Research

Are plant formins integral membrane proteins?

Fatima Cvrčková

Address: Department of Plant Physiology, Faculty of Sciences, Charles University, Viničná 5, CZ 128 44 Praha 2, Czech Republic.
E-mail: fatima@natur.cuni.cz

Published: 27 April 2000

Genome ***Biology*** 2000, **1(1)**:research001.1–001.7

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2000/1/1/research/001

Received: 2 November 1999
Revised: 25 January 2000
Accepted: 28 January 2000

Abstract

Background: The formin family of proteins has been implicated in signaling pathways of cellular morphogenesis in both animals and fungi; in the latter case, at least, they participate in communication between the actin cytoskeleton and the cell surface. Nevertheless, they appear to be cytoplasmic or nuclear proteins, and it is not clear whether they communicate with the plasma membrane, and if so, how. Because nothing is known about formin function in plants, I performed a systematic search for putative *Arabidopsis thaliana* formin homologs.

Results: I found eight putative formin-coding genes in the publicly available part of the *Arabidopsis* genome sequence and analyzed their predicted protein sequences. Surprisingly, some of them lack parts of the conserved formin-homology 2 (FH2) domain and the majority of them seem to have signal sequences and putative transmembrane segments that are not found in yeast or animal formins.

Conclusions: Plant formins define a distinct subfamily. The presence in most *Arabidopsis* formins of sequence motifs typical of transmembrane proteins suggests a mechanism of membrane attachment that may be specific to plant formins, and indicates an unexpected evolutionary flexibility of the conserved formin domain.

Background

Some mechanisms involved in cell morphogenesis, such as membrane vesicle transport, are conserved at least among crown eukaryotes (metazoa, fungi and plants) [1,2], whereas others, such as those involving extracellular structures or the precise roles of different Rho-like GTPases [3], are not. Yet other cellular processes, such as cytokinesis, often recruit conserved proteins to accomplish superficially dissimilar tasks (for example, budding, cleavage or phragmoplast-based cell division of plant cells) [4]. For many morphogenetic mechanisms, the question of evolutionary conservation remains unresolved because available information is limited to one or a few model organisms. For example, this is the case for the molecular mechanisms that ensure the communication between the cytoskeleton and the surface of the cell. However, the recent increase in the data available from a number of genome projects allows wide-ranging searches for homologs of known components of signaling and morphogenetic pathways. The results of such searches can lead both to experimentally testable hypotheses and to general conclusions regarding the evolution of morphogenetic processes.

Formins, also known as formin homology (FH) proteins, are proteins implicated in cellular and organismal morphogenesis of both metazoa and fungi. On the cellular level, they are involved in the establishment and maintenance of cell and/or tissue polarity [5,6], in cytokinesis [4] and in the positioning of the mitotic spindle [7]. They interact directly or indirectly with actin, profilin, Rho-like GTPases [5,6,8,9–11], the yeast Spa2 protein and septins [12,13], proteins containing SH3 or WW domains [10,14], dynein and microtubules [7,15–17]. The yeast formin homolog encoded by *BNI1* is localized to the cell periphery and participates in positioning cortical actin patches towards distinct regions of the plasma membrane

[5,13,18]. Some kind of contact with the plasmalemma (in addition to that mediated by a Rho-like GTPase) might therefore be expected, although there is no evidence as yet for such a contact. Furthermore, metazoan formins are believed to be cytoplasmic or nuclear proteins [19,20].

Nothing is known about formin function in plants, although the existence of two *Arabidopsis thaliana* proteins containing the conserved formin-homology 2 (FH2) domain has been reported recently [6,10]. Given that all known formins represent a well-defined family, this class of proteins may be a good candidate for a systematic genome sequence search. Here, I present the results of such an approach, which has led to the identification of putative plant formin genes, as well as to the finding that the evolutionarily old formin domain may be used in a number of different ways and contexts ('modules' as defined by Hartwell *et al.* [21]) by recent eukaryotes.

