

Research article

Four signature motifs define the first class of structurally related large coiled-coil proteins in plants.

Frank Gindullis¹, Annkatrin Rose², Shalaka Patel² and Iris Meier^{*2}

Address: ¹CellTec Biotechnologie GmbH, Frohmestrasse 110, D-22459 Hamburg, Germany and ²Plant Biotechnology Center and Department of Plant Biology, Ohio State University, 210 Rightmire Hall, 1060 Carmack Road, Columbus, Ohio 43210, USA

E-mail: Frank Gindullis - frank.gindullis@celltec.de; Annkatrin Rose - rose.273@osu.edu; Shalaka Patel - patel.492@osu.edu; Iris Meier* - meier.56@osu.edu

*Corresponding author

Published: 9 April 2002

Received: 23 January 2002

BMC Genomics 2002, 3:9

Accepted: 9 April 2002

This article is available from: <http://www.biomedcentral.com/1471-2164/3/9>

© 2002 Gindullis et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Animal and yeast proteins containing long coiled-coil domains are involved in attaching other proteins to the large, solid-state components of the cell. One subgroup of long coiled-coil proteins are the nuclear lamins, which are involved in attaching chromatin to the nuclear envelope and have recently been implicated in inherited human diseases. In contrast to other eukaryotes, long coiled-coil proteins have been barely investigated in plants.

Results: We have searched the completed Arabidopsis genome and have identified a family of structurally related long coiled-coil proteins. *Filament-like plant proteins (FPP)* were identified by sequence similarity to a tomato cDNA that encodes a coiled-coil protein which interacts with the nuclear envelope-associated protein, MAF1. The FPP family is defined by four novel unique sequence motifs and by two clusters of long coiled-coil domains separated by a non-coiled-coil linker. All family members are expressed in a variety of Arabidopsis tissues. A homolog sharing the structural features was identified in the monocot rice, indicating conservation among angiosperms.

Conclusion: Except for myosins, this is the first characterization of a family of long coiled-coil proteins in plants. The tomato homolog of the FPP family binds in a yeast two-hybrid assay to a nuclear envelope-associated protein. This might suggest that FPP family members function in nuclear envelope biology. Because the full Arabidopsis genome does not appear to contain genes for lamins, it is of interest to investigate other long coiled-coil proteins, which might functionally replace lamins in the plant kingdom.

Background

Coiled-coil domains are protein oligomerization motifs, which consist of two or more alpha helices that twist around one another to form a supercoil [1]. Peptides with the capacity to form coiled coils are characterized by a heptad repeat pattern in which residues in the first and fourth position are hydrophobic, and residues in the fifth

and seventh position are predominantly charged or polar. This pattern can be used by computational methods to predict coiled-coil domains in amino acid sequences [2,3].

Coiled-coil proteins can be grouped into two general classes. The first class is comprised of short coiled-coil do-

mains of six or seven heptad repeats, also called leucine zippers. They are frequently found as homo- and heterodimerization domains in transcription factors (e.g. [4]). The second class is defined by long coiled-coil domains of several hundred amino acids, which are found in a variety of proteins involved in structuring cellular processes [1].

One of the three main classes of cytoskeletal proteins, the intermediate-filament proteins, represents a well-characterized group of coiled-coil proteins. A subgroup of intermediate-filament proteins are the nuclear lamins, which are involved in attaching chromatin to the nuclear envelope and have recently been implicated in inherited human diseases [5,6]. Besides structural proteins of the cytoskeleton, the motor proteins that interact with them also contain coiled-coil motifs. Myosin, the actin motor protein, has an extended coiled-coil domain necessary for the assembly of the muscle thick filaments [7]. Kinesin and dynein, two microtubule motor proteins, also contain coiled-coil domains [8,9].

The coiled-coil motif has also been identified in a variety of proteins associated with centromeres, centrosomes, the nuclear matrix, and chromatin [10–14]. TPR, a protein associated with the inner filaments of the nuclear pore complex is an extended coiled-coil protein [15]. The filamentous structures associated with both the outer and inner surface of the nuclear pore complex suggest that additional unknown coiled-coil proteins might be involved in its assembly. Several large coiled-coil proteins were also found associated with the cytoplasmic surface of the Golgi apparatus, and two of them have been shown to function in the docking of vesicles to the Golgi cisternae [16–18]. AKAPs are adapter proteins that attach protein kinase PKA pathways to cytoskeletal elements. At least one AKAP, AKAP 450 is a large coiled-coil protein attaching PKA to the centrosome [19].

Despite the multitude of cellular functions in which coiled-coil proteins participate, the general theme is the association of proteins with the solid-state components of the cell. In some cases, as in the attachment of PKA, this function appears to be crucial for the spatial regulation of signal transduction. Importantly, mutations in individual coiled-coil proteins lead to specific developmental or even behavioral phenotypes. The *Drosophila* mutations *Mushroom body defect* (a gene involved in the proliferation of neuronal precursor cells) and *quick-to-court* (a mutation that causes elevated sexual behavior) lie in two tissue-specifically expressed coiled-coil proteins of unknown function [20,21]. In humans and mouse models, specific point mutations in nuclear lamins are the cause for autosomal Emery-Dreifuss muscular dystrophy [22,23].

In contrast to animals and yeast, only a small number of long coiled-coil proteins have been identified from plants. MFP1 contains an extended coiled-coil domain and is located at the nuclear periphery of tobacco suspension culture cells [24,25]. The carrot protein NMCP1 is also located at the nuclear rim and has been shown to migrate to the spindle poles in dividing carrot suspension culture cells [26,27]. The cellular function of these plant proteins is not known.

Several candidates for myosins and kinesins have been identified from plants and their function as motor proteins is under investigation [28,29]. Besides these few examples, nothing is presently known about plant long coiled-coils and their potential function in anchoring and structuring of different cellular events.

In the early nineties, several groups reported lamin-like proteins in different plant species [30–32]. A purification protocol adapted from animal lamins and applied to pea nuclei was used to purify four proteins between 49 and 66 kD [31]. These proteins were recognized by antibodies against mammalian intermediate filament proteins, and by antibodies against a peptide derived from lamin B. By immunofluorescence and immunogold labeling, these antigens were located mostly in the internal nuclear matrix, and not predominantly at the nuclear rim. Similar results were obtained by other investigators [32]. Although these early reports were promising, no lamin ortholog has been identified molecularly from plants. In a recent search of all publicly available plant sequences, including the full *Arabidopsis* genome, no homologs of lamins were found ([33] and Meier, unpublished results). Similarly, despite earlier reports of lamin-like proteins in yeast [34], the fully sequenced *S. cerevisiae* genome [35] also contains no lamin genes. It is therefore likely that non-animal eukaryotes have a distinct set of nuclear envelope proteins that functionally replace the lamins.

