

Letter to *MBE* - Discoveries

**Feather development genes and associated regulatory
innovation predate the origin of Dinosauria**

**Craig B. Lowe¹, Julia A. Clarke², Allan J. Baker³, David Haussler⁴,
and Scott V. Edwards^{5*}**

¹Howard Hughes Medical Institute and Stanford University School of Medicine, Stanford, CA 94305

²Department of Geological Sciences, University of Texas at Austin, Austin, TX 78713

³Department of Natural History, Royal Ontario Museum, Toronto, and Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada M5S 2C6

⁴Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064

⁵Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138

*Correspondence: sedwards@fas.harvard.edu

The evolution of avian feathers have recently been illuminated by fossils and the identification of genes involved in feather patterning and morphogenesis. However, molecular studies have focused mainly on protein-coding genes. Using comparative genomics and more than 600,000 conserved regulatory elements, we show that patterns of genome evolution in the vicinity of feather genes are consistent with a major role for regulatory innovation in the evolution of feathers. Rates of innovation at feather regulatory elements exhibit an extended period of innovation with peaks in the ancestors of amniotes and archosaurs. We estimate that 86% of such regulatory elements were present prior to the origin of Dinosauria. On the branch leading to modern birds, we detect a strong signal of regulatory innovation near IGFBP2 and IGFBP5, which have roles in body size reduction, and may represent a genomic signature for the miniaturization of dinosaurian body size preceding the origin of flight.

Feathers constitute complex branched structures that arise through interactions between the dermis and epidermis (Widelitz et al. 2003; Mou et al. 2011; Ng et al. 2012; Li et al. 2013; Lin et al. 2013). Although feathers were long thought to be a key innovation associated with the origin of avian flight, paleontological discoveries over the past fifteen years indicate a more ancient origin; filamentous feather precursors are now known to be present in many lineages of non-avian dinosaurs, and pennaceous feathers clearly arose prior to the origin of flight (Xu et al. 2001; Norell and Xu 2005; Zheng et al. 2009; Kellner et al. 2010; Godefroit et al. 2014). At the same time, the molecular processes underlying feather development and deployment throughout the integument are becoming better known through studies of gene expression patterns (Antin et al. 2014) and natural mutants (Mou et al. 2011; Ng et al. 2012). Comparative genomics can offer insights into the evolutionary history of functional elements in the genome; however, aside from the β -keratins, which are known to have diversified extensively on the lineage leading to birds (Li et al. 2013), we know little about evolutionarily novel genes or noncoding regions associated with feather development. Recent studies have shown that regulatory changes underlie many key phenotypes in vertebrates (Karlsson et al. 2007; Chan et al. 2010; McLean et al. 2011; reviewed in Wray 2013), but regulatory innovations associated with the origins of feathers have not been systematically explored. In particular, conserved non-exonic elements (CNEEs) have emerged as important regulators of gene expression (Visel et al. 2008) and have revealed the evolutionary dynamics of genomic regions associated with novel phenotypes such as mammalian hair (Lowe et al. 2011).

Results and Discussion

Conserved non-exonic elements and constraint in the avian genome. We identified a set of 193 genes that have been associated with feather development through mutant phenotypes or spatiotemporally restricted expression patterns (Supplementary Materials and Supplementary Table 1). To investigate the evolutionary history of these genes and their potential regulatory elements, we constructed a 19-way whole-genome alignment referenced on the chicken genome (Hillier et al. 2004) containing four birds, two crocodilians, two turtles, a lizard, four mammals, a frog, and five actinopterygian (ray-finned) fish. Regions of the genome showing evolutionary constraint were identified using a phylogenetic hidden Markov model to detect regions of the alignment evolving more slowly than synonymous sites in coding regions. Overall, 957,409 conserved elements totaling ~71Mbp and spanning ~7.2% of the chicken genome were identified, a higher percentage than the 5% often reported for the human genome. This result is consistent with the small (1.2 Gb) size of the chicken genome relative to the human genome, making the total amount of sequence annotated as constrained about half of what is currently reported for human (Siepel et al. 2005; Lindblad-Toh et al. 2011). To identify putative regulatory elements we removed any regions overlapping an exon annotated in chicken, or another species, resulting in 602,539 CNEEs covering 4.4% of the chicken genome. We identified the gene that each CNEE is likely to regulate by assigning each CNEE to the gene with the closest transcription start site, and found that 13,307 of the CNEEs were associated with the 193 feather-related genes in the data set. Although regulatory elements can act over long genomic distances that include genes not regulated by the elements (Kleinjan and