Results and discussion

Formins are defined by the presence of two sequence domains — the low-complexity, proline-rich FH1 and the carboxy-terminal FH2 [6,10,22]. A third domain — the amino-terminal FH3 motif — has been characterized biochemically but is rather poorly delimited in sequence terms [23]. Despite a conflicting consensus definition, this motif appears to be identical to the amino-terminal conserved block found in some formins by Wasserman [10]. I have used the L-x-x-G-N-x-M-N (single-letter amino-acid notation; x is any amino acid) motif present in the FH2 domain of most fungal and metazoan formins [10] to search for putative *Arabidopsis* formin homologs and found eight such inter-related genes (see Materials and methods and Table 1). All of them correspond either to hypothetical open reading frames (ORFs) or to unannotated genomic or cDNA clones, indicating that at least some of them are expressed *in vivo*. These putative genes and their predicted protein products will be referred to henceforth as AtFORMINs 1 to 8.

Sequence comparison with known formins revealed the presence of a genuine FH2 domain in all *Arabidopsis* formins (Figure 1). However, even the longest predicted proteins, encoded by the AtFORMIN3, -4 and -5 genes, lack parts of the FH2 region ubiquitously conserved among corresponding genes of fungi and metazoa (Figures 1 and 2), although not necessarily among their protein products, because some formin mRNAs undergo complex splicing [24]. Sequence motifs corresponding to the missing regions were found in all cases within the predicted introns by visual inspection of three-frame translation data. Because the reliability of mRNA structure prediction is limited [25], failure to identify exons correctly may explain the apparent deletion of this region of the FH2 domain. The possibly mispredicted intron encoding subdomain g of AtFORMIN4 is split by a frameshift mutation, however. Although this could reflect a sequencing error, the possibility remains that plant formin homologs have a modular structure within the FH2 domain at the gene level, and that at least some of the FH2-related sequences within predicted introns are vestiges of exons lost by mutation.

Proline-rich regions corresponding to FH1 were identified in all *Arabidopsis* formins. Surprisingly, there are two such regions in AtFORMINs 2, 6 and 8 — a feature not observed in the non-plant formins examined (listed in Materials and methods). Neither motifs corresponding to FH3 nor coiled-coil regions flanking FH1 (common but not ubiquitous in non-plant formins [10]) were found. The structure of FH2, the overall protein size (smaller than most non-plant formins) and the domain layout of *Arabidopsis* formins therefore show possible plant-specific features (Figure 2). This idea is supported by the topology of an evolutionary tree that consistently places *Arabidopsis* formins in a branch separate from other members of the formin family (Figure 3).

As in the non-plant formins, the amino-terminal portions of all *Arabidopsis* formins are divergent, although there is 63% identity between AtFORMINs 1 and 4 in the overlaping parts of their sequences. Analysis of AtFORMIN sequences with SMART [26,27] revealed no previously characterized domains outside the FH2 region. However, putative amino-terminal membrane insertion signals (signal peptides) followed by a segment highly likely to be membrane-spanning and a variable number of possible transmembrane domains were found in AtFORMINs 1, 2, 4, 6 and 8. A possible membrane insertion signal was also identified in AtFORMIN5 by one of the two methods used (see Materials and methods, and Figures 2,4). The length of predicted signal peptides suggests that they may represent membrane anchors rather than secretion signals [28]. A putative transmembrane segment was also found in the apparently amino-terminally truncated sequence of AtFORMIN3. In contrast, no signal peptides were found in 12 fungal and animal formins listed in Materials and methods, although transmembrane-like segments were observed in some. Surprisingly, the putative transmembrane segment lies between the two Pro-rich regions in AtFORMINs 2, 6 and 8. Obviously, only the cytoplasmic one of these two motifs can act as a conventional FH1 domain. Its size ranges from 106 to 423 amino acids, with proline content of 13 to 41% and multiple stretches of five to nine consecutive proline residues. This structure roughly corresponds to that of previously characterized FH1 domains [10]. Interestingly, the FH1 domains of AtFORMINs 2, 7 and 8 are extremely rich in serine (up to 20%) and contain stretches of up to seven consecutive serine residues.

