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

Letter to *MBE* - Discoveries

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

Craig B. Lowe[1], Julia A. Clarke[2], Allan J. Baker[3], David Haussler[4], and Scott V. Edwards[5*]

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

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

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

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

[5]Department 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 spaciotemporally 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 Szestak, 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)

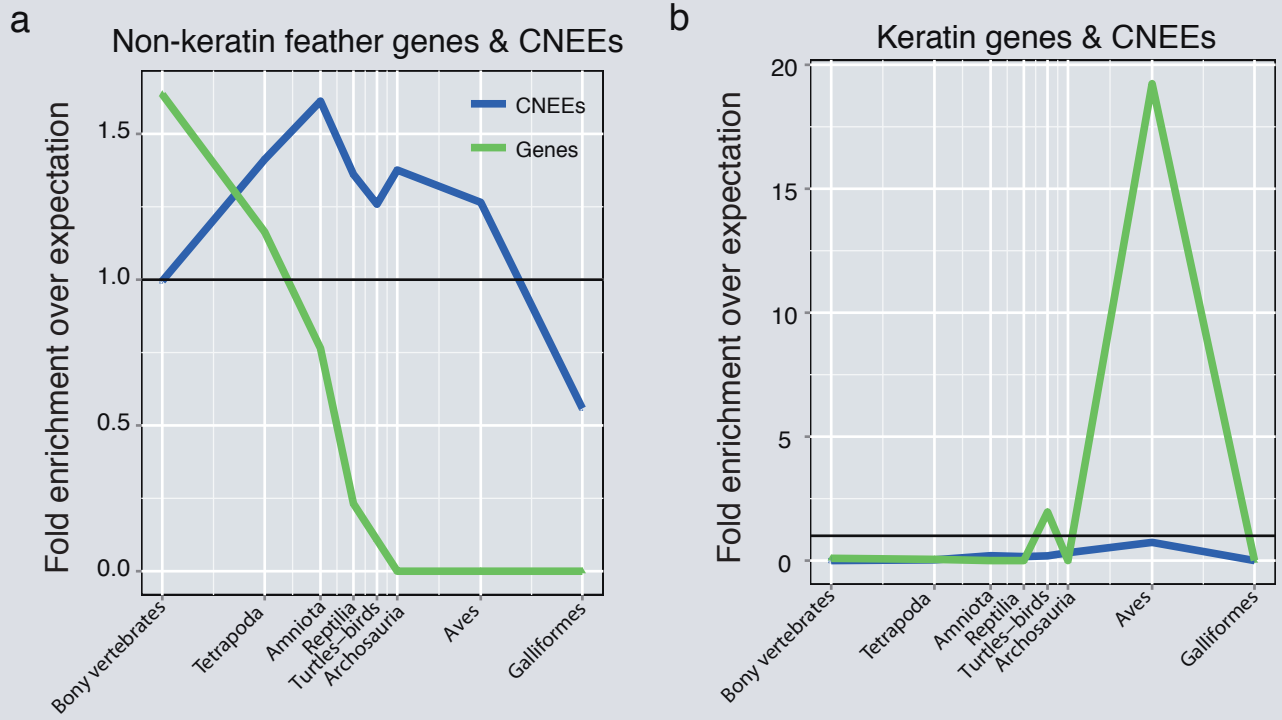
Figure 1. Feather development genes are ancient whereas associated CNEEs peak in the amniote ancestor. Evolutionary dynamics of a) non-keratin feather development genes and associated CNEEs (n = 126 genes) and b) keratin genes and associated CNEEs (n = 67 genes). The black horizontal line indicates the 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).

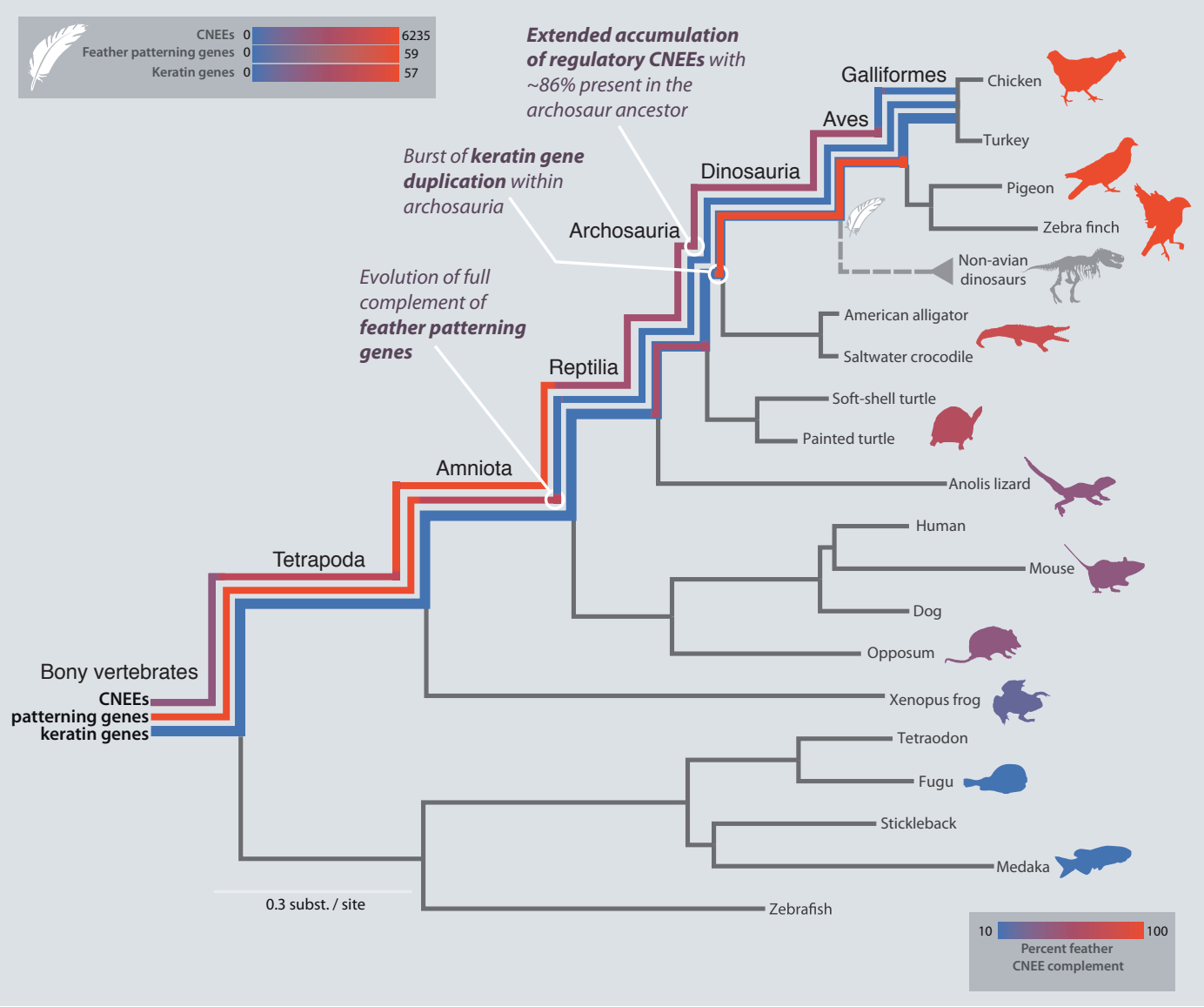
Figure 2. Major genomic events underlying the origin of feathers. The colored backbone of the tree is comprised of three tracks: CNEEs, non-keratin feather genes (n=126), and keratin genes (n=67). Rates of origination of these three genomic classes are indicated by the colors for each stem internode 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.