van Heyningen 2005), experimentally identified enhancers tend to be closest to genes with expression in the same tissues and at the same times in development (Visel et al. 2009). Additionally, many regulatory regions undergo rapid evolution and turnover (Wray 2007; Wray 2013), and these will be missed by our analysis. Due to their different functions, we split the list of 193 feather related genes and their associated CNEEs into a structural set of 67 keratin genes and a patterning set of 126 non-keratin genes and analyzed these groups separately.

An ancient genic toolkit and extended regulatory evolution are associated with feather origins. The genic and regulatory components of the keratin and non-keratin sets show very different patterns across the 500My backbone of our tree, on the lineage leading from the common ancestor of vertebrates to the chicken in our tree (Figs. 1 and 2, Supplementary Fig. 1). The most ancient branch in our analysis, leading to the common ancestor of ray-finned fishes and other vertebrates, shows the strongest enrichment for the non-keratin feather genes (1.7 times expected), with smaller numbers of non-keratin feather genes arising on branches leading to tetrapods and less inclusive clades (Fig. 1A, Fig. 2). No members of this non-keratin feather gene set are reconstructed to have arisen after the ancestor of birds and turtles. Although ancient genes are more likely to be studied during chick development, the non-keratin genes in our study were even more ancient than we would expect taking into account this bias (Mann-Whitney U test; $p < 0.022$; Supplementary Figure 2). The inferred first appearance of non-keratin protein-coding regions that are involved, for example, in placode patterning and feather ontogeny in birds is consistent with these genes being part of an ancient developmental toolkit (Figs. 1 and 2).

Surprisingly, the CNEEs associated with non-keratin feather related genes show the highest rate of origin not on the internode between the ancestral archosaur and birds, where they exhibit a 25% higher-than-expected rate of origination, but instead on the branch leading to amniotes, where they exhibit a rate of origination 60% higher than expected (Figs. 1 and 2, Supplementary Fig. 1). The rate of origination for these CNEEs is greater than what would be expected from CNEEs uniformly distributed throughout the genome for 6 of the 8 branches along the lineage leading to chicken, suggesting a large amount of regulatory innovation over an extended time period (Figs. 1A and 2). Thus, the non-keratin genic component of feather development arose deep in vertebrates and the greatest signal of regulatory innovation was coincident with the burst of phenotypic change associated with the transition to land. Although information on the integument of the ancestral amniote remains exceptionally limited (Alibardi et al. 2009; Alibardi 2012), the accumulation of CNEEs inferred to have occurred at this time indicates a key role for regulatory change during this transition and in the subsequent evolution of vertebrate integumentary diversity. Consistent with this hypothesis, 32 genes in our feather gene set are here identified as shared with those involved in the development of mammalian hair (Lowe et al. 2011) (hypergeometric distribution, $p < 1e-80$; Supplementary Table 3) and as present in the amniote ancestor. Genes driving hair development have been previously shown to exhibit an increase in regulatory innovation on the branch leading to amniotes, followed by a peak on the branch leading to mammals and a decline more recently (Lowe et al. 2011).

Our analysis suggests that non-avian dinosaurs, as part of Archosauria, possessed the entirety of the known non-keratin protein-coding toolkit for making

feathers. Moreover, assuming a constant rate of genome-wide accumulation of CNEEs throughout vertebrates, we estimate that 86% of non-keratin feather gene CNEEs were also present in the archosaur ancestor. The CNEEs present in this ancestor may have less to do with feather origins but instead could be linked to the earlier amniote transition to land, with later, bird-specific CNEEs having feather-specific functions. These results are also consistent with new data on integumentary innovation and diversity in Archosauria: filamentous or bristle structures either originated once early in the clade or three or more times (Clarke 2013) in pterosaurs (Kellner et al. 2010), ornithischian (Zheng et al. 2009; Godefroit et al. 2014) and theropod dinosaurs (Norell and Xu 2005). Thus, the genic and regulatory complement identified in the ancestral archosaur was either a flexible toolkit coopted in multiple origins of new structures including feathers, or indicates an ancient origin in that clade for filamentous integumentary structures, often called feather precursors, on some part of the body or stage in development more than 100 million years before the origin of pinnate feathers in dinosaurs.