The other proline-rich domain of AtFORMINs 2, 6 and 8 is predicted to be exposed to a non-cytoplasmic compartment. Given that polyproline stretches are characteristic for a class of structural cell-wall proteins known as extensins [29], it is tempting to speculate about a possible role for this domain in communication between formins and structures within the cell wall. Apart from this, few predictions of function can

Table 1

Putative formin-related genes of *Arabidopsis thaliana*

Gene	Primary accession	Protein sequence accession	ORF location*	ORF size†	Type	Chr‡	Number of introns	Notes
AtFORMIN1	gb|AC002062 (emb|T43335)	gb|AAB61101	47 640...48 637; 48 716...50 000	760	Genomic (EST)	I	1	AtORF2 in [6]
AtFORMIN2	gb|AC002333 (gb|AI997606)	gb|AAB64026	28 161...26 738; 26 653...26 466; 26 314...26 061; 25 979...25 161 (R)	894	Genomic (EST)	II	3	AtORF1 in [6]
AtFORMIN3	emb|Z97338	gb|CAB10299	30 407... 30 285; 30 171... 29 688; 29 608... 29 146; 29 075... 28 870; 28 800... 28 683; 28 602... 28 566; 28 485... 28 147 (R)	589	Genomic	IV	6	Sequencing error at the 5′ end leading to ORF truncation?
AtFORMIN4	gb|AC002396 (gb|AI998115)	gb|AAC00575	28 830... 29 848; 29 951... 30 296; 30 542... 31 218; 31 885... 32 320	825	Genomic (EST)	I	3	ORF extends 15 base pairs upstream of the reported ATG; alternative splicing possible
AtFORMIN5	dbj|AB016879		67 574... 67 401; 66 710... 66 520; 66 171... 66 092; 66 004... 65 389; 65 298... 65 099; 64 637... 63 784 (R)	705	Genomic	V	5	Alternative splicing possible
AtFORMIN6	dbj|AB013390 (emb|F19772)		6 001... 7 470; 7 550... 7 757; 8 244... 8 506; 8 587... 9 378	910	Genomic (EST)	V	3	
AtFORMIN7	gb|AC007258	gb|AAD39332	121 331... 120 011; 122 896... 121 428; (R)	929	Genomic	I	1	
AtFORMIN8	dbj|AB025639 (emb|Z18512)		41 595... 39 722; 39 635... 39 430; 39 248... 39 004; 38 919... 38 092 (R)	1051	Genomic (EST)	III	3	

*ORF coordinates refer to the longest putative ORF in the first sequence (first primary accession) listed. (R), reverse complement. †The ORF size is given in codons. ‡Chr, chromosome number.

be made on the basis of the sequence data. Although formins are well conserved with respect to their molecular structure, we do not know the extent of their conservation within signaling or structural modules [21]. As the relationships between protein structure, module structure and biological function are far from straightforward [30], we can at present neither prove nor exclude the possibility that plant formins contribute to similar functional modules to their animal and fungal counterparts. The question of whether these proteins have a direct role in cytokinesis, in mitotic spindle localization, or in some other cellular process, possibly involving cytoskeleton rearrangement or cell-surface growth, will have to be answered experimentally.

Conclusions

A systematic search of the available *Arabidopsis* genomic and cDNA sequences revealed the presence of eight genes encoding proteins that define a novel subfamily of the formin family. At least six out of eight *Arabidopsis* formins appear to be integral membrane proteins. This indicates a mechanism of membrane localization that may be specific to

Figure 1
Alignment of the FH2 domain of selected formins and definition of the subdomain modules. Subdomain modules (a–j) are marked in color. Red dots denote the position of introns (not shown in MFORMIN, for which only mRNA sequence is available). The consensus line shows 80% consensus of the EMBL DS39866 alignment. Numbers indicate positions within the sequence and the size of unaligned insertions; residues corresponding to unambiguous consensus and/or shared by all *Arabidopsis* formins are highlighted. For gene terminology see Table 1 and Materials and methods.

