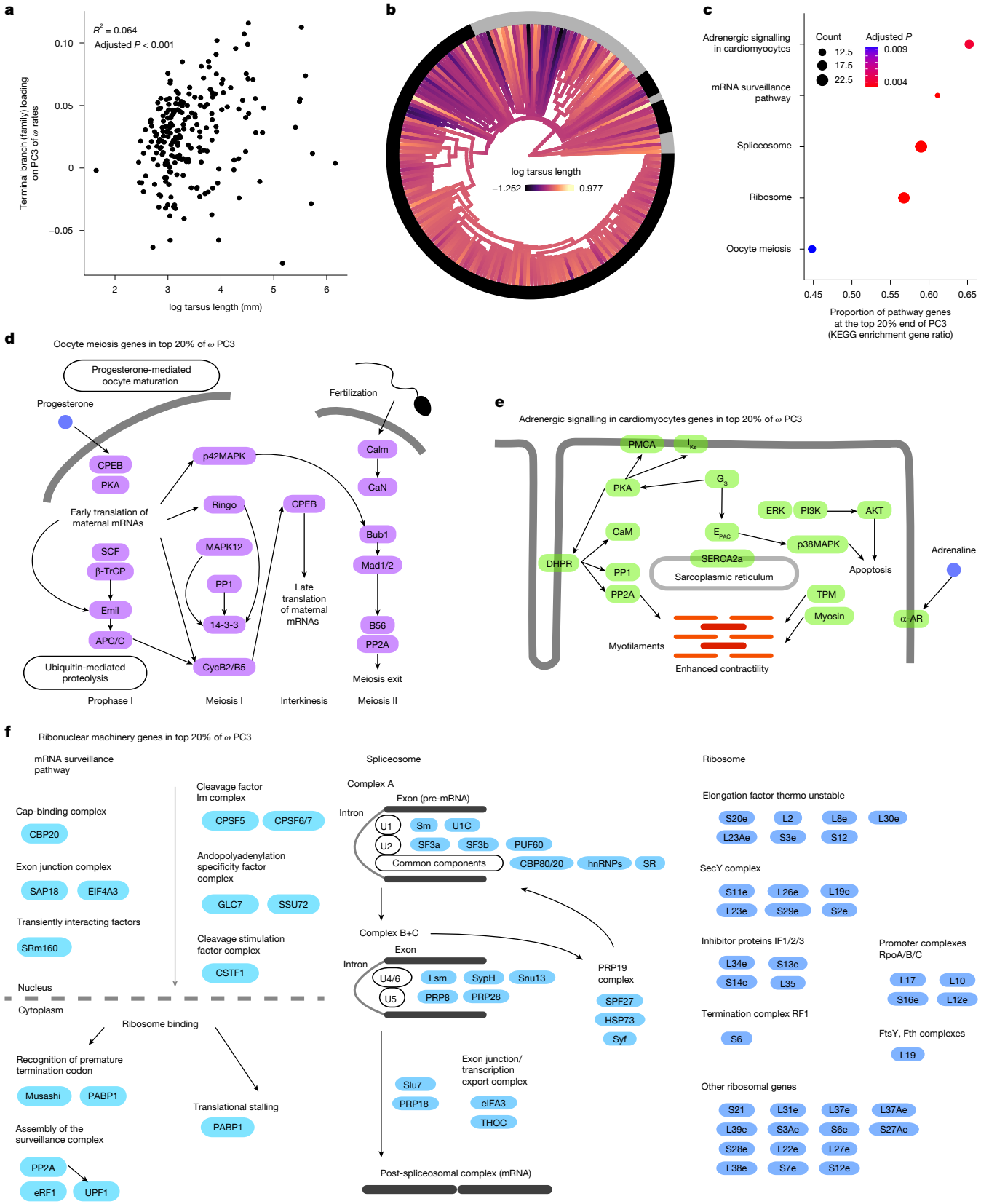


Fig. 4 | Trait regression and gene enrichment associated with the third principal component of *ω* decomposition. a, Regression *P* value and R^2 are shown for the single-variable regression among family-level (terminal branches) loadings on *ω* third principal component (PC3) against log mean tarsus length. This variable was also uniquely significant in multiple regression involving other traits as partial terms. **b**, The phylogenetic distribution of tarsus length shows the increase in diversity in this trait immediately after the K–Pg transition. **c**–**e**, Genes with extreme values of PC3 of *ω* were enriched for multiple functional terms (**c**), including oocyte meiosis, putatively linked to DNA repair mechanisms (**d**); and adrenergic signalling in cardiomyocytes, linked with heart function (**e**). **f**, Additional enriched terms included the ribonuclear machineries for splicing, mRNA surveillance and ribosomal activity. Only the genes inside the extreme values of PC3 of *ω* are shown, such that their metabolic links are approximate.

(Supplementary Data 4). These terms suggest that albatrosses and some birds of prey underwent substantial genomic changes in body growth patterns and fuel usage[55], possibly allowing endurance locomotion in large-bodied taxa. The sixth principal component of ω was linked with several lineages of waterbirds with long tarsi (95% CI: −2.15 to −0.44) and inhabiting low-rainfall environments (95% CI: −1.52 to −0.27), and revealed accelerated rates in gene terms including cellular senescence and lectin signalling. These could be related to accelerated evolution in immune systems, which might be needed to sustain the large colonial lifestyles of many water birds. Analyses of these patterns of decomposed rates across genomes at finer taxonomic scales will provide a refined picture of the dominant drivers of adaptations within avian orders and families.

Conclusions

Our study has demonstrated the effectiveness of using a framework based on genomic evolutionary rates to disentangle processes of mutation and selection across deep timescales. By identifying the dominant axes of evolutionary rate variation, our analyses have shed light on the relationships between biological diversity and genomic evolution in birds. To achieve this, we performed a series of genomic comparative analyses that revolved around the decomposition of evolutionary rates across regions of the genome and across branches of the bird phylogeny. These analyses provide evidence that the key drivers of genome evolution in birds include clutch size, generation length and tarsus length[2,19,31,47]—traits long considered important in avian life history and ecology—along with GC content in genes[22]. Our decomposition of evolutionary rates also revealed the importance of meiotic, cardiac and ribonuclear machinery evolution along some deep branches in the avian tree. These shifts in evolutionary rates took place primarily in microchromosomes and upon the evolution of tarsus length, a trait linked to terrestrial locomotion. Collectively, our analyses indicate that meiotic and ribonuclear machineries and microchromosomes, with crucial roles in DNA replication, gene usage and housekeeping, underwent rapid evolution that is likely to have contributed to the explosion of avian diversity in the immediate aftermath of the K–Pg transition. The focus of this study on nucleotide evolution provides a foundation for testing the drivers of evolutionary rates in other genomic components, such as karyotype, protein structure, gene rearrangement and transposition. We expect that analyses of genomes gathered from large numbers of taxa will provide valuable insights into the evolutionary processes shaping phenotypic variation across the avian tree of life.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-08777-7.

1. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
2. Sackton, T. B. et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
3. Whitney, O. et al. Core and region-enriched networks of behaviorally regulated genes and the singing genome. *Science* **346**, 1256780 (2014).
4. Yusuf, L. et al. Noncoding regions underpin avian bill shape diversification at macroevolutionary scales. *Genome Res.* **30**, 553–565 (2020).
5. Feng, S. et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
6. Stiller, J. et al. Complexity of avian evolution revealed by family-level genomes. *Nature* **629**, 851–860 (2024).
7. Wright, N. A., Gregory, T. R. & Witt, C. C. Metabolic "engines" of flight drive genome size reduction in birds. *Proc. Biol. Sci.* **281**, 20132780 (2014).
8. 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).
9. Pfenning, A. R. et al. Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014).
10. Mason, N. A. et al. Song evolution, speciation, and vocal learning in passerine birds. *Evolution* **71**, 786–796 (2017).
11. Toomey, M. B. et al. High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proc. Natl Acad. Sci. USA* **114**, 5219–5224 (2017).
12. Aguillon, S. M., Walsh, J. & Lovette, I. J. Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proc. Biol. Sci.* **288**, 20201805 (2021).
13. Abzhanov, A. et al. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442**, 563–567 (2006).
14. Bravo, G. A., Schmitt, C. J. & Edwards, S. V. What have we learned from the first 500 avian genomes? *Annu. Rev. Ecol. Evol. Syst.* **52**, 611–639 (2021).
15. Gregory, T. R., Andrews, C. B., McGuire, J. A. & Witt, C. C. The smallest avian genomes are found in hummingbirds. *Proc. Biol. Sci.* **276**, 3753–3757 (2009).
16. Price-Waldman, R. & Stoddard, M. C. Avian coloration genetics: recent advances and emerging questions. *J. Hered.* **112**, 395–416 (2021).
17. Smith, S. D., Pennell, M. W., Dunn, C. W. & Edwards, S. V. Phylogenetics is the new genetics (for most of biodiversity). *Trends Ecol. Evol.* **35**, 415–425 (2020).
18. Lanfear, R., Ho, S. Y. W., Love, D. & Bromham, L. Mutation rate is linked to diversification in birds. *Proc. Natl Acad. Sci. USA* **107**, 20423–20428 (2010).
19. Berv, J. S. et al. Genome and life-history evolution link bird diversification to the end-Cretaceous mass extinction. *Sci. Adv.* **10**, eadp0114 (2024).
20. Bromham, L. Why do species vary in their rate of molecular evolution? *Biol. Lett.* **5**, 401–404 (2009).
21. Weber, C. C., Nabholz, B., Romiguier, J. & Ellegren, H. Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol.* **15**, 542 (2014).
22. Botero-Castro, F., Figuet, E., Tilak, M.-K., Nabholz, B. & Galtier, N. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. *Mol. Biol. Evol.* **34**, 3123–3131 (2017).
23. Berv, J. S. & Field, D. J. Genomic signature of an avian Lilliput effect across the K–Pg extinction. *Syst. Biol.* **67**, 1–13 (2018).
24. Montoya, P., Cadena, C. D., Claramunt, S. & Duchêne, D. A. Environmental niche and flight intensity are associated with molecular evolutionary rates in a large avian radiation. *BMC Ecol. Evol.* **22**, 95 (2022).
25. Iglesias-Carrasco, M., Jennions, M. D., Ho, S. Y. W. & Duchêne, D. A. Sexual selection, body mass and molecular evolution interact to predict diversification in birds. *Proc. Biol. Sci.* **286**, 20190172 (2019).
26. Duchêne, D. A. et al. Linking branch lengths across sets of loci provides the highest statistical support for phylogenetic inference. *Mol. Biol. Evol.* **37**, 1202–1210 (2020).
27. Snir, S., Wolf, Y. I. & Koonin, E. V. Universal pacemaker of genome evolution. *PLoS Comput. Biol.* **8**, e1002785 (2012).
28. Lanfear, R., Welch, J. J. & Bromham, L. Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* **25**, 495–503 (2010).
29. Duchêne, D. A., Duchêne, S., Stiller, J., Heller, R. & Ho, S. Y. W. ClockstaRX: testing molecular clock hypotheses with genomic data. *Genome Biol. Evol.* **16**, evae064 (2024).
30. Ohta, T. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992).
31. Bergeron, L. A. et al. Evolution of the germline mutation rate across vertebrates. *Nature* **615**, 285–291 (2023).
32. Jetz, W., Sekercioglu, C. H. & Böhning-Gaese, K. The worldwide variation in avian clutch size across species and space. *PLoS Biol.* **6**, 2650–2657 (2008).
33. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18**, 225–239 (1982).
34. Welch, J. J., Bininda-Emonds, O. R. P. & Bromham, L. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* **8**, 53 (2008).
35. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
36. Germain, R. R. et al. Species-specific traits mediate avian demographic responses under past climate change. *Nat. Ecol. Evol.* **7**, 862–872 (2023).
37. Claramunt, S., Derryberry, E. P., Remsen, J. V. Jr & Brumfield, R. T. High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proc. Biol. Sci.* **279**, 1567–1574 (2012).
38. Ksepka, D. T. et al. Tempo and pattern of avian brain size evolution. *Curr. Biol.* **30**, 2026–2036.e3 (2020).
39. Harries, P. J. & Knorr, P. O. What does the 'Lilliput effect' mean? *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **284**, 4–10 (2009).
40. Brown, J. W., Rest, J. S., García-Moreno, J., Sorenson, M. D. & Mindell, D. P. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol.* **6**, 6 (2008).
41. Waters, P. D. et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl Acad. Sci. USA* **118**, e2112494118 (2021).
42. Liu, J. et al. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Res.* **31**, 497–511 (2021).
43. Germain, R. R. et al. Changes in the functional diversity of modern bird species over the last million years. *Proc. Natl Acad. Sci. USA* **120**, e2201945119 (2023).
44. Bryant, D. & Hahn, M. W. in *Phylogenetics in the Genomic Era* (eds Scornavacca, C., Delsuc, F. & Galtier, N.) 3.4:1–3.4:23 (HAL, 2020).
45. Xu, Y. et al. Ecological predictors of interspecific variation in bird bill and leg lengths on a global scale. *Proc. Biol. Sci.* **290**, 20231387 (2023).
46. Tobias, J. A. et al. AVONET: morphological, ecological and geographical data for all birds. *Ecol. Lett.* **25**, 581–597 (2022).
47. Shakya, S. B., Edwards, S. V. & Sackton, T. B. Convergent evolution of noncoding elements associated with short tarsus length in birds. Preprint at *bioRxiv* https://doi.org/10.1101/2024.04.30.591925 (2024).
48. Rezatabar, S. et al. RAS/MAPK signaling functions in oxidative stress, DNA damage response and cancer progression. *J. Cell. Physiol.* **234**, 14951–14965 (2019).

