Integrative and Comparative Biology Advance Access published July 3, 2008

Evolutionary relationships among copies of feather beta (β) keratin genes from several avian orders§

Travis C. Glenn,[1,*,†] Jeffrey O. French,[*] Traci J. Heincelman,[*,‡] Kenneth L. Jones[†] and Roger H. Sawyer[*]

[*]Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA; [†]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA; [‡]Marine Science Program, University of South Carolina, Columbia, SC 29208, USA

Synopsis The feather beta (β) keratins of the white leghorn chicken (order Galliformes, *Gallus gallus domesticus*) are the products of a multigene family that includes claw, feather, feather-like, and scale genes (Presland *et al.* 1989a). Here we characterize the feather β-keratin genes in additional bird species. We designed primers for polymerase chain reactions (PCR) using sequences available from chicken, cloned the resulting amplicons to isolate individual copies, and sequenced multiple clones from each PCR reaction for which we obtained amplicons of the expected size. Feather β-keratins of 18 species from eight avian orders demonstrate DNA sequence variation within and among taxa, even in the protein-coding regions of the genes. Phylogenies of these data suggest that Galliformes (fowl-like birds), Psittaciformes (parrots), and possibly Falconiformes (birds of prey) existed as separate lineages before duplication of the feather β-keratin gene began in Ciconiiformes (herons, storks, and allies), Gruiformes (cranes, rails, and allies), and Piciformes (woodpeckers and allies). Sequences from single species of Coraciiformes (kingfishers) and Columbiformes (pigeons) are monophyletic and strikingly divergent, suggesting feather β-keratin genes in these birds also diverged after these species last shared a common ancestor with the other taxa investigated. Overall, these data demonstrate considerable variation in this structural protein in the relatively recent history of birds, and raise questions concerning the origin and homology of claw, feather-like, and scale β-keratins of birds and the reptilian β-keratins.

Introduction

An important use of reptilian genomic information is to illuminate major evolutionary events in the adaptation of amniotes to living on land. The appearance of feathers in birds is a striking evolutionary development that has fascinated scientists for centuries. Previous studies have attempted to track the evolution of feathers from dinosaurs to birds using the fossil record (Chatterjee 1997), with a commonly accepted relationship relying upon the fossil *Archaeopteryx* to link Neornithes (modern birds) with the Saurichia (Chatterjee 1997). As evidenced in the fossil record, a few recognizable orders (primarily the recent ancestors of ducks, loons, auks, a variety of seabirds, and tinamous and ratites) had appeared by the end of the Cretaceous period 65 million years ago, and most other extant orders appeared within the next 30 million years (Feduccia 1996; Padian and Chiappe 1998). The Neornithine birds date back ~50 million years (Chatterjee 1997) and include both Palaeognathae (ratites and tinamous) and Neognathae (other modern birds). Unfortunately, the fossil record is incomplete and determining composition and homology of even spectacularly well-preserved feather-like structures in the fossil record is a daunting task.

An alternative approach is to use feather proteins and DNA sequences to assess the evolutionary events that must have occurred to produce the modern genomic organization that is characteristic of modern feathers. Characterization and comparison of the feather β-keratin gene family within and among extant avian and nonavian reptilian orders would allow us to make some observations about the evolutionary history of feather β-keratins (Sawyer et al. 2000, 2003a, 2003b, 2005). Although proteins similar to avian β-keratins are known to occur in embryonic alligators (Sawyer et al. 2003a), the DNA sequence of these genes remains elusive. Obtaining DNA sequences from species representing a diverse range of extant avian orders would provide much needed information about feather genes and eventually the entire multigene family of β-keratins in birds and reptiles. The primary

From the symposium "Reptile Genomics and Evolutionary Genetics" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 2–6, 2008 at San Antonio, Texas.
[1]E-mail: Travis.Glenn@gmail.com
§Dedicated to the Memory of Traci Jeanne Heincelman (deceased).
Integrative and Comparative Biology, pp. 1–13
doi:10.1093/icb/icn061

purpose of this study is to add to the pool of data from which we approach the larger questions of feather evolution, and to make some preliminary statements about the new contributions that feather β-keratin data can make.

Feather structure and β-keratins

The reptilian and avian classes are unique compared with other vertebrates because the keratinocytes in their epidermal appendages express β-keratins (Alibardi and Sawyer 2002; Alibardi et al. 2007 and references therein). The feathers of all extant birds examined to date are constructed of feather-type β-keratins (Astbury and Marwick 1932; Bear and Rugo 1951; Fraser and MacRae 1959, 1962; Fraser et al. 1971). Knowledge of the evolutionary relationships within this large multigene family of structural proteins may add to an understanding of the evolutionary origin of feathers.

Researchers interested in avian development have known for almost three decades that there is a variety of keratin proteins in the avian integument, and that the keratin proteins found in feathers are different from those found in other avian epidermal appendages, such as scales (Kemp and Rogers 1972). Walker and Rogers (1976) suggested that there were at least 22 different feather keratins, most of which were closely related. The existence of different but related feather β-keratin proteins implies the presence of a family of similar, but separate, feather β-keratin genes or gene copies. Studies by Gregg and Rogers (1986) suggested that there were about 20 genes in the feather-keratin family in the chicken. The feather-keratin family most likely developed through a series of duplication events, leaving the copies in extant birds relatively clustered in the genome. This hypothesis has been supported by other publications (Lockett et al. 1979; Molloy et al. 1982). Additionally, the areas flanking the coding region are similar among copies, leading to the conclusion that feather β-keratin genes did not arise from convergence of originally independent genes (Gregg et al. 1983).

In addition to the first publications of amino acid sequences from feather proteins (O'Donnell 1973, O'Donnell and Inglis 1974), there have been subsequent attempts to obtain DNA sequences that code for the feather β-keratins. Two of the most notable studies by Gregg and Rogers (1986) and Presland et al. (1989a) presented the most complete and informative description of the feather β-keratin genes from white leghorn chickens (order Galliformes, *Gallus gallus domesticus*). Figure 1 summarizes the structure of feather β-keratin genes in chicken, as described by Presland et al. (1989a) and Gregg and Rogers (1986).

The coding exon is flanked at the 5′ end by a noncoding exon that is 57–58 bp long and interrupted after base 37 by an intron. The first segment is almost 100% conserved whereas the second is 80% conserved. The intron varies from 324–341 bp and despite frequent divergence among gene copies, 90–100% of this intron is conserved, including both splice sites and several segments of undetermined significance (Presland et al. 1989a). Including the initiation (ATG) and termination (TAA) codons, the coding exon is 297 bp long and produces a protein of 97 residues. There is 95% similarity among amino acid sequences (Presland et al. 1989a) with most of the variation in the codons near the amino- and carboxyl-terminus (Gregg and Rogers 1986).

A noncoding region of 431–450 bp follows the 3′ end of the coding region. In addition to the segment around the polyadenylation signal, this 3′ noncoding region includes one other conserved segment, 18 bp in length, of undetermined significance. Gregg and Rogers (1986) assumed that only a role in gene maintenance could result in the high level of conservation in this 18 bp segment, but the exact nature of that role has not been explored. Overall, the gene copies sequenced by Presland et al. were between 1109–1141 bp in length. After ~340 bases (the 5′ intron) are removed by splicing, the length of the transcription product minus the poly (A) tail is ~800 bases (Gregg and Rogers 1986).

