

Phylogenetic Analysis of the Formin Homology 2 Domain[□]

Henry N. Higgs^{*†} and Kevin J. Peterson[‡]

^{*}Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03755; and [‡]Department of Biology, Dartmouth College, Hanover, NH 03755

Submitted July 7, 2004; Accepted October 12, 2004
Monitoring Editor: Thomas Pollard

Formin proteins are key regulators of eukaryotic actin filament assembly and elongation, and many species possess multiple formin isoforms. A nomenclature system based on fundamental features would be desirable, to aid the rapid identification and characterization of novel formins. In this article, we attempt to systematize the formin family by performing phylogenetic analyses of the formin homology 2 (FH2) domain, an independently folding region common to all formins, which alone can influence actin dynamics. Through database searches, we identify 101 FH2 domains from 26 eukaryotic species, including 15 in mice. Sequence alignments reveal a highly conserved yeast-specific insert in the “knob loop” region of the FH2 domain, with unknown functional consequences. Phylogenetic analysis using minimum evolution (ME), maximum parsimony (MP), and maximum likelihood (ML) algorithms strongly supports the existence of seven metazoan groups. Yeast FH2 domains segregate from all other eukaryotes, including metazoans, other fungi, plants, and protists. Sequence comparisons of non-FH2 regions support relationships between three metazoan groups (Dia, DAAM, and FRL) and examine previously identified coiled-coil and Diaphanous auto-regulatory domain sequences. This analysis allows for a formin nomenclature system based on sequence relationships, as well as suggesting strategies for the determination of biochemical and cellular activities of these proteins.

INTRODUCTION

The formin protein family regulates actin filament assembly and growth (Waller and Alberts, 2003; Zigmond, 2004). These proteins are present in all eukaryotes examined, with many species possessing multiple isoforms. Biochemical studies of bacterially expressed fragments from four formins, two from yeast and two from mammals, suggest that they serve a general role in the acceleration filament assembly. The details of this polymerization-accelerating activity differ between formins, with two formins nucleating filaments *de novo* (Pruyne *et al.*, 2002; Sagot *et al.*, 2002; Li and Higgs, 2003); one formin requiring an additional protein, profilin, for effective nucleation (Kovar *et al.*, 2003); and a fourth formin seeming to use filament severing as its chief mechanism (Harris *et al.*, 2004).

Several general points can be made about the current state of formin research. First, very few formins have been characterized biochemically or cellularly, and thus little can be said concerning generalizations about formin function. Second, given the different experimental approaches taken by different laboratories, accurate comparison of the biochemical and cellular data for any two formins is difficult at this stage. Third, the biochemical functions studied to date might represent a subset of actual activities and might be strongly influenced by as-yet unidentified binding proteins.

These facts cause difficulties when additional formins are identified or studied. Formin names and comparisons with other formins are being made based on the best information available at the time, which is generally minimal. This situation is certainly not unique to formins, with one good example being the enormous myosin superfamily. For myosins, order has been established in the field by broadly accepted phylogenetic analyses of the motor domain, shared by all myosins (Hodge and Cope, 2000; Sellers, 2000; Berg *et al.*, 2001). Significantly, the phylogenetic relationships between myosin motor domains parallel the relationships between the highly divergent nonmotor domains (Berg *et al.*, 2001), demonstrating that the motor domain phylogeny is robust.

In this article, we mimic these studies of myosin by conducting phylogenetic analyses of the formin homology 2 (FH2) domain. The FH2 domain, ~400 amino acids in length, defines the formin protein family and is sufficient for many of the effects of formins on actin (reviewed in Waller and Alberts, 2003; Zigmond, 2004). As with myosins, the highly diverse nature of the sequences outside of the FH2 domains makes phylogenetic analysis of complete formin sequences difficult. In addition, the position of the FH2 domain in relation to the N and C termini can vary greatly, forcing alignment of sequence blocks instead of full sequences. Other domains have been defined in formins, including FH1, FH3, GTPase binding domain (GBD) and Diaphanous auto-regulatory domain (DAD). However, each of these domains presents problems for phylogenetic analysis. FH1 domains are proline rich, but possess no clear consensus sequence. An FH3 domain has been proposed for the fission yeast formin Fus1 (Petersen *et al.*, 1998), but the existence of this domain in other formins is unclear. Many formins possess at least one GBD, but not all do. Finally, the DAD is a short (~20 residue) sequence with a core consen-

Article published online ahead of print. Mol. Biol. Cell 10.1091/mbc.E04-07-0565. Article and publication date are available at www.molbiolcell.org/cgi/doi/10.1091/mbc.E04-07-0565.

[□] The online version of this article contains supplemental material at MBC Online (<http://www.molbiolcell.org>).

[†] Corresponding author. E-mail address: henry.n.higgs@dartmouth.edu.

sus motif of MDXLLXL (Alberts, 2001). The short length of this motif makes it difficult to identify with confidence and prevents meaningful phylogenetic analysis.

Our work is divided into four sections. First, we describe the process by which we identified FH2 domains from sequence databases. Second, we discuss the sequence features of FH2 domains themselves and the insights we have obtained in aligning them. This alignment significantly extends previous FH2 domain alignments (Zeller *et al.*, 1999; Bate-man *et al.*, 2002), enhancing the prediction of residues important to all formins as well as sequences unique to subsets of formins. Such predictive ability has been useful already in determining biochemical relevance of specific positions (Xu *et al.*, 2004), and alignments covering more diverse groups of eukaryotes increase this potential. Third, we discuss the phylogenetic analysis derived from these alignments. Grouping FH2 domains by phylogeny is both complementary to and independent from biochemical analysis and reveals some striking findings that can guide subsequent experimentation. Fourth, we conduct targeted similarity searches of non-FH2 regions to test the phylogeny-based groupings and probe for additional relationships.