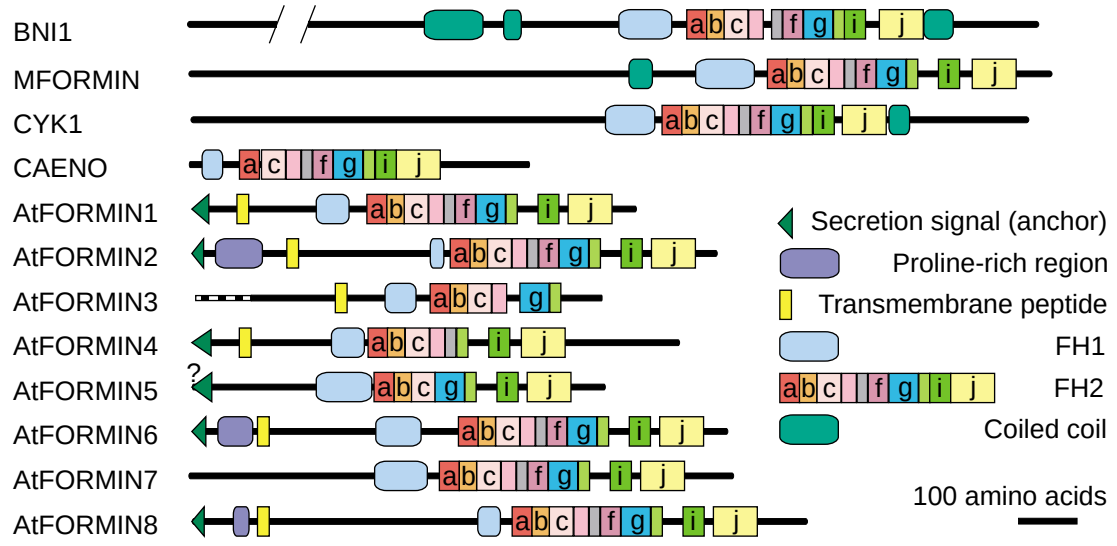

Figure 2
Domain structure of *Arabidopsis* and selected yeast and animal formins. Letters denote subdomain modules within FH2 as defined in Figure 1. Only the 'highly likely' membrane-spanning segments are shown.

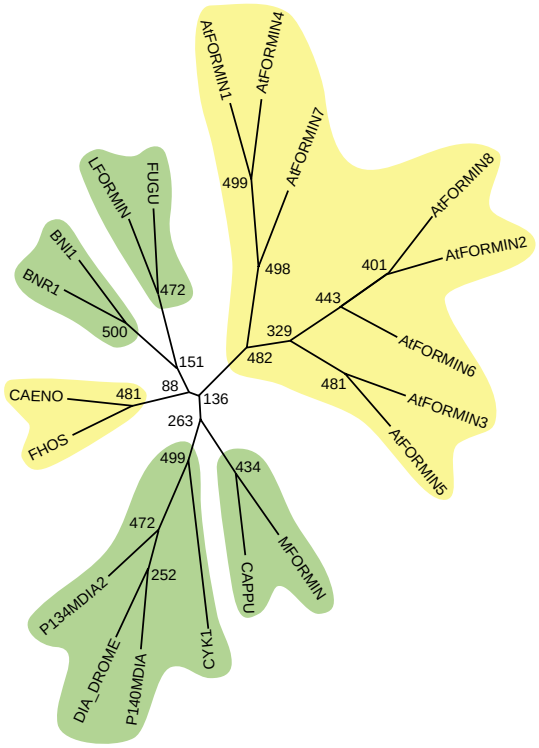

Figure 3
Unrooted evolutionary tree of FH2 subdomains a, c and h constructed by the neighbor-joining method. Numbers at nodes indicate bootstrap values. Branches in agreement with the tree previously reported by Zeller *et al.* [6] are highlighted in green, novel branches in yellow.

plants and functionally related to a possible role for formins in the communication between the plant cell and extracellular structures.