49. Cannell, I. G. et al. p38 MAPK/MK2-mediated induction of miR-34c following DNA damage prevents Myc-dependent DNA replication. *Proc. Natl Acad. Sci. USA* **107**, 5375–5380 (2010).
50. Kim, J. et al. β-Arrestin 1 regulates β2-adrenergic receptor-mediated skeletal muscle hypertrophy and contractility. *Skelet. Muscle* **8**, 39 (2018).
51. Najafi, A., Sequeira, V., Kuster, D. W. D. & van der Velden, J. β-Adrenergic receptor signalling and its functional consequences in the diseased heart. *Eur. J. Clin. Invest.* **46**, 362–374 (2016).
52. Lynch, M. & Kewalramani, A. Messenger RNA surveillance and the evolutionary proliferation of introns. *Mol. Biol. Evol.* **20**, 563–571 (2003).
53. Wolin, S. L. & Maquat, L. E. Cellular RNA surveillance in health and disease. *Science* **366**, 822–827 (2019).
54. Singh, P., Saha, U., Paira, S. & Das, B. Nuclear mRNA surveillance mechanisms: function and links to human disease. *J. Mol. Biol.* **430**, 1993–2013 (2018).
55. Pan, S. et al. Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. *Nat. Commun.* **10**, 2756 (2019).
56. Oliveros, C. H. et al. Earth history and the passerine superradiation. *Proc. Natl Acad. Sci. USA* **116**, 7916–7925 (2019).

Methods

Family-level phylogenomic data for avian families were collected from the database of the B10K consortium[5]. These included 63,430 evenly spaced intergenic loci, each with a length of 1 kb, as well as the coding regions of 15,093 orthologous genes. For each data type, we selected the representative of each avian family for which the greatest nucleotide completeness was available, providing a sample of 218 sampled tips. Other samples within families were excluded to minimize node-density effects that can mislead family-level inferences. To make reliable estimates of branch lengths in expected synonymous (d_S) and non-synonymous (d_N) substitutions, we filtered coding regions to exclude codons where any of the three positions was missing for more than 50% of the taxa and where the most common amino acid occurred in less than 50% of taxa. In each region, we also excluded taxa for which less than 30% of the nucleotides were available. Regions with fewer than three taxa remaining after this step were excluded from further analyses, producing a final dataset with 63,174 intergenic and 14,849 coding regions.

To infer evolutionary rates for each genomic region, we used phylogenetic estimates of branch lengths on a fixed tree topology. We forced the tree topology for each region to conform to the family-level relationships reported by the B10K consortium, using the species tree inferred from intergenic regions under the multi-species coalescent[6]. Phylogenetic branch lengths from intergenic regions were estimated using the best-fitting model from the GTR + F + R[57] family in IQ-TREE (v2.1.2)[58]. Coding regions were used for inference of d_N and d_S branch lengths. We first estimated distance matrices for d_N and d_S using maximum likelihood in PAML using the models HKY + Γ and F3 × 4 for nucleotide and codon substitutions, respectively[59]. For each gene, maximum-likelihood d_N and d_S distance matrices were then transformed to the lengths of branches in the species tree using ERaBLE[60].

We used ClockstaRX[29] to extract branch lengths for each locus at each species-tree branch, which we then converted into rates using the family-level time-tree estimate (Supplementary Data 5). To avoid adding error due to lack of signal in the molecular data, we excluded branches with estimated lengths of less than 5×10^{-6} substitutions per site. Analyses in which each locus has a free topology are also possible in ClockstaRX, but led to a vast loss of data, including the majority of genes at deep branches, as expected in ancient systems with large amounts of incomplete lineage sorting and compounded phylogenetic error. These analyses were not considered further. Analyses using a forced topology must acknowledge that although the data might show patterns of evolutionary rate variation when the gene trees are summarized in this way, these patterns might be due to processes that occurred on nearby branches in gene trees with discordant topologies[40]. Nonetheless, the species tree inferred from whole-genome data provides a reasonable hypothesis for the evolutionary history that any given locus has followed.

To explore whether specific species traits explained molecular rates, we collected trait data from a series of databases and primary literature (see Supplementary Data 1 for a description of each variable). We collected a total of 23 traits with complete or near-complete sampling across bird species, including 5 life-history and demography traits (clutch size, longevity, generation length, abundance and developmental mode from altricial to precocial), 7 morphological traits (beak size, tarsus length, wing length, hand-wing index, tail length, body mass and brain mass), 5 ecological traits (habitat, habitat density, trophic level, migration and primary lifestyle), 4 geographical traits (absolute centroid latitude, latitudinal span, centroid longitude and range size) and 2 environmental traits (mean annual temperature and annual precipitation). The variable of developmental mode comes from previous work[61], and is the first principal component explaining around 60% of variance in a set of traits likely to discriminate altricial and precocial species. We calculated beak size as the first principal component of beak length, width and height, explaining 89% of variance. Missing values for wing length and developmental mode were added using phylogenetic imputation assuming a Brownian motion model for trait evolution, as implemented in Rphylopars[62]. Across variables, the family mean and mode were used for continuous and discrete traits, respectively. Continuous variables were verified to follow logarithmic scaling. This led us to perform natural log transformations of most variables, with the exception of temperature and longitude.

Subsequent regression analyses were performed on two treatments of trait data, using the trait value of the single species sampled for genomes, or taking the mean value across all species in the corresponding family. Results from both types of analyses are provided (Supplementary Data 3). A focus on family averages reflects the fact that estimates of molecular rates correspond to the average value along the evolutionary path between the sampled individual up to the common ancestor with its sister family, following existing literature on evolutionary rates[63]. This approach addresses the potential effect of internal nodes when estimating evolutionary rate, but does not explicitly consider the effect of punctuated molecular change on diversification events[64,65]. In agreement with this signal, models explaining molecular rates using the family-average trait data of family means led to overall higher regression R^2 values than those using trait data from the species sampled.

