Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Cell Biology 84 (2005) 623–635
European Journal of Cell Biology
www.elsevier.de/ejcb

Terrestrial vertebrates have two keratin gene clusters; striking differences in teleost fish

Alexander Zimek, Klaus Weber*

Department of Biochemistry and Cell Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

Received 16 December 2004; received in revised form 25 January 2005; accepted 25 January 2005

Abstract

Keratins I and II form the largest subgroups of mammalian intermediate filament (IF) proteins and account as obligatory heteropolymers for the keratin filaments of epithelia. All human type I genes except for the K18 gene are clustered on chromosome 17q21, while all type II genes form a cluster on chromosome 12q13, that ends with the type I gene K18. Highly related keratin gene clusters are found in rat and mouse. Since fish seem to lack a keratin II cluster we screened the recently established draft genomes of a bird (chicken) and an amphibian (*Xenopus*). The results show that keratin I and II gene clusters are a feature of all terrestrial vertebrates. Because hair with its multiple hair keratins and inner root sheath keratins is a mammalian acquisition, the keratin gene clusters of chicken and *Xenopus tropicalis* have only about half the number of genes found in mammals. Within the type I clusters all genes have the same orientation. In type II clusters there is a rare gene of opposite orientation. Finally we show that the genes for keratins 8 and 18, which are the first expression pair in embryology, are not only adjacent in mammals, but also in *Xenopus* and three different fish. Thus neighboring K8 and K18 genes seem a feature shared by all vertebrates. In contrast to the two well defined keratin gene clusters of terrestrial vertebrates, three teleost fish show an excess of type I over type II genes, the lack of a keratin type II gene cluster and a striking dispersal of type I genes, that are probably the result of the teleost-specific whole genome duplication followed by a massive gene loss. This raises the question whether keratin gene clusters extend beyond the ancestral bony vertebrate to cartilage fish and lamprey. We also analyzed the complement of non-keratin IF genes of the chicken. Surprisingly, an additional nuclear lamin gene, previously overlooked by cDNA cloning, is documented on chromosome 10. The two splice variants closely resemble the lamin LIII a + b of amphibia and fish. This lamin gene is lost on the mammalian lineage.

Keywords: Amphibia; Birds; Fish; Intermediate filaments; Keratin gene clusters; Lamin genes; Mammals; Teleosts; Vertebrate genomes; Zebrafish

Introduction

In mammals the family of genes encoding the structural proteins of the cytoplasmic intermediate filaments (IFs) has nearly 70 members and is one of the 100 largest multigene families (Hesse et al., 2001,

*Corresponding author. Tel.: +49 551 201 1486; fax: +49 551 201 1578.
E-mail address: office.weber@mpibpc.gwdg.de (K. Weber).

0171-9335/$ - see front matter © 2005 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ejcb.2005.01.007

2004). Sequence identity levels of IF proteins, their biochemical properties, the organization of the corresponding genes and their expression patterns define several IF subtypes (Coulombe et al., 2001; Fuchs and Weber, 1994; Herrmann and Aebi, 2000). Type I and II keratins are by far the largest subfamilies. They give rise to the epithelial keratin filaments that are based on obligate heteromeric double-stranded coiled coils formed by a type I and a type II keratin. Type III covers four proteins able to form homopolymeric IF, which are mesenchymally derived. The genes for the seven type IV proteins show an entirely different intron pattern than do type I to III genes. The nuclear lamins form the type V, whereas the eye lens proteins filensin and phakinin constitute a separate group. Point mutations in a still growing number of IF genes are connected with human diseases. Mutations in at least 14 epidermal keratin genes cause skin fragility syndromes and demonstrate that IF protect against mechanical stress (Irvine and McLean, 1999).

Type I to IV IF proteins are also known for birds (Sato and Yasugi, 1997; Vanhoutteghem et al., 2004), amphibia and fish (see (Schaffeld et al., 2002a, b)). Type I to III genes are not restricted to vertebrates, but have also been documented in the early chordates (Karabinos et al., 2002, 2004a; Wang et al., 2002). Cepholochordate and urochordates seem, however, to lack type IV genes. In protostomic animals the cytoplasmic IF proteins show a striking relation to nuclear lamins. They keep the increased length of the coil 1b domain and often display additionally a lamin homology region in the carboxy-terminal tail domain (Dodemont et al., 1990; Erber et al., 1998; Weber et al., 1989). Nearly all of the 11 IF genes of *Caenorhabditis elegans* are essential for normal nematode development (Karabinos et al., 2001, 2003, 2004b). Interestingly, the only sequence of a cytoplasmic IF protein of a lower deuterostomic species, the hemichordate *Saccoglossus*, is more closely related to protostomic IF proteins than to IF proteins of the chordates (Zimek and Weber, 2002).

A striking feature of mammalian keratin genes is their chromosomal clustering documented so far for man, mouse and rat. In man, all 27 type I keratin genes except for K18, that is expressed in interior epithelia, are clustered on chromosome 17q21 where they are arranged in the same orientation. Similarly, the 27 type II keratin genes form a cluster on chromosome 12q13 which ends with the K18 gene, a type I keratin (Hesse et al., 2004). Similar keratin clusters are not known outside the mammals.

Since the early chordate *Ciona intestinalis* has only one keratin I and one keratin II gene (Karabinos et al., 2004a) keratin gene clusters probably evolved with the vertebrates. Unexpectedly, the preliminary gene complement of the teleost fish *Fugu rubripes* indicates a sizeable excess of type I over type II keratin genes and implies the absence of a keratin II gene cluster (Zimek et al., 2003).

To resolve the question whether keratin gene clusters exist in vertebrate classes other than the mammals, we have now analyzed the recently established genomic sequence database for chicken, as a representative of the birds, and the amphibian *Xenopus tropicalis*. Our search establishes a keratin type I cluster on chicken chromosome 27 and a type II cluster of yet unknown chromosomal location. The smaller number of keratin genes for each chicken and *Xenopus* gene cluster seems explained by the fact that hair keratins are a mammalian acquisition. The overall similarity between mammalian and non-mammalian keratin clusters is illustrated by the nearly perfect unidirectionality of the genes in the clusters. However, the genome drafts of three teleost fish show remarkable changes in keratin gene distribution versus the terrestrial vertebrates.

Materials and methods

We analyzed the genome of two vertebrate species, *Xenopus tropicalis*, the western clawed frog, and *Gallus gallus*, the chicken. The chicken draft assembly by the Genome Sequencing Center at Washington University School of Medicine in St. Louis was produced in February 2004. The assembly has a 6.63 X coverage, 88% of the genome are mapped to the chromosomes 1–24, 26–28, 32 and the gonosomes W and Z. The remaining sequences are assembled in the virtual chromosome "Un". Version 3.0 of the *Xenopus* whole genome shotgun assembly is available since October 2004. Sequencing and assembly are performed by the DOE Joint Genome Institute (JGI). The raw sequence data with a 7.4 X coverage is assembled into 27,064 scaffolds without chromosomal mapping. Roughly half of the genome is covered by 392 scaffolds with at least 1.2 Mbps of length.

