PNAS RESEARCH ARTICLE | EVOLUTION
BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Homeotic and nonhomeotic patterns in the tetrapod vertebral formula

Rory T. Cerbus[a,b,1], Ichiro Hiratani[a], and Kyogo Kawaguchi[b,c,d,e,1]

Edited by Denis Duboule, College de France, Paris, France; received June 7, 2024; accepted September 10, 2024

Vertebrate development and phylogeny are intimately connected through the vertebral formula, the numerical distribution of vertebrae along the body axis into different categories such as neck and chest. A key window into this relationship is through the conserved *Hox* gene clusters. *Hox* gene expression boundaries align with vertebral boundaries, and their manipulation in model organisms often results in the transformation of one vertebral type into its neighbor, a homeotic transformation. If the variety in the vertebrate body plan is produced by homeotic shifts, then the number of adjacent vertebrae will be inversely related when making interspecies comparisons since the gain in one vertebra is due to the loss in its neighbor. To date, such a pattern across species consistent with homeotic transitions has only been found in the thoracolumbar vertebral count of mammals. To further investigate potential homeotic relationships in other vertebrate classes and along the entire body axis, we compiled a comprehensive dataset of complete tetrapod vertebral formulas and systematically searched for patterns by analyzing combinations of vertebrae. We uncovered mammalian homeotic patterns and found balances between distal vertebrae not anticipated by a *Hox*-vertebral homeotic relationship, including one that emerged during the progression from theropods to birds. We also identified correlations between vertebral counts and intergenic distances in the *HoxB* gene cluster which do not align with the common picture of a colinear relationship between *Hox* expression and vertebral categories. This quantitative approach revises our expectations for the diversity of a *Hox*-mediated vertebrate body plan.

vertebral column | evolution | *Hox* code

Significance

Tetrapods have different numbers of vertebrae in the neck, torso, and lower parts of their bodies. These numbers can follow patterns anticipated by the relationship between vertebrae and the highly conserved *Hox* genes, but to date, the only confirmed example is the inversely related fluctuations in the thoracic and lumbar counts. Here, we collected the most comprehensive dataset of complete vertebral counts to date to look for further patterns across tetrapods. We uncover how the entire mammalian vertebral axis displays inverse patterns like the thoracolumbar constraint, while other tetrapod classes exhibit distinct, unanticipated patterns. We study one of these, an anterior–posterior balance, which is present in birds, bats, and some extinct flying theropods.

Patterns in vertebrae both unite and separate a wide range of animal groups. Based on its location, function, and morphology, an individual vertebra can be classified into one of five different categories: *cervical* (neck), *thoracic* (rib-bearing dorsal), *lumbar* (rib-less dorsal), *sacral* (pelvic), and *caudal* (tail) (Fig. 1*A*). The numerical distribution of the vertebra in these categories is termed the vertebral formula, which has been studied since the inception of modern comparative anatomy (1–8). On account of the vertebral formula alone, a bird can easily be distinguished from a mammal (1), and even orders within mammals are differentiable (9–11).

Patterns in the vertebral formula are identified in the context of a phylogenetic tree (Fig. 1*B*). Most mammals have seven cervical vertebrae while birds exhibit broad variability (3, 4, 7), justifying the significant attention given to the neck. Along with more recent discoveries of patterns further along the body axis, such as the weak invariance of the mammalian thoracolumbar count (9, 11), these vertebral patterns have generated important insights and questions about anatomy (12, 13) and trait evolution (10, 11, 14).

A major reason for the attention given to vertebral patterns is their manifest but mysterious relationship with the *Hox* genes, a set of strongly conserved homeobox genes that control the body plan of vertebrates and even invertebrates including Cnidaria (15). Within tetrapods these *Hox* genes are arranged in a fixed linear order in four clusters (16), and their sequential expression pattern reflects the serialization of the vertebrae, the phenomenon of colinearity (17). Although colinearity suggests a simple relationship between *Hox* genes and vertebral boundaries, a *Hox* code (18), a number of experiments have demonstrated that the *Hox*-vertebral relationship is not so straightforward (19–21). Irrespective of the detailed mechanism, in experiments manipulating *Hox* gene expression, the dominant phenotype is a shift in the vertebral boundaries, which can be interpreted as homeotic transformations from one vertebra into its neighbor (22–32). In some cases, even the manipulation of other genes involved in the prevertebral somite formation will yield homeotic transformations (25), further suggesting the ubiquity of this phenomenon.

Author affiliations: [a]Nonequilibrium Physics of Living Matter RIKEN Hakubi Research Team, RIKEN Center for Biosystems Dynamics Research, Kobe 650-0047, Japan; [b]Laboratory for Developmental Epigenetics, RIKEN Center for Biosystems Dynamics Research, Kobe 650-0047, Japan; [c]RIKEN Cluster for Pioneering Research, Kobe, Japan; [d]Institute for Physics of Intelligence, The University of Tokyo, Hongo, Tokyo 113-0033, Japan; and [e]Universal Biology Institute, The University of Tokyo, Tokyo 113-0033, Japan

Author contributions: R.T.C. and K.K. designed research; R.T.C., I.H., and K.K. performed research; R.T.C. and K.K. analyzed data; and R.T.C. and K.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: rory.cerbus@riken.jp or kyogo.kawaguchi@riken.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2411421121/-/DCSupplemental.

Published November 11, 2024.

Although experiments on the *Hox*-vertebral relationship are typically performed on a single model species, it is natural to suppose that homeotic transformations act as an important source of variation in the vertebrate body plan across species, giving rise to the different vertebral formulas found among tetrapods (11, 33–35). We can test this hypothesis by looking for signatures or patterns when comparing the vertebral formula of different species; these homeotic shifts should yield a relatively fixed sum of combined counts between adjacent vertebrae since the sum total of the two vertebrae is not changed by one transforming into the other (9). More generally, these transformations can produce an inverse relation or anticorrelation since an increase in one vertebra will correlate with a decrease in the other which is not necessarily of the same magnitude.

In principle, five vertebral categories with homeotic shifts between neighbors can yield four inverse relationships, which we thus describe as "homeotic" vertebral patterns (Fig. 1*C*). Using the shorthands C, T, L, S, and Ca to represent the number of vertebrae in each vertebral category (Cervical, Thoracic, Lumbar, Sacral, and Caudal, respectively; Fig. 1*A*), we can represent combined sums as $C + T$, $T + L$, $L + S$, and $S + Ca$. To date, the only well-established example of a homeotic pattern is the relatively fixed thoracolumbar count ($T+L$) in mammals (9), as well as the possibility that a homeotic transformation is behind the deviation from the mammalian "rule of seven" ($C = 7$) for sloths and manatees (14, 39–41) (see also *SI Appendix*, Fig. S1). And yet the expectation of *Hox*-mediated homeotic patterns is so strong that it is frequently assumed in the analysis of even extinct species whose genomic sequence, and thus direct information about the *Hox* genes, is unavailable (11, 33–35). Here, we ask whether additional homeotic vertebral patterns, or any other patterns, are present further along the body axis in not only mammals but other tetrapods as well.

To search for vertebral patterns we construct a broad dataset of complete vertebral formulas for 388 species, representing four tetrapod classes (Mammalia, Aves, Reptilia, Amphibia), 42 orders, and 175 families (42). This dataset was assembled from a myriad of data sources such as CT scans, photographs of