Limited role of protein evolution in feather origins. Our analysis detects the well-known burst of duplication in β -keratin genes within Archosauria (Greenwold and Sawyer 2010; Li et al. 2013) on the branch leading to birds (Figs. 1B, 2). The larger peak for keratin innovation is comprised of 57 β -keratins arising as an expansion of a gene cluster on chicken chromosome 27 and 5 β -keratins from duplications on chromosome 2. The small peak in the turtle-bird ancestor is due to the expansion of a β -keratin gene cluster on chromosome 25. Both of these results are consistent with

previous studies of β -keratin evolution (Greenwold and Sawyer 2010; Li et al. 2013). However, this keratin burst constitutes the only, albeit substantial, signal of innovation at the protein level in pinnate feather origins. Notably, there is little evidence for regulatory innovation in the vicinity of β -keratin genes. We detected little additional cross-species constraint outside of the exonic regions in the keratin clusters than we would expect if CNEEs were randomly distributed in the genome. We only detected 15 CNEEs neighboring feather-related keratins in the lineage leading to birds, suggesting that regulatory evolution near β -keratins is not exceptional. Although the signature of CNEEs is likely complicated by a history of duplication and gene conversion in this multigene family, either the regulatory landscape around β -keratins does not appear noteworthy or their regulatory elements are under less severe constraint. These data are consistent with the idea that the keratin component of feathers arose primarily as a result of genic innovations.

Aside from β -keratin evolution, protein evolution appears to play a limited role in pinnate feather origins. We searched for signals of positive selection with respect to amino acid substitutions. After Bonferroni correction, only 3 of the 126 non-keratin feather genes showed signatures of positive selection on the archosaurian branch leading to birds (Supplementary Table 2). These results indicate that most non-keratin genes related to feather development exhibit regulatory, not protein-coding, innovations in the avian stem lineage, including living birds and non-avian dinosaurs, consistent with the hypothesis that regulatory innovations underlie adaptations in skin patterning and feather morphology.

Body size genes exhibit exceptional regulatory innovation in Dinosauria: Genes with an anomalously large number of regulatory elements arising in birds after their divergence from extant crocodilians may contribute to the origin of avian phenotypes. A genome-wide survey of 1 Mb genomic windows revealed 23 segments of the chicken genome possessing anomalously high numbers of CNEEs arising on the branch leading to birds (Fig. 3a; corrected $p < 0.01$; Supplementary Table 4). Although gene ontology analysis does not reveal significant enrichment for any functions for the set of genes near these innovation-rich segments, a number of these segments flank genes involved in body size, limb development, and integument (Fig. 3a). The region showing the greatest enrichment for bird-specific CNEEs in the entire chicken genome, over 500 percent more than expected ($p < 1^{-53}$), is centered in a 400-kb gene desert with insulin-like growth factor binding protein (IGFBP) 2 and 5 being the two closest genes (Fig 3b and c). IGFBP2 is expressed in the chick apical ectodermal ridge and at the tips of the growth plates in the wing bud, contains single nucleotide polymorphisms linked to phenotypic variation in the limbs of chickens (McQueeney and Dealy 2001; Li et al. 2006), and lies in the signaling pathway of both body size and limb length in mammals and birds (Fisher et al. 2005; Sutter et al. 2007). IGFBP5 also plays important roles in limb development (McQueeney and Dealy 2001) and the reduction of body size (Salih et al. 2004). Its widespread expression during chick development (Antin et al. 2014) is consistent with a role for IGFBP5-associated regulatory elements in body size reduction. Body size and limb length are known to vary extensively across Dinosauria and have been proposed to play a key role in dinosaur evolutionary dynamics (Benson et al.

2014), with miniaturization indicated by the fossil record to have preceded the origin of flight in Paraves (Turner et al. 2007; Lee et al. 2014), and changes in limb scaling within Maniraptora and continuing into birds associated with the origin of flight (Xu et al. 2001). Thus, analysis of patterns of regulatory innovation offer the potential to link genome evolution to key shifts in shape and form occurring in deep time.