Both genomes are available in the UCSC genome browser (http://genome.ucsc.edu) and can be searched with the BLAT alignment tool. Starting with keratin sequences from chicken (GK-19: AB016281; gKb12: AY574984) and *Xenopus* (xKa18: BC061366; xKa56: BC074628) the keratin type I and II clusters were located in the genome and completely analyzed with the GCG Package Version 10.3 (http://www.accelrys.com/products/gcg_wisconsin_package). We aligned the identified keratin sequences with the "pileup" program and checked the predicted protein sequences for obvious variations in the conserved IF patterns. This was usually a hint for misassembled sequences. In some cases gaps in the genomic assembly prevented the identification of exons. Sometimes the 3′ parts of the IF genes, coding for the tail domain of the protein, could not

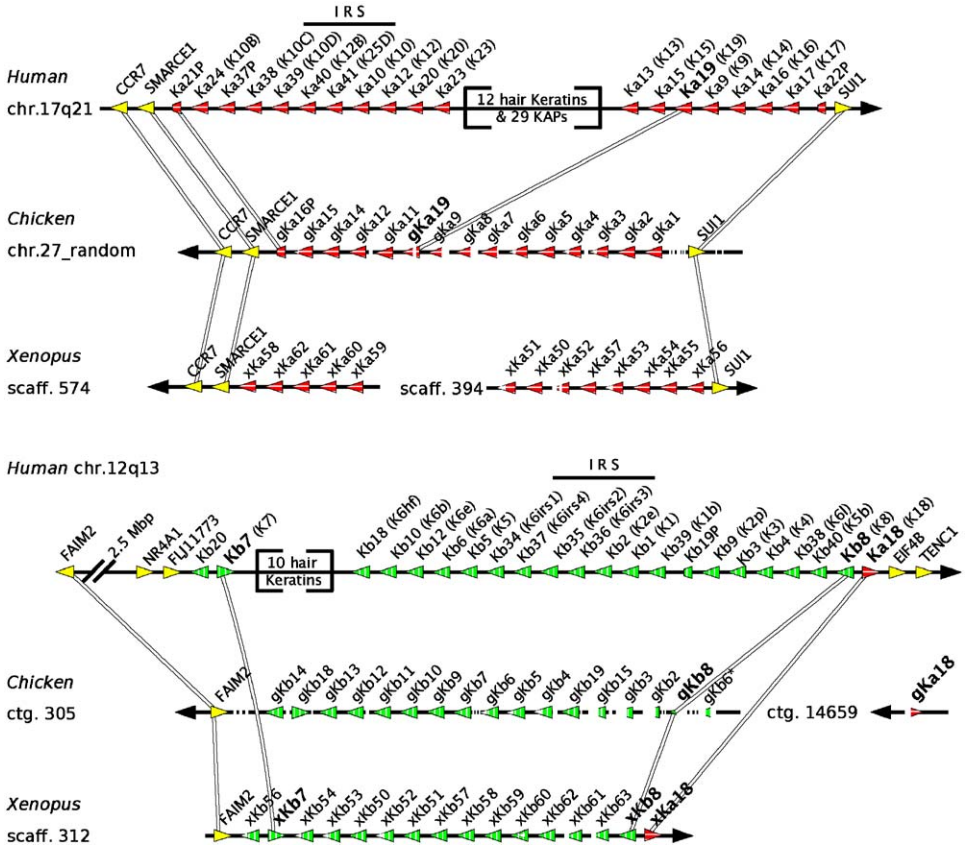

Fig. 1. Comparison of keratin gene clusters in man, chicken and *Xenopus tropicalis*. Type I keratin genes are marked as red triangles, type II keratins as green triangles and non-keratin genes as yellow triangles. The tips of the triangles correspond to the 3′ end of the genes. Pseudogenes are drawn with a blunted triangle. Long black arrows display the 5′–3′ orientation of the DNA strand. Rectangular brackets replace the genes for human hair keratins and keratin-associated proteins (KAPs). The mammalian-specific inner root sheath keratins (IRS) are marked by a line above the gene symbols. Gaps in the genome assembly are indicated by white spaces. Keratin genes with unidentified tails or heads have white spaces inside the triangle (see text and Tables 1 and 2). Names for human keratins follow the proposed nomenclature for keratins (Hesse et al., 2004). Numbers for the chicken and *Xenopus* keratins do not imply homology versus the mammalian keratins (see text) except when printed in bold type. The *Xenopus* type I gene cluster is covered by two scaffolds as indicated. For individual non-mammalian keratin genes, see Tables 1 and 2. Non-keratin genes (yellow) are like their human orthologs: CCR7: chemokine (C–C motif) receptor 7 (BC035343); SMARCE1: SWI/SNF-related matrix-associated actin-dependent regulator of chromatin member 1 (NM_003079); SUI1: putative translation initiation factor (NM_005801); FAIM2: Fas apoptotic inhibitory molecule 2 (NM_012306); NR4A1: nuclear receptor subfamily 4 group A member 1 (NM_002135); FLJ11773 (AK021835); EIF4B: eukaryotic translation initiation factor 4B (NM_001417); TENC1 tensin-like C1 containing phosphatase (NM_015319) and neurodg1 prognosticated neural developmental protein (NM_214794). Note that the non-mammalian keratin type I and II gene clusters have the same neighboring genes as the mammalian clusters (Hesse et al., 2004). Conservation of synteny is indicated by the gray lines connecting corresponding human/chicken/*Xenopus* genes. Note also the strong reduction in keratin genes for avian and amphibian gene clusters compared to their mammalian counterparts (see text).

be identified due to high sequence variability (Fig. 1; Tables 1, 2). We extracted 18 complete and 15 incomplete keratin genes from the chicken genome and 23 keratin genes in the frog with 6 genes supported by published cDNAs.

We consulted additionally two fish genomes. The Zv4 assembly (September 2004) of the zebrafish (*Danio rerio*) is available at the Ensemble Genome Browser (http://www.ensembl.org/Danio_rerio). The spotted green pufferfish (*Tetraodon nigroviridis*) V7 assembly (February 2004) is available at http://genome.ucsc.edu.