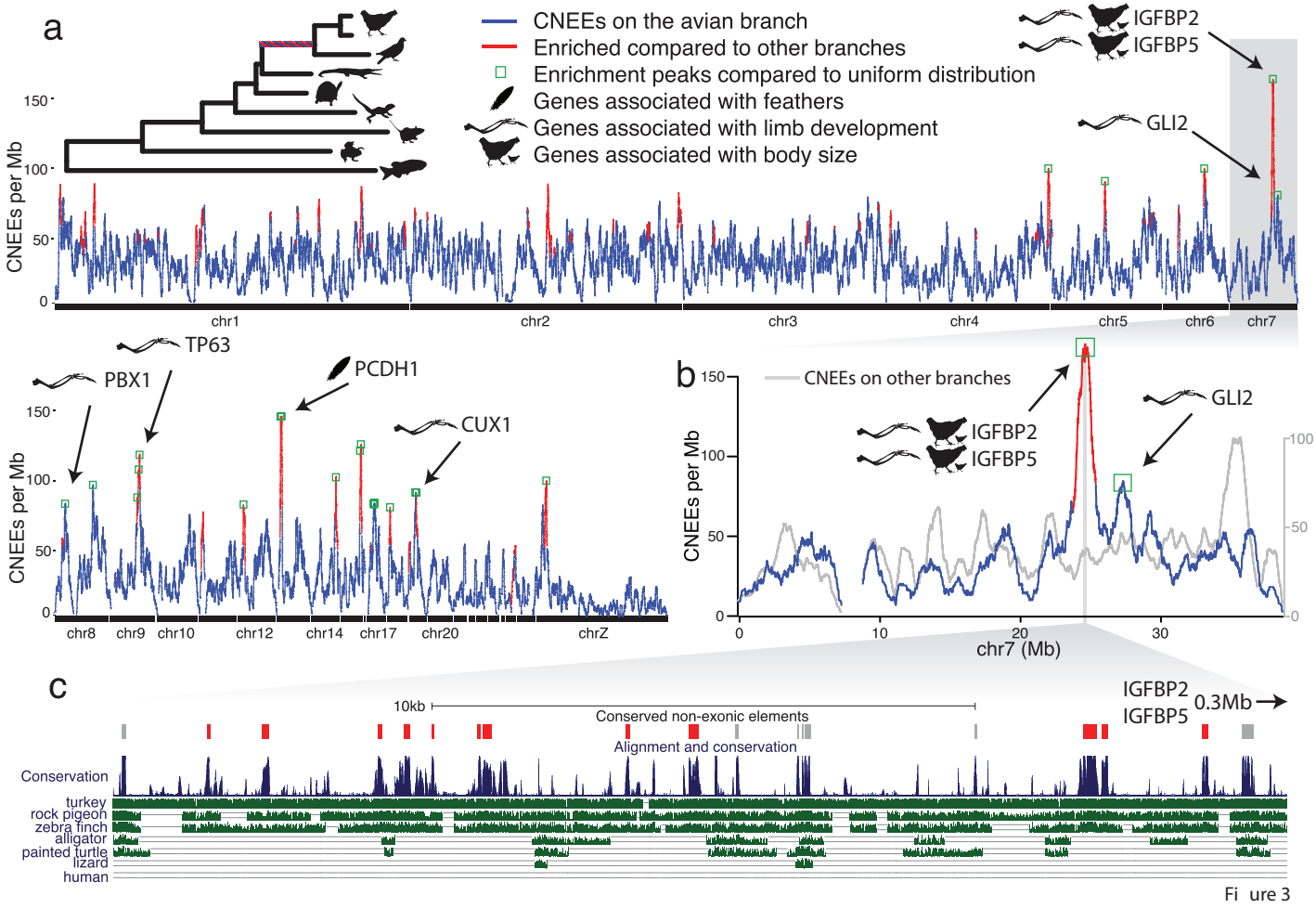
The colors of the silhouettes at right indicate the percent of the feather regulatory component present in the chicken genome inferred to have arisen in the ancestor of each indicated taxon. For example, the fish are inferred to possess about 28% of the CNEEs associated with feather genes in chicken, 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
Fold enrichment over expectation
1.5
1.0
0.5
0.0
CNEEs
Genes
Bony vertebrates
Tetrapoda
Amniota
Reptilia
Turtles–birds
Archosauria
Aves
Galliformes
b
Keratin genes & CNEEs
Fold enrichment over expectation
20
15
10
5
0
Bony vertebrates
Tetrapoda
Amniota
Reptilia
Turtles–birds
Archosauria
Aves
Galliformes