Further studies shed some light on the relationship between feather keratins and the other avian keratins, scale and claw. Presland et al. (1989b) described the arrangement of 18 feather β-keratin genes in a 53-kb

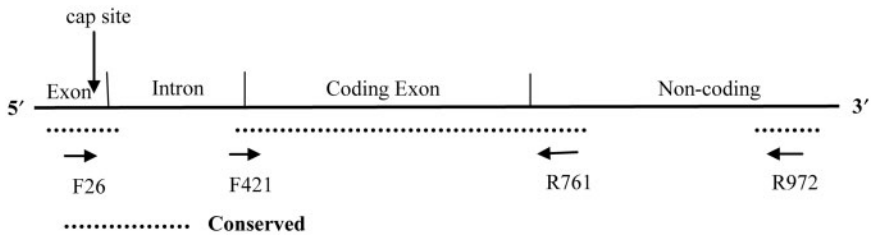


Fig. 1 Feather β-keratin gene structure (summarizing Presland et al. 1989a) and locations of the four primers tested in PCR. The numbers refer to the position of the 3′ base of the primers relative to an alignment of all sequences available in Presland et al. (1989a).

segment of the chicken genome, flanked downstream by 4 claw β-keratin genes and upstream by at least three copies of a related β-keratin gene designated ''feather-like.'' For many of the feather β-keratin genes, there are regions of similarity with the genes that code for claw and scale β-keratins. In scale keratin genes, however, there is a region of 156 bp, encoding 4 repeats of 13 amino acids (Gregg et al. 1984; Whitbread et al. 1991), which has a high Gly content and is not present in feather β-keratin genes. It is plausible that a single deletion event could have removed this region from a scale gene to produce an ancestor of the feather gene. Relevant studies of avian scale β-keratin and claw β-keratin genes, however, are still limited (Whitbread et al. 1991).

Specifically, our goals are: (1) to use novel PCR primers to attempt to amplify the feather β-keratin genes from as many species, representing as many orders, as could be obtained; (2) to determine, depending on the number of unique cloned PCR products, whether the data support the assumption that many avian species carry multiple variable copies of the feather β-keratin gene; (3) to expand the body of existing data by determining the DNA sequence of feather β-keratin genes in additional species; and (4) to make some observations about evolutionary relationships among gene copies.

Methods

Sample collection

In an effort to optimize new PCR primers for a range of avian species that was as inclusive as possible, we collected genomic DNA from a variety of species, both native and exotic. Most of our samples were collected as whole blood samples from live captive birds at Riverbanks Zoo and Garden, Columbia, SC, USA. With the cooperation of zoo hospital staff we were able to obtain samples from 33 species of birds representing 12 orders (Cracraft 1988; Sibley and Monroe 1990). Information from additional species and orders was obtained from genbank. Additional samples came from native birds, living or dead, which were brought to the zoo rehabilitation facility. Since chickens represent a domesticated species, we used several sources of genomic DNA as ''chicken.'' Unless otherwise noted, ''chicken'' refers to the breed maintained by Riverbanks Zoo as a relatively pure representation of wild red jungle fowl. Other chicken and domesticated game (quail, and game hen) samples came from store-bought meats. Table 1 lists all species included in this study.

Genomic DNA was isolated from 15 µl of whole blood or tissue by digesting in 900 µl of standard proteinase K digestion buffer (Sambrook et al. 1989) followed by an extraction protocol modified from Carter and Milton (1993) using guanidine thiocyanate with diatomaceous earth. The extraction products were run with a positive control for 20 min on a 1% agarose gel stained with ethidium bromide to check quality and quantity of the DNA isolated.

PCR amplification

Using aligned feather β-keratin sequences from chickens (Presland et al. 1989a), primers for the polymerase chain reaction (PCR) were designed in four conserved regions, including coding and non-coding exons (Table 2, Fig. 1; for details see French 2001, Appendix B). These primers were tested by PCR with chicken genomic DNA as the template. Primers F26 and R972 successfully amplified a fragment of DNA ~950 base pairs long. These locations were selected because they showed little variation among all known β-keratin sequences (all from chicken, at the time they were designed; Presland et al. 1989a) and thus required minimal degeneracy.

The initial sample set consisted of all species listed in Table 3. The primers F26 and R972 were used to amplify the target region of the feather β-keratin gene in one individual from each species sampled. Reactions were carried out in 50 µl volumes with final reaction concentrations of 50 mM KCl, 10 mM Tris–HCl pH 9, 1% Triton X-100, 2.0 mM $MgCl_2$, 150 µM of each dNTP, 250 µg/ml BSA (Fraction V, Sigma, St. Louis, MO, USA), 0.25 µM of each primer, 1 unit Taq DNA polymerase (Promega, Madison, WI, USA), and 40 ng of DNA. Thermocycling parameters were as follows: initial denaturation at 95°C for 3 min, followed by 35 cycles of 95°C for 20 s, 52°C for 20 s, and 72°C for 60 s. Eight microliters of the resulting products were examined for amplicons by electrophoresis for 45 min through 1.5% agarose gels containing ethidium bromide and 1× TBE buffer (Sambrook et al. 1989, p. 6.7–6.13).

As enumerated subsequently, we carried the products of the successful amplifications through cloning and sequencing steps. The successful sequences make up the initial dataset. We then repeated the entire process to produce a dataset focused on the protein coding sequence. All PCR products were precipitated by adding an equal volume of 20% PEG with 2.5 M NaCl, centrifuging, and washing twice with 80% EtOH (Kusukawa et al. 1990), and were quantified on a 1.5% agarose gel using lambda DNA as a standard.

Cloning of PCR products

To improve ligation efficiency, A tailing reactions were carried out for 30 min at 70°C in 10 µl volumes

