

Research

Open Access

## A novel and ancient group of type I keratins with members in bichir, sturgeon and gar

Michael Schaffeld\*<sup>1</sup>, Mark Haberkamp<sup>1</sup>, Sonja Schätzlein<sup>2</sup>,  
Sebastian Neumann<sup>1</sup> and Christian Hunzinger<sup>3</sup>

Address: <sup>1</sup>Institute of Zoology, Johannes-von-Müller-Weg 6, Johannes Gutenberg University, D-55099 Mainz, Germany, <sup>2</sup>Dept. of Gastroenterology, Medical School Hannover, Carl Neuberg Str. 1, K11, E01, R1400, 30629 Hannover, Germany and <sup>3</sup>Merck KGaA, Central Services Analytics, Central Product Analytics/Bioanalytics, Frankfurter Str. 250, D-64293 Darmstadt, Germany

Email: Michael Schaffeld\* - schaffeld@uni-mainz.de; Mark Haberkamp - haberkam@uni-mainz.de; Sonja Schätzlein - schatzlein.sonja@mh-hannover.de; Sebastian Neumann - sneumann@students.uni-mainz.de; Christian Hunzinger - christian.hunzinger@merck.de

\* Corresponding author

Published: 6 June 2007

Received: 22 December 2006

Frontiers in Zoology 2007, 4:16 doi:10.1186/1742-9994-4-16

Accepted: 6 June 2007

This article is available from: <http://www.frontiersinzoology.com/content/4/1/16>

© 2007 Schaffeld et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**1. Background:** Vertebrate epithelial cells typically express a specific set of keratins. In teleosts, keratins are also present in a variety of mesenchymal cells, which usually express vimentin. Significantly, our previous studies revealed that virtually all known teleost keratins evolved independently from those present in terrestrial vertebrates. To further elucidate the evolutionary scenario that led to the large variety of keratins and their complex expression patterns in present day teleosts, we have investigated their presence in bichir, sturgeon and gar.

**2. Results:** We have discovered a novel group of type I keratins with members in all three of these ancient ray-finned fish, but apparently no counterparts are present in any other vertebrate class so far investigated, including the modern teleost fish. From sturgeon and gar we sequenced one and from bichir two members of this novel keratin group. By complementary keratin blot-binding assays and peptide mass fingerprinting using MALDI-TOF mass spectrometry, in sturgeon we were able to assign the sequence to a prominent protein spot, present exclusively in a two-dimensionally separated cytoskeletal preparation of skin, thus identifying it as an epidermally expressed type I keratin. In contrast to the other keratins we have so far sequenced from bichir, sturgeon and gar, these new sequences occupy a rather basal position within the phylogenetic tree of type I keratins, in a close vicinity to the keratins we previously cloned from river lamprey.

**3. Conclusion:** Thus, this new K14 group seem to belong to a very ancient keratin branch, whose functional role has still to be further elucidated. Furthermore, the exclusive presence of this keratin group in bichir, sturgeon and gar points to the close phylogenetic relationship of these ray-finned fish, an issue still under debate among taxonomists.

### Background

In vertebrates the cytoskeleton of epithelial cell types is typically reinforced by a specific set of type I and type II keratins that assemble into 10 nm thick filaments formed

from type I/II heterodimers. The keratins are members of the large multigene family of intermediate filament proteins (IF proteins) of which they form by far the most complex group. In human, 53 of the hitherto nearly 70

identified IF protein genes code for keratins [1-4] that are expressed in tissue and developmental specific patterns. Without including human hair and nail forming keratins, the number of keratin genes found in teleost fish is comparably high, but in contrast to human and other tetrapods, teleost fish possess a large excess of type I keratin genes [5-7]. By analysing the molecular evolution of keratins, as well as the evolution of their expression patterns in lower vertebrates, we want to further elucidate the scenario and probable evolutionary forces that led to this extraordinary variety of keratins in vertebrates.

