



Flanking genes of an essential gene give information about the evolution of metazoa

Alexander Zimek*, Klaus Weber

Max Planck Institute for Biophysical Chemistry, 37077 Goettingen, Germany

ARTICLE INFO

Article history:

Received 2 September 2010
Received in revised form 6 October 2010
Accepted 8 October 2010

Keywords:

Metazoa
Molecular phylogeny
Lamin genes
Flanking genes
Genome drift

ABSTRACT

We collected as much information as possible on new lamin genes and their flanking genes. The number of lamin genes varies from 1 to 4 depending more or less on the phylogenetic position of the species. Strong genome drift is recognised by fewer and unusually placed introns and a change in flanking genes. This applies to the nematode *Caenorhabditis elegans*, the insect *Drosophila melanogaster*, the urochordate *Ciona intestinalis*, the annelid *Capitella teleta* and the planaria *Schmidtea mediterranea*. In contrast stable genomes show astonishing conservation of the flanking genes. These are identical in the sea anemone *Nematostella vectensis* and the cephalochordate *Branchiostoma floridae* lamin B1 gene. Even in the lamin B1 genes from *Xenopus tropicalis* and man one of the flanking genes is conserved. Finally our analysis forms the basis for a molecular analysis of metazoan phylogeny.

© 2010 Elsevier GmbH. All rights reserved.

Introduction

Text books tell us how a comparison of specific protein sequences can be used to calculate an evolutionary tree (Wehring and Gehring, 2007). A recent specific example is the tree calculated for 21 arthropods from sequences of motor proteins (Odrionitz et al., 2009). We have taken a different approach based on intron patterns in lamin genes (Zimek and Weber, 2008).

The large multigene family of metazoan intermediate filament proteins (IF proteins) covers two types: the small group of nuclear lamins and the large group of cytoplasmic IF proteins having sequence homologies, that are particularly obvious in the protozoan cytoplasmic invertebrate IF proteins (Dodemont et al., 1990; Doring and Stick, 1990; Zimek and Weber, 2002). The common structural element is a long central rod domain consisting of the smaller coil 1a, the long coil 1b, the small coil 2a and the long coil 2b domain. This describes a segmented double stranded coiled coil (for review see Gruenbaum et al., 2005), from which the non-helical N-terminal head and C-terminal tail domains extend into the exterior. A major difference between the lamins and the cytoplasmic IF proteins is a nuclear location signal (NLS), which usually consists of four consecutive basic residues. Lamins usually have a C-terminal isoprenylation site C-a-a M, which in a stepwise reaction gives rise to a chargeless carboxy terminal cysteine (see Gruenbaum et al., 2005).

Lamin A is only expressed once cells differentiate (Rober et al., 1989). Many cytoplasmic IF proteins show mutations which give rise to human diseases (Bonifas et al., 1991; Vassar et al., 1991). The nematode *Caenorhabditis elegans* (*C. elegans*) has a single lamin and 11 cytoplasmic IF proteins, that in part arise by differential splicing. RNA interference studies showed that at least five of them are essential for embryonic development (Karabinos et al., 2001). *C. elegans* lamin is crucially involved already in the earliest stages of embryonic development (Liu et al., 2000). The number of cytoplasmic IF proteins ranges from 11 in *C. elegans* (Karabinos et al., 2001) to 65 in man (Hesse et al., 2001).

One striking difference between A type lamins and B lamins is seen before mitosis when the nuclear envelope disassembles into vesicles. Lamins B1 and B2 stay bound to these vesicles and are used in the formation of nuclear envelopes in the daughter cells. A-type lamins are of course initially also isoprenylated, but the modified C-terminal end is subsequently cleaved by a specific metalloproteinase to form the mature lamin A molecule which has lost 10 amino acid residues (Gruenbaum et al., 2005; Weber et al., 1989). Thus, the differentially spliced lamin C as well as the mature lamin A molecules lack the isoprenylation site and become soluble. The sequence around the cleavage site is highly specific so that any mutations at this site give rise to laminopathies (Worman and Bonne, 2007).

Electron microscopical studies have shown that arthropods and their close relatives lack cytoplasmic intermediate filaments. Here, thick bundles of specialised microtubules substitute for the bundles of cytoplasmic intermediate filaments, that are observed in other phyla (Bartnik and Weber, 1989). This finding suggested that cytoplasmic IF proteins arose in evolution from a nuclear lamin. In

* Corresponding author at: Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Goettingen, Germany. Tel.: +49 551 201 1157; fax: +49 551 201 1175.

E-mail address: azimek@gwdg.de (A. Zimek).

Table 1

Intron sizes of radiata and placozoa lamins starting at the 5'-end of the gene.

Nematostella vectensis Predicted protein: EDO37225	The 9 introns of the gene have a size starting from the 5'-end of the gene of 1.2 kb, 0.3 kb, 0.2 kb, 0.8 kb, 0.2 kb, 0.5 kb, 0.4 kb, 0.5 kb and 0.6 kb
Hydra magnipapillata On genomic contig NW.002157556.1	The 9 introns of the gene have a size starting from the 5'-end of the gene of 389 bp, 11.33 kbp, 99 bp, 2.38 kbp, 5.45 kbp, 155 bp, 114 bp, 865 bp and 78 bp
Trichoplax adhaerens scaffold.3: 3298016–3306114	The 9 introns of the gene have a size starting from the 5'-end of the gene of 3107 bp, 145 bp, 105 bp, 190 bp, 129 bp, 105 bp, 143 bp, 111 bp and 977 bp

Table 2

Intron sizes of ecdysozoa lamins starting at the 5'-end of the gene.

Caenorhabditis elegans lmn-1 NM.059970	The 5 introns of the gene have a size starting from the 5'-end of the gene of 90 bp, 48 bp, 157 bp, 52 bp and 84 bp
Drosophila melanogaster Lamin Dmo NT.033779 Lamin C NT.033778	The 2 introns of the gene have a size starting from the 5'-end of the gene of 243 bps and 64 bps The 3 introns of the gene have a size starting from the 5'-end of the gene of 2.4 kbp, 60 bps and 148 bps
Daphnia pulex scaffold.90: 311661–315862	The 10 introns of the gene have a size starting from the 5'-end of the gene of 910 bp, 89 bp, 63 bp, 60 bp, 78 bp, 87 bp, 62 bp, 113 bp, 91 bp and 166 bp

Table 3

Intron sizes of lophotrochozoa lamins starting at the 5'-end of the gene.