CNEEs 0 6235
Feather patterning genes 0 59
Keratin genes 0 57
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
Galliformes
Aves
Dinosauria
Archosauria
Reptilia
Amniota
Tetrapoda
Bony vertebrates
CNEEs
patterning genes
keratin genes
Chicken
Turkey
Pigeon
Zebra finch
Non-avian dinosaurs
American alligator
Saltwater crocodile
Soft-shell turtle
Painted turtle
Anolis lizard
Human
Mouse
Dog
Opposum
Xenopus frog
Tetraodon
Fugu
Stickleback
Medaka
Zebrafish
0.3 subst. / site
10 100
Percent feather CNEE complement


a
CNEEs on the avian branch
Enriched compared to other branches
Enrichment peaks compared to uniform distribution
Genes associated with feathers
Genes associated with limb development
Genes associated with body size
IGFBP2
IGFBP5
GLI2
CNEEs per Mb
150
100
50
0
chr1
chr2
chr3
chr4
chr5
chr6
chr7
TP63
PBX1
PCDH1
CUX1
chr8
chr9
chr10
chr12
chr14
chr17
chr20
chrZ
b
CNEEs on other branches
IGFBP2
IGFBP5
GLI2
0
10
20
30
chr7 (Mb)
c
10kb
Conserved non-exonic elements
Alignment and conservation
IGFBP2
IGFBP5
0.3Mb
Conservation
turkey
rock pigeon
zebra finch
alligator
painted turtle
lizard
human

Fi ure 3

Supplementary Material

Lowe et al. "Feather development genes and associated regulatory innovation predate the origin of Dinosauria"

Supplementary Materials and Methods:

Genome-wide Multiple Alignment. Our genome-wide alignment referenced on the chicken genome contains the following species and build numbers: Chicken (*Gallus gallus*; galGal3), Turkey (*Meleagris pavo*; melGal1), Pigeon (*Columba livia*; colLiv1), Zebra Finch (*Taeniopygia guttata*; taeGut1), American Alligator (*Alligator mississippiensis*; allMis1), American Crocodile (*Crocodylus acutus;* croPor0), Soft-shelled Turtle (*Pelodiscus sinensis*; pelSin1), Painted Turtle (*Chrysemys picta bellii*; chrPic1), Green Anole (*Anolis carolinensis*; anoCar2), Human (*Homo sapiens*; hg18), House Mouse (*Mus domesticus*; mm9), Dog (*Canis lupus familiaris*; canFam2), Opossum (*Monodelphis domestica*; monDom4), Xenopus (*Xenopus tropicalis*; xenTro3), Tetraodon (*Tetraodon nigriviridis*; tetNig1), Fugu (*Takifugu rubripes*; fr2), Stickleback (*Gasterosteus aculeatus*; gasAcu1), Medaka (*Oryzias latipes*; oryLat2), and Zebrafish (*Danio rerio*; danRer5). Initial pairwise alignments were done between the chicken genome and all other genomes using Blastz(Schwartz et al. 2003). We then chained these local alignments together to construct whole-genome alignments(Kent et al. 2003). We applied a synteny filter that requires a consistent order and orientation between the species, which helped distinguish orthologs from paralogs(Kent et al. 2003). We then used Multiz(Blanchette et al. 2004) to create the final multi-species genome-wide

alignment from the filtered pairwise alignments. We expect that additional avian genomes(Ellegren et al. 2012; Huang et al. 2013; Qu et al. 2013; Rands et al. 2013; Zhan et al. 2013) will add resolution to this analysis, but we suspect that the fraction of the genome estimated as constrained in our study will be similar to those employing additional genomes.

Identification of Conserved Non-Exonic Elements. We used a phylogenetic hidden Markov model (HMM) to define regions of cross-species conservation in our genome-wide alignments (Siepel, Haussler 2004). The HMM had both a neutral state and a conserved state that resisted substitutions. We defined the neutral state by fitting a time-reversible model of nucleotide substitution to four-fold degenerate sites in codons(Siepel, Haussler 2004). The conserved state has the same tree topology, but with the branch lengths reduced to three tenths of the neutral rate. We set the transition probabilities of the HMM by expecting a length of 45 base pairs and a coverage of 0.3. We then filtered this set of conserved regions by removing any elements overlapping exons according to spliced expressed sequence tags, Ensembl genes(Hubbard et al. 2002), RefSeq genes(Pruitt, Tatusova, Maglott 2005) or Transmap(Zhu et al. 2007).

Enrichment calculations for feather-related genes. There is variable statistical power to identify conserved elements depending on the evolutionary age of the element. For this reason we report the fraction of all CNEEs that arose on a branch that are associated with a given set of genes. We report enrichment as the fraction of CNEEs

arising on a branch near a set of genes divided by the fraction that would be expected near this set if CNEEs were uniformly distributed across the genome(Lowe, Bejerano, Haussler 2007). P-values for CNEE enrichment are calculated using the binomial distribution where the probability of success is the number of bases closest to a transcription start site for the set of genes being analyzed, divided by the number of bases in the genome. The number of trials is the number of CNEEs and the number of successes is the number of CNEEs closest to a transcription start site of a gene in the set being analyzed. For protein coding genes we report enrichment as the fraction of genes arising on a branch that are members of the gene set divided by the fraction of all genes that are in the gene set.