Acknowledgments We thank Jacob Musser, Gunter Wagner and Rick Prum for discussion and comments on an earlier draft of this manuscript, and Lucas Moreira for conducting the PAML analysis. Two anonymous reviewers provided helpful comments. Niclas Backström, Matt Fujita, Clemens Küpper, Frank Rheindt, Miguel Alcaide, Mark Liu, Moos Blom and Daria Shipilina helped collect Ensemble IDs. CBL and DH were supported by the Howard Hughes Medical Institute. AJB acknowledges support from NSERC grant 200-12. SVE and JC acknowledge support from the US National Science Foundation (DEB-1355343/DEB-1355292).

Author Contributions DH, CBL and SVE conceived the study; CBL and SVE collected and analyzed data; CBL, SVE and JC wrote the paper, with comments from all other authors.

Author Information: All alignments and genomic coordinates of genes and CNEEs used in the analysis can be found at <http://hgwdev-lowec.cse.ucsc.edu/>

Competing financial interests statement. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to sedwards@fas.harvard.edu, lowec@stanford.edu, or Julia_Clarke@jsg.utexas.edu.

References

- Alibardi, L. 2012. Perspectives on hair evolution based on some comparative studies on vertebrate cornification. *Journal of experimental zoology. Part B, Molecular and developmental evolution* 318:325-343.
- Alibardi, L, L Dalla Valle, A Nardi M Toni. 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. *Journal of Anatomy* 214:560-586.
- Antin, PB, TA Yatskievych, S Davey DK Darnell. 2014. GEISHA: an evolving gene expression resource for the chicken embryo. *Nucleic Acids Research* 42:D933-937.
- Benson, RBJ, NE Campione, MT Carrano, PD Mannion, C Sullivan, P Upchurch DC Evans. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *Plos Biology* 12:e1001853-e1001853.
- Chan, YF, ME Marks, FC Jones, et al. 2010. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science* 327:302-305.
- Clarke, J. 2013. Feathers Before Flight. *Science* 340:690-692.
- Fisher, MC, C Meyer, G Garber CN Dealy. 2005. Role of IGFBP2, IGF-I and IGF-II in regulating long bone growth. *Bone* 37:741-750.
- Godefroit, P, SM Sinitsa, D Dhouailly, YL Bolotsky, AV Sizov, ME McNamara, MJ Benton P Spagna. 2014. A Jurassic ornithischian dinosaur from Siberia with both feathers and scales. *Science* 345:451-455.
- Greenwold, MJ, RH Sawyer. 2010. Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*): implications for feather evolution. *Bmc Evolutionary Biology* 10.
- Hedges, SB, J Dudley S Kumar. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971-2972.
- Hillier, LW, W Miller, E Birney, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.
- Karlsson, EK, I Baranowska, CM Wade, et al. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* 39:1321-1328.
- Kellner, AWA, XL Wang, H Tischlinger, DD Campos, DWE Hone X Meng. 2010. The soft tissue of *Jeholopterus* (Pterosauria, Anurognathidae, Batrachognathinae) and the structure of the pterosaur wing membrane. *Proceedings of the Royal Society B-Biological Sciences* 277:321-329.
- Kleinjan, DA, V van Heyningen. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *American journal of human genetics* 76:8-32.