Here, we have used a tomato protein containing long coiled-coil domains to search the *Arabidopsis* genome for related sequences. We have identified seven novel *Arabidopsis* proteins with extended coiled-coil domains, which form a protein family distinct from other *Arabidopsis* long coiled-coil proteins. They are characterized by four novel motifs, which are highly conserved in sequence and position, and which were not found in proteins outside of this family. In addition, a rice homolog was identified, indicating conservation between dicots and monocots.

Results

LeFPP, a novel plant coiled-coil protein

A yeast two-hybrid screen was performed to identify interaction partners of the tomato nuclear envelope-associated protein LeMAF1 [36]. A prey plasmid (pAD-LeFPP) was

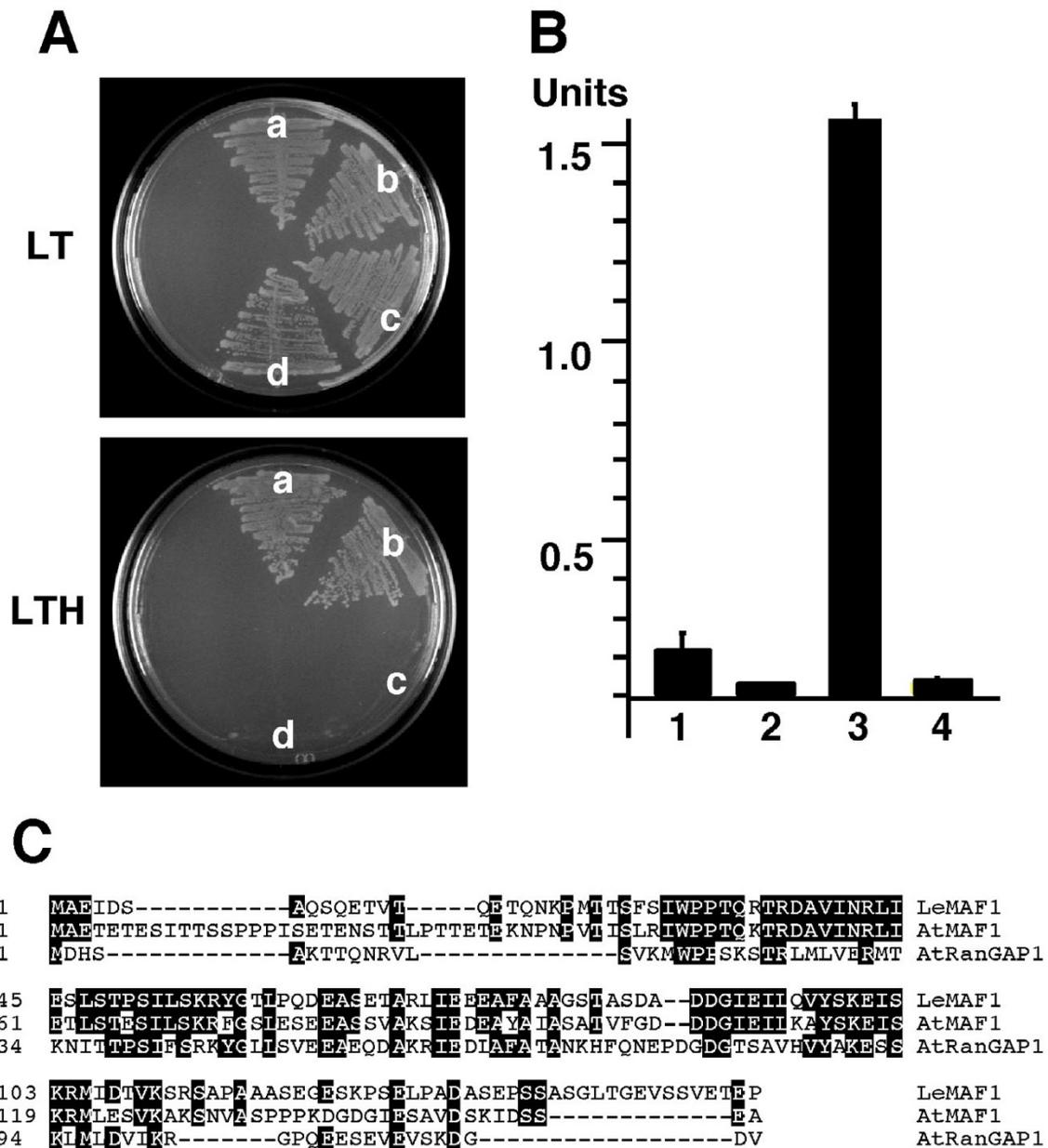


Figure 1

Yeast two-hybrid interaction of LeFPP with MAF1. (A) Yeast colonies co-expressing bait and prey plasmids were streaked on dropout plates without leucine and tryptophane (LT) or dropout plates without leucine, tryptophane, and histidine (LTH). (a) pBD-LeMAF1/pAD-LeFPP; (b) pBD-AtMAF1/pAD-LeFPP; (c) pBD-AtRanGAP1/pAD-LeFPP; (d) pBD-GAL4/pAD-LeFPP. (B) Quantification of β -Galactosidase activity in yeast strains harboring combinations of bait and prey plasmids. (1) p53/pSV40 (positive control); (2) pLaminC/pSV40 (negative control); (3) pBD-LeMAF1/pAD-LeFPP; (4) pBD-GAL4/pAD-LeFPP. Units, relative β -Galactosidase units (see Methods). (C) Sequence alignment between LeMAF1, AtMAF1, and the N-terminal domain of AtRanGAP1. Identical amino acids are shaded in black.