Schmidtea mediterranea contig 1379: mk4.001379.07	The 4 introns of the gene have a size starting from the 5'-end of the gene of 51 bp, 1970 bp, 53 bp and 55 bp
Capitella teleta scaffold 373: 172264–179935	The 6 introns of the gene have a size starting from the 5'-end of the gene of 3.36 kbp, 50 bp, 49 bp, 58 bp, 50 bp and 910 bp

Table 4

Intron sizes of lower deuterostomes lamins starting at the 5'-end of the gene.

Strongylocentrotus purpuratus NM.214500	The 9 introns of the gene have a size starting from the 5'-end of the gene of 21 kbp, 2.4 kbp, 1.9 kbp, 1.8 kbp, 2.5 kbp, 1.2 kbp, 1050 bp, 540 bp and 910 bp
Ciona intestinalis Lamin L2 AJ251957	The 10 introns of the Lamin L2 gene have a size starting from the 5'-end of the gene of 0.6 kbp, 1.8 kbp, 2.6 kbp, 0.6 kbp, 0.5 kbp, 0.25 kbp, 1.1 kbp, 0.35 kbp, 0.3 kbp and 0.75 kbp
Branchiostoma floridae chrUn: 258492196–258513674	The 10 introns of the gene have a size starting from the 5'-end of the gene of 7744 bp, 324 bp, 519 bp, 962 bp, 1175 bp, 885 bp, 1857 bp, 396 bp, 4227 bp and 1410 bp

addition the full genome of *Drosophila melanogaster* proved this point much later on the molecular level. Once the genome of the sea anemone *Nematostella vectensis* became available we found that it had a single nuclear lamin gene and lacked genes for cytoplasmic IF proteins. The *Nematostella* lamin gene has 9 introns, which are strikingly conserved in the human lamin B genes, which have only 1 (lamin B1) or 2 (lamin B2) additional introns. Using the information on one neighbouring gene we concluded that the human B1 gene on chromosome 5 and the human B2 gene on chromo-

some 19 arose later with the vertebrate lineage (Zimek and Weber, 2008).

Here we have assembled information on new lamin genes and determined when possible both flanking genes.

Materials and methods

We analysed the public available genome assemblies of various metazoans using online tools of the respective websites and

Table 5

Intron sizes of vertebrate lamins starting at the 5'-end of the gene.

Homo sapiens Lamin B1 NM.005573	The 10 introns of the gene have a size starting from the 5'-end of the gene of 26.9 kbp, 0.63 kbp, 4.44 kbp, 1.38 kbp, 6.97 kbp, 1.73 kbp, 1.55 kbp, 3 kbp, 6.5 kbp and 3.38 kbp
Lamin B2 NM.032737	The 9 introns of the gene have a size starting from the 5'-end of the gene of 27.6 kb, 4.45 kb, 1.4 kb, 7 kb, 1.75 kb, 1.6 kb, 3.1 kb, 6.65 kb and 3.4 kb
Lamin A NM.170707	The 11 introns of the gene have a size starting from the 5'-end of the gene of 15.3 kbp, 3.5 kbp, 275 bp, 210 bp, 0.58 kbp, 90 bp, 0.48 kbp, 85 bp, 0.42 kbp, 0.74 kbp and 0.32 kbp
Xenopus tropicalis Lamin B1 NM.203867	The 10 introns of the gene have a size starting from the 5'-end of the gene of 18.5 kbp, 650 bp, 230 bp, 320 bp, 770 bp, 1.5 kbp, 0.8 kbp, 0.7 kbp, 0.7 kbp and 0.35 kbp
Lamin B2 scaffold.555: 694,555–707,819	The 10 introns of the gene have a size starting from the 5'-end of the gene of 3.45 kbp, 1.41 kbp, 0.77 kbp, 0.75 kbp, 120 bp, 175 bp, 0.98 kbp, 85 bp, 0.9 kbp and 1.78 kbp
Lamin B3 NM.001083354	The 10 introns of the gene have a size starting from the 5'-end of the gene of 1.9 kbp, 280 bp, 180 bp, 3.2 kbp, 90 bp, 630 bp, 0.8 kbp, 820 bp, 2.18 kbp and 95 bp
Lamin A CR942559	The 11 introns of the gene have a size starting from the 5'-end of the gene of 37.8 kbp, 2.1 kbp, 1.37 kbp, 75 bp, 250 bp, 770 bp, 150 bp, 1.5 kbp, 120 bp, 2.1 kbp and 200 bp

```

1
nematostel MATATKSPAS SSSTPKTPVS SSRIMGSSPP GSAKFTRAQE KAELOHLNDR LATYIDRVKN LEQENSKLRS EVTVSRKTVE REVDMSKSLY ETELADARRL
hydra MA..... .STP..DVR SPRYGYQSP..NVR.TRVQE KEELGGLNDR LATYIDRMRY LEQONTRLSA EITTSKESVT REGVNPKAMF EAELEGEARRL
trichoplax M..... .SSPSM... ..ASP SPLKIKRVQE KEELKNLTDNR LASYIEKSRF LESRNARLNE EIRSSRRSGD ENIKSLKALY EQELQEARKT

nematostel LDETAKKAK QQIESSKNSN DAQEFKNKFD EAAAARKKAE KELNDVRKLL HDKENQITRR NQEAALNLESV LRELQGECEE LKDALKAAY ALEQETITRV
hydra LDETAKDKAL LQLENNKLTLS LVEDLRHLE KEVANRSRIE BELKRVERRL LEKESLNIVL SKERKDLHQ VKDLEDQLKE YQDDLEKEKS DHEKEIIRRV
trichoplax IDEMANEKSK ITITVMKYTK QVEELTSLLE KVKASKSSLQ VALNNAEKRA TELEGEVNVNE RKNVAELREV VIGHKEKIQR SSEELESANK LAEQQALRRV

nematostel DLENRCQSLQ EEQNFKQMY EELSDIRSQ LKTIVETKRVV VETDYK.DKY EGLMAEKLEQ LREYDSDAR SFKEETELLY SSEEELRIQ RERDSEALAK
hydra DAENRAQTLQ EEVDFNEQVH AKEMDLRAQ...TESLHFS LQMDKVDNY DNILRDKLQE LREFEFEED NAKRELEDAY QNPFETLRAE SDRDRSHVTR
trichoplax QIENEYQTLK ESSEFSTEMH QCEIRSLKKN LKTVDQORST LQDDPQSE.Y DTKMTEALQQ LRRDNEENSK RLKQVEEELY RSQVELESQ RERDTKIITE

nematostel LREENRNLK SVLELSSQVH QLEAKNNALV SRVSDLQGLR AQDKKDHNE ILLRENEIAE LRTSIDDALR DYEDLMGVKV ALDMEITAYR KLESEETRL
hydra LIESNTNLKS ELDKLRSDYS NLESKVVQLQ KRISDLDSLQ QHDKDDPHKQ LQVRDDKNRL LQDKIDELQN EYEALLGSKI ALDVEIAAYR KLEGEETERL
trichoplax LKNECRKFRD KSESISSEVH VLRAENSSLQ TRKRELEESL IEERQKNTK FEYEMEMRA LQLKLEEQST EYKELLDVVKI ELDNVEIAAFR KLEGEETERL

nematostel EI.TPPASPV LGGSSVSQT RRGNKARTE ETESTMTTIT TAEGAIQFTE ADPDGKYIKI YNSGKDEAL GGWTIQRVV.GTEDPSV.YK FTKPVLYKSK
hydra NISTPPR... .GSTS... KRGTKRSRAD DEEQP.HVEN NAIGSIEIAE CDSDAKFVRL INNSDKPEPL GGWSVQRIVD GKDEQAVEYK FTKPVLYKSK
trichoplax S.SSP..... .SSPFAQR KR.....RRI DEDSDYIVSS HATGDVQISD VDKHGLYVRL LNTSKTCHI GNFKVHKQVD NKNE...ISYK FNTKALKLGN

nematostel SHVIVMSAQG GTTHKPPSDL VFKQLPSWGS GNEARTALVN AGCEMATLL EEK.VFQHYS TDVIDSGRRG RCRR...G EVAKGCVA.M*
hydra QTVIIVSQNS GVKQKPPPTDF VLKQ.EDWSH GSSMTTSLIN PDPEIVA... .K.HTIALP TPPVGT.RRI RRR...G GERESCKL.M*
trichoplax CVTIVVAADS GQSHKPPANY LWKQNNFGL GEEMVTRLVN AGCEPAIFN LTKAQVEIK DELFQSYSI IGFKQLPQFI ETTEAVNIKR*