Bayesian and frequentist regressions were used for assessing whether molecular evolutionary rates were explained by the set of traits sampled. We defined a parametric model for each of the estimates of mean rates in each family (in all d_N, d_S, ω and intergenic regions) as response variables. Using a Bayesian mixed-effects linear modelling framework, parameter inference was performed via the no U-turn sampler[66,67] as implemented in the package brms[68]. By using a Bayesian model, we were able to include a shrinkage prior and mitigate finite sample sizes more adequately than through a frequentist framework. In addition, the framework naturally incorporates uncertainty around significant coefficients via credible intervals, instead of *P* values. Using log-transformed molecular rate estimates $y \in \mathbb{R}^n$, we assumed a Gaussian likelihood,

$$y_i \sim \text{Normal}(\mu_i, \sigma_e^2)$$

The linear predictor μ is then assumed to have a mixed-effect structure. The fixed effects included each of the 19 continuous traits collected, forming a matrix $X \in \mathbb{R}^{n \times m}$ with an intercept included in X. We normalized covariates by centring and scaling by twice their standard deviation[69]. Random effects were then included for three categorical variables collected: the habitat ($u_{1,i}$), trophic level ($u_{2,i}$) and primary lifestyle ($u_{3,i}$), modelled as Gaussian random effects with independent standard deviation parameters:

$$u_{1,i} \sim \text{Normal}(0, \sigma_1^2)$$
$$u_{2,i} \sim \text{Normal}(0, \sigma_2^2)$$
$$u_{3,i} \sim \text{Normal}(0, \sigma_3^2)$$

The phylogenetic effect was captured as a multivariate Gaussian random effect:

$$y_i \sim \text{Normal}(0, \sigma_y^2 \Sigma)$$

where Σ is a variance–covariance matrix from our fixed time-tree assuming a Brownian motion model of trait evolution, and then scaled during analysis by the effect of phylogeny on traits, using the σ_y^2 parameter. To add posterior convergence of the scaling parameter, the original Σ matrix was scaled to a correlation matrix with time of most recent common ancestry of 1. The final linear predictor was therefore:

$$\mu_i = \beta X^T + u_{1,i} + u_{2,i} + u_{3,i} + \gamma_i$$

Prior distributions were chosen such that they ensured shrinkage of parameters to zero, and thereby assuming a null model where the fixed effects and random effects have no effect on the response variable.

$$\beta \sim \text{Normal}(0, 5), \text{ and } \sigma_e, \sigma_1, \sigma_2, \sigma_3, \sigma_4 \sim \text{Student}T_{\upsilon=3}(0, 5)$$

Parameter sampling using Markov chain Monte Carlo was run across four chains for 104 iterations each. Posterior convergence to a stationary distribution was assessed via the Rhat statistic[70] as well as via visual inspection of the Markov chain Monte Carlo trace.

To complement Bayesian regression analyses, we implemented the frequentist counterparts and only interpret results that are consistent between the two approaches. Models with each molecular rate as response and including trait variables as covariates were implemented using a Pagel's lambda model of trait evolution along the phylogeny[71], implemented as phylogenetic regression in the package phylolm[72]. In frequentist analysis, the variable of trophic level was not included due to imbalance across small numbers of factor levels that complicates parameter optimization. The distribution of trait data was visualized across branches in the avian phylogeny using fast maximum-likelihood ancestral state reconstruction as implemented in phytools[73]. Phylogenetic data were processed with the support of phangorn[74].

The variables explaining evolutionary rates were also tested at the level of gene rates. For each gene, we calculated the mean of each type of rate (d_N, d_S, ω and intergenic regions) and mean GC proportion across all families. We then calculated the location along chromosomes as the proportional distance from the end of each chromosome. Genes with a distance of 0 are the closest to a chromosome ending, whereas genes with a distance of 1 indicate the closest to the chromosome middle. Bayesian and frequentist multiple regressions were used to assess whether each of the molecular rate types was explained by GC content, gene length and gene location. We also included in these models the two-way interactions between these variables. To address the possible effect of model misspecification on these analyses, we used matched-pairs tests of homogeneity[75], as implemented in IQ-TREE2 (ref. 76), and parametric regression to assess the goodness of fit[77] and base compositional non-stationarity[78,79], as implemented in PhyloMAd[80]. We added the resulting statistics from each of these tests as covariates, and did not find that including any of them had a substantial effect on the results. We report the results of models including tests of marginal symmetry as covariates accounting for possible phylogenetic model mis-specification, and make all model mis-specification metrics available online (Supplementary Data 2).

Random forests provide an alternative for comparative analysis, identifying predictors while capturing the possible interactions and non-linearities. We examined the importance scores of traits as predictors in a random forest framework. We included 500-tree forests with randomly selected sets of fivefold cross-validation resampling with five iterations for hyperparameter tuning, and using the root mean squared error as a metric for regression performance (Supplementary Data 6), as implemented in caret[81]. Although random forests can capture additional predictive links from those shown by linear regression, they do not incorporate non-independence across samples arising from phylogenetic relatedness, such that they can also introduce bias.

We tested whether the mean molecular rate of each chromosome was different from the mean expectation from loci across the genome. To do this, we performed permutations per chromosome, involving a random draw of 1,000 sets of loci of identical size to the chromosome. We performed this permutation for each chromosome and calculated the *Z*-score of the mean empirical chromosome rate with respect to the permutation distribution.

We performed rate decomposition analyses to identify the lineages and genes that dominate molecular rate variation. This was done in ClockstaRX[29] for each set of loci independently (d_N, d_S, ω and intergenic regions). To allow computational efficiency when examining intergenic regions, we analysed three random samples of 10,000 loci, and confirmed that results were consistent across each sample. The basic method of decomposition collects rate estimates in a matrix with rows representing loci and columns representing the full set of lineages (branches) in the species-tree estimate. This data structure is then decomposed for identifying the main axes of variation using principal component analysis. The software performs a permutation of the data matrix, summarizing the eigenvalues at each component using two test statistics ψ and φ[82]. These statistics, together with lineage loadings on each principal component, allow for tests of whether each principal component and lineage loading contributes more variation than expected from a random sample across the data[82] ($\alpha = 0.01$). A primary assumption under this method is that the species tree is accurate and largely concordant with gene trees. As this assumption is often violated due to phylogenetic error and gene-tree discordance, all inferences under this method refer to either the branch in question or neighbouring branches from similar phylogenetic trees, where substitutions might have actually occurred[83].

We explored whether lineage contribution to variance on rate axes was associated with any traits in our comprehensive set, such that a principal component might be driven by the ecology or life history of families. To do this, we first extracted the loadings of families (terminal branches) on each of the principal components that significantly explained variation in rates. We then proceeded to perform simple and multiple phylogenetic regressions where traits were tested to explain principal component loadings, as performed for mean genome-wide rate metrics.

We then explored whether principal components are associated with any particular metabolic pathway or gene function. We extracted the 20% of loci with the maximum and minimum values at the principal components that significantly explained variation in rates. We used gene set enrichment analysis to evaluate any overrepresentation of metabolic functions of gene products independently in loci with extreme high and low rates on each significant rate axis. Each overrepresentation analysis is equivalent to a one-sided Fisher's exact test of the disproportionate presence of a set of genes inside another set[84]. Gene identities were inferred using the best blastn match[85] and used as input for testing enrichment of KEGG terms using clusterProfiler[86]. Genes in pathways were visualized and interpreted using pathview[87]. Bonferroni correction was used for *P* values of significant terms.