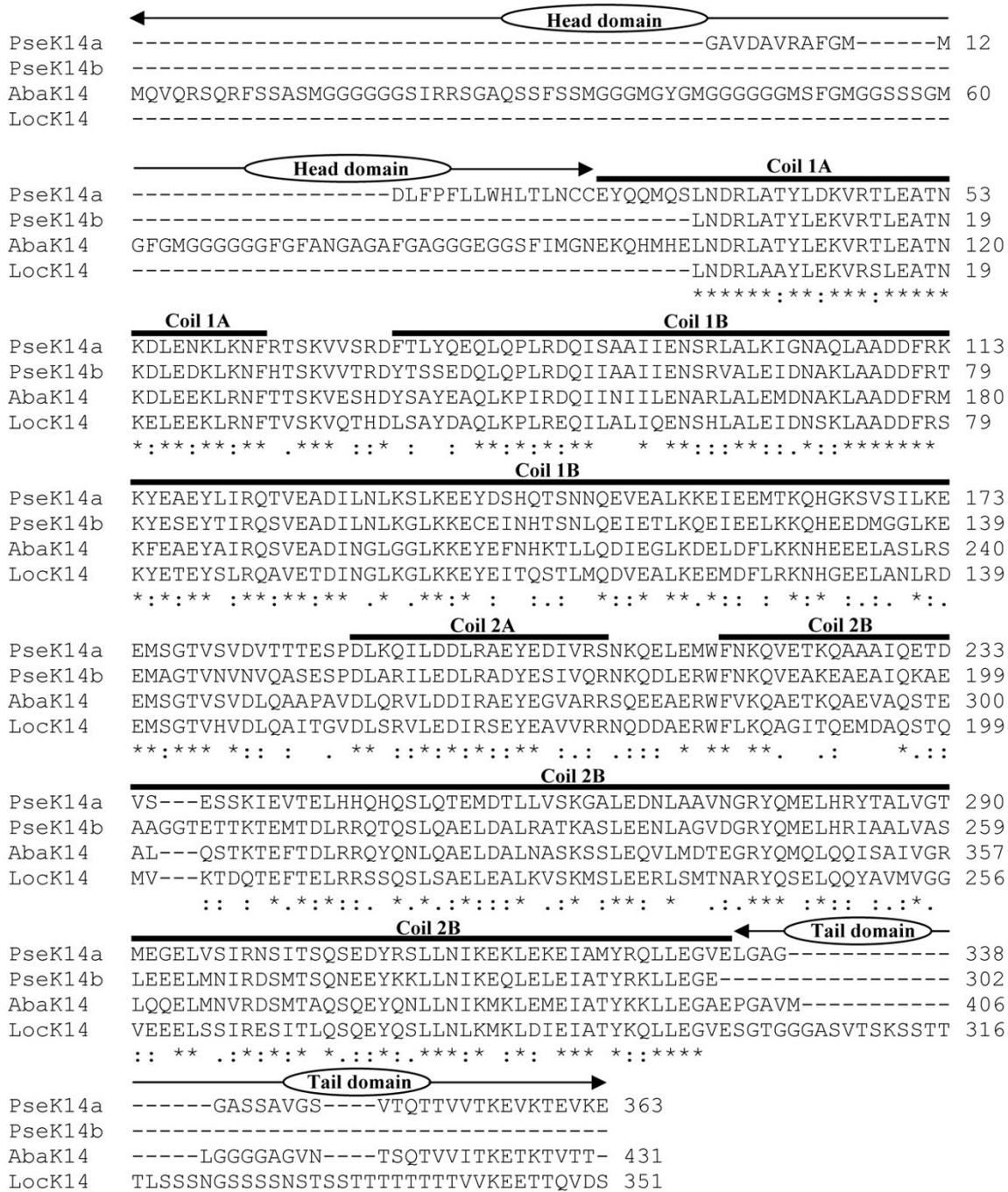
Our investigations of the keratin systems in lamprey, shark, trout, zebrafish, carp, goldfish and lungfish have so far revealed that type I and type II keratins are apparently present in all classes of vertebrates and that the various keratins can generally be subdivided into the "E" keratins, expressed in epidermal keratinocytes and other stratified epithelia, and those appearing in cells forming simple epithelia, thus named "S" keratins [8-19]. Nevertheless, our data based on cDNA sequence analysis, followed by thorough phylogenetic analyses [13-18] as well as the studies based on the recently available genome data from man and teleost fish [1,2,5-7], strongly support the view of largely independent origins of the keratin genes found in fish and man. According to our present data, solely the typical "S" keratin pair K8 and K18 can at least be found in all gnathostomian vertebrate groups, indicating the unique and general importance of these "ancient" keratins. In contrast to other vertebrates investigated so far, in modern teleost fish keratins, including K8 and K18, in addition to their typical IF epithelial appearance show a widespread IF occurrence in mesenchymally derived cells and tissues, such as fibroblasts, chondrocytes and blood vessel endothelia (for review see [11]). The latter in the non-teleost vertebrates usually do not express keratins but the type III IF protein vimentin [20-26]. To further trace the origin of the different "E" and "S" keratins as well as the evolution of the mesenchymal keratin expression in teleosts, we have investigated the keratin systems in a bichir, a sturgeon and a gar that are believed to represent the most ancient groups of the extant ray-finned fish. In the course of these studies, from all three species we obtained sequences that apparently belong to a novel branch of type I keratins, without counterparts in any other vertebrate group investigated so far, including the teleost fish. Here we present and discuss their sequences as well as their phylogenetically relationships to the other members of the type I keratin subfamily, which may also provide clues to the early evolution of ray-finned fish. The latter is still strongly debated among taxonomists, whether on the basis of molecular or morphological data (for an overview see [27]).

## Results and discussion

### Novel type I keratin sequences from bichir, sturgeon and gar

Only recently we have discovered that a rather ancient and distant group of keratin-related sequences, the extracellularly secreted thread keratins TK and TK<sub>2</sub>, are not only present in hagfish (the assumedly most ancient vertebrate group), but also in lamprey, teleosts and amphibians. This provided major clues relating to keratin evolution in vertebrates, but also pointed to a more general role of this previously considered highly specific IF protein group in vertebrates [28]. By combination of RT-PCR experiments and cDNA library screening (for details see Methods), from sturgeon and gar we have now isolated one and from bichir two cDNA sequences that, according to our phylogenetic analysis, code for members of another novel keratin group (Fig. 1, Table 1). The 1858 bp long cDNA clone we isolated from the sturgeon cDNA library (*abak14*; [EMBL: [AJ493259](#)]) contains the complete coding sequence for a type I keratin of 46759Da (431 amino acids) and a calculated pI of 5.1, which we now term *AbaK14* (from *Acipenser baeri* keratin). However, the corresponding sequences we have so far obtained from bichir and gar are still incomplete. A 1369 bp long clone (*pseK14a*; [EMBL: [AM419452](#)]) isolated from the bichir cDNA library encodes a type I keratin that we term *PseK14a* (*Polypterus senegalus* keratin). It still lacks a portion of its head encoding sequence in addition to its 5' UTR. By RT-PCR using degenerate primers, from bichir we additionally recovered a 908 bp long cDNA sequence (*pseK14b*; [EMBL: [AM419453](#)]), comprising almost the complete rod encoding segment of a second K14 counterpart in this species (*PseK14b*). In addition, we found five further incomplete cDNA clones that apparently encode different variants of *PseK14a* (not shown here). The latter only slightly vary in DNA sequence, from 0.1 – 3.1% (amino acid variance of 0.4 – 4.9%). In a similar way we were also able to amplify a cDNA fragment encoding the rod domain of a K14 counterpart in gar, which we term *LockK14* (from *Lepisosteus oculatus* keratin). By RACE-PCR we additionally recovered its tail encoding sequence and 3' UTR. Its assembled sequence (*lock14*; [EMBL: [AM419454](#)]) overall comprises 1207 bp, but still lacks the complete head encoding segment in addition to the section coding for the first seven residues of the rod domain.