```

Fig. 1. Alignment of predicted lamin protein sequences for the cnidaria *Nematostella vectensis* (upper line), *Hydra magnipapillata* (middle) and the placozoan *Trichoplax adhaerens*. Intron positions are marked by inverted letters (white on black) in the sequence, for intron sizes see Table 1.

published lamin sequences. Some gene sequences were manually extracted from databases for the respective animals and assembly versions. The collection of these putative lamin coding sequences is available as [Supplementary material](#) (mmc2.zip to mmc8.zip). Locations, either GenBank accession numbers or genomic database locations are given in the respective Tables 1–5.

Genome browsers, analysed animals and assembly releases

<http://genome.ucsc.edu/>

- *Strongylocentrotus purpuratus*: Sep. 2006 (Baylor 2.1/strPur2)
- *Branchiostoma floridae*: Mar. 2006 (JGI 1.0/braFlo1)
- *Caenorhabditis elegans*: May 2008 (WS190/ce6)
- *Ciona intestinalis*: Mar. 2005 (JGI 2.1/ci2)

<http://genome.jgi-psf.org>

- *Nematostella vectensis*: v1.0 (March 2008)
- *Capitella teleta*: v1.0 (October 2007)
- *Trichoplax adhaerens*: v1.0 (October 2007)

<http://www.ncbi.nlm.nih.gov/genomes/gebblast.cgi?gi=6140>

- *Hydra magnipapillata*: NZ_ABRM00000000, WGS draft (2009/01/02)

<http://smedgd.neuro.utah.edu>

- *Schmidtea mediterranea*: v3.1 (Sofia M.C. Robb, Eric Ross and Alejandro Sánchez Alvarado (2007) SmedGD: the Schmidtea mediterranea Genome Database Nucleic Acids Research, 36:D599–D606).

Used software

Sequence analysis: GCG Version 11.1 (Accelrys Inc., San Diego, CA).

Graphics: penOffice.org v. 3.2.1 (Oracle).

Results

The lamin gene of *Nematostella*

The lamin gene has 9 introns (Zimek and Weber, 2008). Their sizes and positions are given in Table 1 and Fig. 1 respectively. The flanking genes are genes encoding methylmalonyl-CoA epimerase

and the membrane associated ring-finger protein 2/3 on scaffold 145 (Fig. 6).

A lamin genes from Placozoa

With the genome of *Trichoplax adhaerens* the genome of an evolutionary very old metazoan became available (Wehner and Gehring, 2007). Using the predicted cDNA sequence of the sea anemone lamin gene we have located the *Trichoplax* lamin gene on scaffold 3. It has the same 9 intron positions as the *Nematostella* gene. Their sizes and positions are given in Table 1 and Fig. 1 respectively. One flanking gene is as for *Nematostella* the gene encoding methylmalonyl-CoA epimerase (Fig. 6). The second flanking gene cannot be identified because of a gap in the genome information. Potentially interesting is the finding that the deduced lamin sequence lacks the carboxy terminal isoprenylation C-a-a-M so far found in all other lamins (see Discussion).

The *Hydra* lamin gene, a strong drift of *Hydras* versus the sister group of the *Anthozoa*

Fig. 1 shows the lamin gene of *Hydra magnipapillata* (Chapman et al., 2010) versus the lamin gene of *Nematostella*. The *Hydra magnipapillata* lamin gene is located on the genomic contig NW_002157556 and has 9 introns. Their sizes and positions are given in Table 1 and Fig. 1 respectively.

We had previously cloned the *Hydra* gene from *Hydra attenuata* (Erber et al., 1999). The deduced cDNA and protein sequences show 99% identity with the new data from *Hydra magnipapillata*. However, contrary to our report, the gene has the same 9 intron positions as the *Nematostella* gene and not only the last 3 positions. We were therefore surprised when we interpreted the *Nematostella* lamin gene and did not put our *Hydra* gene in the summary, but mentioned it only in the text with the note that we did not know whether our intron mapping procedure had been correct (Zimek and Weber, 2008). It was clearly not and this most likely concerns also the *Priapul* *caudatus* lamin gene where we reported also only 3 introns (Erber et al., 1999).

