Regulatory Divergence among Beta-Keratin Genes during Bird Evolution

Maloyjo Joyraj Bhattacharjee,[1] Chun-Ping Yu,[1] Jinn-Jy Lin,[1,2,3] Chen Siang Ng,[1,3] Tzi-Yuan Wang,[1] Hsin-Hung Lin,[1] and Wen-Hsiung Li*[,1,4,5]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2]Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Sciences, Academia Sinica, Nankang, Taipei, Taiwan

[3]Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

[4]Center for the Integrative and Evolutionary Galliformes Genomics (iEGG Center), National Chung Hsing University, Taichung, Taiwan

[5]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois

***Corresponding author:** E-mail: whli@sinica.edu.tw.

Associate editor: Gunter Wagner

The RNA-seq data used in this study are available at National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Bioproject: PRJNA245063.

Abstract

Feathers, which are mainly composed of α- and β-keratins, are highly diversified, largely owing to duplication and diversification of β-keratin genes during bird evolution. However, little is known about the regulatory changes that contributed to the expressional diversification of β-keratin genes. To address this issue, we studied transcriptomes from five different parts of chicken contour and flight feathers. From these transcriptomes we inferred β-keratin enriched co-expression modules of genes and predicted transcription factors (TFs) of β-keratin genes. In total, we predicted 262 TF–target gene relationships in which 56 TFs regulate 91 β-keratin genes; we validated 14 of them by in vitro tests. A dual criterion of TF enrichment and "TF–target gene" expression correlation identified 26 TFs as the major regulators of β-keratin genes. According to our predictions, the ancestral scale and claw β-keratin genes have common and unique regulators, whereas most feather β-keratin genes show chromosome-wise regulation, distinct from scale and claw β-keratin genes. Thus, after expansion from the β-keratin gene on Chr7 to other chromosomes, which still shares a TF with scale and claw β-keratin genes, most feather β-keratin genes have recruited distinct or chromosome-specific regulators. Moreover, our data showed correlated gene expression profiles, positive or negative, between predicted TFs and their target genes over the five studied feather regions. Therefore, regulatory divergences among feather β-keratin genes have contributed to structural differences among different parts of feathers. Our study sheds light on how feather β-keratin genes have diverged in regulation from scale and claw β-keratin genes and among themselves.

Key words: **feather evolution, keratin genes, regulatory evolution, cis elements.**

Introduction

Feather was an extraordinary innovation in dinosaurs but has become highly diversified in birds, evolving into various forms for protection, camouflage, heat retention, mate attraction and flight (Pettingill and Breckenridge 2012). Feather diversification is presumed to be key to the success of modern birds, allowing them to adapt and radiate into various ecological niches. It is the most complex integumentary structure found in the avian lineage of vertebrates (Prum 1999; Prum and Brush 2002, 2003; Pabisch et al. 2010; Weiss et al. 2011) and a premier example of a complex evolutionary novelty.

In vertebrates all epidermal appendages, including the feather, are formed mostly by alpha (α) and beta (β) keratin proteins (Alibardi and Toni 2008; Alibardi et al. 2009; Wu et al. 2015). Both types of keratin genes have been expanded by gene duplication and functional diversification during vertebrate evolution. The α-keratins are associated with hair, nail, horns, and hoofs (Vandebergh and Bossuyt 2012). The β-keratins emerged in reptiles and underwent expansion in chelonians and birds to form structures such as scale, claw and shell of turtles and scale, claw, beak and feather of birds. In particular, the avian lineage underwent many β-keratin gene duplications, contributing to feather diversification (Dalla Valle, Nardi, Belvedere, et al. 2007; Dalla Valle,Nardi, Toffolo, et al. 2007; Greenwold and Sawyer 2010, 2013; Li et al. 2013; ZhaNg et al. 2014).

The β-keratins in birds are divided into scale, claw, keratinocyte, and feather based on sequence heterogeneity and tissue specific expression (Presland et al. 1989; Li et al. 2013; Greenwold et al. 2014). It has been hypothesized that feather β-keratins evolved from reptilian scale β-keratins by losing glycine and tyrosine rich tail-moieties (Regal 1975; Zhang and Zhou 2000; Greenwold and Sawyer 2010). This conferred feather structures more flexibility compared with scales and claws in birds. It is reasonable to assume from fossil

Article

Fast Track

discoveries that some ancient types of feather β-keratin genes were already present in feathered dinosaurs (Xu et al. 2012, 2015). Later, more complex feather β-keratins genes evolved in birds, leading to feather diversification for various functions such as thermoregulation, and flight. Greenwold et al. (2014) studied 48 bird genomes and found that in most bird species, scale, claw and keratinocyte β-keratin genes were localized in Chr25, whereas most feather β-keratin genes were localized in Chr25 and Chr27 along with few others in Chr1, Chr2, Chr6, Chr7, and Chr10. Ng et al. (2014) proposed that the diversity of feather is partly due to differential expression of feather β-keratin genes in different feather structures.

While much research has been conducted on the duplication and functional diversification of β-keratin genes, little has been done on the regulatory network that controls the expression of different classes (claw, scale, feather and keratinocyte) of β-keratin genes and on the differential expression of feather β-keratin genes in different feather parts.

A transcription factor (TF) binds to specific DNA sequences (TF binding sites, abbreviated as TFBSs) in the promoter of a gene, thereby controlling the expression timing and level of the gene. TF–TFBS pairs therefore are important elements of a gene regulatory network. The purpose of this study is to identify the TFBSs in β-keratin genes and their cognate TFs and to understand the common and unique regulation of different classes of β-keratin genes. For this purpose, we shall reanalyze the transcriptomes obtained previously from different parts of developing contour feather and flight feather of chicken (Ng et al. 2014). These transcriptomes of feather development are excellent materials for inferring strongly co-expressed genes with potentially common regulatory basis and for predicting TF–TFBS pairs (Yu et al. 2015).

Results and Discussion

Gene Co-Expression Modules

Supplementary figure S1, Supplementary Material online, shows the different parts of body contour feather and flight feather from which the transcriptomes (RNA-seq data) were obtained previously (Ng et al. 2014). The two parts of contour feather were defined as early-body (EB) and late-body (LB), whereas the three parts of flight feather were defined by early-flight (EF), middle-flight (MF) and late-flight (LF). Three biological replicates were obtained for each of these five feather parts, so that in total 15 transcriptomes were obtained. A gene is considered expressed if its FPKM (fragments per kilobase of exon per million mapped reads) is ≥ 1 in at least 1 of the 15 transcriptomes. In total, 11,533 genes, including 130 β-keratin genes, were expressed (supplementary table S1, Supplementary Material online). In terms of the sub-classification of β-keratins (Greenwold et al. 2014), the 130 expressed β-keratin genes, which represent ~87% of the β-keratin genes annotated in Ng et al. (2014), include 18 scale, 11 claw, 15 keratinocyte, and 86 feather β-keratin genes.

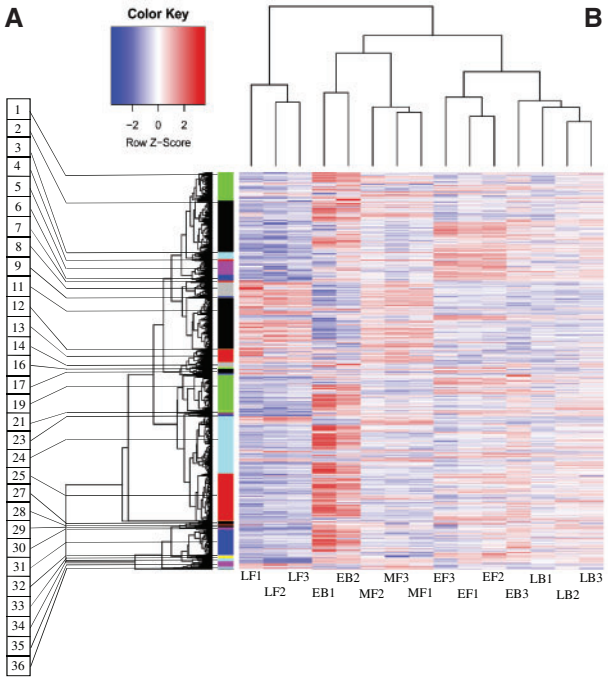
For the expressed genes (11,533), the Pearson correlation coefficient (PCC) in expression level is high (PCC > 0.97) between the biological replicates except for LB (PCC $= 0.93$, 0.91, 0.76) and EF (PCC $= 0.96$, 0.95, 0.86) (supplementary fig. S2, Supplementary Material online, above the diagonal). This implies that the LB and EF replicates are more heterogeneous. Between the conditions, EB and EF have a high mean PCC (0.96 ± 0.02), and so do MF and LF (mean PCC $= 0.97 \pm 0.01$). By applying the criterion of PCC ≥ 0.8 in expression level between β-keratin and other genes (supplementary table S2, Supplementary Material online), we extracted from the 11,533 expressed genes a subset of 2314 genes (supplementary table S3, Supplementary Material online). This subset includes 128 β-keratin genes and 2,186 nonkeratin genes (only 2 expressed β-keratin genes did not satisfy PCC ≥ 0.8). The PCC pattern for the subset of 2,314 genes (supplementary fig. S2, Supplementary Material online, below the diagonal) is consistent with that for the 11,533 expressed genes.