Identifying the Branch of Origin for Genes and CNEEs. For each gene or CNEE we used the branch leading to the most recent common ancestor of chicken and the most divergent species with an ortholog present in the alignment as the time of origin. For CNEEs, an ortholog was annotated as present if the species had orthologous bases for at least one-third of the bases in the CNEE. For genes, the species had to have orthologous bases present for at least 50 percent of the isoform with the most exons.

Genomic locations experiencing anomalous regulatory evolution leading to Aves.

We compared the genomic locations of CNEEs appearing on the branch leading to Aves against a uniform distribution along the genome. For each base in the genome we

calculated the number of CNEEs arising on the avian branch within 0.5Mb in either direction (roughly 1Mb windows). To identify the genomic location of peaks in this data set we fit a cubic spline with a smoothing parameter of 0.35. We calculated the significance of each peak based on the binomial distribution with the number of avian CNEEs divided by the number of bases being the probability of success, 1,000,001 being the number of trials, and Bonferroni correcting based on the number of bases in the genome. This resulted in a significance threshold of 76 CNEEs/Mb ($p < 0.01$). We also tested for significance against the distribution of CNEEs originating outside the avian branch being the null hypothesis instead of a uniform distribution across the genome. In this case we performed a fishers exact test for every base in the genome, comparing the number of avian CNEEs within the window to the number of non-avian CNEEs within the same range. We again used a p-value threshold of 0.01 after correcting for the number of tests being the number of bases in the genome.

Estimating the percentage of CNEEs near feather genes present in archosaurs.

To estimate the percentage of CNEEs present in Aves near non-keratin feather genes that were already present in archosaurs we needed to adjust the number, but not percentage, of feather-related CNEEs detected on each branch for the power to detect conserved elements that originated at that time. This is because for very recently originated sequence it is difficult to notice a resistance to substitutions, and for very ancient elements it is difficult to align over long distances, even when the sequence is

resisting many substitutions. To do this we assume that CNEEs in general, but not those near feather-related genes, have accumulated at a constant rate through time relative to the speed of substitutions in neutral DNA. For each branch in our analysis we maintain the fraction of CNEEs near feather-related genes, but adjust the counts so that the total number of CNEEs arising on each branch, divided by the length of the branch in substitutions per site, is a constant for all branches in our analysis.

j ={1, 2, 3, 4, 5, 6, 7} with 1 being the branch leading to Aves and higher numbers representing progressively more ancient branches.

b_j = length in substitutions per site of branch #j

f_j = number of feather-related CNEEs detect as originating on branch #j

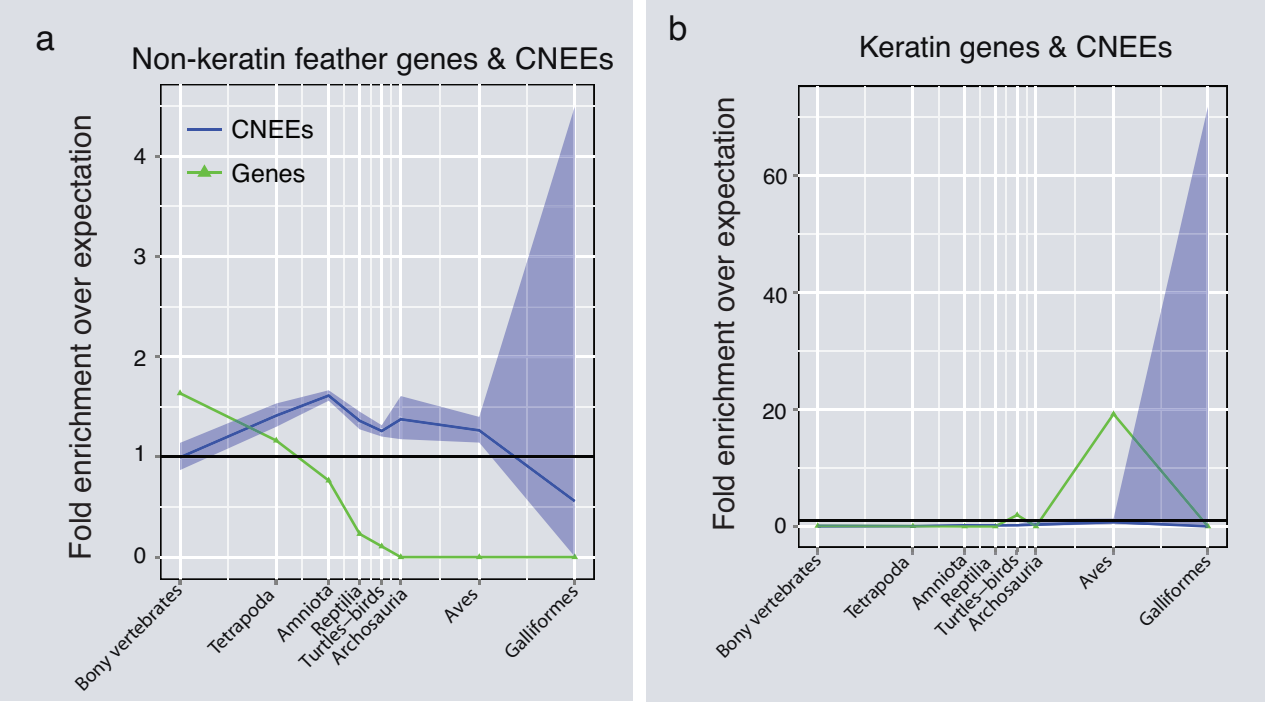
a_j = number of CNEEs detected as originating on branch #j