Subsequent mining of the available genome and EST databases for K14 counterparts in other vertebrates such as teleosts, amphibians, birds and mammals, so far has not yet yielded any matches, suggesting that this keratin group may only be present in the ancient groups of ray-finned fish. We only found two matches encoding K14 of another sturgeon, notably *Acipenser transmontanus* (white sturgeon; [EMBL: [DR975435](#), [DR975694](#)]), which both stem from a skin-derived cDNA library.



**Figure 1**  
**Sequence comparison of K14 from bichir, sturgeon and gar.** Multiple alignment of the keratin 14 (K14) sequences we obtained from bichir, sturgeon and gar. Thick black lines mark the four helical subdomains (coils 1A to 2B), which are typical for the central rod domain of all known IF-proteins. Asterisks indicate identical amino acids; double dots indicate a high and single dots a lower degree of amino acid conservation. Pse, *Polypterus senegalus* (bichir); Aba, *Acipenser baeri* (sturgeon); Loc, *Lepisosteus oculatus* (gar). Note that only AbaK14 comprises the complete amino acid sequence. From PseK14a at least a section and from LocK14 the complete head sequence is still missing. From PseK14b we still have to recover both, the complete head and tail sequence.

**Table 1: Properties of the isolated cDNA clones encoding K14 from bichir, sturgeon and gar.**

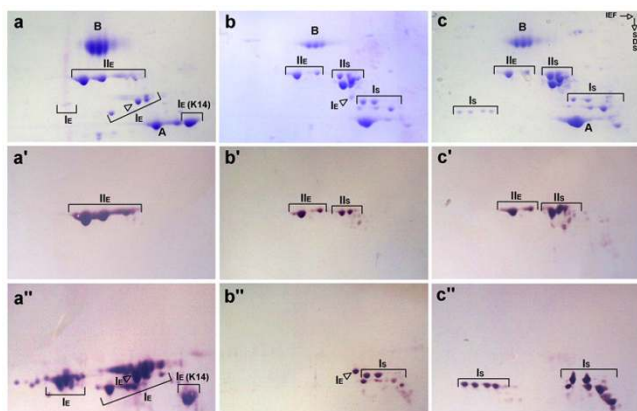
cDNA clone	Keratin	EMBL accession number	Size of cDNA (bp)	Number of encoded amino acids	Mr (Da)	pI
<i>abaK14</i>	AbaK14	<a href="#">AJ493259</a>	1858	431	46759 <sup>4)</sup>	5.1 <sup>4)</sup>
<i>pseK14a</i>	PseK14a	<a href="#">AM419452</a>	1369	363 <sup>1)</sup>	-	-
<i>pseK14b</i>	PseK14b	<a href="#">AM419453</a>	908	302 <sup>2)</sup>	-	-
<i>locK14</i>	LocK14	<a href="#">AM419454</a>	1207	351 <sup>3)</sup>	-	-

Pse, *Polypterus senegalus* (bichir); Aba, *Acipenser baeri* (sturgeon); Loc, *Lepisosteus oculatus* (gar); <sup>1)</sup> a portion of the head domain is still missing; <sup>2)</sup> head and tail domain are still missing; <sup>3)</sup> complete head domain is still missing; <sup>4)</sup> as predicted from the cDNA sequence.

**Biochemical identification of K14 in sturgeon**

To analyse the general set of keratins expressed in sturgeon, we extracted the cytoskeletal proteins from different tissues, including skin, liver, intestine, stomach and gill, separated them by 2D-PAGE and subsequently analysed the patterns by CKBB assays and immunoblotting (results for skin, stomach and intestine are shown in Fig. 2). The major spots were additionally analysed by peptide mass fingerprinting (PMF) to reveal similarities between identified keratin spots and to assign them to the sequences we obtained from the sturgeon by cDNA library screening. In the course of these investigations we were clearly able to

assign a single protein spot to the amino acid sequence derived from the cloned *abak14* cDNA sequence. The matching spot was solely found as a major component in the cytoskeletal preparation of skin (Fig. 2a), in which it was firmly identified as a type I keratin in the CKBB assay (Fig. 2a"). It showed a positive reaction with the anti-trout-keratin antiserum GPpoly (data not shown), which we previously introduced as a general keratin marker in fish [8,10-15,19]. The two-dimensional position of this protein spot fits the theoretical values calculated from the sequence of AbaK14 (see above and Table 1). Peptide mass fingerprint (PMF) analysis of this protein yielded 29 matching peptide masses for AbaK14, of which 26 were specific for AbaK14 in comparison to the other six type I keratins we so far sequenced from sturgeon. Overall amino acid sequence coverage for AbaK14 was 61%.



**Figure 2**  
**Biochemical identification of K14 in sturgeon.** 2D-PAGE of cytoskeletal proteins extracted from sturgeon skin (a), stomach (b) and intestine (c). Isoelectric focusing (IEF) was used in the first dimension, in the second dimension we applied SDS-PAGE. Bovine serum albumin (B) and rabbit -actin (A) were added to the samples as marker proteins. E, epidermal keratins; S, simple keratins. (a-c) Coomassie Blue stained gels, showing the polypeptide patterns, which were subsequently used for complementary keratin blot-binding (CKBB) assays. (a'-c') CKBB test employing biotinylated human keratin K18, thereby identifying type II keratins (II). (a''-c'') CKBB assay using biotinylated human keratin K8 which identifies type I keratins (I). Note that the K14 spot is not present as a major component in intestine or stomach and that both tissues express a mixture of E and S keratins.