MATERIALS AND METHODS

Searching for FH2 Domain Sequences

Searches for human, mouse, chicken, *Drosophila melanogaster*, *Caenorhabditis elegans*, most protists (eukaryotes that are not metazoans, not plants, and not fungi), and most Ascomycota (yeast) were conducted primarily using National Center for Biotechnology Information BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>), either in protein-protein BLAST (blastp), protein query versus translated database (tblastn), or tblastn of individual genomes. Searches for *Takifugu rubripes* (puffer fish), *Ciona intestinalis* (sea squirt), *Chlamydomonas reinhardtii*, *Thalassiosira pseudonana* (a diatom), and *Phanerochaete chrysosporium* (a Basidiomycete) were conducted using the Department of Energy Joint Genome Institute Genome Portal (<http://www.jgi.doe.gov/genomes/index/html>). For *Ashbya gossypii* (a yeast), <http://data.cgt.duke.edu/ashbya/blast.html> was used. For *Neurospora crassa*, <http://www.broad.mit.edu/annotation/fungi/neurospora/> was used. For *Arabidopsis*, most sequences were obtained from Cvrckova *et al.* (2004), with additional sequences found by National Center for Biotechnology Information BLAST. Initial searches were conducted using budding yeast Bni1p FH2 (1348-1766) as the query sequence. As metazoan groups were defined, additional searches were conducted of metazoan species for specific groups. Individual protist species also were searched for formins from specific metazoan groups.

Aligning FH2 Domains

Initial alignment was conducted using the ClustalW program in MacVector (www.accelrys.com). Subsequently, sequences were edited to align residues known to be functionally relevant from structural and biochemical studies (Shimada *et al.*, 2004; Xu *et al.*, 2004) and to minimize gaps. After editing, all sequence N-terminal to that aligning with residue 1348 of Bni1p, and C-terminal to 1760, was removed.

Phylogenetic Analysis of FH2 Domain Alignment

Eight hundred fifty four amino acid positions from 101 sequences were analyzed with PAUP version 4.0b10 for Macintosh (Swofford, 2002). Distance analysis used minimum evolution (ME) as the optimality criterion and mean character difference as the distance measure. Bootstrap analysis for both maximum parsimony (MP) and ME used 1000 replicates by using heuristic search with tree-bisection-reconnection and random addition sequence with 10 replications per bootstrap. Gaps are treated as missing. Maximum likelihood analysis of 17 sequences (two species from each metazoan group, underlined in Figure 2; Sc Bni1p; and one member of *Arabidopsis* group 1 [At1a] and group 2 ([At1]) used Quartet Puzzle version 5.0 (JTT model of amino acid substitution; 1000 puzzling steps). Bootstrap values from ML are boxed in Figure 2 for each metazoan group.

Comparing non-FH2 Domain Sequences

To compare non-FH2 domain sequences within metazoan groups, we aligned sequences by using ClustalW within the MacVector program. We then evaluated the sequences for regions of similarity by the following criteria. First, the region must be ≥ 26 residues in length, which is of sufficient length to minimize the likelihood of false positives. Second, $\geq 50\%$ of the positions in this region must be identical in n-1 sequences. For example, the Dia group

alignment contained six sequences, so each position counted as positive needed to be identical for five of the sequences. Third, there must be no gaps for any individual sequence in this region. For this analysis, we included only sequences that we judged to be full length or close to full length. Thus, a number of sequences suitable for FH2 domain comparison were excluded from the present analysis. The sequences included in this analysis were (refer to Table 1 for nomenclature) as follows: Dia group: Mm mDia1, Mm mDia2, Mm mDia3, Gg1, Ci1, and Dm Dia; FRL group: Mm FRL1, Mm FRL2, Mm FRL3, Ci4, Tr1, Tr6, Dm2, and Ce3; DAAM group: Mm DAAM1, Mm DAAM2, Tr5, Dm1, and Ce5; INF group: Mm INF1, Mm INF2, Tr20, Ci6, Dm7, and Ce2; delphilin group: Mm delphilin, Tr15, Tr16, Dr, and Ci2; FHOD group: Mm FHOD1, Mm FHOD2, Tr21, Tr22, Dm6, and Ce1; and FMN group: Mm FMN1, Mm FMN1 IV, Mm FMN2, Gg1, and Dm cappu. Although clearly in the Dia group, Ce6 was not included in analysis because it possessed many short insertions (<5 residues) that caused several regions that were clearly similar to be judged as not similar by our criteria. To compare non-FH2 domain sequences between metazoan groups, we aligned all of the sequences included in analysis of individual groups for pairs of groups. When judging similarity between metazoan groups, we allowed gaps to be inserted. Regions were judged to be similar between groups if $>20\%$ of the positions were identical in n-2 of the sequences. For example, the alignment of the Dia and DAAM groups contained 11 sequences, and each position counted as positive needed to be identical for nine of the sequences. Yeast, Basidiomycota, and *Dictyostelium* sequences were analyzed by similar criteria, and then examined for regions identified within metazoan groups. Coiled-coil analysis was conducted using <http://www.russell.embl.de/cgi-bin/coils-svr.pl>, which uses an algorithm developed by Lupas *et al.* (1991). All sequences used for non-FH2 analysis were analyzed. Only scores >0.9 (out of 1) in the 28-residue window were counted as suggestive of coiled-coil region. Although most sequences possessed strong coiled-coil predicted sequence toward the C terminus of the FH2 domain (supporting structural results, Shimada *et al.*, 2004; Xu *et al.*, 2004), this region was not reported because the focus was on characterizing non-FH2 domain sequences.