Fraction present in archosaurs =

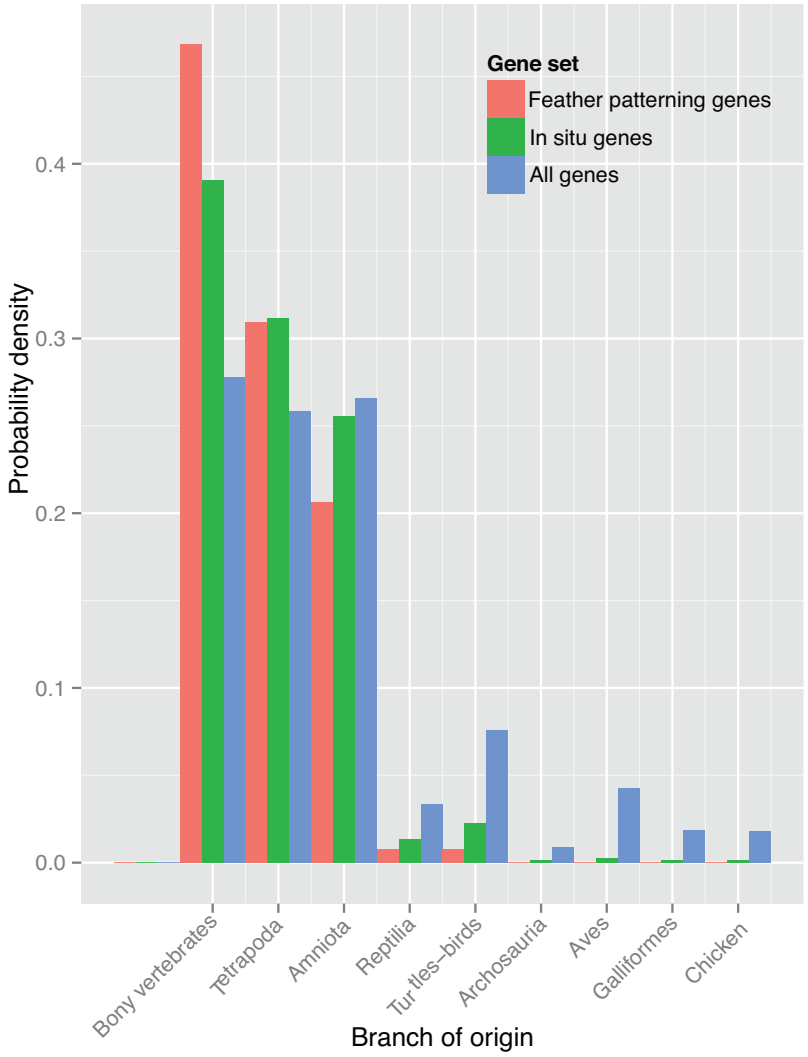
$$\frac{\sum_{j=2}^{7} \frac{f_j}{a_j} \times b_j}{\sum_{j=1}^{7} \frac{f_j}{a_j} \times b_j}$$

Analysis of protein evolution. Alignments of the protein-coding exons for the 126 non-keratin feather genes were extracted from the 19-way genomic alignment. To test the hypothesis that adaptive evolution of amino acid substitutions occurred on the branch leading to extant birds, we ran two branch-site models under Model A in the

package Phylogenetic Analysis using Maximum Likelihood 4 (PAML4, version 4.4). The null (neutral or purifying selection) model allows two values of ω, the ratio of nonsynonymous to synonymous substitutions, to vary between 0 and 1 (ω_0, site classes 0 and 2a) or to be constrained to equal 1 (ω_1, site classes 1 and 2b). The alternate model (allowing for positive selection) enables ω to be equal to or exceed 1 for some sites (classes 2a and 2b) for the avian branch. For each gene we recorded the likelihood of both models, conducted a likelihood ratio test, and used Bonferroni correction with the number of tests being equal to the number of genes. The results suggest that 3 out of 126 genes rejected the null model at $p < 0.01$.



Supplementary Figure 1 - Dynamics of feather-related gene and CNEE innovation across vertebrates. The key indicates symbols used to designate gene and CNEE trajectories. In both panels, the black horizontal line indicates the null expectation of the number of genes or CNEEs and the blue area indicates the 99% confidence limits on the fold-enrichment for CNEEs on a given branch. a, Dynamics of non-keratin feather-related genes and associated CNEEs (n = 126 genes). b, Dynamics of keratin genes and associated CNEEs (n = 67 genes). The green spike at the right consists of β-keratin genes inferred to have duplicated along the lineage leading to birds.



Supplementary Figure 2 - Feather related genes are more ancient than expected for the chicken genome. The distribution of origin for the 126 feather patterning genes (red) is more ancient when compared to both the entire set of Ensembl genes (blue; Mann-Whitney U test; $p < 1e\text{-}9$; see Hubbard et al. 2002) as well as the set of genes with Ensembl identifiers that have been deposited in a data base of in situ hybridizations during chick development (green; Mann-Whitney U test; $p < 0.022$; (Antin et al. 2014)).

Supplementary Table 1 - Keratin and non-keratin feather gene set.