**Molecular evolution of keratins – the ancient origin of K14**

To infer the phylogenetic relationships of the novel K14 group to the other currently known type I keratins, by different methods we thoroughly analysed a comprehensive data set of 118 polypeptides, including the type I keratins from lancelet, lamprey, shark, bichir, sturgeon, gar, zebrafish, trout, *Xenopus*, lungfish and man (for further details see Methods; the available accession numbers of the employed sequences are listed in Fig. 3). Most of the fish sequences stem from our own data, notably those from lamprey, shark, bichir, sturgeon, gar, trout and lungfish in addition to K8 and K18 from zebrafish [13-18,23,29,30]. We rooted the trees with the type I keratin sequences available from the lancelets, which are believed to represent the most ancient group of living chordates [31]. In general we received the same basal tree topology when applying Neighbor Joining (NJ), Maximum Likelihood (ML) or Bayesian (B) methods for our phylogenetic analysis. The phylogenetic tree shown in Fig. 4 is based on Bayesian inference and clearly shows that the K14 sequences form a separate and basal branch within the type I keratins, phylogenetically close to the sequences we obtained from the river lamprey. They even branch off prior to the twig formed by the gnathostomian K18 sequences, that apparently emerged before the separation of cartilaginous and bony fish [13-15]. This ancestral ori-

gin of the K14 twig together with the assumedly exclusive presence of K14 in ancient ray-finned fish (see above) raises the question as to the complementary binding partner(s) of this keratin group. So far, analysis of bichir and sturgeon has not revealed a group of type II keratins occupying a similar basal position equivalent to the K14 sequences in the type I keratin tree [8,29,30]. Moreover, the suggested restriction of K14 sequences to bichir, sturgeon and gar indicates the close phylogenetic relationship of these fish groups, an issue still under debate among taxonomists. Based on recent molecular data it can be concluded that sturgeons, gars and bowfin together may form

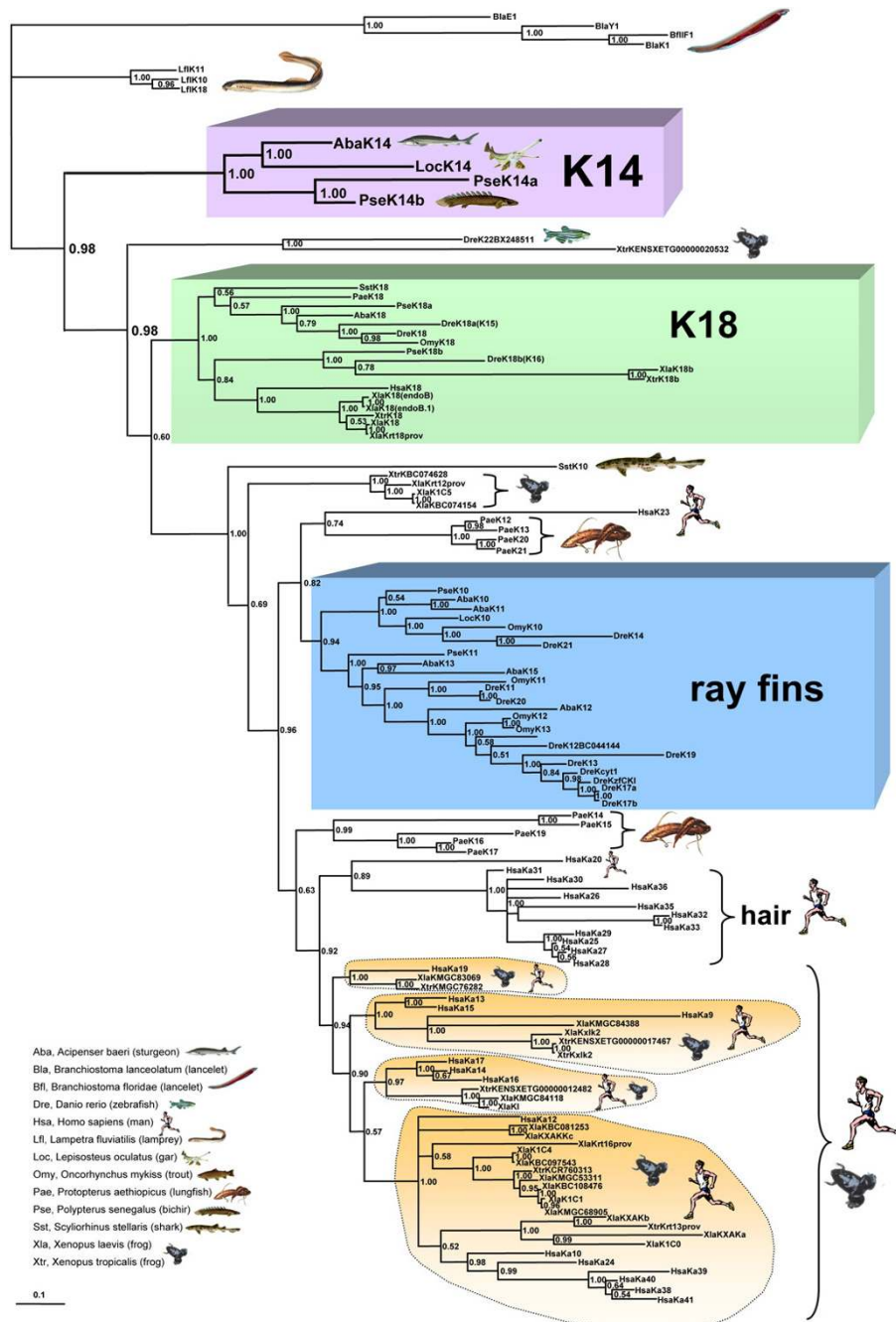
a sister group to the teleost fish and that bichirs represent the most ancient group of ray-finned fish (for review see [27]). Since our current studies do not contribute to the resolution of this problem, the keratin system in bowfin should be investigated, which might shed more light on this issue.