Table 1 List of species used in this study

Order	Common name	Scientific name	Zoo ID	PCR Positive
Anseri-	Barrow's goldeneye	*Bucephala islandica*	6508	No
Ciconii-	American flamingo	*Phoenicopterus ruber*	0036, 5195, 6404	Yes
	Great blue heron	*Ardea herodias*	97-94	Yes
	Roseate spoonbill	*Ajaia ajaja*	3600, 4620	Yes
	Wood stork	*Mycteria americana*	98-144	Yes
Columbi-	Domestic pigeon	*Columba livia*	[a]	Not tested
Coracii-	Micronesian (Guam) kingfisher	*Todirhamphus cinnamominus*	6181, 6292, 6293, 6445	Yes
Cuculi-	Lesser roadrunner	*Geococcyx velox*	5307	No
Falconi-	Bald eagle	*Haliaeetus leucocephalus*	0552, 2035, 98-45	Yes
	Cinereous vulture	*Aegypius monachus*	0176, 0177	Yes
	Red-tailed hawk	*Buteo jamaicensis*	97-109	Yes
	Turkey vulture[c]	*Cathartes aura*	[a]	Yes
Galli-	Bobwhite quail	*Colinus virginianus*	[a]	Yes
	Red junglefowl	*Gallus gallus*	[b]	Yes
	Sumatran hill partridge	*Arborophila orientalis*	4812	Yes
Grui-	Crested seriema	*Cariama cristata*	1167	Yes
	Crowned crane	*Balearica pavonina*	6536, 6537	Yes
	Purple gallinule	*Porphyrio martinicus*	98-4	Yes
	Whooping crane	*Grus americana*	[a]	Yes
Passeri-	Bali myna	*Leucopsar rothschildi*	6584	No
	Fairy bluebird	*Irena puella*	6572	No
	Superb starling	*Lamprotornis superbus*	6070, 6509, 6547	No
	Yellow-faced (Dumont's) myna	*Mino dumontii*	6580, 6581	No
Pici-	Green aracari	*Pteroglossus viridis*	6346, 6533	Yes
	Keel-billed toucan	*Ramphastos sulfuratus*	6555, 6556	Yes
Psittaci-	Eclectus parrot	*Eclectus roratus*	2410, 3132	Yes
	Hyacinth macaw	*Anodorhynchus hyacinthinus*	3714, 4285, 6810	Yes
	Palm cockatoo	*Probosciger aterrimus*	5826, 6402, 6440, 6442	Yes
	Stella's lorikeet	*Charmosyna papou*	5467	Yes
	Thick-billed parrot	*Rhynchopsitta pachyrhyncha*	0386, 0528, 0531, 5439	Yes
Sphenesci-	Jackass penguin	*Spheniscus demersus*	0576, 0577, 0597, 2548	No
Strigi-	Barred owl	*Strix varia*	97-102, 97-104	No
	Burrowing owl	*Speotyto cunicularia*	5281	No
	Great horned owl	*Bubo virginianus*	97-41	No

The far right column indicates whether PCR amplicons of the expected size were obtained in any trial.
[a]DNA sample was obtained from a source other than the Riverbanks Zoo.
[b]The zoo does not assign ID numbers to every individual junglefowl chick.
[c]New world vultures are sometimes placed within Ciconiiformes.

with final concentrations of 50 mM KCl, 10 mM Tris–HCl pH 9, 1% Triton X-100, 2.0 mM $MgCl_2$, 0.2 mM dATP, 2 units Taq DNA Polymerase (Promega, Madison, WI, USA), and 250 ng of purified DNA. Three microliters (75 ng) of the A tailing reaction were used in to the ligation reaction for each sample. Ligations were then performed using pGEM-T® Vector System kits (Promega, Madison, WI, USA) according to the manufacturer's protocol. Two volumes (20 µl) of TE were added to each reaction, and each sample was incubated at 65°C for 20 min to dissociate the ligase from the DNA. Transformations were performed using Epicurian Coli® XL-10 Gold™ Ultracompetent Cells (Stratagene, La Jolla, CA, USA) according to the manufacturer's protocol or by electroporation of this strain. For each sample, 48–192 colonies, identified as potentially positive for inserts by color screening,

Table 2 Primers used for amplification and sequencing of β-keratins from six avian orders

Primer	Sequence (5′ to 3′)
Feather Beta K F26	CGCCCTCATCCACGTCTCTT
Feather Beta K F421[a]	TCCCDCMRCC**ATGTCCTG**
Feather Beta K F433	CC**ATGTCCTGCTATGATCTGTG**
Feather Beta K R761	GGCACCTGGCTTTGGGCTTC
Feather Beta K R972	CTCAACTTGCTTCAGGATYAA

PCR primers F26 and R972 were derived from aligned feather β-keratin sequences from chickens (adapted from Presland et al. 1989a). Internal primers were used for sequencing and were developed after obtaining the initial dataset (Table 3).
[a]Nucleotides in the protein coding region are in **bold.**

Table 3 Summary of the avian species and orders included in the initial dataset

Order	Species, common name (source, if not zoo)	No. of clones
Ciconiiformes	Great blue heron	2
	Wood stork	2
Coraciiformes	Micronesian (Guam) kingfisher	3
Falconiformes	Cinereous vulture	6
	Turkey vulture[a]	2
Galliformes	Chicken (store)	2
	Game hen (store)	2
	Quail (store)	3
Gruiformes	Crested seriema	4
Piciformes	Green aracari	1
Psittaciformes	Palm cockatoo	2
		29

The far right column indicates the number of clones successfully sequenced for each species.
[a]New-world vultures are sometimes placed within Ciconiiformes.

were picked and transferred to 40 µl of water. These colonies were screened again for inserts by performing PCR as above using M13 Forward and Reverse primers (cf. Stratagene, La Jolla, CA, USA) and colony suspensions as templates with an initial denaturation period of 10 min, and were examined on a 1.5% agarose gel. Of these PCR products, those that contained inserts of 800–1100 bp were used as templates for sequencing reactions.

DNA sequencing

Two methods of automated DNA sequencing were used. To produce clean reads over 1 kb in length, the initial dataset was sequenced on a LICOR automated sequencer (Lincoln, NE, USA). For this set, sequences from both DNA strands were determined from the products of the colony screening PCRs, which had been purified by PEG precipitation and quantified. Cycle sequencing (Amersham, Cleveland, OH, USA) of purified products was conducted with M13 Forward and Reverse primers labeled with IRD41 (LICOR, Lincoln, NE, USA). Secondarily, the focused dataset was sequenced using an ABI 377XL automated sequencer (Applied Biosystems, Foster City, CA, USA). For this set of sequences we used primer R972 or R761, and F421 or F433 (Table 2). The internal primers were designed specifically for the purpose of allowing us to focus on the protein coding region of the gene and obtain accurate sequences from the shorter reads obtained from the Applied Biosystems instrument. Sequences from both DNA strands were determined directly from the products of the colony PCRs using Big Dye terminator chemistry. For the focused dataset, sequencing reactions were carried out in 10 µl volumes with 0.32 µM primer, ~50 ng of PCR product, and Big Dye Terminators using ABI specifications except that the terminator mix was diluted 1:1 with halfBD^{TM} (GENPAK Ltd, Stony Brook, NY, USA) or a home-made equivalent (400 mM Tris–HCl pH 9, 10 mM $MgCl_2$).

Chromatograms of forward and reverse sequences from each sample were imported into Sequencher 3.1 (Gene Codes, Ann Arbor, MI, USA), which we used to edit and align the individual sequences into contigs. All computations and analysis reported below were performed using DNA sequences assembled in and exported from Sequencher 3.1.

Alignment and tree building

Sequences from the initial dataset were aligned in Sequence Navigator® 1.0.1 (Perkin Elmer Applied Biosystems, Foster City, CA, USA) using the Clustal algorithm (gap penalty = 16, gap length penalty = 6, all other settings at default values of 10). Exported sequences were brought manually in to BBEdit (Bare Bones Software, Inc., Bedford, MA USA) to be exported as Nexus files for use in PAUP* 4.0. We used Modeltest 3.04 (Posada and Crandall 1998) to select the best model of molecular evolution for the datasets. The model TrN + G (Tamura and Nei 1993; two rate categories for transitions, one rate category for transversions, gamma shape parameter of 0.5766; see French 2001, Appendix C for details) was chosen for the initial dataset using an alpha of 0.01.

