Radiation and Functional Diversification of Alpha Keratins during Early Vertebrate Evolution

Wim Vandebergh[1] and Franky Bossuyt*[,1]

[1]Biology Department, Unit of Ecology and Systematics, Amphibian Evolution Lab, Vrije Universiteit Brussel, Brussels, Belgium

*Corresponding author: E-mail: fbossuyt@vub.ac.be.

Associate editor: David Irwin

Abstract

The conquest of land was arguably one of the most fundamental ecological transitions in vertebrates and entailed significant changes in skin structure and appendages to cope with the new environment. In extant tetrapods, the rigidity of the integument is largely created by type I and type II keratins, which are structural proteins essential in forming a strong cytoplasmic network. It is expected that such proteins have undergone fundamental changes in both stem and crown tetrapods. Here, we integrate genomic, phylogenetic, and expression data in a comprehensive study on the early evolution and functional diversification of tetrapod keratins. Our analyses reveal that all type I and type II tetrapod keratins evolved from only two genes that were present in the ancestor of extant vertebrates. Subsequently, the water-to-land transition in the stem lineage of tetrapods was associated with a major radiation and functional diversification of keratin genes. These duplications acquired functions that serve rigidity in integumental hard structures and were the prime for subsequent independent keratin diversification in tetrapod lineages.

Key words: gene family evolution, comparative genomics, keratin clusters, tetrapod land colonization.

Introduction

The water-to-land transition of tetrapods was characterized by fundamental morphological changes in response to a totally different terrestrial environment. The integument gained a more resistant framework to cope with water loss and mechanical forces (Alibardi 2009), whereas the evolution of its variety in appendages (e.g., hairs or feathers) considerably contributed to the ecological diversification of each of the tetrapod lineages (Chuong and Homberger 2003). An important component of integument rigidity in tetrapods is formed by two classes of intermediate filament molecules termed type I and type II alpha keratins. These keratins are major structural proteins that form a cytoplasmic network essential in creating toughness and are a component of structures as diverse as hairs (Langbein et al. 2007), nails (Perrin et al. 2004), wool (Wilson et al. 1988), and baleen (Szewciw et al. 2010) in mammals, claws of reptiles (Eckhart et al. 2008), and larval beaks of amphibians (Olsen et al. 2004). Tetrapod alpha keratins are organized in two large gene clusters (Zimek and Weber 2005), which in humans together contain 54 keratin genes (Moll et al. 2008). Beta keratins are, despite their name, structurally unrelated to the alpha keratins. They have only been found in the skin and skin appendages of sauropsids (e.g., claws, scales, and feathers). Like the keratin-associated proteins of mammals, they were likely absent in stem tetrapods and only emerged after the divergence between sauropsids and mammals (Wu et al. 2008; Greenwold and Sawyer 2010).

The conquest of land by early tetrapods is anticipated to have been associated with substantial changes in integument keratinization, which in turn are likely to have been the consequence of duplications and functional shifts of alpha keratin genes (Schaffeld et al. 2005; Krushna Padhi et al. 2006). On the other hand, extensive keratinization in crown tetrapods is most evident in amniotes (Alibardi 2006). Furthermore, the origin of morphologically complex epidermal appendages, such as sebaceous glands, sweat glands, nails, hair follicles, and hair shafts in mammals, concurs with the appearance of tissue-specific keratins that have not been found in amphibians (Kurokawa et al. 2011). Amphibians use their skin as primary respiratory organ, and skin keratinization is therefore less extensive in this group (Alibardi 2003). These combined observations leave the possibility that the major diversification of keratins happened in individual tetrapod lineages and was of less importance during tetrapod conquest of land.

Our understanding of the origin and early diversification of type I and type II alpha keratins in tetrapods is still incomplete, for several reasons. First, available phylogenetic studies have shown that lampreys, fishes, and lungfishes expanded their keratin repertoire by lineage-specific gene radiations (Schaffeld, Haberkamp, et al. 2002; Schaffeld, Höffling, et al. 2002; Schaffeld et al. 2004, 2005, 2007; Krushna Padhi et al. 2006; Schaffeld and Schultess 2006), which contribute little information on keratin diversification in early tetrapods. Second, although genomic screening of tetrapods pointed out the presence of multiple keratin genes in amphibians (Zimek and Weber 2005), most of them have not been included in phylogenetic studies. Keratinization occurs in a variety of amphibian structures, like larval beaks and the nuptial pads in some anurans, but their molecular nature is largely unknown (Alibardi 2006). However, amphibians play a pivotal role in reconstructing early tetrapod evolution, and the

phylogenetic position of their keratin genes is essential to understand the overall origin of keratinized structures in tetrapods. Third, recent studies detected alpha keratins in claws and digital pads of reptiles (Eckhart et al. 2008; Hallahan et al. 2009; Alibardi et al. 2011), but a comprehensive screening for sauropsid alpha keratins is not available. This is mainly because research has focused on sauropsid-specific beta keratins, the major constituents of feathers, claws, and toe pads (Sawyer et al. 2000; Wu et al. 2004; Hallahan et al. 2009).

All together, knowledge on the type I and type II alpha keratin clusters in some vertebrate lineages strongly contrasts with the absence of a phylogenetic hypothesis for the all keratins and consequent lack of a theory for their diversification in early tetrapod evolution. Here, we provide a comprehensive study of all available type I and type II alpha keratins of eight species belonging to the major classes of tetrapods. We combine the available genomic and transcriptomic information with phylogenetic analyses to construct an evolutionary framework and use that information to study how keratin differentiation evolved in stem and early crown tetrapods. Finally, we compare expression data of amphibians and mammals to estimate functional diversification of alpha keratins in early tetrapods.