Results

Keratin genes

We searched the 2004 assembly of the chicken and *Xenopus tropicalis* genomes for keratin type I and II genes. Following the recent terminology on mammalian keratins (Hesse et al., 2004) type I keratins are abbreviated as Ka and type II keratins as Kb. The species is indicated by g (gallus gallus) and x (*Xenopus*), respectively. Since a full keratin gene expression pattern

Table 1. Chicken keratin I and II genes and their genomic localization (2004 assembly)

Type I	AC	Exons	Chr	Area	Contig	Gaps	Ori	Comment	Tissue
gKa18	BX935060	5–8	Un	45643382–45646524	14659.1 + 2	1–4	—	Complete cDNA	Trunk
gKa14		1–8	27_random	622283–626298	183.11		+	Complete	
gKa11	AY574986	1–8	27_random	596251–600632	183.12		+	Complete type I alpha-keratin15	Epid. keratinocytes
gKa2	AY574987	1–8	27_random	5195559–523433	183.18 + 19		+	Complete type I alpha-keratin 14	Epid. keratinocytes
gKa19	AB016281	1–8	27_random	588439–593288	183.12	3	+	Complete cDNA, GK-19,	Chicken embryonic gut
gKa1		1–8	27_random	512952–518757	183.20 + 19		+	Complete	
gKa12		1–8	27_random	603395–612734	183.11 + 12		+	Complete	
gKa5	bu285805	1–8	27_random	544534–549262	183.18		+	Complete	Brain, not cerebrum or cerebellum
gKa15		1–6	27_random	628528–632128	183.11		+	e7 + e8 not identifiable	
gKa16P	bu412444 bi393024	4–7	27_random	636420–641500	183.11		+	Exons 1–3 + 8 not identifiable; good Identity on nt.-level to human pseudogene Ka21P	Ovary, pituitary + pineal gland/ Hypothalamus
gKa4	bm491097	1–6	27_random	535346–538801	183.18		+	Exons 7 + 8 not identifiable	Pituitary + pineal gland/hypothalamus
gKa6		1–6	27_random	552745–555788	183.18		+	Exons 7 + 8 not identifiable	
gKa7		1–7	27_random	563493–567346	183.17		+	Exon 8 not identifiable	
gKa3	bu270429	1–6	27_random	527522–529932	183.18		+	Exons 7 + 8 not identifiable	Limbs
gKa9	BX932871	2–8	27_random	579006–581867	183.15 + 14	1	+		Ovary
gKa8	BU447565	2–8	27_random	570584–572700	183.16 + 15	1	+		Ovary
Type II									
gKb16	AY574985	1–9	Un	45197658–45202695	1784.2		+	Complete, type II alpha-keratin IIC	Epid. keratinocytes
gKb7		1–9	Un	7787521–7793215	305.22		+	Complete	
gKb9		1–9	Un	7807229–7815707	305.25–27		+	Complete	
gKb10		1–9	Un	7820122–7824990	305.27		+	Complete	
		1 + 3–6	Un	131758699–131760926	5096.1	2 + 7–9	+	Partial duplication	
gKb11	AY574983	1–9	Un	7829723–7835739	305.27		+	Complete, type II alpha-keratin IIA	Epid. keratinocytes
gKb12	AY574984	1–9	Un	7840872–7844313	305.28		+	Complete, type II alpha-keratin IIB	Epid. keratinocyte
gKb18	AF072698	1–8	Un	7861015–7870045	305.31–33		—	Complete, otokeratin (ear-specific cytokeratin)	Basilar Papilla
gKb14		1–9	Un	7876386–7890427	305.36		+	Complete	
gKb13	bm491222	1–9	Un	7852572–7855819	305.28		+	Complete	Pituitary + pineal gland/hypothalamus
gKb3		1–7	Un	7737027–7740103	305.15	8 + 9	+	Complete	
		2–7	Un	42244445–42246127	17066.1	1 + 8–9	+	Complementing	
		4–9	Un	42241158–42242928	17065.1	1–3	—	Duplications	
gKb6		1–7	Un	7774526–7778174	305.21		+	Exons 8 + 9 not identifiable	
		5–7	Un	7704407–7705950	305.1	1–4 + 8–9	+	Partial duplication	
gKb8	bu215235	6–8	Un	7711643–7713704	305.5 + 6	1–5	+		Whole embryo
gKb2		1	Un	7722455–7722958	305.9	2–9	+		
gKb4		1–3	Un	7745268–7746685	305.16		+		
gKb5		1–7	Un	7760067–7764637	305.18 + 19		+		
gKb15		1–2	Un	7731083–7732536	305.13	3–9	+		
gKb19		1–2	Un	7731083–7732536	305.13	3–9	+		

AC, accession number; Chr, chromosome; Ori, orientation; Un, unanchored contig.

Table 2. *Xenopus tropicalis* keratin I and II genes and their genomic localization (October 2004 assembly)

	AC	Exons	Scaff	Area	Gaps	Ori	Comment
Type I							
xKa18	BC06136	1–7	312	1346254–1349670		+	Complete
xKa50	BC081376	1–8	394	333629–341763		—	Complete
xKa54		1–8	394	432044–437316		—	Complete
xKa55		1–8	394	447552–451915		—	Complete
xKa56	BC074628	1–8	394	491943–501041		—	Complete
xKa57	BC061624	1–8	394	369710–380742		—	Complete
xKa58	cf344471	1–8	574	352117–356585		+	Complete
xKa59	cf345537	1–8	574	218144–226123		+	Complete
xKa61	BC087788	1–8	574	301687–308999		+	Complete
xKa60		1–4	574	284823–291893		+	Exons 5–8 unidentifiable
xKa62		1–3 + 6–7	574	314040–321029	4 + 5	+	Exon 8 unidentifiable
xKa63	cf152125	1–6	471	1007184–1018168		—	Exons 7–8 unidentifiable
xKa51		1–6	394	319926–324168		—	Exons 7–8 unidentifiable
xKa52		1–4 + 6	394	349019–354276	5	—	Exons 7–8 unidentifiable
xKa53		1–6	394	422254–429271		—	Exons 7–8 unidentifiable
Type II							
xKb50	cf225163		312	1055340–1064193		—	Complete
xKb51	cf524359		312	1083990–1092186		—	Complete
xKb57			312	1100751–1106747		—	Complete
xKb58			312	1108384–1120612		—	Complete
xKb59			312	1145623–1155872		—	Complete
xKb62	BC075443		312	1183287–1190094		—	Complete
xKb7	al676621		312	953386–958752		—	Complete
xKb8	BC067939	1–9	312	1278806–1287246		—	Complete
xKb60		1–7	312	1161265–1172520		—	Exons 8–9 unidentifiable
xKb61	cf343661	2–7	312	1204084–1207277	1,8,9	—	
xKb63	cf374565	1–5	312	1217960–1220145	6–9	—	
xKb52		1–7	312	1074469–1083989		—	Exons 8–9 unidentifiable
xKb53		1–7	312	1042148–1046205		—	Exons 8–9 unidentifiable
xKb54		1–7	312	989807–993577		—	Exons 8–9 unidentifiable
xKb56	cf591838	1–2 + 4–8	312	861485–884711	3 + 9	—	

AC, accession number; Scaff, scaffold; Ori, orientation.

is not available for chicken and *Xenopus*, a final nomenclature is not possible at this stage. Thus the numbers following Ka or Kb refer to our file numbers and do not imply a special relation to particular mammalian keratins. Exceptions are indicated by bold letter type in Tables 1 and 2 and Fig. 1. Thus Kb8 has a unique 20-residue signature at the C-terminal end which is highly conserved from fish to man (Schaffeld et al., 2003). Similarly gKa19 is well related to its mammalian counterparts by the protein sequence including the presence of a very short tail domain and by the expression pattern of the gene (Sato and Yasugi, 1997). Finally the assignment of Ka18 is supported by the observation that this is the only type I gene, which like its mammalian counterparts, lies outside the type I cluster. Our search for chicken keratin genes is summarized in Table 1, which provides the areas occupied by the genes, their orientation and where possible the chromosomal location. For several genes accession numbers of cDNAs (capital letters) or EST sequences (small letters) are provided by the database.