Keratin or non-Keratin	Gene ID	Gene Symbol	PMID
Non-keratin	ENSGALG00000004212	Adam10	21780243
Non-keratin	ENSGALG00000012740	Adam12	21158755
Non-keratin	ENSGALG00000016427	Adam17	21780243
Non-keratin	ENSGALG00000008976	Adam22	21780243
Non-keratin	ENSGALG00000008582	Adam23	21780243
Non-keratin	ENSGALG00000003398	Adam9	21780243
Non-keratin	ENSGALG00000004270	Aldh1a2	21423653
Non-keratin	ENSGALG00000007129	Aldh1a3	21423653
Non-keratin	ENSGALG00000023542	Alpl	8563025
Non-keratin	ENSGALG00000006995	ApoD	15580625
Non-keratin	ENSGALG00000021455	Asip	18287407
Non-keratin	ENSGALG00000008830	Bmp2	17948257
Non-keratin	ENSGALG00000012429	Bmp4	8823374
Non-keratin	ENSGALG00000007668	Bmp7	15272377
Non-keratin	ENSGALG00000000608	Cdh1	18394491
Non-keratin	ENSGALG00000015132	Cdh2	19557684
Non-keratin	ENSGALG00000013034	Cecr2	20336606
Non-keratin	ENSGALG00000007280	Cldn1	15749086
Non-keratin	ENSGALG00000009641	Col1a2	15272377
Non-keratin	ENSGALG00000002552	Col3a1	15272377
Non-keratin	ENSGALG00000008494	Creb	7556946
Non-keratin	ENSGALG00000010257	Crp2	12128214
Non-keratin	ENSGALG00000011905	Ctnnb1	12949772
Non-keratin	ENSGALG00000011182	Delta1	9435296
Non-keratin	ENSGALG00000004256	Dhrs3	21423653
Non-keratin	ENSGALG00000004481	Eda	17948257
Non-keratin	ENSGALG00000004599	Eda2r	17362907
Non-keratin	ENSGALG00000016809	Edar	17948257
Non-keratin	ENSGALG00000014369	Edaradd	15673574
Non-keratin	ENSGALG00000007466	Ednrb2	18625062
Non-keratin	ENSGALG00000012155	Egf	12586066
Non-keratin	ENSGALG00000012363	Egfr	12586066
Non-keratin	ENSGALG00000012120	En1	9012506
Non-keratin	ENSGALG00000016428	Enpp2	17366625
Non-keratin	ENSGALG00000005256	Epha4	12949772
Non-keratin	ENSGALG00000001143	Ets1	9074942
Non-keratin	ENSGALG00000014872	Fgf10	15201222
Non-keratin	ENSGALG00000002203	Fgf18	11669372
Non-keratin	ENSGALG00000011835	Fgf2	15201222
Non-keratin	ENSGALG00000013663	Fgf20	22712610
Non-keratin	ENSGALG00000007563	Fgf3	15201222
Non-keratin	ENSGALG00000007562	Fgf4	15201222
Non-keratin	ENSGALG00000007706	Fgf8	9012506
Non-keratin	ENSGALG00000003311	Fgfr1	15201222
Non-keratin	ENSGALG00000009495	Fgfr2	21302260

Keratin	ENSGALG00000024142	Fker	22297689
Non-keratin	ENSGALG00000003578	Fn1	21497985
Non-keratin	ENSGALG00000023293	Foxe1	21999483
Non-keratin	ENSGALG00000003654	Foxn1	17675842
Non-keratin	ENSGALG00000014908	Fst	12949772
Non-keratin	ENSGALG00000006902	Fzf1	8602858
Non-keratin	ENSGALG00000023647	Gbx1	22826214
Non-keratin	ENSGALG00000013342	Gbx2	9767154
Non-keratin	ENSGALG00000019869	Gdf7	21423653
Non-keratin	ENSGALG00000009724	Grem1	10556075
Non-keratin	ENSGALG00000006778	Hhex	16172986
Non-keratin	ENSGALG00000023129	Hoxa13	21437328
Non-keratin	ENSGALG00000011059	Hoxa6	2100252
Non-keratin	ENSGALG00000009273	Hoxd11	17675842
Non-keratin	ENSGALG00000009270	Hoxd8	11934149
Non-keratin	ENSGALG00000009546	Htra1	18570253
Non-keratin	ENSGALG00000005499	Il17rd	15844098
Non-keratin	ENSGALG00000001405	Irf6	16245336
Non-keratin	ENSGALG00000007145	Itgb1	1371480
Non-keratin	ENSGALG00000009020	Jag1	12949772
Non-keratin	ENSGALG00000011696	Jag2	9435296
Non-keratin	ENSGALG00000013925	Kit	7544170
Non-keratin	ENSGALG00000011206	Kitlg	7544170
Keratin	ENSGALG00000019719	Krt19	22297689
Keratin	ENSGALG00000019716	Krt20	22297689
Keratin	ENSGALG00000002349	Krt75	22829773
Non-keratin	ENSGALG00000004284	Lfng	10704863
Non-keratin	ENSGALG00000000936	Lmx1	16330160
Keratin	ENSGALG00000024227	LOC425854	22297689
Keratin	ENSGALG00000024220	LOC427060	22297689
Keratin	ENSGALG00000024185	LOC770926	22297689
Non-keratin	ENSGALG00000023459	Mc1r	18625062
Non-keratin	ENSGALG00000007679	Mitf	18625062
Non-keratin	ENSGALG00000003580	Mmp2	21497985
Non-keratin	ENSGALG00000006438	Mnx1	11027598
Non-keratin	ENSGALG00000015013	Msx1	7537773
Non-keratin	ENSGALG00000002947	Msx2	7537773
Non-keratin	ENSGALG00000005911	Mtf2	15294861
Non-keratin	ENSGALG00000013956	Myb	1879337
Non-keratin	ENSGALG00000016308	Myc	1879337
Non-keratin	ENSGALG00000007839	Ncam1	8823374
Non-keratin	ENSGALG00000003114	Nog	12949772
Non-keratin	ENSGALG00000002375	Notch1	12949772
Non-keratin	ENSGALG00000002922	Notch2	9435296
Non-keratin	ENSGALG00000001024	P63	11118893
Non-keratin	ENSGALG00000002553	Pcdh1	23318466
Non-keratin	ENSGALG00000016941	Pcdh17	23318466
Non-keratin	ENSGALG00000009732	Pcdh18	23318466