RESULTS

Searching for FH2 Domain Sequences

The most studied formin to date is Bni1p, with genetic, cellular, and biochemical information available (Waller and Alberts, 2003; Zigmond, 2004). In addition, the atomic structure of Bni1p's dimeric FH2 domain (residues 1348–1766) is known (Xu *et al.*, 2004). This structure demonstrates that dimerization occurs by one loop of amino acids (the "lasso") encircling another loop (the "post") in a highly stable interaction. A "knob" region between the lasso and post on each monomer projects above the plane of the lasso/post interface, giving the dimer a boat-like shape (Figure 1A). The structure of a segment of the FH2 domain from the mammalian formin mDia1 is also known (Shimada *et al.*, 2004). The construct used for this work lacks the N-terminal lasso region, and the resulting protein is monomeric. However, the mDia1 partial FH2 domain adopts a structure largely similar to that of Bni1p FH2. An mDia1 construct similar to Bni1p 1348-1766 seems dimeric (Li and Higgs, 2003; Xu *et al.*, 2004), but structural information for this construct is unavailable.

We searched multiple protein and nucleic acid databases for FH2 domains in an iterative manner, varying the search sequence between known yeast, metazoan, and "protist" FH2 domains (we define protist as nonfungal, nonmetazoan, and nonplant). To distinguish between true FH2 domains and false positives, we used several criteria. First, the sequence had to contain a plausible post-region, corresponding to the previously defined "core FH2" (Wasserman, 1998), from R1528 to F1589 (Bni1p numbering used unless otherwise stated). This is the most highly conserved region in FH2 domains, with R1528, L1573, G1576, N1577, N1580, and F1589 being particularly conserved. Second, key residues in the N-terminal lasso region were examined, including aromatics at 1363 and 1374, and an aliphatic at 1373. Third, the sequence should have an aliphatic (almost always an I) at 1431. Fourth, the sequence should extend 150–200 residues beyond the core FH2 region. Although the C termini of FH2