By hierarchical clustering, we classified the 2,314 genes in the above subset into 47 clusters (fig. 1A, see also supplementary fig. S3 and table S4, Supplementary Material online). To increase the probability of co-expressed genes being co-regulated, we added the condition that the genes in a cluster share the same Gene Ontology (GO) term, so that they are likely involved in the same biological function. The GO functional enrichment analysis showed that 20 (ID: 1, 2, 5, 7, 9, 17, 21 23–25, 27–36) of the 47 clusters, where mostly β-keratin genes are present, are significantly enriched in functional biological terms (fig. 1B, supplementary table S5, Supplementary Material online). Note that the majority of the terms enriched are related to keratin-based cellular structural processes such as cytoskeleton, intermediate filament, and keratin (Kreplak et al. 2004; Alibardi et al. 2009; Herrmann et al. 2009; Lee et al. 2012). Furthermore, 10 other clusters (ID: 3, 4, 6, 8, 11–14, 16, 19), which lack β-keratin genes, also showed enriched functional terms. Therefore, we consider the 30 ($20 + 10$) significant clusters as co-expressed modules, suitable for inferring regulatory elements and annotation of regulatory network of genes, particularly for β-keratin genes.

Predicted TF–TFBS Pairs of β-Keratin Genes and Inferred TF–Target Gene Relationships

The TF–TFBS pairs of 496 TFs of chicken in the CIS-BP database (cisbp.ccbr.utoronto.ca/, last assessed 14 April 2016) (Weirauch et al. 2014) served as a valuable resource to scan and select enriched TFBSs (P value < 0.0001, FDR < 0.00001) in each of the co-expression module (supplementary fig. S4, Supplementary Material online). To increase the probability of a selected TF being true, we added the condition that the gene coding for the TF recognizing the motif and the gene in which the TFBS is overrepresented have a high PCC (PCC $> |0.8|$). This condition helped us to identify the TF–target gene relationship (summarized in fig. 2A and detailed in supplementary table S6, Supplementary Material online). Overall, we identified 13,624 pTF–target gene pairs (full data not shown), in which 213 TFs (pTFs) putatively regulate 1857 genes by positive or negative interactions. In terms of β-keratin genes, we found 372 TF–target gene pairs in which 81 TFs putatively regulate 91 β-keratin genes, which may be further categorized as: for scale, 13 putative TFs (pTFs)/

A



B

	Cluster ID	Enriched keratin related GO terms	Other enriched GO terms
β-keratin related clusters	1	Yes	Wnt signaling pathway, developmental protein, differentiation
	2	Yes	Catalytic activity, hydrolase activity, mitochondrion, protein stabilization
	5	No	Endodermal cell differentiation, ion transport, signal transducer activity
	7	No	Hydrolase activity, acting on carbon-nitrogen bonds, transmembrane signaling
	9	No	Circadian rhythm, response to endoplasmic reticulum stress, chemotaxis
	17	Yes	NA
	21	Yes	Microtubule-based process, phosphatidylinositol-mediated signaling, pigmentation
	23	Yes	Protein stabilization, hydrolase activity, acting on carbon-nitrogen bond
	24	No	Intracellular protein transport, neuron apoptotic process, cytoplasm
	25	Yes	NA
	27	Yes	NA
	28	Yes	NA
	29	Yes	NA
	30	Yes	NA
	31	No	Extracellular vesicular exosome, fatty acid metabolic process, translation
	32	Yes	NA
	33	Yes	NA
	34	Yes	NA
	35	No	Poly(A) RNA binding, extracellular vesicular exosome
	36	Yes	NA
Other clusters	3	NA	Transmembrane transport, sensory perception of pain
	4		Proteinaceous extracellular matrix, signal transduction, transaminase activity
	6		Membrane, exocytosis, synaptic vesicle, G-protein coupled receptor activity
	8		Signal transducer activity, immune system process, locomotion
	11		Cell adhesion, nucleolus, poly(A) RNA binding, heparin binding
	12		Protein dimerization activity, organelle, nucleus
	13		Potassium channel activity, ion channel activity
	14		Immune response, determination of left/right symmetry, immune system process
	16		Calcium ion import, phosphatidylinositol-mediated signaling
	19		Cofactor metabolic process, cell death, DNA helicase activity

Fig. 1. Hierarchical clustering and heatmap of gene-expression level and enriched functional categories in coexpression clusters. (*A*) The clustering of genes is based on expression similarity of the 2314 sorted genes in different feather parts. The color bars on the left side denote the clusters obtained after cutting the hierarchical cluster tree at specific height (see supplementary fig. S2, Supplementary Material online). The heatmap represents expression pattern of the gene in each cluster. High expression is shown in red and low expression is shown in blue (see the color bar at the left top). The entries below the heatmap denote transcriptome ID as given in figure S1, Supplementary Material online, and are explained in the text. The clusters which have significantly enriched GO terms are numbered in the figure. (*B*) The entries in the table represent the enriched functions corresponding to the significant clusters. The clusters are categorized as "β-keratin related" and "others". For keratin genes, the child GO terms such as "feather", "scale", "claw", "keratinocytes", and "intermediate filament" are assigned to its parental GO term "keratin". "NA" denotes "not available".

regulate 17 target genes; for claw, 7 pTFs/9 target genes; for keratinocytes, 18 pTFs/9 target genes; and for feather, 63 pTFs/56 target genes. Interestingly, feather β-keratins were found regulated by many repressors. The Venn diagram depicts that each of the β-keratin sub-classes (scale, claw, keratinocytes and feather) have common as well as unique regulators (fig. 2*A*). As the β-keratin genes expanded in birds by gene duplication, the common and unique regulators found herein may be useful to unveil their regulatory transition, as discussed below.

β-Keratin Genes in Turkey and Evolutionary Conservation of Predicted TFBS

We further consider the condition that the TFBS present in the promoter region of a gene has been conserved in the promoter region of its orthologue in Turkey. For this purpose, we scanned the genome of Turkey (release 5.0) using chicken β-keratin genes as queries. This approach identified 142 β-keratin genes in Turkey, which includes 14 scale, 12 claw, 2 keratinocyte and 114 feather β-keratin genes (supplementary table S7 and fig. S5, Supplementary Material online). Thus, chicken and Turkey have similar numbers of β-keratin genes. Unfortunately, only ~70 of the annotated Turkey β-keratin genes could be assessed for upstream 1,000 bp and another 10 genes for upstream 200-bp promoter sequences from the Turkey genome. We found 192 predicted TF–target gene pairs (out of 372 pairs) satisfied the conservation criterion (supplementary table S6 and fig. S5, Supplementary Material online). Importantly, the TFBS found in the promoter of a particular sub-class of β-keratin (scale, claw, keratinocyte and feather) in chicken was conserved on the promoter of the corresponding sub-class in Turkey β-keratin genes. This supported the predicted TF–TFBS relationships of β-keratin genes in chicken.

Regulatory Transition of β-Keratin Genes

The pTF–target gene relationships of different sub-classes of β-keratin genes are presented in figure 3*A* and sorted from supplementary tables S6–S8, Supplementary Material online. We also did maximum likelihood phylogenetic reconstruction of β-keratins to understand the regulatory transition associated with the evolution of β-keratin genes (fig. 3*B*, and see supplementary fig. S6, Supplementary Material online, for an expanded phylogeny). The phylogeny is concordant with previous studies with scale and claw β-keratins being the basal clade of feather β-keratins (Greenwold et al. 2014; Zhag et al. 2014).