Chapman et al. (2010) report that the sequence divergence between a *Hydra* peptide and its *Nematostella* orthologue is typically greater than the sequence divergence between the *Nematostella* and its human orthologue. For the lamin protein we find a sequence identity between the two cnidarian proteins of 43.5% while the corresponding value for the *Nematostella* and the human lamin B1 in spite of the long evolutionary distance is still 40% (see also Zimek and Weber, 2008). Chapman et al. (2010) also report a strong loss of introns in *Hydra* versus *Nematostella*. Because of



Fig. 2. Alignment of predicted lamin protein sequences for the nematode *C. elegans* (upper line), the two lamins of the fruitfly *Drosophila melanogaster* (lamin Dmo and C, middle lines) and the small crustacean *Daphnia pulex* (lower line). Intron positions are marked by inverted letters (white on black) in the sequence, for intron sizes see Table 2.

sequence gaps in the Hydra genomic DNA the nearest neighbours of the Hydra lamin gene are not available.

While the Hydra lamin has a nuclear location signal with only 3 basic amino acid residues (KRSR) and an unusual basic cluster RRIRKRR, cell biological experiments show that this cluster is not responsible for nuclear uptake (Erber et al., 1999). We now recognise in the Nematostella lamin also a nuclear location signal with only three basic residues (KRAR) and also a basic stretch close to the carboxy terminal end (RRGRGRR). Both features seem specific for cnidaria (see Fig. 1).

A normal lamin gene from the Ecdysozoa branch

Using the Nematostella gene as a probe we have located a lamin gene in the genomic information of *Daphnia pulex* on scaffold 90. The gene has 10 introns. Their sizes and positions are given in Table 2 and Fig. 2 respectively. The flanking genes of the gene are peptidyl-glycine- α amidating monooxygenase and a gene encoding a pdz domain containing protein (Fig. 6). Fig. 2 and Table 2 also contain the information on the single lamin gene of the nematode *C. elegans*. It has 5 introns. The sizes and positions of these introns are given in Table 2 and Fig. 2 respectively. The flanking genes of the single lamin gene are a gene encoding a member of the family for the transport to the inner mitochondrial membrane and a gene encoding a member of the helix-loop-helix family (Fig. 6).

Included in Fig. 2 is also the information on the 2 lamins of *Drosophila melanogaster*. The constitutively expressed lamin Dmo gene located on chromosome 2L has 2 introns. The organisation of this gene was already established by Osman et al. (1990) long before the *Drosophila* genome was established. Their sizes and positions are given in Table 2 and Fig. 2 respectively. The flanking genes of the lamin Dmo gene are a gene encoding a helicase at 25E and oscillin (Fig. 6). Finally there is also the information on the *Drosophila* lamin C gene. It is expressed only in differentiated cells (Riemer et al., 1995). The gene lies on chromosome 2R. It has 3 introns. Their sizes and positions are given in Table 2 and Fig. 2 respectively. The flanking genes of the *Drosophila* lamin C gene encode the cuticular protein 51A and an RNA polymerase subunit. It is striking that lamin genes from organisms with a strong genome drift show this not only in a reduced number of introns, often at new positions

(Zimek and Weber, 2008), but additionally lose the identity of the flanking genes.

Lamin genes from the Lophotrochozoa branch

We located 2 lamin genes of species from the Lophotrochozoa branch by a screen with the Nematostella gene.

In the planarian *Schmidtea mediterranea*, which is a member of the plathelminthes, we located a lamin gene on v31.001379. The gene has 4 introns. Their sizes and positions are given in Table 3 and Fig. 3 respectively. Thus the lamin gene keeps only one of the usually conserved intron positions. It has lost 8 introns and gained 3 new intron positions, indicating a strong genomic drift of this metazoan species. Since the genome analysis of *Schmidtea mediterranea* has started only recently, no information is available on flanking genes.

We have also found a lamin gene in the genomic information on the annelid *Capitella teleta* on scaffold 373. The gene has 6 introns. Their sizes and positions are given in Table 3 and Fig. 3 respectively. The lamin gene of *Capitella teleta* has lost 4 typical intron positions and acquired one new intron position indicating a strong drift in the genome of the organism. The deduced protein sequence ends with M-L-R rather than the usually found C-A-V-M. The flanking genes are genes encoding the zinc finger protein 44 and a gene encoding a protein with a pdz domain (Fig. 6).

Lamin genes from the lower Deuterostomia

The lamin gene of *Branchiostoma floridae* found by a screen with the Nematostella gene on chromosome UN has 10 introns. Their sizes and positions are given in Table 4 and Fig. 4 respectively. The flanking genes are as for the Nematostella gene. One of them encodes the ring finger protein 2/3 and the other encodes the methylmalonyl CoA epimerase (Fig. 6). This lamin gene is clearly a lamin B1. A second lamin gene indicated by cDNA accession number AJ271077 cannot be found in the currently available genomic sequence.

The urochordate *Ciona intestinalis* has a lamin gene on chromosome 12. It has 10 introns (Zimek and Weber, 2008). Their sizes and positions are given in Table 4 and Fig. 4 respectively. They show that

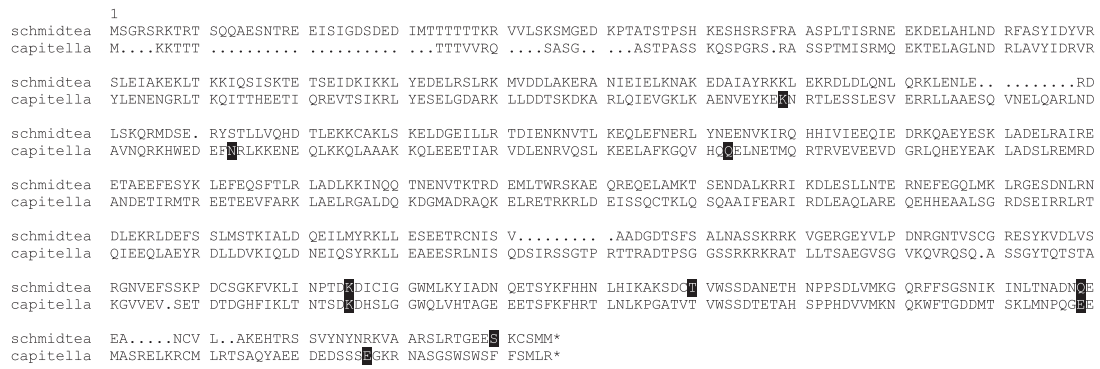


Fig. 3. Alignment of predicted lamin protein sequences for the planaria *Schmidtea mediterranea* (upper line) and the polychaete *Capitella teleta*. Intron positions are marked by inverted letters (white on black) in the sequence, for intron sizes see Table 3.