Overrepresentation of chromosomes in the 20% extremes of significant rate axes (principal components) was evaluated using binomial tests. Therefore, the null hypothesis was a Bernoulli experiment[88] in which the proportion of loci in a given chromosome is expected to be equal in the 20% maximum and minimum loci of each axis. This test was performed for each chromosome and each significant principal component, with *P* values corrected using false discovery rates within principal components.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genomic assemblies and annotations from avian families originally from Feng et al.[5] are available in the NCBI Sequence Read Archive and GenBank under accession ID PRJNA545868. Gene trees and species tree data are available via the Figshare data availability repository for this publication[89] (https://doi.org/10.6084/m9.figshare.27323229).

Code availability

Trait and molecular data and analysis scripts[90] are available as Supplementary Data 1–6 (https://doi.org/10.5281/zenodo.14848360). The repository also contains instructions for extraction and additional exploration of results.

57. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
58. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
59. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
60. Binet, M., Gascuel, O., Scornavacca, C., Douzery, E. J. P. & Pardi, F. Fast and accurate branch lengths estimation for phylogenomic trees. *BMC Bioinformatics* **17**, 23 (2016).
61. Ducatez, S. & Field, D. J. Disentangling the avian altricial-precocial spectrum: quantitative assessment of developmental mode, phylogenetic signal, and dimensionality. *Evolution* **75**, 2717–2735 (2021).
62. Goolsby, E. W., Bruggeman, J. & Ané, C. Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods Ecol. Evol.* **8**, 22–27 (2017).
63. Lanfear, R. et al. Taller plants have lower rates of molecular evolution. *Nat. Commun.* **4**, 1879 (2013).
64. Venditti, C. & Pagel, M. Speciation as an active force in promoting genetic evolution. *Trends Ecol. Evol.* **25**, 14–20 (2010).
65. Duchêne, D. A., Hua, X. & Bromham, L. Phylogenetic estimates of diversification rate are affected by molecular rate variation. *J. Evol. Biol.* **30**, 1884–1897 (2017).
66. Hoffman, M. & Gelman, A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2011).
67. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1 (2017).
68. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
69. Gelman, A. Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* **27**, 2865–2873 (2008).
70. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian Data Analysis* 2nd edn (CRC Press, 2003).
71. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
72. Ho, L. S. T. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
73. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
74. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2010).
75. Jermiin, L. S., Jayaswal, V., Ababneh, F. M. & Robinson, J. Identifying optimal models of evolution. *Methods Mol. Biol.* **1525**, 379–420 (2017).
76. Naser-Khdour, S., Minh, B. Q., Zhang, W., Stone, E. A. & Lanfear, R. The prevalence and impact of model violations in phylogenetic analysis. *Genome Biol. Evol.* **11**, 3341–3352 (2019).
77. Goldman, N. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**, 182–198 (1993).
78. Duchêne, D. A., Duchêne, S. & Ho, S. Y. W. New statistical criteria detect phylogenetic bias caused by compositional heterogeneity. *Mol. Biol. Evol.* **34**, 1529–1534 (2017).
79. Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).
80. Duchêne, D. A., Duchêne, S. & Ho, S. Y. W. PhyloMAd: efficient assessment of phylogenomic model adequacy. *Bioinformatics* **34**, 2300–2301 (2018).
81. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
82. Björklund, M. Be careful with your principal components. *Evolution* **73**, 2151–2158 (2019).
83. Mendes, F. K. & Hahn, M. W. Gene tree discordance causes apparent substitution rate variation. *Syst. Biol.* **65**, 711–721 (2016).
84. Boyle, E. I. et al. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715 (2004).
85. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
86. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
87. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
88. Clopper, C. J. & Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413 (1934).
89. Duchêne, D. A. et al. Drivers of avian genomic change revealed by evolutionary rate decomposition. *figshare* https://doi.org/10.6084/m9.figshare.27323229 (2025).
90. Duchene, D. duchene/b10k_family_rates: v1.0. *Zenodo* https://doi.org/10.5281/zenodo.14848361 (2025).

Acknowledgements This work was supported by grants from the Carlsbergfondet of Denmark (grant CF18-0223), an European Research Executive Agency Marie Sklodowska-Curie Action (H2020-MSCA-IF-2019-883832) and an Emerging Data Science Investigator Award from the Novo Nordisk Foundation (NNF23OC0084647) to D.A.D. We also acknowledge support from the Danish National Research Fund 'Center for Evolutionary Hologenomics' award to M.T.P.G. (DNRF143). S.F. was supported by a National Natural Science Foundation of China grant (32170626) and a National Key Research and Development Program of China grant (2023YFA1800500). This work was also supported by a research grant (42153) from VILLUM FONDEN to J.S. The international collaboration was supported by a Partnership Collaboration Award from the University of Sydney and University of Copenhagen to D.A.D., M.T.P.G., G.Z. and S.Y.W.H. S.B. acknowledges support from the Novo Nordisk Foundation via The Novo Nordisk Young Investigator Award (NNF20OC0059309), from The Eric and Wendy Schmidt Fund For Strategic Innovation via the Schmidt Polymath Award (G-22-63345), and the National Institute for Health and Care Research (NIHR) Health Protection Research Unit in Modelling and Health Economics (grant code NIHR200908). Images of metabolic pathways are interpretations from the R package pathview. Silhouettes of birds and bird parts are originals or modifications of images dedicated to the public domain under the CC0 1.0 Universal Public Domain Dedication and Public Domain Mark 1.0 licenses from PhyloPic (https://phylopic.org).

Author contributions D.A.D., S.Y.W.H., A.-A.C., J.A.T., M.T.P.G., G.Z. and S.F. conceived and designed the study. D.A.D., A.-A.C., J.S., J.A.T. and J.Y. contributed to data collection and preparation. D.A.D., A.-A.C. and S.B. performed the genomic analyses and regression analyses. D.A.D., S.Y.W.H., A.-A.C., M.I.-C., J.A.T., G.Z. and S.F. contributed to the data interpretation. D.A.D., S.Y.W.H., M.I.-C. and S.B. wrote the manuscript with input from all co-authors.