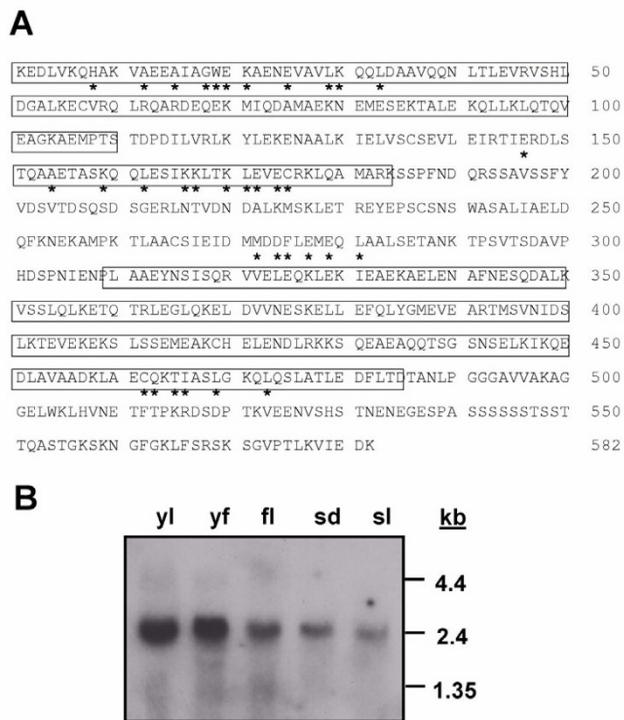


Figure 2
LeFPP is a ubiquitously expressed coiled-coil protein. (A) Non full-length amino acid sequence of LeFPP (GenBank accession number AF405309). The Lysine in position 1 is the first amino acid encoded by the LeFPP cDNA in frame with the GAL4 activation domain. The coiled-coil regions are boxed. The residues correlating to motifs I, II, III, and IV are indicated by asterisks below the sequence (see text). (B) RNA blot analysis of LeFPP RNA accumulation in different tomato tissues. Approximately equal amounts of total RNA (20 µg) were loaded per lane, as confirmed by ethidium-bromide staining (data not shown). yl, young leaves; yf, young fruits; fl, fully open flowers; sd, light-grown seedlings; sl, dark-grown seedlings. The positions of RNA size markers are indicated on the right.

identified that led to activation of the two reporter genes *HIS* and *LacZ* after retransformation with the pBD-LeMAF1 bait. We tested the specificity of this interaction by co-transforming pAD-LeFPP with two additional bait plasmids, pBD-AtMAF1 and pBD-AtRanGAP1. AtMAF1 is one of three Arabidopsis homologs of LeMAF1 (Patel and Meier, unpublished results), and has 45% amino acid identity with LeMAF1. AtRanGAP1 is one of the two Arabidopsis Ran GTPase activating proteins. Its N-terminus has 28% amino acid identity with LeMAF1 and 30% with AtMAF1 (Fig. 1C; [33,37]). Fig. 1A shows that pAD-LeFPP leads to activation of the *HIS* gene in combination with pBD-LeMAF1 (sector a) and with pBD-AtMAF1 (sector b), but not with pBDAtRanGAP1 (sector c) or the binding do-

main vector pBD-GAL4 alone (sector d). This demonstrates that the observed interaction is specific for MAF1-like proteins.

Fig. 1B shows the quantification of β -Galactosidase activity in the pBD-LeMAF1/pAD-LeFPP strain. Lanes one and two are the positive and negative control provided with the kit, respectively, lane three shows the activity of the pBD-LeMAF1/pAD-LeFPP strain and lane 4 the activity of the pBD-GAL4/pAD-LeFPP strain. Although yeast two-hybrid interactions can not be easily quantified, the very high level of β -Galactosidase expressed by the pBD-LeMAF1/pAD-LeFPP stain suggests a comparably strong interaction in the yeast two-hybrid screen.

The 1927 bp pAD-LeFPP cDNA was sequenced and was shown to contain an uninterrupted open reading frame in frame with the GAL4 activation domain. The conceptual translation of this sequence leads to a protein of 582 amino acids (aa) with a calculated molecular weight of 64.4 kDa and a pI of 4.7. The first methionine is at position 71, and does not have a plant start ATG consensus sequence [38], suggesting that the cDNA is not full length.

A BLAST search with the pAD-LeFPP open reading frame indicated that it represents a novel protein with weak similarity to filament-like proteins from animals and yeast (see below). Secondary structure prediction algorithms showed that it is organized almost entirely in alpha-helical domains and contains extensive stretches of coiled-coil domains (Fig. 2A and Fig 3). The protein was named LeFPP for tomato filament-like plant protein. RNA blot analysis showed a single species of mRNA of ca. 2.4 kb. It is present in tomato leaves, fruits, flowers, light-grown seedlings, and dark-grown seedlings (Fig. 2B).

A family of LeFPP-like proteins in Arabidopsis

A BLAST search was performed with the LeFPP amino acid and DNA sequence. In a protein BLAST, sequences with highest scores represented a number of uncharacterized open reading frames in the Arabidopsis genome, followed by sequences for animal filament-like proteins such as myosin, kinesin, ankyrin, etc. While the e-values for the Arabidopsis ORFs were between $2e-56$ and $1e-16$, the best e-value for a non-plant sequence was 0.003 for an unknown human open reading frame. The Arabidopsis "hits" represented seven unique genes, some of which had repeated entries in GenBank (see Table 1).

A translated BLAST search identified one additional "hit" in the rice BAC H0212B02 of chromosome 4 (GenBank accession number AL442007). The "hit" lay between the annotated open reading frames H0212B02.17 and H0212B02.18 in a region of DNA not predicted to have coding capacity. However, running GenScan [39] on this

Table 1: Genomic organization and features of the Arabidopsis FPP gene family.

Name	GenBank Accession (TAIR name ^a)	Alias ^b	Chr. # (Map pos. (Mb))	Size ORF (aa) (Size protein (kDa))	pI	EST ^c	library
AtFPP1	AGG51666 (AT1G77580)	T5M16.17	1 (28.9)	779 (87.8)	4.7	AV524363	Seedlings, aboveground org.
						APZL64f04F	Mixed tissues
AtFPP2	AAF16547 (AT1G21812)	T26F17.2	1 (7.65)	629 (71.1)	4.9	AV524363	Seedlings, aboveground org.
		F8K7.24					
AtFPP3	AAF27033 (AT3G05270)	T12H1.24	3 (1.5)	603 (68.3)	5.3	AV547438	Roots
		T15D2.30				T22555	Mixed tissues
AtFPP4	AAG12549 (AT1G19835)	F6F9.12	1 (6.8)	1024 (114.5)	5.2	AV528396	Seedlings, aboveground org.
		F14P1.46				AV556649	Green siliques
		F6F9.17				AV546174	Roots
						AV547135	Roots
AtFPP5	CAB81521 (AT4G36120)	T19K4.250	4 (16.3)	981 (110.4)	6.3	RAFL16-63- O05 ^d	unknown
AtFPP6	AAG51782 (AT1G47900)	T6B12.1	1 (17.2)	1054 (118.5)	4.8	AA042379	Hypocotyls
						N96153	Mixed tissues
AtFPP7	AAC23780	F26B6.1	2	886	5.7	W43775	Hypocotyls
	(AT2G23360)	T20D16.1	(10.24)	(99.9)			

^aThe Arabidopsis Information Resource. ^bAdditional early annotation names of the same protein. ^cGenBank accession numbers. ^dRIKEN Arabidopsis Full-Length cDNA [http://www.gsc.riken.go.jp/Plant/]

part of the BAC, an open reading frame of 1074 amino acids was identified on the opposite strand that contained the "hit". The predicted coding region begins at position 114,922 of the H0212B02 DNA sequence, ends at position 107,317, and contains seven predicted introns. This rice open reading frame was named OsFPP and was included in the sequence and structural analysis shown in Fig. 3 and Fig. 4.