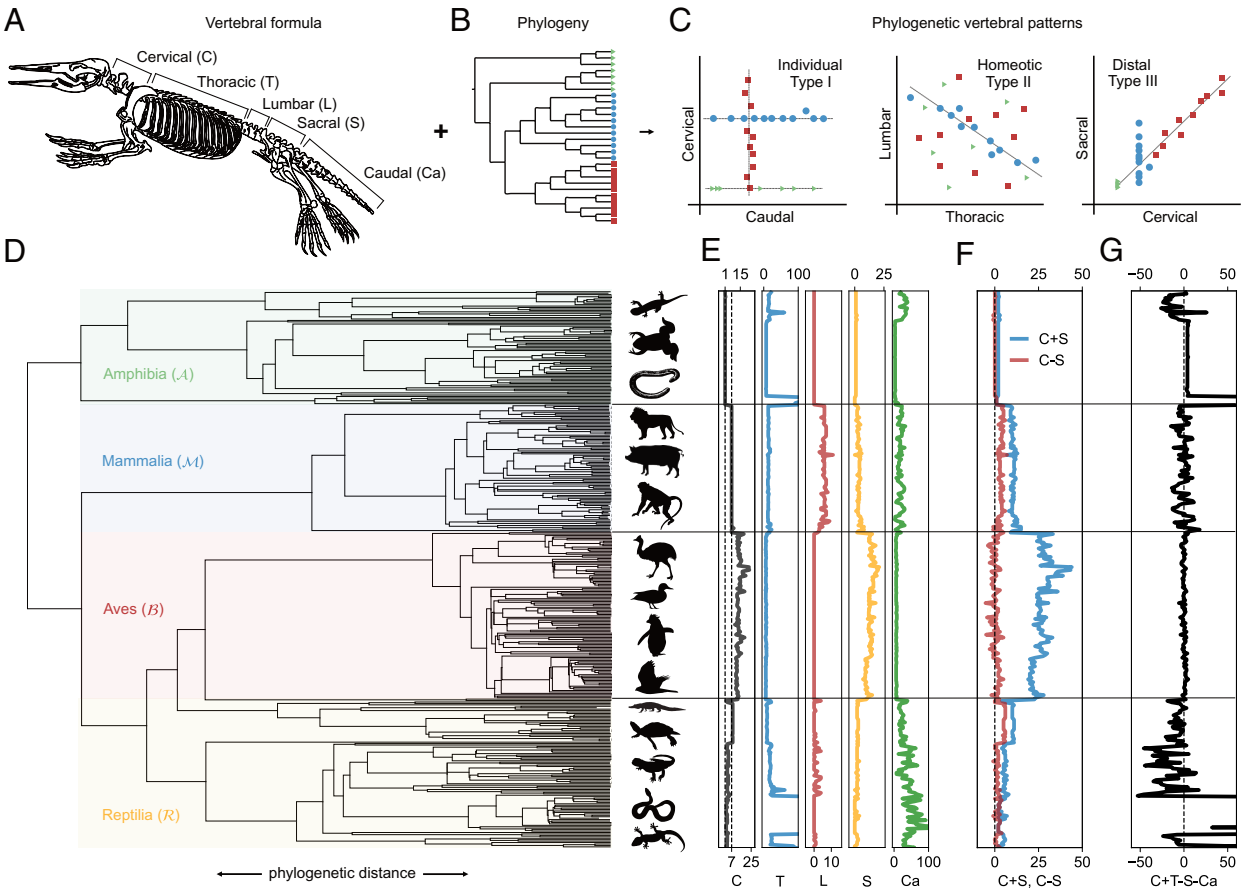

Fig. 1. Vertebral formula and phylogeny. (*A*–*C*) Schematic of the approach to determine vertebral patterns. (*A*) Sketch of a platypus (*Ornithorhynchus anatinus*) skeleton reproduced from Cuvier (36). The numbers of vertebrae for each category yield the vertebral formula. (*B*) Schematic example of a phylogenetic tree distinguishing three different groups. Analyzing the vertebral formulas in the context of the tree yields the vertebral patterns. (*C*) We identify at least three types of constraints. For individual or Type-I constraints, the vertebral count in a single category is nearly constant within a specific branch of the tree. This same category can be very plastic or take on different constant values in other branches. In homeotic or Type-II constraints, adjacent vertebrae are anticorrelated and sum to a nearly constant value. In distal or Type-III constraints, nonadjacent vertebrae are positively correlated and can be numerically balanced. (*D*–*G*) Plots of the vertebral data arranged according to a phylogenetic tree (37, 38). (*D*) A phylogenetic tree containing all the species for which we have the full vertebral count. (We exclude some bird data to avoid unbalancing the tree, but it is used in later analysis.) The tree has been organized so that the four classes of tetrapods, Amphibia, Mammalia, Aves (birds), and Reptiles, are arranged vertically. The silhouettes to the right of the tree are a nonexhaustive representation of groups within the tetrapod classes. From top to bottom, they are (with individual examples in parentheses), within Amphibia: Urodela (salamanders), Anura (frogs), Gymnophiona (caecilians); Mammalia: Carnivora (lions), Artiodactyla (pigs), Primates (monkeys); Aves: Palaeognathae (emus), Anseriformes (ducks), Sphenisciformes (penguins), Accipitriformes (hawks); Reptilia: Crocodilia (alligators), Testudines (turtles), Iguania (lizards), Ophidia (snakes), Gekkonomorpha (geckos). (*E*) The individual vertebrae plotted vertically. (*F*) Vertical plots of two constraints and plasticities identified with our approach. $C - S$ is a global constraint found when comparing all tetrapods. $C + S$ is a plasticity found within birds. (*G*) Vertical plot of a constraint, $C + T - S - Ca$ found for all birds.

skeletons, scientific literature, and manual inspection of skeletons (*Materials and Methods*). We here examine the full vertebral count across species representative of all major tetrapod clades, in contrast to previous studies of the vertebral formula focusing on subsets of tetrapods, typically mammals, and most often on presacral counts (9, 12, 13, 40, 43, 44). As represented schematically in Fig. 1 *A*–*C*, examining these vertebral formulas in the context of a phylogenetic tree yields several types of patterns, both classic and unfamiliar. We observe branches of the tree with nearly constant or strongly varying individual counts, such as the nearly fixed *C* or vigorously fluctuating *Ca* in mammals. We also observe branches with anticorrelations between adjacent vertebrae, corresponding to the pattern we have termed homeotic since they suggest a *Hox*-mediated homeotic transformation between adjacent vertebral categories. We additionally find an unexpected type, namely positive correlations between vertebral counts that are in distal (not adjacent) regions such as the cervical and sacral, as well as a balance between the counts in the anterior and posterior vertebrae in birds. Finally, we further probe *Hox*-related vertebral patterns by comparing the variation of the vertebral formula with the *Hox* cluster genetic sequence of the corresponding species, finding regions in the *Hox* cluster for which the intergenic distance correlates with the counts.

Results

Patterns in the Vertebral Formula. There is a variety of nomenclature to describe patterns such as those we treat here, with terms often implying a particular origin (45). Here, we do not determine a pattern's origin, such as whether it is due to a shared developmental program, physical restriction, stabilizing selection, or phylogenetic inertia. With the null hypothesis that vertebral counts are free to vary and independent, we label any significant pattern with low variation a "constraint" and any significant pattern with large variation a "plasticity" (46).

We generate a two-dimensional array V_{ij} with five columns representing the vertebral categories. Each row corresponds to a species, with a vertebral formula $(C_i, T_i, L_i, S_i, Ca_i)$, where each shorthand (e.g., *C*) represents the count in that category. To analyze patterns in the tetrapod phylogenetic tree, we utilize TimeTree (37) to obtain a consensus tree. TimeTree takes the median value of divergence times from multiple studies which are typically determined by sequence alignments calibrated against the fossil record. A representation of this tree generated using MEGA11 (38) is shown in Fig. 1*D*. Individual components V_{ij} are plotted for each species (Fig. 1*E*), revealing established patterns like $C \simeq 7$ in Mammalia (9) and $C \simeq 1$ in Amphibia (5). Patterns involving combinations of vertebral counts, such as the thoracolumbar pattern (9), are not immediately obvious from Fig. 1*E*, and require different methods of detection.

To identify constraints and plasticities, we iterate through branches *B* with at least $N = 20$ tips, extracting a subarray V^B_{ij} for each branch. The variance of the sum of linear combinations $\sum_j c_j V^B_{ij}$ is used to determine constraints or plasticities. Principal component analysis (PCA) is used to identify c_j which serve as constraint or plasticity candidates, which are then confirmed using a threshold criterion. Manual screening is additionally performed by computing $\sigma^2_{c,B}$ for all c_j with integer coefficients between [−3,3]. Phylogenetic tests [PIC (48)] are conducted to validate identified patterns (*Materials and Methods*). An annotated tree and tables with all patterns detected in this way are shown in *SI Appendix*, Figs. S2 and S4.

Our method identifies well-known patterns in the vertebral formula (Fig. 2*A*) as well as additional plasticities and constraints, the latter of which come in at least three flavors (Figs. 1*C* and 2*B*). The first (I) and most familiar is when a single vertebra is nearly fixed, as with the highly conserved mammalian cervical count ($C \simeq 7$, Fig. 2*C*). As Figs. 2 *C* and *D* show, a vertebra can be a constraint in one branch and plastic in another; we find plasticities in a single vertebra such as the highly variable caudal count in mammals, reptiles, and some amphibians (Urodela, i.e., salamanders) (Fig. 2*C*).

The second (II) type of constraint is an anticorrelation of vertebral counts in adjacent domains within a group of species, which also manifest as a conserved sum of counts in the two adjacent vertebral categories. Since this is a possible manifestation of the *Hox*-mediated homeotic transformations which motivated our search (11), we refer to these as homeotic constraints. The only previously confirmed example of this type is the mammalian $T + L$ (9) (Fig. 2*E*), while here we have found additional weak anticorrelations within Mammalia further along the body axis corresponding to constraints on $L + S$ (Fig. 2*F*), and a stronger constraint corresponding to $3S + Ca$, or three times the sacral count plus the caudal count (Fig. 2*G*). We note that a homeotic transition between the lumbar and sacral vertebra has been discussed in the context of intraspecies variability related to different modes of motion (12, 13). The peculiar coefficient on the constraint $3S + Ca$ (Fig. 2*G*) suggests that the gain or loss of a single sacral vertebra yields a corresponding change of three caudal vertebrae. An unequal conversion by homeotic transformation may be taking place to increase the variability in the caudal vertebrae in mammals. A similar unequal homeotic transformation has been observed between thoracic and lumbar vertebrae in quadruple (paralogous group) *Hox9* mutant mice (32).

These results suggest that the entire mammalian axis follows a homeotic pattern, where the adjacent vertebral counts are inversely related. We thus also observe a near-constant precaudal count compared with the caudal count in mammals (Fig. 2*H*), but with a small yet significant negative correlation between them. Outside Mammalia, there are other examples of type II constraints such as the $C + T$ in Testudinata and Archosauria (birds and alligators, Fig. 2*I*), which we also visually detect when including all reptiles and mammals, especially those not obeying the $C = 7$ rule (sloths and manatees, *SI Appendix*, Fig. S1). In summary, we find that only the mammalian body plan can be said to follow the patterns anticipated if variations in the vertebral numbers are primarily due to homeotic transformations (11, 33–35). Whether the mammalian homeotic patterns are *Hox*-mediated, and what is the reason for the absence of homeotic patterns in other classes will require further genetic perturbation experiments in nonmammalian species.