Materials and Methods

Keratin Cluster Organization in Tetrapods
The genomic organization of the keratin gene clusters from the African clawed frog (Silurana tropicalis), zebra finch (Taeniopygia guttata), chicken (Gallus gallus), anole (Anolis carolinensis), platypus (Ornithorhynchus anatinus), opossum (Monodelphis domestica), mouse (Mus musculus), and human (Homo sapiens) were assembled by BLAT searches (Kent 2002) implemented in the University of California Santa Cruz Genome Browser database (http://genome.ucsc.edu). Keratins retrieved from the Unigene database at the National Center for Biotechnology Innovation (http://www.ncbi.nlm.nih.gov/) were used in primary BLAT searches on the genomic assemblies of the respective species to locate type I and type II keratins in the genome. Proteins sequences or predicted proteins, determined by local Genscan gene predictions (Burge and Karlin 1997), were extracted and used in subsequent BLAT searches, that were reiterated until no new keratins were found. All keratin genes were accordingly located on one of the two keratin clusters. The genomic organization of keratins from pufferfish (Tetraodon nigroviridis) and zebrafish (Danio rerio) was determined to aid reconstructing the origin of the tetrapod gene cluster. Human keratins were named according to the revised protein keratin nomenclature (Schweizer et al. 2006). In this study, keratins with the same name in different species have shown not to be one-on-one orthologs. We therefore named the other vertebrate keratins according to their position in the genome. Genomic position, genomic assembly information, and the accession number (if available) are provided in supplementary table 1 (Supplementary Material online).

Phylogenetic Analyses
Amino acid sequences correspond to the central most conserved helical domain of the keratins. Type I and type II gene clusters were analyzed independently because their separate alignment yielded more informative data. Amphioxus (Branchiostoma floridae and Branchiostoma lanceolatum), lamprey (Lampetra fluviatilis), ciona (Ciona intestinalis), sea squirt (styela clava), zebrafish (D. rerio), pufferfish (T. nigroviridis), shark (Scyliorhinus stellaris), and lungfish (Protopterus aethiopicus) sequences were used to break up early lineages and to determine the number of independent origins of tetrapod keratins. Because screening of the shark and lamprey genome yielded only small fragments, we used the keratins retrieved from mRNA studies on the shark S. stellaris (Schaffeld et al. 1998, 2004) and the lamprey L. fluviatilis (Schaffeld and Schultess 2006). Some Notophthalmus viridescens (salamander) and Gekko gecko keratins were included because of their functional importance related to appendage formation (Ferretti et al. 1991; Hallahan et al. 2009). "Thread keratins" were included to clarify their evolutionary relationship with more terrestrially adapted keratins. Protein accession numbers and genomic information are provided in supplementary table 1 (Supplementary Material online).

Keratin data sets were aligned using ClustalX (Thompson et al. 1997) and manually corrected in MacClade 4.06 (Maddison DR and Maddison WP 2000). Bayesian posterior probabilities (PPs) were estimated using MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) under a mixed amino acid model with gamma correction for rate heterogeneity and invariable sites. Two parallel Markov chain Monte Carlo (MCMC) runs of four chains each were performed, with a length of 10,000,000 generations, a sampling frequency of 1 per 1,000 generations. Tracer 1.2 (http://tree.bio.ed.ac.uk/software/tracer/) was used to define the appropriate burnin and to check the convergence of the MCMC runs. Maximum likelihood (ML) analyses were performed by RAxML, using a JTT + G + I model with four partitions. Clade support was assessed by performing 200 "rapid" bootstrap replicates (Stamatakis et al. 2008).

Gene Tree Reconciliation and Duplication History
To account for phylogenetic uncertainties, we constructed two alternative gene tree phylogenies by rearranging branches with low support (Bootstrap ML < 75 or Bayesian PP < 95). The first rearrangements were conducted with the software Notung 2.6, which incorporates a strict "duplication–loss parsimony" principle and attempts a strict minimum of duplications given a species tree (Chen et al. 2000). In the second approach, synteny was used to minimize the amount of duplications in the ancient most nodes of the species tree (Minimal Early Duplication, MED model). Gene tree reconciliation methods, such as Notung, permit the rearrangements of unsupported clades in the gene tree to minimize the amount of duplications and losses in the species tree. These methods, however, allow rearrangements that are in conflict with the genomic organization (synteny) and can easily overestimate ancient duplications. We therefore designed an alternative model

of gene tree reconciliation to estimate the minimal amount of keratin gains in the most ancient nodes of the vertebrate species tree. We avoided phylogenetic clustering of keratins from highly different loci in the genomes as an additional rearrangement criterium. To do that, we used the genes of supported clades (ML bootstrap support $\geq$ 75% and Bayesian PP $\geq$ 0.95) to delimit borders within the cluster organization. Hence, genes were favored together when they were flanked by supported clades or occurred at the end of a cluster, even when the phylogeny did not support such a clade. Thereafter, the keratin relationships within these clusters were resolved according to a modified duplication/loss parsimony method: to estimate a minimum number of keratins in the oldest lineages of the vertebrate species tree, we performed rearrangements to reduce the overall amount of early duplications. Therefore, the monophyly of a clade containing genes from one species (or taxonomic group) was favored above the breakup of this clade to decrease the amount of duplications/losses (except for the monophyly of keratin 8 in lungfish and tetrapods reported by Schaffeld et al. 2005). For both type I and type II keratins, gene counts through tetrapod evolution were performed on each of the three phylogenetic hypotheses: Species divergence estimates were retrieved from the Timetree of life website (Hedges et al. 2006). When no "expert results" were available, the weighted average was used.

Gene Expression Estimation in Frog and Mouse

We used expressed sequence tag (EST) counts to estimate the level of keratin EST expression in the tropical clawed frog S. tropicalis and the mouse M. musculus. EST profiles were retrieved from the Unigene database (http://www.ncbi.nlm.nih.gov; March 2011). EST counts, in terms of transcripts per million values, were extracted from several body sites (head, stomach, intestine, liver, lung, and skin) and developmental stages (egg, neurula, gastrula, tailbud embryo, tadpole, metamorphosis, and adult), and Unigene IDs were cross-referenced with the proteins extracted from the genome of S. tropicalis. In mouse, the expression level was estimated for keratins from the same clades/groups using a comparable set of body sites and developmental stages.

