nature communications

Article

https://doi.org/10.1038/s41467-025-64134-2

Comparative genomics sheds light on mammalian and avian gene regulation and phenotypic evolution

Received: 2 August 2024

Accepted: 4 September 2025

Published online: 14 October 2025

Check for updates

Anabella P. Trigila[1,2], Paula Beati[1,2], Dante Montini[1], Camila Jovicic[1], Paula de la Vega[1] & Lucía F. Franchini[1]

The discovery of the genetic mechanisms underlying the emergence of phenotypic novelties is a cornerstone of evolutionary biology. To identify sequences in vertebrate genomes leading to the emergence of clade defining traits, we uncover genomic regions conserved across vertebrates that accumulated substitutions in a faster than neutral rate in the Avian or in the Mammalian basal lineages. We identify 2888 and 3476 noncoding avian and mammals accelerated regions, respectively, accumulating in key developmental genes in each lineage. Interestingly, we find that the neuronal transcription factor *NPAS3*, that carries the largest number of human accelerated regions, also accumulates the largest amount of noncoding mammals accelerated regions. We also detect that four *NPAS3* noncoding mammals accelerated regions overlap previously identified human accelerated regions. Thus, we show that noncoding regions of particular large-effect genes are repeatedly targets of accelerated evolution, suggesting the existence of evolutionary hotspots underlying phenotypic innovation in different lineages.

A central goal of evolutionary biology is to understand the genetic bases underlying the emergence of morphological novelties. However, the genetic mechanisms driving the emergence of macroevolutionary novelties in vertebrate groups such as birds and mammals are still unknown.

Previous studies have attempted to elucidate the evolutionary forces underlying the emergence of some phenotypic novelties that characterize mammals and that acted on gene coding regions, including dentition[1], hair development[2], milk generation[3], and high-frequency hearing[4–6]. A common feature of these studies is that they focused on gene groups already known to be involved in a specific trait. This approach may be relatively straightforward for certain phenotypic traits, where tissue transcriptomes can be accessed, but it shows greater challenges for more complex traits, such as homeothermy or parental care. In this sense, it is useful to incorporate more generalist approaches and to comparatively scan entire genomes, identifying signatures of evolutionary forces without focusing exclusively on a particular set of genes. Moreover, the use of comparative genomics enables the study of noncoding, potentially regulatory, regions. One way to identify regulatory regions that may play a role in the emergence of macroevolutionary novelties is to look for conservation signals in noncoding regions of the mammalian and avian genomes. This approach assumes that DNA sequences involved in gene regulation have remained significantly more conserved than non-functional DNA, a pattern observable across a broad phylogenetic spectrum. In addition, comparative genomics methods have been developed that take into account the lineage-specific rates of DNA substitution. These methods compare observed rates with the expected rate across the evolutionary tree, allowing to identify coding and noncoding regions that are under lineage-specific accelerated evolution.[7,8]. Most studies aiming to identify accelerated elements have focused on specific branches of the mammalian tree, including particular species or groups (humans, bats, among others) or certain lineages of birds (flightless birds). For instance, bat accelerated regions

[1]Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina. [2]These authors contributed equally: Anabella P. Trigila, Paula Beati. e-mail: franchini.lucia@gmail.com

(BARs) are 2796 conserved sequences in vertebrates that changed significantly in the common ancestor of bats. These regions could have influenced limb development[9,10]. Similarly, ~3000 human accelerated regions (HARs) have been identified and extensively studied for their roles in brain development and their potential involvement in the etiology of neurological diseases[8,11]. In another approach, accelerated elements in the terminal branches of elephants, bats, dolphins, killer whales, naked rats and squirrels were identified[12]. Additionally, accelerated elements were described in the basal branch of therian (placental and marsupial) mammals (known as TSARs). Some of these elements are located in genes associated with phenotypic features unique to this group, such as hormonal control of milk ejection, uterine contractions, and visual processing[13]. More recently, the Zoonomia project[14], an international collaboration aimed at discovering the genomic basis of shared and specialized traits in mammals, has produced a comprehensive 240-species genome alignment. This alignment includes only regions shared across eutherian mammals and was used to generate a new dataset of human accelerated regions (zooHARs)[15]. However, the study did not explore regions with acceleration signatures in other mammalian branches or in the lineage leading to mammals.

Concerning birds' phenotypic evolution, another group analyzed 48 avian genomes and identified millions of avian-specific highly conserved elements (ASHCEs) residing predominantly in noncoding regions. Notably, one ASHCE was found to be associated with the *Sim1* gene and might be linked to the evolution and development of flight feathers[16]. Moreover, a comparative analysis of 11 genomes of paleognathous birds, including an extinct moa, identified 2355 independent accelerations along lineages of flightless paleognaths, with apparent functional consequences for driving gene expression in the developing forelimb[17]. More recently, in April 2024, the Bird 10,000 Genomes (B10K) Project, an initiative to generate representative draft genome sequences from all extant bird species, reported the completed analysis of 363 bird genomes, including 267 newly sequenced genomes[18]. This resource enabled the reconstruction of a new avian family tree and identified coding gene losses in different bird lineages[19]. However, the alignments produced as part of this project include only avian species and exclude other vertebrates, which could serve as background species for comparative analyses in the lineage leading to the appearance of birds.

Despite significant advances, large-scale genomic studies aimed at identifying the molecular basis of phenotypic traits that have evolved in parallel across vertebrate lineages, such as mammals and birds, remain limited. Extant birds and mammals share several similar traits, including: homeothermy and insulation from the environment via feathers or hair[20–23], similar cardiovascular system[24–27], blood with small-sized erythrocytes and leukocytes coupled with higher blood pressure[28,29], complex behaviors such as offspring care[23,30], improved hearing ability[31], vocal communication[32] and high basal metabolism[33].

Thus, little is known about the genetic changes underlying the evolution of phenotypic traits that define macroevolutionary leaps in vertebrates. In particular, it remains unclear whether the phenotypic changes that arose independently, or the traits that are shared by both lineages resulted from the accelerated evolution of regulatory networks.

In this work, we focus on the identification of genetic changes that may have driven the emergence of phenotypic novelties unique to avian and mammalian lineages, distinguishing them from other vertebrates. The objectives of this work are twofold: first, to detect accelerated evolutionary events in the basal mammalian and avian lineages, and second, to identify potential regulatory regions that may have influenced the evolution of traits specific to mammals, as well as those shared with birds. We identify 2888 and 3476 noncoding (nc) avian (AvARs) and mammals accelerated regions (MARs) that accumulate in key developmental genes, particularly in transcription factors. Most of these accelerated regions emerged by evolutionary mechanisms, since signatures of biased gene conversion are found in a minimum proportion of these elements. We also find that all five of the most accelerated ncMARs behave as transcriptional enhancers in transgenic zebrafish assays, underscoring the functional importance of these regions. We also uncover hotspots of accelerated sequences in genes and genomic regions in the avian and mammalian lineages. A remarkable example is the neuronal transcription factor *Neuronal PAS domain-containing protein 3* (*NPAS3*), which displays 30 ncMARs in its locus. *NPAS3* also carries the largest number of human accelerated sequences (HARs) and accumulates numerous ncAvArs, suggesting that some genes are repeatedly remodeled in different lineages, likely impacting morphological and functional evolution.

Results

Identification of conserved and accelerated regions in mammals and birds

To detect accelerated elements in the basal branches of mammals (Fig. 1a, d) and birds (Fig. 1a, e), we used the phastCons and phyloP programs from the PHAST package[34,35]. First, we scanned whole vertebrate genome alignments to identify sequences that were conserved in vertebrates using PhastCons. For mammals, we required that the most ancestral mammalian species, the platypus (*Ornithorhynchus anatinus*), be present in all alignments and share nucleotide changes with all other mammals in the tree (Fig. 1d). Using this approach, we identified 93,881 conserved mammalian sequences, with a minimum size of 100 bp. In these conserved sequences, we used the phyloP software[34] to detect acceleration signals and identified 24,007 MARs of which 3476 are noncoding, representing 14.4% of the total number of accelerated regions (Fig. 1a–c; Supp. Data 1 and 5). The remaining 20,531 accelerated regions overlap coding exons and were designated as coding mammalian accelerated regions (cMARs) (Fig. 1a–c; Supp. Data 2). Among the 3476 ncMARs identified in this work, only 151 regions overlapped with the 4798 therian mammalian accelerated elements (TSARs) previously reported[13]. This highlights the importance of including the platypus in the multiple alignment, as its sequence and shared changes with other mammals are critical for identifying regions specific to the mammalian basal branch.

For detecting accelerated sequences in birds, we first identified a total of 155,630 conserved sequences in vertebrate genome alignments, each with a minimum size of 100 bp (Fig. 1a, b). In this case, to detect accelerated elements, we required that at least one early diverging bird (either white-throated tinamou, *Tinamus guttatus* or ostrich, *Struthio camelus*) be present in all alignments and share a nucleotide change with other bird species, while differing from the consensus sequence of other tetrapods. Using this approach, we identified 5659 AvARs, of which 2888 were noncoding (Fig. 1c; Supp. Data 3). These ncAvARs represented 51% of the total 5659 accelerated elements, with the remaining 2771 being coding elements (Fig. 1c; Supp. Data 4 and 5). One requirement of our method is that the accelerated sequences (MARs and AvARs) remain highly conserved throughout the evolutionary history of their respective clades (i.e., Mammals and Birds, see Methods and Fig. 1b).

We found striking differences between the proportions of noncoding and coding accelerated sequences in mammals and birds (Fig. 1C and Supp. Data 5). In mammals, 85.6% of accelerated elements (20,531 out of 24,007) and 78% of base pairs (4,261,915 out of 5,449,351 bp) were coding, while only 14.4% (3476 out of 24,007) covering 1,187,436 bp (22% of total) were noncoding. In contrast, birds showed nearly equal proportions of coding and noncoding accelerated elements, with 49% of elements (2771 out of 5659) and 900,855 bp (of 1,981,612 bp) being coding and 51% (2888 out of 5659), including 1,080,757 bp being noncoding (Supp. Data 5). However, it is important to note that these proportions follow a trend already observed in the proportions of conserved coding and noncoding

https://doi.org/10.1038/s41467-025-64134-2

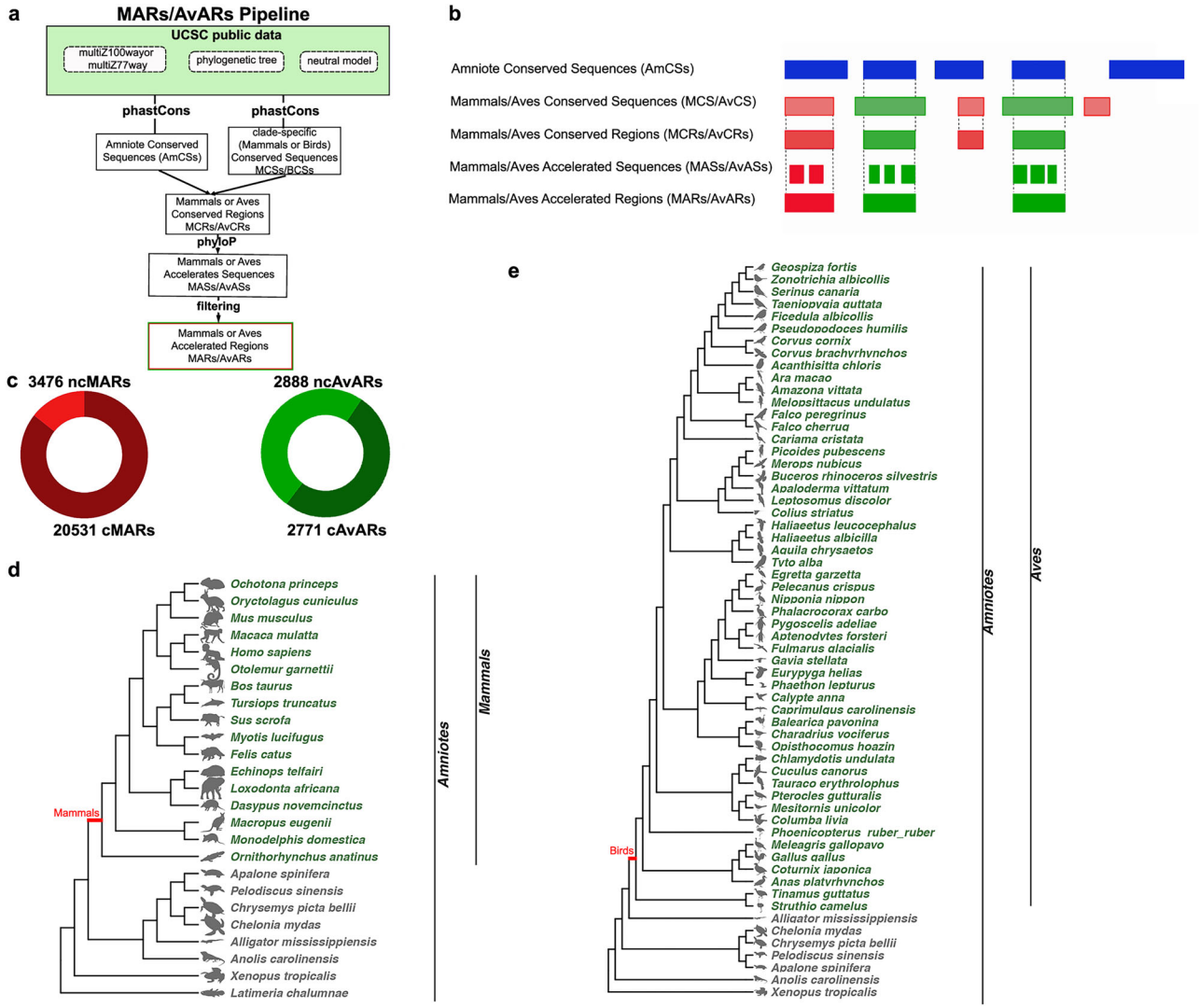

Fig. 1 | Identification of conserved and accelerated regions in mammals and birds. a Diagram illustrating the analysis pipeline followed to obtain vertebrate conserved genomic elements using *phastCons* and to identify accelerated elements through *phyloP* implementation in the mammal or bird basal lineages. **b** Schematic of the procedure to obtain the accelerated regions from conserved sequences. **c** Donut plot showing the total proportion of coding versus noncoding accelerated elements in mammals (left) or birds (right). Phylogenetic trees of mammalian (**d**) and birds (**e**) species used in the accelerated sequence detection analysis. Animal silhouette images were sourced from PhyloPic (phylopic.org[99]). The branch tested in either mammals or birds is highlighted in red.

regions in our mammalian and bird alignments (Supp. Data 5). These data suggest that accelerated evolution is shaping these two functional components of the genome, coding and noncoding, with different intensities in mammals and birds.