A subset of the focused dataset sequences (chicken FBK A, B, D, feather like, and Micronesian kingfisher) were aligned in ClustalW (accessed at http://www.ebi.ac.uk/Tools/clustalw/; KTUP = 5, window length = 10, pairgap = 10, matrix = blosum, gap open = 50, no end gaps = yes, gap extension = 0.5,

all other settings at default values). All other sequences were then aligned manually relative to these. The percent of gene copies attained per species was calculated by $[(1-n/N) \times 100]$, where n is the number of unique clones and N is the total number of clones surveyed (Dang and Lovell 2000). We then used ModelTest 3.4 to determine the best-supported models of molecular evolution. The model TVM + I + G was determined for the complete focused DNA dataset (101 sequences) [base = (0.1734, 0.3455, 0.2695); Nst = 6; Rmat = (0.8785, 4.0604, 1.8121, 0.7357, 4.0604); rates = gamma; shape = 0.8737; pinvar = 0.2149]. Because analysis of 101 sequences is extremely computationally intensive, we then investigated smaller subsets of the data in detail. For the 24 Galliform sequences, the model HKY + G [Hasegawa et al. (1985); base = (0.1650, 0.3491, 0.2608); Nst = 2; TRatio = 1.7976; rates = gamma; shape = 0.3699] was determined and for the pruned DNA dataset (36 sequences) the model HKY + G [base = (0.1608, 0.3381, 0.2665); Nst = 2; TRatio = 1.8473; rates = gamma; shape = 0.4749] was determined. We used PAUP* 4.0 (Swofford 1999), to generate trees with the above parameter estimates in likelihood analyses using heuristic searches. We then ran 100 bootstrap replicates using the same model and search algorithm.

We found that the inferred amino acid sequences from the feather β-keratin coding regions could be aligned without the aid of alignment software. Alignment with chicken feather-like, claw and scale, and alligator scale were also achieved manually, once the repeat regions with high Gly content were excluded (Fig. 2). Trees were constructed from the inferred amino acid sequences using PAUP* 4.0 and uncorrected p distances, heuristic searches, and bootstrapping with the same search algorithm. Alligator scale, chicken scale, and chicken claw were defined as outgroups.

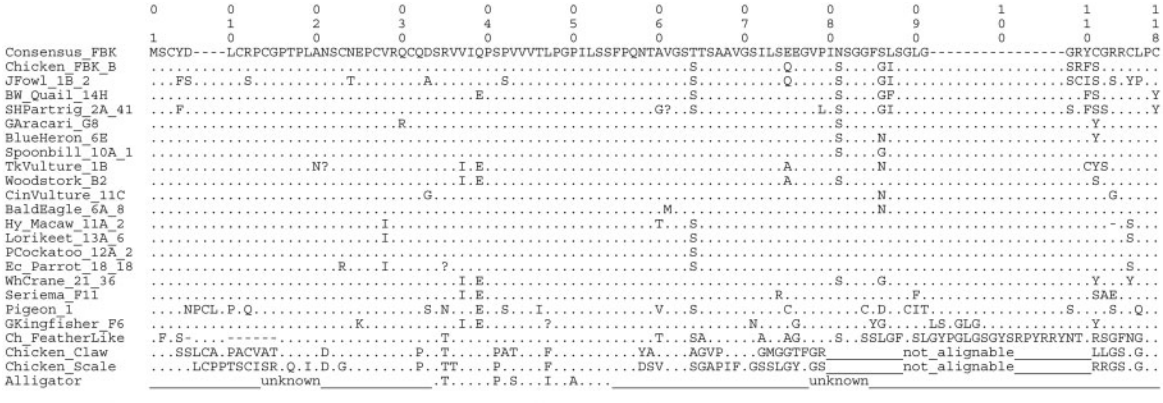
Results

Amplifications

The first indication that feather β-keratins vary significantly among avian taxa is that our primers, designed from highly conserved regions in the DNA sequence of the feather β-keratin gene from chickens, failed to amplify any fragment of DNA from many of the taxa in our sample set (Table 1). For every species for which primers F26 and R972 produced an amplicon, we found that amplicons were produced for every species sampled from that order. There were no cases in which a sample yielded an amplicon when other species from the same order did not. Perhaps most notable is that we were unable to obtain amplicons from any species of the order Passeriformes, which accounts for approximately 5700 of the over 9600 avian species (Sibley and Monroe 1990).

Initial dataset

The initial dataset (summarized in Table 3) represents the attempt to sequence clones from as many species as possible using primers F26 and R972. Comparison of forward and reverse sequences for all successfully sequenced clones produced reliable DNA sequence data for a total of 29 copies from 11 species (counting the store-bought samples as three species), representing seven orders. One of these orders (Piciformes) is represented by one copy from one species. The other six orders served as the primary target groups for the focused dataset described subsequently. The initial dataset allowed us to (1) compare the structure and organization of our



Sequence references: Presland *et al.* 1989a, b (Chicken FBK_B, Chicken Feather-like); Whitbread *et al.* 1991 (chicken claw); unpublished data, Y. I. and R. B. S. (alligator); this study (all others). See Table 1 for species identifications.

Fig. 2 Alignment of one inferred amino acid sequence per species from the focal dataset plus sequences used as outgroups for phylogenetic analyses. A 20 amino acid region is conserved across feather, feather-like, avian scale, avian claw, and alligator claw β-keratins.

amplicons with that described by Presland et al. (1989a), (2) determine which regions of the gene are conserved enough to be easily aligned for further analyses, (3) make some general observations about gene relatedness through phylogenetic analyses, and (4) select the orders of birds for primary use in the focused dataset, based on the function of our primers and with the goal of sampling multiple species from several orders.

The alignment of the DNA sequences (French 2001, Appendix D) suggests an organization of the gene that is consistent with the description by Presland et al. (1989a), in that the placement of the coding region is consistent and the alignment of that region is aided by high levels of conservation. In addition, there is a large intron flanking the 5′ end of the coding region that is difficult to align among sequences of different species, and sometimes even among sequences within species. Alignment of the intron is possible only to the extent that the two conserved regions of the noncoding exon are present (i.e., the coding sequences can be aligned easily but alignment of bases within the intron sequences is poor).

Figure 2 summarizes the inferred amino acid sequence data from the initial and focused datasets, as well as sequences presented by Sawyer et al. (2000). The amino acid sequences demonstrate that when translated, the coding region produces a protein with regions highly conserved among chicken (including inferred sequence from Presland et al. 1989a), other species we sequenced, and duck (Arai et al. 1986), emu (O'Donnell 1973), scale, claw, feather-like, and alligator scale (R. H. Sawyer, unpublished data). However, portions of the proteins are quite variable for such an important structural protein. The full alignment of the initial DNA dataset and trees from the sequences are given by French (2001).