Identification of a keratin I cluster on chicken chromosome 27

All keratin I genes except one lie in a tight cluster on chromosome 27 of the chicken (Fig. 1, Table 1). The cluster of 16 genes extends over 129 kb and has a gene density of 8 kb. Like in its mammalian counterparts (Hesse et al., 2004) all keratin I genes have the same orientation. The non-keratin genes flanking the cluster are the same in mammals and chicken. On one side are the genes encoding CCR7 (C–C chemokine receptor type 7 precursor) and SMARCE1 (SW1/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily e, member 1), while the gene for SUI1

(translation initiation factor SuI1 isolog) occupies the other end.

There are two major differences between the mammalian and chicken type I cluster. First, the chicken type I cluster lacks the central subdomain of 362 kb in man and 466 kb in mouse, which contains the genes for 29 or 34 high/ultrahigh sulfur hair keratin-associated proteins (KAPs) (Hesse et al., 2004). These KAPs have cysteine contents of above (ultra high) or below (high) 30% and their mRNAs locate to the upper cortex of the hair shaft (Rogers et al., 2001). Screening of the entire chicken genome for such KAPs reveals no related proteins. Thus these KAPs seem a mammalian specialization. Second, the number of keratin I genes plus pseudogenes is reduced from 31 (man) to 16 (chicken). This seems primarily due to the lack of 12 genes and pseudogenes for hair keratins I and four genes for inner root sheath keratins I (Fig. 1), which can be considered to be an acquisition of the mammalian lineage. Interestingly, hair keratin I genes form a subdomain into which the KAP region is inserted (Hesse et al., 2004).

Since chicken keratins have been poorly studied in the past we have not tried to relate individual chicken and mammalian genes except in one case. The chicken gene labeled gKa19 reflects the mammalian Ka19 by both the predicted cDNA and protein sequences as well as its expression data already provided by Sato and Yasugi (1997). Interestingly in the mammalian type I cluster the Ka19 gene occupies a less central position than its counterpart in the chicken type I cluster (Fig. 1).

The first gene of the chicken type I cluster following the gene for SMARCE1 seems to be a pseudogene lacking exons 1 to 3 (Table 1; Fig. 1). This fits the observation that the first gene in the mammalian type I cluster is also a pseudogene (Ka21P) and indeed the sequences of the two pseudogenes are well related.

Only one chicken type I keratin gene (gKa18) lies outside the cluster on chromosome 27. It is located on a small contig of unknown chromosomal location (Table 1). This gene could be the counterpart of the mammalian keratin Ka18 gene, which is not contained in the type I cluster and flanks instead the type II cluster (Hesse et al., 2004). This speculation cannot be evaluated until the sequence data on the chicken type II cluster are extended in the future (see below).

Identification of the chicken keratin II gene cluster

A search of the genomic database identified 17 keratin II genes or gene fragments (Table 1). Of these 16 describe a tightly packed gene cluster (Fig. 1) while the other gene lies on a different contig. For the two contigs a chromosomal location is not yet available. In contrast to the type I cluster there are still many sequence gaps in the type II cluster and cDNA and EST sequences are only available for 6 genes. Three of the four cDNAs emerged from a study on epidermal keratinocytes (Vanhoutteghem et al., 2004). Fig. 1 compares the type II clusters of man and chicken. We still lack knowledge of the non-keratin genes flanking the chicken type II cluster (see Discussion). The number of genes plus pseudogenes is reduced from 31 in man to 17 in chicken. This reduction seems primarily due to the lack of 10 genes for hair keratins and four inner root sheath keratins (Fig. 1), that are an acquisition of the mammalian lineage. Strikingly in both clusters nearly all genes have the same orientation. An interesting exception is the second gene in both clusters (gKb18 for the chicken). In the human type II cluster additional genes of opposite polarity are one hair keratin gene and two hair keratin pseudogenes (Hesse et al., 2004). There are no non-keratin II genes in the cluster. This feature also holds for the murine type II cluster while the human cluster contains three pseudogenes for non-keratin proteins (Hesse et al., 2004).

Since expression patterns for most chicken keratin II genes are poorly documented we have not been able to convincingly relate the various chicken and human type II keratins. However, the chicken gene gKb8, although not completely provided by the genomic DNA, is readily defined by the unique amino acid sequence of some 20 residues located at the C-terminal end. This sequence feature is strictly conserved from fish to man (Schaffeld et al., 2003).

Two keratin gene clusters in the amphibian *Xenopus tropicalis*

When our analysis of chicken keratin genes was essentially complete, the draft genome of the amphibian *Xenopus tropicalis* became available (release October 2004). We therefore searched the draft for keratin genes (Table 2). The keratin I gene cluster is provided by contigs 574 and 394. Contig 574 shows the gene for SMARCE1 as in mammalian type I clusters followed by five type I genes in the same orientation (Fig. 1). Contig 394 provides eight type I genes in the same orientation and ends like mammalian type I gene clusters with the gene for SUI1. The gene for Ka18 lies as expected at the end of the type II cluster. The relation of the single type I gene on the large contig 471 to the type I cluster is currently unclear. Since the Ka18 gene is at the end of the type II cluster, the total number of keratin type I genes is at least 15.

The *Xenopus* keratin II gene cluster (Fig. 1) located on scaffold 312 consists of 14 type II genes. It ends with the gene encoding Kb8, which is followed in opposite orientation by the type I keratin gene Ka18 like in mammalian type II clusters (Hesse et al., 2004). The other end of the *Xenopus* as well as the chicken type II

cluster is marked by the gene encoding Fas apoptotic inhibitory molecule 2 (FAIM2) which at the human type II cluster occurs more than 2.5 Mbp upstream (Fig. 1). Interestingly, in all type II clusters the second gene has opposite orientation versus the other genes. The predicted *Xenopus* type II protein seems to resemble in sequence the mammalian keratin Kb7.

Unusual features of the distribution of keratin genes in teleost fish

Analysis of the genome drafts of *Tetraodon nigroviridis* and *Danio rerio* (zebrafish) extend earlier results on *Fugu rubripes* (Zimek et al., 2003), where scaffolds were relatively short and lacked chromosomal anchorage. The two new genome drafts (Tables 3 and 4) support the earlier conclusion that teleosts have an excess of type I over type II genes and show that unlike in the genomes of terrestrial vertebrates, fish type II genes lie at least on two different chromosomes (Fig. 2). These are for *Tetraodon* chromosomes 3 and 9 and for *Danio* chromosomes 2, 6, 17, 22, and 23. In addition, there are a few other type II genes on separate sites in *Tetraodon* in regions for which chromosomal location is lacking. Both fish have a site resembling one end of the mammalian type II cluster. It has only two type II genes in the same orientation followed by a type I gene in opposite orientation and the gene encoding EIF4B. The neighboring type II and I genes are Kb8 and Ka18.