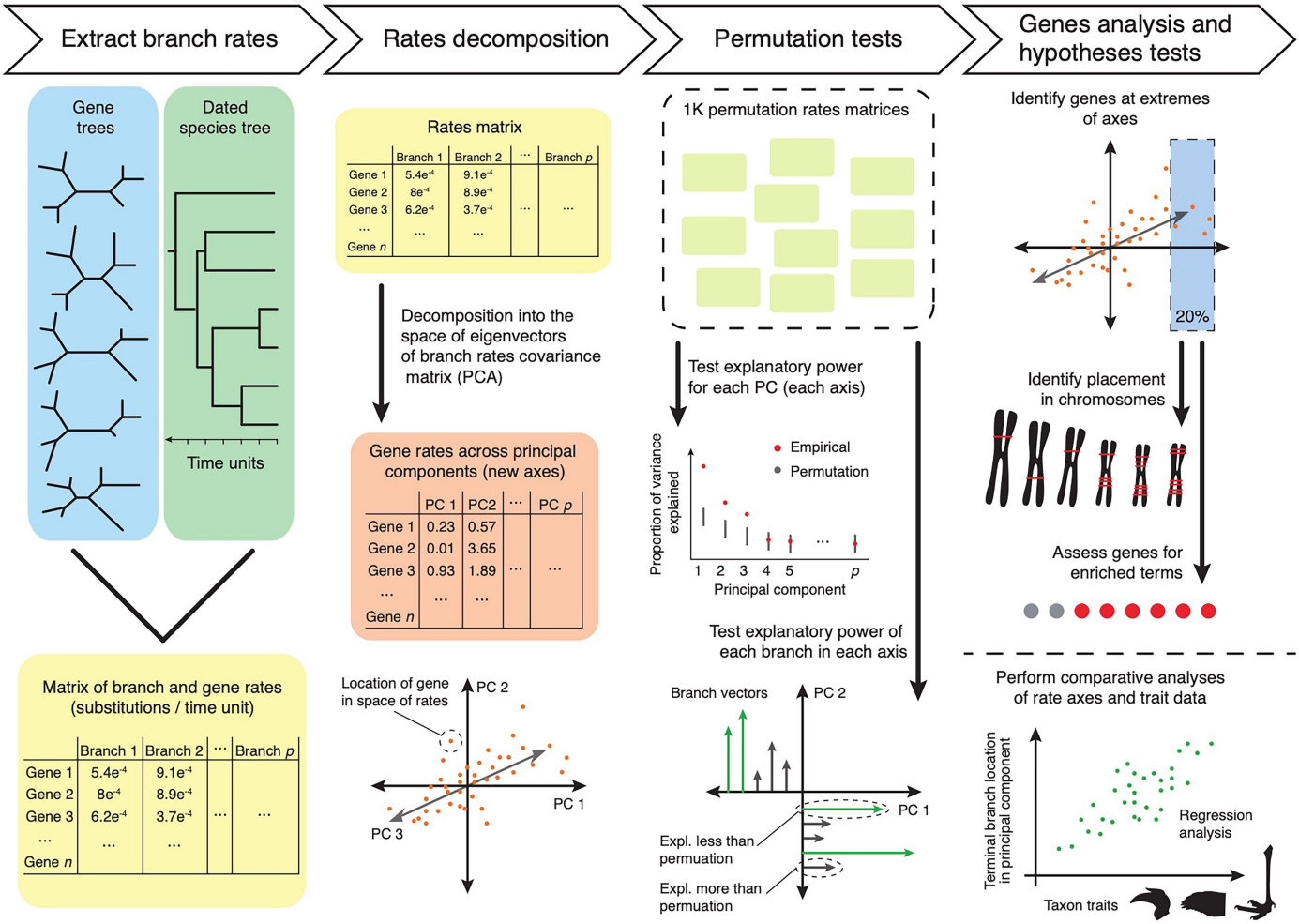
Competing interests M.T.P.G. serves on the Science Advisory Board of Colossal Laboratories & Biosciences. The other authors declare no competing interests.

Additional information

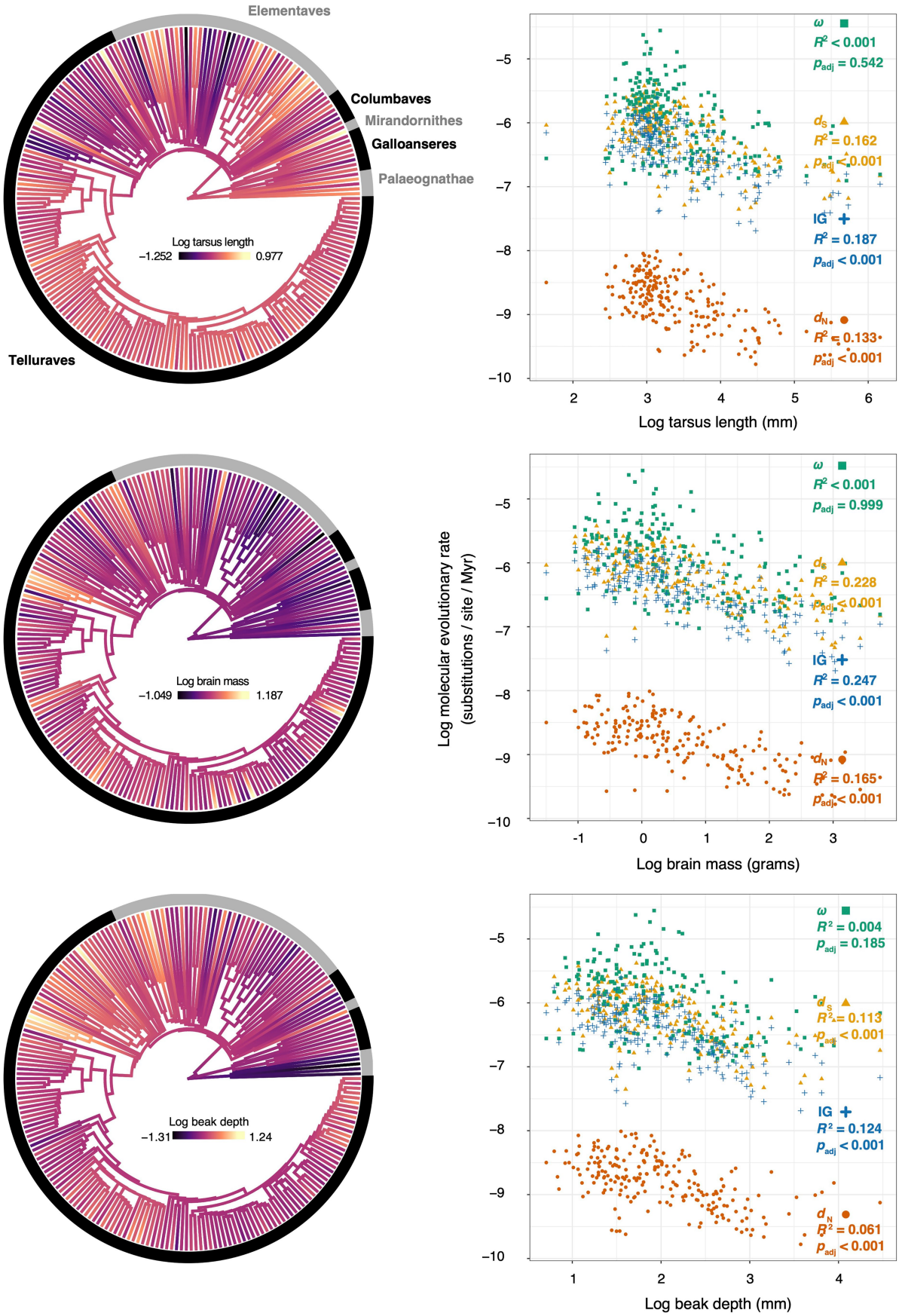
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-08777-7.