Furthermore, the phylogenetic tree illustrated in Fig. 4 suggests that most of the type I keratins from ray-finned fish evolved independently from those present in lungfish, frog or man and that very early in actinopterygian evolution a gene duplication gave rise to at least two dif-

Name	Accession	Species	Name	Accession	Species
AbaK10	AJ493255	<i>Acipenser baeri</i>	Omy-K10	AJ272372	<i>Oncorhynchus mykiss</i>
AbaK11	AJ493256		Omy-K11	AJ272371	
AbaK12	AJ493257		Omy-K12	AJ427868	
AbaK13	AJ493258		Omy-K13	AJ427867	
AbaK14	AJ493259		Omy-K18	Y14289	
AbaK15	AJ493260		Pae-K12	AJ785786	<i>Protopterus aethiopicus</i>
AbaK18	AJ493261		Pae-K13	AJ785787	
BflIF1	AF108192	<i>Branchiostoma floridae</i>	Pae-K14	AJ785788	
BlaE1	AJ010294	<i>Branchiostoma lanceolatum</i>	Pae-K15	AJ785789	
BlaK1	AJ245426		Pae-K16	AJ785790	
BlaY1	AJ245428		Pae-K17	AJ785791	
DreKcyt1	AF084461	<i>Danio rerio</i>	Pae-K18	AJ785792	
DreKzfCKI	AF197880		Pae-K19	AJ785793	
DreK11	BC075874		Pae-K20	AJ785799	
DreK12	BC044144		Pae-K21	AJ785785	
DreK13	BC076059		PseK10	AM419450	<i>Polypterus senegalus</i>
DreK14	BC076485		PseK11	AM419451	
DreK18	AJ493269		PseK14a	AM419452	
DreK18a(K15)	BC066541		PseK14b	AM419453	
DreK18b(K16)	BC078359		PseK18a	AM419448	
DreK17a	BC115108		PseK18b	AM419449	
DreK17b	BC092718		Sst-K10	AJ623268	<i>Scyliorhinus stellaris</i>
DreK19	BC097213		Sst-K18	Y14647	
DreK20	AL645755		Xla-K18	BC054993	<i>Xenopus laevis</i>
DreK21	BC115321		XlaKK18(endoB.1)	BC072305	
DreK22	BX248511		XlaK18(endoB)	Y00230	
HsaKa9	X75015	<i>Homo sapiens</i>	XlaKrt18-prov	BC054993	
HsaKa10	X14487		XlaK18b	BC106430	
HsaKa12	D78367		Xla-K1C0 (XL51)	Y00968	
HsaKa13	X14640		Xla-K1C1 (XK81A1)	X04804	
HsaKa14	BC042437		Xla-KMGC68905	BC055996	
HsaKa15	X07696		Xla-K1C4 (XK81B2)	M18155/ X04807	
HsaKa16	AF061812		Xla-K1C5	M11032	
HsaKa17	Z19574		Xla-KXAKa	AB045600	
HsaKa18	X12881		Xla-KXAKb	AB045601	
HsaKa19	Y00503		Xla-KXAKc	AB086829	
HsaKa20	X73501		XlaKMGC83069	BC070682	
HsaKa23	AF102848		XlaKMGC84118	BC074309	
HsaKa24	AK000268		XlaKMGC84388	BC093567	
HsaKa25	Y16787		XlaKMGC53311	BC042927	
HsaKa26	X90761		XlaKBC081253	BC081253	
HsaKa27	AJ633621		XlaKBC108476	BC108476	
HsaKa28	Y16789		XlaKBC074154	BC074154	
HsaKa29	Y16790		XlaKxIk2	AB218824	
HsaKa30	Y16791		XlaKrt16-prov	BC077649	
HsaKa31	Y16792		XlaKI	BC084340	
HsaKa32	Y16793		XlaKBC097543	BC097543	
HsaKa33	Y16793		XlaKrt1-2-prov	BC045031	
HsaKa35	BK004054		XtrK18	BC061366	<i>Xenopus tropicalis</i>
HsaKa36	BK004055		XtrK18b	BJ072726/Dr860147	
HsaKa38	AJ564204		XtrKENSXETG00000017467	ENSXETG00000017467*	
HsaKa39	AJ564205		XtrKENSXETG00000012482	ENSXETG00000012482*	
HsaKa40	AJ564206		XtrKENSXETG00000020532	ENSXETG00000020532*	
HsaKa41	AJ564207		XtrKxIk2	BC081376	
LocK10	AM419763	<i>Lepisosteus oculatus</i>	XtrKrt13-prov	BC087788	
LocK12	AM419764		XtrKBC074628	BC074628	
LocK14	AM419454		XtrKMGC76282	BC061624	
Lfl-K10	AJ308116	<i>Lampetra fluviatilis</i>	XtrKCR760313	CR760313	
Lfl-K11	AJ308117				
Lfl-K18	AJ308118				

**Figure 3**

**Accession numbers.** EMBL accession numbers of the type I keratin sequences we used for phylogenetic inference. \*For three *Xenopus tropicalis* type I keratin sequences the Ensembl database gene IDs are given.