Ciona intestinalis has genome drift. The flanking genes encode a Ser/Thr specific protein phosphorylase and endoglin, a CD105 antigen. Previous cell biological studies showed that *Ciona* had only 5 intermediate filament proteins: a type I keratin, a type II keratin, a type III IF protein, a unique annexin-IF fusion protein and a single nuclear lamin (Karabinos et al., 2004). Curiously the *Ciona* lamin gene has a unique deletion, which removes the Ig-G-like middle domain of the tail domain. A second *Ciona* lamin gene predicted by the cDNA sequence for *Ciona* L1-□ (Acc. number AJ271075) cannot be found in the currently available *Ciona* genome data.

The lamin gene of *Strongylocentrotus purpuratus* present on scaffold 34693 has 9 introns (Zimek and Weber, 2008). Their sizes and positions are given in Table 4 and Fig. 4 respectively. The flanking genes are VPS33B and SMN1 (Fig. 6). Attempts to find a second lamin gene by a search with the coding sequence (CDS) of the known lamin gave no other gene. Thus it seems that at least this member of the echinoderms has only a single lamin gene and resembles the situation in the nematode *C. elegans* where a single gene is firmly established.

The lamin genes of vertebrates

Table 5 gives the sizes of the introns, and the positions of the introns are indicated along the deduced protein sequences in Fig. 5. The identity of the flanking genes when known is summarised in Fig. 6.

Mammals have 3 genes, lamins B1, B2 and lamin A. The latter is expressed only in differentiated cells (Rober et al., 1989). The human B1 gene is located on chromosome 5 and flanked by genes

encoding the phosphorylated adapter RNA export protein PHAX and the Membrane Associated Ring-finger protein 2/3 (MARCH 2/3-TYPE), which is also one flanking gene of the *Nematostella* lamin gene. The human B1 gene has 10 introns. The human B2 lamin gene lies on chromosome 19. It has 11 introns. The flanking genes encode TIMM13 and BADD45B. The human lamin A gene lies on chromosome 1. It is flanked by genes encoding the Ring-finger 2/3 and KH domain containing protein 4 and the sematophorin B precursor. It is interesting to note that the murine B2 gene maps on the distal arm of chromosome 10 (Zewe et al., 1991). This finding was made more than 10 years before the human and murine genomes were established.

Non mammalian vertebrates have 4 lamin genes due to a germline specific lamin gene. The number of 4 genes has been established only in the genome of *Xenopus tropicalis*. The B1 gene has 10 introns and lies on scaffold 7. It is flanked by the gene encoding Ring-finger protein 2/3 just as in the human B1 and *Nematostella* lamin genes (Fig. 6). The other flanking gene encodes an aldehyde dehydrogenase 7. The lamin B2 gene lies on scaffold 555. It has 11 introns. The flanking genes encode timm13 and RPL36. The lamin A gene has 12 introns and lies on scaffold 1266, which is very small so that no flanking genes are currently known. The germline specific gene B3 lies on scaffold 562. One flanking gene is SHGL1/2/3. The other flanking gene is not yet available. Interestingly the *Xenopus tropicalis* B3 protein shows 13% amino acid sequence identity with the *Xenopus laevis* protein. The corresponding B3 (LIII) gene was first described by Doring and Stick (1990). It is also described for the fish *Danio rerio* (Erber et al., 1999). We now find that the gene is located on chromosome 10. The flanking



Fig. 4. Alignment of predicted lamin protein sequences for the echinoderm *Strongylocentrotus purpuratus* (upper line), the tunicate *Ciona intestinalis* (lamin L2) and the chordate *Branchiostoma floridae* (lower line). Intron positions are marked by inverted letters (white on black) in the sequence, for intron sizes see Table 4.