Focused dataset

Having obtained sequences that were consistent with those of Presland et al. (1989a) and that demonstrated variation in the coding region among copies in the same individual, we then focused on obtaining DNA sequences from the coding region of additional clones and taxa. In this dataset, comparison of forward and reverse sequences for all successfully sequenced clones produced clean, reliable full length coding DNA sequence data for 109 copies from 18 species representing seven orders, although for three species only one clone was sequenced reliably (Table 4). Ten sequences (all six for pigeon and four for chicken [FBK A, B, D, and feather-like) were obtained from Genbank making a total of 119 sequences from 19 species representing eight avian orders. Of the 119 sequences, 101 were unique. No more than half of the possible unique copies of feather β-keratins were sequenced for any species (Table 4).

The alignment of 101 unique, complete-coding DNA sequences contained a total of 354 characters. Variability was highest in the 3′ third of the sequences. Relative to all other sequences, pigeons had an insertion of 9 bp (positions 31–39), kingfishers had an insertion of 15 bp (positions 245–263), hyacinth macaw had a deletion of 3 bp (positions 341–343), and featherlike had a deletion near the 5′ end (overlapping with the insertion observed in pigeon) and two insertions at the 3′ end (one of which overlaps with the insertion observed with kingfisher). The full dataset is available as an on-line supplement (Supplementary Material 1).

The maximum likelihood analysis of relationships among the full 101 DNA sequences was allowed to run for 24 h at which time the analysis was terminated and the two most likely trees saved and reviewed. In general, sequences formed monophyletic groups within orders where relationships of species are noncontroversial. For example, Fig. 3 shows the relationships among the 24 sequences investigated from Galliformes. Not surprisingly, sequences from red junglefowl and chicken are clearly intermixed. There is weak support for the partridge sequences as a monophyletic group, but overall, relationships among the chicken/junglefowl and the quail and partridge are not clear. Within the tree from 101 DNA sequences, some sequences were indicated as more related to sequences in species from other orders than to sequences within the same individual/species or order. To further investigate this issue, the two most distantly related sequences per species were retained for further analysis (i.e., the sequenced were pruned from 101 sequences to 36).

Among the 36 coding DNA sequences investigated in detail (Fig. 4), the sequences for five of the orders were monophyletic (Columbiformes, Coraciformes, Falconiformes, Galliformes, and Pisittaciformes). However, sequences within Falconiformes are only monophyletic if both new- and old-world vultures are defined within this order, and two orders (Columbiformes and Coraciformes) are represented by single individuals/species and thus are better described with species designations (pigeon and Micronesian kingfisher). Sequences from species of two orders are clearly paraphyletic (Ciconiiformes and Gruiformes). Finally, Piciformes is represented by a single sequence (green aracari).

The tree based on amino acid sequences (Fig. 5) is similar to the tree obtained from DNA data, except that the branch lengths are much shorter and the

Table 4 Summary of the number of clones and unique sequences from the avian species and orders included in the focused dataset

Order	Species (common name)	No. of clones	unique	Attained (%)[a]
Ciconiiformes	Great blue heron	4	3	25
	Roseate spoonbill	5	3	40
	Wood stork	8	4	50
Columbiformes	Pigeon[b]	6	5	17
Coraciiformes	Micronesian (Guam) kingfisher	11	9	18
Falconiformes	Bald eagle	2	2	–
	Cinereous vulture	11	11	–
	Turkey vulture[c]	9	8	11
Galliformes	Bobwhite quail	1	1	–
	Chicken[d]	6	6	–
	Red junglefowl	11	11	–
	Sumatran hill partridge	6	6	–
Gruiformes	Crested seriema	11	10	9
	Whooping crane	1	1	–
Piciformes	Green aracari	1	1	–
Psittaciformes	Eclectus parrot	2	2	–
	Hyacinth macaw	9	6	33
	Stella's lorikeet	10	8	20
	Palm cockatoo	5	4	20
Total		119	101	

Percent attained gives an indication of how well the number of clones sequenced cover the total amount of feather β-keratin diversity within each species.

[a]Percent gene copies attained $= [(1-n/N)\times 100]$ where n is the number of unique clones and N is the total number of clones surveyed (Dang and Lovell 2000).

[b]From Genbank (accessions AB017901–AB017906).

[c]New world vultures are sometimes placed within Ciconiiformes.

[d]From Genbank and initial dataset.

turkey–vulture sequences are allied with Ciconiiformes instead of Falconiformes. By defining the scale and claw keratins as outgroups, it is clear that feather-like is primitive, as are pigeon keratins. The full dataset used is available as an on-line supplement (Supplementary Material 2).

Discussion

It is necessary to note that although we accept commonly held ideas about which orders of birds are most primitive, based on diagnostic features of the orders that are recognizable relatively early in the fossil record, it is not our intent in this study to draw conclusions about organismal phylogeny of extant avian species based on DNA or amino acid sequence from avian β-keratins. We acknowledge the difficulty of distinguishing orthologous molecules from paralogous molecules, and we recognize that more than one scenario might cause two sequences to appear homologous (Hillis et al. 1996). Once duplication of a feather β-keratin gene has occurred, the multiple genes that are produced may tend to evolve in either an independent or a concerted manner (Zimmer et al. 1980). We will discuss the conclusions that can be made about duplication and speciation in either case.

In the initial dataset, Galliformes is best represented in terms of sampling of multiple species, although the store-bought samples represent three domesticated species. We used species closely related to chicken in our earliest amplification trials because our primers were designed using the published data on chickens (Presland et al. 1989a), and it was not surprising that Galliformes samples produced positive amplicons at a wide range of stringency in amplification parameters. Attempts at describing the relationships between major groups of birds place Galliformes, with Anseriformes (waterfowl), on the oldest branch of all the neognathous birds (see Fig. 6; Chatterjee 1997; Cracraft 1986). Interestingly, the largest group from which we were unable to obtain amplicons was the order Passeriformes, a lineage that is thought to be highly derived. However, PCR trials using the one species from Anseriformes produced

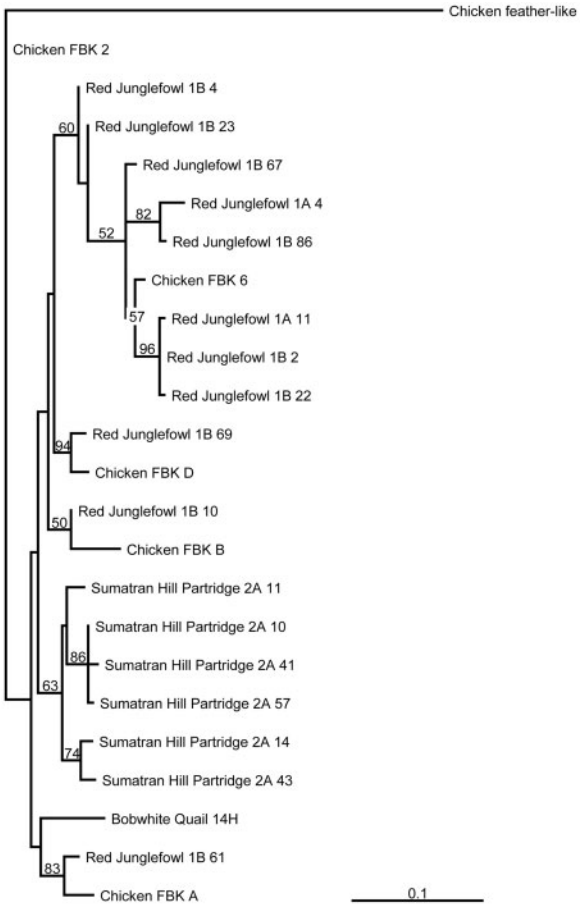


Fig. 3 Relationship of 24 β-keratin DNA sequences from Galliformes. The tree was constructed using maximum likelihood and heuristic searches (see text for details). Numbers on the branches refer to bootstrap values.