Materials and methods

Identification of Arabidopsis formin homologs and protein sequence prediction

The initial search for formin homologues in the non-redundant *Arabidopsis thaliana* protein (NRAT) database, performed using the PatMatch program [31,32] with the query pattern L-x-x-G-N-x-M-N, yielded three potential formin homologs — AtFORMIN1 to AtFORMIN3. AtFORMINs 2 to 8 were found by a TBLASTN 2.0 search [33,34] in GenBank, using the predicted protein sequence of AtFORMIN 1 as query (P(N) values in the range of 5.8×10^{-227} to 1.3×10^{-11}). Known members of the formin family (a human *diaphanous* homolog and *Drosophila melanogaster cappucino*) were found in the same search (P(N) values 1×10^{-21} and 1.3×10^{-13}, respectively), verifying the statistical significance of the initial PatMatch results.

Membrane insertion signals (anchors)

```
AtFORMIN1     1 MAAMFNHPWP----NLTLIYFFFIVVLPFQSLS 29
AtFORMIN2     1              MTTIPFCFLFVAFFFSSSTA 20
AtFORMIN4     1 MAAMLMQPWPPFLPHLTLVFLTLILFFPNQSFS 33
AtFORMIN6     1           MKALQSRFFFFFFYIFFSVSVSS 24
AtFORMIN8     1              MLFFLFFFYLLLSSSSDLVFA 21
```

Transmembrane peptides

```
AtFORMIN1    78 AVLITAASTLLVAGVFFFCLQ  98
AtFORMIN2   157 TASVISAAALLSLFAVFIIFI 177
AtFORMIN3   157 AVASTAVLTFVFVALMFLCCF 177
AtFORMIN4    80 AVLITAASTLLVAAVFFFLVH 100
AtFORMIN6   109 IVISVGIVTLGMLSALAFFLY 129
AtFORMIN8   108 LLIVAISAVSSAALVALLIAL 128
```

Figure 4
Putative membrane anchors and transmembrane domains of *Arabidopsis* formins. Aliphatic (I, L, V), aromatic (F, H, W, Y) and other potentially hydrophobic (A, C, G, K, M, R, T) amino acids are highlighted.

Intron positions in the genomic sequences were determined (or confirmed) using the NetGene2 server [25]. Translation of the DNA sequences was performed on the SIB ExPASy WWW server [35,36]. Only the longest predicted ORFs were subjected to further analysis.

Sequence alignment and domain structure analysis

All sequence comparisons were done on a set of 20 metazoan, yeast and plant formin sequences. These were FUGU, *Fugu rubripes* formin homolog gb|AAC34395.1; LFORMIN, mouse lymphocyte-specific formin gb|AAD01273; BNR1, yeast Bnr1 protein sp|P40450; BNI1, yeast Bni1 protein sp|P41832; FHOS, human formin-like protein gb|AAD39906.1; CAENO, *Caenorhabditis elegans* formin homolog gb|AAB42354.1; CAPPU, *D. melanogaster* Cappuccino gb|AAC46925.1; P140MDIA and P134MDIA2, mouse Diaphanous homologs gb|AAC53280 and gb|AAC71771.1; DIA_DROME, *D. melanogaster* Diaphanous sp|P48608; CYK1, *C. elegans* Cyk1 assembled from gb|AAA81161.1 and gb|AAC17501.1; MFORMIN, mouse formin sp|Q05860; and AtFORMIN 1 to 8. Protein sequences were aligned with the aid of MACAW [37], using the Gibbs sampler and segment pair algorithms, BLOSUM45 matrix. Only blocks with P $<10^{-7}$ were considered. No homology to FH3 as defined by Petersen *et al.* [23] or to the amino-terminal conserved region [10] was revealed by this tool, whereas the FH2 domain was readily identified. Non-aligned parts of the sequence within the FH2 domain were adjusted manually. Consensus of the resulting alignment of FH2 (deposited in the EMBL alignment database, accession number DS39866) has been calculated for each subdomain separately (see Figure 1) by the method of Brown and Lai [38,39].

The SMART program [26,27] was used to examine predicted protein sequences for the presence and location of known

sequence domains, putative secretion signals, transmembrane segments, coiled-coil motifs and low sequence complexity regions (usually representing proline-rich FH1 domains whose location was confirmed by visual inspection). Prediction of signal peptides by the neural network (NN) method [28]) was independently verified by a hidden Markov model-based (HMM) method on the SignalP 2.0 server [40,41]). Results of both methods were in agreement, with the exception of AtFORMIN5, which was predicted to be membrane-anchored by NN but cytoplasmic by HMM.