Table 1. FH2 domain sequences used in this study

Species	Classification	Protein	Identifying #	Group	Common name	Truncated?
<i>S. cerevisiae</i>	Ascomycota	SC Bni1	NP014128		Bni1	
<i>S. cerevisiae</i>	Ascomycota	SC Bnr1	NP012107		Bnr1	
<i>S. pombe</i>	Ascomycota	Sp Cdc12	Q10059		Cdc12	
<i>S. pombe</i>	Ascomycota	Sp For3	O94532		For3	
<i>S. pombe</i>	Ascomycota	Sp Fus1	Q10719		Fus1	
<i>A. gossypii</i> (<i>Eremothecium gossypii</i>)	Ascomycota	Ag 1	AAS54041			
<i>A. gossypii</i> (<i>E. gossypii</i>)	Ascomycota	Ag 2	AAS53672			
<i>A. gossypii</i> (<i>E. gossypii</i>)	Ascomycota	Ag 3	AAS54127			
<i>N. crassa</i>	Ascomycota	Nc	EAA26755.1			
<i>Aspergillus nidulans</i>	Ascomycota	An	eaa57863.1		Cytokinesis sepA.	
<i>Saccharomyces castellii</i>	Ascomycota	Scast	aacf01000011.1			
<i>Kluyveromyces waltii</i>	Ascomycota	Kw	aadm01000174.1			
<i>Candida albicans</i>	Ascomycota	Ca 1	contig 6-2433 9980-11230			
<i>C. albicans</i>	Ascomycota	Ca 2	contig 6-2516 146417-147655			
<i>Encephalitozoon cuniculi</i>	Microsporida	Ec	pfam q8sux7			
<i>Ustilago maydis</i>	Basidiomycota	Um 1	aacp01000040.1			
<i>Ustilago maydis</i>	Basidiomycota	Um 2	aacp01000148.1			
<i>Phanerochaete chrysosporium</i>	Basidiomycota	Pc	JGI 153:13551-13637			
<i>D. melanogaster</i>	Insecta	Dm 1	AAF45601	DAAM		
<i>D. melanogaster</i>	Insecta	Dm 2	AAO39658	FRL		
<i>D. melanogaster</i>	Insecta	Dm Dia	T13170	Dia	Diaphanous	
<i>D. melanogaster</i>	Insecta	Dm Cappu	NP722951	FMN	Cappuccino	
<i>D. melanogaster</i>	Insecta	Dm 6	NP729410	FHOD		
<i>D. melanogaster</i>	Insecta	Dm 7	BAC76439	INF		
<i>A. thaliana</i>	Plant	At 14	TIGR 68069.t00042	AT group 1		
<i>A. thaliana</i>	Plant	At 16	TIGR 67601.t00020	AT group 1		
<i>A. thaliana</i>	Plant	At 18	TIGR 51105.t00023	AT group 1		
<i>A. thaliana</i>	Plant	At 1	gi6503010	AT group 2		
<i>A. thaliana</i>	Plant	At 5	TIGR 67856.t00014	AT group 2		
<i>A. thaliana</i>	Plant	At 8	TIGR 50885.00033	AT group 2		
<i>A. thaliana</i>	Plant	At 2	TIGR 5256.t00032	AT group 2		
<i>A. thaliana</i>	Plant	At 3	TIGR 68154.t01289	AT group 2		
<i>A. thaliana</i>	Plant	At 4	TIGR 50826.t00057	AT group 2		
<i>A. thaliana</i>	Plant	At 6	TIGR 67936.t00019	AT group 2		
<i>A. thaliana</i>	Plant	At 7	TIGR 51305.t00047	AT group 2		
<i>A. thaliana</i>	Plant	At 11	TIGR 60210.t00017	AT group 2		
<i>A. thaliana</i>	Plant	At 19	TIGR 67601.t00022	AT group 1		
<i>A. thaliana</i>	Plant	At 21	pfam q9lvn1	AT group 1		
<i>A. thaliana</i>	Plant	At 22	pfam q95rr2	AT group 1		
<i>A. thaliana</i>	Plant	At 23	pfam q9lk78	AT group 1		C-term
<i>Chlamydomonas reinhardtii</i>		Cr	scaffold 177: 49636			
<i>Dictyostelium discoideum</i>	Dictyosteliida	Dd 3	BAC16796		FHPA	
<i>D. discoideum</i>	Dictyosteliida	Dd4	AAO51197			
<i>D. discoideum</i>	Dictyosteliida	Dd 2	bac16797		FHPB	
<i>D. discoideum</i>	Dictyosteliida	Dd 1	bac16798		FHPC	
<i>Entamoeba histolytica</i>	Entamoebidae	Eh	pfam q9ngx1			
<i>Plasmodium falciparum</i>	Apicomplexa	Pf 1	NP703650			
<i>P. falciparum</i>	Apicomplexa	Pf 2	NP701549			
<i>Thalassiosira pseudonana</i>	Diatoms	Tp 1	JGI 7:112780-112944			N-term
<i>T. pseudonana</i>	Diatoms	Tp 2	JGI 124:5788-6027			
<i>T. pseudonana</i>	Diatoms	Tp 3	JGI 131:10761-10970			C-term
<i>C. elegans</i>	Nematoda	Ce 5	NP503132	DAAM		
<i>C. elegans</i>	Nematoda	Ce 3	NP497505	FRL		
<i>C. elegans</i>	Nematoda	Ce 6	NP741210	Dia	CYK1 (LET-794)	
<i>C. elegans</i>	Nematoda	Ce 1	NP740839	FHOD		
<i>C. elegans</i>	Nematoda	Ce 2	NP497334	INF		
<i>C. elegans</i>	Nematoda	Ce 7	pfam Q19479			
<i>Gallus gallus</i>	Chordata	Gg 2	BAB20321	Dia		
<i>G. gallus</i>	Chordata	Gg 1	A41724	FMN		
<i>Danio rerio</i>	Chordata	Dr	CAE49895	Delphilin		
Mouse*	Chordata	Mm FRL1	AF215666 (np005883)	FRL	FRL1	
Mouse*	Chordata	Mm FRL2	XM_128263 (np443137)	FRL	FRL2	
Mouse*	Chordata	Mm FMN2	NP_062318 (xp371352)	FMN	Formin 2	
Mouse*	Chordata	Mm mDia2	NP_062644 (Q9NSV4)	Dia	mDia2	
Mouse*	Chordata	Mm FMN1	NP_034360 (bac86815)	FMN	Formin 1	
Mouse*	Chordata	Mm mDia1	NM_007858 (o60610)	Dia	mDia1	
Mouse*	Chordata	Mm Delphilin	NP_579933 (xp294249)	delphilin	delphilin	
Mouse*	Chordata	Mm INF1	XP_130991 (xp034262)	INF	INF1	
Mouse*	Chordata	Mm FHOD1	bac27106 (aao38757)	FHOD	FHOD1	

Table 1. (Continued)

Species	Classification	Protein	Identifying #	Group	Common name	Truncated?
Mouse*	Chordata	Mm mDia3	bac40476 (o60879)	Dia	mDia3	
Mouse*	Chordata	Mm DAAM1	aar05118 (np055807)	DAAM	DAAM1	
Mouse*	Chordata	Mm DAAM2	aar05119 (np056160)	DAAM	DAAM2	
Mouse*	Chordata	Mm FHOD2	bac98303 (xp371114)	FHOD	FHOD2	
Mouse*	Chordata	Mm FRL3	XP_288949 (np783863)	FRL	FRL3	
Mouse*	Chordata	Mm INF2	NP_940803 (bc008756)	INF	INF2	
<i>Ciona intestinalis</i>	Chordata	Ci 1	JGI 326:24696-24956	Dia		
<i>C. intestinalis</i>	Chordata	Ci 2	JGI 120:50920-51075	Delphilin		
<i>C. intestinalis</i>	Chordata	Ci 3	JGI 304:93803-93898	DAAM		N-term
<i>C. intestinalis</i>	Chordata	Ci 4	JGI 76:144416-144649	FRL		
<i>C. intestinalis</i>	Chordata	Ci 5	JGI 651:21893-22057	INF		
<i>C. intestinalis</i>	Chordata	Ci 6	JGI 67:139721-139945	INF		
<i>C. intestinalis</i>	Chordata	Ci 8	JGI 1968:1960-2043	FMN		C-term
<i>C. intestinalis</i>	Chordata	Ci 9	JGI 20:174173-174244	FMN		N-term
<i>Takifugu rubripes</i>	Chordata	Tr 1	JGI 93:130960-131265	FRL		
<i>T. rubripes</i>	Chordata	Tr 2	JGI 834:39453-39698	FRL		
<i>T. rubripes</i>	Chordata	Tr 3	JGI 1434:78301-78504	FRL		
<i>T. rubripes</i>	Chordata	Tr 5	JGI 548:116536-116655	DAAM		
<i>T. rubripes</i>	Chordata	Tr 6	JGI 51:230037-230282	FRL		
<i>T. rubripes</i>	Chordata	Tr 15	JGI 562:118846-119067	Delphilin		
<i>T. rubripes</i>	Chordata	Tr 16	JGI 170:265901-266107	Delphilin		
<i>T. rubripes</i>	Chordata	Tr 18	JGI 1675:38617-38718	INF		
<i>T. rubripes</i>	Chordata	Tr 19	JGI 22:165389-165493	INF		
<i>T. rubripes</i>	Chordata	Tr 20	JGI 153:261105-261212	INF		
<i>T. rubripes</i>	Chordata	Tr 21	JGI 795:17295-17477	FHOD		
<i>T. rubripes</i>	Chordata	Tr 22	JGI 340:221413-221607	FHOD		
<i>T. rubripes</i>	Chordata	Tr 23	JGI 237:221413-221607	FHOD		
Trypanosome	Kinetoplastida	Tb 1	TIGR 5691 12c12			
Trypanosome	Kinetoplastida	Tb 2	TIGR 5693 1047053511755			
Trypanosome	Kinetoplastida	Tb 3	TIGR 5693 1047053508641			
<i>Leishmania major</i>	Kinetoplastida	Lm 1	Sanger 5664 LM16.2 contig 192			