In stark contrast to these two constraint categories, our method uncovers an unexpected third type (III) which is a balance between distal (nonadjacent) vertebrae (Fig. 1*C*), yielding a positive correlation as seen in Figs. 2 *J* and *K* and *SI Appendix*, Fig. S4. Such distal correlations are intriguing since they are not consistent with variations in the vertebral formula arising from homeotic transformations (11, 33–35), such as would be anticipated from the *Hox*-vertebral relationship (22–31). We next discuss two prominent examples from this constraint type.

Correlations between Distal Vertebrae. At the phylogenetic scale of the tree, we uncover that $C - S$ is a constraint, with $C \simeq S$ over a wide range of *C* and *S*. In Fig. 1*F*, we plot $C - S$ and $C + S$ with the vertical location corresponding to the

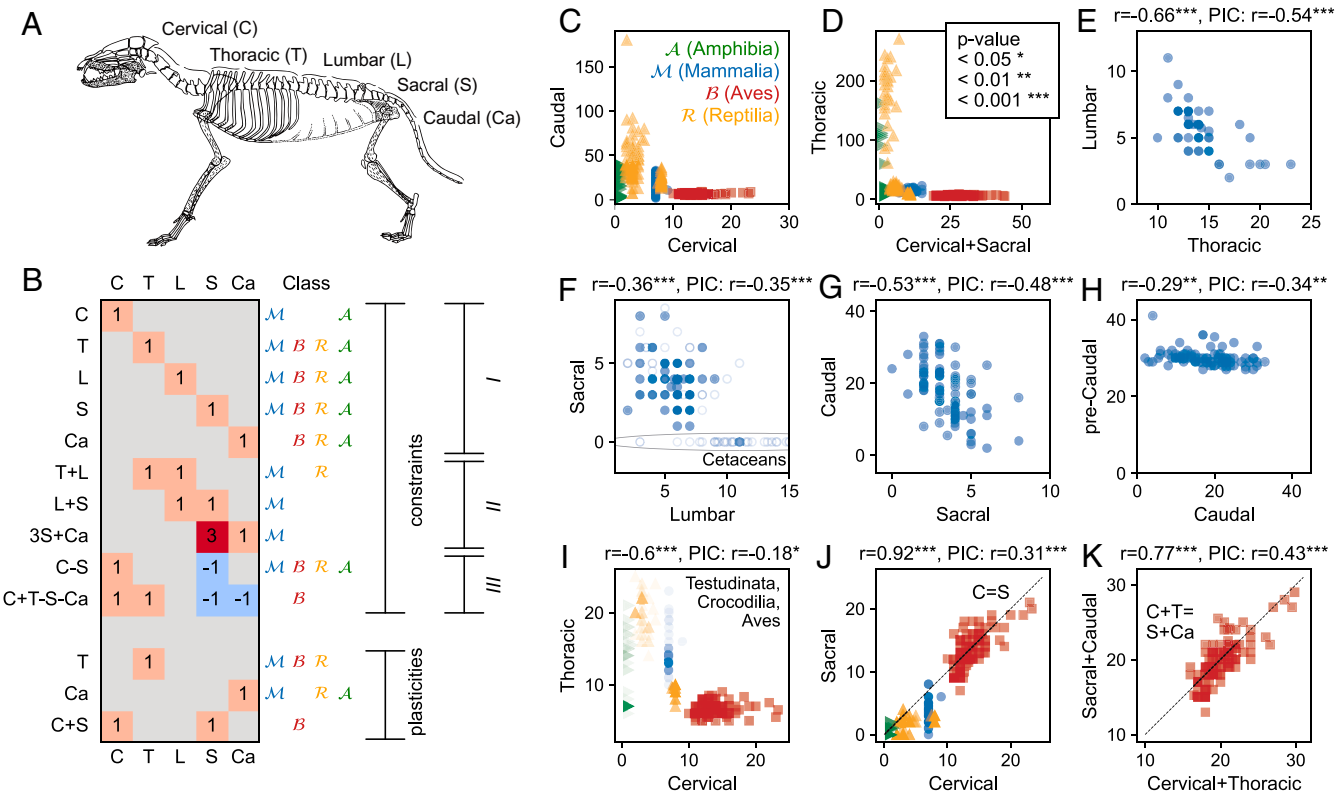

Fig. 2. Representative constraints and plasticities. (*A*) Sketch of a prototypical Mammal redrawn from Owen (47). (*B*) Table with representative constraints and plasticities, including their formulas, the tetrapod class(es) within which they are found, and the constraint types (single: I, homeotic: II, distal balance: III). In order to combine all of our analyses and make the results more readable, we round the linear coefficients to the nearest integer (*Materials and Methods*). The coefficients are also color-coded so that red is positive and blue is negative. Several constraints (plasticities) have been identified. (*C–K*) Plots of representative constraints and plasticities (re)identified using our method. In each plot, the color of the data symbols indicates the tetrapod class according to the key in (*C*). For combination constraints (II, III) we calculate not only the Pearson correlation coefficient (r) between vertebral categories but also between their phylogenetic independent contrasts (PIC) (48). In (*C*) and (*D*), we plot vertebral counts which can be constraints for some classes and plasticities for others, yielding an L-shaped plot. (*E–G*) Mammalian homeotic patterns (II) including the previously known constrained sum $T + L$ and the constrained sums $L + S$ and $3S + Ca$ revealed by our analysis. Constraints on combined sums manifest as a negative correlation when the two categories are plotted against each other. In (*F*) we also plot genus-averaged data from datasets with partial vertebral formulas (13) as well as from Cetaceans (49) as open circles with higher transparency (up to $L = 15$). Two additional homeotic patterns (II) are found for Testudinata (*H*) and Testudinata with Archosauria (birds and crocodilians) (*I*). In the background of (*I*), we additionally plot all tetrapods with a higher transparency. Excluding Amphibia and snakes yields significant correlation between C and T ($r \simeq -0.84$***). (*J* and *K*) Plots of two prominent distal balance constraints (III). (*J*) The $C - S$ constraint spans all tetrapods, while in (*K*) the $C + T - S - Ca$ constraint is unique to birds.

tree in Fig. 1*D*. While within some classes there are significant departures from $C \simeq S$, such as in Testudinata ($C \simeq 8$) and Mammalia ($C \simeq 7$), especially in tree sloths (39) (*SI Appendix*, Fig. S1), at the tetrapod phylogenetic scale we find that small S follows small C (Amphibia) and large S follows large C (Aves). Within Aves, we find that $C \simeq S$ holds even when C and S can individually vary substantially. For example, $C \simeq 11$ and $S \simeq 9$ in the Rook (*Corvus frugilegus*) (8), while $C \simeq 23$ and $S \simeq 21$ in the Whooper swan (*Cygnus cygnus*) (5, 8).

Another exceptional constraint identified by our method is a balance between the anterior ($C + T$) and posterior ($S + Ca$) vertebral counts of birds. It is well known that for birds C can vary considerably from less than 10 to more than 23 (4, 44). And yet it is manifest from Figs. 1*G* and 2*J* that C is strongly correlated with S in Aves, and when combined, the anterior $C+T$ and posterior $S + Ca$ have a significant correlation ($r \simeq 0.772$, P-value $\simeq 3.122 \times 10^{-24}$). Moreover, $C + T - (S + Ca) \simeq 0$, indicating an anterior–posterior balance (Figs. 1*G* and 2*K*).

We further analyze this bird constraint by including additional data, yielding a total of 178 unique bird species (196 total data entries) representing 29 unique orders. All data adhere to the bird constraint $C + T \simeq S + Ca$ regardless of their source (*SI Appendix*, Fig. S3*A*). We portray the balance using curated bird skeletons (Fig. 3 *A–F*) and a visual representation of the vertebral counts (Fig. 3*G*) for birds from 21 different orders, as well as examples from extinct theropods and other tetrapod classes. The birds, the bat (*Myzopoda aurita*), and *Archaeopteryx* show a conspicuous anterior–posterior balance.

With the expectation that this constraint may shed light on the evolution of birds, we also investigate the vertebral formula in the clade Theropoda, which includes extinct nonavian dinosaurs such as *Tyrannosaurus rex*. Indeed, a maximum likelihood reconstruction of the tetrapod ancestor also follows the $C-S \simeq 0$ constraint (*SI Appendix*, Fig. S6), and deviations from $C-S \simeq 0$ appear to be a more recent innovation (*SI Appendix*, Fig. S7). In Figs. 3 *I* and *J*, we plot the individual vertebral counts, the tetrapod constraint $C - S$ (dashed line), and the bird constraint $C+T-(S+Ca)$. We find that the tetrapodian $C-S$ constraint is even satisfied by the dinosaur *Tyrannosaurus rex*. The deviation from the bird constraint, on the other hand, increases as the phylogenetic distance from modern birds increases, primarily due to the caudal count (24).