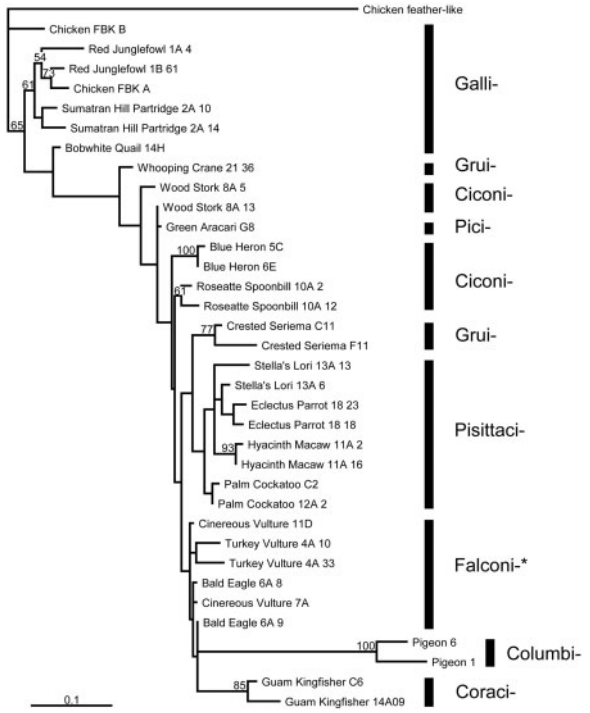


Fig. 4 Relationship of 36 β-keratin DNA sequences from all species in the focused dataset. The two most divergent sequences per species were included in the analysis. The tree was constructed using maximum likelihood and heuristic searches (see text for details). Numbers on the branches refer to bootstrap values.

negative results. So conclusions based on amplification success alone should be viewed with caution.

The most significant observation based on PCR trials follows from the observation that amplification results were negative for all tested species from 5 of the 12 sampled orders. Because primers F26 and R972 were designed to complement highly conserved regions flanking the coding exon as described by Presland et al. (1989a), it was clear from the outset that a significant amount of variation exists across avian taxa, even in this highly expressed gene encoding a structural protein. Indeed, if not for the presence of these conserved flanking regions, the coding region would be the only segment of these DNA sequences that could be aligned with confidence (see French 2001, Appendix F). DNA sequences from the coding region did demonstrate enough variation to warrant further analysis, in that the sequences (both of DNA and of inferred amino acids) were highly conserved across feather β-keratins but were more difficult to align with chicken scale, claw, and feather-like β-keratins. Figure 2 shows a 20-amino acid region that is conserved across all these genes, including the gene for alligator claw (Sawyer et al. 2000). This finding along with recent studies showing how antibodies to regions conserved among all β-keratins and feather β-keratins specifically (Sawyer et al. 2003a, 2003b, 2005; Alibardi et al. 2007) suggest a fertile area for future studies of the relationship between the feather β-keratin gene family and the β-keratins from reptiles.

One goal of this study was to verify that the feather β-keratin gene exists in multiple copies in species other than the chicken. After excluding identical sequences from the 119 sequences in the focused dataset, we found that for every sampled species with multiple sequences, there are at least two, and as many as 11, unique sequences (Table 4). Even with a very simple estimate of the percentage of sequences attained, it is clear that many more unique copies are likely to be discovered by sequencing more clones in each species.

The focused dataset (Table 4) includes sequences from at least two species from five orders: Ciconiiformes, Falconiformes, Galliformes, Gruiformes, and Psittaciformes. Many avian orders are well defined, but others are not as clear. We based our taxonomic groupings on Sibley and Monroe (1990)

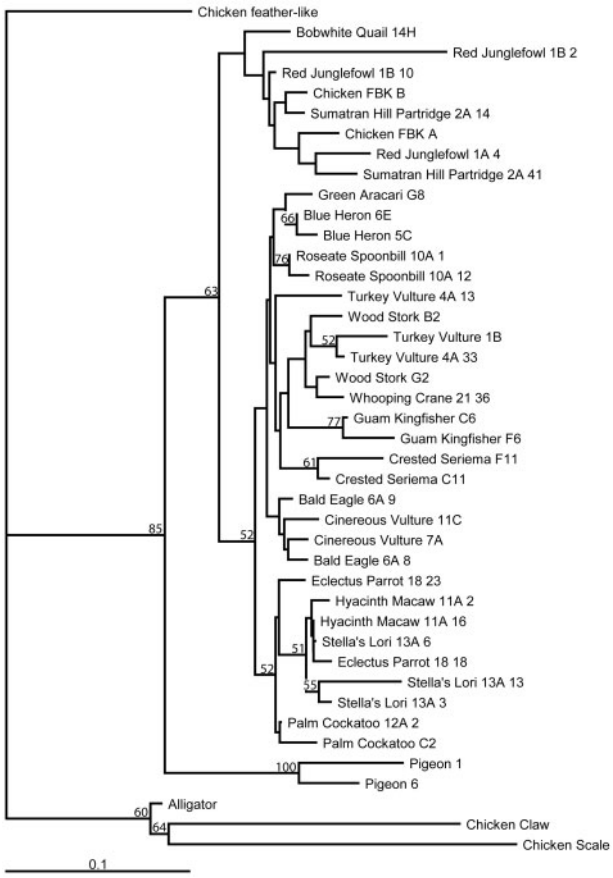


Fig. 5 Relationship of 42 β-keratin amino acid sequences from all species in the focused dataset. The two most divergent sequences per species were included in the analysis. The tree was constructed using distance and heuristic searches (see text for details). Numbers on the branches refer to bootstrap values.