Results and Discussion

Vertebrate Keratin Cluster Organization

A comprehensive screening for keratin genes on the most recent genomic scaffolds of eight tetrapod species and two fishes reveals an improved view on the tetrapod genomic organization. Reiterated rounds of search using BLAT (see Materials and Methods) yielded 348 different keratin gene sequences in tetrapods and resulted in an overview of their genomic position (fig. 1a). For both type I and type II clusters, recovered keratin sequences were situated between the same flanking genes, that is, SMARCE1 and EIF1, and FAIM2 and EIF4B, respectively (fig. 1a, black symbols), indicating homology of genes in these respective clusters. Our analyses show that tetrapod clusters contain between

14 and 34 genes on cluster one and 12 and 28 on cluster two and thus indicate considerably sized keratin clusters for all main tetrapod lineages. The most recent genomic scaffold of frog, the Joint Genome Institute (JGI) Silurana tropicalis genome assembly 4.1 (August 2005), reveals 36 keratin genes, which is seven more than the previous version (JGI S. tropicalis genome assembly 3.0 in Zimek and Weber [2005]). The anole genomic content adds up to 41 keratin genes, exceeding that of birds (27 and 28 in zebra finch and chicken, respectively), and indicates a more important role for alpha keratins in reptiles than perceived. Platypus revealed minimum 47 genes and indicates that the enhanced number of keratins in mammals (compared with other tetrapods) likely arose in the stem lineage. In line with previous findings (Zimek and Weber 2005; Krushna Padhi et al. 2006), we retrieved a more scattered pattern of keratin gene organization in the fishes D. rerio and T. nigroviridis. In Tetraodon, type I and type II keratins are situated on three pairs of assembled chromosomes (fig. 1b, chr. 18 & 3; 9 & 11; and 4 & 12) that diverged due to the whole-genome duplication in the lineage of teleost fishes (Jaillon et al. 2004). The fish ancestor probably contained three keratin gene loci, two of which having the same non-keratin flanking genes as the tetrapod gene clusters (fig. 1b). The typical organization of keratins and keratin-flanking genes thus likely originated before the divergence of fish and tetrapods.

Phylogenetic Hypotheses

To understand the early evolution of keratin genes in tetrapods, we combined the retrieved keratin genes of our eight species in phylogenetic analyses. We determined the phylogenetic position of type I and II keratins with respect to other types of intermediate filament proteins by constructing a data set with representatives of these proteins. These analyses provide high support for the monophyly of type I and type II keratins, with the exclusion of the urochordate Ciona and Styela intermediate filament proteins (see supplementary fig. 1, Supplementary Material online). Thread keratins have only been found in lamprey, hagfish, fish, and frog and represent a functionally different set of keratins (Schaffeld and Schultess 2006). Thread keratins (TH) diverged before the type I and type II keratin diversification in vertebrates (supplementary fig. 1, Supplementary Material online). We used this information to construct two larger data sets with choice of accurate outgroups. Our type I and type II data sets consist of 298 and 316 unambiguously aligned amino acids, largely corresponding to the conserved rod (helix) regions of keratins. ML and Bayesian analyses supported highly consistent trees, showing strong support for multiple clades in our type I and type II phylogenies (figs. 2 and 3, respectively). However, several unsupported nodes assume an unnecessary large amount of keratin duplications and losses during vertebrate evolution. We therefore constructed two alternative phylogenetic hypotheses, in which well-supported relationships (Bootstrap ML $\geq$ 75 and Bayesian PP $\geq$ 95) were used as a scaffold, whereas unsupported branches were allowed to shift according to two criteria. We used the
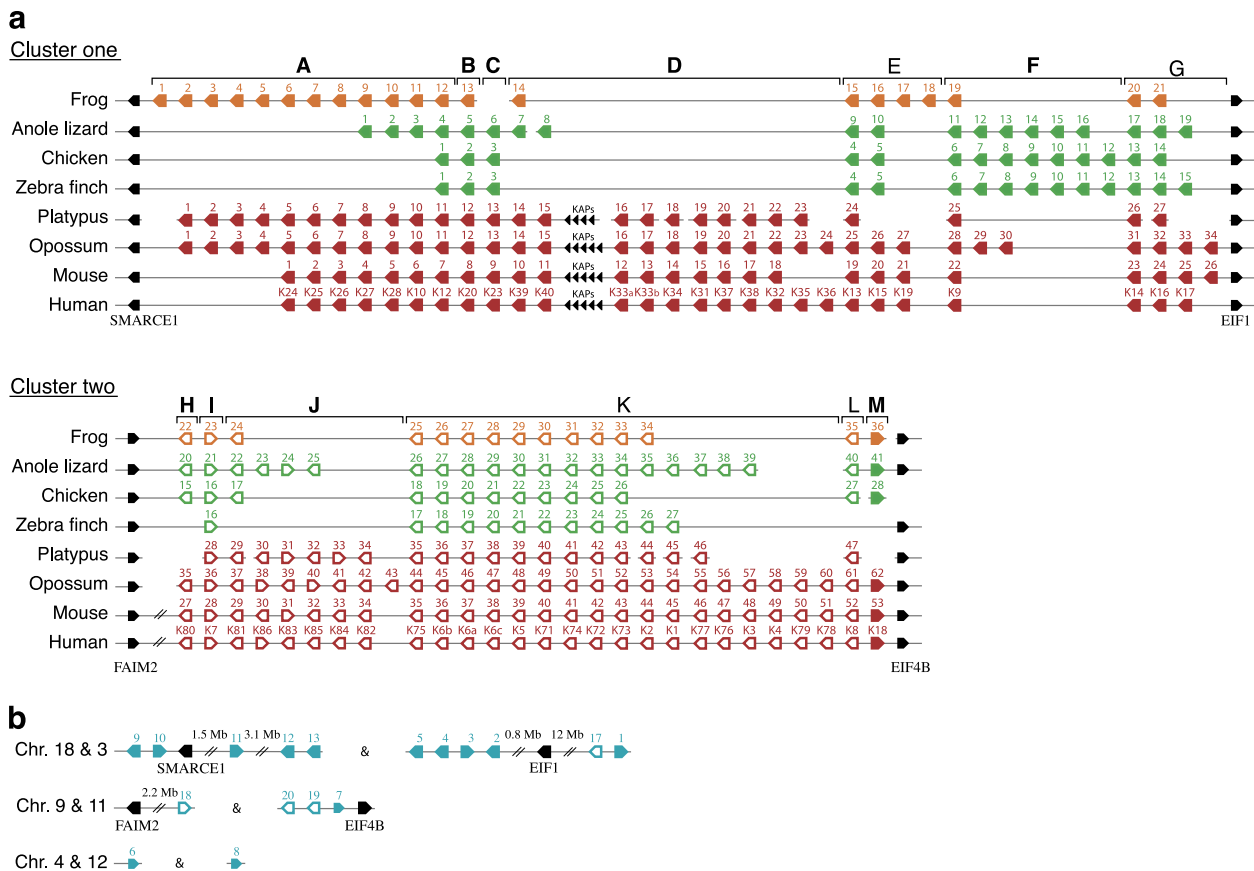
FIG. 1. Genomic organization of type I and type II keratins (*a*) Keratins in tetrapods. Arrows indicate the direction of keratin genes in the genome. Protein names follow (Schweizer et al. 2006) for human keratins and are numbered according to their genomic position for other tetrapods. Connecting lines indicate keratins genes on the same genomic scaffold. Letters denote clades supported by our phylogenetic analyses (bold) or groups between supported clades (regular). (*b*) Keratins in pufferfish (*Tetraodon nigroviridis*) are located on six assembled chromosomes. Keratins from unassembled chromosomes were not included. Filled triangles = type I keratins; Empty triangles = type II keratins; Black triangles = non-keratin flanking genes; and KAPs = keratin-associated proteins. Colors correspond to those in figures 2 and 3.