Regarding the mechanisms underlying accelerated evolution and given the potential for GC-biased gene conversion (gBGC) to mimic adaptive evolution signals, we aimed to detect its influence in the obtained accelerated regions. We used phastBias to detect informative regions of gBGC in the mammalian and avian lineages. We detected 276 gBGC tracts in 182 ncMARs (~5% of total ncMARs; Supp. Data 6) and 37 gBGC tracts in 31 ncAvARs (~1% of total ncAvARs; Supp. Data 7). We found that some of the most accelerated sequences in the mammalian lineage, ncMAR-1, ncMAR-6 and ncMAR-7 display gBGC signatures. In ncMARs, the lengths of gBGC tracts ranged from a minimum of 1 bp to a maximum of 600 bp, with scores varying between 1.03 and 507.5. The highest-scoring tract (chr3:70582725-70583325, score: 507.5), encompassing ncMAR-1859, suggests a strong signature of GC-biased gene conversion in this element. Similarly, in ncAvARs, tracts varied in gBGC influence, with the highest-scoring tract in ncAvAR-2863 (chr2:8892519-8892810), with a score of 272.3. For both mammals and birds, many of these tracts of gBGC signatures appeared in adjacent segments in the same accelerated element (such as in ncMAR-1818 or ncAvAR-1203), which suggests some accelerated regions were strongly affected by gBGC. Overall, our data indicate that the effect of gBGC underlies acceleration in a small proportion of elements in both mammal and bird basal lineages, suggesting that other mechanisms shaped the evolution of these regions.

Hotspots of rapidly evolving genomic regions and genes in mammals and birds

In order to detect hotspots or accumulation of accelerated elements in the genome, we calculated the frequency of accelerated noncoding elements in intronic regions per gene transcriptional unit (defined as the genomic regions comprising the transcriptional start and end of genes). We found that the gene carrying the largest amount of ncMARs is the transcription factor *NPAS3* (Fig. 2a; Supp. Data 8). This gene encodes a bHLH-PAS protein that plays a key role in the developing nervous system, and its dysfunction is associated with the etiology of schizophrenia and bipolar disorders[36–39]. Surprisingly, *NPAS3* was also identified as the locus with the highest number of human accelerated

https://doi.org/10.1038/s41467-025-64134-2

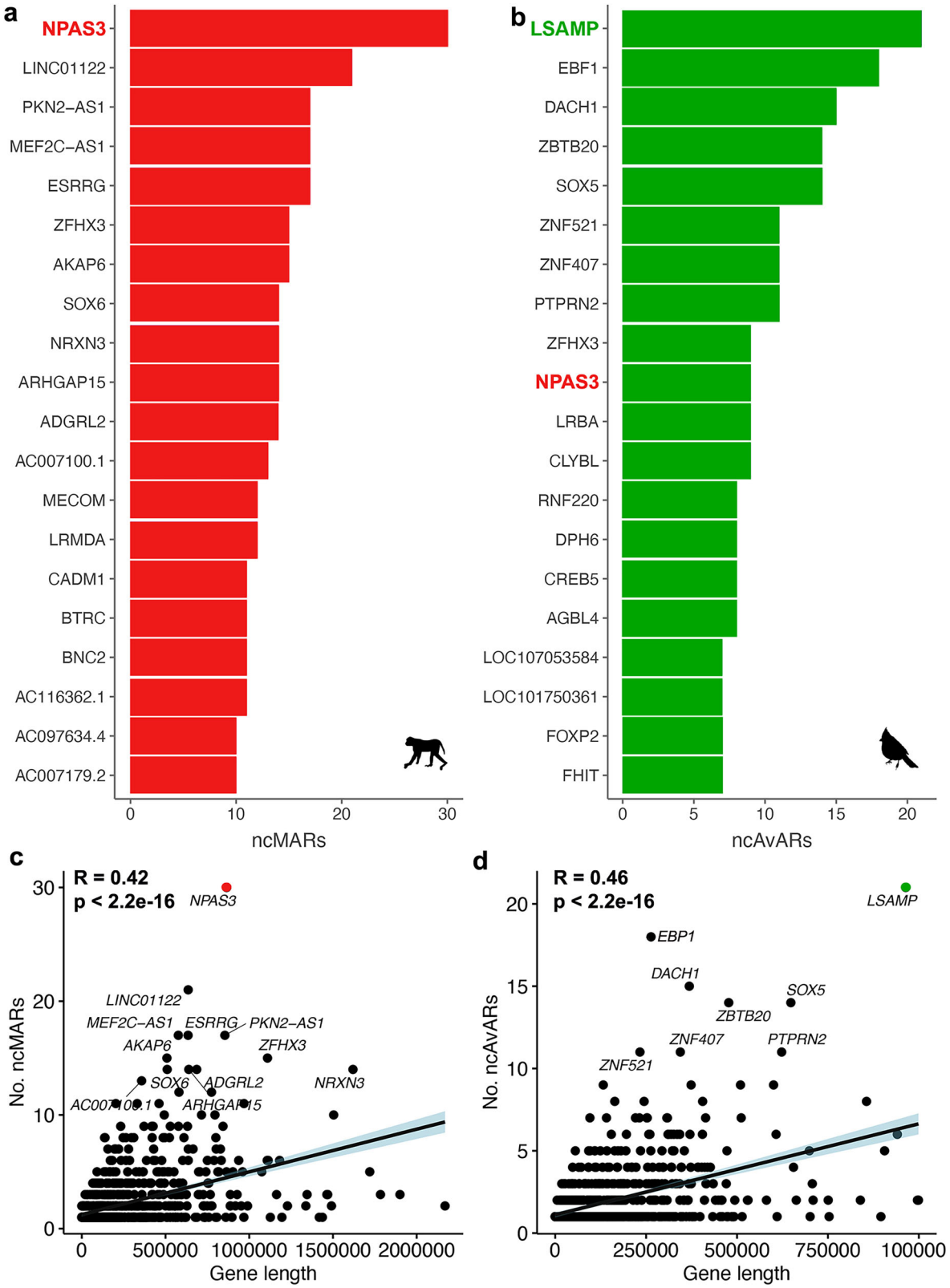

Fig. 2 | Accelerated regions accumulate in big genes. Diagrams showing the top twenty genes accumulating the largest numbers of ncMARs (**a**) and ncAvARs (**b**) in the vertebrate genome. The transcription factor *NPAS3* is highlighted in red in both diagrams. Graphs showing the number of ncMARs (**c**) and ncAVARs (**d**) by gene length (in base pairs). A positive Pearson correlation is observed, **c** $R = 0.42$ p value < 2.2e-16, **d** $R = 0.46$ p value < 2.2e-16. Linear regression lines and their 95% two-sided confidence regions are shown. Gene symbols are displayed for those genes showing the largest numbers of accelerated elements. The top genes accumulating ncMARs (*NPAS3*) and ncAvARs (*LSAMP*) are highlighted in each diagram.

elements (HARs), many of which function as transcriptional enhancers[40]. An independent analysis also identified clusters of accelerated regions in several ape lineages in *NPAS3*[41]. A comparable study also found a cluster of numerous flightless birds accelerated regions in *NPAS3*[17]. It is interesting to note that in our analysis, *NPAS3* is among the ten genes accumulating the largest amount of ncAvAR (Fig. 2b). Taken together, these findings suggest that *NPAS3* is a locus with a high number of conserved, potentially regulatory elements that have been repeatedly modified across multiple lineages. In contrast, the distribution of cMARs along genes differed substantially. The genes with the largest number of cMARs did not overlap with those carrying the highest numbers of ncMARs (Supp. Fig. 1a) and *NPAS3* is absent from the list of top genes accumulating cMARs.

The next two gene transcriptional units accumulating the largest number of accelerated elements in the lineage leading to mammals were *LINC01122* (*long intergenic non-coding RNA 1122*) and *ESRRG* (*Estrogen-related receptor gamma*) (Fig. 2a, Supp. Data 8). *LINC01122* is classified as a long intergenic noncoding RNA and is expressed in several regions of the brain (nucleus accumbens, cortex, putamen and basal ganglia) (https://gtexportal.org/home/gene/LINC01122). ESRRG is a transcription factor involved in the estrogen response. In humans, the mRNA encoded by this gene is expressed in the brain, parathyroid glands and placenta of adult humans (https://www.proteinatlas.org/ENSG00000196482-ESRRG/tissue), although protein expression shows low consistency. In mice, during development, this gene is also expressed in the genitourinary system, gonads and central nervous system (http://www.informatics.jax.org/marker/MGI:1347056). Undoubtedly, many more experimental studies will be necessary to determine the role that these genes played in the evolution of mammalian traits.

In the lineage leading to birds, the top four genes displaying the largest number of ncAvARs were *LSAMP*, *EBF1*, *DACH1*, and *SOX5* (Fig. 2b, Supp. Data 8). The *LSAMP* (Limbic System Associated Membrane Protein) gene encodes a preprotein that generates a neuronal surface glycoprotein upon processing, which is thought to act during axonal migration and neuronal growth in the limbic system of mammals[42]. There is no information about *LSAMP* expression or function in chicken (*Gallus gallus*). In mice, the absence of this gene leads to abnormal grooming, impaired social interaction and hyperactivity (http://www.informatics.jax.org/marker/MGI:1261760). The *EBF1* (*Early B-cell Factor 1*) gene is expressed in chickens in the mesenchyme of the hind limbs, surrounding the perichondrial layer[43]. Expression was also observed in the pineal gland, diencephalon, midbrain and telencephalon[44,45]. In mice, *EBF1* deletion causes incomplete dominance lethality before weaning. Surviving mice show abnormalities in cardiovascular tissues, facial morphology, and a reduced freezing response to a stimulus (https://www.mousephenotype.org/data/genes/MGI:95275). In chickens, *Dachshund family transcription factor 1* or *DACH1* is expressed during early development in the ectoderm and neural crest. At later stages, it shows expression in many tissues, including the central nervous system, cranial ganglia, otic and optic placodes, eye, brain, mesenchyme of the head and limbs, pharyngeal arches, spinal cord and wing mesenchyme[46]. In addition, it has been shown that the *Dach1* gene accumulates a large number of accelerated elements in ratites (flightless paleognath birds)[17]. This result is highly relevant considering that *Dach1* is a known regulator of limb development[46]. Regarding *SOX5 (SRY-Box Transcription Factor 5)*, limited information exists on its role in birds but it is known that in chicken this transcription factor is expressed during development in the neural tube and somites, and then in the otic and nasal placodes, mesenchyme of the head and limbs (including wings), wing bone, nerves and spinal ganglia, telencephalon and midbrain (GEISHA ID: ChEST909e22). In mammals, homozygous *Sox5* mutant mice die at birth due to respiratory failure, a small rib cage and bone mineralization problems. Its expression during development is observed in the genitourinary system, visceral organs and nervous system (http://www.informatics.jax.org/marker/MGI:98367). These results suggest that these four genes, but particularly *Dach1*, are strong candidates for further studies exploring their roles in the evolution of avian phenotypic traits, including wing development. Besides, among the 20 genes accumulating ncAvARs we found *FOXP2*, a gene that has been largely involved in the development and evolution of speech in humans[47] and the evolution of song-learning in birds[48]. Noticeably, we also found in an independent study that this transcription factor accumulates HARs that act as transcriptional enhancers during brain development[49]. As with mammals, the distribution of coding avian accelerated regions (cAvARs) did not overlap with the genes carrying the largest numbers of ncAvARs (Supp. Fig. 1b).

Since gene size could influence the number and length of intronic regions, and consequently, the number of accelerated elements, we analyzed the distribution of accelerated elements as a function of gene size. We observed a trend toward a greater accumulation of accelerated elements in larger genes (longer transcriptional units). We observed that, for ncMARs, *NPAS3* stands out with the highest number of accelerated elements compared to other genes (Fig. 2c). Similarly, for ncAvARs, this pattern is observed for *LSAMP*, *EBF1* and *DACH1* (Fig. 2d). For coding regions, we calculated the distribution of cMARs and cAvARs and observed that these regions tend to accumulate in genes with a large number of exons (Supp. Fig. 1c, d).

To further investigate the distribution of accelerated sequences, we analyzed their location within topologically associated domains (TADs; Supp. Data 9). A TAD is a genomic region in which DNA sequences within it physically interact with each other more frequently than with sequences outside the TAD[50]. That is, they are genomic regions bounded by physical chromatin structures and thus could represent three-dimensional structures where there is a higher probability of enhancer-promoter interaction. Since TADs are relatively conserved in different tissues, we used data from the fetal human cortical plate (GSE77565) as a reference for the human genome (hg38 version). The three regions with the highest frequency of ncMARs by TAD size (bp) were: chr2:57652865-60692865 (a region of the genome populated by several long noncoding RNAs; lncRNAs), chr9:78305084-83025085 (a very large chromosomal interval with a high gene density) and chr14:32570794-34450794 (the region corresponding to the *NPAS3* locus).

In the chicken genome, we used chromatin contact maps derived from Hi-C studies performed on chicken fibroblasts[51], annotated on the galGal6 genome (Supp. Data 9). The three regions showing the highest frequency of ncAvAR were: chr2:90246100-91807271 (a gene desert according to *G. gallus* gene annotations), chr1:157601983-158881376 (the *DACH1* locus) and chr2:103022509-104142533 (a genomic region with high density of coding genes).

Hotspots of accelerated elements in genomic regions could suggest that these clustered elements are changing in a coordinated manner to regulate or modify the same gene. To investigate this, we identified coding and noncoding clusters in mammalian and avian genomes. In order to compare with previously identified accelerated element databases[13], a group or "cluster" was considered to be a set of three or more accelerated elements where the nearest neighboring element was at least at a distance less than or equal to 50 kb. In mammals, we found 801 ncMARs organized in 145 clusters (Supp. Data 10). The largest ncMAR cluster ($n = 15$) is scattered in the distal 5′ intronic regions of *AKAP6*, near the *NPAS3* promoter. The second largest cluster ($n = 13$) is located in the human chromosome 2, a region comprising the *ZEB2* gene (Supp. Data 10). Additionally, our analysis revealed that clusters of noncoding and coding accelerated elements do not overlap. For instance, the largest cluster of cMARs is located in the human chromosome 3 and contains 62 elements (Supp. Data 11). We performed the same analysis for avian accelerated elements, identifying 962 ncAvARs organized into 150 clusters (Supp. Data 12).

The largest ncAvARs cluster ($n = 26$) is located in the chicken chromosome 2 (genomic region: chr2: 91,249,238-91,668,202;galGal6). Predicted NCBI RefSeq genes found in this region include *Tshz1*, *Zadh2* and *Znf407*. On the other hand, cAvARs are organized in 90 clusters containing 436 elements. The largest cluster ($n = 12$) is located in the chicken chromosome 1, showing again that clusters of coding and noncoding elements do not overlap (Supp. Data 13).

Overall, our combined results demonstrate that certain loci, such as *NPAS3* in mammals and *DACH1* in birds, underwent a high frequency of evolutionary reshaping in noncoding regions, highlighting their potential roles in lineage-specific phenotypic evolution.