Truncated? refers to whether the sequence is missing either a portion of the N or C terminus. In cases where truncated sequences are analyzed, we strongly suspect that the reason lies in the database and not the fact that the putative protein is missing this portion of its FH2 domain.

* Human accession numbers given in parentheses.

domains diverge significantly, this region seems important for structure and function. Whereas some sequences contain only a subset of these features, all those included for phylogenetic analyses possess sufficient features to be defined as FH2 domains with a high level of confidence. Several sequences are truncated at either N or C termini (Table 1), but they are included because truncation is likely due to incomplete database information rather than actual termination.

These searches identified 101 putative FH2 domains from 26 diverse eukaryotic species (Table 1). Fifteen mouse FH2-containing genes were identified, of which four (INF1, INF2, FRL2, and FRL3) had not been previously recognized as formins. Humans possess homologues to all 15 mouse sequences (Table 1). Although our searches in mouse and human databases were extensive, additional FH2 domain-containing proteins may remain to be identified in these organisms. Furthermore, some mammalian species may possess additional formins not found in mouse or humans.

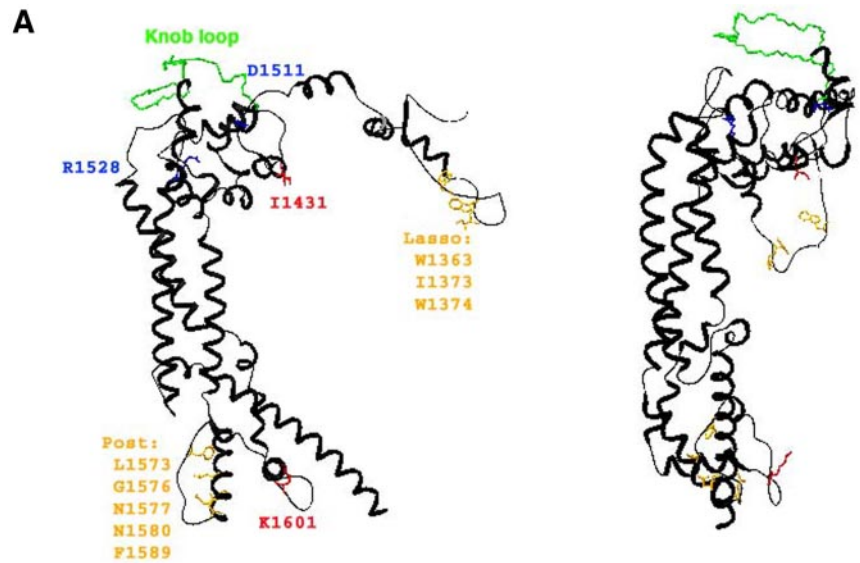
Four other metazoans were examined in detail: *T. rubripes* (15 sequences), *C. intestinalis* (8 sequences), *D. melanogaster* (6 sequences), and *C. elegans* (7 sequences). Although their genomes are not complete, the fish and tunicate (sea squirt) sequences provide useful data points bridging the fully sequenced mammalian, insect, and worm genomes, fish being vertebrate chordates and tunicates being invertebrate chordates (Hedges, 2002). Their inclusion allows us to evaluate phylogenetic groupings with more confidence, because se-

quence relatedness should follow an order of (nematode (fly (ascidian (fish (chick, mouse))))).

From fungi, we identified 18 putative FH2 domains, including 14 from Ascomycota (yeast, including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), three from Basidiomycota, and one from Microsporidia. From protists, we identified 15 FH2s, with organisms including trypanosomes (3 sequences), *Chlamydomonas* (1 sequence), *Dictyostelium* (4 sequences), *Entamoeba* (1 sequence), *Plasmodium* (2 sequences), and diatoms (3 sequences). Although the genomes of many of the fungal and protist species used are not fully sequenced, their inclusion allows the possibility of detecting relationships within and between groups. For plants, we used the FH2 domains previously identified for *Arabidopsis* (Cvrckova *et al.*, 2004).

FH2 Domain Alignments

We next aligned the 101 FH2 domains using ClustalW and subsequently refined this alignment to place gaps at plausible regions, given the available structural information. Specifically, gaps were not placed in regions found to be helical in the Bni1p or mDia1 structures. This work extends previous analyses (Zeller *et al.*, 1999; Bateman *et al.*, 2002) in 1) the breadth of sequences used, many of which have not been previously identified as formins; and 2) the use of the FH2 domain atomic structures, giving more rigor to the placement of indels (insertions and deletions).