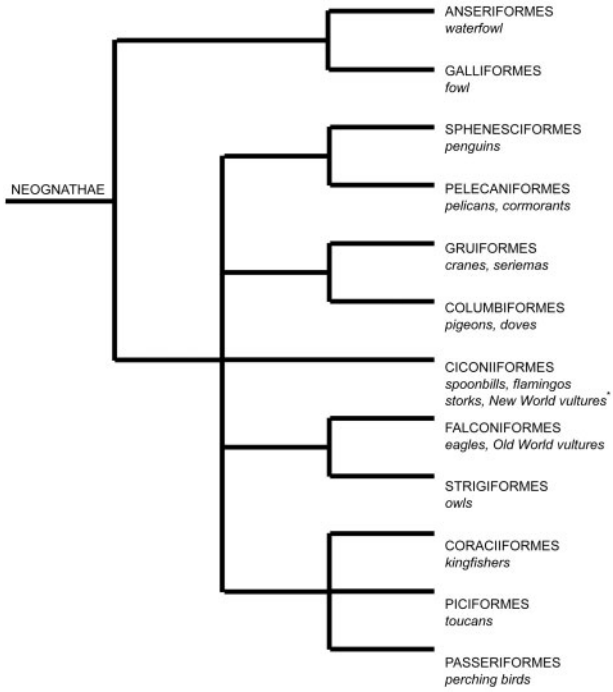


Fig. 6 Abbreviated phylogeny of neognathous birds, (modified from Cracraft 1988 and Chatterjee 1997) showing primarily the orders mentioned in this study. *See Discussion section for comments relative to the placement of vultures in Ciconiiformes or Falconiformes.

and Cracraft (1986, 1988). Cracraft (1986, 1988) separated old world vultures and new world vultures, placing new world vultures in Ciconiiformes, but Sibley and Monroe (1990) placed all falcons in Ciconiiformes, whereas more recent work suggests new world vultures should be in Falconiformes (http://www.aou.org/checklist/index.php3#cath; accessed 5/14/08). Thus, exactly what species should be in Ciconiiformes and Falconiformes is controversial. Similarly, the crested seriema is also very distant from whooping crane and is listed within the category of ''doubtful Gruiformes''. (http://en.wikipedia.org/wiki/Gruiformes; accessed May 14, 2008)

Our results support the assumption that gene duplication events have occurred within all avian orders, when two or more gene copies arose from one by some mechanism such as unequal crossing over (Smith 1976). In fact, it is possible that multiple copies in an individual could have identical sequences, causing two or more copies to be counted as one, especially because we focused on coding sequences and would have missed variation in the introns. Additionally, a duplication event might have occurred in the relatively recent past without time for subsequent divergence, or concerted evolution might have driven two copies to become more similar through periodic homogenization of multiple copies.

Based on our gene geneologies, we can begin to make a few observations about these duplication events and the evolutionary history of these taxa if we consider the order of events that would lead to these patterns of variation. Figure 7, a modification of diagrams given by Hillis et al. (1996), illustrates the possible scenarios. Duplication events produce two genes that are paralogous, and speciation events produce two genes that are orthologous (Hillis et al. 1996; Fitch 1970). Because we cannot be sure if two gene copies from different species are related by paralogy or orthology, our trees represent the molecular phylogenies, or the relatedness of the genes, not of the organisms (Hillis et al. 1996).

In the focused dataset, three orders, Galliformes, Ciconiiformes, and Psittaciformes, are well represented. First consider Psittaciformes. All the sequences from Psittaciformes are grouped together on the tree, and copies from the same species (or individual) are grouped together. In this case, an individual copy of the gene is most similar to copies from the same species.

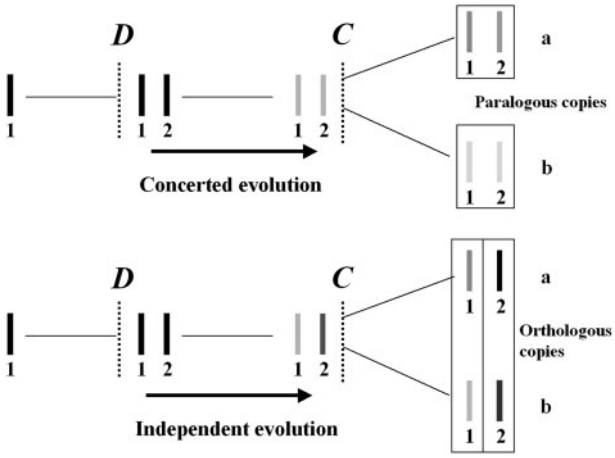


Fig. 7 Diagrams of alternative scenarios for evolution of multiple copies of DNA segments (modified from Hillis et al. 1996). D and C indicate duplication and cladogenic events, respectively; 1 and 2 are copies and a and b are lineages. Boxes denote similarity.

Two possibilities could have created this pattern of variation: either the duplication events did not occur until after the cladistic events that led to these species, or the copies generated by duplication events evolved in a concerted manner (not independently) before the cladistic events that led to these species.

Next consider Galliformes. All copies from Galliformes are grouped together on the tree, but within that order, some gene copies are more similar to copies in other species than to other copies within the same species. If any duplication events occurred before the cladistic events producing the Galliformes lineage, then duplicated copies from those events underwent concerted evolution until the divergence of Galliformes. However, after the divergence of Galliformes, at least some duplication events, followed by independent evolution of copies, must have occurred before the divergence of the species. This is the only scenario that would cause some copies to demonstrate more similarity across species than within species.

Finally, we consider Ciconiiformes, Falconiformes, and Gruiformes. Sequences among species in these taxa appear to be paraphyletic. This should be viewed with caution, however, because these orders of birds may be paraphyletic themselves (see Discussion section).

It was our goal to make some observations about duplication events of feather β-keratin genes in light of the evolutionary history of these taxa. Feather β-keratin genes did not evolve at a single time in ancestral birds and remain constrained by selection thereafter. Our data clearly indicate that feather β-keratin genes are evolving at a rapid pace throughout modern avian taxa. The observations we have made suggest that both Galliformes and Psittaciformes may have existed as separate lineages from the other tested orders before duplication of the feather β-keratin gene began. Duplication may have begun in the Galliformes lineage before the appearance of the species recognized today, but if duplication began in the Psittaciformes lineage before the divergence of the tested species, then that duplication was followed by concerted evolution of copies.

Further exploration of the origin and evolution of feather β-keratin genes will require more extensive sampling of avian species and nonavian reptiles. These efforts will benefit tremendously from sequences of whole genomes because the genes can be identified without the development of new, more universal primers for the feather β-keratin genes. This and previous studies have described several regions of the feather β-keratin gene where high conservation of sequences implies transcriptional or translational significance, such as the 18-bp segment in the 3′ noncoding region (Gregg and Rogers 1986; Presland et al. 1989a). This study relied on such segments for alignment of sequences. Future studies could not only explore the importance of these conserved regions and their specific roles in expression, but would also benefit from more experimentation with primers designed at those regions. Because our data extend gene conservation to include alligator claw, an attempt to compare reptilian and avian β-keratins more thoroughly is justified. Because the variation that we report has mostly occurred since the divergence of neognathous birds, the data presented in this study provide the start for a comprehensive database to understand the evolution of β-keratins and thus, feathers.

Supplementary data

Supplementary data are available at ICB online.

Acknowledgments

We dedicate this manuscript to T.H., who collected the focused dataset presented herein as part of her undergraduate honors thesis. T.H. was a McNair Scholar in the Honors College at the University of South Carolina where she finished her BS in Marine Science and her coursework for the M.S. in Biological Sciences. T.H. died in a tragic automobile accident before this work could be published. We thank Daniel Janes and Chris Organ for organizing the Reptilian Genome symposium where this was presented; Lisa Davis, Belinda Bade, and other members of the Sawyer lab who contributed to this work; Regina York and Lynette Washington for sequencing support; the staff of the Riverbanks Zoological Garden for their assistance in obtaining samples; and Julie Glenn for editing the manuscript.