Correspondence and requests for materials should be addressed to David A. Duchêne.

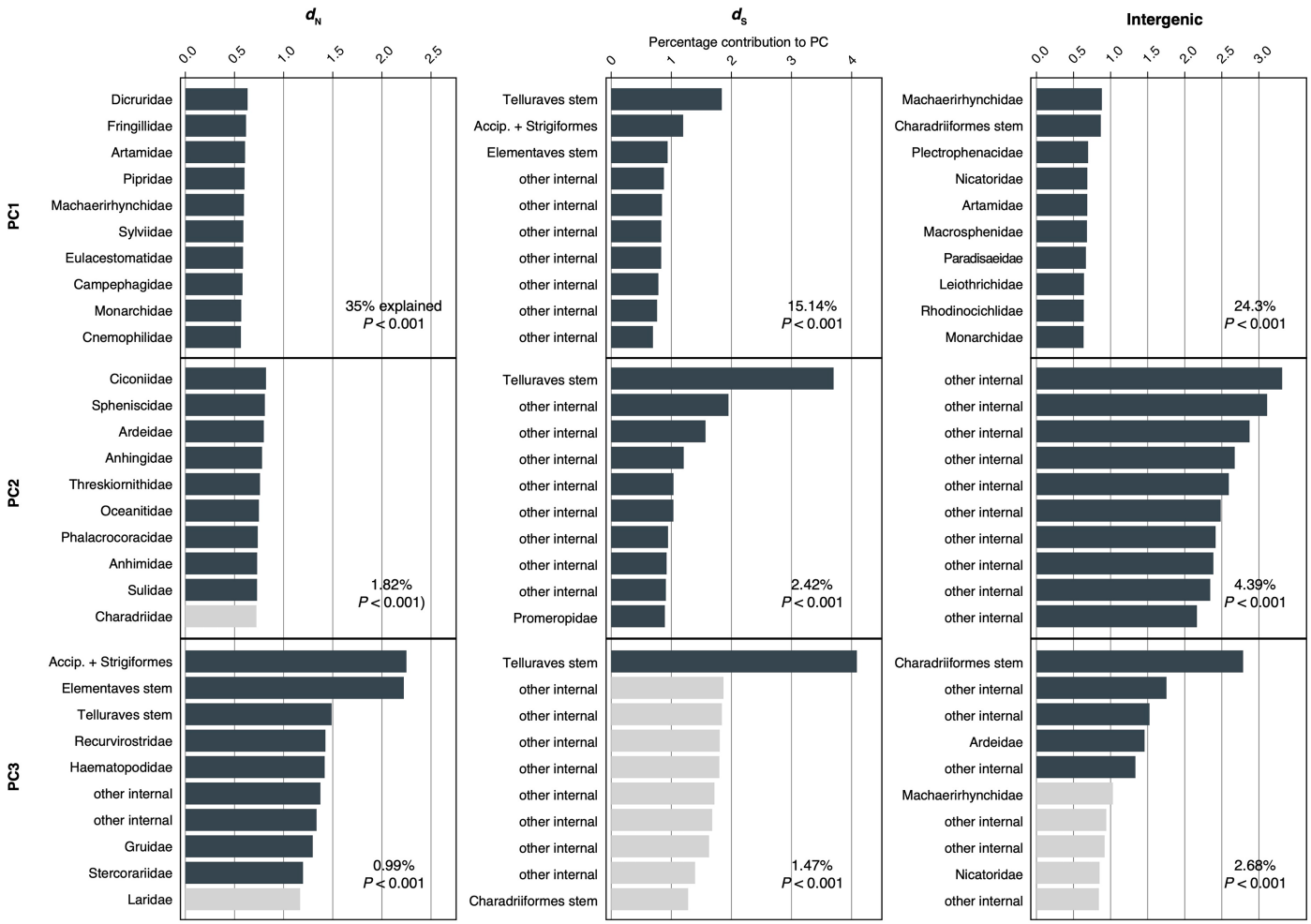
Peer review information *Nature* thanks Jacob Berv and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



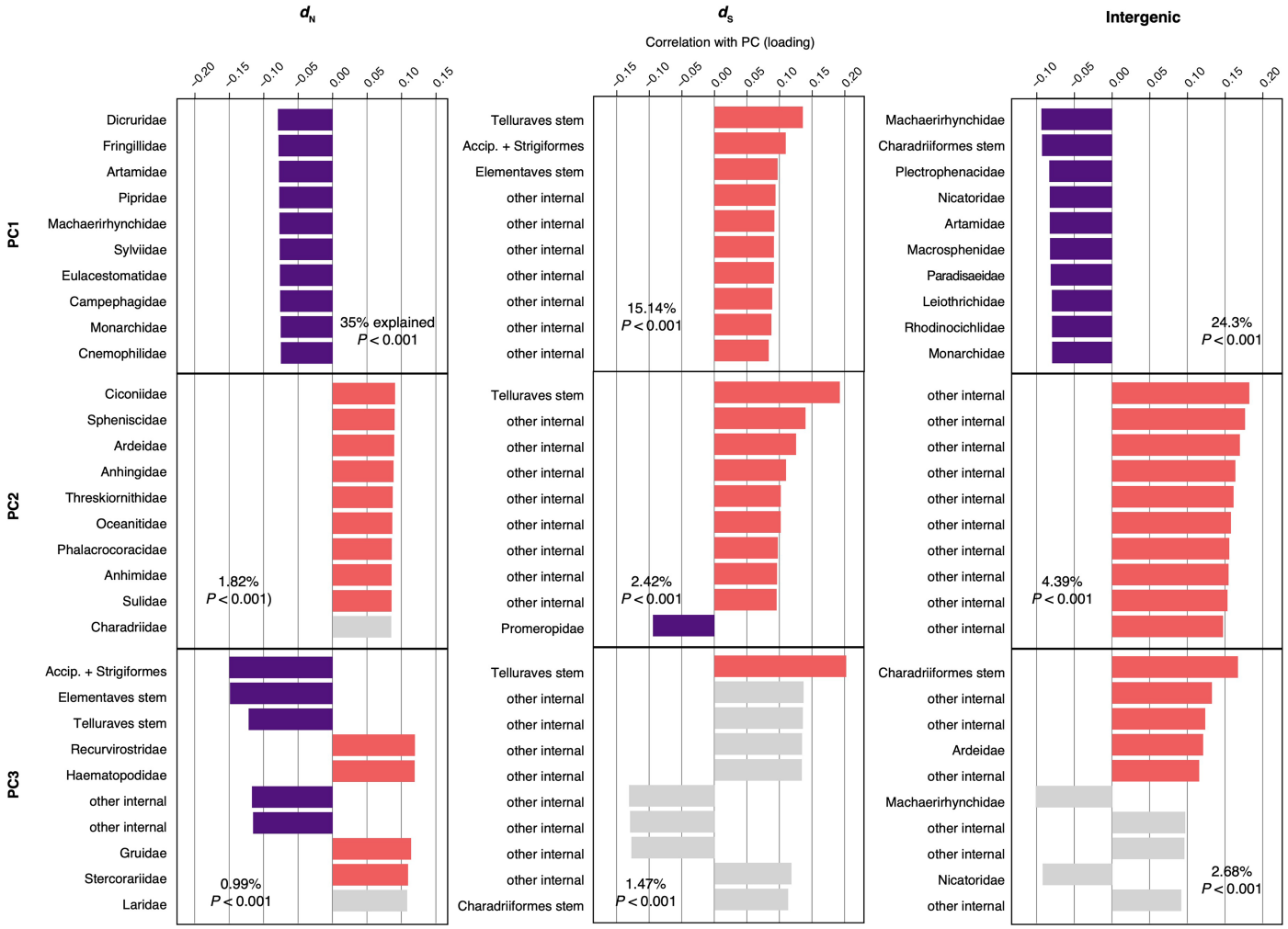
Extended Data Fig. 1 | Overview of rate decomposition analyses. Data from gene trees and the species tree are the starting point, and are used as input in ClockstaRX to produce a rates matrix and perform rates decomposition and permutation analyses. The output can be used in a range of other downstream analyses, such as of gene enrichment and phylogenetic regression. This figure was adapted from the original description of the method[29]. Silhouettes were adapted from PhyloPic (https://phylopic.org) under CC0 1.0 (*Buceros rhinoceros* and *Accipiter nisus*) and Public Domain Mark 1.0 (*Cariama cristata*) licences.