To seek the possible meaning of the bird constraint, we modify our comparison of birds with other tetrapod classes by including the lumbar count (L), which is always 0 in Aves (*Materials and Methods*). Developmental studies suggest

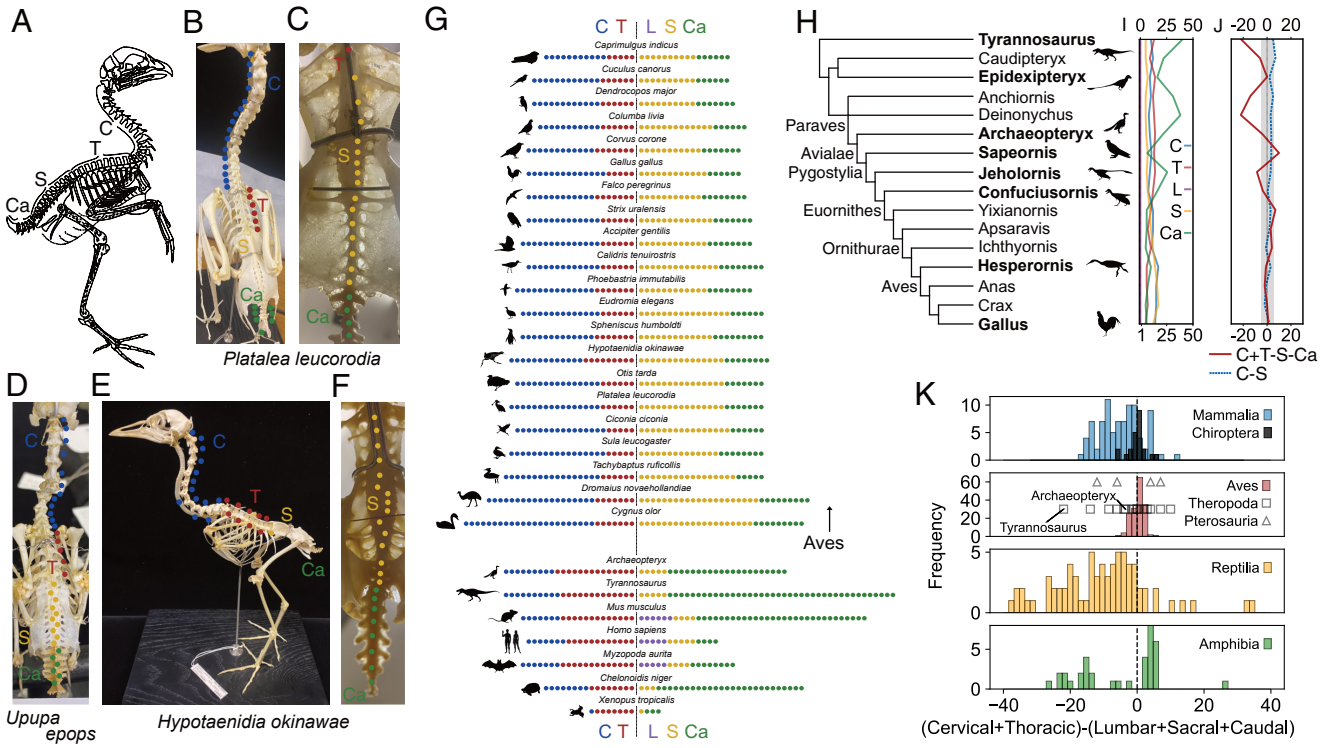

Fig. 3. Analysis of the bird constraint. (*A*) Sketch of a prototypical bird redrawn from Owen (47). (*A–F*) Images of the skeletons of *Platalea leucorodia* (Eurasian spoonbill), *Upupa epops* (Eurasian hoopoe), and *Hypotaenidia okinawae* (basionym: *Rallus okinawae*, Okinawa rail) from Abiko bird museum indicating *C*, *T*, *S*, and *Ca*. (*G*) Plot of the vertebral formula represented by a number of different colored circles. Here, we show birds from 21 different orders in Aves, as well as extinct Theropod species and extant species from the three other tetrapod classes. The vertical dashed line is an anterior–posterior boundary located after the thoracic (*T*). Birds show a balance between the two sides. (*H*) Cladogram of extinct and extant theropods extracted from a strict consensus tree (50). (*I*) Plot of the individual vertebral counts corresponding to the tree. (*J*) Plots of $C-S$ (dashed line) and $C+T-(S+Ca)$ for the species on the left. The gray band denotes three SD of $C+T-(S+Ca)$ around the average value of $\simeq$0.084 for extant birds. All taxa conform to the tetrapodian $C-S$ constraint while more phylogenetically distant theropods do not conform to the bird constraint. (*K*) Histograms of $(C+T)-(L+S+Ca)$ differentiated by tetrapod class to test a "modified" bird constraint ($L=0$ for Aves). The values of the modified constraint for theropods (□) and pterosaurs (△) are fixed at a vertical location for better visualization. All other groups of tetrapods save the putative flyers such as birds, bats (Chiroptera), and Pterosauria deviate strongly from the modified constraint.

that the synsacrum of birds may contain somites that originally had affinity with lumbar vertebrae (51), so in Fig. 3*K* we plot histograms of a modified bird constraint: $(C+T)-(L+S+Ca)$. When we test another extant group of flyers, Chiroptera (bats) (52, 53), and an extinct group of putative flyers, Pterosauria (54–59), we find that the mean deviation from the modified bird constraint, $\Delta=\sqrt{\langle[C+T-(L+S+Ca)]^2\rangle}$, is $\simeq$3.27 and $\simeq$7.83, respectively, while it is $\simeq$8.29 in mammals as a whole and $\simeq$1.82 in birds. We conjecture that the bird constraint is a result of selection for flight biomechanics, which will be important in the discussions of the flight capabilities of extinct theropods and pterosaurs.

Genetic Patterns. We next ask whether any of the vertebral variations can be directly connected with the genetic pathways controlling vertebral formation using a similar many-species comparison (65, 66). The vertebrae derive from the somites, which are themselves segmented fragments of the paraxial mesoderm (25, 67, 68). The somites are generated rhythmically with a pace set by a species-specific segmentation clock (68). *Wnt* signaling, among other signaling pathways (25, 68), is important for the initiation of the clock and the later expression of the *Hox* genes, whose functional and expression boundaries largely colocate with vertebral boundaries. The total number of somites, and thus vertebrae, is controlled by a complicated interplay of the speed of axis elongation, rate of somitogenesis, as well as the timing of termination (68, 69). While for most of the body axis somitogenesis is largely independent of *Hox* genes, which mainly act later to differentiate the vertebrae, the *Hox9-13* terminal *Hox* genes have also been shown to play a role in terminating somitogenesis (68, 69).

In Fig. 4 *A* and *B* we review a literature survey of species-specific studies on the correlation between individual *Hox* gene's anterior expression boundaries or perturbations (knock-ins and knock-outs) and homeotic vertebral transformations (22–24, 51, 60–64). Fig. 4*B* visually demonstrates the overall colinearity between the *Hox* genes and the vertebrae. As the somites destined to become vertebrae are formed from head to tail, the *Hox* genes are expressed in parallel and in all clusters from paralogous gene group (*PG*) 1 to 13 (27, 70). However, the colinearity is not perfect, and some genes, such as *PG9-11*, correlate with several vertebral transitions. Moreover, Fig. 4*B* only portrays the existence of a correlation and not its magnitude, which has been shown to be important in some cases (19). Likewise, there is also evidence that the sequential arrangement of the *Hox* genes is not a strictly necessary condition for establishing the vertebral sequence (20).

The apparent species-independence of the *Hox*-vertebral relationship, while a testament to its robustness (17), also begs the question as to how species-specific vertebral formulas can arise. While the *Hox* genes themselves are conserved, the sequence around them is highly variable in terms of transposable elements and intergenic distance (63). Recent work has shown that the number of CCCTC-binding factor (CTCF) sites between