Construction of the evolutionary tree

The tree (Figure 3) was calculated from the three FH2 subdomains present in all formins studied, using programs from the PHYLIP package [42,43] version 3.573. An input file was prepared by joining subdomains a, c and h and was used to produce a bootstrapped data set by SEQBOOT with 500 sampling cycles. Distances were calculated using PROTDIST with the PAM distance matrix, and the results were used for tree construction using the neighbor-joining method [44] by NEIGHBOR. The consensus tree was determined by CONSENSE and plotted using DRAWTREE.

Acknowledgements

This work has been supported by the Grant Agency of the Czech Republic Grant 204/98/0482 and by the Czech Ministry of Education Program J13/98:113100003. I thank J. Flegr for helpful discussion.

References

1. Žárský V, Cvrčková F: **Small GTPases in the morphogenesis of yeast and plant cells.** In *Molecular Mechanisms of Signalling and Membrane Transport*. Edited by Wirtz KWA. Berlin: Springer, 1997, 75-88.
2. Sanderfoot AA, Raikhel N: **The specificity of vesicle trafficking: coat proteins and SNAREs.** *Plant Cell* 1999, **11:**629-642.
3. Li H, Wu G, Ware D, Davis KR, Yang Z: ***Arabidopsis*** **Rho-related GTPases: differential gene expression in pollen and polar localization in fission yeast.** *Plant Physiol* 1998, **118:**407-417.
4. Field C, Li R, Oegema K: **Cytokinesis in eukaryotes: a mechanistic comparison.** *Curr Opin Cell Biol* 1999, **11:**68-90.
5. Evangelista M, Blundell K, Longtine MS, Chow CJ, Adames N, Pringle JR, Peter M, Boone C: **Bni1p, a yeast formin linking Cdc42p and the actin cytoskeleton during polarized morphogenesis.** *Science* 1997, **276:**118-122.
6. Zeller R, Haramis AG, Zuniga A, McGuigan C, Dono R, Davidson G, Chabanis S, Gibson T: **Formin defines a large family of morphoregulatory genes and functions in establishment of the polarising region.** *Cell Tissue Res* 1999, **296:**85-93.
7. Heil-Chapdelaine R, Adames N, Cooper JA: **Formin' the connection between microtubules and the cell cortex.** *J Cell Biol.* 1999, **144:**809-811.
8. Frazier J, Field C: **Actin cytoskeleton: are FH proteins local organizers?** *Curr Biol* 1997, **7:**R414-R417.
9. Fujiwara T, Tanaka K, Mino A, Kikyo M, Takahashi K, Shimizu K, Takai Y: **Rho1p-Bni1p-Spa2p interactions: implication in localization of Bni1p at the bud site and regulation of the actin cytoskeleton in *Saccharomyces cerevisiae*.** *Mol Biol Cell* 1998, **9:**1221-1233.
10. Wasserman S: **FH proteins as cytoskeletal organizers.** *Trends Cell Biol* 1998, **8:**111-115.
11. Johnson DI: **Cdc42: an essential Rho-type GTPase controlling eukaryotic cell polarity.** *Microbiol Mol Biol Rev* 1999, **63:**54-105.
12. Zahner JE, Harkins HA, Pringle JR: **Genetic analysis of the bipolar pattern of bud site selection in the yeast Saccharomyces cerevisiae.** *Mol Cell Biol.* 1996, **16:**1857-1870.
13. Mino A, Tanaka K, Kamei T, Umikawa M, Fujiwara T, Takai Y: **Shs1p: a novel member of septin that interacts with Spa2p, involved in polarized growth in Saccharomyces cerevisiae.** *Biochem Biophys Res Commun* 1998, **251:**732-736.
14. Kamei T, Tanaka K, Hihara T, Umikawa M, Imamura H, Kikyo M, Ozaki K, Takai Y: **Interaction of Bnr1p with a novel Src homology 3 domain-containing Hof1p. Implication in cytokinesis in *Saccharomyces cerevisiae*.** *J Biol Chem* 1998, **273:**28341-28345.
15. Miller RK, Matheos D, Rose MD: **The cortical localization of the microtubule orientation protein, Kar9p, is dependent upon actin and proteins required for polarization.** *J Cell Biol.* 1999, **144:**963-965.