Fig. 3 shows a comparison of the secondary structure of LeFPP with the seven Arabidopsis ORFs and the rice ORF. All open reading frames are characterized by extended alpha-helical and coiled-coil domains. The extension of the coiled-coil domains shown in Fig. 3 is derived from the "coiled-coil" analysis software in the DNASTAR sequence

analysis software package, and was confirmed in all cases by the MultiCoil algorithm [2]. In addition, MultiCoil predicted that all coiled-coil regions have a high probability to form dimers and a low probability to form trimers. The seven Arabidopsis ORFs were named AtFPP1 – AtFPP7 (Fig. 3 and Table 1). They fall into two groups according to size and extension of coiled-coil domains. AtFPP1, AtFPP2, and AtFPP3 are shorter (603 aa to 779 aa) and contain two extended coiled-coil domains connected by a short non-coiled-coil linker. AtFPP4 to AtFPP7 are longer (866 aa to 1054 aa) and contain a longer N-terminal and a shorter C-terminal coiled-coil domain, separated by a long non-coiled-coil linker domain (Fig. 3). The rice ORF (OsFPP) resembles structurally more the second

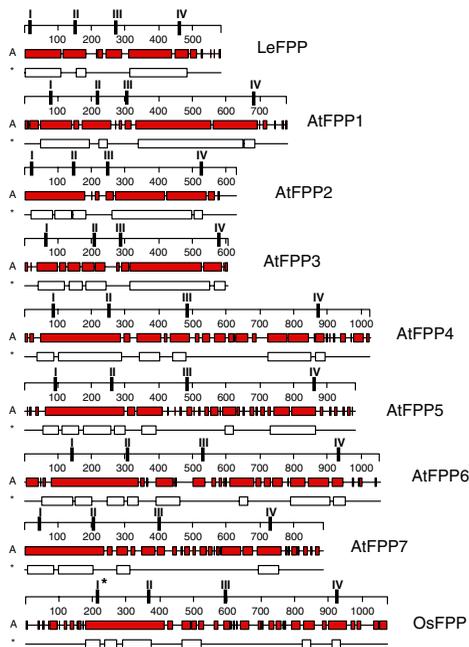


Figure 3

Comparison of predicted secondary structure of FPP proteins. All proteins are drawn to scale. The numbers on the scales represent amino acids. Alpha helices are indicated as gray bars (A); coiled-coil domains are indicated as white bars (*). The positions of the four conserved sequence motifs are indicated as I, II, III, and IV on the scales. I* indicates the variant of motif I found in OsFPP.

group of Arabidopsis ORFs, with two shorter stretches of coiled-coil domains separated by a long linker.

Novel sequence motifs define a subfamily of plant coiled-coil proteins

Alignment of all nine amino acid sequences showed the presence of four novel sequence motifs, which are 100% conserved between the tomato and Arabidopsis proteins (Fig. 4A). Motifs II, III, and IV are also 100% conserved in OsFPP, while motif I is present, but slightly less conserved in the monocot sequence. Their location in each protein is indicated in Fig. 3. Their position with respect to the arrangement of the coiled-coil domains is strikingly conserved, despite differences in spacing between the four motifs due to the different length of the proteins and the coiled-coil domains.

Motifs I and II always correlate with the N-terminal cluster of coiled-coil domains, with motif I being located at the beginning and motif II at the end of this cluster. Motif III is always located in the linker between the N-terminal and C-terminal coiled-coil domains. Motif IV is always located at the very end of the most C-terminal coiled-coil domain. This positioning is conserved even in the rice sequence, where the C-terminal coiled-coil domain is extremely short (Fig. 3).

The amino acid identities between the eight dicot proteins are in the range of 20% to 50% (Fig. 4B), with the rice sequence showing between 13% and 26% identity with the dicot sequences. The Arabidopsis protein with the greatest similarity to LeFPP is AtFPP1 (40% identity).

Table 1 shows an overview of the genomic organization of *AtFPP1* through *AtFPP7* and of the features of the predicted proteins. The genes are not clustered, and no two of them are arranged in tandem array. Except for chromosome 5, at least one member of this gene family is present on each chromosome. ESTs have been identified for all genes, indicating that they are expressed (Table 1). Consistent with the ubiquitous expression pattern found for *LeFPP*, *AtFPP* expression covers a variety of tissues and developmental stages including shoots, hypocotyls, roots, and siliques.

In order to identify more long coiled-coil proteins in the Arabidopsis genome, the complete list of annotated ORFs was searched with the keyword "myosin" to identify ORFs that had been annotated as myosin-like proteins. Because of the long stretches of conserved heptad repeats, BLAST searches with large coiled-coil proteins usually produce significant e-values with at least some of the large number of myosin heavy-chain sequences in GenBank, frequently leading to this annotation.

The 40 ORFs identified this way were analyzed by Multi-Coil and 37 were found to contain coiled-coil domains of at least 200 amino acids. 3 ORFs showed no coiled-coil motifs and were discarded. AtFPP2 and AtFPP5 had been annotated as "myosin-like" too and were therefore identified in this search. They were discarded to avoid duplication. None of the remaining 35 new coiled-coil proteins contained any of the conserved motifs identified in the FPP family.

