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

1Department of Developmental Biology and Howard Hughes Medical Institute, Stanford University, Stanford, CA
2Department of Geological Sciences, University of Texas at Austin, Austin TX
3Department of Natural History, Royal Ontario Museum, Toronto, and Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada (Deceased)
4Center for Biomolecular Science and Engineering, University of California, Santa Cruz
5Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA

*Corresponding author: sedwards@fas.harvard.edu.
Associate editor: Katja Nowick

Abstract

The evolution of avian feathers has 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 and 100% of the nonkeratin feather gene set were present prior to the origin of Dinosauria. On the branch leading to modern birds, we detect a strong signal of regulatory innovation near insulin-like growth factor binding protein (IGFBP) 2 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.

Key words: enhancer, gene regulation, comparative genomics, integument, body size, dinosaur.

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 15 years indicate a more ancient origin; filamentous feather precursors are now known to be present in many lineages of nonavian 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 nonexonic 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

CNEEs 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 methods and supplementary table S1, Supplementary Material online). 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 approximately...
71 Mb and spanning approximately 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, 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 nonkeratin genes and analyzed these groups separately.

**Fig. 1.** Feather development genes are ancient whereas associated CNEEs peak in the amniote ancestor. Evolutionary dynamics of (a) nonkeratin 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 figure 2, with spacing proportional to divergence times as recorded in timetree.org (Hedges et al. 2006). In (b), the larger peak comprised 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).

An Ancient Genic Toolkit and Extended Regulatory Evolution Are Associated with Feather Origins

The genic and regulatory components of the keratin and nonkeratin sets show very different patterns across the 500-My backbone of our tree, on the lineage leading from the common ancestor of vertebrates to the chicken (figs. 1 and 2, supplementary fig. S1, Supplementary Material online). 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 nonkeratin feather genes (1.7 times expected), with smaller numbers of nonkeratin feather genes arising on branches leading to tetrapods and less inclusive clades (figs. 1a and 2). No members of this nonkeratin 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 nonkeratin 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 fig. S2, Supplementary Material online). The inferred first appearance of nonkeratin 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 nonkeratin 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. S1, Supplementary Material online). The rate of origination for these CNEEs is greater than what would be expected from CNEEs uniformly distributed throughout the genome for six of the eight branches along the lineage leading to chicken, suggesting a large amount of regulatory innovation over an extended time period (figs. 1 and 2). Thus, the nonkeratin 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 S3, Supplementary Material online) and 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 nonavian dinosaurs, as part of Archosauromorpha, possessed the entirety of the known nonkeratin protein-coding toolkit for making feathers. Moreover, assuming a constant rate of genome-wide accumulation of CNEEs throughout vertebrates, we estimate that 86% of nonkeratin 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 genetic and regulatory complement identified in the ancestral archosaur was either a flexible toolkit co-opted 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 My 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 \( \beta \)-keratin genes within Archosauria (Greenwold and Sawyer 2010; Li et al. 2013) on the branch leading to birds (figs. 1b and 2). The larger peak for keratin innovation comprised 57 \( \beta \)-keratins arising as an expansion of a gene cluster on chicken chromosome 27 and 5 \( \beta \)-keratins from duplications on chromosome 2. The small peak in the turtle-bird ancestor is due to the expansion of a \( \beta \)-keratin gene cluster on chromosome 25. Both of these results are consistent with previous studies of \( \beta \)-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 \( \beta \)-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 on the branch leading to birds, suggesting that...
regulatory evolution near \( \beta \)-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 \( \beta \)-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 \( \beta \)-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 nonkeratin feather genes showed signatures of positive selection on the archosaurian branch leading to birds (\textit{supplementary} \textit{table} \textit{S2}, \textit{Supplementary Material} online). These results indicate that most nonkeratin genes related to feather development exhibit regulatory, not protein-coding, innovations in the avian stem lineage, including living birds and nonavian dinosaurs, consistent with the hypothesis that regulatory innovations underlie adaptations in skin patterning and feather morphology.

\textbf{Body Size Genes Exhibit Exceptional Regulatory Innovation in Dinosaurs}

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; \textit{supplementary} \textit{table} \textit{S4}, \textit{Supplementary Material} online). 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\% more than expected (\( P < 10^{-58} \)), 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 offers the potential to link genome evolution to key shifts in shape and form occurring in deep time.

\textbf{Supplementary Material}

\textit{Supplementary} \textit{materials} and \textit{methods}, \textit{tables} \textit{S1}–\textit{S4}, and \textit{figures} \textit{S1} and \textit{S2} are available at \textit{Molecular Biology and Evolution} online (http://www.mbe.oxfordjournals.org/).

\textbf{Acknowledgments}

The authors 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 Shchipilina helped collect Ensemble IDs. C.B.L. and D.H. were supported by the Howard Hughes Medical Institute. A.J.B. acknowledges support from NSERC grant 200-12. S.V.E. and J.C. acknowledge support from the US National Science Foundation (EAR-1355343/DEB-1355292). D.H., C.B.L., and S.V.E. conceived the study; C.B.L. and S.V.E. collected and analyzed data; C.B.L., S.V.E., and J.C. wrote the paper, with comments from all other authors. All alignments and genomic coordinates of genes and CNEEs used in the analysis can be found at http://fhgdevd-lowecesse.ucsc.edu/. The authors declare no competing financial interests.

\textbf{References}