Figure 3*A* shows that the basal scale and claw β-keratin genes have common as well as unique regulators. Specific TFs of homeodomain, SMAD, T-box, AP-2, Sox, nuclear receptor, and C2H2 ZF family uniquely regulate scale β-keratin genes (supplementary table S8, Supplementary Material

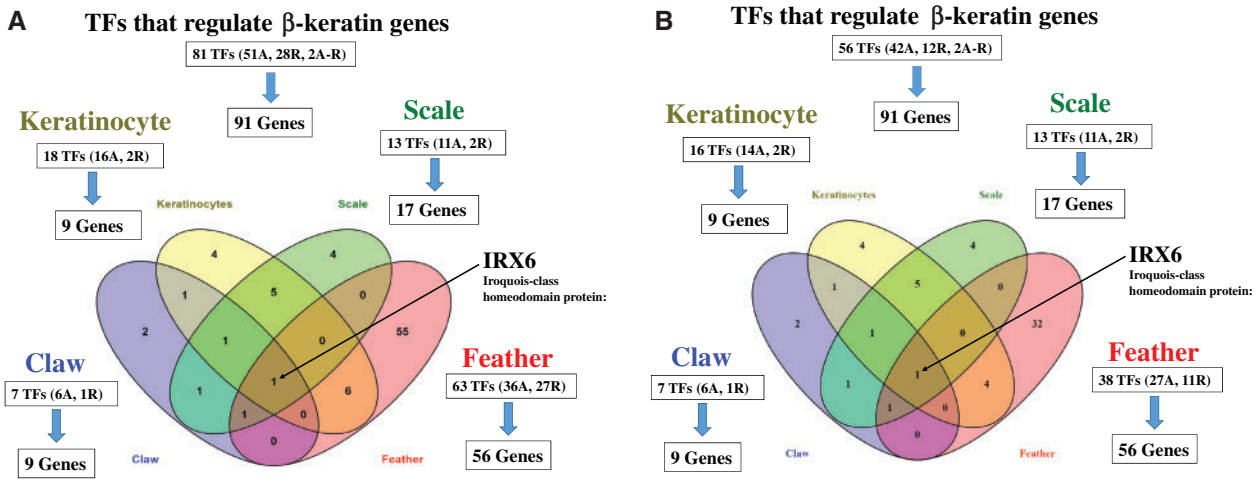


Fig. 2. Summary of annotated TF–target gene relationships. (*A*) Overall predictions and (*B*) predictions supported by differential expression of pTF–target gene. The figure summarizes the number of annotated TFs and their corresponding numbers of target β-keratin genes. The annotated TFs are further classified according to the sub-class (scale, claw, keratinocytes and feather) of the β-keratin genes they regulate. The entries "A" and "R" denotes "activator" and "repressor", respectively. The Venn-diagram shows the numbers of common and unique regulators among the four sub-classes of β-keratins. Iroquois-class homeodomain (IRX6) was found the sole TF protein which commonly regulates all the four sub-classes of β-keratin genes.

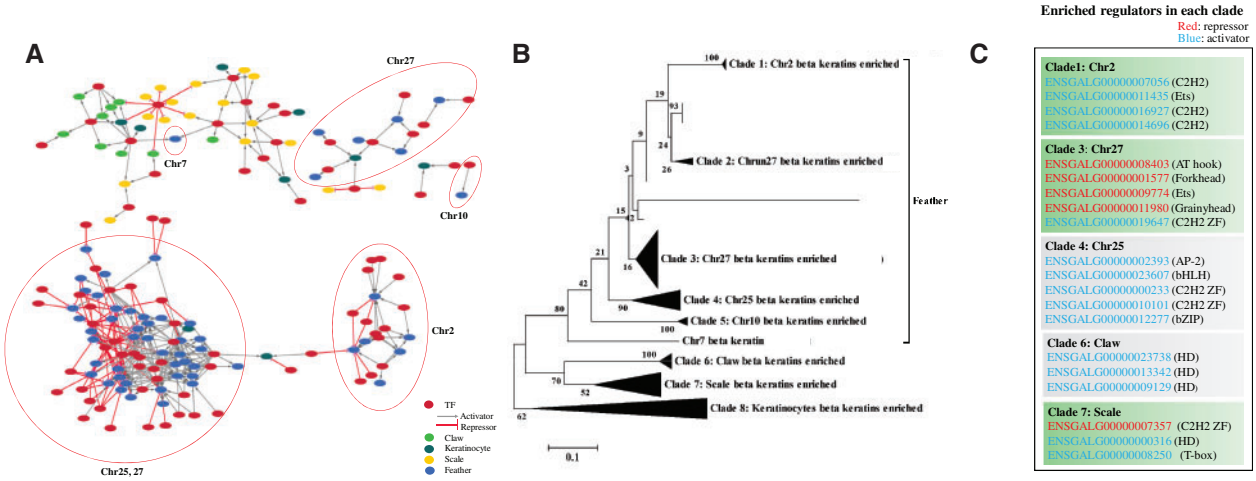


Fig. 3. Regulator–target gene relationships, phylogenetic tree of β-keratin genes and chromosome/sub-class-wise enriched regulators of β-keratin genes. (*A*) Regulators (TF's) and their target genes. A specific color is used to denote a sub-class of target genes (scale, claw, feather or keratinocytes). TFs are denoted by red color, and activators (gray arrows) and repressors (red T shapes) are shown in conventional notations. Most feather β-keratin genes have chromosome-wise specific regulators and only the Chr7 β-keratin gene shares a regulator with scale and claw β-keratin genes. (*B*) Maximum likelihood phylogeny of the annotated β-keratin genes. Scale and claw β-keratin genes are basal to feather β-keratin genes, whereas the Chr7 β-keratin gene is basal to all feather β-keratin genes. The clustering pattern of β-keratin genes showed consistency with the regulation pattern. (*C*) Enriched regulators of each clade of β-keratin genes identified by phylogenetic analysis. The enrichment is based on finding overrepresented regulators on each clade and is done to infer the major regulators of each clade that regulates multiple β-keratin genes.

online). Similarly, specific TFs of the homeodomain family uniquely regulate claw β-keratin genes. A few TFs of homeodomain (ENSGALG00000023738, ENSGALG00000023195) and C2H2 ZF (ENSGALG00000016016, ENSGALG00000007357) commonly regulate scale and claw β-keratin genes. Among the above TFs, two are repressors, one being common (ENSGALG00000007357, C2H2 ZF family) and the other (ENSGALG00000003759, nuclear receptor family) unique for scale. As in Greenwold et al. (2014), Chr7 feather β-keratin ENSGALG00000006133 (named as Chr7 in fig. 3*A*) is basal of all feather β-keratins (fig. 3*B*). Notably, this gene is regulated by two TFs of the homeodomain family (ENSGALG00000023738, ENSGALG00000023195) common with ancestral scale and claw β-keratin genes. ENSGALG00000023738 and ENSGALG00000023195 are members of the Iroquois (*IRX-6*) and Distal-less (*DLX-6*) homeobox 6 protein families, respectively, both having important roles in pattern formation, limb development and appendage growth in vertebrates (Ogura et al. 2001; Velinov et al. 2012). ENSGALG00000007357 and ENSGALG00000003759 are B-cell lymphoma 6 (*BCL6*) and RAR-related orphan receptor A (*RORALPHA1*) protein,

respectively. *BCL6* has been reported to function as a repressor for germinal center formation and influences apoptosis (Huynh et al. 2000). *RORALPHA1* also has been reported as a repressor and is involved in organogenesis, differentiation, and circadian rhythm in vertebrates (Xiao et al. 2015). This implies that the TFs taking part in structural framework of ancestral scale and claw as well as proto-feathers may have been recruited from related networks regulating pattern formation, limb development, appendage growth, and organogenesis.