**Figure 4**  
**Molecular evolution of type I keratins.** Phylogenetic tree based on Bayesian inference, illustrating the relationships of the K14 sequences from bichir, sturgeon and gar to the other type I keratins known from vertebrates. The tree was rooted with the lancelet type I keratin sequences. It clearly shows that the K14 sequences form a separate branch (boxed in violet) close to the sequences we cloned from the river lamprey. They even branch off prior to the twig formed by the gnathostomian K18 sequences (boxed in green) that apparently emerged before the separation of cartilaginous and bony fish. The tree, furthermore, suggests that most of the ray-finned fish type I keratins (boxed in blue) evolved independently from those present in lungfish, frog or man and that early in actinopterygian evolution gene duplications already gave rise to at least two different type I keratin groups with members in both, ancient and modern ray-finned fish. Importantly, within the tetrapod lineage the Bayesian analysis revealed four highly supported keratin subgroups, each with members in both, frog and man (encircled by dotted lines and coloured orange). Bar, 0.1 substitutions per site.

ferent type I keratin branches, each with members in both ancient and modern ray-finned fish (see the boxed ray-fins twig in Fig. 4). Compared to our recent analyses [14-18,28], we have now included additional type I keratin sequences from zebrafish and *Xenopus*; therefore, in this study the K18 branch shows a more complex branching pattern, not allowing a clear identification of the authentic K18 counterparts in the different vertebrates solely on the basis of their position in the tree. However, our previous identification of K18 in shark, bichir, sturgeon, trout and zebrafish was additionally based on its typical occurrence in simple epithelia, such as liver hepatocytes or intestinal mucosal epithelium, corresponding to the situation in man and other tetrapods [10,12-16,19,29]. Nevertheless, these "functional K18 counterparts" identified in the different vertebrate groups have to be considered as paralogous protein sequences. The branching pattern of the K18 twig supports the suggestion that early as well as more recent duplication events led to the various K18 related genes. Future analysis of their expression patterns may provide further clues for possible functions of these K18 relatives.

Compared to the evolution of type II keratins, the phylogenetic tree of type I keratins appears rather complex and several nodes cannot be resolved. Moreover, in teleosts the number of detected type I keratin genes virtually triples the number of those coding for type II keratins [5-7], in contrast to the almost equal number of type I and type II keratin genes detected in the tetrapod genomes. However, when including the hair- and nail-forming keratins the total number of keratin genes in tetrapods is clearly higher than in teleost fish [1-4]. A group of two sequences, one from zebrafish, the other from *Xenopus*, also occupy a rather basal position in the type I keratin tree, directly branching off between the K14 and K18 twig. But it has to be taken into consideration that both sequences stem from genomic DNA sequencing and may only represent non-active "relics" of ancient keratin genes. From the current Bayesian phylogenetic analysis (but not from our NJ and ML analysis) another phenomenon emerged for the first time: There are now four highly supported keratin subgroups in the tetrapod twig, each with members in both, frog and man (in Fig. 4 encircled by dotted lines). Moreover, the tree topology indicates that the human hair keratins may have emerged prior to these radiation events.

## Conclusion

The type I and type II keratins represent the two most complex and abundant groups of intermediate filament (IF) proteins among vertebrates and their structure, function and extraordinary variety cannot be completely understood without phylogenetic considerations. Therefore, we investigated the keratin systems in more ancient representatives of the vertebrate lineage. In particular, the

keratins present in skin may play an important role in the transition of vertebrates from water- to land- living animals. Here we present a novel group of epidermal type I keratins, which we termed keratin 14 and so far have only found them in the basal groups of living ray-finned fish, notably bichir, sturgeon and gar. Our phylogenetic analysis revealed a rather basal position of this keratin group in the tree of type I keratin evolution. The keratin 14 group even emerged prior to the gnathostomian K18 sequences, which together with its binding partner K8 and with the exception of the recently discovered thread keratins in teleosts and amphibians, was hitherto considered as most ancient group of gnathostomian keratins. Future analyses will hopefully shed more light on the expression and functional role of K14 in the fish epidermis and clarify the identity of its type II keratin binding partner.

## Methods

### Preparation of tissues and cytoskeletal proteins

Sturgeons (*Acipenser baeri*) were purchased from a local hatching farm (Fischzucht Rhönforelle, Gersfeld, Germany), and gars (*Lepisosteus oculatus*) from a local pet shop (Fauna Exotica, Mainz-Kostheim, Germany). The bichirs (*Polypterus senegalus*) were a gift from Dr. Latz (Institute of Zoology, University of Mainz). Within the scope of this study we employed one specimen for each species. Animals were killed by cutting the neck or the tail artery after MS 222 narcosis (0.5 g/l). For further procedures, the tissues were excised and used immediately or snap frozen according to [8]. Cytoskeletal proteins were extracted as described in [8].

### Electrophoresis, immunoblotting, CKBB, PMF and immunofluorescence microscopy

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), complementary keratin blot-binding (CKBB) assays, Western blotting and indirect immunofluorescence microscopy were essentially performed as described in [8]. For immunoblotting we used 10% milk powder as the blocking reagent and the antibody incubations were performed overnight at 8°C. Peptide mass fingerprinting (PMF) using matrix-assisted laserdesorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed as described in [16,18]. The mass tolerance for matching fragments was set to 100 ppm.

### Preparation of RNA

Total RNA from bichir was extracted according to a protocol modified from [32], including a GTC extraction and subsequent sedimentation of RNA by ultracentrifugation through a dense cushion of caesium chloride (detailed protocol given in [8]). For purification of sturgeon RNA we applied a protocol that included consecutive steps of guanidinium thiocyanate (GTC) homogenizations, acidic phenol/chloroform extractions and precipitations. The

final RNA precipitation was accomplished with 8 M LiCl. Gar RNA was prepared using the GeneMATRIX Universal RNA Purification Kit purchased from Roboklon (Germany) according to the suppliers' instructions. Isolation of mRNA was generally performed with the "PolyAtract®-mRNA isolation system" from Promega.