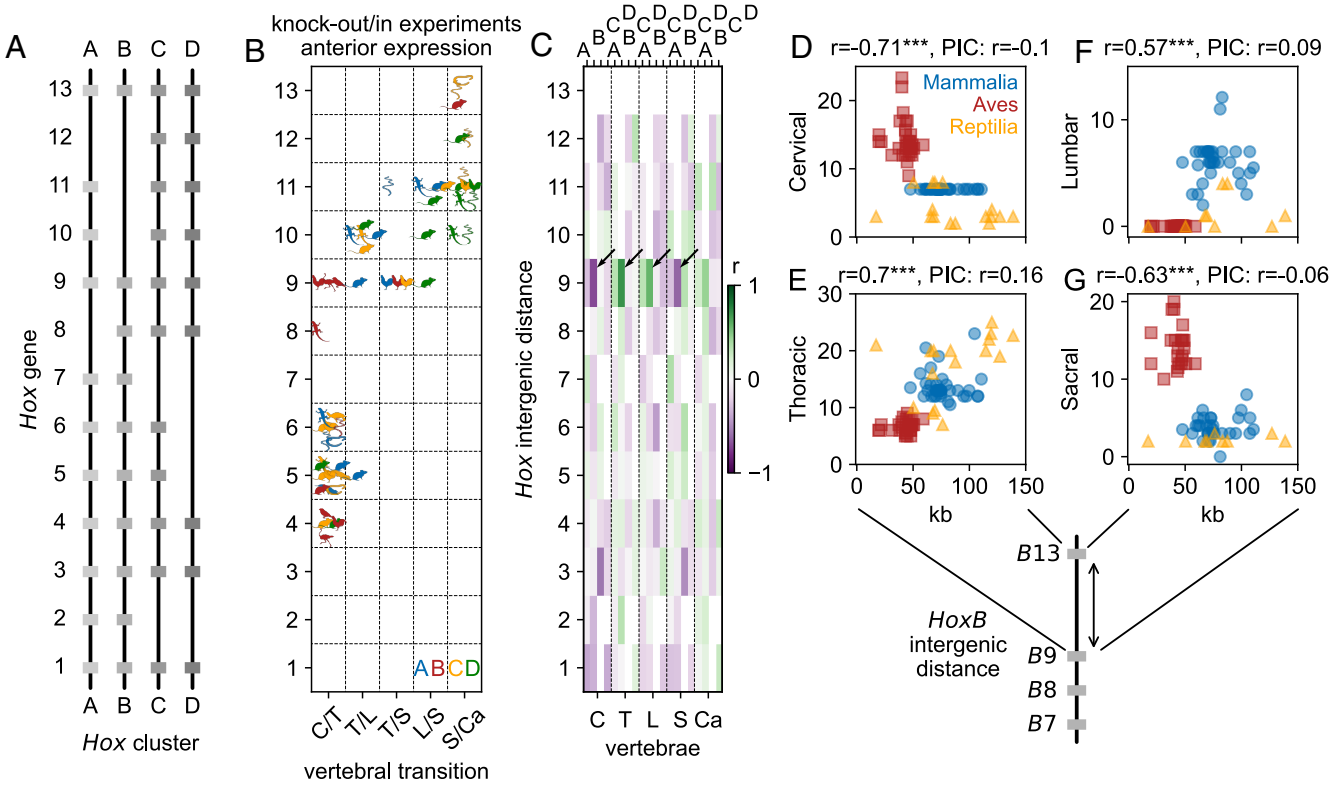

Fig. 4. Plots of the correspondence between the *Hox* genes and vertebrae. (*A*) Schematic inventory of the *Hox* genes in the four clusters present in the putative tetrapod ancestor (16). (*B*) Sampling of knock-out/in experiments and experimentally determined anterior expression boundaries of individual *Hox* genes that correlate with vertebral transitions (23, 24, 51, 60–64). The species used in the study is indicated by the data point shape, and the specific *Hox* cluster is denoted by the color. Studies often focus on specific *Hox* regions, and so, for example, the higher density of data points near the anterior *Hox* genes does not represent a stronger correlation. In general, anterior *Hox* genes correlate with anterior vertebral boundaries, while posterior genes correlate with posterior boundaries. Notable exceptions to this are *Hox9-11*. (*C*) A plot of the Pearson correlation coefficient (*r*) between the *Hox* intergenic distances and the vertebral numbers for all classes. The strongest correlations we find are between the *HoxB9-B13* intergenic distance and the *C*, *T*, *L*, and *S* counts, which we indicate by arrows. (*D–G*) *C*, *T*, *L*, and *S* vs. intergenic distance between *HoxB9* and *HoxB13*. While the correlations are large and significant, they do not survive the PIC tests. Amphibians are not included in (*D–G*) since *HoxB13* is missing in many of them (16).

HoxD genes can significantly modify their expression in mouse gastruloids (71), suggesting CTCF binding sites as important elements in the regulation of the vertebral count through the *Hox* genes. The CTCF binding sites can serve as anchors for loops in chromatin, which may modify the timing and amount of transcription (67, 71). Thus more CTCF binding sites after a *Hox* gene mediating the transition from, for example, *C* to *T* could yield *C*/*T* homeotic transitions. Indeed, earlier work has shown that adding or removing CTCF binding sites from *HoxA*, *HoxC*, or *HoxD* mouse embryos in vivo can result in homeotic transformations of cervical into thoracic, thoracic into lumbar, and even lumbar into sacral (72, 73).

Experimental information about CTCF binding sites is mostly restricted to mammals and a few model organisms, thus prohibiting us from directly testing this hypothesis with diverse tetrapod species. Instead, we use the intergenic distance as a proxy for the number of CTCF binding sites. We justify this indirect approach by comparing the size of *Hox* clusters in several model organisms with the number of experimentally determined CTCF binding sites therein and found a significant correlation ($r \simeq 0.81$, P-value $\simeq 1.53 \times 10^{-4}$, see *SI Appendix*, Fig. S9*A*). We thus proceed to use the intergenic distances as a stand-in for the number of CTCF binding sites. However, it is important to note that variations in intergenic distances likely also correlate with other factors that may be involved in the regulation of the *Hox* gene expression, so that any significant relationships between *Hox* intergenic distances and vertebral counts could also be due to other regulating sequences. Moreover, there is evidence that species-specific patterning outputs can be modulated by single base pair differences in *Hox*-dependent enhancers in trans (21), a feature that will not be captured by this approach.

Using genome assemblies available from NCBI, we generate a database of species for which we have at least partial vertebral formulas and for which the *Hox* clusters are each located on a single chromosome or scaffold (42). We determine *Hox* gene locations from existing annotations or by a *BLAST* search (74), yielding a combined database of *Hox* intergenic distances and (at least partial) vertebral formulas with 106 species across the four tetrapod classes (*Materials and Methods*). The intergenic distances are calculated from the centers of the genes. We exclude snakes from this analysis since their unusual vertebral count is apparently controlled by other factors (21, 69).

We compare the intergenic distances with the individual vertebral counts and produce the Pearson correlation matrix shown in Fig. 4*C* (we show the correlation within classes in *SI Appendix*, Fig. S8). The strongest and most significant correlations are between the *HoxB9-13* intergenic distance and the vertebral counts (Fig. 4 *D–G* and *SI Appendix*, Fig. S9). The correlation signs of adjacent vertebrae are opposite with respect to this intergenic distance. That is, *C* and *S* have a positive correlation, while *T* and *L* have a negative correlation, in accord with the anticorrelation between *C* vs. *T* and *L* vs. *S* suggestive of a homeotic transformation shown in Fig. 2 *F* and *I*. Although mammals have an essentially fixed $C \simeq 7$, they still fall along

a trend line that goes from birds to reptiles in Fig. 4*D*, just as mammals fell along the trend with the C/S distal vertebral correlation and the C/T homeotic pattern (Fig. 2 *D* and *I* and *SI Appendix*, Fig. S1). Within the separate tetrapod classes, the data do not tend to follow this trend; we have a low correlation ($r \simeq 0.17$, $P > 0.05$) after the PIC test, indicating that the pattern could be explained by the phylogenetic structure and random variation. We note that the correlation between *C*, *T*, and *HoxB9* is consistent with a previous finding that the forelimb, typically located at the C/T transition, coincides with the position of switching in the expression from *HoxB4* to *HoxB9* in several bird species (64). Likewise, Burke et al. (51) found that *HoxB9* is expressed near the end of the thoracic series in both chicken and mouse. Our finding supplies a possible explanation for how the same *Hox* gene can generate different vertebral counts in different species via the regulation of nearby sequences.

In general, we did not find significant correlations that directly support the idea that the vertebral formula is controlled by the *Hox*-cluster genetic structure through intergenic sequence elements. Within the correlations between the intergenic distance and the vertebral counts, almost none survive the phylogenetic test (PIC), with a notable exception being the correlation between the thoracic and *HoxD1-D3* intergenic distance in amphibians (*SI Appendix*, Fig. S9*E*). However, just as there are branch-specific vertebral patterns (Fig. 2), we found scattered correlations specific to certain tetrapod classes (*SI Appendix*, Figs. S8 and S9). On the other hand, the significant correlations between anterior vertebrae and posterior intergenic distances (*HoxB9-B13*) seem to align with the picture of posterior prevalence (75). Namely a posterior *Hox* gene such as *HoxB9* can control anterior vertebral boundaries which generally colocalize with anterior *Hox* gene expression boundaries. Our findings suggest the need to investigate additional factors beyond the *Hox* cluster genetic structure itself to explain the distinct timing of *Hox* gene expression that contributes to vertebral variations across species.

Discussion

By generating a database (42) of complete vertebral formulas for tetrapods, we identified three main constraint patterns; constant single count (I), homeotic (II), and distal correlations (III). We found that the vertebral formula is largely constructed according to homeotic patterns within mammals, but generally not in other tetrapod classes. Thus while the expression boundaries of the highly conserved and homologous clusters of *Hox* genes are largely the same throughout tetrapods, either their utilization or the influence of other factors makes their relationship with the vertebral formula more complicated. More concretely, we identified correlations between distal vertebrae for tetrapods and birds that are not compatible with the homeotic transformations anticipated by the *Hox*-vertebral relationship.

Important next steps will be to further investigate the origin or nature of the vertebral patterns. In this context, ancestral state reconstruction of vertebral formulas could serve as a useful tool to trace patterns found in extant groups (*SI Appendix*, Fig. S6). Direct investigation through developmental perturbations can ascertain whether constraints are developmental or evolutionary (45). For example, manipulating the expression of *HoxB9* in chickens, which may affect the cervical count (64), could permit a test of the C/S distal constraint. Alternatively, patterning in organoid models could provide a more accessible system for manipulating *Hox* gene expression to test its relationship with vertebral patterns (71, 76).

Indirect approaches, such as correlating co-occurring features, may offer additional insights. Examples include the relationship between the cervical count and mammalian cancer (77) as well as the mammalian diaphragm (78), or the consistent vertebral counts in *Anura* despite significant phylogenetic diversity (Fig. 1). The possible connection between avian flight and the unique bird constraint exemplifies this approach. It will be also important to consider whether patterns arise at the point of segmentation or much later, as may be the case for the avian synsacrum (79).

Furthermore, exploring intraspecies vertebral count variability could yield valuable insights. Investigating intraspecies variations of lumbosacral counts in mammals has yielded potential relationships with mobility and environment (12, 13). Likewise, within the mammals that disobey the rule of seven or which have entirely lost their sacral vertebrae, the intraspecies variations suggest a homeotic origin (39, 40, 49) (see also *SI Appendix*, Fig. S1). We anticipate that probing intraspecies variations in birds relative to their unique constraint may also shed light on their origins. Finally, extending our analysis to encompass the entire vertebral tree including fish, where extensive data have been collected for the vertebral counts and their intraspecies variation (80), may uncover further vertebral patterning and provide clues for understanding the evolutionary and developmental factors which produced them.