The remaining feather β-keratin genes showed consistent genomic organization, clustering pattern, and regulation distinct from the ancestral scale and claw β-keratin genes. For example, all Chr2 β-keratins are clustered together as clade-1 and uniquely regulated by TFs of eight families (Ets, E2F, C2H2 ZF, nuclear receptor, bZIP, bHLH, homeodomain-POU, and forkhead). The two biggest loci, Chr27 and 25 β-keratin genes formed two (clade-2 and 3) and one clade (clade-4), respectively. Both Chr27 and Chr25 β-keratin genes have common as well as unique regulators. Specific TFs of six families (bHLH, RFX, homeodomain, C2H2 ZF, bZIP, and forkhead) commonly regulate Chr25 and Chr27 β-keratin genes. TFs of five families (bHLH, homeodomain, AP-2, E2F, and C2H2 ZF) uniquely regulate Chr25 β-keratin genes. TFs of 15 families (ARID/BRIGHT, Ets, Rel, C2H2 ZF, bZIP, grainyhead, AT hook, homeodomain, bHLH, HSF, TBP, SMAD, nuclear receptor, DP,E2F, and forkhead) uniquely regulate Chr27 β-keratin genes. Notably, many of the Chr27 unique regulators are repressors. All Chr10 β-keratin genes were clustered as clade-5, although our approach annotated only one Chr10 β-keratin gene being regulated by specific TFs of two families (Ets and bHLH). Lastly, keratinocyte β-keratin genes that together were clustered as clade-8 are found having springy regulation, common with scale, claw and feather β-keratin genes. This is consistent with their diverse roles related to scale, claw and feather (Greenwold et al. 2014). It is important to note that the Animal Transcription factor database annotated at least 858 chicken genes as TFs (Zhang et al. 2012), whereas the CIS-BP database (Weirauch et al. 2014) has known or predicted TF–TFBS relationships for 496 TFs of chicken, with paucity of TF–TFBS information for the remaining TFs. Fortunately, the TF–target gene relationships we have obtained include representatives of all genomic loci (except one feather β-keratin of Chr1 and one jun-transformed β-keratin of Chr6) and sub-classes for understanding the characteristic regulation of β-keratin genes.

The typical regulation of β-keratin genes observed above can be explained by the following innovative regulatory transition events. First, the transition from scale to claw or vice versa involved both retention and gain/loss of TFBSs. The motifs corresponding to the common TFs are the cases of retention; notably, the common repressor motif for ENSGALG00000007357 (C2H2 ZF family) might be important to repress expression of scale and claw β-keratin genes in feather. The unique repressor motif of scale β-keratins corresponding to ENSGALG00000003759 (nuclear receptor family) might be a gain or loss event, depending on whether the scale or claw β-keratins were ancestral, and might repress scale β-keratin gene expression in claw. Second, the Chr7 β-keratin gene, which is the first locus where feather β-keratin evolved from a scale or claw β-keratin gene, retains the ancestral cis motifs for ENSGALG00000023738 and ENSGALG00000023195. Third, most of the feather β-keratin genes were expanded from the basal Chr7 feather β-keratin gene and lost motifs recognized by ancestral scale and claw TFs and gained new motifs to perform new function. In summary, feather β-keratin genes underwent chromosome-wise expansions and recruited chromosome-wise regulators. Similar to the evolution of claw and scale β-keratin genes, the transition of feather β-keratin genes from Chr25 to Chr27 or the reverse involved both retention and gain/loss of TFBSs, as they showed common and unique regulation. The characteristic regulatory control of feather β-keratin genes observed here may be essential to effect their mosaic expression in diverse feather parts as has been suggested in a previous study (Ng et al. 2014).

Major Regulators of β-Keratin Genes

By a major regulator we mean a TF that regulates multiple β-keratin genes. Following the annotation of β-keratin regulators, we inferred the major regulators of β-keratin genes in two ways.

Major TFs Predicted by Enrichment Analysis

First, we identified over-represented regulators in each clade of β-keratin genes (fig. 3C, also see supplementary table S9, Supplementary Material online) and found that the clades of Chr2, Chr27, Chr25, claw, and scale β-keratin genes have significantly enriched TFs (P value < 0.001, FDR < 0.0001); clades 2 (having few Chr27 β-keratins) and 5 (having few Chr10 β-keratins), being very small, showed no over-represented TFs. Clade 8, in which all keratinocyte β-keratin genes are clustered, also lacks over-represented TFs as keratinocyte β-keratin genes are regulated in small clusters together with scale, claw and feather β-keratin genes. In details, among the 15 annotated TFs that regulate Chr2 β-keratin genes (supplementary table S8, Supplementary Material online), 4 specific TFs (all activators) of 2 families (C2H2 and Ets) were over-represented in the Chr2 clade. Similarly, five specific TFs (four repressors and one activator) of five families (AT hook, forkhead, Ets, grainyhead and C2H2 ZF) and five specific TFs (all activators) of four families (AP-2, bHLH, C2H2 ZF and bZIP) were over-represented in the Chr27 and Chr25 clades, respectively. As mentioned earlier, Chr25 and Chr27 β-keratin genes have common as well as unique TFs, and we found that the five over-represented TFs of the Chr25 clade may be categorized as 2 unique (ENSGALG00000023607 and ENSGALG00000002393) and 3 (ENSGALG00000000233, ENSGALG00000010101 and ENSGALG00000012277) common with Chr27 β-keratin genes. This implies that the three common over-represented TFs have more targets on Chr25 and fewer on Chr27 and therefore were not over-represented in the Chr27 clade. Similarly, the five

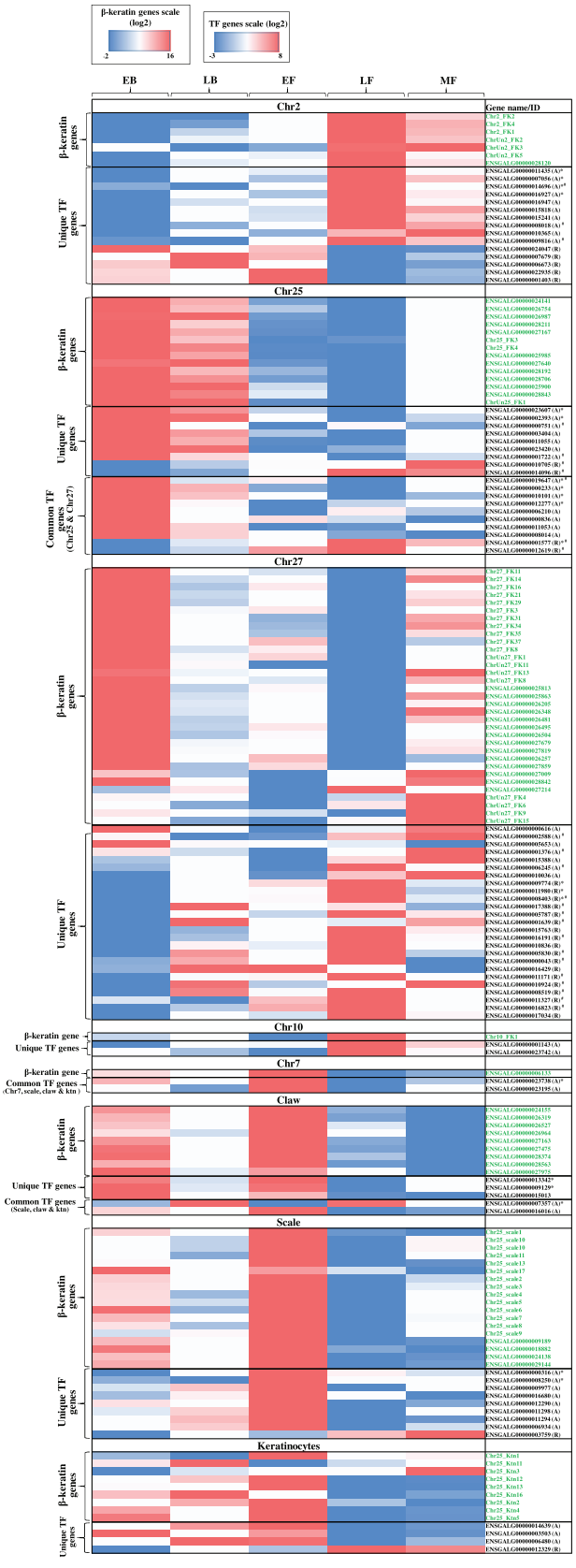


Fig. 4. Heatmap showing dynamic expression of β-keratin genes and their putative regulators. The heatmap is based on an average expression level (normalized) of the genes under each condition. Low to high expression is shown in a gradient scale from blue to red (see the color bar at top). The entries on the top of the heatmap denote the conditions from which the transcriptomes were obtained as detailed in figure S1, Supplementary Material online. The entries in top middle that interrupted the heatmap denote β–keratin genes on a chromosome or in a subclass. The extreme right entries denote β–keratins (green) corresponding to each chromosome or subclass (shown in top middle) and their corresponding TFs (black). As the expression differences between TF and target genes differ in magnitude, we used independent scales for them for better resolution of synchronous expression. TFs that regulate β-keratin genes of a particular subclass (scale, claw or feather) along with few keratinocytes are considered under the particular subclass. About 25 of the 81 pTFs did not show differential expression and are marked by "#" symbol. The over-represented TFs are indicated by "*".