strict duplication/loss parsimony method implemented in the software Notung (Chen et al. 2000) to construct a phylogenetic hypothesis that assumes a minimum amount of duplications and losses over the whole tree. Because this method favors ancient duplications (Hahn 2007), we also took advantage of genomic information (synteny) and tetrapod evolutionary relationships to construct a phylogenetic hypothesis in which the amount of duplications early in tetrapod evolution was minimized (MED model). These trees complement the strength of Likelihood and Bayesian models of phylogeny with the parsimony principle of minimal gains and losses of genes on syntenic positions. Below, we use them jointly to further investigate keratin evolution.

Origin of Tetrapod Cluster Organization

Our phylogenetic reconstructions and associated estimates of speciation versus duplication for nodes in the trees allow us to estimate the number of ancestral keratins. First, both lamprey keratin types split off as a single supported clade (figs. 2 and 3, nodes S1), indicating that they diversified in the lineage to Petromyzontida itself. This suggests that type I and type II keratin clusters each arose from a single gene in the vertebrate ancestor (fig. 4a). Second, one of the intriguing characteristics of the tetrapod clusters is the presence of

a type I keratin on keratin cluster 2 (fig. 1). Our trees show strong phylogenetic support for a clade containing only shark and Osteichthyes type I keratins (fig. 2, node S2a) and place the second shark type I keratin together with tetrapod clade M (fig. 2, node S2b). These observations indicate that a duplication in stem gnathostomes can explain the two shark type I keratins and marks the origin of keratin organization on two different genomic loci (one of which containing two different keratin types) (fig. 4a). Third, the phylogenetic uncertainty for some nodes in our keratin gene trees does not allow us to unequivocally estimate the number of keratins in tetrapod ancestors. To pinpoint along which branches in tetrapod evolution the major keratin gains occurred and to calculate the rate at which keratin genes were gained throughout tetrapod evolution, we plotted the estimated amount of ancestral keratins at every node of the tetrapod species tree against time (fig. 4b), using our three hypotheses of keratin evolution. Our analyses reveal an important increase in keratin genes in the stem lineage of tetrapods, even under the most conservative approach (MED). This shows that both keratin types underwent major diversifications in the tetrapod stem lineage, that is, the period where early land-colonizing tetrapods had to optimize their relationship with the terrestrial environment.