Materials and Methods

Vertebral Categorization. In order to compare the vertebral formula between different classes of tetrapods, it was necessary to identify categories of homologous vertebrae. Different nomenclatures and conventions exist for different classes and even orders, and so we determined a consensus definition for each vertebral category. The cervical (*C*) bones were defined as those vertebrae between the skull and the first rib-bearing vertebrae, including the axis and atlas. In mammals and amphibians, this definition leaves no ambiguity as to the identity of the cervical vertebrae. In some reptiles (e.g., Crocodilia) and birds, however, other authors have identified the cervical vertebrae as all vertebrae between the skull and the forelimb. We next define the thoracic vertebrae (*T*) as all rib-bearing vertebrae in the dorsal region, anterior to the pelvis or sacrum. This includes free or "false" ribs not attached to the sternum and excludes vertebrae with ribs that are fused with the synsacrum in birds (8). The lumbar (*L*) vertebrae are all remaining dorsal vertebrae posterior to the rib-bearing thoracic and anterior to the pelvis or sacrum. With this definition, birds and amphibians and many reptiles do not have lumbar vertebrae, but most mammals do. We define the sacral vertebrae (*S*) as those vertebrae which are (typically) fused together to form the pelvic region to which the hindlimbs are attached. Some authors have made further categorizations of vertebrae within the fused synsacrum of birds (51), but we go no further than to identify them as sacral. In this, we essentially follow Eyton's identification (8). Identifying the individual sacral vertebrae in birds is especially challenging due to their reduced size and complex ossification (Fig. 3 and *SI Appendix*, Fig. S13). For snakes, which do not have a hindlimb, the sacrum is identified with the cloacal when possible (69). Finally, the caudal vertebrae (*Ca*) are identified as all free vertebrae posterior to the sacrum. We counted the bird's fused pygostyle as a single vertebra, and following Ročková et al. (81), we assigned a caudal count of 3 for Anura unless the data source notes otherwise.

Extant Data Collection. The vertebral counts for extant species were obtained from a variety of sources, including the literature where the counts are already given or could be surmised from figures, CT scans, and images of skeletons from online repositories (MorphoSource, DigiMorph, and Sketchfab), and from manual inspection of bird skeletons at the Abiko City Museum of Birds in Abiko, Japan. A full list of sources can be found in the electronic repository (42). The precise age and sex of the majority of the specimens are unknown, but based on overall size, we presume that most of the specimens are adults. Some of the counts were taken from studies of embryos, for which the authors identified

which vertebral type the somites eventually became. We also provide several example images of skeletons, X-rays, and CT data in *SI Appendix*, Figs. S12 and S13.

Algorithm for Finding Vertebral Patterns. We search for vertebral patterns within the vertebral data V_{ij} in the context of an existing phylogenetic tree (37). For each branch B of the tree with at least $N = 20$ tips (species), we identify which linear combinations c_j of the vertebral data columns have large or small variation $\sigma^2_{c,B}$ compared to both the total variation of this branch σ^2_B and the variation of the same linear combination for all species (tips) outside this branch $\sigma^2_{c,\notin B}$. Because the number of linear combinations c_j to test is in principle infinite, we use two tools to focus our search. First, we use principle component analysis (PCA) (82) to reorganize and rotate the vertebral data and sort it by variance. PCA will naturally pick out features (c_j) with large and small variance, making it ideal for this task. Second, we test all linear combinations with integer coefficients ranging from -3 to 3. This yields a total number of $7^5 = 16{,}807$ possible c_j, which reduces to 8,403 unique c_j. We compile a list of c_j from both techniques and for each branch B that satisfy the criteria for either a constraint, $\sigma^2_{c,B} < 0.1\sigma^2_B, \sigma^2_{c,B} < 0.1\sigma^2_{c,\notin B}$, or a plasticity $\sigma^2_{c,B} > 0.7\sigma^2_B, \sigma^2_{c,B} > 10\sigma^2_{c,\notin B}$. We provide this master list in an electronic repository for the criteria we use (42). If a constraint or plasticity satisfies both respective threshold criteria (inside and outside the branch), then we consider it a "strong" constraint (or plasticity). We label these few cases in *SI Appendix*, Fig. S2*A*. For certain c_j where possible (a constraint with more than one nonzero coefficient), we also tested whether the Pearson correlation coefficient r determined after renormalizing the counts using the method of PIC (Section) (48), is greater than some threshold value (for most vertebral data, we required $r_{\text{PIC}} > 0.2$, which we relaxed for the combined vertebral and intergenic data). We combine the results from the integer c_j and the c_j determined by PCA, which are typically not integers, by rounding the PCA c_j coefficients to the nearest integer. For the combined list, we then proceeded to identify all unique constraints and plasticities and determine the highest (earliest) node on the tree at which they hold. Some constraints appeared independently in the tree, such as the constraint on $T + L$ in both Mammalia and Testudinata. This final list of constraints and plasticities are labeled in Fig. 1*A*, tabulated in *SI Appendix*, Fig. S4, and plotted in part in Fig. 2. In *SI Appendix*, Fig. S5*A*, we show the total number of constraints and plasticities identified as a function of these different variance thresholds, which show kinks that we used to determine the thresholds. We provide the Python code used to analyze the vertebral data in a repository (42).

PIC. It is possible that correlations between the phenotypic characters of different species may be manifestations of the structure of the phylogenetic tree to which the species belong (48). The placement of the species in the tree is itself determined by morphological and genetic differences. For the correlation of the characters in question to be significant, their relationship must differ from what is expected purely by their respective phylogenetic distance. Put another way, we could imagine the two characters under scrutiny take some original value at the root of the phylogenetic tree, and we let the value of these characters undergo a random walk along the branches. If their final values at the tips are correlated in the same way as we observe in the same measure as the actual data, our correlation is not significant.

We use the method of PIC originally proposed in ref. 48 to deal with this potential problem. In this method, the character values at the tree tips are subtracted from their nearest neighbor and normalized by their respective phylogenetic distance. The contrasted tips are then merged using a weighted average and the process is repeated until all tips are accounted for. A tree with N tips will thus yield $N - 1$ contrasts. The subtraction or contrast is designed to remove the shared influence of the tree before their common branching point, and the normalization makes it easier to compare contrasts with different (original) branch lengths. We used a relatively low threshold for the Pearson correlation coefficient of $r = 0.3$ after the PIC for screening the vertebral patterns belonging to types II and III (it is not possible to test type I constraints with PIC). We wrote a custom Python implementation for PIC to keep track of the species contributing to each contrast and to ensure transparency of method, which we include in a repository (42). The ability to monitor the original species was useful when testing correlations on subsets of the entire phylogenetic tree.

We confirmed our calculations with the widely used *ape* (83) and *phytools* (84) R libraries and include the scripts in an electronic repository (42). It has been noted that the PIC can also serve as a proxy for rates of evolutionary change (85). Using the PIC for the different vertebrae at the tree's nodes, in *SI Appendix*, Fig. S7, we show how certain branch points exhibit higher rates of change in values occurring after those points.

Analysis of Vertebral Data from Extinct Theropods and Paraves. The vertebral counts for extinct Theropod and Paraves species were obtained from a number of literature sources which had analyzed the fossil remains (86–98). We combined and averaged the counts for all taxa at the genus level. We used the phylogenetic tree or cladogram from ref. 50 and manually trimmed it to the taxa for which we have the full vertebral count in the tree shown in Fig. 3*C*. We provide a table of all the vertebral data in an online repository as well as the corresponding source (42).

BLASTing Unannotated Genomes. Many of the genome assemblies in the NCBI database have not yet been fully annotated. We determined the approximate location of the *Hox* genes by performing a BLAST search (blastn) (74) on each genome using the gene sequence from the corresponding *Hox* gene from a model organism in that tetrapod class as the search sequence. Thus for all unannotated bird genomes, for example, we used the chicken *Hox* genes to search for the genomic coordinates of their *Hox* genes. We then manually curated the search results and only retained species for which we were able to find all four *Hox* clusters with the entire cluster located on a single chromosome or scaffold. We include the list of annotated and unannotated genomes used in our study in an electronic repository (42), as well as several measures of the genome quality (coverage), contiguity (scaffold N50), and RefSeq annotation completeness when available (BUSCO score). In addition to requiring that the BLAST-identified *Hox* genes be on a single scaffold or chromosome, we also required that their significance or "E-value" be $< 1 \times 10^{-4}$. The average E-value for all results was $\simeq 8.09 \times 10^{-7}$. We include statistics about the genome quality and BLAST results in *SI Appendix*, Fig. S11. In addition, we also plot in *SI Appendix*, Fig. S10 comparisons between the vertebral counts and *Hox* intergenic distances for a restricted set of higher quality genomes with the same result as in Fig. 4. For the tuatara (*Sphenodon punctatus*) the location of the top BLAST hit for HoxB13 was ~231 kb away from the HoxB9 location, which was nearly twice the distance as other Squamata. Thus for this species and gene alone we determined the location of HoxB13 using an alternative alignment tool, progressiveMauve (99).