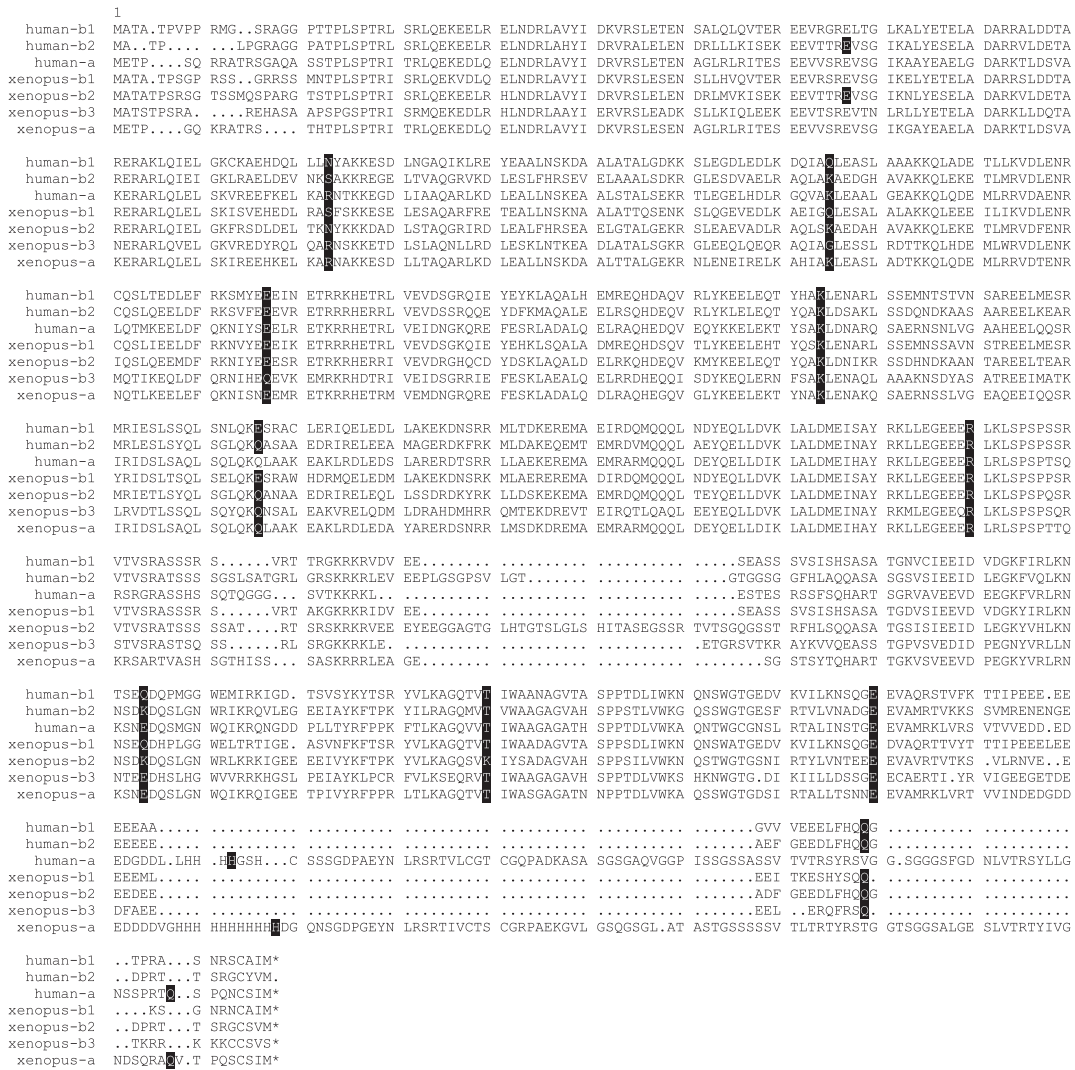


Fig. 5. Alignment of predicted lamin protein sequences for the three human lamin genes (B1, B2 and A/C) and the Xenopus lamins (B1, B2, B3 and A). Intron positions are marked by inverted letters (white on black) in the sequence, for intron sizes see Table 1.

gene encodes the phosphorylated adapter RNA export protein. The flanking gene on the other side is not yet available.

Phylogenetic basis of the number of lamin genes of metazoa

Radiata like the cnidarian *Nematostella* (Zimek and Weber, 2008), *Hydra* (this study) and the primitive placozoan *Trichoplax adhaerens* (this study) have a single lamin gene. Mammals have 3 lamin genes (Fig. 5) and non mammalian vertebrates have 4 genes because of a germline specific lamin gene. Only in *Xenopus tropicalis* are all 4 genes documented (this study). Other bilateral metazoa have only 2 lamin genes although the nematode *C. elegans* has only 1 gene.

This germline specific lamin gene with its 12 introns was originally described by Doring and Stick (1990) for *Xenopus laevis* and by Erber et al. (1999) for the fish *Danio rerio*. In the case of platypus *Ornithorhynchus anatinus* the genome provides currently only 3 genes, but practically no information on flanking genes. This will change in the future.

Discussion

We report the gene sequence of a larger number of additional lamin genes and where possible have identified the flanking genes.

This has provided some understanding on the number of lamin genes per species, which is 1 to 4. It is now established that radiata have only a single lamin gene. This was first seen in our analysis of the genome of *Nematostella vectensis* (Zimek and Weber, 2008) and is now extended to *Hydra magnipapillata* and to *Trichoplax adhaerens*. Equally well established is that mammals have three lamin genes B1, B2 and A. Non mammalian vertebrates have 4 lamin genes due to an additional germ line specific gene (Doring and Stick, 1990). The 4 lamin genes of non mammalian vertebrates are currently only documented for *Xenopus tropicalis*. More problematic is the question which bilateral animals express 2 or only 1 lamin. Thus the genome of the nematode *C. elegans* has only 1 lamin gene, while *Drosophila melanogaster* clearly has 2, the generally expressed lamin *Dmo* gene and the gene C, which is additionally expressed in differentiated cells. The sea urchin *Strongylocentrotus purpuratus*, a member of the echinoderms, may have only a single lamin gene. A further candidate for only one gene may be *Schmidtea mediterranea*, a member of the plathelminthes. However, we note that for many bilateral animals only one lamin gene is known and attempts to find a second one have not been made. So one can only say that currently the number of lamin genes per species follows more or less metazoan phylogeny.

The importance of identifying the nature of the genes flanking the lamin genes is illustrated by two observations. Already