The *Tetraodon* type I genes locate to at least three chromosomes (numbers 4, 12 and several sites on chromosome 18) and several independent loci of yet unknown chromosomal location. This is in contrast to the situation in terrestrial vertebrates where all type I genes except Ka18 form the type I cluster. The preliminary genome draft for *Danio* makes another important point. Chromosome 19 harbors a minicluster of seven type I genes all with the same orientation flanked by the gene for SUI1. This is precisely the arrangement found at one end of the type I clusters in the genomes of terrestrial vertebrates (see Fig. 1). This particular arrangement of seven *Danio* type I genes plus the flanking genes is verified by a well characterized BAC clone (see Fig. 2). On *Danio* chromosome 11 are five neighboring type I genes (Fig. 2). For reasons currently unknown we have not detected counterparts of these two patterns in either *Tetraodon* or *Fugu*). In *Tetraodon* the single copy gene for SUI1 lies on chromosome 3, which lacks genes for type I keratins. Similarly, in *Fugu* the SUI1 gene located on scaffold 482 lacks neighboring keratin genes. Thus the two puffer fish and *Danio* differ in gene arrangements.

Table 3. *Tetraodon nigroviridis* keratin I and II genes and their genomic localization (V7 assembly; Feb. 2004)

	AC	Exons	Chr	Area	Scaff	Gaps	Ori	Comment
Type I								
tnKa1		1–8	18	939745–941792	14786		—	Complete
tnKa6	CR632782.1	1–8	4	1145917–1148657	14575		+	Complete
		6–8	Un_random	78780186–78780841	6076		+	Partial duplication
tnKa8	CR693944.1	1–8	Un_random	46052897–46054757	22868		+	Complete
		1–3	Un_random	163130490–163131271	22867	e3–e8	—	Partial duplication
tnKa9		1–8	18	2501066–2503222	14637		+	Complete
tnKa10	CR726301.1	1–8	18	5132454–5134696	15124		—	Complete
tnKa14	CR657988.1	1–8	Un_random	46063527–46066365	7511		+	Complete
tnKa15	CR683995.1	1–7	Un_random	27407226–27409216	10361		+	Complete
tnKa19	CR687908.1	1–8	18	5125903–5128707	15124		—	Complete
tnKa18	CR685085.1	1–7	Un_random	43141208–43144443	14505		—	Complete
tnKa21	CR730787	1–7	Un_random	119587651–119589896	12162		+	Complete
tnKa2		1–3 + 5–7	18	947010–952206	14786		+	
tnKa3		1–7	Un_random	157946477–157948125	21093		+	
tnKa4	CR633893.1	1–4	Un_random	119582437–119586777	12162		+	
tnKa7		1–2 + 4–8	12	5072416–5075819	14996		+	
Type II								
tnKb8		1–9	Un_random	43149835–43155841	14505		+	Complete
tnKb3	CR728800.1	1–8	Un_random	43165135–43168007	14505		+	Complete
tnKb4		1–10	Un_random	12014767–12020666	14139		+	Complete
tnKb18	CR732986.1	1–7	Un_random	119565913–119568195	12162	e8–e9	—	Complete
tnKb5		1–8	9	6303029–6305106	15033		+	
tnKb6		2–7	3	2288820–2290046	14553		+	

AC, accession number; Chr, chromosome; Scaff, scaffold; Ori, orientation; Un_random, unmapped scaffolds and ultracontigs.

Table 4. *Danio rerio* keratin I and II genes and their chromosomal location (Zv4 assembly)

	AC	Exons	Chr	Area	Scaffold/AC	Area	Gaps	Ori	Comment
Type I									
Ka10	cn014897	1–8	5	64835837–64847239	BX255921.2	76191–87593		+	Complete
Ka2	NM_200568	1–7	6	19582721–19587092	Zv4_Scaffold512	13321–17692		—	Complete
Ka11[a]	ca474603	1–8	10	7599405–7615125	Zv4_Scaffold865	766920–79224		+	Complete
Ka11[b]			10	6911594–6928276	BX248511.4	79109–95792		—	
Ka6	BC044144	1–5 + 7	11	7566578–7568890	Zv4_Scaffold954.3	143336–145648	6	—	Complete cDNA
Ka12	cf999501	1–8	11	7504998–7508577	Zv4_Scaffold954.3	82058–85458		—	Complete
Ka18	NM_178437	1–7	23	5405547–5409014	Zv4_Scaffold1874.5	1417016–1420482		+	Complete, distance to EIF4b*: 1.15 mbp
Ka13	BC076059	1–7	19	581422–583448	BX323079.7	85779–87545		+	Complete
Ka14	NM_131107	1–7	19	597533–599480	BX323079.7	69488–71434		+	Complete
Ka17	ck870397	1–7	19	610192–612200	BX323079.7	56767–58775		+	Complete
Ka26	ck024288	1–8	19	629349–637381	BX323079.7	31586–39618		+	Complete
Ka7	ck353300	1–5	11	7555504–7563886	Zv4_Scaffold954.3	132262–140644		+	Complete
Ka23	BC076485	1–8	11	8488101–8493197	BX469912.5	39426–49942		+	Complete
		1–8	23	17415146–17423116	Zv4_Scaffold1891	142002–150972		—	
Ka25	cv120414	1–5	NA	88593616–88596894	Zv4_NA6373	1000–4279		+	Exons 1–5
	ck704473	6–8	11	7613434–7616368	Zv4_Scaffold954.3	406130–408880	7	—	Exons 6 + 7, gap in exon 7 but sequence in ck704473
Ka3[c]		1–8	22	9502668–9506880	AL645755.20.1	4880–8988		+	Complete
Ka3[d]		4–8	25	20367845–20368925	Zv4_Scaffold2070.6	80861–82111		—	Duplication
Ka3[e]		1–7	25	20382873–20385600	Zv4_Scaffold2070.6	95161–98859	1 + 8	+	Duplication
Ka22	BC075874	1–3	25	20363492–20364877	Zv4_Scaffold2070.6			+	
Ka16	be017725	1–7	19	592694–594688	BX323079.7	74297–76367		+	Complete, exons 1–7
Ka19	ck714234	1–8	19	616100–622800	BX323079.7	46167–52867		+	Complete
Ka1	bq263412	1–4	11	7517964–7519502	Zv4_Scaffold954.3	310652–312014		—	e4–e7 missing
Ka27			19	603730–605690	BX323079.7	63287–65237		+	
Type II									
Kb4	NM_131509	1–9	6	22408626–22412190	BX323816.5	133772–137336		+	Complete
Kb8	NM_200080	1–9	23	5314261–5318177	Zv4_Scaffold1874.5	1325730–1329646		—	Complete
Kb10	be016996	1–9	2	31839261–31850062	Zv4_Scaffold150	2665293–2676094		—	Complete
Kb3	cn174840	1–9	6	19602731–19610074	Zv4_Scaffold512.1	33331–40674		+	Complete
Kb5[i]	ck687786	1–8	22	4599797–4610379	Zv4_Scaffold1804	7768–20527		+	
Kb5[h]			17	32316482–32326942	Zv4_Scaffold1499	749092–761731		+	
Kb1	NM_131156	1 + 3–9	23	5273573–5276814	Zv4_Scaffold1874.5	1513365–1516606	2	—	

AC, accession number; Chr, chromosome; Ori, orientation.