Relevance of accelerated elements in the evolution of mammalian and avian traits

We used gene ontology (GO) analysis to interpret the biological significance of the identified accelerated elements. We employed rGREAT[52] to inspect the molecular functions (MF), biological process (BP) and cellular component (CC) of ncMARs and ncAvARs (Fig. 3, Supp. Fig. 2 and Supp. Data 14–19). Given the non-random distribution of conserved elements in the genome, we implemented a strategy comparing the results of rGREAT obtained from the set of accelerated regions with the distribution of those obtained from equivalent random sets of conserved regions. Using this approach we found that the most enriched ontological terms for the ncMARs with respect to their biological processes (GO:BP) fall into categories related to behavioral and cognitive processes (such as behavioral fear response, social behavior, learning or memory, suckling behavior), nervous system development (cerebral cortex regionalization or neuron migration), cell proliferation, differentiation, and migration (such as mammary gland epithelial cell proliferation, endothelial cell migration, and skeletal muscle cell differentiation), and cardiovascular and muscular system development (including cardiac muscle tissue growth, regulation of heart growth, and ventricular cardiac muscle cell differentiation) (Fig. 3a and Supp. Data 14).

Particularly interesting is the presence of several terms related to the morphogenesis of the heart, an organ that underwent intense remodeling in the mammalian lineage. In fact, we found ncMARs related to genes like *TBX5*, a key regulator of heart development[24,53], and to other genes such as *TBX2*, *TBX20* and *BMP2,* largely involved in heart evolution in vertebrates[25]. It is also noteworthy, that the presence of terms related to brain development, particularly to cerebral cortex regionalization, is linked to ncMARs associated with genes such as *EMX1/2*[54] and *DMRTA*[55]. We also found GO terms indicating significant enrichment in accelerated regions associated with genes involved in mammary gland development, including *STAT6*[56], *WNT5A*[57], and *GATA3*[58].

We performed the same procedure as indicated above to perform GO enrichment analysis on ncAvARs (Fig. 3b and Supp. Data 17). ncAvARs were enriched into regions carrying genes related to nervous system development (e.g., neural tube patterning, hypothalamus development, hypothalamus development, cerebellar Purkinje cell layer development, cerebellar cortex development), immune response (e.g., regulation of type I interferon production), regulation of hormonal and neurotransmitter pathways (e.g., regulation of androgen receptor signaling pathway, negative regulation of androgen receptor signaling pathway, positive regulation of intracellular steroid hormone receptor signaling pathway, tyrosine metabolic process) cell proliferation and differentiation (e.g., epithelial cell fate commitment, epithelial-mesenchymal cell signaling, endothelial cell fate commitment, lymphatic endothelial cell differentiation, cell differentiation involved in embryonic placenta development, trophoblast giant cell differentiation), intracellular components rearrangements (e.g., regulation of inclusion body assembly, negative regulation of inclusion body assembly, inclusion body assembly, spliceosomal snRNP assembly, regulation of lysosomal lumen pH, intracellular pH elevation, protein localization to perinuclear region of cytoplasm) and organ morphogenesis (e.g., middle ear morphogenesis, atrial cardiac muscle tissue morphogenesis, ventricular cardiac muscle cell development). It is interesting to note that genes related to the ear morphogenesis, particularly the middle ear, are significantly overrepresented in the ncAvARs group, particularly *EDNRA*[59], *EYA1*[60] and *NKX3-2*, a known transcription factor involved in the morphological evolution of the middle ear[61]. Among the terms related to nervous system development, it is noteworthy that the presence of cerebellar development genes linked to accelerated sequences, including *LDB1*[62], *SKOR-2* [63], is particularly significant because of the link between the cerebellum and the ability to power flight in birds[64]. It should be emphasized that our analysis found ncAvARs enriched in the genomic regions of genes involved in heart development, like *BMP4*, *TBX5* and *TBX3,* among others. This finding is particularly interesting considering that these genes also show enrichment in ncMARs, suggesting that genes involved in heart morphogenesis were shaped by accelerated evolution of noncoding regions in both lineages.

Our analyses of molecular function indicated that the genes linked to both ncMARs and ncAvARs show a strong representation of terms related to transcriptional regulation (Supp. Fig. 2a, b, Supp. Data 15 and 18), suggesting that accelerated elements could modify transcriptional networks of genes underlying the development of key anatomical structures involved in the biological processes described above. Regarding cellular components, whereas ncMARs are overrepresented in genes that act in the nucleus of the cell (Supp. Fig. 2c; Supp. Data 13), ncAvARs are linked to a diversity of cellular structures (Supp. Fig. 2d; Supp. Data 16 and 17).

Signatures of regulatory activity in noncoding accelerated elements

If accelerated noncoding elements function as regulatory elements, we expect them to be located in regions that exhibit signals indicating a chromatin openness state or epigenetic marks associated with regulatory activity. Based on chromatin features potentially indicating regulatory function, we assembled a collection of putative regulatory elements identified in the human genome and verified the statistical significance of the associations with noncoding accelerated elements. First, we used the LOLA (Locus Overlap Analysis) library[65] with the LOLAcore catalog (hg38) that pools curated regulatory element information identified in different cell types (http://databio.org/regiondb). We additionally incorporated databases that pool transcriptional enhancers identified through functional experiments (VISTA enhancers, https://enhancer.lbl.gov/) to assess the possible regulatory role of accelerated noncoding elements (Supp. Data 20 and 21).

We found that DNase I hypersensitive sites, p300-tagged enhancers, and VISTA enhancers are statistically overrepresented in mammalian accelerated noncoding elements (ncMARs) compared to their genome-wide conserved counterparts. That is, ncMARs are more frequently associated with open regions of chromatin (DNAse I) and enhancers than the rest of the phastCons (Supp. Data 20).

In the same way, we analyzed the regulatory potential of ncAvARs. We liftedOver the data from accelerated noncoding regions (ncAvARs in galGal6) to human genome coordinates (hg38) and used the combination of epigenetic data available in LOLAcore. ncAvARs showed enrichment in VISTA enhancers, and transcription factor binding in cell lines (TCF7L2 in HEK293, GATA3 in SH-SY5Y) (Supp. Data 21). Additionally, using chicken-specific regulatory data from the FAANG project[66], ncAvARs showed enrichment in regulatory elements identified specifically in hypothalamus, cerebellum, and gizzard (Supp. Data 22).

These results suggest that the proportion of ncMARs and ncAvARs showing signatures compatible with enhancer function is greater than that of the genomic set that gave rise to them (conserved elements).

https://doi.org/10.1038/s41467-025-64134-2

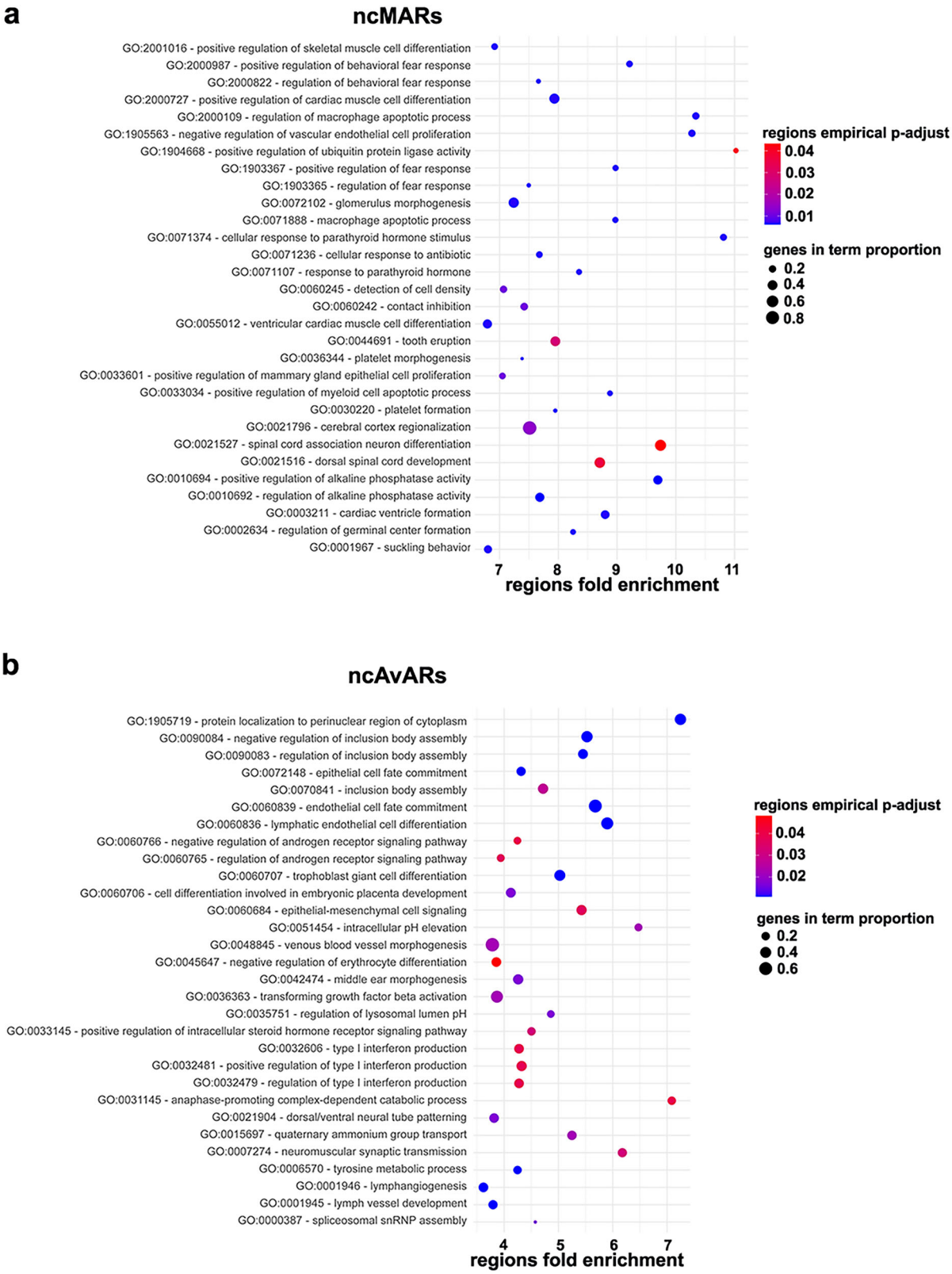

Fig. 3 | Gene ontology (GO) analyses for ncMARs and ncAvARs. Dot-plots showing the top thirty biological processes GO terms for ncMARs (**a**) and ncAvARs (**b**) ordered by descending fold-enrichment, with bubble size proportional to the fraction of genes associated with each term that are also linked to the noncoding genomic regions, color representing −log10(*p* adjust) (target/background region hits). Empirical *p* value for each GO term is calculated as the proportion of 'observed_regions_hits' values associated with conserved elements distribution that are greater than the 'observed_regions_hits' value associated with accelerated elements set (one-sided). These *p* values are corrected for multiple comparisons by Benjamini–Hochberg (BH) procedure. For the complete GO enrichment analysis results, see Supplementary Fig. 2 and Supp. Data 14–19.

https://doi.org/10.1038/s41467-025-64134-2

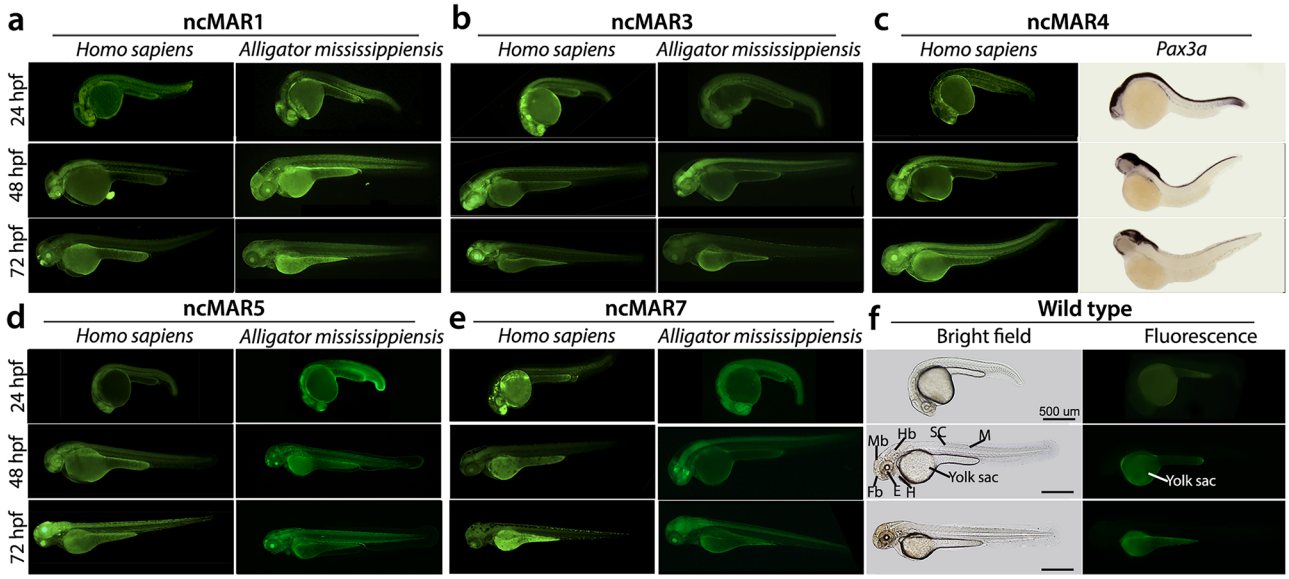

Fig. 4 | Transgenic zebrafish enhancer assays for selected ncMARs. Fluorescent photographs of selected zebrafish transgenic stable lines of human and alligator versions of ncMAR-1 (**a**), ncMAR-3 (**b**), ncMAR-5 (**d**) and ncMAR-7 (**e**) showing EGFP expression at 24, 48, and 72 hpf. **c** Fluorescent photographs of a selected transgenic zebrafish stable line of ncMAR4-human showing EGFP expression at 24, 48, and 72 hpf and in situ hybridizations of the *Pax3a* gene in zebrafish, performed by the laboratory of Drs. Bernard and Christine Thisse (https://zfin.org/ZDB-PUB-051025-1). **f** Wild-type zebrafish microphotographs in bright field and epifluorescence showing anatomic details. At least three transgenic lines per transgene were generated. See all transgenic lines and sequence details in Supp. Figs. 3–7. Fb forebrain, Mb midbrain, Hb hindbrain, E eye, H heart, SC spinal cord, M trunk muscle.

Identification of common accelerated noncoding regions in avian and mammalian genomes

We investigated whether conserved noncoding regions evolved at an accelerated rate in both the avian and mammalian lineages in parallel. To analyze this, we cross-referenced mammalian and avian noncoding accelerated regions, identifying 285 common accelerated noncoding regions (CARs) (Supp. Data 23).

We found that most (262/285) of the transcriptional units containing CARs had only one, whereas 23 units had two or more. Notably, the transcriptional units of *AC007100.1* and *DACH1* genes contained the highest number of CAR regions, with four elements each. These results indicate that most of the common accelerated elements are not found in clusters but are dispersed throughout the genome. Our results indicate that certain conserved noncoding elements in vertebrates underwent accelerated evolution in parallel in the lineages leading to mammals and birds, probably because they serve a function that was modified in both lineages.