**B**

mDia1	891	SELKEEYDD-----LAESQ-F
mDia2	750	SQFRSDYNS-----LCEPEQ-F
mDia3	762	AQLKSEYDD-----LCEPEQ-F
DAAM1	744	EEHKHELDL-----AKADR-F
DAAM2	778	EEHKHEIER-----ARADR-F
FRL1	761	ARFEKEQRP-----MEELSEDR-F
FRL2	651	RQYERERQP-----LEELAAEDR-F
FRL3	809	RLYERERKP-----LENLSDEDR-F
FHOD1	782	EEAQLANPD-----VPLGPAEN-F
FHOD2	1021	QEAQLANPE-----VPLGSAEQ-F
Delphilin	764	QAFREAPGR-----LSEPDQ-F
INF1	227	KAFNGDVSK-----LSLADS-F
INF2	187	RAFTEERAK-----LSNADQ-F
FMN1	1120	RKYETSKE-----EDLKLLDKPEQ-F
FMN2	1261	EKHSRSSKD-----KENAKSLDKPEQ-F
At5	562	RLYCGEIAQ-----LGS AER-F
At14	921	RNYTGDKEM-----LGKCEQ-V
Dd1	743	HELHSNEVI-----LDKPER-W
Dd3	888	NEFLREEKD-----ITKLGPEQ-F
Um1	1724	RDYDGDISA-----LSKADQ-F
Um2	1283	GEYRNASDE-----ELRLHPADR-L
		4 KNOB LOOP 5
Bni1p	1475	VNLARNYAPYSTDWEGVRNLEDAKPPKDPNDLQRADQIY
Nc	1286	DNTAKVMAPYSKDWG--PDAGKDTREQDPNELTRQDQIY
An	1274	ENVSKLMAPYSKDWG--PDAANTEREQDPSELTRQDQIY
Scast	1415	VNLARNYAPYSTDWEGVKSVEDAKAPEKDPNELQRADQLY
Kw	1442	NNLARNFAPYTTDWEGITNVKDAKPPKDPSELQRADRLY
Ag2	1494	TNMARSFAPYTTDWEGVSSVEDAKSPEKDPSELQRSQDLF
Cdc12	1103	PKLKADLEPYRIDWNG---GDLVNSAEKDASELSRWYDLY
Fus1	991	EGLKKQLKPYMSSRNE-----EVMEVSEKLELSRFDQIY
For3	970	NAYTDPFIPYSTDYKK----NPKEPTTADVNSLSYFEKFF
Bnr1p	1011	KSMLNKYEPYSQKDG-----KAVSDLQRADRIF
Ag1	1289	QSIQRLFGPYETNYLTG--ERPEKDPALERADRIYLELF
Ag3	895	OSTQRMFAPYETNYLTG--KTPDKDPVLERADRIYLELF
		I KNOB LOOP J

Figure 1. Summary of FH2 domain alignment results. (A) Models of one subunit from the dimeric Bni1p FH2 domain structure, made using DeepView (www.expasy.org/spdbv/) based on published structure (Xu *et al.*, 2004; PDB file 1UX5). Left, highly conserved residues are shown in orange for dimerization interface (W1363 = 95/101 sequences identical; I1373 = 85/101 for aliphatic and 97/101 for aromatic; W1374 = 97/101 for aromatic; L1573 = 95/101 identity and 99/101 for aliphatic; G1576 = 98/101 identity; N1577 = 100/101 identity; N1580 = 98/101 identity; F1589 = 95/101 as F or Y), red for residues key to effects on actin (I1431 = 93/101 identity and 98/101 as I or V; K1601 = 98/101 as base), and blue for residues important for knob structure (D1511 = 93/101 as acid; R1528 = 100/101 as R or K). The extended knob loop, unique to the Ascomycota fungi, is shown in green. Right, rotated 90° to show the extent to which the Ascomycota knob loop (green) projects over the rest of the knob. (B) Alignment of a segment of the knob region for a selection of FH2 domains. The first set of sequences is from non-Ascomycota, including 15 mouse sequences (to FMN2), one member of each *Arabidopsis* group, two *Dictpostelium* sequences, and two Basidiomycota sequences. The second set of sequences is from Ascomycota. Refer to Table 1 for sequence abbreviations. Below the alignment, bars depicting helices (black) and the knob loop regions (gray) in the structures of mDia1 and Bni1p, respectively, are illustrated.

Our alignment is consistent with many aspects of the Bni1p and mDia1 structures. Although no amino acid position is 100% identical in all sequences, several key residues are highly conserved (Figure 1A). Positions with >90% iden-

tity include W1363 (numbering for Bni1p), L1573, G1576, N1577, N1580, and F1589, all being important to the lasso/postdimerization interaction. Two other residues with >90% identity, I1431 and K1601 (conserved as a basic residue), are

surface exposed in the Bni1p structure and crucial for effects on actin (Xu *et al.*, 2004). Two residues important for knob region structure, D1511 and R1528, also are highly conserved. D1511 is an acidic residue in all but seven sequences, whereas R1528 is a basic residues in all formins.

One of the least conserved regions of FH2 domains is the knob, particularly at a region called the “knob loop” for the Bni1p structure (Xu *et al.*, 2004; Figure 1, A and B). Intriguingly, all 14 Ascomycota FH2 domains present in the alignment (including Bni1p) possess a long knob loop (25 residues in Bni1p, median = 24), whereas the 87 non-Ascomycota (including the four other fungal sequences) contain much more modest loops (9 residues in mDia1, median = 9). The function of the knob loop is unknown, but the evolutionary maintenance of such dramatic differences in loop length between Ascomycota and non-Ascomycota is suggestive of functional relevance. The Bni1p knob loop projects over a significant area of the knob outer surface (Figure 1A).