Non-keratin	ENSGALG00000006822	Pcdh19	23318466
Non-keratin	ENSGALG00000016944	Pcdh8	23318466
Non-keratin	ENSGALG00000016908	Pcdh9	23318466
Non-keratin	ENSGALG00000014634	Pmel	17106652
Non-keratin	ENSGALG00000016600	Pomc	21187100
Non-keratin	ENSGALG00000003446	Prlr	22297689
Non-keratin	ENSGALG00000012620	Ptc1	16330160
Non-keratin	ENSGALG00000010133	Ptch2	11784016
Non-keratin	ENSGALG00000007130	Rac1	15136143
Non-keratin	ENSGALG00000002859	Rac3	15136143
Non-keratin	ENSGALG00000011298	Rarb	8619957
Non-keratin	ENSGALG00000002841	Rfng	10704863
Non-keratin	ENSGALG00000016485	Rhob	15136143
Non-keratin	ENSGALG00000000569	Sdc3	7729580
Non-keratin	ENSGALG00000009241	Sfrp2	15272377
Non-keratin	ENSGALG00000006379	Shh	18079472
Non-keratin	ENSGALG00000016698	Shox	16904661
Non-keratin	ENSGALG00000004885	Slc24a5	18287407
Non-keratin	ENSGALG00000003310	Slc45a2	18625062
Non-keratin	ENSGALG00000007870	Smad3	16330160
Non-keratin	ENSGALG00000018639	Smad7	16330160
Non-keratin	ENSGALG00000015310	Sobp	22802190
Non-keratin	ENSGALG00000010797	Sostdc1	21423653
Non-keratin	ENSGALG00000005984	Sox18	11418236
Non-keratin	ENSGALG00000005285	Tbx4	16330160
Non-keratin	ENSGALG00000009612	Tgfb2	8903351
Non-keratin	ENSGALG00000011442	Tgfbr2	8903351
Non-keratin	ENSGALG00000011294	Thrb	10828844
Non-keratin	ENSGALG00000008120	Tle3	12666196
Non-keratin	ENSGALG00000007113	Tnc	12949772
Non-keratin	ENSGALG00000017119	Tnfrsf19	17362907
Non-keratin	ENSGALG00000016971	TSC-22	11803572
Non-keratin	ENSGALG00000004274	Twist2	11744368
Non-keratin	ENSGALG00000017237	Tyr	8612714
Non-keratin	ENSGALG00000015205	Tyrp1	18625062
Non-keratin	ENSGALG00000000839	Wnt11	7779076
Non-keratin	ENSGALG00000001079	Wnt3	18394491
Non-keratin	ENSGALG00000011358	Wnt6	17948257
Non-keratin	ENSGALG00000005123	Wnt7a	12949772
Non-keratin	ENSGALG00000005668	Wsb1	10354473
Keratin	ENSGALG00000009196		20482795
Keratin	ENSGALG00000009987		20482795
Keratin	ENSGALG00000018679		20482795
Keratin	ENSGALG00000018914		20482795
Keratin	ENSGALG00000022753		20482795
Keratin	ENSGALG00000022754		20482795
Keratin	ENSGALG00000022755		20482795
Keratin	ENSGALG00000022756		20482795

Keratin	ENSGALG00000022757		20482795
Keratin	ENSGALG00000024140		20482795
Keratin	ENSGALG00000024143		20482795
Keratin	ENSGALG00000024144		20482795
Keratin	ENSGALG00000024145		20482795
Keratin	ENSGALG00000024147		20482795
Keratin	ENSGALG00000024154		20482795
Keratin	ENSGALG00000024174		20482795
Keratin	ENSGALG00000024179		20482795
Keratin	ENSGALG00000024180		20482795
Keratin	ENSGALG00000024181		20482795
Keratin	ENSGALG00000024184		20482795
Keratin	ENSGALG00000024186		20482795
Keratin	ENSGALG00000024187		20482795
Keratin	ENSGALG00000024188		20482795
Keratin	ENSGALG00000024189		20482795
Keratin	ENSGALG00000024190		20482795
Keratin	ENSGALG00000024191		20482795
Keratin	ENSGALG00000024192		20482795
Keratin	ENSGALG00000024194		20482795
Keratin	ENSGALG00000024197		20482795
Keratin	ENSGALG00000024198		20482795
Keratin	ENSGALG00000024200		20482795
Keratin	ENSGALG00000024201		20482795
Keratin	ENSGALG00000024202		20482795
Keratin	ENSGALG00000024205		20482795
Keratin	ENSGALG00000024206		20482795
Keratin	ENSGALG00000024208		20482795
Keratin	ENSGALG00000024209		20482795
Keratin	ENSGALG00000024212		20482795
Keratin	ENSGALG00000024215		20482795
Keratin	ENSGALG00000024216		20482795
Keratin	ENSGALG00000024218		20482795
Keratin	ENSGALG00000024219		20482795
Keratin	ENSGALG00000024221		20482795
Keratin	ENSGALG00000024222		20482795
Keratin	ENSGALG00000024225		20482795
Keratin	ENSGALG00000024226		20482795
Keratin	ENSGALG00000024228		20482795
Keratin	ENSGALG00000024229		20482795
Keratin	ENSGALG00000024230		20482795
Keratin	ENSGALG00000024233		20482795
Keratin	ENSGALG00000024236		20482795
Keratin	ENSGALG00000024237		20482795
Keratin	ENSGALG00000024238		20482795
Keratin	ENSGALG00000024239		20482795
Keratin	ENSGALG00000024240		20482795
Keratin	ENSGALG00000024241		20482795

Keratin	ENSGALG00000024242		20482795
Keratin	ENSGALG00000024245		20482795
Keratin	ENSGALG00000024246		20482795
Keratin	ENSGALG00000024247		20482795

Gene ID provides Ensemble ID numbers for chicken. PMID, Pubmed ID, for source publication. To construct our candidate gene set for feather-related genes, we conducted a literature search in Web of Science through the search engine in Endnote X5, using “feather AND gene*” as keywords, with no constraints on year of publication. We conducted our last such systematic search on February 18, 2013, yielding approximately 400 abstracts, but added additional abstracts as recently as November 2013. We recorded genes for which a mutation is known to cause a feather phenotype. We also included genes where the pattern of expression was spatially restricted to placodes, feather buds, or feathers during development.

Supplementary Table 2 – Feather-related genes showing evidence of positive selection on the branch leading to birds.