- Lee, MSY, A Cau, D Naish GJ Dyke. 2014. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* 345:562-566.
- Li, YI, LS Kong, CP Ponting W Haerty. 2013. Rapid Evolution of Beta-Keratin Genes Contribute to Phenotypic Differences That Distinguish Turtles and Birds from Other Reptiles. *Genome Biology and Evolution* 5:923-933.
- Li, ZH, H Li, H Zhang, SZ Wang, QG Wang YX Wang. 2006. Identification of a single nucleotide polymorphism of the insulin-like growth factor binding protein 2 gene and its association with growth and body composition traits in the chicken. *Journal of animal science* 84:2902-2906.
- Lin, SJ, RB Wideliz, Z Yue, A Li, X Wu, TX Jiang, P Wu CM Chuong. 2013. Feather regeneration as a model for organogenesis. *Dev Growth Differ* 55:139-148.
- Lindblad-Toh, K, M Garber, O Zuk, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478:476-482.
- Lowe, CB, M Kellis, A Siepel, BJ Raney, Clamp, Michele, SR Salama, DM Kingsley, K Lindblad-Toh D Haussler. 2011. Three periods of regulatory innovation during vertebrate evolution. *Science* 333:1019-1024.
- McLean, CY, PL Reno, AA Pollen, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471:216-219.
- McQueeney, K, CN Dealy. 2001. Roles of insulin-like growth factor-I (IGF-I) and IGF-I binding protein-2 (IGFBP2) and -5 (IGFBP5) in developing chick limbs. *Growth Hormone & IGF Research* 11:346-363.
- Mou, C, F Pitel, D Gourichon, et al. 2011. Cryptic Patterning of Avian Skin Confers a Developmental Facility for Loss of Neck Feathering. *Plos Biology* 9.
- Ng, CS, P Wu, J Foley, et al. 2012. The Chicken Frizzle Feather Is Due to an alpha-Keratin (KRT75) Mutation That Causes a Defective Rachis. *Plos Genetics* 8.
- Norell, MA, X Xu. 2005. Feathered dinosaurs. *Annual Review of Earth and Planetary Sciences* 33:277-299.
- Salih, DAM, G Tripathi, C Holding, TAM Szesztak, MI Gonzalez, EJ Carter, LJ Cobb, JE Eisemann JM Pell. 2004. Insulin-like growth factor-binding protein 5 (Igfbp5) compromises survival, growth, muscle development, and fertility in mice. *Proceedings of the National Academy of Sciences of the United States of America* 101:4314-4319.
- Siepel, A, G Bejerano, JS Pedersen, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research* 15:1034-1050.
- Sutter, NB, CD Bustamante, K Chase, et al. 2007. A single IGF1 allele is a major determinant of small size in dogs. *Science* 316:112-115.
- Turner, AH, D Pol, JA Clarke, GM Erickson MA Norell. 2007. A basal dromaeosaurid and size evolution preceding avian flight. *Science* 317:1378-1381.
- Visel, A, MJ Blow, Z Li, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854-858.
- Visel, A, S Prabhakar, JA Akiyama, M Shoukry, KD Lewis, A Holt, I Plajzer-Frick, V Afzal, EM Rubin LA Pennacchio. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genetics* 40:158-160.
- Widelitz, RB, TX Jiang, MK Yu, T Shen, JY Shen, P Wu, ZC Yu CM Chuong. 2003. Molecular biology of feather morphogenesis: A testable model for evo-devo research. *Journal*

- of Experimental Zoology Part B-Molecular and Developmental Evolution 298B:109-122.
- Wray, GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* 8:206-216.
- Wray, GA. 2013. Genomics and the Evolution of Phenotypic Traits. *Annual Review of Ecology, Evolution, and Systematics* 44:51-72.
- Xu, X, HH Zhou RO Prum. 2001. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* 410:200-204.
- Zheng, X-T, H-L You, X Xu Z-M Dong. 2009. An Early Cretaceous heterodontosaurid dinosaur with filamentous integumentary structures. *Nature* 458:333-336.

Figures (main text)

Figure 1. Feather development genes are ancient whereas associated CNEEs peak in the amniote ancestor.

Evolutionary dynamics of a) non-keratin feather development genes and associated CNEEs (n = 126 genes) and b) keratin genes and associated CNEEs (n = 67 genes). The black horizontal line indicates the null expectation of the number of new genes (comparison to all genes in the genome) or CNEEs (a uniform distribution throughout the genome). Points above this line indicate lineages on which a higher-than-expected number of genes or CNEEs have arisen. Points on the X-axis correspond to the ancestors depicted in Fig. 2, with spacing proportional to divergence times as recorded in timetree.org (Hedges et al. 2006). In b, the larger peak is comprised of β -keratins arising from expansions of gene clusters on chicken chromosomes 27 and 2. The small peak in the turtle-bird ancestor is due to the expansion of a β -keratin gene cluster on chromosome 25. Both of these results are consistent with previous studies of β -keratin evolution (Greenwold and Sawyer 2010; Li et al. 2013).