Accelerated elements derived from transposable elements

To gain more insight into the evolutionary origin of ncMARs and ncAvARs, we analyzed whether these elements are derived from transposable elements. We found that 245 ncMARs overlap with repetitive elements (Supp. Data 24), noticeably 25 of them contain more than 40% of their extent derived from transposable elements of several types (Supp. Data 25). Five of these sequences overlap with ENCODE cis regulatory sequences (CRE) of various types including distal and proximal enhancers and promoters.

We also found that 288 ncAvARs overlap with repetitive elements (Supp. Data 26) and in 19 of them the transposable elements cover at least 40% of their extent (Supp. Data 27), whereas seven of them are characterized as ENCODE CREs, including promoters, proximal and distal enhancers. Our data indicate that transposable elements provided raw material that was co-opted as functional sequences and probably shaped through accelerated evolution to serve as regulatory sequences of different types.

Functional characterization of selected accelerated elements as transcriptional enhancers

We explored the enhancer-like regulatory function of selected ncMARs using a transgenic assay in zebrafish. We chose five elements (ncMAR-1, ncMAR-3, ncMAR-4, ncMAR-5, ncMAR-7) that show the highest nucleotide substitution rates in mammals relative to other vertebrates. ncMAR-2 and ncMAR-6 were not included because the first shows epigenetic signatures of promoter function, and the second intersects a pseudogene. For each element, we cloned the conserved region spanning the accelerated elements into a plasmid containing Tol2 recognition sites flanking a cassette containing a murine *cFos* minimal promoter fused to the *EGFP* reporter gene. The injected zebrafish embryos (F0) are mosaic transgenic animals, meaning that they carry the transgene in some cells but not in others. To overcome this problem, we evaluated the expression patterns in the progeny (F1), generating stable transgenic lines that incorporate the constructs with the accelerated elements into their genome. At least 3 independent lines per element were established, and the *EGFP* expression pattern was analyzed at 24, 48 and 72 h post fertilization (hpf) in F1 animals, classifying expression domains using general anatomical terms (Fig. 4).

The ncMAR-1 element is the mammalian accelerated noncoding element with the highest number of mammalian-specific substitutions (Supp. Fig. 3a). It is located on human chromosome 2, at position chr2:163762291-163762419 (hg38), at the *AC016766.1* (ENSG00000237844.2) locus, upstream of the coding gene *Fidgetin*, *Microtubule Severing Factor* or *FIGN* (Supp. Fig. 3a). Our results show that ncMAR1-human behaves as a strong transcriptional enhancer that drives reproducible expression of *EGFP* to the eye, forebrain, midbrain, hindbrain and spinal cord in the three transgenic lines generated (Fig. 4a, Supp. Fig. 3b). We also evaluated the expression pattern driven by the American alligator (*Alligator mississippiensis*) sequence and found that most of the expression domains are conserved between the human and alligator transcriptional enhancers (Fig. 4a, Supp. Fig. 3b) despite the large number of changes that the mammalian sequence presents compared to the tetrapod outgroup (Supp. Fig. 3a). Regarding the genes that this enhancer could be

controlling, the *AC01676666.1* gene is classified as a long noncoding RNA (lncRNA) with unknown expression and function. On the other hand, *FIGN* is a coding gene that in zebrafish is expressed in somites, central nervous system, brain (diencephalon, telencephalon, tegmentum, hindbrain and cerebellum) and eyes[67] (Supp. Fig. 3e) and its expression pattern resembles the one shown by ncMAR-1 at the same developmental stages. In mammals, *FIGN* is involved in axon outgrowth, cell division and neuronal regeneration[68].

The human and alligator sequences of ncMAR-3 (chr14:76,089,219-76,089,323; hg38; Supp. Fig. 4a), drove reproducible *EGFP* expression in the eye, forebrain, midbrain, hindbrain, spinal cord, otic capsule and lateral line in the transgenic lines analyzed (Fig. 4b, Supp. Fig. 4b, c). At 72 hpf, both versions of ncMAR-3 also drive the expression of the reporter gene to the developing hearing and balance system (Fig. 4b, Supp. Fig. 4b–d). This element is classified by ENCODE as a regulatory element with distal enhancer properties using DNAsa I hypersensitive site assays (DHSs) from 706 experiments in human cells in the SCREEN project (Accession: EH38E1729226). If we assume that its enhancer function is potentially performed on the promoter of the nearest transcription start site (TSS), it would correspond to the Intraflagellar Transport 43 gene (*IFT43)*, located approximately at 4000 bp downstream of this regulatory element (hg38; Supp. Fig. 4a). The protein encoded by this gene, is expressed in adult humans in male reproductive organs (testicle, epididymis), fallopian tubes, kidney, soft tissues and cerebellum (https://www.proteinatlas.org/ENSG00000119650-IFT43/tissue). In zebrafish, its expression pattern is unknown, and in mice, the absence of this gene causes lethality and abnormal midbrain/hindbrain development, poor neural tube closure, increased corneal size and abnormal pharyngeal arches in the embryo (https://www.mousephenotype.org/data/genes/MGI:1923661).

We obtained five independent transgenic lines carrying the human sequence of ncMAR-4 (chr2:222,100,655-222,100,767; hg38; Supp. Fig. 5a), which reproducibly directed *EGFP* expression to the brain, spinal cord and eyes (Fig. 4c, Supp. Fig. 5b, c). The element is located within a gene desert, where the closest gene is *PAX3* (Supp. Fig. 5a). There are two *PAX3* paralogs in zebrafish: *pax3a* and *pax3b*. *Pax3a* is widely expressed in the central nervous system, including the spinal cord and the brain (diencephalon, forebrain and hindbrain), neural cristae, eyes and muscle, (https://zfin.org/ZDB-GENE-980526-52) (Supp. Fig. 5d). *Pax3b* has fewer expression studies and is known to be expressed in neural cristae and somites (https://zfin.org/ZDB-GENE-080917-53). In mouse embryos, *Pax3* is widely expressed during development in several regions, especially in those tissues derived from ectoderm (brain and spinal cord) and mesoderm (paraxial mesenchyme such as somites, mesenchyme derived from neural crests and limbs) (http://www.informatics.jax.org/marker/MGI:97487). This transcription factor has a key role in the development of skeletal muscle, the central nervous system and neural crest derivatives[69].

The ncMAR-5 element is located within the human lncRNA locus AL110292.1 (ENSG00000258081.4) at position (hg38) chr14:26,987,886-26,987,992 (Supp. Fig. 6a). The human and alligator versions of ncMAR-5 directed their expression to similar domains, including the entire central nervous system and the eye (Fig. 4d, Supp. Fig. 6b, c). This element was classified as a distal enhancer for its epigenetic characteristics in the ENCODE SCREEN study (reference: EH38E1704813). Regarding coding genes that are located close to this element, at 400 Kbp is the *NOVA1* gene. *NOVA1* encodes a neuronal RNA-binding protein expressed in the central nervous system that is essential for survival in mice and normal development in humans[70]. *Nova1* is also expressed in the developing nervous system in zebrafish (Supp. Fig. 6e) in a pattern that resembles the one directed by ncMAR-5, and morpholinos directed against *nova1* in zebrafish resulted in abnormal brain morphology and abnormal optic tectum[71].

The ncMAR-7 region (chr11:45,726,523-45,726,695; hg38) is located within the intronic portion of lncRNA *AC044839.1* (ENSG0000 0254519.5) between two coding genes (*CHST1* and *SLC35C1*) (Supp. Fig. 7a). The human and alligator sequences of ncMAR-7 directed reproducible expression to the forebrain, spinal cord, and eyes (Fig. 4e; Supp. Fig. 7b–d). This accelerated element has distal enhancer features (EH38E1536858), defined by the ENCODE consortium in their SCREEN project and, in addition, has been associated with the *CHST1* (*Carbohydrate sulfotransferase 1*) coding gene in eQTL studies in tibial artery samples[72]. In adult human tissue, *CHST1* mRNA is enriched in the brain and, to a lesser extent, in endocrine tissues (https://www.proteinatlas.org/ENSG00000175264-CHST1). Mice that are homozygous mutants for this gene show increased aggression among males and abnormally sized lymph nodes (http://www.informatics.jax.org/marker/MGI:1924219). The *SLC35C1* gene in adult humans is expressed in the liver in high concentration, while the protein encoded by this gene is found in various parts of the digestive system and brain (https://www.proteinatlas.org/ENSG00000181830-SLC35C1). Mutant mice for this gene show perinatal lethality and growth retardation, defects in immune-related blood parameters, and defective development of the lungs and lymph nodes (http://www.informatics.jax.org/marker/MGI:2443301). Thus, this evidence indicates that the expression pattern driven by ncMAR-7 is more related to the one displayed by the *CHST1* gene and its function. CHST1 is also called Keratan Sulfate Gal-6 Sulfotransferase 1 (KSGal6ST) and is responsible for the sulfation of Keratan Sulfate in the developing mouse brain[73] where it plays key roles as a modulator of extracellular matrix components during the development of the brain[74].

In summary, analysis of the five ncMARs expression in transgenic fish showed that all the elements analyzed were able to distinctively and reproducibly activate *EGFP* expression, mainly in different domains of the developing central nervous system. We were able to compare the expression of mammalian and reptile sequences for four of them, and our results indicate that despite the strong signature of accelerated evolution that distinguishes these elements in the lineage leading to mammals, the expression domains controlled by the mammal and the reptile versions are similar in transgenic assays. However, we can not rule out the possibility that changes in expression in different cellular types or in additional developmental stages where the enhancers are active exist, and more detailed analysis will be necessary to assess these hypotheses.

Functional characterization of accelerated elements in the *NPAS3* locus

As mentioned before, our hotspot analysis showed that the gene comprising the largest number (30) of ncMARs is *NPAS3* (Figs. 2 and 5 and Supp. Data 8). The *NPAS3* locus also contains 10 ncAvARs (Figs. 2b and 5), which makes it one of the top genes accumulating accelerated noncoding sequences in the lineage leading to birds. It is interesting to note that when we analyzed the distribution of ncMARs across chromosome 14, we found that the region containing *NPAS3* stands out for the accumulation of this kind of element (Fig. 5a). On the other hand, the distribution of cMARs is more homogenous along the chromosome and the *NPAS3* region does not accumulate accelerated coding regions (Fig. 5a). We have also previously found that this transcription factor is the *locus* with the largest number of HARs and that most of these noncoding sequences act as transcriptional enhancers during nervous system development[40]. Comparing the distributions of HARs/MARs we found that there are four ncMARs that overlap HAR elements in the *NPAS3* locus: HAR189/HACNS490/ncMAR-2474/ncMAR-381, HAR89/ncMAR-3386, HAR173/ncMAR-506, 2xHAR157/ncMAR-1551 (Fig. 5b). In order to analyze the function of these NPAS3-HARs/ncMARs we cloned the human, chimpanzee and alligator sequences of these regions and performed enhancers assays in transgenic zebrafish.

The human and chimpanzee versions of HAR173/ncMAR-506 (Fig. 5b, Supp. Fig. 8b, d) drive *EGFP* expression to several domains of the developing central nervous system, including the forebrain,

https://doi.org/10.1038/s41467-025-64134-2

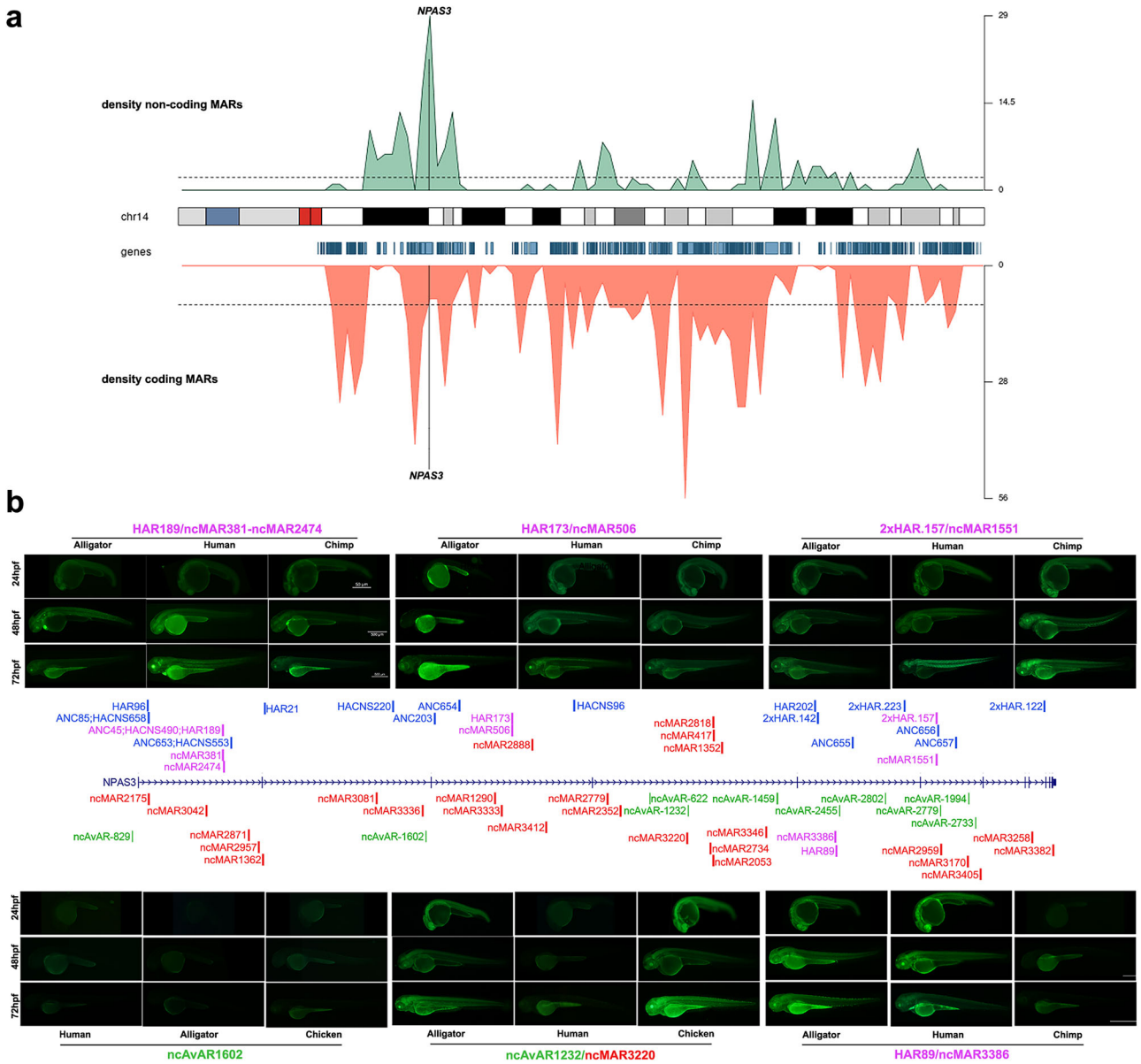

Fig. 5 | The *NPAS3* gene displays the largest amount of ncMARs. a Distribution of coding and noncoding accelerated elements in the human chromosome 14 where the location of the *NPAS3* locus is indicated by the vertical black bar. **b** In the center, the *NPAS3* genomic region showing the distribution of noncoding accelerated elements in the human genome (hg38) is shown. In blue, HARs; in green, ncAvARs; in red, ncMARs; and in purple, ncMARs that coincide with HARs. Generated through the UCSC Genome Browser. On top and below representative zebrafish stable transgenic lines at 24, 48, and 72 hpf are shown for the four ncMARs/HARs, one selected ncMAR/ncAvARs (ncMAR-3220/ncAvAR-1232) and ncAvAr-1602. At least three transgenic lines per transgene were generated. See all transgenic lines and sequence details in Supp. Figs. 8–13.

midbrain, hindbrain and spinal cord. In addition, both versions of the enhancer direct strong expression of the reporter gene to the otic capsule and sensory ganglia. We observed that the human version directs a more widespread expression pattern at 72 hpf compared to the chimpanzee sequence, including more territories particularly in the frontal regions of the developing brain (forebrain) and also to the branchial arches (Supp. Fig. 8e). In contrast, the alligator sequence is inactive as a transcriptional enhancer at the stages analyzed in transgenic zebrafish (Fig. 5b, Supp. Fig. 8c, f), suggesting that this sequence gained enhancer activity in the mammalian lineage.