The 35 new sequences were aligned with the seven AtFPP sequences in a CLUSTAL analysis. Fig. 5 shows the results represented as a phylogenetic tree. The AtFPP family clearly forms a family of sequences separate from the other coiled-coil proteins (boxed in Fig. 5). The three "short" AtFPPs, AtFPP1, AtFPP2, and AtFPP3 form a sub-family, as do the four "long" AtFPPs AtFPP4, AtFPP5, AtFPP6, and

A

Motif I

```

86  EKVVAEEAVAGWEKAEENEVVELKKEPEAADDKNRVLEDRVSHLDGALKECVRO AtFPP1
23  EAKVAEDAVAGWEKAEENEVVELKQKPEAADDKNIVLEDRVSHLDGALKECVRO AtFPP2
71  EAKVAEEAVSWEKAEENEAALKQKPEASTSKVSALEDRNSHLDGALKECVRO AtFPP3
89  ESKVAEEAVTGWKAEPEASALKTHPETITLAKLTVEDRAAHLDDALKECMRO AtFPP4
99  EAKVAEEAVSWEKADAEPLALKRQPESVTLKLTAEERASHLDDALKECTRO AtFPP5
139 ESKVAEDAVSWEKADAEALALKNTPESVTLKLTAEERASHLDDALKECMRO AtFPP6
40  EETEAOEATVWEKTKPEVASLKKKPEALNEKHRSEERSSHTDAGLKECVQO AtFPP7
8   EAKVAEEAIAWEKAEENEVAVLKQKPEAAVQONLTLEVRVSHLDGALKECVRO LeFPP
228 QAKVEEALIGWFAECPAIAIKTQDDTLDQKAAQKICQDEANVAMVBOsFPP
    
```

Motif II

Motif III

Motif IV

```

213 ERDLSTQAAETASQOQDSIKKVAKLEAEC 302 MDDFLEMERT 667 CCKTIASGKQI AtFPP1
143 ERDLSTQAAETASQOHLDIKKLAKLEAEC 244 MDDFLEMERT 514 CCKTIASGKQI AtFPP2
206 ERDLSTQAAESASQOLEGIRKLTAKLEAEC 294 MDDFLEMERT 579 CCKTIASGQRI AtFPP3
251 EKNMSMRSAAEAAKQOHEGVKKIAKLEAEC 477 MDDFLEMERT 878 CCKTIILGKQI AtFPP4
261 EKNMSLKSADIANQOHEGVKKIAKLEAEC 486 MDDFLETEKL 876 CCKTIHLGKQI AtFPP5
301 EKNMCIRSAESANKQOHEGVKKIAKLEAEC 523 MDDFLEMERT 937 CCKTIIVGKQI AtFPP6
202 EREFSRRTAEASHKLEHENVKRVAKLESEC 400 MDDFLEMERT 731 CCKTILNIGKQI AtFPP7
146 ERDLSTQAAETASQOLESIKKLTAKLEVEC 272 MDDFLEMEQL 462 CCKTIASGKQI LeFPP
8   ERKFNLKSADAABQOHEENVKRTAKLEAEC 592 MDDFLETEKL 921 CCKTILNIGKQI OsFPP
    
```

B

Percent Identity										
1	2	3	4	5	6	7	8	9		
████	49.8	34.2	24.1	19.3	19.6	16.9	40.0	12.7	1	AtFPP1
	████	29.2	20.3	20.0	21.9	21.5	35.7	18.0	2	AtFPP2
		████	26.5	24.2	26.0	20.1	30.4	19.7	3	AtFPP3
			████	35.3	45.7	23.0	25.4	18.9	4	AtFPP4
				████	32.7	19.5	21.1	20.1	5	AtFPP5
					████	22.3	24.4	19.7	6	AtFPP6
						████	21.1	26.0	7	AtFPP7
							████	16.6	8	LeFPP
								████	9	OsFPP

Figure 4

Four novel sequence motifs define a sub family of plant coiled-coil proteins. (A) Sequence alignment of motifs I, II, III, and IV. In all four alignments, amino acid residues conserved among all dicot sequences are indicated in black. Amino acids in the rice sequence that match these conserved residues are also indicated in black. All residues are matched by the rice sequence in motifs II, III, and IV, while 6 residues are exchanged in the rice sequence in motif I. Numbers on the left indicate the position of the first residue of each motif in the amino acid sequence of the respective protein. (B) Percent amino acid identity between FPP proteins.

AtFPP7 (see Fig. 3). Interestingly, the two most closely related *AtFPP* genes, *AtFPP1* and *AtFPP2* are located in comparable positions on two duplicated segments of chromosome 1, indicating that they might be derived from a recent duplication event.

Discussion

A novel class of plant coiled-coil proteins has been identified that are conserved in dicots and monocots. Seven members of the FPP protein family are present in the Ara-

bidopsis genome. They are characterized by the presence of two clusters of coiled-coil domains separated by a linker domain of variable size. Four highly conserved sequence motifs were identified, which are a signature feature of this protein family and which are also conserved in the homologs LeFPP and OsFPP from tomato and rice, respectively. The tomato homolog of the FPP family binds in a yeast two-hybrid assay to a nuclear envelope-associated protein, indicating that they might function in nuclear envelope biology.