Extended Data Fig. 2 | Additional relationships between traits and molecular rates. Various treatments of the data led to variables of lower effect emerging as significant. Values and colours in phylogenetic trees are mass-regressed residuals as correction for visualization purposes. See effect sizes in Fig. 1 and Supplementary Data 3.



Extended Data Fig. 3 | Taxon contribution to principal components. Rate decomposition for each molecular data set produces distinct results, yet consistently show a strong contribution of early neoavian branches (e.g., stem Telluraves). Results across all principal components can be found in the data repository listed in Data Availability.



Extended Data Fig. 4 | Correlation of taxa with each principal component. All branches had the same correlation sign with the first principal component of each data set, indicating that the principal component describes overall gene rate (rather than lineage-specific rates). Subsequent principal components contain more nuanced results, but frequently have large numbers of branches that do not explain variance significantly more than under permutation of the data.

Extended Data Fig. 5 | Branch loadings across principal components. Colours indicate the full range of loadings, yet first principal components consistently had the same loading sign across all branches. Branches shown in grey did not explain greater variance in the principal component than under permutation of the data.

nature portfolio

Corresponding author(s): David A. Duchene

Last updated by author(s): Feb 11, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All data are freely available and were collected by hand from previous publications.
Data analysis	All custom code for data analysis unique to this study is available at https://github.com/duchene/avian_family_rates.git. Other software used for data analysis includes the following, including the specific use in this study: IQ-TREE v2.1.2 - Phylogenetic analysis PAML v4.9 - Inference of pairwise distances using codon models ERaBLE v1.0 - Inference of branch lengths using pairwise distance data. ClockstaRX v1.1 - Rate decomposition (described in Extended Data Figure 1). phylolm v2.6.2 - Phylogenetic regression analyses. brms v2.2 - Bayesian regression analyses. caret v6.0 - Random forests analyses. phytools v1.5.1 - Phylogenetic data visualization and ancestral state reconstruction. phangorn v2.11.1 - Phylogenetic data handling. clusterProfiler v4.6.2 - Gene set enrichment analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Genomic assemblies and annotations form avian families orginally from Feng et al.14 are available in the NCBI SRA and GenBank under accession PRJNA545868. Gene trees and species tree data are available via https://sid.erda.dk/cgi-sid/ls.py?share_id=hVS3naBtJ6 while trait data and analysis scripts are available at https://github.com/duchene/avian_family_rates.git and will receive a permanent DOI upon acceptance of the article. This latter repository also contains copies of supplementary files and instructions on extended visualisation of results.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No sex or gender data were involved in this study.
Reporting on race, ethnicity, or other socially relevant groupings	No data on race, ethnicity, or social grouping were involved in this study.
Population characteristics	The data in this study does not involve human participants.
Recruitment	No human participants were recruited in this study.
Ethics oversight	An ethics approval was not applicable to this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Genome-scale sequence alignments were used for inferring molecular evolutionary rates across branches of the family-level avian tree of life. These data were used to test the relationship between molecular rates and a comprehensive set of biological traits. Rates were also examined using PCA as described in Extended Data Figure 1, which we term 'rate decomposition' and includes permutation tests on principal components and their loadings. Genes at extremes of components were examined for gene term enrichment and over-representation of genes in particular chromosomes. Rates were also compared across chromosomes using permutation tests.
Research sample	The taxonomic sample includes 218 species representing extant avian families. The genomic sample includes 63,430 evenly spaced intergenic loci each with 1 kbp in length, as well as the coding regions of 15,093 orthologous genes, before undergoing the filtering described in the methods. The trait data includes 29 traits with complete or near-complete sampling across bird species, described in detail in the methods and extensive detail in the Supplementary File 1.
Sampling strategy	Data were drawn from the 363 avian genomes published by Feng et al. (2020, Nature). We selected the representatives with the greatest gene-completeness from each of the 218 families in those data.
Data collection	The data used are accessible online as described in the data availability statement.
Timing and spatial scale	Global scale, with timing ranging the evolution of the extant members of class Aves.
Data exclusions	As also stated in the methods, to make reliable estimates of branch lengths in expected synonymous (dS) and non-synonymous (dN) substitutions, coding regions were filtered to exclude codons where any of the three positions was missing for >50% of the taxa, and where the most common amino acid occurred in <50% of taxa. In each region, we also excluded taxa for which <30% of the nucleotides were available. Regions with fewer than three taxa remaining after this step were also excluded from further analyses,

producing a final data set with 63,174 intergenic and 14,849 coding regions.

Reproducibility	The description of procedures in the Methods section and in the code available is guided at maximizing detail and annotation, allowing exact reproducibility of all analyses. Similarly, the data are freely available at various stages of analysis (molecular data, inferred gene trees, rates and traits data, PCA outputs, and full permutation and regression outputs).
Randomization	Not applicable.
Blinding	Not applicable.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable.
Novel plant genotypes	Not applicable.
Authentication	Not applicable.