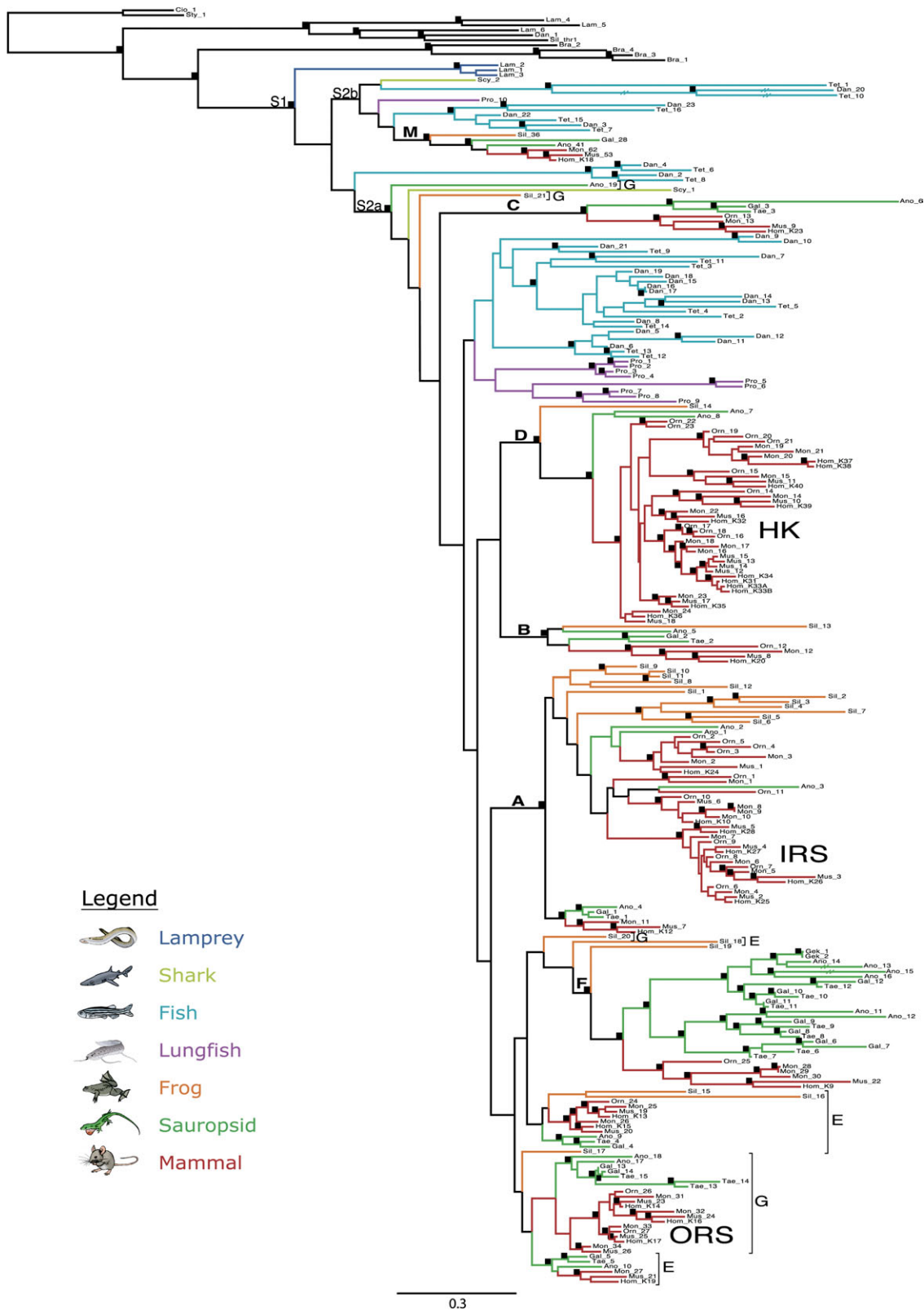
Fig. 2. Phylogeny of keratin type I genes. Phylogenetic reconstructions of type I keratins were inferred from ML searches. Branches interrupted by two parallel lines correspond to half the length of that branch. Black filled squares represent ML bootstrap support ≥ 75% and Bayesian PP ≥ 0.95. Speciation events discussed in the text are indicated with S. HK = hair keratins; IRS = keratins present in the inner root sheath; ORS = keratins present in the outer root sheath + companion layer. Gene names represent the species name followed by the number of their position in the genome.
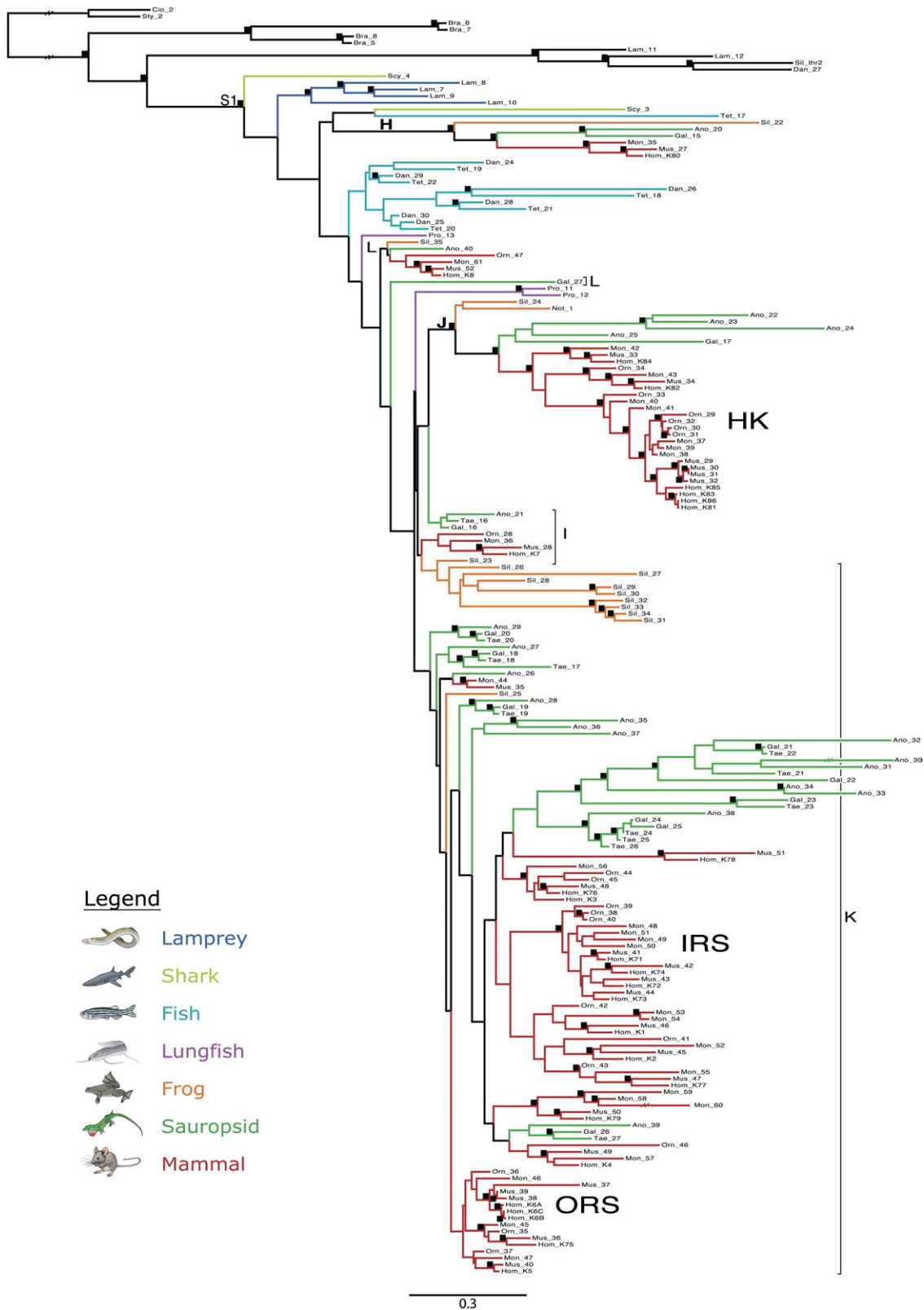
FIG. 3. Phylogeny of keratin type II genes. Phylogenetic reconstructions of type II keratins were inferred from ML searches. Branches interrupted by two parallel lines correspond to half the length of that branch. Black filled squares represent ML bootstrap support ≥ 75% and Bayesian PP ≥ 0.95. Speciation events discussed in the text are indicated with S. HK = hair keratins; IRS = keratins present in the inner root sheath; ORS = keratins present in the outer root sheath + companion layer. Gene names represent the species name followed by the number of their position in the genome.
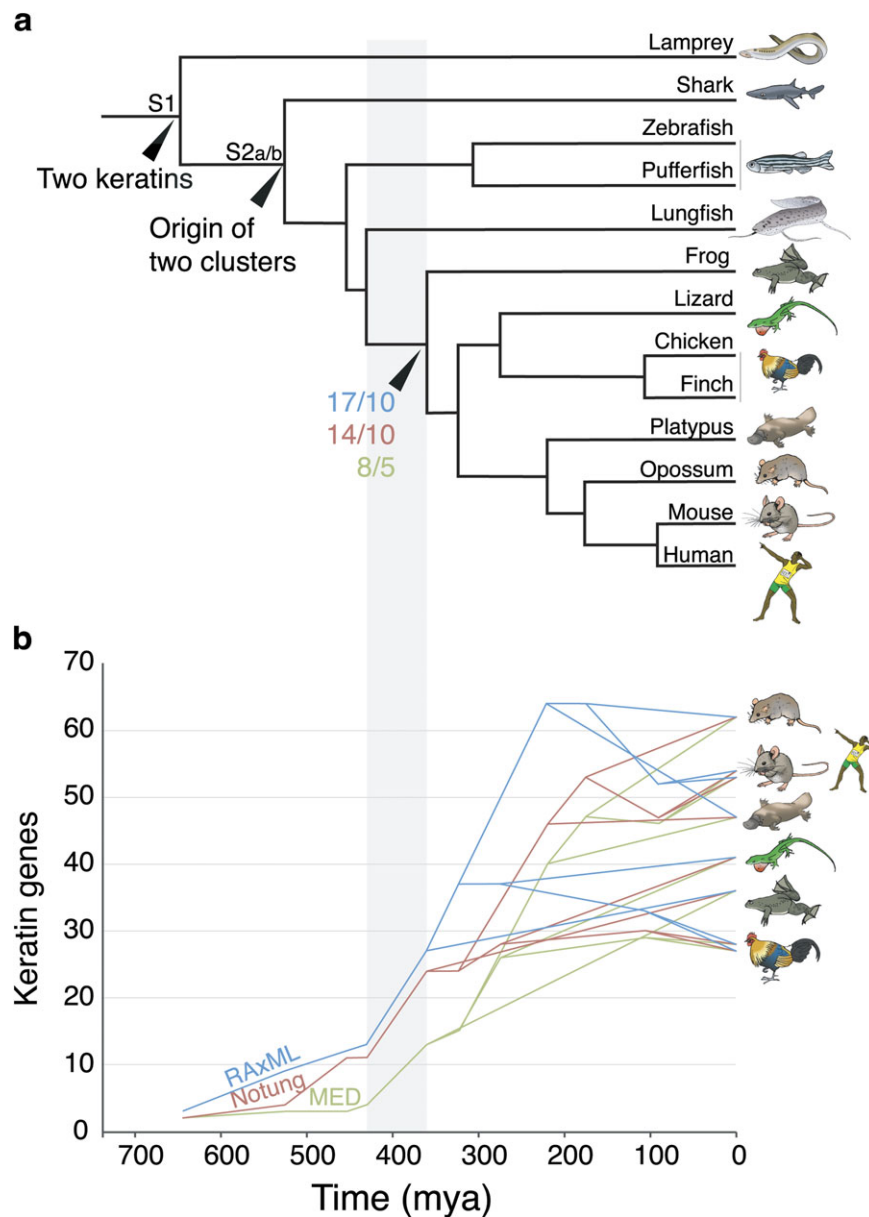
FIG. 4. Origin and expansion of the tetrapod keratin gene clusters (*a*) Vertebrate timetree with indication of keratins according to our phylogenetic reconstructions: S1 and S2 indicate the presence of two ancestral keratins in the stem lineage of vertebrates and the origin of the clusters on two different genomic locations, respectively. The colored numbers indicate the estimated amount of keratins in type I and type II in the tetrapod ancestor (gray shade) according to three phylogenetic hypotheses. (*b*) Keratin gene diversity through time. The total amount of keratins at every node in the species tree is plotted against the time of speciation for each node and thus corresponds to the duplication rate in tetrapod evolution. Colors correspond to the reconstruction method used.

