**Subtracted cDNA library construction**

RNA was extracted from tissues by TRI reagent (Sigma Aldrich), followed up by an extra purification using the RNAqueous-micro kit (Ambion). Suppression subtractive hybridization (SSH) was performed using a SMART PCR cDNA Synthesis Kit and a PCR-Select Subtractive Hybridization Kit (Clontech). We created a subtracted cDNA library of genes expressed in the toe pad (complete distal part) versus genes expressed in the same digits minus the toe pad of the tree frog *Hyla cinerea*. Amplification products were cloned using a pGEM-T Easy cloning vector (Promega) and vectors were transformed into One Shot GeneHogs Electrocompetent cells (Invitrogen). Colonies were picked randomly and inserts were amplified and cycle-sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit and visualized on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Expressed sequence tags (ESTs) of the toe pad enriched library were assembled into unique contigs using Codoncode Aligner (Codoncode corporation) and screened using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Relevant toe pad ESTs were extended by rapid amplification of cDNA ends (RACE) PCR, using cDNA created with Smart PCR cDNA Synthesis Kit (Clontech) (see above) as a PCR template, with one gene specific and one adaptor-specific primer. All primers used are shown in electronic supplementary material, S2.

**Semi-quantitative RT-PCR reactions**

We used a semi-quantitative RT-PCRs to determine tissue specificity of newly found toe pad keratins. Various tissues and organs, dissected from *H. cinerea* and *Xenopus laevis*, were stored in RNAlater (Sigma). RNA was extracted using TRI reagent (Sigma), followed by the RNAqueous-micro kit (Ambion). In-between, a Dnase I treatment was used to avoid DNA contamination. cDNA was synthesized with a SuperScript III First-Strand Synthesis System (Invitrogen) according to the manufacturer's instructions. Oligo dT primers were used to create the first strand cDNA. The amount of the RNA starting material was determined spectrometrically and normalized at 250 ng total RNA per cDNA reaction. 0.5 µl of cDNA was used in a 25 µl amplification reaction with FastStart TaqDNA Polymerase (Roche). Cycles were determined to fall within the exponential phase of amplification and amplification products were analysed by agarose-gel electrophoresis. The expression levels of the keratin genes were standardized with that of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). All primers used are shown in electronic supplementary material, S2.
**Sequence alignment and phylogenetic reconstructions**

Tetrapod keratins from lamprey (*Lampetra fluviatilis*), pufferfish (*Tetraodon nigroviridis*), tropical clawed frog (*Silurana tropicalis*), chicken (*Gallus gallus*), anole (*Anolis carolinensis*), platypus (*Ornithorhynchus anatinus*), opossum (*Monodelphis domestica*), mouse (*Mus musculus*) and human (*Homo sapiens*) were derived from the protein database of ‘National Center for Biotechnology Information’ and from genomic screening of public libraries described in [1]. Lamprey, shark, fish, lungfish, and some tetrapod keratins (e.g. K8 & K18) that are shown to have diverged earlier in keratin evolution were used as outgroup in this study [1, 2].

Keratin datasets were aligned using ClustalX v1.81 [3] and manually corrected in MacClade 4.06 [4]. Phylogenetic relationships and branch support were estimated using three different model-based methods. First, Bayesian posterior probabilities for branch support (BPPs) were estimated using MrBayes 3.1.2 [5] 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 and 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 burn-in and to check the convergence of the MCMC runs. Second, heuristic Maximum likelihood (ML) searches were performed by RAxML [6]. Clade support was assessed by performing 200 “rapid” bootstrap replicates. Third, phylogenetic relationships were estimated by the metapopulation genetic algorithm implemented in MetaPIGA v2 [7], using probability consensus pruning among four populations of four individuals each. The posterior probability distribution of possible trees was estimated by performing replicated MetaGA searches that were stopped when the mean relative error (MRE) among ten consecutive consensus trees remained under five percent. Because the Bayesian analyses converged on the JTT+G+I model of AA substitution using a mixed model prior, we applied this model in both RAxML and MetaPIGA.

The three methods (Figure 1 and electronic supplementary material, S5) resulted in similar consensus trees and support the finding that toe pad keratins did not originate by a series of gene duplications within amphibians, but from ancestral tetrapod keratins that were later differently used in amphibians and mammals. MetaPIGA yielded overall lower branch support (electronic supplementary material, S5) than the other two methods, potentially due to the low number of sites compared to the high number of taxa analysed.
References