The human version of HAR89/ncMAR-3386 (Supp. Fig. 9a) functions as a strong transcriptional enhancer driving the expression of *EGFP* at the entire developing central nervous system (Fig. 5b, Supp. Fig. 9c). The alligator version of this element also directs the expression of *EGFP* to similar domains in the nervous system but also drives expression to the developing branchial arches (Fig. 5b, Supp. Fig. 9b). Interestingly, the chimpanzee version of this regulatory region is inactive in driving the expression of the reporter gene at the three stages analyzed in three stable transgenic lines. (Fig. 5b, Supp. Fig. 9d, e).

The human, chimpanzee and alligator versions of HAR189/HACNS490/ncMAR-2474/ncMAR-381show strong enhancer activity driving the expression of the reporter gene to the developing central nervous system, eye, heart and somitic muscle at the three stages analyzed (Fig. 5b, Supp. Fig. 10b–d). Although the three versions of this sequence work as transcriptional enhancers, we observed some differences in expression pattern among them, particularly in the developing inner ear and in the lateral line system. However, a more detailed analysis will be necessary to assess more subtle lineage-specific changes in expression domains.

The human, chimpanzee and alligator sequences of 2xHAR157/ncMAR-1551 drove strong expression of EGFP to the developing central nervous system, eye, heart, somitic muscle and branchial arches (Fig. 5b, Supp. Fig. 11). Although the transgenic lines carrying the human and chimpanzee versions show stronger expression in the hindbrain, spinal cord and somitic muscle of the trunk (Supp. Fig. 11b, c) compared to the alligator lines (Supp. Fig. 11d) more studies will help to clarify this apparent difference shown by the ortholog sequences.

We also observed that one *NPAS3*-ncMARs overlapped with an ncAvARs (ncMAR-3220nc/AvAR-1232), and analyzed its function as an enhancer in transgenic zebrafish. We found that the human version of this region does not function as a transcriptional enhancer whereas the chicken and alligator sequences drive the expression of the *EGFP* reporter gene to several domains of the developing nervous system in zebrafish (Fig. 5b, Supp. Fig. 12). Our results suggest that this sequence lost enhancer function in the mammalian lineage (Supp. Fig. 12b) whereas substitutions in the avian lineage had a compensatory role maintaining enhancer function (Supp. Fig. 12c, d).

In addition we characterized the ncAvAR-1602 chicken, alligator and human sequences in enhancer assays in transgenic zebrafish but none of them displayed activity in the stages analyzed (Fig. 5b, Supp. Fig. 13).

Discussion

In this work, we used comparative genomics to identify changes in noncoding and coding conserved sequences that could drive the evolution of morphological and physiological traits in mammals and birds. In mammals, we identified a total of 3476 ncMARs and 20,531 cMARs. When comparing our results to the study that identified accelerated sequences in therian mammals or TSARs[13], we found that only 151 ncMARs overlapped ncTSARs. This finding underscores the novelty of the ncMARs dataset and highlights the importance of including the platypus genome in the experimental design to fully capture mammalian genome evolution. In birds, we identified 2888 ncAvARs from a total of 5659 accelerated elements. Our findings reveal a distinctive difference in the proportion of noncoding accelerated elements between mammals (3476/24,007; 14.4%) and birds (2888/5659; 51%), suggesting that the role of accelerated evolution in shaping genome function differs between these two vertebrate groups. It was previously found that most TSARs were located in coding regions, and these coding accelerated regions show high rates of nonsynonymous (dN) substitution compared to synonymous (dS) substitution in the branch of the therian ancestor, which may indicate adaptive evolution of the proteins involved[13]. These and our results suggest that in mammals, coding evolution appears to play a key role and is readily detectable with accelerated evolution studies. In contrast, studies of avian genome evolution have found that functional noncoding regions are more plastic and less strongly conserved than coding regions, which are more frequently subject to acceleration[19].

Regarding the mechanisms underlying the generation of accelerated sequences, we found that only 182 ncMARs (~5% of total ncMARs) and 31 ncAvARs (~1% of total ncAvARs) displayed signatures of gBGC, suggesting that other, possible evolutionary mechanisms were responsible for the rapid evolution of ncMARs and ncAvARs. In contrast, an independent study identified accelerated sequences in nine mammalian and ten internal bird branches and found that more than 60 % of them resulted from gBGC[75]. On the other hand, it has been shown that most HARs identified by ref. 76 show signatures of positive selection[77]. In fact, none of the NPAS3-HARs/MARs analyzed in this work showed an indication of gBGC in the lineage leading to mammals, but two of them show nucleotide changes compatible with gBGC (HAR173 and HAR189), and two of them show signatures of positive selection in the human lineage (HAR89 and 2xHAR.157). Thus, it is possible that a combination of different mechanisms underlie accelerated evolution in different lineages.

We found that 245 ncMARs and 288 ncAvARs overlap with regions of the genome derived from transposable elements. Some of them also overlap ENCODE regions indicating that they provided raw material that was co-opted and probably shaped through accelerated evolution to serve regulatory function. This finding shows once again that transposable elements build our genomes providing putative functional sequences, however, further experimental studies will be necessary to fully understand the role played by these sequences.

To find patterns that could help us understand how accelerated elements shaped the evolution of the genomes, we used different methodologies to reduce data dimensionality. On the one hand, we identified the genomic regions that accumulate the largest numbers of accelerated elements. Specifically, we used two strategies to identify such regions: by transcriptional units and by TAD. We found that the loci with the highest accumulation of accelerated elements were generally similar despite the method used in both lineages analyzed. In the lineage leading to mammals, *NPAS3* stood out as the transcriptional unit with the largest number of accelerated elements. Moreover, this transcription factor is also among the top twenty genes accumulating accelerated sequences in the lineage leading to birds. This result was surprising given that *NPAS3* is also among the most accelerated regions in other lineages, which include humans[40], apes[41] and ratites[17]. Our results thus identify a new lineage where this locus exhibits one of the highest numbers of noncoding accelerated elements. Studies on *NPAS3* coding regions have found no evidence of positive selection in humans, other primates and vertebrates more broadly[40]. Furthermore, our analysis of cMARs distribution did not detect this locus as a target of accelerated evolution in coding regions, indicating that the acceleration process occurred specifically in certain noncoding regions of *NPAS3* introns and not as a consequence of large-scale evolutionary phenomena affecting the entire locus.

In the birds' lineage, we found that both the transcriptional unit and the TAD comprising the limb developmental regulator, *DACH1*, are overrepresented in ncAvARs. As mentioned before, this gene was previously reported as one of the genomic regions carrying the largest number of accelerated elements in the ratite lineage[17]. Given *DACH1*'s role in the regulation of morphogenesis and limb differentiation[46], along with its acceleration in the lineage leading to flightless birds, and, in our study, in the avian basal lineage, we might speculate that this transcription factor plays a role in wing formation. However, caution is warranted, as it is unlikely that a single gene, whose full range of functions remains unknown, can solely explain the evolution of wings. What we can say with certainty, and in agreement with previous studies[17], is that certain key developmental genes appear to be used repeatedly to induce morphological evolution in amniotes. It is also possible that some syntenic regions or TADs possess a high density of regulatory genes, making them prone to repeated evolutionary modifications. Either way, our data support the hypothesis that evolutionary hotspots underlie phenotypic innovation across different lineages.

To search for parallel changes remodeling the same genomic regions in mammals and birds, we identified 285 accelerated noncoding elements common to mammals and birds, or CARs. We found that the transcriptional units of *AC007100.1* and *DACH1* genes contained the highest number of CAR regions, with four elements each. Our results indicate that certain conserved noncoding elements in vertebrates underwent accelerated evolution in parallel in mammals and birds, probably because they serve a function that was modified in both lineages. This dataset of CARs is a starting point that, through functional characterization, would probably help to understand their role in the coevolution of shared morphological and physiological traits in mammals and birds.

In order to investigate the role that accelerated elements may play in each lineage, we used GO analysis. At this point, it is important to highlight one of the technical limitations of our work and that of

others, which is associating a noncoding accelerated element with the nearest gene. To overcome this limitation, we used the rGREAT tool, which constructs region sets for each GO term and evaluates the enrichment of target genomic regions within these sets using a binomial model test. Thus, following this strategy, our target genomic regions are associated with noncoding accelerated elements, which, by design, are also linked to conserved elements. Concerning ncMARs, we found that terms related to the morphogenesis of the heart are overrepresented. Noteworthy, we found ncMARs linked to transcription factors including *TBX5, TBX2*, *TBX20* and *BMP2* largely involved in heart development and evolution in vertebrates [24,25,53]. These results suggest that gene regulatory networks involved in heart development were remodeled through noncoding evolution, which is noticeable since the heart underwent intense morphological and functional transformation in the mammalian lineage. It is also worthy of mention, that the overrepresentation of terms related to brain development, particularly to cerebral cortex regionalization, linked to ncMARs associated with genes like *EMX1/2*[54] and *DMRTA*[55]. Since these and other genes linked to ncMARs play a key role in the development of the cerebral cortex in mammals, our findings suggest that changes in the expression pattern of them induced by regulatory regions accelerated evolution could underlie some aspects of the tremendous morphological reshaping that the neocortex underwent in this lineage[78–80]. We also found GO terms indicating significant enrichment of accelerated regions associated with genes involved in mammary gland development, including *STAT6*, *WNT5A* and *GATA3*[56–58]. Undoubtedly, mammary glands are a defining morphological feature of mammals, thus, the discovery of noncoding regions that underwent accelerated evolution linked to genes involved in mammary gland development is a remarkable finding. However, more functional studies will be necessary to address whether accelerated evolution translated into functional changes that reshaped the expression of these genes, impacting the development of mammary glands in the mammalian lineage.

Regarding ncAvARs, it is interesting to note that genes related to the ear morphogenesis, particularly the middle ear, are significantly overrepresented in the ncAvARs group, particularly *EDNRA*[59], and *EYA1*[60] and *NKX3-2*, a known transcription factor involved in the morphological evolution of the middle ear[61]. Among the terms related to nervous system development, it is noteworthy the presence of cerebellar development genes linked to accelerated sequences, including *LDB1*[62], *SKOR-2* [63], particularly because of the link between cerebellum evolution and the ability to power flight in birds[64].

If accelerated regions identified in this work act as regulatory elements, as has been observed for accelerated sequences identified in other lineages[9,40,49,81–86], we would expect them to show enrichment with sequences showing evidence of regulatory function. We found that ncMARs and ncAvARs were overrepresented in VISTA enhancers, DNase I hypersensitive sites and p300-tagged enhancers. In the case of ncMARs, regulatory elements that overlap with accelerated elements are more frequently found in cells associated with the nervous system. These findings, based on sequence characteristics and functional genomic data, suggested that many accelerated elements are predicted to be transcriptional enhancers that regulate active genes in embryonic tissues. To test the function of these noncoding sequences in vivo, we analyzed the regulatory function of a selection of these elements using reporter assays in transgenic zebrafish. For these assays, we selected five of the mammalian accelerated elements with the highest number of mammalian-specific substitutions (ncMAR-1, ncMAR-3, ncMAR-4, ncMAR-5, ncMAR-7). We found that they were able to direct the expression of the reporter gene *EGFP* when cloned upstream of a minimal cFos promoter in transgenic zebrafish assays, indicating that they are, in fact, transcriptional enhancers, active during development. In addition, we comparatively analyzed the expression pattern of the human and the alligator versions of four of these sequences. We found that in general, the alligator sequences analyzed are also transcriptional enhancers, driving expression of the reporter gene to the same territories as the human sequences during the developmental stages analyzed. Although future studies will be necessary to evaluate whether the orthologous versions of ncMARs direct the expression to different cell types in the same territories, our results suggest that some of the many nucleotide changes observed arose to compensate for other changes that indeed modified the expression pattern.

In addition, using enhancer assays, we comparatively analyzed the human, chimpanzee and alligator sequences of the accelerated regions that are at the same time ncMARs and HARs in the *NPAS3* locus. We found that all of them are developmental transcriptional enhancers in at least one of the species analyzed. One particularly interesting case is the HAR173/ncMAR-506 element that shows gain-of-function as a transcriptional enhancer in mammals and displays some changes in expression pattern comparing human and chimpanzee sequences, suggesting that acceleration can act in different lineages on the same region to shape the function of regulatory sequences. Another interesting example is the element HAR89/ncMAR-3386, where the human and alligator versions function as strong transcriptional enhancers in the entire developing central nervous system, in contrast to the chimpanzee version, which is inactive. We hypothesize that two possible scenarios could explain these results, in the first place HAR89/ncMAR-3386 could have lost enhancer function in the chimpanzee lineage or alternatively, this sequence lost function in the mammalian lineage and recovered it in humans due to the effect of accelerated evolution. The cloning and study of more sequences will enable us to test these hypotheses. We also found that a common accelerated region (CAR) in mammals and birds in the *NPAS3* locus (ncMAR-3220/ncAvAR-1232) lost enhancer function in mammals, whereas the chicken and alligator sequences drive the expression of the reporter gene to several domains of the developing nervous system in zebrafish. Our results suggest that the substitutions in the avian lineage could have had a compensatory role in maintaining enhancer function. To summarize, of the five NPAS3-ncMARs tested, two showed gain or loss of function in the mammalian lineage. Taking into account these results and the previous findings testing the most accelerated ncMARs, we hypothesize that accelerated evolution could have an impact on function, but this is not always the rule. In our experience, between one-third and half of HAR sequences display functional changes between human and chimpanzee orthologs when tested in enhancer assays in transgenic zebrafish or mice[40,49,82]. Similar results have also been observed by other researchers who have found that 31 % of tested human and chimpanzee HAR orthologs show expression differences in transgenic mice[81]. In addition, this trend has also been shown using other experimental approaches like massive parallel reporter assays (MPRA) in different cell types. In fact, around 30 % of 1363 HARs showed human-chimp differences when tested in neural stem cells[86] and 50 % of 293 active HARs showed human-chimp differences when tested in human or chimpanzee neural precursor cells (NPCs)[87]. In this regard, recent predictions using machine learning indicate that 43% of HARs have variants with putative large opposing effects on chromatin state and 14% on neurodevelopmental enhancer activity, suggesting that compensatory evolution could underlie the evolution of some HARs in the human lineage[87]. Thus, we can conclude that accelerated noncoding regions could potentially drive expression changes on target genes, but functional studies are necessary to really assess this possibility. This is a very important message that reinforces the idea that functional studies are keen to drive conclusions about the impact of sequence evolution on functional and morphological changes. In other words, bioinformatics analyses alone are not enough to fully understand the link between molecular and phenotypic evolution.