over-represented TFs of the Chr27 clade may be categorized as three unique (ENSGALG00000009774, ENSGALG00000011980, and ENSGALG00000008403) and two common (ENSGALG00000019647 and ENSGALG00000001577) with Chr25. This implies that the two over-represented TFs of the Chr27 clade common with Chr25 mostly regulate Chr27 β-keratin genes along with much fewer Chr25 β-keratin genes and therefore were not over-represented in the Chr25 clade. Three TFs (all activators) of one family (HD) and three TFs (one repressor and two activators) of three families (C2H2 ZF, HD and T-box) were over-represented in the claw and scale β-keratin clades, respectively. Among them, the repressor TF (ENSGALG00000007357) commonly regulates scale and claw β-keratin genes. As the repressor TF was over-represented in the scale clade, it clearly implied that the repressors have more targets in scale β-keratin genes than in claw β-keratin genes. Overall, the enrichment analysis provided a set of major TFs, narrowing down the relatively large network (supplementary table S8, Supplementary Material online) to a few TFs (fig. 3C).

Major TFs Predicted by Differential Expression of β-Keratin Genes and Their pTFs

We conducted differential expression analysis of both β-keratin genes and their pTF genes (fig. 4, also see supplementary tables S10 and S11, Supplementary Material online) to: (1) cross check overall predicted TF–target gene relationship, (2) find additional major TFs regulating a small cluster of β-keratin genes of two clades or within a clade, skipped by enrichment analysis, and (3) understand the structural diversity of feather beyond the previous study (Ng et al. 2014). We found that among the five different feather structures (EB, LB, EF, LF and MF), Chr2 β-keratin genes were up-regulated in LF and MF (log fold change (FC) $\geq$ 2.59, $P < 0.02$) due to the up-regulation of 12 of the 15 pTFs of Chr2 (including 3 of the 4 over-represented TFs of the Chr2 clade). Similarly, Chr25 β-keratin genes were up-regulated in EB and LB (logFC $\geq$ 1.66, $P < 0.02$) and Chr27 β-keratin genes were down-regulated in LF (logFC ≥ -2.58, $P < 0.02$), due to up-regulation of some pTFs of Chr25 and down-regulation of some pTFs of Chr27 β-keratin genes, respectively (4 of 9 unique pTF of Chr25, 15 of 25 unique pTF and 3 of 10 common pTFs of Chr25

and Chr27 were not differentially expressed). While a few of the Chr25 and Chr27 differentially expressed pTFs are unique (supplementary tables S8 and S11, Supplementary Material online), others are common. The differentially expressed pTFs that commonly regulate Chr25 and Chr27 β-keratin genes include three (ENSGALG00000000233, ENSGALG00000010101 and ENSGALG00000012277) of the five over-represented TFs of the Chr25 clade but none of the Chr27 clade. The unique differentially expressed TFs of the Chr25 clade include two over-represented TFs (ENSGALG00000023607 and ENSGALG00000002393) of the Chr25 clade along with three other activators (ENSGALG00000003404, ENSGALG00000011055 and ENSGALG00000023420), whereas the differentially expressed unique TFs of Chr27 β-keratin genes are mostly repressors, including two (ENSGALG00000009774, ENSGALG00000011980) of the four over-represented repressor TFs of Chr27 β-keratin genes. This implies that for Chr25 β-keratin genes, two unique over-represented TFs along with other unique activators TFs effect up-regulation in EB and LB. For Chr27 β-keratin genes, two unique over-represented repressor TFs together with other unique repressor TFs effect down-regulation of β-keratin genes in LF, which is the barbs and barbule-less rachis portion of feather. Interestingly, the Chr7 β-keratin gene together with scale, claw and few keratinocyte β-keratin genes showed up-regulation in EB and EF (Chr7: logFC $\geq$ 2.48; claw: logFC $\geq$ 1.87; scale: logFC $\geq$ 1.37; keratinocytes: logFC $\geq$ 1.08; $P < 0.02$) relative to those in LB, MF and LF. Most pTFs of the scale, claw, keratinocyte and Chr7 β-keratin genes, including enriched regulators of claw and scale β-keratin genes, differentially expressed with their target genes.

Differential expression of a pTF gene with the target gene supported 262 (out of total 372) predicted TF–target gene relationships involving 56 of the 81 pTFs (figs. 2B and 4). Thus, compared with the overall prediction (fig. 2A), we found: for scale, 13 putative TFs (pTFs)/regulate 17 target genes; for claw, 7 pTFs/9 target genes; for keratinocyte, 16 pTFs/9 target genes; and for feather, 38 pTFs/56 target genes. Twenty-five TFs (9 activators and 16 repressors) that mostly regulate feather β-keratin genes, although they satisfied PCC $> |0.8|$ with their target gene, did not satisfy differential gene expression with their target gene. Based on this robust rule, we identified 56 TFs that regulate 91 β-keratin genes. We found that β-keratin genes on a chromosome or in a sub-class that formed clades have over-represented TFs. Differential expression analysis further supported 16 of the 20 predicted over-represented TFs and provided additional 10 major TFs that either regulate two clades/sub-classes, or a sub-clade of β-keratin genes within a clade skipped by enrichment analysis. The major TFs predicted by enrichment analysis (for clades) and further supported by differential analysis are summarized in table 1, which includes 22 activators but only four repressors. The major regulators predicted herein are described with their unique and common tentative roles, and are important candidates for functional characterization of β-keratin genes.

Table 1. Inferred major regulators of β-keratin genes with putative roles. The listed TFs are the key regulators of β-keratin genes predicted by a combined approach of clade-wise enrichment and synchronous differential expression of TF-target genes.

TF gene ID	TF Family	Activator/ repressor	Enriched	Diff. expressed	No. of target genes	Tentative role
ENSGALG00000007056	C2H2 ZF	activator	Yes	Yes	5	Uniquely regulates Chr2 β-keratins
ENSGALG00000011435	Ets	activator	Yes	Yes	4	
ENSGALG00000016927	C2H2 ZF	activator	Yes	Yes	4	
ENSGALG00000009774	Ets	repressor	Yes	Yes	15	Uniquely regulates Chr27 β-keratins
ENSGALG00000011980	Grainyhead	repressor	Yes	Yes	12	
ENSGALG00000000233	C2H2 ZF	activator	Yes	Yes	17	Commonly regulates Chr25 and Chr27 β-keratins
ENSGALG00000010101	C2H2 ZF	activator	Yes	Yes	21	
ENSGALG00000012277	bZIP	activator	Yes	Yes	17	
ENSGALG00000008014	bZIP	activator	No	Yes	13	
ENSGALG00000006210	bHLH	activator	No	Yes	4	
ENSGALG00000011053	Homeodomain	activator	No	Yes	10	
ENSGALG00000000836	RFX	activator	No	Yes	6	
ENSGALG00000023607	bHLH	activator	Yes	Yes	10	Uniquely regulates Chr25 β-keratins
ENSGALG00000002393	AP-2	activator	Yes	Yes	13	
ENSGALG00000003404		activator	No	Yes	3	
ENSGALG00000011055	Homeodomain	activator	No	Yes	3	
ENSGALG00000001143	Ets	activator	No	Yes	1	Uniquely regulates Chr10 β-keratins
ENSGALG00000023742	bHLH	activator	No	Yes	1	
ENSGALG00000023738	Homeodomain	activator	Yes	Yes	9	Commonly regulates scale, claw, Chr7 β-keratins
ENSGALG00000023195	Homeodomain	activator	No	Yes	6	
ENSGALG00000013342	Homeodomain	activator	Yes	Yes	6	Uniquely regulates claw β-keratins
ENSGALG00000009129	Homeodomain	activator	Yes	Yes	4	
ENSGALG00000007357	C2H2 ZF	repressor	Yes	Yes	10	Commonly regulates scale and claw β-keratins
ENSGALG00000000316	Homeodomain	activator	Yes	Yes	5	Uniquely regulates scale β-keratins
ENSGALG00000008250	T-box	activator	Yes	Yes	6	
ENSGALG00000003759	Nuclear receptor	repressor	No	Yes	3	