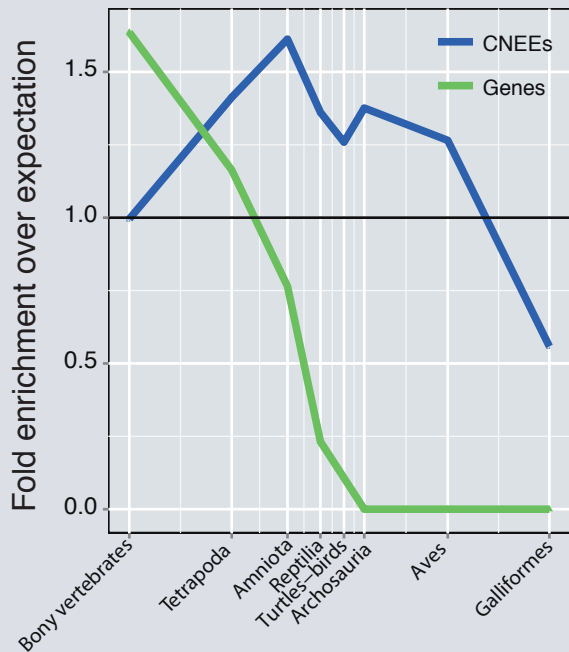
Figure 2. Major genomic events underlying the origin of feathers. The colored backbone of the tree is comprised of three tracks: CNEEs, non-keratin feather genes (n=126), and keratin genes (n=67). Rates of origination of these three genomic classes are indicated by the colors for each stem internode and track in the tree, with blue colors indicating low origination rates and red colors indicating high origination rates. Key events at the level of coding regions (genes) and regulatory elements are indicated.

The colors of the silhouettes at right indicate the percent of the feather regulatory component present in the chicken genome inferred to have arisen in the ancestor of each indicated taxon. For example, the fish are inferred to possess about 28% of the CNEEs associated with feather genes in chicken, whereas 86% of the observed chicken CNEEs are inferred to have arisen by the ancestral archosaur, including non-avian dinosaurs.

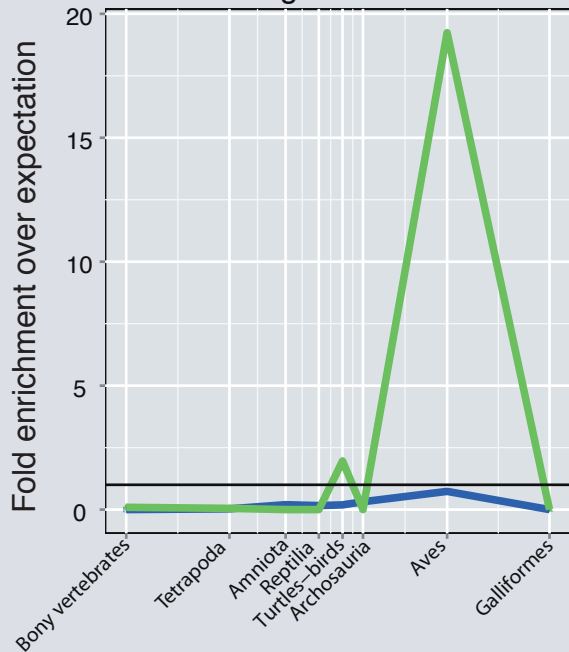
Figure 3. Identification of regions of the avian genome with signatures for exceptional regulatory innovation on the archosaur lineage that includes birds and other dinosaurs. a) A genome-wide plot of the density of conserved nonexonic elements (CNEEs) arising on the archosaurian branch leading to the avian ancestor. Red regions indicate those areas enriched compared to the distribution of CNEEs on other branches (gray line in ‘b’) and green squares indicate the 23 significant peaks of enrichment for bird-specific CNEEs relative to a uniform distribution throughout the genome. We examined the closest upstream and closest downstream genes and for select peaks a flanking gene is indicated along with a proposed role in avian morphological evolution (key at top); regulatory innovation may also have played a role in earlier dinosaur-lineage evolutionary dynamics. b) The densest region for bird-specific CNEEs in the chicken genome is in a gene desert on chromosome 7 with IGFBP2 being the closest well-annotated refseq gene and IGFBP5 being the closest gene prediction. CNEE density on all branches other than the one leading to birds is indicated in grey. c) UCSC Genome Browser shot of a CNEE-rich region in the vicinity