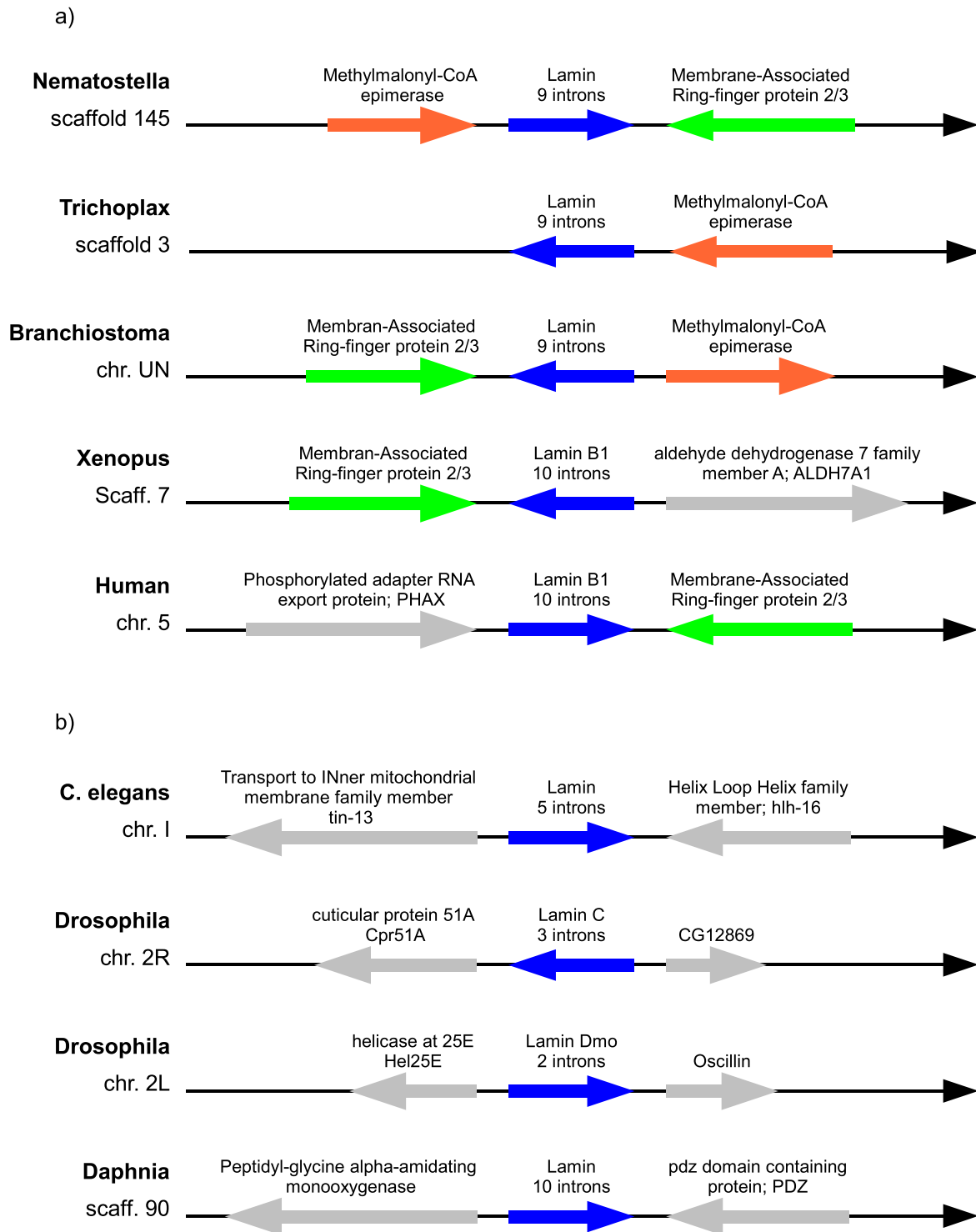


Fig. 6. The flanking genes of metazoan lamins indicate whether the species have a strong genome drift (b) or not (a). Data on flanking genes has been retrieved from the online genome resources (see Materials and methods).

Erber et al. (1999) noted that the lamin protein sequence of the nematode *C. elegans* was the most remote sequence of metazoan known at that time and coined the phrase evolutionary or genomic drift. This observation was extended in the study of Raible et al. (2005) to *Drosophila melanogaster* and related insects. We showed subsequently (Zimek and Weber, 2008) that the genes of the nematode *C. elegans* and the two *Drosophila* lamins had either very many fewer introns and at totally new positions (*C. elegans* lamin and *Drosophila* lamin Dmo) or very few introns

with only some located at otherwise known positions (*Drosophila* lamin C gene). Here we have shown that for the lamin genes of *Drosophila* and the single lamin gene of *C. elegans* the flanking genes are not only totally different than in another member of the ecdysozoa (*Daphnia pulex*), but have also no relation to each other. This is particularly striking for the single *C. elegans* lamin gene and the *Drosophila* Dmo gene, which fulfil the same function. Using now one, two or even three of these criteria we know that not only *Drosophila* and *C. elegans* have a strong genome

drift, but that such a drift occurred also in the urochordate *Ciona intestinalis*, the annelid *Capitella* and the planarian *Schmidtea mediterranea*.

In contrast in species without a strong evolutionary drift there are astonishing examples for the same flanking genes. Thus the single *Nematostella* gene has like the *Branchiostoma floridae* lamin B1 gene characterised here, the same flanking genes - Membrane Associated Ring-finger protein 2/3 and Methylmalonyl-CoA epimerase -. Even in the vertebrates B1 genes of *Xenopus tropicalis* and man the Membrane Associated Ring-finger protein 2/3 is retained as one of the flanking genes. One wonders whether a laboratory with strong computing facilities would not make a major impact when instead of only 2 flanking genes the number would be increased to 3, 4 or even 5. Maybe one would begin to understand how metazoan genomes expanded in the course of normal evolution. Such an analysis could be done with well-documented genomes such as the human or mouse genome and that of *Nematostella*. We note that once genomes are missing information on one of the two flanking genes they cannot be used in a more demanding analysis. We know there are plans to sequence around 20 more animal genomes. While we do not doubt that some of them are important to establish the relation between different metazoa, the real importance of a new genome is apparent only when it approaches completion. The lack of information on flanking genes is a serious drawback. Thus no flanking genes of the *Hydra* lamin gene are known and only one flanking gene for the *Trichoplax* lamin gene. The lamin gene of the mollusc *Lottia gigantea* is not complete at the 3'-end, because of a big gap in which one flanking gene lies. Even for the recently established genome of *Xenopus tropicalis* which provides all 4 non mammalian lamin genes, one flanking gene is not yet available. In our last report (Zimek and Weber, 2008) we pointed out that no information was available on Ctenophora and Porifera and that a genome project would have to be started on a member of the Porifera. Indeed now sequences become available for *Amphimenon queenslandii* (Srivastava et al., 2009; Jackson et al., 2010). This leaves the problem of Ctenophora, the sister group of the cnidaria. If it were only for the lamin gene the procedure we described for the cDNA of *Hydra attenuata* by Erber et al. (1999) could be used.

Lamin genes seem specific for metazoa. No paralogue was found in the yeast *Saccharomyces cerevisiae* (Erber et al., 1999). However, all non metazoan eukaryotes contain a lamin like organisation supporting the inner side of the nuclear membrane. In the parasite *Trypanosoma brucei* the NUP1 antigen of the nuclear envelope is a coiled coil protein containing about 20 near perfect copies of a 144 amino acid sequence. Immuno electron microscopy localised NUP1 to the inner face of the nuclear envelope suggesting that it is a major filamentous component of the Trypanosome lamina (Rout and Field, 2001).

We point out that our approach may be useful for other kingdoms, provided essential genes are used and some guiding information, usually morphogenetic data is available. Alternatively, the procedure may rule out one or the other model discussed in the literature. Finally we speculate on the origin of cytoplasmic IF proteins and their genes. Since they are not present in cnidaria (Zimek and Weber, 2008), they must have originated in an early bilateral species. As porifera are known to have in some cell types desmosomal junctions (Wehner and Gehring, 2007) and these anchor usually keratin filaments, the genome draft of the sponge *Amphimenon queenslandii* may allow this hypothesis to be explored in the future.