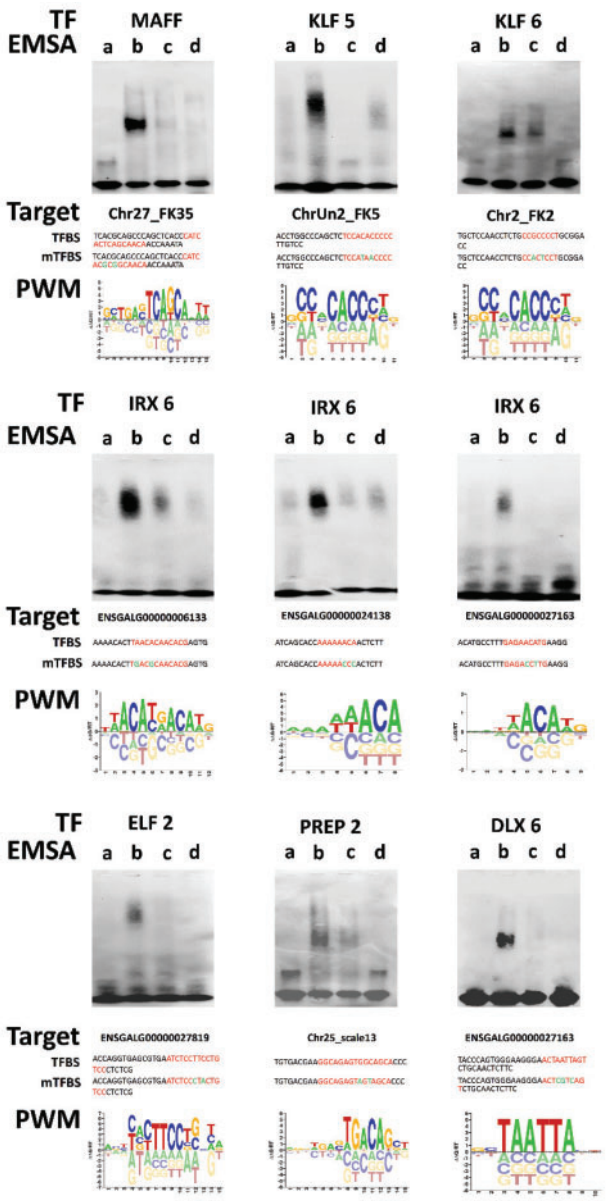


Fig. 5. Assays of the binding of a transcription factor to the cis-elements in its target genes by EMSA. (*A*–*I*) TF proteins purified from *Escherichia coli* were incubated with wild (TFBS) and mutated (mTFBS) cis-element probes in the target gene for EMSA. EMSA was conducted with the labeled probe alone (lane a) or with combined purified TF and the labeled probe (lane b). Combined purified TF with a labeled probe and excess of an un-labeled (10-fold) probe (lane c) showed quantitative competition of probes for binding. Combined purified TF and a mutated probe (lane d) confirmed binding specificity of TF. "Target" denotes the gene with the predicted TFBS in its promoter. "PWM" denotes the sequence logo of the position weight matrix for the binding-site motifs.

Feather β-keratin genes on a chromosome tend to be recent tandem duplications or to be undergoing concerted evolution (e.g., gene conversion). Thus, feather β-keratin genes on the same chromosome tend to be more similar in sequence and regulation than feather β-keratin genes on different chromosomes. This pattern of chromosome-wise differential expression of feather β-keratins is likely a major cause for structural differences among different parts of feathers. The up-regulation of Chr2 and Chr25 β-keratin genes in MF and LF of flight feather and in EB and LB of contour feather, respectively, was noted in a previous study (Ng et al. 2014). We additionally found Chr27 β-keratin genes were down-regulated in barb and barbule-less rachis portion (LF) of flight feather. Most interestingly, we found that ancestral scale and claw β-keratin genes together with the basal Chr7 feather β-keratin gene were significantly up-regulated in EB and EF. As the paleontological record from nonavian dinosaurs strongly suggested that pennaceous feather arose prior to flight (Xu et al. 2003, 2011; Hu et al. 2009), EB might be an early feather structure evolved. The other parts of feather might have evolved by expansion and divergence of β-keratin genes on different chromosomes with the recruitment of specific regulators. As EF is more similar to EB, we assume flight feather a modified contour feather with the distal portions, i.e., MF and LF, modified, whereas EF remains more similar to the EB of contour feather.

Experimental Validation of Predicted TF–Target Gene Relationships

We used Electrophoretic Mobility Shift Assay (EMSA) to verify the predicted TF–target gene relationships for 7 TFs and 14 cognate TFBSs (from target genes) (supplementary table S12, Supplementary Material online). Nine representative cases are shown in figure 5. In each case, in the absence of the cognate TF protein, only the fast-migrating, free biotin-labeled TFBS probe was observed. When the purified TF protein was included in the binding reaction, a strong TF–TFBS complex was observed as a slowly migrating complex, indicating interaction between the TF protein and the corresponding promoter sequence from the target gene. When purified TF protein along with biotin-labeled and excess un-labeled probe was included in the binding reaction, very light or no TF–TFBS complex was observed, implicating competitive binding of the excessive un-labeled probe with the TF protein in comparison to the labeled probe. When a purified TF along with a biotin-labeled mutated probe (mutation done in core TFBS positions) was included in the binding reaction, no shifted TF–TFBS complex was observed, indicating no interaction between the TF protein and the mutated probe. The selected seven TFs and their cognate TFBSs were representatives from each of the predicted categories having unique role, including the important regulatory case (Iroquois homeobox 6, IRX 6) connecting ancestral scale and claw β-keratins with basal feather β-keratin (Chr7: ENSGALG00000006133). Our EMSA tests demonstrated the reliability of our annotated β-keratin gene regulatory network.

Concluding Remarks

Gene duplication provides raw materials for the evolution of new functions or traits (Ohno 1970). However, how the regulatory machinery changes along with the evolution of duplicated genes remains largely unknown. In this study, we have characterized regulatory changes in duplicated β-keratin genes. We observed a β-keratin gene on Chr7 that was

originally derived from an ancestral scale or claw β-keratin gene likely located on Chr25, but has evolved to be a feather β-keratin. This ancestral β-keratin on Chr7 was duplicated to other chromosomes. Subsequently there were more duplications or translocations of feather β-keratin genes to several other chromosomes (Greenwold et al. 2014; Ng et al. 2014), including Chr25. While the Chr7 β-keratin gene still shares a TF with scale and claw β-keratin genes, the β-keratin genes in other chromosomes have recruited distinct TFs and diverged in regulation. In total, we identified 81 pTFs, which regulate 91 β-keratin genes. Among the 81 pTFs, 56 TFs showed differential expression with their target genes. Finally, 26 TFs were identified as major regulators, each regulating multiple β-keratin genes. Our data showed that the up- and down-regulation of a β-keratin gene in a feather region correlates well with the up- and down-regulation of its putative TF genes. From these observations, we propose that the regulatory divergence among β-keratin genes is largely responsible for the diversification of feathers in different body parts of a bird.