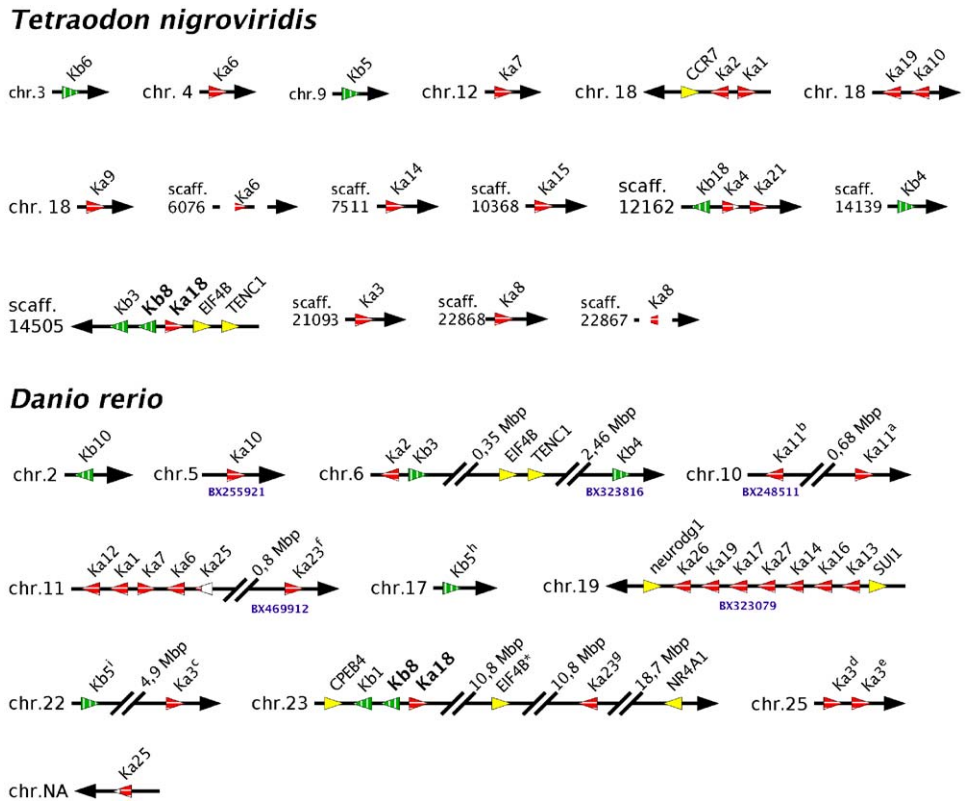

Fig. 2. Comparison of keratin gene arrangements in the teleost fish *Tetraodon nigroviridis* and *Danio rerio*. Type I keratin genes are marked as red triangles, type II keratin genes as green triangles. Non-keratin genes are yellow triangles. Their names are given in the legend to Fig. 1. The tips of the triangles correspond to the 3′ end of the genes. Long black arrows give the 5′ to 3′ orientation of the DNA strand. For *Danio*, gene arrangements supportive information from BAC clones, when available, is indicated in blue. For keratin genes, see Tables 3 and 4. Number for the fish keratins do not imply homology versus the mammalian keratins except when printed in bold type.

The IF gene complement of the chicken

Since the genome draft of the chicken (Hillier et al., 2004) is highly advanced and offers extensive chromosomal mapping, we screened the draft for the non-keratin IF genes using the spectrum of human genes as a base. Table 5 shows that we located all chicken orthologs except for the genes encoding peripherin and α-internexin. Chromosomal locations are known for all genes except those encoding paranemin and lamin A. As indicated in Table 5 corresponding chicken and human genes follow the synteny patterns reported in the chicken genome report (Hillier et al., 2004). As in mammals (Hesse et al., 2004) the genes for neurofilament proteins NF-L and NF-M are adjacent.

Unexpectedly, the gene for lamin B1 locates to chicken chromosome W, which is female specific. Finally we found that the chicken has, just like amphibia and fish (Döring and Stick, 1990; Hofemeister et al., 2002) a gene encoding LIII, the so-called oocyte-specific lamin. Its two alternative splice forms are readily recognized. The chicken LIII gene lies on chromosome 10, and 3 EST sequences document that it is expressed in the ovary and the embryo. In *Xenopus tropicalis* and the chicken the LIII gene is flanked by the genes SH3P13 and ADAMTSL3. These two genes are adjacent on human chromosome 15, mouse chromosome 7 and rat chromosome 1. The same situation holds in the starting genome analysis of the opossum (scaffold 18556). Thus the lamin LIII gene was lost with the mammalian lineage (Hesse et al., 2004).

Discussion

In most mammalian species there is an approximately equal number of keratin type I and II genes, which give rise to so-called expression pairs (Herrmann et al., 2003; Hesse et al., 2004). Unexpectedly, the preliminary gene complement of the teleost fish *Fugu rubripes* indicated a sizeable excess of type I over type II genes (Zimek et al., 2003) and this seems to hold also for the two additional teleosts analyzed here. Our analysis of the genomic information for the chicken and the amphibian *Xenopus*

Table 5. Summary of chicken IF genes (2004 assembly)

Type	Name	AC	Chr	Area	Hchr	Syn
I	Keratin type I gene cluster	See text	27		17	•
II	Keratin type II gene cluster	See text	Un		12	
III	Desmin	AB011672	7	22961587–22962113	2	•
	Vimentin	P09654	2	19100897–19107314	10	•
	GFAP	co422142	27	1597443–1600792	17	•
	Peripherin				12	
IV	NF-H	al589098	15	10988323–10991336	22	•
	NF-M	X05558	22	948820–953332	8	•
	NF-L	bu133465	22	939901–944326	8	•
	Syncoilin	bu413190	23	4660558–4666796	1	•
	Synemin	NM_204809	10	17483158–17507720	15	•
	Paranemin	U59287	Un	121755141–121770471	1	
	α-Internexin				10	
V	Lamin A	X16879		contig6419	1	
	Lamin B1	X16878	W	4084435–4103660	5	•
	Lamin B2	X16880	28	585727–611584	19	•
	Lamin LIIIa + b	bu460809 bu287843 bu206556	10	12048464–12063333		
VI	CP49/Phakinin	X84806	2	41707622–41723088	3	•
	Filensin	X72873	3	5704925–5723571	20	•

Note: EST sequences from lamin LIII are from ovary, kidney and whole embryo.AC, accession number; Chr, chromosome; Hchr, human chromosome; Syn, synteny; Un, unanchored contigs.

tropicalis shows like in mammals about equal numbers of type I and II keratin genes.