Functional Evolution of Keratins in Tetrapods

The acquirement of new keratin genes in land-colonizing tetrapods most likely was associated with functional diversification of these copies. We combined information on keratin gene expression in mouse and frog to estimate the ancestral crown-group function of these genes (Materials and Methods). Although a full functional reconstruction is problematic due to phylogenetic uncertainty and/or unknown functions, a number of well-supported clades allow a cautious estimate of ancestral functions. Our phylogenetic reconstructions show that several keratin genes show limited diversification during tetrapod evolution, that is, there often is one copy in extant taxa (fig. 2: B, C, and

M; fig. 3: H, I, and L). Most of them are simple epithelial keratins and are expressed abundantly in internal tissues and earlier stages of development (fig. 5), where their primary function is regulatory in nature (e.g., osmosis, apoptosis, and cell polarity) (Owens and Lane 2003). Their basic function and conservation as a single copy throughout tetrapod evolution suggest a conservative function for keratin genes in these clades. Our likelihood trees (figs. 2 and 3) suggest that at least some of these acquired their function before the tetrapod stem lineage (clades C and M in type I, clade H and group L in type II). The origin of stratified epithelial keratins is mainly estimated on the tetrapod stem lineage. In development, their expression starts in
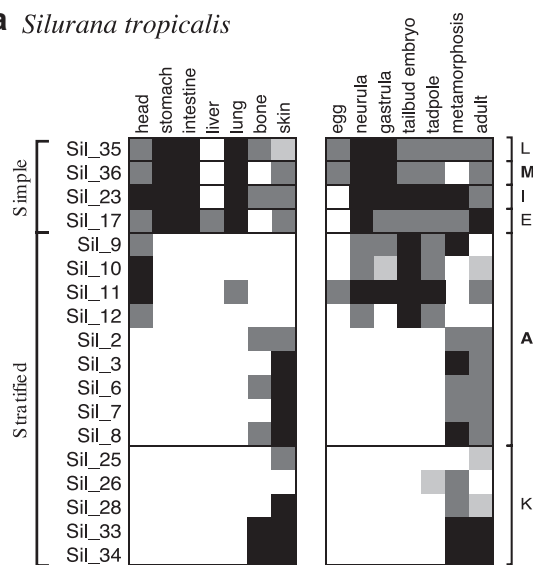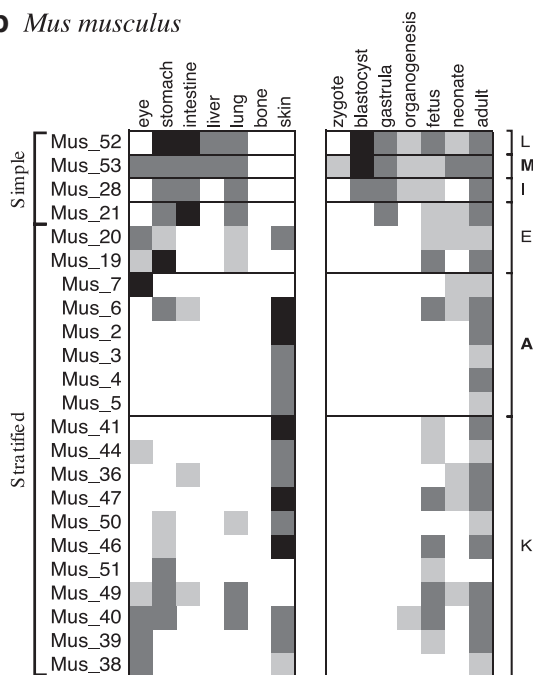
a *Silurana tropicalis*

b *Mus musculus*

FIG. 5. Keratin gene expression patterns in frog and mouse. EST counts were used to estimate the keratin expression level in a variety of tissues (left) and developmental stages (right) from the tropical clawed frog *Silurana tropicalis* (*a*) and mouse *Mus musculus* (*b*). Box shading reflects expression level and is consistent with the transcripts per million values, from white to black: 0–10; 10–50; 51–500; and >500. Letters on the right correspond to the clades defined in figures 1–3. Gene names correspond to those in figures 2 and 3 and supplementary table 1 (Supplementary Material online).