References

Alibardi L, Sawyer RH. 2002. Immunocytochemical analysis of β keratins in the epidermis of chelonians, lepidosaurians, and archosaurians. J Exp Zool 293:27–38.

Alibardi L, Toni M, Valle LD. 2007. Hard cornification in reptilian epidermis in comparison to cornification in mammalian epidermis. Exp Dermatol 16:961–76.

Arai KM, Takahasi R, Yokote Y, Akahane K. 1986. The primary structure of feather keratins from duck (*Anas platyrhynchos*) and pigeon (*Columba livia*). Biochim Biophys Acta 873:6–12.

Astbury WT, Marwick TC. 1932. X-ray interpretation of the molecular structure of feather keratin. Nature 130:309–10.

Bear RS, Rugo HJ. 1951. The results of X-ray diffraction studies on keratin fibres. Ann NY Acad Sci 53:627–48.

Carter MJ, Milton ID. 1993. An inexpensive and simple method for DNA purifications on silica particles. Nucleic Acids Res 21:1044.

Chatterjee S. 1997. The rise of birds. Baltimore, Maryland: Johns Hopkins University Press.

Cracraft J. 1986. The origin and early diversification of birds. Paleobiology 12:383–99.

Cracraft J. 1988. The major clades of birds. In: Benton MJ, editor. The phylogeny and classification of the tetrapods, Vol. 1. Oxford, England: Clarendon Press. p. 339–62.

Dang H, Lovell CR. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. Appl Environ Microbiol 66:467–75.

Feduccia A. 1996. The origin and evolution of birds. New Haven: Yale University Press.

Fitch WM. 1970. Distinguishing homologous from analogous proteins. Syst Zool 19:99–113.

Fraser RDB, MacRae TP. 1959. Molecular organization in feather keratin. J Mol Biol 1:387–97.

Fraser RDB, MacRae TP. 1962. Helical models of feather keratin structure. Nature 195:1167–8.

Fraser RDB, MacRae TP, Parry DAD, Suzuki E. 1971. The structure of feather keratin. Polymer 12:35–56.

French JO. 2001. Characterization and evolutionary relationships among copies of feather β-keratin genes. M.S. Thesis. Columbia: University of South Carolina.

Gregg K, Rogers GE. 1986. Feather keratin: composition, structure, and biogenesis. In: Bereiter-Hahn J, Maltotsy AG, Richards KS, editors. Biology of the integument, Vol. 2. Vertebrates. New York: Springer-Verlag. p. 666–94.

Gregg K, Wilton SD, Parry DAD, Rogers GE. 1984. A comparison of genomic coding sequences for feather and scale keratins: structural and evolutionary implications. EMBO J 3:175–8.

Gregg K, Wilton SD, Rogers GE, Mable PL. 1983. Avian keratin genes: organization and evolutionary inter-relations. In: Nagley P, Linnane AW, Peacock WJ, Pateman JA, editors. Manipulation and expression of genes in eukaryotes. Proceedings of an International Conference held in conjunction with the 12th International Conference of Biochemistry, Monash University. New York: Academic Press. p. 65–72.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 21:160–74.

Hillis DM, Moritz C, Mable BK. 1996. Molecular systematics. 2nd ed. Sunderland, MA: Sinauer Associates.

Kemp DJ, Rogers GE. 1972. Differentiation of avian keratinocytes: characterization and relationship of the keratin proteins of adult and embryonic feathers and scales. Biochemistry 11:969–75.

Kusukawa N, Uemori T, Asada K, Kato I. 1990. Rapid, reliable protocol for direct sequencing of material amplified by the PCR. Biotechniques 9:66–72.

Lockett TJ, Kemp DJ, Rogers GE. 1979. Organization of the unique and repetitive sequences in feather keratin messenger ribonucleic acid. Biochemistry 18:5654–63.

Molloy PL, Powell BC, Gregg K, Barone ED, Rogers GE. 1982. Organization of feather keratin genes in the chick genome. Nucleic Acids Res 10:6007–21.

O'Donnell IJ. 1973. The complete amino acid sequence of a feather keratin from emu (*Dromaius novae-hollandiae*). Aust J Biol Sci 26:415–35.

O'Donnell IJ, Inglis AS. 1974. Amino acid sequence of a feather keratin from silver gull (*Larus novae-hollandiae*) and comparison with one from emu (*Dromaius novae-hollandiae*). Aust J Biol Sci 27:369–82.

Padian K, Chiappe LM. 1998. The origin and early evolution of birds. Biol Rev 73:1–42.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Biochemistry. 14:817–8.

Presland RB, Gregg K, Molloy PL, Morris CP, Crocker LA, Rogers GE. 1989a. Avian keratin genes I: a molecular analysis of the structure and expression of a group of feather keratin genes. J Mol Biol 209:549–59.

Presland RB, Whitbread LA, Rogers GE. 1989b. Avian keratin genes II: chromosomal arrangement and close linkage of three gene families. J Mol Biol 209:561–76.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sawyer RH, Glenn T, French JO, Mays B, Shames RB, Barnes GL Jr, Rhodes W, Ishikawa Y. 2000. The expression of beta (β) keratins in the epidermal appendages of reptiles and birds. Am Zool 40:530–9.

Sawyer RH, Rogers L, Washington L, Glenn TC, Knapp LW. 2005. The evolutionary origin of the feather epidermis. Dev Dyn 232:256–67.

Sawyer RH, Salvatore BA, Potylicki TF, French JO, Glenn TC, Knapp LW. 2003a. Origin of feathers: feather beta (b)

keratins are expressed in discrete epidermal cell populations of embryonic scutate scales. J Exper Zool (Mol Develop Evol) 295B:12–24.

Sawyer RH, Washington LD, Salvatore BA, Glenn TC, Knapp LW. 2003b. Origin of avian epidermal appendages: the bristles of the wild turkey beard express feather-type b keratins. J Exper Zool (Mol Develop Evol) 297B:27–34.

Sibley CG, Monroe BL Jr. 1990. Distribution and taxonomy of birds of the world. New Haven: Yale University Press.

Smith GP. 1976. Evolution of repeated DNA sequences by unequal crossover. Science 191:528–35.

Swofford DL. 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods). Ver. 4. Sunderland, MA: Sinauer Associates.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–26.

Walker ID, Rogers GE. 1976. Differentiation in avian keratinocytes: the properties of the proteins of the chick down feather. Eur J Biochem 69:329–39.

Whitbread LA, Gregg K, Rogers GE. 1991. The structure and expression of gene encoding chick claw keratin. Gene 101:223–9.

Zimmer EA, Martin SL, Beverly SM, Kan YW, Wilson AC. 1980. Rapid duplications and loss of genes coding for α chains of hemoglobin. Proc Natl Acad Sci USA 77:2158–62.