Determining the CTCF Peak Numbers. We tested the relationship between the number of CTCF binding sites and intergenic distances in the *Hox* clusters for several model organisms in different tetrapod classes. We used experimental ChIP-seq data from human H1 cells (100), mouse embryonic stem cells (101), chicken embryonic cells (102), and frog embryonic cells (103). The human, mouse, and frog data were already analyzed and mapped to recent genome assemblies (GRCh38, GRCm38, and xenTro10). We mapped the raw reads for the chicken data to the bGalGal1 genome assembly using bowtie2 (104) with standard parameters and used samtools (105) with standard parameters to remove duplicates and sort the alignments. After smoothing the data, we used the *find_peaks* function from the Python *scipy* signal library (106) to detect peaks with a prominence larger than one SD of the signal. Using the genome annotations from NCBI we determined the locations of the *Hox* genes in each cluster and counted the number of peaks in between each gene as well as determined the intergenic distance. We plot the sum of these counts and intergenic distances for each cluster and species in *SI Appendix*, Fig. S9*A*. The intergenic distances and number of CTCF binding sites have a Pearson correlation coefficient of $r \simeq 0.81$ and P-value $\simeq 1.53 \times 10^{-4}$.

Silhouette and Images. The silhouette images used in the figures were all obtained from www.phylopic.org except for *Hypotaenidia okinawae* which we created ourselves from an image obtained from the Ryuku Shimpo. The *Jeholornis* (*Shenzhouraptor sinensis*) and *Hesperornis* (*regalis*) were created

by Scott Hartman while the *Epidexipteryx* (for which we used *Scansoriopteryx heilmannii*) and *Sapeornis* (*chaoyangensis*) were created by Matthew Martyniuk under a Creative Commons license (https://creativecommons.org/licenses/by/3.0/). All other images are in the public domain unless otherwise noted.

Data, Materials, and Software Availability. The tables of vertebral formulas, *Hox* intergenic distances, code for the implementation of PIC and vertebral analysis, and other relevant information have been deposited in an online repository (42). Data and code have been deposited in GitHub (https://github.com/dbkk/vertebral_pattern).

ACKNOWLEDGMENTS. We thank Y. Odaya and the Abiko City Museum of Birds for their hospitality and help in viewing their bird skeletons. We also thank Shigeru Kuratani, Olivier Pourquié, Asami Oji, and Rawin Poonperm for critically reading the manuscript. We also thank W.M. Cerbus and E.D. Cerbus for their help in redrawing the sketches from Cuvier and Owen. This work was supported by RIKEN Biosystems Dynamics Research intramural grants, MEXT KAKENHI Grant No. JP18H05530 (to I.H.), JSPS KAKENHI Grant Nos. JP20K20582 (to I.H.), JP19H05795, JP19H05275, JP21H01007, and JP23H00095 (to K.K.), and JST CREST Grant Nos. JPMJCR20S5 (to I.H.) and JPMJCR1912 (to K.K.). This work was also supported by the RIKEN Pioneering Project "Genome Building from Topologically Associating Domains.

1. P. Belon, P. Glardon, *L'histoire de la Nature des Oyseaux* (Librairie Droz, 1997), vol. 306.
2. F. J. Cole, N. B. Eales, The history of comparative anatomy: Part I.–A statistical analysis of the literature. *Sci. Prog.* **1916–1919** (11), 578–596 (1917).
3. É. G. Saint-Hilaire, *Cours de L'histoire Naturelle des Mamifères* (Pichon, 1829).
4. G baron Cuvier,, *Leçons D'anatomie Comparée* (Baudouin, 1840), vol. 3.
5. R. Owen, *Descriptive Catalogue of the Osteological Series Contained in the Museum of the Royal College of Surgeons of England* (Taylor & Francis, London, UK, 1853), vol. 1.
6. R. Owen, *Anatomy of Vertebrates, Vol. 1: Fishes and Reptiles* (Longmans, Green, and Company, London, UK, 1866).
7. R. Owen, *On the Anatomy of Vertebrates: Birds and Mammals* (Longmans, Green and Company, 1866), vol. 2.
8. T. C. Eyton, *Osteologia Avium: Or, A Sketch of the Osteology of Birds* (R. Hobson, 1875), vol. 1.
9. Y. Narita, S. Kuratani, Evolution of the vertebral formulae in mammals: A perspective on developmental constraints. *J. Exp. Zool. Part B: Mol. Dev. Evol.* **304**, 91–106 (2005).
10. M. R. Sánchez-Villagra, Y. Narita, S. Kuratani, Thoracolumbar vertebral number: The first skeletal synapomorphy for Afrotherian mammals. *Syst. Biodiversity* **5**, 1–7 (2007).
11. J. Müller *et al*., Homeotic effects, somitogenesis and the evolution of vertebral numbers in recent and fossil amniotes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2118–2123 (2010).
12. F. Galis *et al*., Fast running restricts evolutionary change of the vertebral column in mammals. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11401–11406 (2014).
13. S. A. Williams *et al*., Increased variation in numbers of presacral vertebrae in suspensory mammals. *Nat. Ecol. Evol.* **3**, 949–956 (2019).
14. I. Varela-Lasheras *et al*., Breaking evolutionary and pleiotropic constraints in mammals: On sloths, manatees and homeotic mutations. *EvoDevo* **2**, 1–27 (2011).
15. S. He *et al*., An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*. *Science* **361**, 1377–1380 (2018).
16. D. Liang, R. Wu, J. Geng, C. Wang, P. Zhang, A general scenario of Hoxgene inventory variation among major *Sarcopterygian* lineages. *BMC Evol. Biol.* **11**, 1–13 (2011).
17. D. Duboule, The (unusual) heuristic value of Hox gene clusters; a matter of time? *Dev. Biol.* **484**, 75–87 (2022).
18. M. Kessel, P. Gruss, Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**, 89–104 (1991).
19. J. Zakany, M. Gerard, B. Favier, S. Potter, D. Duboule, Functional equivalence and rescue among group 11hoxgene products in vertebral patterning. *Dev. Biol.* **176**, 325–328 (1996).
20. P. Tschopp, B. Tarchini, F. Spitz, J. Zakany, D. Duboule, Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genet.* **5**, e1000398 (2009).
21. I. Guerreiro *et al*., Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10682–10686 (2013).
22. D. M. Wellik, M. R. Capecchi, Hox10 and hox11 genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363–367 (2003).
23. C. Böhmer, Correlation between Hox code and vertebral morphology in the mouse: Towards a universal model for synapsida. *Zool. Lett.* **3**, 1–11 (2017).
24. D. J. Rashid *et al*., From dinosaurs to birds: A tail of evolution. *EvoDevo* **5**, 1–20 (2014).
25. M. Mallo, T. Vinagre, M. Carapuço, The road to the vertebral formula. *Int. J. Dev. Biol.* **53**, 1469–1481 (2009).
26. M. Mallo, D. M. Wellik, J. Deschamps, Hox genes and regional patterning of the vertebrate body plan. *Dev. Biol.* **344**, 7–15 (2010).
27. M. Carapuço, A. Nóvoa, N. Bobola, M. Mallo, Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev.* **19**, 2116–2121 (2005).
28. G. Horan *et al*., Compound mutants for the paralogous Hoxa-4, Hoxb-4, and Hoxd-4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. *Genes Dev.* **9**, 1667–1677 (1995).
29. G. Horan, K. Wu, D. J. Wolgemuth, R. R. Behringer, Homeotic transformation of cervical vertebrae in Hoxa-4 mutant mice. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12644–12648 (1994).
30. H. Saegusa, N. Takahashi, S. Noguchi, H. Suemori, Targeted disruption in the mousehoxc-4locus results in axial skeleton homeosis and malformation of the xiphoid process. *Dev. Biol.* **174**, 55–64 (1996).
31. D. Kostic, M. R. Capecchi, Targeted disruptions of the murine Hoxa-4 and Hoxa-6 genes result in homeotic transformations of components of the vertebral column. *Mech. Dev.* **46**, 231–247 (1994).
32. D. C. McIntyre *et al*., Hox patterning of the vertebrate rib cage. *Development* **134**, 2981–2989 (2007).
33. T. Szczygielski, Homeotic shift at the dawn of the turtle evolution. *R. Soc. Open Sci.* **4**, 160933 (2017).
34. F. Mao, C. Zhang, C. Liu, J. Meng, Fossoriality and evolutionary development in two cretaceous mammaliamorphs. *Nature* **592**, 577–582 (2021).
35. B. Figueirido *et al*., Body-axis organization in tetrapods: A model-system to disentangle the developmental origins of convergent evolution in deep time. *Biol. Lett.* **18**, 20220047 (2022).
36. G. Cuvier, *Recherches sur les Ossemens Fossiles* (Dufour-D'Ocagne, 1825), vol. 5.
37. S. Kumar *et al*., Timetree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
38. K. Tamura, G. Stecher, S. Kumar, Mega11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
39. E. A. Buchholtz, C. C. Stepien, Anatomical transformation in mammals: Developmental origin of aberrant cervical anatomy in tree sloths. *Evol. Dev.* **11**, 69–79 (2009).
40. L. Hautier, V. Weisbecker, M. R. Sánchez-Villagra, A. Goswami, R. J. Asher, Skeletal development in sloths and the evolution of mammalian vertebral patterning. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18903–18908 (2010).
41. C. Böhmer, E. Amson, P. Arnold, A. H. van Heteren, J. A. Nyakatura, Homeotic transformations reflect departure from the mammalian 'rule of seven' cervical vertebrae in sloths: inferences on the Hox code and morphological modularity of the mammalian neck. *BMC Evol. Biol.* **18**, 1–11 (2018).
42. R. T. Cerbus, I. Hiratani, K. Kawaguchi, Data from "Vertebral pattern dataset." GitHub. https://github.com/dbkk/vertebral_pattern. Deposited 16 August 2024.
43. Y. Li *et al*., Divergent vertebral formulae shape the evolution of axial complexity in mammals. *Nat. Ecol. Evol.* **7**, 367–381 (2023).
44. C. Böhmer, O. Plateau, R. Cornette, A. Abourachid, Correlated evolution of neck length and leg length in birds. *R. Soc. Open Sci.* **6**, 181588 (2019).
45. J. M. Smith *et al*., Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
46. J. Antonovics, P. H. van Tienderen, Ontoecogenophyloconstraints? The chaos of constraint terminology. *Trends Ecol. Evol.* **6**, 166–168 (1991).
47. R. Owen, *On the Nature of Limbs: A Discourse Delivered on Friday, February 9, at an Evening Meeting of the Royal Institution of Great Britain* (J. van Voorst, 1849).
48. J. Felsenstein, Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
49. E. A. Buchholtz, J. K. Gee, Finding sacral: Developmental evolution of the axial skeleton of *Odontocetes* (Cetacea). *Evol. Dev.* **19**, 190–204 (2017).
50. S. L. Brusatte, G. T. Lloyd, S. C. Wang, M. A. Norell, Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur–bird transition. *Curr. Biol.* **24**, 2386–2392 (2014).
51. A. C. Burke, C. E. Nelson, B. A. Morgan, C. Tabin, Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).
52. D. W. Walton, G. M. Walton, Post-cranial osteology of bats. *Fond. Sci. Ser.* **1**, 7 (1970).
53. O. O. Igado, E. R. Ade-Julius, Gross description and osteometrics of the axial skeleton (ribs and vertebrae) of *Eidolon helvum* (African fruit bat). *Niger. J. Physiol. Sci.* **33**, 189–194 (2018).
54. S. W. Williston, *On the osteology of Nyctosaurus (Nyctodactylus), with notes on American pterosaurs. Geological Series* (Field Columbian Museum, 1903), vol. 2.
55. S. C. Bennett, The osteology and functional morphology of the late cretaceous pterosaur pteranodon. Part I. General description of osteology. *Palaeontogr. Abteilung A* **260**, 1–112 (2001).
56. X. Wang *et al*., New evidence from china for the nature of the pterosaur evolutionary transition. *Sci. Rep.* **7**, 42763 (2017).
57. D. M. Unwin, J. Lü, N. N. Bakhurina, On the systematic and stratigraphic significance of pterosaurs from the lower cretaceous Yixian formation (Jehol group) of Liaoning, China. *Fossil Rec.* **3**, 181–206 (2000).
58. D. Hone, D. M. Henderson, F. Therrien, M. B. Habib, A specimen of Rhamphorhynchus with soft tissue preservation, stomach contents and a putative coprolite. *PeerJ* **3**, e1191 (2015).
59. A. Ősi, E. Prondvai, Forgotten pterosaurs in Hungarian collections: First description of Rhamphorhynchus and Pterodactylus specimens. *Neues Jahrb. Geol. Paläontol.-Abh.* **252**, 167–180 (2009).
60. C. Böhmer, O. W. Rauhut, G. Wörheide, Correlation between Hox code and vertebral morphology in archosaurs. *Proc. R. Soc. B: Biol. Sci.* **282**, 20150077 (2015).
61. J. M. Woltering *et al*., Axial patterning in snakes and caecilians: Evidence for an alternative interpretation of the Hox code. *Dev. Biol.* **332**, 82–89 (2009).
62. Y. K. Ohya, S. Kuraku, S. Kuratani, Hox code in embryos of Chinese soft-shelled turtle *Pelodiscus sinensis* correlates with the evolutionary innovation in the turtle. *J. Exp. Zool. Part B: Mol. Dev. Evol.* **304**, 107–118 (2005).
63. N. Di-Poi *et al*., Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99–103 (2010).
64. C. Moreau *et al*., Timed collinear activation of Hox genes during gastrulation controls the avian forelimb position. *Curr. Biol.* **29**, 35–50 (2019).
65. P. Alberch, From genes to phenotype: Dynamical systems and evolvability. *Genetica* **84**, 5–11 (1991).
66. M. Pigliucci, Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos. Trans. R. Soc. B: Biol. Sci.* **365**, 557–566 (2010).
67. J. Deschamps, D. Duboule, Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev.* **31**, 1406–1416 (2017).