Since we are interested in lamin genes and their flanking genes, we have not considered here the lamin gene of the hemichordate *Saccoglossus* and the lamin gene of the mollusc *Lottia gigantea* and the three genes of the non mammalian vertebrate *Ornithorhynchus anatinus*.

Acknowledgements

We thank Dr. Kerstin Bartscherer from the MPI for Molecular Medicine in Muenster for pointing us to the start of the Genome Project of *Schmidtea mediterranea* and Dr. Mary Osborn, this Institute, for helpful comments on the manuscript. Cornelia Maggs provided excellent secretarial help. This study was supported by "Der Fonds der Deutschen Chemischen Industrie" to Klaus Weber.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejcb.2010.10.005.

References

- Bartnik, E., Weber, K., 1989. Widespread occurrence of intermediate filaments in invertebrates; common principles and aspects of diversification. *Eur. J. Cell Biol.* 50, 17–33.
- Bonifas, J.M., Rothman, A.L., Epstein, E.H.J., 1991. Epidermolysis bullosa simplex: evidence in two families for keratin gene abnormalities. *Science* 254, 1202–1205.
- Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., Disbennett, K., Pfannkoch, C., Sumin, N., Sutton, G.G., Viswanathan, L.D., Walenz, B., Goodstein, D.M., Hellsten, U., Kawashima, T., Prochnik, S.E., Putnam, N.H., Shu, S., Blumberg, B., Dana, C.E., Gee, L., Kibler, D.F., Law, L., Lindgens, D., Martinez, D.E., Peng, J., Wigge, P.A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J.S., Ikee, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P.R., Zhang, X., Aufschnaiter, R., Eder, M.K., Gornyy, A.K., Salvenmoser, W., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Bottger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K.A., Strausberg, R.L., Venter, J.C., Technau, U., Hobmayer, B., Bosch, T.C., Holstein, T.W., Fujisawa, T., Bode, H.R., David, C.N., Rokhsar, D.S., Steele, R.E., 2010. The dynamic genome of *Hydra*. *Nature* 464, 592–596.
- 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.
- Doring, 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., Hofemeister, H., Bovenschulte, M., Stick, R., Panopoulou, G., Lehrach, H., Weber, K., 1999. Characterization of the *Hydra* lamin and its gene: a molecular phylogeny of metazoan lamins. *J. Mol. Evol.* 49, 260–271.
- Gruenbaum, Y., Margalit, A., Goldman, R.D., Shumaker, D.K., Wilson, K.L., 2005. The nuclear lamina comes of age. *Nat. Rev. Mol. Cell Biol.* 6, 21–31.
- 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.
- Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466, 720–726.
- 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. U.S.A.* 98, 7863–7868.
- Karabinos, A., Zimek, A., Weber, K., 2004. The genome of the early chordate *Ciona intestinalis* encodes only five cytoplasmic intermediate filament proteins including a single type I and type II keratin and a unique IF-annexin fusion protein. *Gene* 326, 123–129.
- Liu, J., Rolef Ben-Shahar, T., Riemer, D., Treinin, M., Spann, P., Weber, K., Fire, A., Gruenbaum, Y., 2000. Essential roles for *Caenorhabditis elegans* lamin gene in nuclear organization, cell cycle progression, and spatial organization of nuclear pore complexes. *Mol. Biol. Cell* 11, 3937–3947.
- Odriontz, F., Becker, S., Kollmar, M., 2009. Reconstructing the phylogeny of 21 completely sequenced arthropod species based on their motor proteins. *BMC Genomics* 10, 173.
- Osman, M., Paz, M., Landesman, Y., Fainsod, A., Gruenbaum, Y., 1990. Molecular analysis of the *Drosophila* nuclear lamin gene. *Genomics* 8, 217–224.
- Raible, F., Tessmar-Raible, K., Osoegawa, K., Wincker, P., Jubin, C., Balavoine, G., Ferrier, D., Benes, V., de Jong, P., Weissenbach, J., Bork, P., Arendt, D., 2005. Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* 310, 1325–1326.
- Riemer, D., Stuurman, N., Berrios, M., Hunter, C., Fisher, P.A., Weber, K., 1995. Expression of *Drosophila* lamin C is developmentally regulated: analogies with vertebrate A-type lamins. *J. Cell Sci.* 108, 3189–3198.

- Rober, R.A., Weber, K., Osborn, M., 1989. Differential timing of nuclear lamin A/C expression in the various organs of the mouse embryo and the young animal: a developmental study. *Development* 105, 365–378.
- Rout, M.P., Field, M.C., 2001. Isolation and characterization of subnuclear compartments from *Trypanosoma brucei* Identification of a major repetitive nuclear lamina component. *J. Biol. Chem.* 276, 38261–38271.
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, Odronitz, F., Becker, S., Kollmar, M., 2009. Reconstructing the phylogeny of 21 completely sequenced arthropod species based on their motor proteins. *BMC Genomics* 10, 173.C.
- Vassar, R., Coulombe, P.A., Degenstein, L., Albers, K., Fuchs, E., 1991. Mutant keratin expression in transgenic mice causes marked abnormalities resembling a human genetic skin disease. *Cell* 64, 365–380.
- Weber, K., Plessmann, U., Traub, P., 1989. Maturation of nuclear lamin A involves a specific carboxy-terminal trimming, which removes the polyisoprenylation site from the precursor; implications for the structure of the nuclear lamina. *FEBS Lett.* 257, 411–414.
- Wehner, R., Gehring, W., 2007. *Zoologie G.* Thieme Verlag, Stuttgart.
- Worman, H.J., Bonne, G., 2007. "Laminopathies": a wide spectrum of human diseases. *Exp. Cell Res.* 313, 2121–2133.
- Zewe, M., Hoger, T.H., Fink, T., Lichter, P., Krohne, G., Franke, W.W., 1991. Gene structure and chromosomal localization of the murine lamin B2 gene. *Eur. J. Cell Biol.* 56, 342–350.
- 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.
- Zimek, A., Weber, K., 2008. In contrast to the nematode and fruit fly all 9 intron positions of the sea anemone lamin gene are conserved in human lamin genes. *Eur. J. Cell Biol.* 87, 305–309.