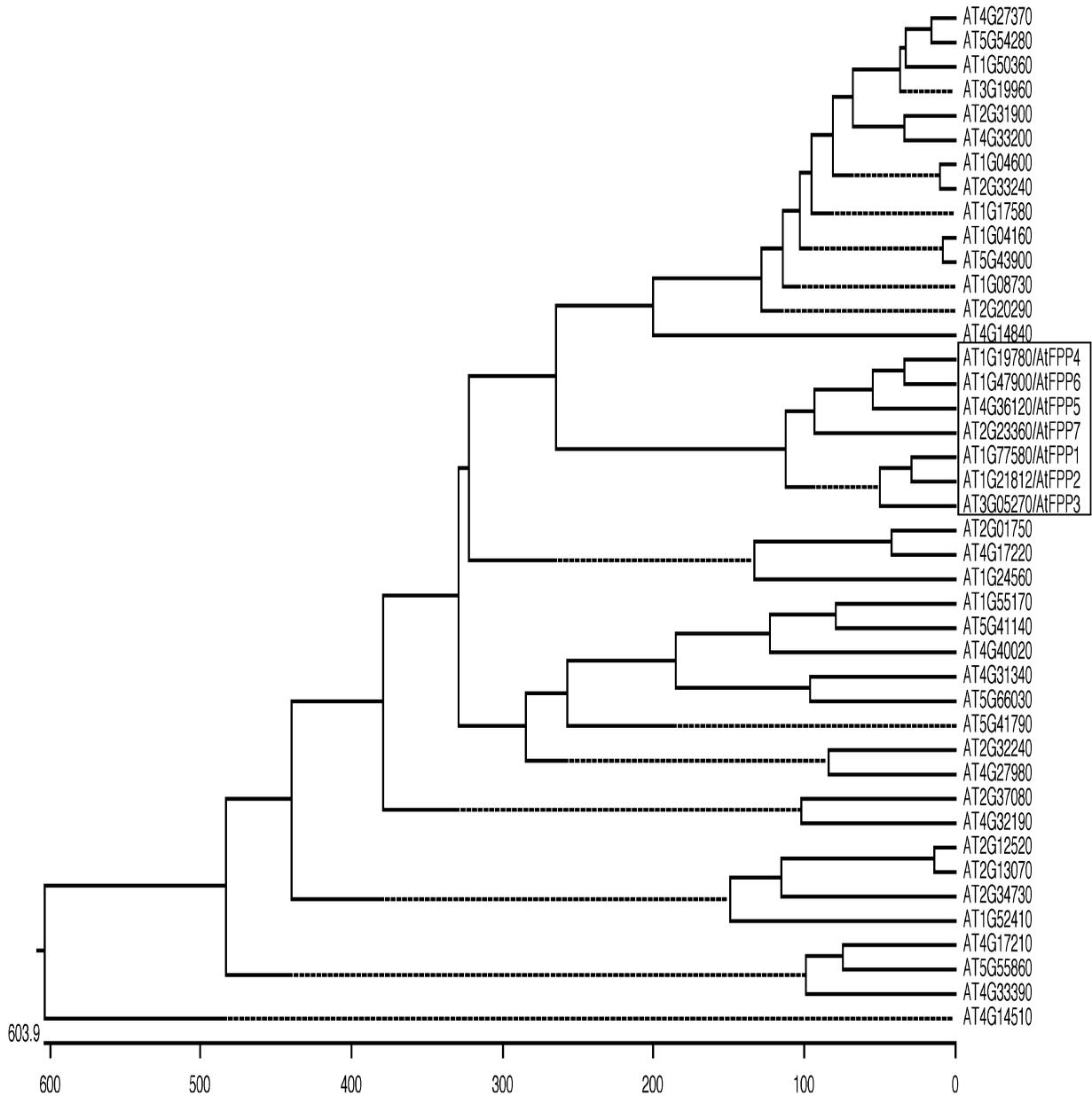


Figure 5

Phylogenetic tree derived from a CLUSTAL analysis of the full ORF sequences of Arabidopsis coiled-coil proteins. The seven AtFPP sequences (boxed) were aligned together with 35 additional Arabidopsis coiled-coil proteins that had been annotated as "myosin-like". The length of each pair of branches measures the distance between sequences. Units indicate the number of computed residue substitution events. Dotted lines have no unit length and were included for alignment purposes. ORFs were identified by their TAIR gene names, accessible through [<http://www.arabidopsis.org>].

Besides the presence of extended coiled-coil domains, the FPP family members have no sequence similarity to the vertebrate lamins. They also do not show significant sequence similarity to other vertebrate nuclear envelope-as-

sociated proteins such as LAP1, LAP2, MAN1, and emerlin (GenBank accession numbers A55649, S55255, Q9WU40, and XP_048410, respectively). None of these proteins appears to have a convincing ortholog in the Ara-

bidopsis genome (data not shown), consistent with the hypothesis that the composition of nuclear envelope-associated proteins differs significantly between the plant and animal kingdom.

Interestingly, a recent report about the protein composition of the nuclear envelope of the highly divergent eukaryote *Trypanosoma brucei* has identified a coiled-coil nuclear envelope protein (TbNUP-1) in this organism that also has no similarity with lamins [40]. TbNUP-1 is a 350 kDa protein with a striking repeat structure. Its full sequence is presently not known but a partial open reading frame of 268 amino acids has been constructed [40]. The near-perfect 144 amino acids repeats contain two coiled-coil domains each. Pairwise alignments of TbNUP-1 with AtFPP1 through AtFPP7 showed no significant sequence similarities beyond the typical low degree of similarity generally observed between coiled-coil proteins (data not shown). The TbNUP-1 open reading frame was also used for a BLAST search of the full Arabidopsis genome. Only comparably weak similarities with other "myosin-like proteins" were found.

The four conserved sequence motifs identified in the FPP family appear to be unique to this group of coiled-coil proteins and are neither shared by other coiled-coil open reading frames present in the Arabidopsis genome nor by coiled-coil proteins such as lamins from other organisms. Their role in the structure or function of the FPP proteins is presently not known, but their high degree of conservation in the tomato and rice sequences imply that they are of relevance. Several functional small conserved motifs have been identified in other families of long coiled-coil proteins. The nuclear lamins contain short, conserved sequences flanking the lamin rod domain, which are phosphorylated by cdc2 kinase [41]. The 90 amino acid PACT domain, which immediately follows the coiled-coil domain of the related proteins pericentrin and AKAP450 and two uncharacterized coiled-coil open reading frames from *Drosophila* and *Schizosaccharomyces pombe*, confers centrosomal localization [19]. A conserved domain of ca. 50 amino acids was found in several Golgi-associated large coiled-coil proteins. It has been named GRIP domain [42] or Golgi localization domain (GLD) [18] and is sufficient to specify Golgi targeting in mammalian cells. Nearly all intermediate filament proteins exhibit a highly conserved amino acid motif (YRKLEGE) at the C-terminal end of their central alpha-helical rod domain which has been shown to be crucial for the formation of authentic tetrameric complexes and for the control of filament width [43].

No sequences with similarity to the consensus sequences of motifs II, III, and IV were found in the databases that might indicate their potential function. However, a se-

quence with some similarity to motif I was identified in the hemagglutinin-esterase fusion glycoprotein from influenza virus C (GenBank accession number S07412). Alignment of this sequence (amino acids 39 to 74) with the central conserved residues of motif I: A--GWEK---E---LK--L-----R--H-D (-- indicating non-conserved residues) shows a good match of identical and conserved amino acid residues: A--GF EK---D---LK--I-----R--H-E. The fusion glycoproteins contain a coiled-coil domain and are involved in fusing the viral envelope with the host cell membrane. The conformational change of the protein necessary for membrane fusion is driven by ionic interactions between residues of the coiled-coil domain, and Glu74, the last amino acid aligning to motif I, is involved in this process [1]. In light of their putative interaction with the nuclear envelope-associated protein MAF1, it will be interesting to investigate if FPPs are associated with membrane systems of the plant cell.