#### **RT-PCR analyses**

Applying 0.5 to 1.0 µg of total RNA, RT-PCR was performed either using the Superscript II reverse transcriptase from Invitrogen (RT for 1 h at 42°C), followed by standard PCR (Taq Polymerase, Invitrogen) or the QIAGEN® OneStep RT-PCR Kit (RT for 0.5 h at 50°C). PCR was done for 30–35 cycles and primer specific annealing temperatures. We used different combinations of degenerate, IF-specific primers, as previously described [8] (notably primer numbers P4, P5, P8, P9, P10). 3' RACE-PCR was accomplished using an oligodT-19mer as downstream and sequence-specific oligos as upstream primers. For gel extraction of PCR products we employed the QIAquick Gel Extraction Kit (Qiagen) or the GeneMATRIX DNA Purification Kit AGAROSE-OUT (Roboklon). Cloning of the isolated DNA fragments was accomplished either with the TOPO-TA Cloning Kit (Invitrogen), the StrataClone™ PCR Cloning Kit (Stratagene) or the pGEM-T Easy vector kit (Promega). All kits were used according to the instruction manual. Primers were synthesized by Roth and Sigma-Ark, respectively. Nucleotide sequencing was performed on both strands using the Taq Dye Deoxy Terminator system. The subsequent gel run was done by commercial services (Centerprise, Germany).

#### **cDNA library construction and screening**

According to the supplier's instructions we constructed - phage cDNA libraries (ZAP-Express®, Stratagene) from sturgeon and bichir, respectively, in each case using 5 µg of mRNA purified from a mixture of tissues, including skin, eyes, brain and internal organs. Clones were isolated using the fish keratin-specific antiserum GPpoly [8] and digoxigenin-labeled cDNA probes derived from obtained RT-PCR fragments. Digoxigenin labeling was either performed using the DIG-High Prime Kit from Roche, according to the instruction manual or by standard PCR, using digoxigenin labeled nucleotides from Roche.

#### **Sequence analyses and phylogenetic inference**

For database searches, we employed the Basic Local Alignment Search Tool (BLAST [33]) of the National Center for Biotechnology Information (NCBI) and of the Ensembl Genome Browser. The multiple sequence alignments were performed with ClustalX version 1.8 [34] using default gap penalties. When necessary, the alignment was edited by hand. The final alignment was analysed by Neighbor Joining (NJ), Maximum Likelihood (ML) and Bayesian methods (B). All analyses were conducted under the

Jones-Taylor-Thornton substitution model of amino acid evolution (JTT [35]). Furthermore, the model applied for the ML and Bayesian analyses included observed amino acid frequencies (F), estimated proportion of invariant sites (I), and estimation of among-site rate variation for the remaining sites according to a gamma distribution (G) that was set to four rate categories. For the NJ analysis we used the programs PROTDIST and NEIGHBOR of the Phylogeny Inference Package (PHYLIP, version 3.6 b [36]). The reliability of the tree topology was then tested by bootstrap analysis [37] with 100 replications, using the PHYLIP programs seqboot, protdist, neighbor and consensus. The ML analysis was conducted with the PHYLIP-like interface PHYML [38]. For bootstrapping we generated 100 pseudo data sets. The Bayesian inference analysis was performed with MRBAYES 3.1 [39] with the frequencies fixed to the Jones frequencies. In each of the two parallel runs four Markov chains (one hot and three cold chains) were run simultaneously for 3,000,000 generations, starting with random trees. Sampling from the trees was set to every 10th generation. Under these conditions the average value for the deviation of split frequencies reached a value < 0.01. The burn-in was set to 115000, based on the stationary phase. All consensus trees were drawn using TREEVIEW version 1.6.6 [40].

#### **Abbreviations**

CKBB Complementary keratin blot-binding

MALDI Matrix-assisted laser desorption/ionisation

TOF Time-of-flight

MS Mass spectrometry

PAGE Polyacrylamide gel electrophoresis

PMF Peptide mass fingerprint

2D Two-dimensional

#### **Competing interests**

The author(s) declare that they have no competing interests.

#### **Authors' contributions**

Mark Haberkamp performed the sequencing work in sturgeon and part of the biochemical work in sturgeon, including the analysis of the provided MALDI-TOF data. Sonja Schätzlein did most of the cloning experiments in bichir. All experiments were accomplished while both authors were members of our group in Mainz. Sebastian Neumann sequenced K14 from gar. Christian Hunzinger performed the peptide mass fingerprint experiments and assisted the analysis of the MALDI-TOF data while he was



in the employ of ProteoSys AG, Carl-Zeiss-Str. 51, 55129 Mainz, Germany. Michael Schaffeld supervised the conception of experiments, carried out part of the molecular and biochemical studies, performed the complete phylogenetic analyses, including sequence alignment and tree reconstruction, and wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Prof. Dr. Harald Herrmann for providing the recombinant human keratins 8 and 18, Simon Höffling for additional sequencing work, Nicole Tappe for additional biochemical experiments and Prof. Dr. J. Robin Harris for proof reading of the manuscript. We especially thank Prof. Dr. Jürgen Markl for initiating and continuously supporting this work and critical reading of the manuscript. Financially aided by a grant to J.M. from the Deutsche Forschungsgemeinschaft (MA 843/5-1).