68. Y. Miao, O. Pourquié, Cellular and molecular control of vertebrate somitogenesis. *Nat. Rev. Mol. Cell Biol.* **25**, 517–533 (2024).
69. C. Gomez *et al*., Control of segment number in vertebrate embryos. *Nature* **454**, 335–339 (2008).
70. T. Iimura, O. Pourquié, Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**, 568–571 (2006).
71. H. Rekaik *et al*., Sequential and directional insulation by conserved CTCF sites underlies the Hox timer in stembryos. *Nat. Genet.* **55**, 1164–1175 (2023).
72. V. Narendra, M. Bulajić, J. Dekker, E. O. Mazzoni, D. Reinberg, CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).
73. A. R. Amândio *et al*., Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse Hoxd cluster. *Genes Dev.* **35**, 1490–1509 (2021).
74. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
75. D. Duboule, G. Morata, Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358–364 (1994).
76. Y. Yamanaka *et al*., Reconstituting human somitogenesis in vitro. *Nature* **614**, 509–520 (2023).
77. F. Galis, Why do almost all mammals have seven cervical vertebrae? Developmental constraints, Hox genes, and cancer. *J. Exp. Zool.* **285**, 19–26 (1999).
78. E. A. Buchholtz *et al*., Fixed cervical count and the origin of the mammalian diaphragm. *Evol. Dev.* **14**, 399–411 (2012).
79. H. N. N. Bui, H. C. Larsson, Development and evolution of regionalization within the avian axial column. *Zool. J. Linnean Soc.* **191**, 302–321 (2021).
80. C. Lindsey, Pleomerism, the widespread tendency among related fish species for vertebral number to be correlated with maximum body length. *J. Fisher. Board Canada* **32**, 2453–2469 (1975).
81. H. Ročková, Z. Roček, Development of the pelvis and posterior part of the vertebral column in the Anura. *J. Anat.* **206**, 17–35 (2005).
82. M. Ringnér, What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
83. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
84. L. J. Revell, phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).
85. T. Garland Jr, Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* **140**, 509–519 (1992).
86. J. H. Ostrom, Archaeopteryx and the origin of birds. *Biol. J. Linnean Soc.* **8**, 91–182 (1976).
87. L. M. Chiappe, S. A. Ji, Q. Ji, M. A. Norell, Anatomy and systematics of the Confuciusornithidae (Theropoda, Aves) from the late Mesozoic of Northeastern China. *Bull. Am. Mus. Nat. Hist.* **242**, 1–89 (1999).
88. Z. Zhou, F. Zhang, Jeholornis compared to Archaeopteryx, with a new understanding of the earliest avian evolution. *Naturwissenschaften* **90**, 220–225 (2003).
89. R. Pei, Q. Li, Q. Meng, M. A. Norell, K. Q. Gao, New specimens of *Anchiornis huxleyi* (Theropoda: Paraves) from the late Jurassic of Northeastern China. *Bull. Am. Mus. Nat. Hist.* **2017**, 1–67 (2017).
90. J. H. Ostrom, J. A. Gauthier, *Osteology of Deinonychus antirrhopus, an Unusual Theropod from the Lower Cretaceous of Montana* (Yale University Press, 2019).
91. Z. Zhong-He, W. Xiao-Lin, Z. Fu-Cheng, X. Xing, Important features of Caudipteryx–Evidence from two nearly complete new specimens. *Verteb. PalAsiatica* **38**, 241 (2000).
92. J. Qiang, P. J. Currie, M. A. Norell, J. Shu-An, Two feathered dinosaurs from Northeastern China. *Nature* **393**, 753–761 (1998).
93. C. A. Brochu, Osteology of Tyrannosaurus rex: Insights from a nearly complete skeleton and high-resolution computed tomographic analysis of the skull. *J. Verteb. Paleontol.* **22**, 1–138 (2003).
94. F. Zhang, Z. Zhou, X. Xu, X. Wang, C. Sullivan, A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
95. J. A. Clarke, Z. Zhou, F. Zhang, Insight into the evolution of avian flight from a new clade of early cretaceous ornithurines from China and the morphology of *Yixianornis grabaui*. *J. Anat.* **208**, 287–308 (2006).
96. J. Benito *et al*., Forty new specimens of ichthyornis provide unprecedented insight into the postcranial morphology of crownward stem group birds. *PeerJ* **10**, e13919 (2022).
97. A. Bell, L. M. Chiappe, Anatomy of Parahesperornis: Evolutionary mosaicism in the cretaceous Hesperornithiformes (Aves). *Life* **10**, 62 (2020).
98. M. A. Norell, J. A. Clarke, Fossil that fills a critical gap in avian evolution. *Nature* **409**, 181–184 (2001).
99. A. E. Darling, B. Mau, N. T. Perna, progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).
100. J. Zhang *et al*., An integrative encode resource for cancer genomics. *Nat. Commun.* **11**, 3696 (2020).
101. B. Bonev *et al*., Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
102. M. Kadota *et al*., CTCF binding landscape in jawless fish with reference to Hox cluster evolution. *Sci. Rep.* **7**, 1–11 (2017).
103. L. Niu *et al*., Three-dimensional folding dynamics of the *Xenopus tropicalis* genome. *Nat. Genet.* **53**, 1075–1087 (2021).
104. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
105. H. Li *et al*., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
106. P. Virtanen *et al*., Scipy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).