Phylogenetic Analysis of FH2 Domains

From our alignment, we constructed phylogenetic trees with both ME and MP algorithms (Figure 2). Because these algorithms use completely different criteria to search for optimal trees (Nei, 1996), comparison of results from both is helpful when assessing the robustness of groupings. In both cases, 1000 bootstraps were executed, with starting trees for branch swapping generated by random stepwise addition (10 replicates/bootstrap). Bootstrapping tests tree precision by repeatedly sampling random positions with replacement, meaning that the same position can be sampled multiple times (Page and Holmes, 1998). The bootstrap values reported represent percentages of how many times an internal node (branch point) is found at the indicated location. Random stepwise addition of sequences (as opposed to neighbor joining) reduces the likelihood that local, instead of global, minima are found, increasing sampling of tree space (Page and Holmes, 1998).

Figure 2A depicts the MP tree in detail, with bootstrap values provided for nodes supported at >50% for both MP and ME. Both algorithms give very similar topologies. Figure 2B depicts a simplified version of the same tree, to highlight our main findings. We will describe our results by discussing first the relationships between mammalian and other metazoan FH2 domains before progressing to their relationships with FH2 domains from other organisms. One important note is that we present cladograms instead of phylograms, thus branch lengths do not represent evolutionary distance.

Mouse formins segregate into seven groups (Figure 2), to which we give the following names: Dia (*Diaphanous*); FMN (*ForMiN*); FHOD (*Formin HOMology Domain*-containing protein); delphilin; INF (*INverted Formin*); FRL (*Formin-Related gene in Leukocytes*); and DAAM (*Dishevelled-Associated Activator of Morphogenesis*). The FMN group contains the mouse proteins originally named “formins,” identified as limb deformity loci (Woychik *et al.*, 1990). The FHOD group had been known as FHOS (*Formin Homologue Overexpressed in Spleen*; Westendorf *et al.*, 1999) but has been changed by agreement between the laboratories concerned. The INF group is novel. Although our analysis uses mouse as the mammalian representative, a similar tree containing the analogous human sequences (accession numbers given in Table 1) produces the same topology and similar bootstrap values (our unpublished data).

The four other metazoans examined in detail, *T. rubripes*, *C. intestinalis*, *D. melanogaster*, and *C. elegans*, possess at least one member in each of the seven mammalian groups, with

three exceptions. First, no delphilin homologue was found for *Drosophila* or *C. elegans*. Second, no FMN group homologue was identified in *C. elegans*, despite extensive searching of *C. elegans* (and *Caenorhabditis briggsae*) protein, nucleotide and genomic databases. Third, *C. elegans* has an FH2-containing gene that does not group with any others (Ce 7), and the phylogeny does not suggest a relationship with the FMN group.

Three points suggest that the overall arrangement of the seven metazoan groups is robust. First, all seven groups are strongly supported by both algorithms (MP and ME). Second, we have tested this result by performing maximum likelihood (ML) analysis on a limited set of sequences (17 sequences, including two members of each group and three out-groups, underlined sequences in Figure 2), again finding strong support for all seven groups. Third, the interrelationships among the metazoan FH2 domains correlates with the generally accepted phylogeny of the species [i.e., (nematode (fly (ascidian (fish (chick, mouse))))))], and thus the gene tree is concordant with the species tree.

We were generally unable to elucidate with any precision the interrelationships among the seven metazoan groups, except that both the ML and ME analyses suggest that the FRL and DAAM groups might be related. This result is further supported by regions of sequence similarity outside of the FH2 domain (see next section).

With respect to the FH2 domains from the other eukaryotes, the 16 Arabidopsis FH2 domains analyzed seem to be two distinct groups, as found by others (Cvrckova *et al.*, 2004). Neither ME nor MP analysis could relate these groups to the seven metazoan groups. Other FH2 domains from Basidiomycota and Microsporidia fungi, Apicomplexa (*Plasmodium*), diatoms, Kinetoplastida (trypanosomes and *Leishmania*), *Chlamydomonas*, and amoeba do not fall into any of the seven metazoan or two plant groups. The one exception is that ME supports at 79% a relationship between *Dictyostelium* formins Dd1 and Dd2 and the FMN group (our unpublished data).

An important result is that all 14 FH2 domains from Ascomycota fungi, including *S. cerevisiae* and *S. pombe*, segregate from the other eukaryotes, including the four other fungal sequences (Figure 2). This distinction is supported with high precision by both phylogenetic algorithms (bootstrap values of 95 and 97 for MP and ME, respectively). Removal of the extra knob loop residues from the yeast sequences does not alter this relationship. Given that fungi are clearly monophyletic (Taylor *et al.*, 2004), our analysis suggests that Ascomycota FH2 domains have diverged from those of other eukaryotes to a degree exceeding their overall evolutionary placement.

Non-FH2 Domain Relationships

We probed further for relationships between FH2 domain-containing proteins by conducting sequence comparisons of non-FH2 regions. First, we aligned members of each metazoan group separately to test the validity of our FH2 phylogeny-based metazoan groupings. Second, we conducted pairwise alignments between metazoan groups to probe for intergroup relationships. Third, we analyzed Ascomycota formins for relationships. Fourth, we analyzed relationships between individual metazoan groups and Ascomycota and conducted similar analysis for one protist (*Dictyostelium*).