In summary, this work provides a first approach to the identification of molecular changes underlying the evolution of morphological innovations in mammalian and avian basal lineages. Many of the

elements identified here are likely to be critical for the evolution of specific traits in the two lineages under study. The association to a regulatory function was performed through different approaches, and we provide evidence that the elements may have contributed to functional changes through their function as transcriptional enhancers. We identified clusters of elements associated with typical mammalian and avian characteristics, where those regions associated with the nervous system predominated. Our work, together with other work describing acceleration processes in other branches of the vertebrate phylogenetic tree, brings us closer to unraveling the molecular mechanisms underlying the emergence of unique aspects of species biology.

Methods

Ethics approval

All the experiments involving animals were carried out following the Guide for the Care and Use of Laboratory Animals and were approved by the INGEBI institutional animal care and use committee.

Calculation of conserved and accelerated elements

We designed a strategy to evaluate acceleration in internal branches (mammals and aves independently) on genomic regions conserved in amniotes and mammals or conserved in amniotes and aves to identify acceleration signatures in putative (coding/noncoding) functional regions. To implement this strategy, we used the RPHAST package[34] (https://github.com/CshlSiepelLab/RPHAST), developed a custom R package (cladeAcc[88], https://github.com/paulati/cladeAcc), and created downstream analysis scripts (https://github.com/paulati/ncomms-24-47757[89]). The PHAST package requires a multiple alignment, a phylogenetic tree and a neutral evolutionary model among other parameters in order to calculate conservation and acceleration[34].

Neutral model

As a null model, we used the same for the conserved and accelerated elements, a continuous Markov model of nucleotide substitutions derived from neutral sequences. Here, we use the REV model fitted to fourfold degenerate sites. In mammals, we used the neutral model file 'hg38.phastCons100way.mod' (https://hgdownload.cse.ucsc.edu/goldenpath/hg38/phastCons100way/hg38.phastCons100way.mod) from 100-way alignment. This all-species tree model was generated using the phyloFit program from the PHAST package (REV model, EM algorithm, medium precision) using multiple alignments of fourfold degenerate sites extracted from the 100-way alignment (msa_view). The fourfold degenerate sites were derived from the RefSeq (Reviewed+Coding) gene set, and filtered to select single-coverage long transcripts. For birds, we used the neutral model 'galGal6.phastCons77way.mod' (https://hgdownload.soe.ucsc.edu/goldenPath/galGal6/phastCons77way/galGal6.phastCons77way.mod) from the 77-way alignment. This all-species tree model was generated using the phyloFit program from the PHAST package (REV model, EM algorithm, medium precision) using multiple alignments of fourfold degenerate sites extracted from the 77-way alignment (msa_view). The fourfold degenerate sites were derived from the NCBI RefSeq gene set, and filtered to select single-coverage long transcripts.

Multiple species alignments and species selection

At the time we started this study (2020) there were premade multiple alignments that included a large number of mammals, birds and other amniotes, so we decided to conduct our search using the intersection of two large independent multiple alignments: one containing primarily mammals along with other amniotes ("100 pathways", https://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz100way/) and another one containing primarily birds and other amniotes ("77 pathways", http://hgdownload.cse.ucsc.edu/goldenPath/galGal6/multiz77way/).

To compute conservation of amniotes in the mammals analysis we extracted from the 100-way multiple alignment of vertebrate genomes the following species: human, (*Homo sapiens*, hg38; as reference genome), alligator (*Alligator mississippiensis*, allMis1), sea turtle (*Chelonia mydas*, cheMyd1), painted turtle (Chrysemys picta belli, chrPic2), Chinese turtle (*Pelodiscus sinensis*, pelSin1), spiny turtle (*Apalone spinifera*, apaSpi1), green lizard (*Anolis carolinensis*, anoCar2), clawed frog (*Xenopus tropicalis*, xenTro7) and coelacanth (*Latimeria chalumnae*, latCha1).

The species extracted from the 77-way multiple alignment of vertebrate genomes to compute conservation of amniotes in the aves analysis are the following: chicken (*Gallus gallus*, galGal6, as reference genome), American alligator (*Alligator mississippiensis*, allMis1), sea turtle (*Chelonia mydas*, cheMyd1), painted turtle (*Chrysemys picta belli*, chrPic2), Chinese turtle, (*Pelodiscus sinensis*, pelSin1), spiny turtle (*Apalone spinifera*, apaSpi), green lizard (*Anolis carolinensis*, anoCar2) and clawed frog (*Xenopus tropicalis*, xenTro9). It is important to note that there is a core of seven shared amniotes between the 77-way and the 100-way alignment involved in the computation of conserved elements.

To obtain the mammalian conserved sequences we extracted sequences of the following mammalian species from the publicly available 100-way multiple alignment keeping a balance in the representation of different branches of mammalian diversity: human (*H. sapiens*, hg38), great bushbaby (*Otolemur garnettii*, otoGar3), rhesus monkey (*Macaca mulatta*, rheMac3), mouse (*Mus musculus*, mm10), rabbit (*Oryctolagus cuniculus*, oryCun2), American pika (*Ochotona princeps*, ochPri3), pig, (*Sus scrofa*, susScr3), dolphin (*Tursiops truncatus*, turTru2), cow (*Bos taurus*, bosTau8), cat (*Felis catus*, felCat8), microbat (*Myotis lucifugus*, myoLuc2), elephant (*Loxodonta africana*, loxAfr3), tenrec (*Echinops telfairi*, echTel2), armadillo (*Dasypus novemcinctus*, dasNov3), opossum (*Monodelphis domestica*, monDom5), wallaby (*Macropus eugenii*, macEug2), platypus (*Ornithorhynchus anatinus*, ornAna1). This multiple alignment is named Mammals/Amniotes from now on.

To identify the avian conserved sequences we selected the following birds species from the 77-way multiple alignment of vertebrate genomes keeping a balance in the representation of different branches of avian diversity: chicken (*Gallus gallus*, galGal6), Japanese quail (*Coturnix japonica*, cotJap2), turkey (*Meleagris gallopavo*, melGal5), owl (*Tyto alba*, tytAlb1), rhinoceros hornbill (*Buceros rhinoceros silvestris*, bucRhi1), mallard duck (*Anas platyrhynchos*, anaPla1), bartailed trogon (*Apaloderma vittatum*, apaVit1), Anna's hummingbird (*Calypte anna*, calAnn1), common cuckoo (*Cuculus canorus*, cucCan1), killdeer (*Charadrius vociferus*, chaVoc2), northern fulmar (*Fulmarus glacialis*, fulGla1), red-crested turaco (*Tauraco erythrolophus*, tauEry1), hoatzin (*Opisthocomus hoazin*, opiHoa1), American flamingo (*Phoenicopterus_ruber_ruber*, phoRub1), rock pigeon (*Columba livia*, colLiv1), cuckoo-roller (*Leptosomus discolor*, lepDis1), northern carmine bee-eater (*Merops nubicus*, merNub1), Dalmatian pelican (*Pelecanus crispus*, pelCri1), great cormorant (*Phalacrocorax carbo*, phaCar1), white-tailed tropicbird (*Phaethon lepturus*, phaLep1), yellow-throated sandgrouse (*Pterocles gutturalis*, pteGut1), crested ibis (*Nipponia nippon*, nipNip1), little egret (*Egretta garzetta*, egrGar1), Adélie penguin (*Pygoscelis adeliae*, pygAde1), emperor penguin (*Aptenodytes forsteri*, aptFor1), red-legged seriema (*Cariama cristata*, carCri1), brown roatelo (*Mesitornis unicolor*, mesUni1), sunbittern (*Eurypyga helias*, eurHel1), black crowned crane (*Balearica pavonina*, balPav1), African houbara (*Chlamydotis undulata*, chlUnd1), saker falcon (*Falco cherrug*, falChe1), peregrine falcon (*Falco peregrinus*, falPer1), golden eagle (*Aquila chrysaetos*, aquChr2), white-tailed eagle (*Haliaeetus albicilla*, halAlb1), bald eagle (*Haliaeetus leucocephalus*, halLeu1), American crow (*Corvus brachyrhynchos*, corBra1), hooded crow (*Corvus cornix*, corCor1), rifleman (*Acanthisitta chloris*, acaChl1), collared flycatcher (*Ficedula albicollis*, ficAlb2), canary (*Serinus canaria*, serCan1), white-throated

sparrow (*Zonotrichia albicollis*, zonAlb1), medium ground finch (*Geospiza fortis*, geoFor1), zebra finch (*Taeniopygia guttata*, taeGut2), Tibetan ground jay (*Pseudopodoces humilis*, pseHum1), red-throated loon (*Gavia stellata*, gavSte1), chuck will's widow or nightjar (*Caprimulgus carolinensis*, capCar1), budgerigar parakeet (*Melopsittacus undulatus*, melUnd1), Puerto Rican parrot (*Amazona vittata*, amaVit1), scarlet macaw (*Ara macao*, araMac1), common mousebird (*Colius striatus*, colStr1), lesser downy woodpecker (*Picoides pubescens*, picPub1), ostrich (*Struthio camelus*, strCam1), white-throated tinamou (*Tinamus guttatus*, tinGut2). This multiple alignment is named Aves/Amniotes from now on.

Conserved elements computation

To identify conserved regions in amniotes (AmCSs; Fig. 1), we used the phastCons function program from the RPHAST package[34,90] with the following parameters: expected.length = 45, target.coverage = 0.3, rho = 0. 3, viterbi=TRUE, for each of the two sub-alignments described above (Mammals/Amniotes; Aves/Amniotes). The phastCons function returns a list containing parameter estimates, including an object (most.conserved) of type feat which describes conserved elements detected by the Viterbi algorithm. These conserved elements were then filtered according to our criteria for informative regions. A region was considered informative when alligator, lizard, and at least a turtle were present in the alignment and had no gaps in these species. The reported amniotes conserved elements resulted from the intersection of the conserved elements detected by the Viterbi algorithm and the informative regions in the alignment.

Mammalian conserved sequences (MCSs)

We identified conserved regions in mammals using the phastCons function with the same parameters values used for amniotes conserved elements computation (expected.length = 45, target.coverage = 0.3, rho = 0. 3, viterbi=TRUE), for a sub-alignment of "100 pathways" containing the mammalian species listed above. PhastCons most.-conserved results (conserved elements detected by the Viterbi algorithm) were then filtered according to our criteria for informative regions. A region was considered informative when the associated alignment fulfilled the following three conditions: i) hg38, otoGar3, mm10, oryCun2, ochPri3, susScr3, turTru2, bosTau8, felCat8, myoLuc2, ornAna1 were present in the alignment and these species had no gaps, ii) at least one of loxAfr3 or echTel2 was present in the alignment and this species had no gaps, iii) at least one of dasNov3, monDom5 or macEug2 was present in the alignment and this species had no gaps. The reported mammals' conserved elements resulted from the intersection of the conserved elements detected by the Viterbi algorithm and the informative regions in the alignment.

Aves conserved sequences (AvCSs)

We identified conserved regions in Aves using the phastCons function[34,90] with the same parameters values used for amniotes conserved elements computation (expected.length = 45, target.coverage = 0.3, rho = 0. 3, viterbi=TRUE), for a sub-alignment of "77 pathways" containing the aves species listed above. Conserved elements (phastCons most.conserved results detected by the Viterbi algorithm) were then filtered according to our criteria for informative regions. A region was considered informative when the associated alignment fulfilled the following four conditions: i) at least one of falChe1, falPer1, halLeu1 was present in the alignment and this species had no gaps, iii) at least one of taeGut2, geoFor1 was present in the alignment and this species had no gaps, iii) at least one of strCam1, tinGut2 was present in the alignment and this species had no gaps, iv) at least eleven species were present in the alignment and these species had no gaps. The reported Aves Conserved Sequences resulted from the intersection of the conserved elements detected by the Viterbi algorithm and the informative regions in the alignment.

Accelerated regions computation

The set of candidate regions to evaluate acceleration in mammals or aves was the result of the intersection between the amniotes conserved sequences (AmCSs) and the MCSs or the aves conserved regions (AvCSs). The two intersections of these sets resulted in conserved regions in both amniotes and mammals (MCRs) or amniotes and aves (AvCRs). This intersection is useful for ruling out those elements only conserved in amniotes (which could have been lost in mammals) or only conserved in mammals or aves (probably arisen de novo). Since we are interested in shifts specific to the transition from amniotes to either aves or mammals, computing conservation scores directly using the alignments that include both amniote and mammalian species or amniote and avian species would fail to detect these shift elements, as regions without shifts would show higher conservation scores. We split the conserved regions into fragments of size 25 bp and size 50 bp and ran phyloP to detect acceleration on these sets of split elements. Regularizing the lengths of the conserved elements simplifies the likelihood ratio tests, since, otherwise, we would have to consider the length dependency of the likelihood ratios when evaluating significance. To run the test, we added "MAMMALS" label to Newick tree in hg38.phastCons100way.mod and "AVES" label to the tree in galGal6.phastCons77way.mod to be able to specify the internal branch where acceleration was evaluated. We used the phyloP function program from the RPHAST package with the internal branch test (--branch option) to identify mammalian (parameter branches = "MAMMALS") or avian (parameter branches = "AVES") accelerated elements in the set of split sequences mentioned above. The neutral model in phyloP function was the same one used for the conserved elements computation (see above). To assess the significance of accelerated elements obtained as a result of phyloP function, we computed empirical p values using non-parametric simulations instead of relying on the assumption of a chi-square null distribution (see vignette 2 from http://compgen.cshl.edu/rphast/vignette2.pdf).

We began by extracting the regions from the original alignment that correspond to the conserved elements set (candidate regions to evaluate acceleration mentioned above), using the extract.feature.msa function.