Materials and Methods

Transcriptome Analysis and Construction of Gene Co-Expression Modules

We used the transcriptome data obtained previously from two developing conditions of body contour feather and three developing conditions of flight feather (Ng et al. 2014). In total 15 transcriptomes data were available for 5 developing conditions of feather; each condition has 3 biological replicates (supplementary fig. S1, Supplementary Material online). The two parts of contour feather were defined as early-body (EB) and late-body (LB), whereas the three parts of flight feather were defined by early-flight (EF), middle-flight (MF) and late-flight (LF). The transcriptomes from these five feather regions were re-analyzed using the current chicken genome assembly (*Gallus gallus*. galGal4 Ensembl release 77) and gene annotation (*Gallus gallus*.galGal4.77.gtf); the previous analysis was based on release 72. For each transcriptome, low-quality reads were trimmed and adapters were removed, using Trimmomatic version 0.32 (Bolger et al. 2014). The processed paired end reads were mapped to the chicken genome using Tophat version 2.0.13 (Trapnell et al. 2009), allowing mismatches = 2, maximum multi-hits = 5 and *Gallus gallus*. galGal4.77.gtf as the reference gene annotation. The missing β-keratin genes in the previous annotation (Ng et al. 2014) were manually updated in *Gallus gallus*.galGal4.77.gtf. Moreover, mapping of reads was also allowed in novel regions beyond the transcriptome-index fetched from supplied Ensembl gene annotation *Gallus gallus*.galGal4.77.gtf. The alignment quality and distribution of the reads were estimated using SAM tools version 2.0.0 (Li et al. 2009). From the aligned reads, the transcripts were assembled by reference guided (RABT) policy using Cufflinks version 2.1.1 (Trapnell et al. 2013). In addition, de novo transcript assembly was also performed to cross validate RABT assembled transcripts. The gene expression level was calculated in terms of fragments per kilobase of exon per million mapped reads (FPKM). For FPKM calculation, uniquely mapped reads were first assigned to the genes. Multiple-hit reads were then redistributed to genes based on their relative abundance of uniquely mapped reads.

Genes with FPKM ≥ 1 in at least one of the conditions were considered "expressed" and selected for further analysis. To compare expression of the selected genes across the conditions, we applied the upper quartile normalization procedure (Bullard et al. 2010). Early flight replicate 3 (EF3) transcriptome was used as the reference for normalization because its FPKM values were most evenly distributed among the 15 transcriptomes.

The Pearson correlation coefficient (PCC) between the expression profiles of the genes was used as the distance metric for gene expression differences. To obtain β-keratin associated co-expression cluster, we applied two steps. First, we extracted a sub-set of β-keratin and associated genes that have PCC ≥ 0.8 with the β-keratin genes. Second, the genes in the sub-set were clustered based on the expression profiles by the hierarchical clustering-complete linkage distance policy, using heatmap.2 function in the "gplots" package of R (Warnes et al. 2015). The hierarchical cluster was cleaved to get clusters/groups of co-expressed gene, using the "hclust" function of R (R development core team, https://www.r-project.org/, version 3.1.1) (last assessed on April 17, 2016). Because co-expression does not always imply co-regulation, we added the condition that the genes in a cluster should belong to the same functional category (same GO term). The Gene Ontology (GO) annotation of the chicken genome was downloaded from Ensembl (release 77) via BioMart (Smedley et al. 2015). We performed Fisher's exact test (Benjamini and Hochberg 1995) to examine the functional categories that enriched in different clusters (P value < 0.05, FDR < 0.05). For a cluster that may more likely be related to keratin, we added the condition that the genes in a cluster should have the keratin-related functional terms (keratin, feather keratin, claw keratin, scale keratin, keratinocyte, intermediate filament, structural constituent of cytoskeleton) with the lowest P value and FDR in Fisher's exact test. A cluster of co-expressed genes with significant functional enrichment was termed a co-expressed module.

Discovering TF Binding Motifs and Inferring TF–Target Gene Relationships

The putative promoter region was defined as the region from the transcription start sites (TSS) or, if unavailable, the translation start site, to its upstream 1,000 bp, as most of the regulatory motifs are concentrated in this region. To identify whether a TFBS exists in the promoter of a gene and to infer TF–target gene relationship, our method includes three steps (supplementary figs. S4 and S5, Supplementary Material online):

First, for motif discovery we used the known or predicted TF–TFBS pairs of *Gallus gallus* from the CIS-BP database (cisbp.ccbr.utoronto.ca/; v1.01) (last assessed on April 14, 2016) (Weirauch et al. 2014) . Each TFBS, in the form of a position weight matrix, was used to scan the promoter region

of a set of strongly coexpressed genes (obtained by the preceding method), using FIMO (PMID: 21330290) in the MEME suite with *P* value < 1e-4. We assumed that genes in a co-expressed module with significant functional enrichment are co-regulated at the transcriptional level and their promoter regions might contain common binding motifs for the TF. For the motif enrichment analysis in each coexpressed module, Fisher's exact test was used (FDR < 1e-5).

Second, the TFs corresponding to the enriched motifs were extracted from the TF–TFBS pairs of *Gallus gallus* in the CIS-BP database. To verify an over-represented motif found in the promoter of a gene being true, the gene coding the TF protein was required to be correlated (PCC > |0.8|) with the gene whose promoter has the TFBS of the TF. If the above condition was satisfied, we called the TF/TFBS as putative TF/TFBS (pTF/pTFBS). In summary, the first two steps provided putative TF–target relationship of β-keratin and some other genes and putative TF–TFBS interactions.

Third, the conservation of a putative motif (TFBS) was assessed by finding the homologous sequences of chicken β-keratin genes in Turkey (*Meleagris gallopavo*). For this, we annotated the β-keratin genes in Turkey via a search method similar to Greenwold et al. (2014) but with some modifications. First of all, the chicken protein sequences encoded by all of the expressed β-keratin genes were used as queries to search against the Turkey genome (release 5.0) by TBLASTN (Altschul et al. 1990; Mount 2007), and the hits with similarity > 70% (*e* value < 10^{-5}) were selected. For each genomic region with significant hits, redundant hits that completely coincided with other hits were filtered out. For the remaining hits, the nucleotide sequences of the aligned region together with 100 bp upstream and downstream to it were retrieved using Galaxy (Goecks et al. 2010) and GeneWise (Birney et al. 2004) was used to verify the gene structure. Regions that contained an early stop codon (< 80 amino acids) were considered as pseudogenes. β-keratin genes express and functions as clusters and therefore several groups have common regulation and over-represented motifs. This property is useful to assess conservation of a β-keratin motif by finding a group of orthologous (homologous) β-keratin genes within another species rather than assessing the conservation among several species. Therefore, we find homologous relationships of chicken β-keratin genes in Turkey using BLASTP with chicken β-keratins as "queries" and Turkey β-keratins as "subjects". For each chicken β-keratin, hits with similarity > 70% (*E* value <1e-10) were considered significant. Almost 90% of Turkey β-keratins showed significant hits with *E* value <1e-25 and we choose the top five (based on lowest *E* value) Turkey β-keratin genes that showed similarity with a chicken β-keratin gene. The conservation of a TFBS was tested by aligning the chicken β-keratin promoter and its five homologous promoters of Turkey β-keratin genes using MUSCLE version 3.8.31 (Edgar 2004) and finding whether the TFBS also appeared (FIMO *P* value < 1e-4) in at least one of them on the same strand within a distance of 200 bp.

The pTF–target gene relationship was visualized with the software Cytoscape (Shannon et al. 2003) and categorized to groups based on a set of genes having common specific regulators distinct from other groups.

The phylogenetic analysis of the β-keratin genes was done using the maximum likelihood method in MEGA-7 (Kumar et al. 2016). The best model for phylogentic analysis was found through the model test option of MEGA-7 using Akaike and Bayesian information criteria (AIC and BIC). The phylogenetic tree was then compared with pTF–target gene categories to understand regulatory transition associated with the evolution of β-keratin genes. The major putative regulators were found through enrichment analysis of the TFs among the pTF–target gene category to which it belonged. To further confirm our prediction, we did differential expression analysis of the pTF gene and their target gene at different developing condition of the feather using the R package edgeR (Robinson et al. 2010; McCarthy et al. 2012). We consider the pTF–target gene relationship true if both show differential expression at the same time.

EMSA Assay to Validate TF–TFBS Binding

EMSA was conducted to validate the predicted TF–TFBS relationship. For the production of a recombinant TF protein and the subsequent EMSA experiment, we followed the protocol in a previous study (Yu et al. 2015). The primers used for TF amplification/cloning and production of TFBS probe is given in supplementary table S13, Supplementary Material online (sheets 1 and 2). For the EMSA assay between a TF and the TFBS present in the promoter of the putative target gene, we designed three types of probes. The first two types are wild type biotin-labeled and un-labeled probes, respectively. The third type is the biotin-labeled mutated probe, designed by modifying conserved nucleotides, using the position weight matrix of the TFBS (corresponding to a particular TF) as the reference. The EMSA mixtures of TF and TFBS were separated by 3.75% native polyacrylamide gel, transferred to Hybond N^+ membranes and visualized by the BioSpectrum imaging system (UVP).