16. Lee L, Klee SK, Evangelista M, Boone C, Pellman D: **Control of mitotic spindle position by the *Saccharomyces cerevisiae* formin Bni1p.** *J Cell Biol* 1999, **144:**947-961.
17. Chang F: **Movement of a cytokinesis factor cdc12p to the site of cell division.** *Curr Biol* 1999, **9:**849-852.
18. Schmidt A, Hall MN: **Signaling to the actin cytoskeleton.** *Annu Rev Cell Dev Biol* 1998, **14:**305-338.
19. Trumpp A, Blundell PA, de la Pompa JL, Zeller R: **The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis.** *Genes Dev* 1992, **6:**14-28.
20. de la Pompa JL, James D, Zeller R: **The limb deformity proteins during avian neurulation and sense organ development.** *Dev Dyn* 1995, **204:**156-167.
21. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402 Supp:**C47-C52.
22. Castrillon D, Wasserman S: **Diaphanous is required for cytokinesis in *Drosophila* and shares domains of similarity with the limb deformity gene.** *Development* 1994, **20:**3367-3377.
23. Petersen J, Nielsen O, Egel R, Hagan IM: **FH3, a domain found in formins, targets the fission yeast formin FUS1 to the projection tip during conjugation.** *J Cell Biol.* 1998, **141:**1217-1228.
24. Wang CC, Chan DC, Leder P: **The mouse formin (Fmn) gene: genomic structure, novel exons, and genetic mapping.** *Genomics* 1997, **39:**303-311.
25. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: **Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information.** *Nucleic Acids Res* 1996, **24:**3439-3452.
26. Schultz J, Milpetz F, Bork P, Ponting C: **SMART, a simple modular architecture research tool: Identification of signalling domains.** *Proc Natl Acad Sci U S A* 1998, **95:**5857-5864.
27. SMART – simple modular architecture research tool [http://smart.embl-heidelberg.de/].
28. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Prot Engng* 1997, **10:**1-6.
29. Keller B: **Structural cell wall proteins.** *Plant Physiol* 1993, **101:**1127-1130.
30. Oota S, Saitou N: **Phylogenetic relationship of muscle tissue deduced from superimposition of gene trees.** *Mol Biol Evol* 1999, **16:**856-867.
31. Weng S: **PatMatch, Pattern matching software for *Saccharomyces* genome database and *Arabidopsis thaliana* database.** 1998.
32. PatMatch [http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl].
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
34. Gish W, States DJ: **Identification of protein coding regions by database similarity search.** *Nat Genet* 1993, **3:**266-272.
35. Appel RD, Bairoch A, Hochstrasser DF: **A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server.** *Trends Biochem Sci* 1994, **19:**258-260.
36. ExPASy Molecular Biology Server [http://www.expasy.ch/].
37. Schuler GD, Altschul SF, Lipman DJ: **A workbench for multiple alignment construction analysis.** *Prot Struct Funct Genet* 1991, **9:**180-190.
38. Ponting C, Aravind L: **START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins.** *Trends Biochem Sci* 1999, **24:**130-132.
39. Consensus server [http://www.bork.embl-heidelberg.de/Alignment/consensus.html].

40. Nielsen H, Krogh A: **Prediction of signal peptides and signal anchors by a hidden Markov model.** In *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6)*. Menlo Park, California: AAAI Press, 1998, 122-130.
41. Signal P v2.0b2 World Wide Web Prediction Server [http://www.cbs.dtu.dk/services/SignalP-2.0/].
42. Felsenstein J: **PHYLIP — Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5:**164-166.
43. PHYLIP [http://evolution.genetics.washington.edu/phylip.html].
44. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4:**406-425.