## References

- Hesse M, Magin TM, Weber K: **Genes for intermediate filament proteins and the draft sequence of the human genome: novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18.** *J Cell Sci* 2001, **114**:2569-2575.
- Hesse M, Zimek A, Weber K, Magin TM: **Comprehensive analysis of keratin gene clusters in humans and rodents.** *Eur J Cell Biol* 2004, **83**:19-26.
- Rogers MA, Winter H, Langbein L, Bleiler R, Schweizer J: **The human type I keratin gene family: characterization of new hair follicle specific members and evaluation of the chromosome 17q21.2 gene domain.** *Differentiation* 2004, **72**:527-540.
- Rogers MA, Edler L, Winter H, Langbein L, Beckmann I, Schweizer J: **Characterization of new members of the human type II keratin gene family and a general evaluation of the keratin gene domain on chromosome 12q13.13.** *J Invest Dermatol* 2005, **124**:536-544.
- Zimek A, Stick R, Weber K: **Genes coding for intermediate filament proteins: common features and unexpected differences in the genomes of humans and the teleost fish *Fugu rubripes*.** *J Cell Sci* 2003, **116**:2295-2302.
- Zimek A, Weber K: **Terrestrial vertebrates have two keratin gene clusters; striking differences in teleost fish.** *Eur J Cell Biol* 2005, **84**:623-635.
- Krushna Padhi B, Akimenko MA, Ekker M: **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* 2006, **368**:37-45.
- Schaffeld M, Markl J: **Fish keratins.** *Methods Cell Biol* 2004, **78**:627-671.
- Markl J, Franke WW: **Localization of cytokeratins in tissues of the rainbow trout: fundamental differences in expression pattern between fish and higher vertebrates.** *Differentiation* 1988, **39**:97-122.
- Markl J, Winter S, Franke WW: **The catalog and the expression complexity of cytokeratins in a teleost fish, the rainbow trout.** *Eur J Cell Biol* 1989, **50**:1-16.
- Markl J, Schechter N: **Fish intermediate filament proteins in structure, function and evolution.** *Subcell Biochem* 1998, **31**:1-33.
- Conrad M, Lemb K, Schubert T, Markl J: **Biochemical identification and tissue-specific expression patterns of keratins in the zebrafish *Danio rerio*.** *Cell Tissue Res* 1998, **293**:195-205.
- Schaffeld M, Löbbbecke A, Lieb B, Markl J: **Tracing keratin evolution: Catalog, expression patterns and primary structure of shark (*Scyliorhinus stellaris*) keratins.** *Eur J Cell Biol* 1998, **77**:69-80.
- Schaffeld M, Haberkamp M, Brazilius E, Lieb B, Markl J: **Type II keratin cDNAs from the rainbow trout: Implications for keratin evolution.** *Differentiation* 2002, **70**(6):292-299.
- Schaffeld M, Höffling S, Haberkamp M, Conrad M, Markl J: **Type I keratin cDNAs from the rainbow trout: independent radiation of keratins in fish.** *Differentiation* 2002, **70**(6):282-291.
- Schaffeld M, Knappe M, Hunzinger C, Markl J: **cDNA sequences of the authentic keratins 8 and 18 in zebrafish.** *Differentiation* 2003, **71**:73-82.
- Schaffeld M, Höffling S, Markl J: **Sequence, evolution and tissue expression patterns of an epidermal type I keratin from the shark *Scyliorhinus stellaris*.** *Eur J Cell Biol* 2004, **83**:359-368.
- Schaffeld M, Bremer M, Hunzinger C, Markl J: **Evolution of tissue-specific keratins as deduced from novel cDNA sequences of the lungfish *Protopterus aethiopicus*.** *Eur J Cell Biol* 2005, **84**:363-377.
- García DM, Bauer H, Dietz T, Schubert T, Markl J, Schaffeld M: **Identification of keratins and analysis of their expression in carp and goldfish: comparison with the zebrafish and trout keratin catalog.** *Cell Tissue Res* 2005, **322**(2):245-56.
- Herrmann H, Fouquet B, Franke WW: **Expression of intermediate filament proteins during development of *Xenopus laevis*. I. cDNA clones encoding different forms of vimentin.** *Development* 1989, **105**:279-298.
- Lazarides E: **Intermediate filaments: a chemically heterogeneous, developmentally regulated class of proteins.** *Annu Rev Biochem* 1982, **51**:219-250.
- Schaffeld M, Herrmann H, Schultess J, Markl J: **Vimentin and desmin of a cartilaginous fish, the shark *Scyliorhinus stellaris*: Sequence, expression patterns and in vitro assembly.** *Eur J Cell Biol* 2001, **80**:692-702.
- Schultess J: **Molekulare Evolution der Intermediärfilament-Proteine des Flussneunauges *Lampetra fluviatilis*.** In *Phd thesis* University of Mainz, Faculty of Biology; 2001.
- Schultess J, Schaffeld M, Markl J: **Intermediate filament proteins type I to III in the lamprey *Lampetra fluviatilis*: Searching for the origin of the IF protein evolution in vertebrates.** *Zoology* 2000, **103**(Suppl.III):51.
- Schultess J, Schaffeld M, Markl J: **Intermediate filament protein sequences of the cyclostome *Lampetra fluviatilis*.** *Biol Cell* 2001, **93**:235. (First Joint French-German Congress on Cell Biology)
- Schultess J, Schaffeld M, Scheinert I, Markl J: **Intermediate filament protein sequences of the cyclostome *Lampetra fluviatilis*.** *Zoology* 2001, **104**(Suppl.IV):36.
- Zardoya R, Meyer A: **Recent advances in the (molecular) phylogeny of vertebrates.** *Annu Rev Ecol Syst* 2003, **34**:311-338.
- Schaffeld M, Schultess J: **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* 2006, **312**:1447-1462.
- Schaffeld M, Schultess J, Haberkamp M, Bremer M, Markl J: **Intermediate Filament Protein Evolution in Fish: Sequences from lamprey, shark, bichir, sturgeon and trout.** *Biol Cell* 2001, **93**:235. (First Joint French-German Congress on Cell Biology)
- Haberkamp M: **Zur Evolution des Cytoskeletts beim Sibirischen Stör *Acipenser baeri*.** In *Phd thesis* University of Mainz, Faculty of Biology; 2002.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**:965-8.
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: **Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease.** *Biochemistry* 1979, **18**:5294-5299.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
- Jones DT, Taylor VWR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8**:275-282.
- Felsenstein J: **PHYLIP (phylogeny inference package). Version 3.6b.** Edited by: . Distributed by the author, Department of Genetics, University of Washington, Seattle; 2004.
- Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
- Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
- Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
- Page RDM: **TREEVIEW: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357-358.