preparation of the adult life stage (metamorphosis in amphibians; fetus and neonate in mammals) and is often confined to the skin or its appendages (fig. 5). These ancestral keratin genes were the primers to allow further differential ecological radiations in amphibians, sauropsids, and mammals.

Amphibian keratin genes radiations mainly belong to clade A and group K (figs. 1–3). Our gene expression survey

indicates that some of them are expressed in the head of tadpoles and in early developmental stages (fig. 5, clade A, Sil_9 to Sil_12), whereas the rest are dominantly expressed in the skin of metamorphosing or adult frogs (fig. 5, remaining groups A and K). The former might be involved in the development of larval beaks, one of the most keratinized and variable structures of amphibian tadpoles (Altig and McDiarmid 1999), but detailed expression information on amphibians is largely lacking.

The largest keratin radiations in sauropsids belong to clade F and group K (figs. 1–3). These genes form an unexpected diversity of alpha keratins in addition to the already large amount of sauropsid-specific beta keratins. In type I, we find strong support for a sauropsid-specific alpha keratin gene clade with orthologs in mammals and amphibians (fig. 2, clade F). The two gecko keratins in this clade were discovered in adhesive toe pads (Hallahan et al. 2009), whereas the sole human ortholog functions almost exclusively in the reinforcement of hand palms and foot soles (Swensson et al. 1998; Moll et al. 2008). Moreover, this clade also includes six sauropsid cysteine-rich keratins (four anole and two chicken keratins) that are considered important in hard keratin structures (Eckhart et al. 2008). This suggests that keratins of clade F may have served protection to friction at least from stem amniotes onward. Finally, our phylogenetic reconstructions indicate a larger amount of hair keratins (HK) in anoles than in birds, likely as a result of duplications in the lineage to Squamata (fig. 2: D; fig. 3: J). On the contrary, beta keratin duplications occurred more frequently in birds than in reptiles (Greenwold and Sawyer 2010, 2011), indicating that beta keratins may have played a compensational role for the lack of alpha HK in the formation of hard appendages in birds.

In mammals, keratin gene radiations (fig. 2: A, D, and G; fig. 3: J and K) correspond to keratins that are mainly expressed in the hair structure, that is, outer root sheath, inner root sheath, and HK (figs. 2 and 3) (Moll et al. 2008). Until now, HK were considered to be amniote specific because nonmammalian HK were known only from sauropsid claws and scales (Eckhart et al. 2008). Here, our phylogenetic analyses show strong support for the presence of an ortholog in amphibians in both types, indicating that type I and type II HK already originated in stem tetrapods. In newts, type II HK are restricted to the distal part of normal and regenerating limbs and tails (Ferretti et al. 1991). This combined evidence of expression in extant tetrapods is suggestive of an expression in the distal part of epidermal appendages that must have been important during the conquest of land.

The prevalent hypothesis on the origin of hairs is that they evolved from sensory appendages situated in the hinge region of scales, after the divergence of sauropsids from mammals (Maderson 2003; Chang et al. 2009). Our study, however, shows that the main components of hair existed long before the origin of hairs and suggests an alternative scenario on the origin of hair. In line with findings of Dhouailly (2009), our results suggest that hair may have

originated from glandular alpha-keratinzed bumps in stem tetrapods, which evolved into the glandular unit and the dermal papilla of the hair structure in mammals. The considerable keratin gene repertoire in stem tetrapods revealed in this study reflects the ability to form these heavily keratined bumps. Additionally, most keratins that originated in stem mammals are involved in the formation of hair follicles, associated glands, and hair fibers.

Our study shows that keratin genes radiated extensively during the ecological water-to-land transition of tetrapods. The expansion of both their keratin clusters was essential in acquiring new functions in skin or appendages as an adaptation to the enhanced friction imposed by the new terrestrial lifestyle. Furthermore, subsequent asymmetrical radiation of these genes allowed each of the tetrapod classes to further develop specific adaptations. Radiations and functional shifts in protein families generally played an important role in plants and animals. The parallel characterization of gene radiations in multiple protein families is a strong tool toward a better understanding of some of the major evolutionary transitions.

Supplementary Material

Supplementary table 1 and figure 1 are available at *Molecular Biology and Evolution* online (http://www.mbe. oxfordjournals.org/).

Acknowledgments

References

Alibardi L. 2003. Adaptation to the land: the skin of reptiles in comparison to that of amphibians and endotherm amniotes. *J Exp Zool B Mol Dev Evol.* 298:12–41.

Alibardi L. 2006. Structural and immunocytochemical characterization of keratinization in vertebrate epidermis and epidermal derivatives. *Int Rev Cytol.* 253:177–259.

Alibardi L. 2009. Embryonic keratinization in vertebrates in relation to land colonization. *Acta Zool.* 90:1–17.

Alibardi L, Jaeger K, Valle LD, Eckhart L. 2011. Ultrastructural localization of hair keratin homologs in the claw of the lizard Anolis carolinensis. *J Morphol.* 272:363–370.

Altig R, McDiarmid RW. 1999. Tadpoles: the biology of anuran larvae. Chicago (IL): University of Chicago Press.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol.* 268:78–94.

Chang C, Wu P, Baker RE, Maini PK, Alibardi L, Chuong CM. 2009. Reptile scale paradigm: Evo-Devo, pattern formation and regeneration. *Int J Dev Biol.* 53:813–826.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol.* 7:429–447.