Terrestrial animals have two keratin gene clusters

The screen of the recently established draft genomes of the chicken and the amphibian *Xenopus* shows that birds and amphibians display like the mammals a keratin type I gene cluster and a keratin type II gene cluster. While the chicken type I gene cluster lies on chromosome 27, the chromosomal location of the other non-mammalian keratin gene clusters is not yet known. The strong reduction in gene number per chicken and *Xenopus* cluster compared to mammalian gene clusters seems in line with the absence of genes encoding the type I and type II keratins of hair, which is a mammalian specialization. Similarly, the type I clusters of birds and amphibia lack the subdomain encoding some 30 keratin-associated proteins (KAPs) in mammals. Again KAPs are like hair keratins and inner root sheath keratins a mammalian acquisition. Strikingly, all type I clusters so far established show that the genes display the same orientation. In type II clusters there is the occasional gene with opposite orientation (Fig. 1).

Keratin genes Kb8 and Ka18 are neighboring genes in all vertebrate classes

Keratins Kb8 and Ka18 are the first keratin pair expressed in mammalian embryogenesis. This keratin pair is characteristically found in interior epithelia. Interestingly the Ka18 gene is the only type I keratin gene, which is not present in the type I gene cluster of mammals and occurs instead at one end of the type II gene cluster (Hesse et al., 2004). Thus the genes for Ka18 and Kb8 are adjacent. This arrangement also holds for the amphibian *Xenopus*, while in the current genome draft the chicken Ka18 gene lies on a small contig outside the type I cluster. Whether this contig extends the chicken type II cluster is not known, but emerging results on the genome drafts of several fishes show that also here keratin Kb8 and Ka18 genes are adjacent. While the teleost fish seem to lack a type II gene cluster, they show an arrangement of two type II keratin genes followed by a type I keratin gene in opposite orientation and the gene encoding the non-keratin protein EIF4B. The neighboring type II and type I gene encode Kb8 and Ka18, respectively (Zimek et al., 2003). This gene arrangement, which we also observe in the emerging genomes of the teleosts *Tetradon nigroviridis* and *Danio rerio* (Fig. 2), reflects one end of the mammalian type II

cluster (Fig. 1). Since the genes for Kb8 and Ka18 are adjacent in three mammalian genomes, three fish genomes as well as the *Xenopus* genome, we tentatively assume that it will also emerge in the future for the bird genome. The current genome draft already documents that the chicken Ka18 gene lies as expected outside the keratin type I gene cluster (Fig. 1).

Further advance of the chicken genome draft should also clarify the relation of the type II gene gKb16, which lies on a contig separate from the contig housing the type II gene cluster. Since the borders of the chicken type II gene cluster are still not defined the gene Kb16 could be connected with the cluster at its open end.

Unique features of teleost fish

Although there are three genome drafts on teleost fish, several problems prevent a detailed understanding of the arrangements of keratin genes at present. There are limitations in each draft. The *Fugu* scaffolds already discussed (Zimek et al., 2003) are rather short and chromosomal anchorage is lacking. The *Danio* draft was unfortunately derived from DNA of one thousand 5-day-old embryos. The zebrafish consortium warns that there is a high level of misassembly, false duplications and assembly dropouts. The draft provides however some chromosomal anchorage. Finally, the *Tetraodon* draft, although highly advanced, is still incomplete and gives only 64% in chromosomal anchorage.

All three teleost drafts indicate in contrast to mammalian/terrestrial vertebrate genomes an excess of type I over type II genes and also indicate the absence of a type II keratin gene cluster. In *Tetraodon* and *Danio,* at least two of the six type II genes locate to different chromosomes (Fig. 2; Tables 3 and 4) while the others occur separately on contigs of unknown chromosomal location. Among these is a gene arrangement resembling one end of the mammalian type II cluster with two type II genes followed by one type I gene (Ka18) and the gene encoding EIF4B. This arrangement occurs in all three drafts and was discussed above to show that Kb8 and Ka18 are neighboring genes in all vertebrates.

The display of type I keratin genes in teleosts and terrestrial vertebrates differs in that teleost type I genes are spread over many more positions. In addition to the general location of the Ka18 gene next to the gene Kb8, already discussed, type I genes occur on at least three chromosomes (*Tetraodon* chromosomes 4, 12 and several distinct sites of chromosome 18) plus several sites of unknown chromosomal location. Curiously, the striking pattern of seven adjacent type I genes followed by the SUI1 gene on *Danio* chromosome 19 does not emerge in the *Tetraodon* draft. The *Danio* midi cluster of seven type I keratin genes capped by the SUI1 gene clearly resembles the much longer keratin type I cluster of terrestrial vertebrates. It is now well established that a whole genome duplication and subsequent massive gene loss and local gene shuffling occurred in the teleost fish lineage after its divergence from the line leading to terrestrial vertebrates. As elegantly shown by (Jaillon et al., 2004), the processes of genomic adaptation were accompanied by chromosomal rearrangements. Within this scenario we tentatively assume that an evolutionary older keratin type I gene cluster and possibly type II gene cluster of the ancestral bony vertebrate were dispersed on the teleost branch. This in turn raises the question whether keratin gene clusters possibly of shorter length exist also in cartilage fish or in the lamprey, the most primitive vertebrate, or even in the cephalochordate amphioxus. In contrast the urochordate *Ciona* has only a single type I and type II gene (Karabinos et al., 2004a). An important answer as to the evolutionary age of keratin gene clusters is therefore expected from current genome projects on amphioxus and the sea lamprey (Mulley and Holland, 2004). Because of the importance of the zebrafish as a model organism for developmental biology one can also expect with time the emergence of a genome draft without the current shortcomings.

Because of the high quality of the chicken genome draft we extended our analysis to the non-keratin IF genes. Nearly all counterparts of the human genes were located and for nearly all chicken genes chromosomal locations are known (Table 5) and follow the synteny patterns established for human and chicken chromosomes (Hillier et al., 2004). Chromosomal anchorage is still missing for the genes encoding paranemin, lamin A and the keratin type II gene cluster. Unlike in mammals (Hesse et al., 2001) the gene for chicken lamin B1 locates to the sex chromosome W, which is only present in female animals. Provided this is correct and that there is not an additional copy of the B1 gene, one would expect that lamin B1 is missing in male chickens. Entirely unexpected was the finding of a lamin LIII gene, which so far was thought to be restricted to amphibia and fish (Döring and Stick, 1990; Hofemeister et al., 2002). This gene is lost on the lineage leading to mammals (Hesse et al., 2004).