Gene ID	Gene Symbol	omega class 2 (dn/ds)	ln(null model)	ln(allowing positive selection)	Test statistic (D)	P-value
ENSGALG00000002375	NOTCH1	17.7	-60953.0571	-60942.8249	20.4645	<8E-4
ENSGALG00000002552	COL3A1	ds~=0	-30962.3467	-30953.8837	16.9259	<5E-3
ENSGALG00000016941	PCDH17	7	-17124.3081	-17052.9813	142.6535	<1E-30

The test statistic D is twice the difference in log likelihood of the two models tested. P-value is adjusted for Bonferroni correction. Omega (ω) indicates the signal for adaptive evolution on the branch leading to birds relative to the null model $\omega = 1$.

Supplementary Table 3 - 32 Genes shared between feather and hair gene sets.

Gene ID	Gene Symbol
ENSGALG00000002375	Notch1
ENSGALG00000002922	Notch2
ENSGALG00000002947	Msx2
ENSGALG00000003310	Slc45a2
ENSGALG00000003446	Prlr
ENSGALG00000003654	Foxn1
ENSGALG00000004274	Twist2
ENSGALG00000004481	Eda
ENSGALG00000004885	Slc24a5
ENSGALG00000006379	Shh
ENSGALG00000007130	Rac1
ENSGALG00000007145	Itgb1
ENSGALG00000007280	Cldn1
ENSGALG00000007679	Mitf
ENSGALG00000007870	Smad3
ENSGALG00000011206	Kitlg
ENSGALG00000011905	Ctnnb1
ENSGALG00000012155	Egf
ENSGALG00000012363	Egfr
ENSGALG00000012620	Ptc1
ENSGALG00000013925	Kit
ENSGALG00000014369	Edaradd
ENSGALG00000014634	Pmel
ENSGALG00000015205	Tyrp1
ENSGALG00000016308	Myc
ENSGALG00000016427	Adam17
ENSGALG00000016600	Pomc
ENSGALG00000016809	Edar
ENSGALG00000017119	Tnfrsf19
ENSGALG00000017237	Tyr
ENSGALG00000021455	Asip
ENSGALG00000023293	Foxe1

Hair gene set from Lowe et al. (2011).

Supplementary Table 4 - Details of 23 regions of the chicken genome enriched for CNEEs arising on the branch leading to birds.

chromosome	position	CNEEs/Mb	Gene upstream	Gene downstream	Gene prediction upstream (if closer)	Gene prediction downstream (if closer)
chr7	24496029	160.68	TNS1	IGFBP2	_	IGFBP5
chr13	2187175	143.48	DNAJC18	PAIP2	_	SLC23A1
chr13	1725780	138.48	HDAC3	PCDH1	STARD10	_
chr15	10772514	122.29	DGCR6	SF3A1	RTN4R	OSBP2
chr15	10596727	119.21	DGCR6	SF3A1	PRODH	RTN4R
chr9	15448741	110.90	TP63	LPP	_	_
chr9	14989167	103.35	OSTN	TMEM207	IL1RAP	_
chr6	23553695	96.69	ATP6V0E2	GOT1	SLC25A28	_
chr14	12900949	94.14	AXIN1	CLCN7	_	LUC7L
chr8	20835190	90.99	HYI	ST3GAL3	PTPRF	_
chrZ	7694631	88.22	CNTFR	IL11RA	_	_
chr19	4220049	86.95	YWHAG	HSPB1	_	_
chr4	92222023	84.44	AVP	HTR7	DDRGK1	_
chr9	14259981	84.09	ATP13A4	FGF12	HRASLS	MB21D2
chr5	29829705	82.61	COX16	SLC8A3	_	_
chr17	3783301	79.43	TNC	TLR4	TRIM32	ASTN2
chr17	3932933	79.38	TNC	TLR4	TRIM32	ASTN2
chr12	2017381	79.16	CISH	MST1	DOCK3	MANF
chr17	4269608	78.57	TLR4	BRINP1	_	_
chr7	27109460	77.65	MRAS	GLI2	RALB	_
chr18	1148963	77.32	MYOCD	NDEL1	DNAH9	SHISA6
chr19	3930149	76.73	CUX1	PRKRIP1	_	SH2B2
chr8	5404210	76.17	LMX1A	PBX1	_	_

References:

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.

Blanchette, M, WJ Kent, C Riemer, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Research 14:708-715.

Ellegren, H, L Smeds, R Burri, et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756-760.

Huang, YH, YR Li, DW Burt, et al. 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nature Genetics 45:776-+.

Hubbard, T, D Barker, E Birney, et al. 2002. The Ensembl genome database project. Nucleic Acids Research 30:38-41.

Kent, WJ, R Baertsch, A Hinrichs, W Miller, D Haussler. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proceedings of the National Academy of Sciences of the United States of America 100:11484-11489.

Lowe, CB, G Bejerano, D Haussler. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proceedings of the National Academy of Sciences of the United States of America 104:8005-8010.

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.
Pruitt, KD, T Tatusova, DR Maglott. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 33:D501-504.
Qu, YH, HW Zhao, NJ Han, et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nature Communications 4.
Rands, CM, A Darling, M Fujita, et al. 2013. Insights into the evolution of Darwin's finches from comparative analysis of the Geospiza magnirostris genome sequence. Bmc Genomics 14.
Schwartz, S, WJ Kent, A Smit, Z Zhang, R Baertsch, RC Hardison, D Haussler, W Miller. 2003. Human-mouse alignments with BLASTZ. Genome Research 13:103-107.
Siepel, A, D Haussler. 2004. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. Molecular Biology and Evolution 21:468-488.
Zhan, XJ, SK Pan, JY Wang, et al. 2013. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. Nature Genetics 45:563-U142.
Zhu, J, JZ Sanborn, M Diekhans, CB Lowe, TH Pringle, D Haussler. 2007. Comparative genomics search for losses of long-established genes on the human lineage. Plos Computational Biology 3:e247-e247.