Supplementary Material

Supplementary tables S1–S13 and figures S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This study was supported by the Ministry of Science and Technology of Taiwan (MOST 104-2621-B-001-003-MY3). M.J.B. was supported by postdoctoral fellowship of Academia Sinica. M.J.B. and W.H.L. conceived this study. M.J.B., C.P.Y., J.J.L., and W.H.L. designed the experiments and performed computational analysis. M.J.B. carried out the EMSA experiments. C.S.N., T.Y.W. and H.H.L. offered technical consultation and professional suggestions. M.J.B., W.H.L. and C.S.N. wrote the article. All authors read and revised the article.

References

Alibardi L, Dalla Valle L, Nardi A, Toni M. 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. *J Anat*. 214:560–586.

Alibardi L, Toni M. 2008. Cytochemical and molecular characteristics of the process of cornification during feather morphogenesis. *Prog Histochem Cytochem*. 43:1–69.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 57:289–300.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res*. 14:988–995.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94.

Dalla Valle L, Nardi A, Belvedere P, Toni M, Alibardi L. 2007. Beta-keratins of differentiating epidermis of snake comprise glycine-proline-serine-rich proteins with an avian-like gene organization. *Dev Dyn* 236:1939–1953.

Dalla Valle L, Nardi A, Toffolo V, Niero C, Toni M, Alibardi L. 2007. Cloning and characterization of scale beta-keratins in the differentiating epidermis of geckoes show they are glycine-proline-serine-rich proteins with a central motif homologous to avian beta-keratins. *Dev Dyn*. 236:374–388.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 11:R86.

Greenwold MJ, Bao W, Jarvis ED, Hu H, Li C, Gilbert MT, Zhang G, Sawyer RH. 2014. Dynamic evolution of the alpha (alpha) and beta (beta) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. *BMC Evol Biol*. 14:249.

Greenwold MJ, Sawyer RH. 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 Evol Biol*. 10:148.

Greenwold MJ, Sawyer RH. 2013. Molecular evolution and expression of archosaurian beta-keratins: diversification and expansion of archosaurian beta-keratins and the origin of feather beta-keratins. *J Exp Zool B Mol Dev Evol*. 320:393–405.

Herrmann H, Strelkov SV, Burkhard P, Aebi U. 2009. Intermediate filaments: primary determinants of cell architecture and plasticity. *J Clin Invest*. 119:1772–1783.

Hu D, Hou L, Zhang L, Xu X. 2009. A pre-Archaeopteryx troodontid theropod from China with long feathers on the metatarsus. *Nature* 461:640–643.

Huynh KD, Fischle W, Verdin E, Bardwell VJ. 2000. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev*. 14:1810–1823.

Kreplak L, Doucet J, Dumas P, Briki F. 2004. New aspects of the alpha-helix to beta-sheet transition in stretched hard alpha-keratin fibers. *Biophys J*. 87:640–647.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33:1870–1874.

Lee CH, Kim MS, Chung BM, Leahy DJ, Coulombe PA. 2012. Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nat Struct Mol Biol*. 19:707–715.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li YI, Kong L, Ponting CP, Haerty W. 2013. Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biol Evol*. 5:923–933.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 40:4288–4297.

Mount DW. 2007. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc*. 2007:pdb top17.

Ng CS, Wu P, Fan WL, Yan J, Chen CK, Lai YT, Wu SM, Mao CT, Chen JJ, Lu MY, et al. 2014. Genomic organization, transcriptomic analysis, and functional characterization of avian alpha- and beta-keratins in diverse feather forms. *Genome Biol Evol*. 6:2258–2273.

Ogura K, Matsumoto K, Kuroiwa A, Isobe T, Otoguro T, Jurecic V, Baldini A, Matsuda Y, Ogura T. 2001. Cloning and chromosome mapping of human and chicken Iroquois (IRX) genes. *Cytogenet Cell Genet*. 92:320–325.

Ohno S. 1970. Evolution by gene duplication. Berlin (Germany): Springer-Verlag.

Pabisch S, Puchegger S, Kirchner HO, Weiss IM, Peterlik H. 2010. Keratin homogeneity in the tail feathers of Pavo cristatus and Pavo cristatus mut. alba. *J Struct Biol*. 172:270–275.

Pettingill OS, Breckenridge WJ. 2012. Ornithology in laboratory and field. Orlando: Academic press. Elsevier Science.

Presland RB, Whitbread LA, Rogers GE. 1989. Avian keratin genes. II. Chromosomal arrangement and close linkage of three gene families. *J Mol Biol*. 209:561–576.

Prum RO. 1999. Development and evolutionary origin of feathers. *J Exp Zool*. 285:291–306.

Prum RO, Brush AH. 2002. The evolutionary origin and diversification of feathers. *Q Rev Biol*. 77:261–295.

Prum RO, Brush AH. 2003. Which came first, the feather or the bird? *Sci Am*. 288:84–93.

Regal PJ. 1975. The evolutionary origin of feathers. *Q Rev Biol*. 50:35–66.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13:2498–2504.

Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, et al. 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*. 43:W589–W598.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 31:46–53.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111.

Vandebergh W, Bossuyt F. 2012. Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Mol Biol Evol*. 29:995–1004.

Velinov M, Ahmad A, Brown-Kipphut B, Shafiq M, Blau J, Cooma R, Roth P, Iqbal MA. 2012. A 0.7 Mb de novo duplication at 7q21.3 including the genes DLX5 and DLX6 in a patient with split-hand/split-foot malformation. *Am J Med Genet A*. 158A:3201–3206.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechle M, Magnusson A, Moelle S, Schwartz M, Venables B. 2015. Various R programming tools for plotting data. *R Package Version* gplots:2.17.0.7.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158:1431–1443.

Weiss IM, Schmitt KP, Kirchner HO. 2011. The peacock's train (Pavo cristatus and Pavo cristatus mut. alba) II. The molecular parameters of feather keratin plasticity. *J Exp Zool A Ecol Genet Physiol*. 315:266–273.

Wu P, Ng CS, Yan J, Lai YC, Chen CK, Lai YT, Wu SM, Chen JJ, Luo W, Widelitz RB, et al. 2015. Topographical mapping of alpha- and beta-

keratins on developing chicken skin integuments: functional interaction and evolutionary perspectives. *Proc Natl Acad Sci USA*. 112:E6770–E6779.

Xiao L, Wang J, Li J, Chen X, Xu P, Sun S, He D, Cong Y, Zhai Y. 2015. RORalpha inhibits adipocyte-conditioned medium-induced colorectal cancer cell proliferation and migration and chick embryo chorioallantoic membrane angiopoiesis. *Am J Physiol Cell Physiol*. 308:C385–C396.

Xu X, Wang K, Zhang K, Ma Q, Xing L, Sullivan C, Hu D, Cheng S, Wang S. 2012. A gigantic feathered dinosaur from the lower cretaceous of China. *Nature* 484:92–95.

Xu X, You H, Du K, Han F. 2011. An Archaeopteryx-like theropod from China and the origin of Avialae. *Nature* 475:465–470.

Xu X, Zheng X, Sullivan C, Wang X, Xing L, Wang Y, Zhang X, O'Connor JK, Zhang F, Pan Y. 2015. A bizarre Jurassic maniraptoran theropod with preserved evidence of membranous wings. *Nature* 521:70–73.

Xu X, Zhou Z, Wang X, Kuang X, Zhang F, Du X. 2003. Four-winged dinosaurs from China. *Nature* 421:335–340.

Yu CP, Chen SC, Chang YM, Liu WY, Lin HH, Lin JJ, Chen HJ, Lu YJ, Wu YH, Lu MY, et al. 2015. Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *Proc Natl Acad Sci USA*. 112:E2477–E2486.

Zhang F, Zhou Z. 2000. A primitive enantiornithine bird and the origin of feathers. *Science* 290:1955–1959.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.

Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, Guo AY. 2012. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res*. 40:D144–D149.