References

Coulombe, P.A., Ma, L., Yamada, S., Wawersik, M., 2001. Intermediate filaments at a glance. J. Cell Sci. 114, 4345–4347.

Dodemont, H., Riemer, D., Weber, K., 1990. Structure of an invertebrate gene encoding cytoplasmic intermediate filament (IF) proteins: implications for the origin and the diversification of IF proteins. EMBO J. 9, 4083–4094.

Döring, V., Stick, R., 1990. Gene structure of nuclear lamin LIII of *Xenopus laevis*; a model for the evolution

of IF proteins from a lamin-like ancestor. EMBO J. 9, 4073–4081.

Erber, A., Riemer, D., Bovenschulte, M., Weber, K., 1998. Molecular phylogeny of metazoan intermediate filament proteins. J. Mol. Evol. 47, 751–762.

Fuchs, E., Weber, K., 1994. Intermediate filaments: structure, dynamics, function, and disease. Annu. Rev. Biochem. 63, 345–382.

Herrmann, H., Aebi, U., 2000. Intermediate filaments and their associates: multi-talented structural elements specifying cytoarchitecture and cytodynamics. Curr. Opin. Cell Biol. 12, 79–90.

Herrmann, H., Hesse, M., Reichenzeller, M., Aebi, U., Magin, T.M., 2003. Functional complexity of intermediate filament cytoskeletons: from structure to assembly to gene ablation. Int. Rev. Cytol. 223, 83–175.

Hesse, M., Magin, T.M., Weber, K., 2001. 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. 114, 2569–2575.

Hesse, M., Zimek, A., Weber, K., Magin, T.M., 2004. Comprehensive analysis of keratin gene clusters in humans and rodents. Eur. J. Cell Biol. 83, 19–26.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., Dodgson, J.B., Chinwalla, A.T., Cliften, P.F., Clifton, S.W., Delehaunty, K.D., Fronick, C., Fulton, R.S., Graves, T.A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J.D., Miner, T.L., Minx, P., Nash, W.E., Nhan, M.N., Nelson, J.O., Oddy, L.G., Pohl, C.S., Randall-Maher, J., Smith, S.M., Wallis, J.W., Yang, S.P., Romanov, M.N., Rondelli, C.M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H.G., Robertson, L., Masabanda, J.S., Griffin, D.K., Vignal, A., Fillon, V., Jacobbson, L., Kerje, S., Andersson, L., Crooijmans, R.P., Aerts, J., van der Poel, J.J., Ellegren, H., Caldwell, R.B., Hubbard, S.J., Grafham, D.V., Kierzek, A.M., McLaren, S.R., Overton, I.M., Arakawa, H., Beattie, K.J., Bezzubov, Y., Boardman, P.E., Bonfield, J.K., Croning, M.D., Davies, R.M., Francis, M.D., Humphray, S.J., Scott, C.E., Taylor, R.G., Tickle, C., Brown, W.R., Rogers, J., Buerstedde, J.M., Wilson, S.A., Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M.M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoedt, K., Wong, G.K., Wang, J., Liu, B., Yu, J., Yang, H., Nefedov, M., Koriabine, M., Dejong, P.J., Goodstadt, L., Webber, C., Dickens, N.J., Letunic, I., Suyama, M., Torrents, D., von Mering, C., Zdobnov, E.M., et al., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Hofemeister, H., Kuhn, C., Franke, W.W., Weber, K., Stick, R., 2002. Conservation of the gene structure and membrane-targeting signals of germ cell-specific lamin LIII in amphibians and fish. Eur. J. Cell Biol. 81, 51–60.

Irvine, A.D., McLean, W.H., 1999. Human keratin diseases: the increasing spectrum of disease and subtlety of the phenotype-genotype correlation. Br. J. Dermatol. 140, 815–828.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J., Roest Crollius, H., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431, 946–957.

Karabinos, A., Schmidt, H., Harborth, J., Schnabel, R., Weber, K., 2001. Essential roles for four cytoplasmic intermediate filament proteins in *Caenorhabditis elegans* development. Proc. Natl. Acad. Sci. USA 98, 7863–7868.

Karabinos, A., Schünemann, J., Parry, D.A., Weber, K., 2002. Tissue specific coexpression and in vitro heteropolymer formation of the two small *Branchiostoma* intermediate filament proteins A3 and B2. J. Mol. Biol. 316, 127–137.

Karabinos, A., Schulze, E., Schünemann, J., Parry, D.A., Weber, K., 2003. In vivo and in vitro evidence that the four essential intermediate filament (IF) proteins A1, A2, A3 and B1 of the nematode *Caenorhabditis elegans* form an obligate heteropolymeric IF system. J. Mol. Biol. 333, 307–319.

Karabinos, A., Zimek, A., Weber, K., 2004a. The genome of the early chordate *Ciona intestinalis* encodes only five cytoplasmic intermediate filament (IF) proteins including a single type I and type II keratin and a unique IF-annexin fusion protein. Gene 326, 123–129.

Karabinos, A., Schünemann, J., Weber, K., 2004b. Most genes encoding cytoplasmic intermediate filament (IF) proteins of the nematode *Caenorhabditis elegans* are required in late embryogenesis. Eur. J. Cell Biol. 83, 457–468.

Mulley, J., Holland, P., 2004. Comparative genomics: small genome, big insights. Nature 431, 916–917.

Rogers, M.A., Langbein, L., Winter, H., Ehmann, C., Praetzel, S., Korn, B., Schweizer, J., 2001. Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21. J. Biol. Chem. 276, 19440–19451.

Sato, K., Yasugi, S., 1997. Chicken keratin-19: cloning of cDNA and analysis of expression in the chicken embryonic gut. Dev. Growth Differ. 39, 751–761.

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., Höffling, 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., Knappe, M., Hunzinger, C., Markl, J., 2003. cDNA sequences of the authentic keratins 8 and 18 in zebrafish. Differentiation 71, 73–82.

Vanhoutteghem, A., Londero, T., Ghinea, N., Djian, P., 2004. Serial cultivation of chicken keratinocytes, a composite cell type that accumulates lipids and synthesizes a novel beta-keratin. Differentiation 72, 123–137.

Wang, J., Karabinos, A., Zimek, A., Meyer, M., Riemer, D., Hudson, C., Lemaire, P., Weber, K., 2002. Cytoplasmic intermediate filament protein expression in tunicate development: a specific marker for the test cells. Eur. J. Cell Biol. 81, 302–311.

Weber, K., Plessmann, U., Ulrich, W., 1989. Cytoplasmic intermediate filament proteins of invertebrates are closer to nuclear lamins than are vertebrate intermediate filament proteins; sequence characterization of two muscle proteins of a nematode. EMBO J. 8, 3221–3227.

Zimek, A., Stick, R., Weber, K., 2003. 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. 116, 2295–2302.

Zimek, A., Weber, K., 2002. The gene for a cytoplasmic intermediate filament (IF) protein of the hemichordate *Saccoglossus kowalevskii*; definition of the unique features of chordate IF proteins. Gene 288, 187–193.