We then generated 100,000 synthetic alignments by sampling with replacement from this set of regions and ran phyloP on these alignments to obtain our null distribution of log-likelihood ratios, applying phyloP in the same way as for the real data. We calculated empirical p values for the result of the phyloP function on real data and adjusted them for multiple comparisons using p.adjust with Benjamini–Hochberg method to compute approximate false discovery rates (FDRs) for each element. We extracted elements with FDR < 0.05 and treated them as accelerated elements in mammals or aves and putative functional regions. The genomic region may evolve faster or slower in all vertebrates; therefore, all branch lengths of the phylogenetic tree in this model are again fitted to a parameter to reflect the expected local substitution rate. Therefore, the internal branch test is a one-sided likelihood ratio test that compares the probability of the data given the null model to the probability under a model that allows for acceleration along the mammalian ancestor branch. P values from the internal branch test were adjusted for multiple comparisons using the FDR method (FDR[91];. Statistically significant accelerated regions were defined with an FDR threshold of 5%.

From these initial conservation and acceleration results, we defined each accelerated region in the mammalian (MARs) or the avian (AvARs) lineage as a set of conserved regions (i.e., the result of running the phastCons program on the alignment of amniote genomes) separated by a distance of <20 bp from each other and with a minimum size of 100 bp. We applied an additional filter to unify these regions into one accelerated region, taking into account that the accelerated changes remain fixed in the lineage of interest. Those regions whose

change was a "gap" were discarded, allowing only nucleotide changes to be obtained.

Classification of coding and noncoding accelerated sequences

To split elements into coding and noncoding sets we used annotations downloaded as bed files from the UCSC Table Browser. For mammals, annotations were built from genome hg38, track NCBI RefSeq, table RefSeq All, and output one BED record per Coding Exons. For Aves, annotations were built from genome galGal6, track NCBI RefSeq, table RefSeq All, and output one BED record per Coding Exons.

Elements' coordinates having a non-empty intersection with the coding exons annotation file were labelled as coding, and those having an empty intersection were labelled as noncoding elements.

GC-biased gene conversion detection

To identify regions under GC-biased gene conversion (gBGC), we used the phastBias method in the RPHAST package described by ref. 92. This approach employs a phylogenetic Hidden Markov Model (phyloHMM) to classify genomic regions into four states: conserved, neutral, conserved with gBGC, and neutral with gBGC. We analyzed the multiple sequence alignment of each accelerated region and used the previously defined neutral models with foreground branches being either mammals or aves. The resulting gBGC tracts, regions with posterior probability of being in a gBGC state > 0.5, were determined using default parameters (align, mod, foreground = NULL, do.bgc = TRUE, bgc = 3, estimate.bgc = FALSE, bgc.expected.length = 1000, estimate.bgc.expected.length = FALSE, bgc.target.coverage = 0.01, estimate.bgc.target.coverage = TRUE, sel = -2.01483, cons.expected.length = 45, cons.target.coverage = 0.3, estimate.scale = FALSE, post.probs = TRUE).

Density of accelerated elements in genomic regions

To calculate the density of accelerated elements, defined as the number of elements per window, we used the Bioconductor package karyoploteR[93] with a window size of 1e6 bp. We used the genomic coordinates for the accelerated regions in mammals (ncMARs/cMARs; hg38) and the accelerated regions in birds (ncAvARs/cAvARs; galGal6) as BED files or GRanges objects. Transcriptional unit genomic coordinates for human or chicken genes were obtained by creating one BED record for each whole gene, using the transcriptional start site and the transcription end values from the knownCanonical table from the UCSC Genes track. This canonical transcript reports the longest coding sequence for each entry and includes all the noncoding introns. We evaluated how many of the accelerated elements (ncMARs/cMARs and ncAvARs/cAvARs) overlapped these transcriptional units using the GRanges intersect operation in Bioconductor. We also calculated the distribution of accelerated sequences in topologically associated domains (TADs). For TADs, we used data from the fetal human cortical plate (GSE77565) as a reference in the human genome (hg38 version) and intersected this with ncMARs. For ncAvARs, we used chromatin contact maps derived from Hi-C studies performed on chicken (*G. gallus*) fibroblasts[51], annotated on the galGal6 genome.

GO analysis

While coding genes often have clearly annotated biological roles, noncoding regions usually lack functional annotation. To overcome this limitation, we used the rGREAT library (https://github.com/jokergoo/rGREAT), an implementation of the GREAT software (http://great.stanford.edu/public[94]), to assign biological meaning to noncoding genomic regions based on their proximity to annotated genes. rGREAT constructs region sets for each GO term and evaluates the enrichment of target genomic regions within these sets using a binomial model test. The null hypothesis of the rGREAT binomial model assumes that input regions are uniformly distributed across the genome. Our target genomic regions are associated with noncoding accelerated elements, which, by design, are also linked to conserved elements. So, the goal is to identify GO terms associated with enrichment in target regions due to acceleration rather than conservation. This strategy allows reporting a GO term as associated with the accelerated regions set if the behavior of the accelerated regions set is independent of its conserved nature, and the adjusted *p*-value returned from rGREAT on the accelerated regions set is less than 0.05. To address the first condition, we calculated the distribution of 'observed_regions_hits' associated with conserved elements, as returned by rGREAT for each GO term. This distribution was derived from 5000 samples of conserved regions, each matched in size and chromosome distribution to the accelerated regions set we aim to evaluate later. The ontologies considered included GO:BP, GO:CC, GO:MF. The rGREAT analysis was performed using the following parameter values: gr: each of the 5000 conserved elements sample; gene_sets: GO:BP, GO:CC, GO:MF; tss_source: TxDb.Hsapiens.UCSC.hg38.knownGene, TxDb.Ggallus.UCSC.galGal6.refGene; biomart_dataset: NULL, 'hsapiens_gene_ensembl', 'Ggallus_gene_ensembl'; default parameters: min_gene_set_size = 5, mode = "basalPlusExt", extend_from = "TSS", basal_upstream = 5000, basal_downstream = 1000, extension = 1000000, extended_tss = NULL, background = NULL, exclude = "gap". The results of these simulations were stored in matrices with 5000 rows, where each column represents a GO term. Each element (i, j) in the matrix corresponds to the 'observed_regions_hits' value for GO term j in simulation i. The distribution of 'observed_regions_hits' values for conserved elements associated with a specific GO term corresponds to the values in the column representing that term in the matrix. We then use these distribution values to compute the empirical *p*-value for accelerated regions. To evaluate GO term enrichment in genomic regions associated with accelerated elements, we perform the rGREAT analysis using the same parameters as in calculating the 'observed_regions_hits' distribution in conserved elements. The empirical *p*-value for each GO term in the accelerated region set is computed as the proportion of 'observed_regions_hits' values in its distribution within conserved elements that are greater than the obtained 'observed_regions_hits' value for the accelerated region set. These *p* values are corrected for multiple comparisons, and the Benjamini–Hochberg method was applied to compute approximate false discovery rates (FDRs) for each term. We considered a term significant when its adjusted *p*-value returned from rGREAT was less than 0.05 and the adjusted empirical *p*-value was less than 0.05. The rGREAT 'biomart_dataset' parameter allows us to specify different GO gene set annotations. If this parameter is set to null, the default annotations databases used are org.Gg.egGO2ALLEGS (https://doi.org/10.18129/B9.bioc.org.Gg.eg.db) and org.Hs.egGO2ALLEGS (https://doi.org/10.18129/B9.bioc.org.Hs.eg.db). If 'biomart_dataset' is set to 'hsapiens_gene_ensembl' or 'ggallus_gene_ensembl' annotations are retrieved from the Bioconductor package BioMartGOGeneSets (https://doi.org/10.18129/B9.bioc.BioMartGOGeneSets). Since these two annotation databases (default and BioMartGOGeneSets) produce different sets of enriched GO terms, we report the union of significant results from both methods.

To summarize the GO terms result from this analysis, we generated a dot plot for each ontology (GO:BP, GO:MF, GO:CC) using ggplot2 library[95] (https://ggplot2.tidyverse.org). We used the getRegionGeneAssociations function from the rGREAT library to associate noncoding regions with genes. Then, for each GO term, we calculated the gene proportion as the number of genes associated with the noncoding genomic regions divided by the number of genes annotated for that term. Each dot plot displays the top 30 significant GO terms ordered by region fold enrichment descending, colour represents empirical *p*-value adjusted, and dot size indicates the proportion of genes for the term.

Signatures of regulatory activity of accelerated elements

To assess the regulatory potential of accelerated elements, we used the package LOLA (Locus Overlap Analysis) that performs enrichment analysis on genomic regions integrating many epigenetic regulatory databases. LOLA identifies significant overlaps or intersections between different data sets[65]. To this end, a set of genomic regions of interest is compared to a large compendium of existing biologically annotated data, and the observed patterns of overlap are used to interpret the initial region set. One of the advantages of LOLA is that it includes a central reference database assembled from public data, including the CODEX database[96], ENCODE transcription factor binding sites (ENCODE Uniform TFBS composite track in UCSC), the Cistrome database[97], and DNase I hypersensitive sites[65,98]. The enrichment analysis performed by the LOLA library is based on three lists of genomic regions: the target set (in our case, noncoding accelerated elements), the 'universe' set of background genomic regions (in this case, noncoding conserved elements) and a set of genomic regions against which they overlap (in this case, curated regulatory elements from the LOLAcore database or chicken-specific regulatory data from the FAANG project, available at https://figshare.com/articles/dataset/Chicken_FAANG/20032103). LOLA identifies all genomic regions from a query set that overlap with each region set in the reference database. A single shared base pair is sufficient for regions to count as overlapping. Considering each region as independent, LOLA uses Fisher's exact test with false discovery rate correction to assess the significance of overlap in each pairwise comparison. To perform the test, a contingency table is constructed with columns representing the categories 'regions in the reference database' and 'regions not in the reference database', and rows representing the categories 'regions in the query set' and 'regions not in the query set (but within the universe)'. Then, the contingency table elements ((a, b), (c, d)) are defined as follows:

a. The number of regions both in the query and in the reference database (reported as support).
b. The number of regions in the query that are not in the reference database.
c. The number of regions in the reference database and in the universe, but not in the query.
d. The number of regions in the universe that are in neither the query nor the reference database.

Fisher's exact test is used to evaluate whether there is a significant difference in the proportion of hits with the reference database between two sets of genomic regions: those defined in the query and those in the universe excluding the query. In our analysis, we apply this test to assess whether noncoding accelerated elements differ in their proportion of hits compared to noncoding conserved elements (excluding the accelerated elements from this set). The odds ratio indicates how many times more likely positive cases are compared to negative cases. Positive cases are 'regions both in the query and in the reference database' (a) and 'regions in the universe that are in neither the query nor the reference database' (d). Negative cases are 'regions in the query that are not in the reference database' (b) and 'regions in the reference database and in the universe but not in the query' (c). Then, it is calculated as (a*d)/(b*c), and we include only those reference databases where the number of positive cases is greater than the number of negative cases. LOLA returns 'pValueLog', 'oddsRatio', and 'support' as part of its output. We consider a reference database to be informative if it meets all of the following criteria: pValueLog > 1.3 (equivalent to p value < 0.05), oddsRatio > 1, and support ≥6 for mammals or ≥4 for aves. The 'description' column, included in the analysis output, helps interpret the type and biological context of the regions annotated in each reference database.

Enhancer assays in transgenic zebrafish

All experiments were performed on the wild-type AB strain from the Zebrafish International Resource Centre at the University of Oregon, and a wild-type strain from the IBR-CONICET aquarium in the city of Rosario. Adult zebrafish were maintained at 28 °C in a 14/10 h light/dark cycle in a fully automatic aquatic habitat aquarium. All experiments were performed following the Guide for the Care and Use of Laboratory Animals and were approved by the local institutional animal care and use committee of INGEBI. Conservation blocks (phastCons) within the genome are considered potentially functional regions, so we considered this entire block to test its activity. Therefore, genomic regions of phastCons (cloned regions) containing each accelerated region were amplified by PCR from human and *Alligator mississippiensis* genomic DNA samples using primers described in Supp. Data 28. These sequences were then individually cloned first into the pENTR/D-TOPO vector (ThermoFisher) and subsequently subcloned into the pXIG_cFos vector containing the minimal cFos promoter fused to the EGFP reporter gene that was kindly donated by Andy McCallion. Embryos were obtained by natural mating, and each construct was co-injected with transposase mRNA into one- to two-cell zebrafish embryos. Embryos were raised in E3 medium (NaCl 5 mM, KCl 0.17 mM, $CaCl_2$ 0.33 mM, 0.33 mM MgSO4) at 28.3° C until adulthood and raised for stable germline insertion. At least three independent transgenic lines were established per element. We defined an element as an active enhancer when at least three independent lines showed a consistent EGFP expression pattern. When necessary, 0.1 mM 1-phenyl-2-thiourea was added to E3 medium to prevent pigment formation. Microscopy was performed on tricaine-anesthetized embryos mounted in 3% methylcellulose. Whole-mount images were taken on an Olympus BX41 fluorescence microscope with an Olympus DP71 digital camera or a Leica DM2500 with a Leica DFC camera using the LAS X software. In cases where *EGFP* expression was not detected, we performed PCR to confirm the insertion of the cassette into the zebrafish genome using primers specific for EGFP (Supp. Data 28) and the tested transgene. Genomic DNA was obtained from 20 embryos at 72 hpf collected using the Wizard genomic DNA purification kit (Promega). Images were analyzed, cropped and edited for displaying purposes using Adobe Photoshop 26.9.0.

Reporting summary

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

Data availability

The datasets generated and/or analyzed during the current study are included in this published article (and its supplementary data files) and available in the ncomms-24-47757 repository, https://github.com/paulati/ncomms-24-47757. Genomic aligned sequences are public and contained in two large independent multiple alignments: one containing primarily mammals along with other amniotes ("100 pathways", https://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz100way/) and another one containing primarily birds and other amniotes ("77 pathways", http://hgdownload.cse.ucsc.edu/goldenPath/galGal6/multiz77way/). Neutral model: In mammals, we used the neutral model file 'hg38.phastCons100way.mod' (https://hgdownload.cse.ucsc.edu/goldenpath/hg38/phastCons100way/hg38.phastCons100way.mod) from 100-way alignment. For birds, we used the neutral model 'galGal6.phastCons77way.mod' (https://hgdownload.soe.ucsc.edu/goldenPath/galGal6/phastCons77way/galGal6.phastCons77way.mod) from the 77-way alignment. For mammalian epigenetic data we used the CODEX database, ENCODE and the Cistrome database. Chicken-specific regulatory data is from the FAANG project, available at https://figshare.com/articles/dataset/Chicken_FAANG/20032103).

Code availability

The computation of accelerated elements was performed on Amazon Web Services using t2.2xlarge EC2 instances running Ubuntu 20.04. The project code and data are available at refs. 88 and 89.