Chuong CM, Homberger DG. 2003. Development and evolution of the amniote integument: current landscape and future horizon. *J Exp Zool B Mol Dev Evol.* 298:1–11.

Dhouailly D. 2009. A new scenario for the evolutionary origin of hair, feather, and avian scales. *J Anat.* 214:587–606.

Eckhart L, Valle LD, Jaeger K, Ballaun C, Szabo S, Nardi A, Buchberger M, Hermann M, Alibardi L, Tschachler E. 2008. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. *Proc Natl Acad Sci U S A.* 105:18419–18423.

Ferretti P, Brockes JP, Brown R. 1991. A newt type II keratin restricted to normal and regenerating limbs and tails is responsive to retinoic acid. *Development* 111:497–507.

Greenwold MJ, Sawyer RH. 2010. Genomic organization and molecular phylogenies of the beta (β) 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. 2011. Linking the molecular evolution of avian beta (β) keratins to the evolution of feathers. *J Exp Zool B Mol Dev Evol.* 316B:609–616.

Hahn MW. 2007. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. *Genome Biol.* 8:R141.

Hallahan DL, Keiper-Hrynko NM, Shang TQ, Ganzke TS, Toni M, Dalla Valle L, Alibardi L. 2009. Analysis of gene expression in gecko digital adhesive pads indicates significant production of cysteine- and glycine-rich beta-keratins. *J Exp Zool B Mol Dev Evol.* 312:58–73.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledgebase of divergence times among organisms. *Bioinformatics* 22:2971–2972.

Jaillon O, Aury JM, Brunet F, et al. (61 co-authors). 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.

Krushna Padhi B, Akimenko MA, Ekker M. 2006. Independent expansion of the keratin gene family in teleostean fish and mammals: an insight from phylogenetic analysis and radiation hybrid mapping of keratin genes in zebrafish. *Gene* 368:37–45.

Kurokawa I, Takahashi K, Moll I, Moll R. 2011. Expression of keratins in cutaneous epithelial tumors and related disorders—distribution and clinical significance. *Exp Dermatol.* 20:217–228.

Langbein L, Rogers MA, Praetzel-Wunder S, Böckler D, Schirmacher P, Schweizer J. 2007. Novel type I hair keratins K39 and K40 are the last to be expressed in differentiation of the hair: completion of the human hair keratin catalog. *J Invest Dermatol.* 127:1532–1535.

Maddison DR, Maddison WP. 2000. MacClade 4: Analysis of phylogeny and character evolution. Sunderland (MA): Sinauer Associates.

Maderson PF. 2003. Mammalian skin evolution: a reevaluation. *Exp Dermatol.* 12:233–236.

Moll R, Divo M, Langbein L. 2008. The human keratins: biology and pathology. *Histochem Cell Biol.* 129:705–733.

Olsen V, Hyatt AD, Boyle DG, Mendez D. 2004. Co-localisation of Batrachochytrium dendrobatidis and keratin for enhanced diagnosis of chytridiomycosis in frogs. *Dis Aquat Organ.* 61:85–88.

Owens DW, Lane EB. 2003. The quest for the function of simple epithelial keratins. *Bioessays* 25:748–758.

Perrin C, Langbein L, Schweizer J. 2004. Expression of hair keratins in the adult nail unit: an immunohistochemical analysis of the onychogenesis in the proximal nail fold, matrix and nail bed. *Br J Dermatol.* 151:362–371.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

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. *Integr Comp Biol.* 40:530–539.

Schaffeld M, Bremer M, Hunzinger C, Markl J. 2005. Evolution of tissue-specific keratins as deduced from novel cDNA sequences of the lungfish Protopterus aethiopicus. *Eur J Cell Biol.* 84:363–377.

Schaffeld M, Haberkamp M, Braziulis E, Lieb B, Markl J. 2002a. Type II keratin cDNAs from the rainbow trout: implications for keratin evolution. *Differentiation* 70:292–299.

Schaffeld M, Haberkamp M, Schatzlein S, Neumann S, Hunzinger C. 2007. A novel and ancient group of type I keratins with members in bichir, sturgeon and gar. *Front Zool.* 4:16.

Schaffeld M, Hoffling S, Haberkamp M, Conrad M, Markl J. 2002b. Type I keratin cDNAs from the rainbow trout: independent radiation of keratins in fish. *Differentiation* 70:282–291.

Schaffeld M, Hoffling S, Jurgen M. 2004. Sequence, evolution and tissue expression patterns of an epidermal type I keratin from the shark Scyliorhinus stellaris. *Eur J Cell Biol.* 83:359–368.

Schaffeld M, Lobbecke A, Lieb B, Markl J. 1998. Tracing keratin evolution: catalog, expression patterns and primary structure of shark (Scyliorhinus stellaris) keratins. *Eur J Cell Biol.* 77:69–80.

Schaffeld M, Schultess J. 2006. Genes coding for intermediate filament proteins closely related to the hagfish "thread keratins (TK)" alpha and gamma also exist in lamprey, teleosts and amphibians. *Exp Cell Res.* 312:1447–1462.

Schweizer J, Bowden PE, Coulombe PA, et al. (11 co-authors). 2006. New consensus nomenclature for mammalian keratins. *J Cell Biol.* 174:169–174.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 57:758–771.

Swensson O, Langbein L, McMillan JR, Stevens HP, Leigh IM, McLean WH, Lane EB, Eady RA. 1998. Specialized keratin expression pattern in human ridged skin as an adaptation to high physical stress. *Br J Dermatol.* 139:767–775.

Szewciw LJ, de Kerckhove DG, Grime GW, Fudge DS. 2010. Calcification provides mechanical reinforcement to whale baleen alpha-keratin. *Proc Biol Sci.* 277:2597–2605.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.

Wilson BW, Edwards KJ, Sleigh MJ, Byrne CR, Ward KA. 1988. Complete sequence of a type-I microfibrillar wool keratin gene. *Gene* 73:21–31.

Wu DD, Irwin DM, Zhang YP. 2008. Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. *BMC Evol Biol.* 8:241.

Wu P, Hou L, Plikus M, Hughes M, Scehnet J, Suksaweang S, Widelitz R, Jiang TX, Chuong CM. 2004. Evo-Devo of amniote integuments and appendages. *Int J Dev Biol.* 48:249–270.

Zimek A, Weber K. 2005. Terrestrial vertebrates have two keratin gene clusters; striking differences in teleost fish. *Eur J Cell Biol.* 84:623–635.