Conclusions

A family of novel long coiled-coil proteins has been identified from plants. They are characterized by two clusters of long coiled-coil domains separated by a non-coiled-coil linker and by four novel sequence motifs. Seven members of the *filament-like plant protein (FPP)* family are present in the Arabidopsis genome, and one homolog each has been identified from tomato and rice. Thus, this family of proteins appears to be conserved among higher plants, but we have not found convincing homologs from yeast or animals by sequence similarity searches. Tomato FPP was originally isolated in a yeast two hybrid screen with MAF1, a small plant-specific protein associated with the nuclear envelope [36]. Because the full Arabidopsis genome does not appear to contain genes for lamins, it is of particular interest to investigate which other long coiled-coil proteins might functionally replace lamins in the plant kingdom. The FPP family represents an exciting group of candidates for such proteins.

Materials and methods

Yeast two-hybrid screen

Total RNA from young tomato leaves (*Lycopersicon esculentum*) was isolated as described by Wanner and Gruissem [44]. mRNA was purified using the mRNA purification kit (Pharmacia Biotech, Piscataway, NJ). To construct the yeast two-hybrid library the cDNA synthesis kit, the Giga-pack III gold packaging extract and the HybriZAP two-hybrid predigested vector kit (all Stratagene, La Jolla, CA) were used according to the protocol of the manufacturer. The size of the primary library was determined to be 1.5×10^6 plaque forming units. The complete cDNA of LeMAF1 (GenBank accession number AF118113) was cloned as an EcoRI/XhoI fragment between the EcoRI and Sall sites of the bait vector pBD-GAL4 (Statagene). Approximately one million primary transformants were screened for growth

on histidine dropout plates. All steps of the library screening including the transformation, the isolation of clones, and the verification of the interactions, and the β -Galactosidase assays were performed as described by the manufacturer. pBD-AtRanGAP1 [33] and pBD-AtMAF1 were both constructed by cloning an EcoRI fragment containing the complete open reading frames into the EcoRI site of pBD-GAL4. The AtMAF1 open reading frame was isolated by PCR from Arabidopsis genomic DNA using the 5' primer 5'-TCC ATG GCC GAA ACC GAA-3' and the 3' primer 5'-CTA AGT TCA CTT CGA ACT GCT C-3', which had been designed to match the Arabidopsis MAF1 homolog on the P1 clone MGG4 of chromosome 5 (GenBank accession number AB008267).

RNA isolation and RNA blotting

Total RNA from tomato leaves (5–15 mm), fruits (3–8 mm in diameter), whole flowers, and light-grown and dark-grown tomato seedlings was isolated with the Trizol Reagent from Gibco BRL (Gaithersburg, MD). Total RNA (20 μ g each) was separated on a formaldehyde gel, blotted to a nitrocellulose membrane, and hybridized with a radioactively labeled fragment corresponding to bp 1140 – 1926 of the LeFPP cDNA, essentially as described by [45].

Database searches, secondary structure prediction, and sequence comparison

Sequences with similarity to LeFPP were identified by conducting BLAST (Basic Local Alignment Search Tool; [46]) searches for similarity to sequences contained in GenBank. GenBank data were accessed through NCBI [http://www.ncbi.nlm.nih.gov/]. Arabidopsis genome data were accessed through the TAIR database [http://www.arabidopsis.org/]. Arabidopsis ESTs were accessed through the TIGR Arabidopsis thaliana database [http://www.tigr.org/tdb/e2k1/ath1/]. Multiple sequence alignments were performed with the DNASTAR protein alignment protocol (DNASTAR, Madison, WI) using the CLUSTAL algorithm. Parameters for CLUSTAL alignments were: Pairwise alignment: Ktuple, 1; gap Penalty, 3; windows, 5; diagonals saved, 5. Multiple alignment: gap penalty, 10; gap length penalty, 10. The PAM250 weight table was used. Percent identities and phylogenetic relationships were derived from the CLUSTAL output using the MEGALIGN software package in DNASTAR. The secondary structures of the proteins, α -helical, and coiled-coil regions were analyzed using DNASTAR PROTEAN software and the coiled-coil regions were confirmed with the MultiCoil algorithm [2]. Coiled-coil proteins from Arabidopsis were identified by searching the lists of complete-genome ORF annotations [ftp://ftp.arabidopsis.org/home/tair/Genes/ORF_annotations] for the keywords "myosin" (40 hits) and "coiled-coil" (0 hits). Arabidopsis ORFs for CLUSTAL analysis were downloaded either using their GenBank ac-

cession numbers or, if not yet available, by searching the TAIR database with the respective gene name.