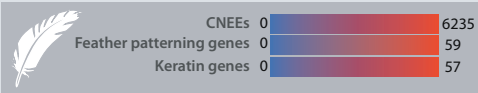
of IGFBP2 and IGFBP5, which function in limb development and body size regulation (see main text, Supplementary Table 4), showing CNEEs found only in birds (red boxes) or arising on deeper branches in the vertebrate tree (gray boxes). Regions of aligning sequence for representatives of the 19 included taxa are in green.

a Non-keratin feather genes & CNEEs



b Keratin genes & CNEEs

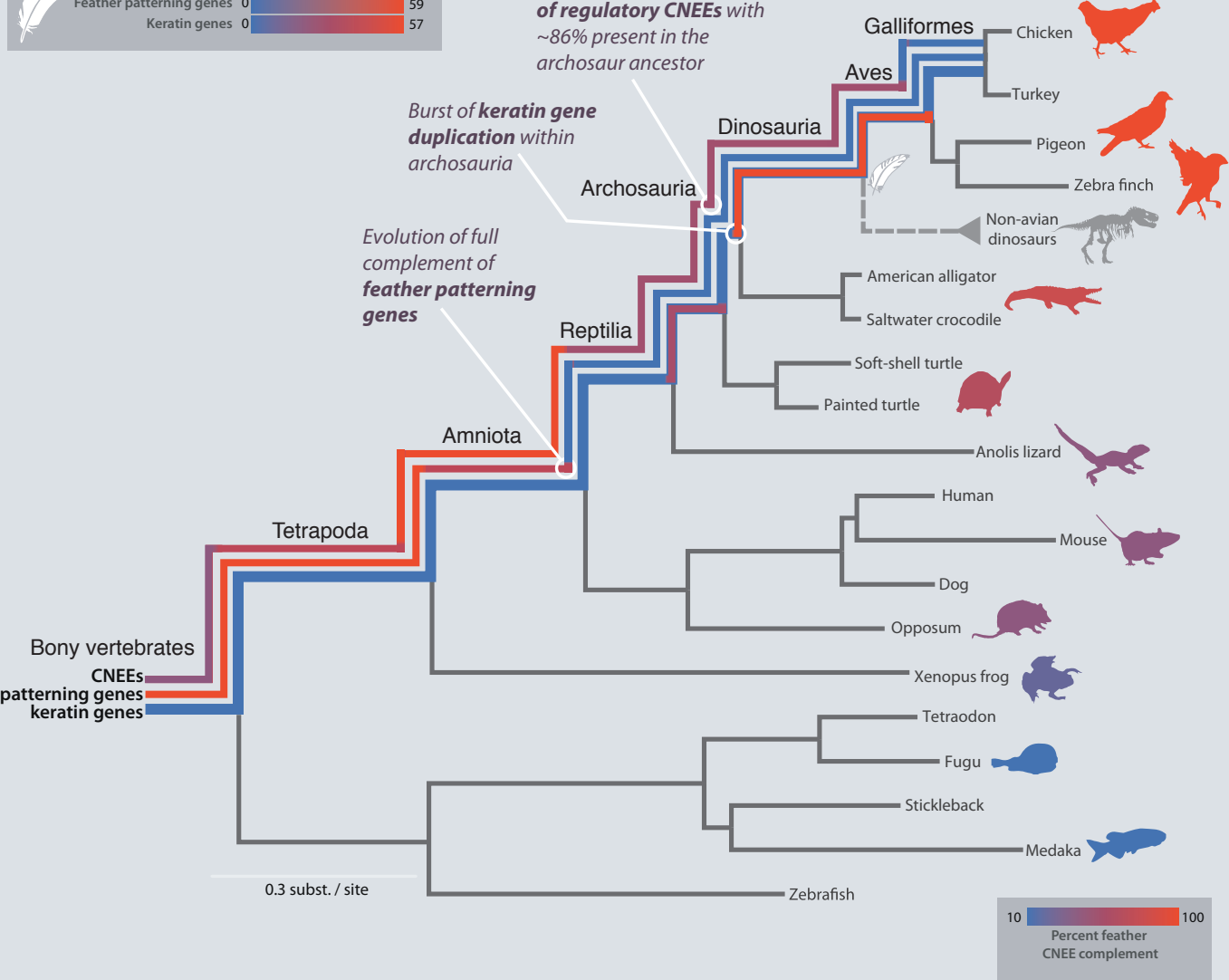




**Extended accumulation
of regulatory CNEEs with
~86% present in the
archosaur ancestor**

**Burst of *keratin gene
duplication* within
archosauria**

**Evolution of full
complement of
*feather patterning
genes***



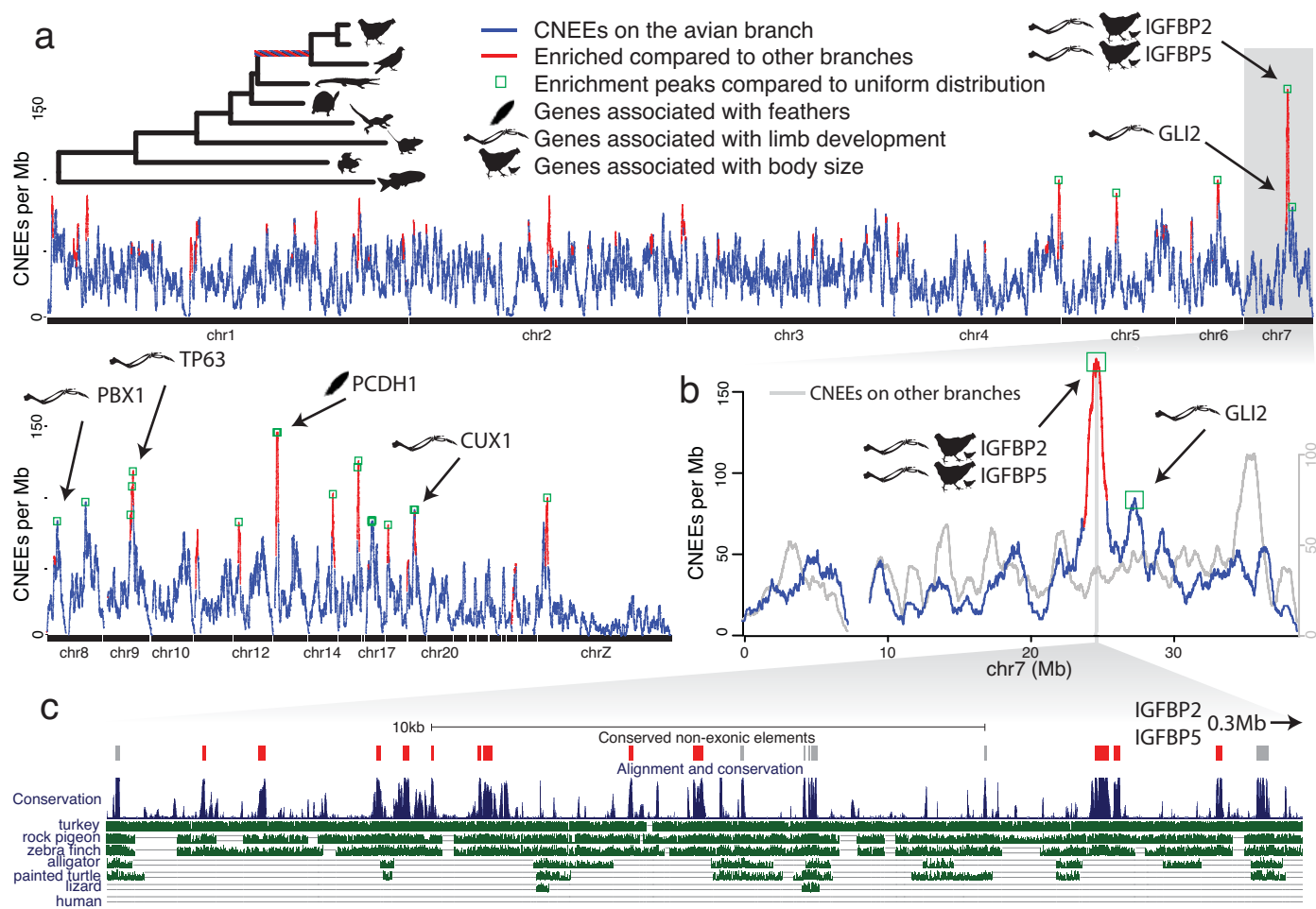


Figure 3