References

1. Machado, J. P. et al. Positive selection linked with generation of novel mammalian dentition patterns. *Genome Biol. Evol.* **8**, 2748–2759 (2016).
2. Khan, I. et al. Mammalian keratin associated proteins (KRTAPs) subgenomes: disentangling hair diversity and adaptation to terrestrial and aquatic environments. *BMC Genomics* **15**, 779 (2014).
3. Taye, M. et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mamm. Genome* **28**, 528–541 (2017).
4. Pisciottano, F. et al. Inner ear genes underwent positive selection and adaptation in the mammalian lineage. *Mol. Biol. Evol.* **36**, 1653–1670 (2019).
5. Trigila, A. P. et al. Accelerated evolution analysis uncovers PKNOX2 as a key transcription factor in the mammalian cochlea. *Mol. Biol. Evol.* **40**, sn (2023).
6. Trigila, A. P., Pisciottano, F. & Franchini, L. F. Hearing loss genes reveal patterns of adaptive evolution at the coding and non-coding levels in mammals. *BMC Biol.* **19**, 244 (2021).
7. Hubisz, M. J. & Pollard, K. S. Exploring the genesis and functions of human accelerated regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).
8. Franchini, L. F. & Pollard, K. S. Human evolution: the non-coding revolution. *BMC Biol.* **15**, 89 (2017).
9. Booker, B. M. et al. Bat accelerated regions identify a bat forelimb specific enhancer in the hoxd locus. *PLoS Genet.* **12**, e1005738 (2016).
10. Eckalbar, W. L. et al. Transcriptomic and epigenomic characterization of the developing bat wing. *Nat. Genet.* **48**, 528–536 (2016).
11. Whalen, S. & Pollard, K. S. Enhancer function and evolutionary roles of human accelerated regions. *Annu. Rev. Genet.* **56**, 423–439 (2022).
12. Ferris, E., Abegglen, L. M., Schiffman, J. D. & Gregg, C. Accelerated evolution in distinctive species reveals candidate elements for clinically relevant traits, including mutation and cancer resistance. *Cell Rep.* **22**, 2742–2755 (2018).
13. Holloway, A. K., Bruneau, B. G., Sukonnik, T., Rubenstein, J. L. & Pollard, K. S. Accelerated evolution of enhancer hotspots in the mammal ancestor. *Mol. Biol. Evol.* **33**, 1008–1018 (2016).
14. Christmas, M. J. et al. Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023).
15. Keough, K. C. et al. Three-dimensional genome rewiring in loci with human accelerated regions. *Science* **380**, eabm1696 (2023).
16. Seki, R. et al. Functional roles of Aves class-specific cis-regulatory elements on macroevolution of bird-specific features. *Nat. Commun.* **8**, 14229 (2017).
17. Sackton, T. B. et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
18. Stiller, J. et al. Complexity of avian evolution revealed by family-level genomes. *Nature* **629**, 851–860 (2024).
19. Feng, S. et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
20. Crompton, A. W., Taylor, C. R. & Jagger, J. A. Evolution of homeothermy in mammals. *Nature* **272**, 333–336 (1978).
21. Bennett, A. F. & Ruben, J. A. Endothermy and activity in vertebrates. *Science* **206**, 649–654 (1979).
22. Hillenius, W. J. & Ruben, J. A. The evolution of endothermy in terrestrial vertebrates: who? when? why?. *Physiol. Biochem. Zool.* **77**, 1019–1042 (2004).
23. Koteja, P. Energy assimilation, parental care and the evolution of endothermy. *Proc. Biol. Sci.* **267**, 479–484 (2000).
24. Koshiba-Takeuchi, K. et al. Reptilian heart development and the molecular basis of cardiac chamber evolution. *Nature* **461**, 95–98 (2009).
25. Jensen, B., Wang, T., Christoffels, V. M. & Moorman, A. F. M. Evolution and development of the building plan of the vertebrate heart. *Biochim. Biophys. Acta* **1833**, 783–794 (2013).
26. Wang, T. EVOLUTION The beat goes on: Why is the alligator heart so similar to the hearts of birds and mammals?. *Elife* **7**, e36882 (2018).
27. Stephenson, A., Adams, J. W. & Vaccarezza, M. The vertebrate heart: an evolutionary perspective. *J. Anat.* **231**, 787–797 (2017).
28. Snyder, G. K. & Sheafor, B. A. Red blood cells: centerpiece in the evolution of the vertebrate circulatory system1. *Integr. Comp. Biol.* **39**, 189–198 (2015).
29. Hawkey, C. M., Bennett, P. M., Gascoyne, S. C., Hart, M. G. & Kirkwood, J. K. Erythrocyte size, number and haemoglobin content in vertebrates. *Br. J. Haematol.* **77**, 392–397 (1991).
30. Farmer, C. G. Parental care: the key to understanding endothermy and other convergent features in birds and mammals. *Am. Nat.* **155**, 326–334 (2000).
31. Manley, G. A. Comparative auditory neuroscience: understanding the evolution and function of ears. *J. Assoc. Res. Otolaryngol.* **18**, 1–24 (2017).
32. Tyack, P. L. A taxonomy for vocal learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20180406 (2020).
33. Wu, Y. & Wang, H. Convergent evolution of bird-mammal shared characteristics for adapting to nocturnality. *Proc. Biol. Sci.* **286**, 20182185 (2019).
34. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
35. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
36. Macintyre, G. et al. Association of NPAS3 exonic variation with schizophrenia. *Schizophr. Res.* **120**, 143–149 (2010).
37. Michaelson, J. J. et al. Neuronal PAS domain proteins 1 and 3 are master regulators of neuropsychiatric risk genes. *Biol. Psychiatry* **82**, 213–223 (2017).
38. Li, Y. et al. Npas3 deficiency impairs cortical astrogenesis and induces autistic-like behaviors. *Cell Rep.* **41**, 111767 (2022).
39. Stanco, A. et al. NPAS1 represses the generation of specific subtypes of cortical interneurons. *Neuron* **84**, 940–953 (2014).
40. Kamm, G. B., Pisciottano, F., Kliger, R. & Franchini, L. F. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol. Biol. Evol.* **30**, 1088–1102 (2013).
41. Kostka, D., Holloway, A. K. & Pollard, K. S. Developmental loci harbor clusters of accelerated regions that evolved independently in ape lineages. *Mol. Biol. Evol.* **35**, 2034–2045 (2018).
42. Salluzzo, M. et al. The role of IgLON cell adhesion molecules in neurodegenerative diseases. *Genes* **14**, 1886 (2023).
43. Mella, S., Soula, C., Morello, D., Crozatier, M. & Vincent, A. Expression patterns of the coe/ebf transcription factor genes during chicken and mouse limb development. *Gene Expr. Patterns* **4**, 537–542 (2004).
44. Merchán, P., Bardet, S. M., Puelles, L. & Ferran, J. L. Comparison of pretectal genoarchitectonic pattern between quail and chicken embryos. *Front. Neuroanat.* **5**, 23 (2011).
45. Bielle, F. et al. Slit2 activity in the migration of guidepost neurons shapes thalamic projections during development and evolution. *Neuron* **69**, 1085–1098 (2011).
46. Kida, Y., Maeda, Y., Shiraishi, T., Suzuki, T. & Ogura, T. Chick Dach1 interacts with the Smad complex and Sin3a to control AER formation and limb development along the proximodistal axis. *Development* **131**, 4179–4187 (2004).
47. Enard, W. et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).

48. Scharff, C. & Haesler, S. An evolutionary perspective on FoxP2: strictly for the birds?. *Curr. Opin. Neurobiol.* **15**, 694–703 (2005).
49. Caporale, A. L., Gonda, C. M. & Franchini, L. F. Transcriptional enhancers in the FOXP2 locus underwent accelerated evolution in the human lineage. *Mol. Biol. Evol.* **36**, 2432–2450 (2019).
50. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
51. Fishman, V. et al. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* **47**, 648–665 (2019).
52. Gu, Z. & Hübschmann, D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics* **39**, btac745 (2023).
53. Steimle, J. D. & Moskowitz, I. P. TBX5: A key regulator of heart development. *Curr. Top. Dev. Biol.* **122**, 195–221 (2017).
54. Shinozaki, K. et al. Absence of Cajal-Retzius cells and subplate neurons associated with defects of tangential cell migration from ganglionic eminence in Emx1/2 double mutant cerebral cortex. *Development* **129**, 3479–3492 (2002).
55. Konno, D. et al. The mammalian DM domain transcription factor Dmrta2 is required for early embryonic development of the cerebral cortex. *PLoS One* **7**, e46577 (2012).
56. Haricharan, S. & Li, Y. STAT signaling in mammary gland differentiation, cell survival and tumorigenesis. *Mol. Cell. Endocrinol.* **382**, 560–569 (2014).
57. Roarty, K. & Serra, R. Wnt5a is required for proper mammary gland development and TGF-beta-mediated inhibition of ductal growth. *Development* **134**, 3929–3939 (2007).
58. Slepicka, P. F., Somasundara, A. V. H. & Dos Santos, C. O. The molecular basis of mammary gland development and epithelial differentiation. *Semin. Cell Dev. Biol.* **114**, 93–112 (2021).
59. Ruest, L.-B. & Clouthier, D. E. Elucidating timing and function of endothelin-A receptor signaling during craniofacial development using neural crest cell-specific gene deletion and receptor antagonism. *Dev. Biol.* **328**, 94–108 (2009).
60. Zou, D. et al. Eya1 gene dosage critically affects the development of sensory epithelia in the mammalian inner ear. *Hum. Mol. Genet.* **17**, 3340–3356 (2008).
61. Anthwal, N., Joshi, L. & Tucker, A. S. Evolution of the mammalian middle ear and jaw: adaptations and novel structures. *J. Anat.* **222**, 147–160 (2013).
62. Zhao, Y. et al. LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc. Natl. Acad. Sci. USA* **104**, 13182–13186 (2007).
63. Nakatani, T., Minaki, Y., Kumai, M., Nitta, C. & Ono, Y. The c-Ski family member and transcriptional regulator Corl2/Skor2 promotes early differentiation of cerebellar Purkinje cells. *Dev. Biol.* **388**, 68–80 (2014).
64. Balanoff, A. et al. Quantitative functional imaging of the pigeon brain: implications for the evolution of avian powered flight. *Proc. Biol. Sci.* **291**, 20232172 (2024).
65. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).
66. Pan, Z. et al. An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. *Sci. Adv.* **9**, eade1204 (2023).
67. Thisse, B. et al. Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol.* **77**, 505–519 (2004).
68. Ma, C. et al. Fidgetin interacting with microtubule end binding protein EB3 affects axonal regrowth in spinal cord injury. *Neural Regen. Res.* **18**, 2727–2732 (2023).
69. Boudjadi, S., Chatterjee, B., Sun, W., Vemu, P. & Barr, F. G. The expression and function of PAX3 in development and disease. *Gene* **666**, 145–157 (2018).
70. Jensen, K. B. et al. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
71. Mattioli, F. et al. De novo frameshift variants in the neuronal splicing factor NOVA2 result in a common c-terminal extension and cause a severe form of neurodevelopmental disorder. *Am. J. Hum. Genet.* **106**, 438–452 (2020).
72. GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
73. Hoshino, H. et al. KSGal6ST is essential for the 6-sulfation of galactose within keratan sulfate in early postnatal brain. *J. Histochem. Cytochem.* **62**, 145–156 (2014).
74. Ocampo Daza, D. & Haitina, T. Reconstruction of the carbohydrate 6-O sulfotransferase gene family evolution in vertebrates reveals novel member, CHST16, lost in amniotes. *Genome Biol. Evol.* **12**, 993–1012 (2020).
75. Liu, A. et al. GC-biased gene conversion drives accelerated evolution of ultraconserved elements in mammalian and avian genomes. *Genome Res.* **33**, 1673–1689 (2023).
76. Pollard, K. S. et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
77. Kostka, D., Hubisz, M. J., Siepel, A. & Pollard, K. S. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. *Mol. Biol. Evol.* **29**, 1047–1057 (2012).
78. Goffinet, A. M. The evolution of cortical development: the synapsid-diapsid divergence. *Development* **144**, 4061–4077 (2017).
79. Franchini, L. F. Genetic mechanisms underlying cortical evolution in mammals. *Front Cell Dev. Biol.* **9**, 591017 (2021).
80. Molnár, Z. et al. Evolution and development of the mammalian cerebral cortex. *Brain Behav. Evol.* **83**, 126–139 (2014).
81. Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. R. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130025 (2013).
82. Berasain, L., Beati, P., Trigila, A. P., Rubinstein, M. & Franchini, L. F. Accelerated evolution in the human lineage led to gain and loss of transcriptional enhancers in the RBFOX1 locus. *Sci. Adv.* **10**, eadl1049 (2024).
83. Caporale, A. L., Cinalli, A. R., Rubinstein, M. & Franchini, L. F. The human accelerated region HAR202 controls NPAS3 expression in the developing forebrain displaying differential enhancer activity between modern and archaic human sequences. *Mol. Biol. Evol.* **41**, msae186 (2024).
84. Kamm, G. B., López-Leal, R., Lorenzo, J. R. & Franchini, L. F. A fast-evolving human NPAS3 enhancer gained reporter expression in the developing forebrain of transgenic mice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130019 (2013).
85. Deng, C. et al. Massively parallel characterization of regulatory elements in the developing human cortex. *Science* **384**, eadh0559 (2024).
86. Uebbing, S. et al. Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. USA.* **118**, e2007049118 (2021).
87. Whalen, S. et al. Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* **111**, 857–873.e8 (2023).
88. Beati, P. & Franchini, L. F. cladeAcc. *Figshare* https://doi.org/10.6084/m9.figshare.29941031 (2025).
89. Trigila, A. T. et al. ncomms-24-47757. *Figshare* https://doi.org/10.6084/m9.figshare.29941478 (2025).
90. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

91. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* (1995).
92. Capra, J. A., Hubisz, M. J., Kostka, D., Pollard, K. S. & Siepel, A. A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet.* **9**, e1003684 (2013).
93. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).
94. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
95. Wickham, H. *Ggplot2*. (Springer International Publishing, Basel, Switzerland, 2016).
96. Sánchez-Castillo, M. et al. CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* **43**, D1117–D1123 (2015).
97. Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).
98. Sheffield, N. C. et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* **23**, 777–788 (2013).
99. Gearty, W. & Jones, L. A. rphylopic: An R package for fetching, transforming, and visualising PhyloPic silhouettes. *Methods Ecol. Evol.* **14**, 2700–2708 (2023).

Acknowledgements

A.P.T. had a doctoral fellowship from CONICET. This work was supported by Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT-2021-I-A-00686 to L.F.F.).

Author contributions

L.F.F. designed and supervised the project. A.P.T. conducted experiments and bioinformatics analysis. P.B. contributed to experimental design, calculation of the accelerated elements, and GO enrichment analysis. A.P.T., P.B., and L.F.F. analyzed data. D.M. was essential for animal care and microinjections. D.M., Pd.l.V. and C.J. performed particular zebrafish experiments. A.P.T., P.B., and L.F.F. wrote the manuscript. L.F.F. provided reagents and analytical tools. All authors edited and approved the final version of this report.

Competing interests

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-64134-2.

Correspondence and requests for materials should be addressed to Lucía F. Franchini.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.