References

- Burkhard P, Strelkov SV, Stetefeld J: **Coiled coils: a highly versatile protein folding motif.** *Trends Cell Biol* 2001, **11**:82-8
- Wolf E, Kim PS, Berger B: **MultiCoil: a program for predicting two- and three-stranded coiled coils.** *Protein Sci* 1997, **6**:1179-89
- Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS: **Predicting coiled coils by use of pairwise residue correlations.** *Proc Natl Acad Sci U S A* 1995, **92**:8259-63
- Bornberg-Bauer E, Rivals E, Vingron M: **Computational approaches to identify leucine zippers.** *Nucleic Acids Res* 1998, **26**:2740-6
- Wilson KL, Zastrow MS, Lee KK: **Lamins and disease: insights into nuclear infrastructure.** *Cell* 2001, **104**:647-50
- Hutchison CJ, Alvarez-Reyes M, Vaughan OA: **Lamins in disease: why do ubiquitously expressed nuclear envelope proteins give rise to tissue-specific disease phenotypes?** *J Cell Sci* 2001, **114**:9-19
- Cohen C, Parry DA: **A conserved C-terminal assembly region in paramyosin and myosin rods.** *J Struct Biol* 1998, **122**:180-7
- Kozielecki F, Sack S, Marx A, Thormahlen M, Schonbrunn E, Biou V, Thompson A, Mandelkow EM, Mandelkow E: **The crystal structure of dimeric kinesin and implications for microtubule-dependent motility.** *Cell* 1997, **91**:985-94
- Gee MA, Heuser JE, Vallee RB: **An extended microtubule-binding structure within the dynein motor domain.** *Nature* 1997, **390**:636-9
- Bouckson-Castaing V, Moudjou M, Ferguson DJ, Mucklow S, Belkaid Y, G Milon, Crocker PR: **Molecular characterisation of ninein, a new coiled-coil protein of the centrosome.** *J Cell Sci* 1996, **109**:179-90
- Meuwissen RL, Offenbergh HH, Dietrich AJ, Riesewijk A, van Iersel M, Heyting C: **A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes.** *Embo J* 1992, **11**:5091-100
- Zeng C: **NuMA: a nuclear protein involved in mitotic centrosome function.** *Microsc Res Tech* 2000, **49**:467-77
- Sundberg HA, Davis TN: **A mutational analysis identifies three functional regions of the spindle pole component Spc110p in Saccharomyces cerevisiae.** *Mol Biol Cell* 1997, **8**:2575-90
- Yao X, Abrieu A, Zheng Y, Sullivan KF, Cleveland DW: **CENP-E forms a link between attachment of spindle microtubules to kinetochores and the mitotic checkpoint.** *Nat Cell Biol* 2000, **2**:484-91
- Strambio-de-Castilla C, Blobel G, Rout MP: **Proteins connecting the nuclear pore complex with the nuclear interior.** *J Cell Biol* 1999, **144**:839-55
- Munro S: **Localization of proteins to the Golgi apparatus.** *Trends Cell Biol* 1998, **8**:11-5
- Kjer-Nielsen L, van Vliet C, Erlich R, Toh BH, Gleeson PA: **The Golgi-targeting sequence of the peripheral membrane protein p230.** *J Cell Sci* 1999, **112**:1645-54
- Kjer-Nielsen L, Teasdale RD, van Vliet C, Gleeson PA: **A novel Golgi-localisation domain shared by a class of coiled-coil peripheral membrane proteins.** *Curr Biol* 1999, **9**:385-8
- Gillingham AK, Munro S: **The PACT domain, a conserved centrosomal targeting motif in the coiled-coil proteins AKAP450 and pericentrin.** *EMBO Rep* 2000, **1**:524-9
- Gaines P, Tompkins L, Woodard CT, Carlson JR: **quick-to-court, a Drosophila mutant with elevated levels of sexual behavior, is defective in a predicted coiled-coil protein.** *Genetics* 2000, **154**:1627-37
- Guan Z, Prado A, Melzig J, Heisenberg M, Nash HA, Raabe T: **Mushroom body defect, a gene involved in the control of neuroblast proliferation in Drosophila, encodes a coiled-coil protein.** *Proc Natl Acad Sci U S A* 2000, **97**:8122-7
- Nagano A, Arahata K: **Nuclear envelope proteins and associated diseases.** *Curr Opin Neurol* 2000, **13**:533-9
- Mounkes LC, Burke B, Stewart CL: **The a-type lamins. Nuclear structural proteins as a focus for muscular dystrophy and cardiovascular diseases.** *Trends Cardiovasc Med* 2001, **11**:280-5
- Meier I, Phelan T, Gruissem W, Spiker S, Schneider D: **MFPI, a novel plant filament-like protein with affinity for matrix attachment region DNA.** *Plant Cell* 1996, **8**:2105-15

25. Gindullis F, Meier I: **Matrix attachment region binding protein MFPI is localized in discrete domains at the nuclear envelope.** *Plant Cell* 1999, **11**:117-28
26. Masuda K, Haruyama S, Fujino K: **Assembly and disassembly of the peripheral architecture of the plant cell nucleus during mitosis.** *Planta* 1999, **210**:165-7
27. Masuda K, Xu ZJ, Takahashi S, Ito A, Ono M, Nomura K, Inoue M: **Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain.** *Exp Cell Res* 1997, **232**:173-81
28. Reddy AS: **Molecular motors and their functions in plants.** *Int Rev Cytol* 2001, **204**:97-178
29. Reddy AS, Day IS: **Analysis of the myosins encoded in the recently completed Arabidopsis thaliana genome sequence.** *Genome Biol* 2001, **2**:
30. Minguez A, Moreno Diaz de la Espina S: **Immunological characterization of lamins in the nuclear matrix of onion cells.** *J. Cell Sci.* 1993, **106**:431-439
31. McNulty AK, Saunders MJ: **Purification and immunological detection of pea nuclear intermediate filaments: evidence for plant nuclear lamins.** *J. Cell Sci.* 1992, **103**:407-414
32. Beven A, Guan Y, Peart J, Cooper C, Shaw P: **Monoclonal antibodies to plant nuclear matrix reveal intermediate filament-related compounds within the nucleus.** *J. Cell Science* 1991, **98**:293-302
33. Meier I: **A novel link between ran signal transduction and nuclear envelope proteins in plants.** *Plant Physiol* 2000, **124**:1507-10
34. Georgatos SD, Maroulakou I, Blobel G: **Lamin A, lamin B, and lamin B receptor analogues in yeast.** *J Cell Biol* 1989, **108**:2069-2082
35. Mewes HW, Hani J, Pfeiffer F, Frishman D: **MIPS: a database for protein sequences and complete genomes.** *Nucleic Acids Res.* 1998, **26**:33-37
36. Gindullis F, Peffer NJ, Meier I: **MAFI, a novel plant protein interacting with matrix attachment region binding protein MFPI, is located at the nuclear envelope.** *Plant Cell* 1999, **11**:1755-68
37. Rose A, Meier I: **A domain unique to plant RanGAP is responsible for its targeting to the plant nuclear rim.** *Proc. Natl. Acad. Sci. U.S.A.* 2001, **98**:15377-15382
38. Lutcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA: **Selection of AUG initiation codons differs in plants and animals.** *Embo J* 1987, **6**:43-8
39. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268**:78-94
40. Rout MP, Field MC: **Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component.** *J Biol Chem* 2001, **276**:38261-71
41. Stuurman N, Heins S, Aebi U: **Nuclear lamins: their structure, assembly, and interactions.** *J Struct Biol* 1998, **122**:42-66
42. Munro S, Nichols BJ: **The GRIP domain – a novel Golgi-targeting domain found in several coiled-coil proteins.** *Curr Biol* 1999, **9**:377-80
43. Herrmann H, Strelkov SV, Feja B, Rogers KR, Brettel M, Lustig A, Haner M, Parry DA, Steinert PM, Burkhard P, et al: **The intermediate filament protein consensus motif of helix 2B: its atomic structure and contribution to assembly.** *J Mol Biol* 2000, **298**:817-32
44. Wanner LA, Gruissem W: **Expression dynamics of the tomato rbcS gene family during development.** *Plant Cell* 1991, **3**:1289-303
45. Sambrook J, Fritsch EM, Maniatis T: *Cold Spring Harbor Laboratory, Cold Spring Harbor.* 1989
46. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-402

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>



editorial@biomedcentral.com