To judge similarity within groups, we aligned all group members for which we possessed apparent full-length sequence. We then scanned the alignment for regions of at least 26 residues that displayed $\geq 50\%$ sequence identity in n-1 sequences, with no gaps for any individual sequence (see

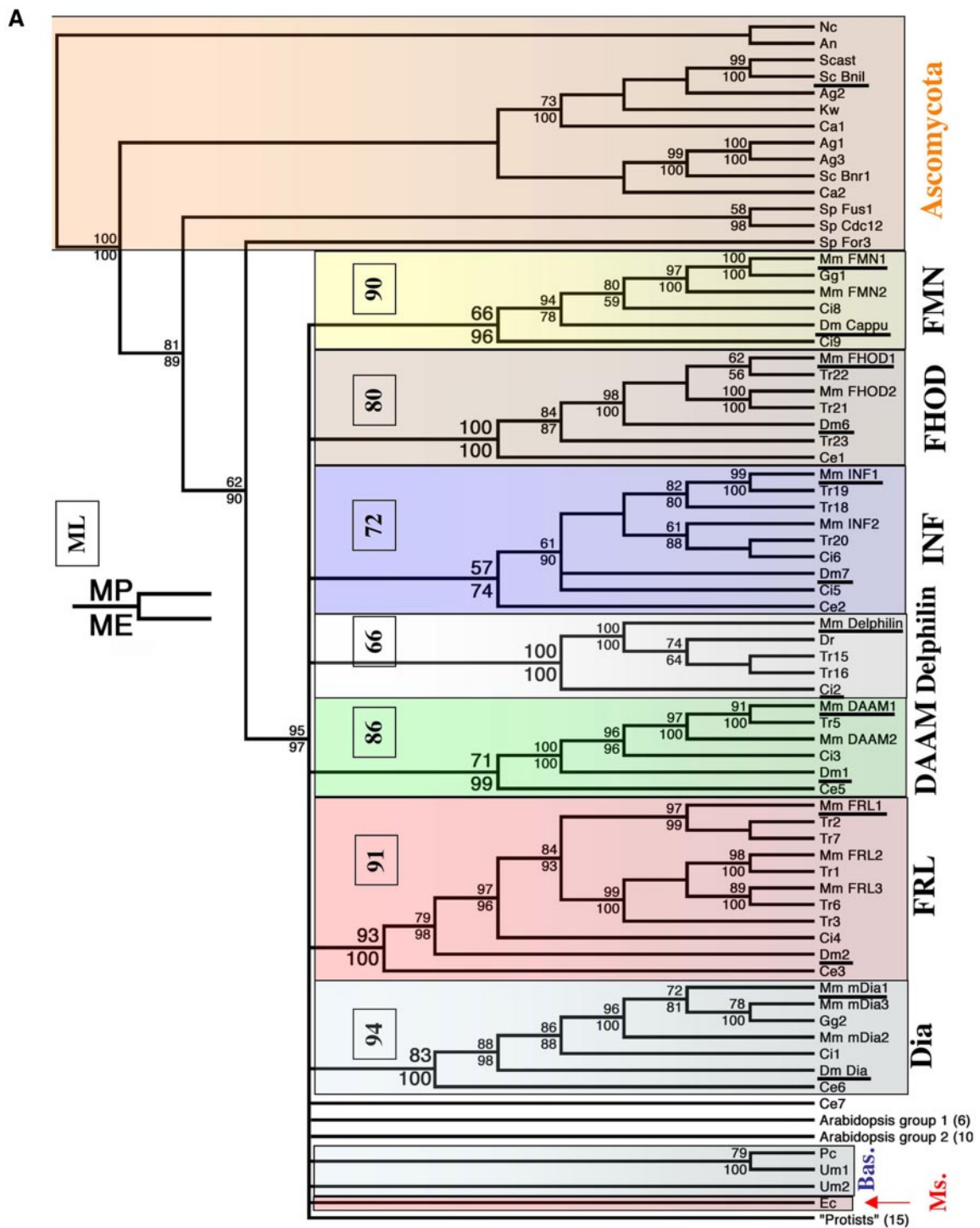


Figure 2. Cladograms of FH2 domains. (A) MP cladogram shown. Branch lengths of cladograms do not depict evolutionary distance. The FH2 domain alignment was analyzed with PAUP version 4.0b10 for Macintosh (Swofford, 2002). Distance analysis used ME as the optimality criterion, and mean character difference as the distance measure. Bootstrap analysis for both MP and ME used 1000 replicates by using heuristic search with tree-bisection-reconnection and random addition sequence with 10 replications per bootstrap. MP bootstrap is given at the top of the node and ME at the bottom. Bootstrap values are not given at nodes that are not supported at $>50\%$ for both algorithms. Maximum-likelihood analysis of 17 sequences (two species from each metazoan group [underlined], Sc Bn1p, and one member of Arabidopsis group 1 [At14] and group 2 [At1]) used Quartet Puzzle version 5.0 (JTT model of amino acid substitution; 1000 puzzling steps); these values are shown in the small box for each of the seven families, which are boxed in color. The fungal branches also are boxed: Ascomycota (orange), Basidiomycota (Bas., blue), and Microsporida (Ms., red). Dm Dia and Dm Cappu indicate the Diaphanous and Cappuccino proteins, respectively, from *Drosophila* (Castrillon and Wasserman, 1994; Emmons *et al.*, 1995). The branch labeled "protists" contains sequences from Apicomplexa (*Plasmodium*), diatoms, Kinetoplastida (trypanosomes and *Leishmania*), *Chlamydomonas*, and amoeba. Numbers in parentheses below the Arabidopsis and protist groups denote the number of sequences in each branch.

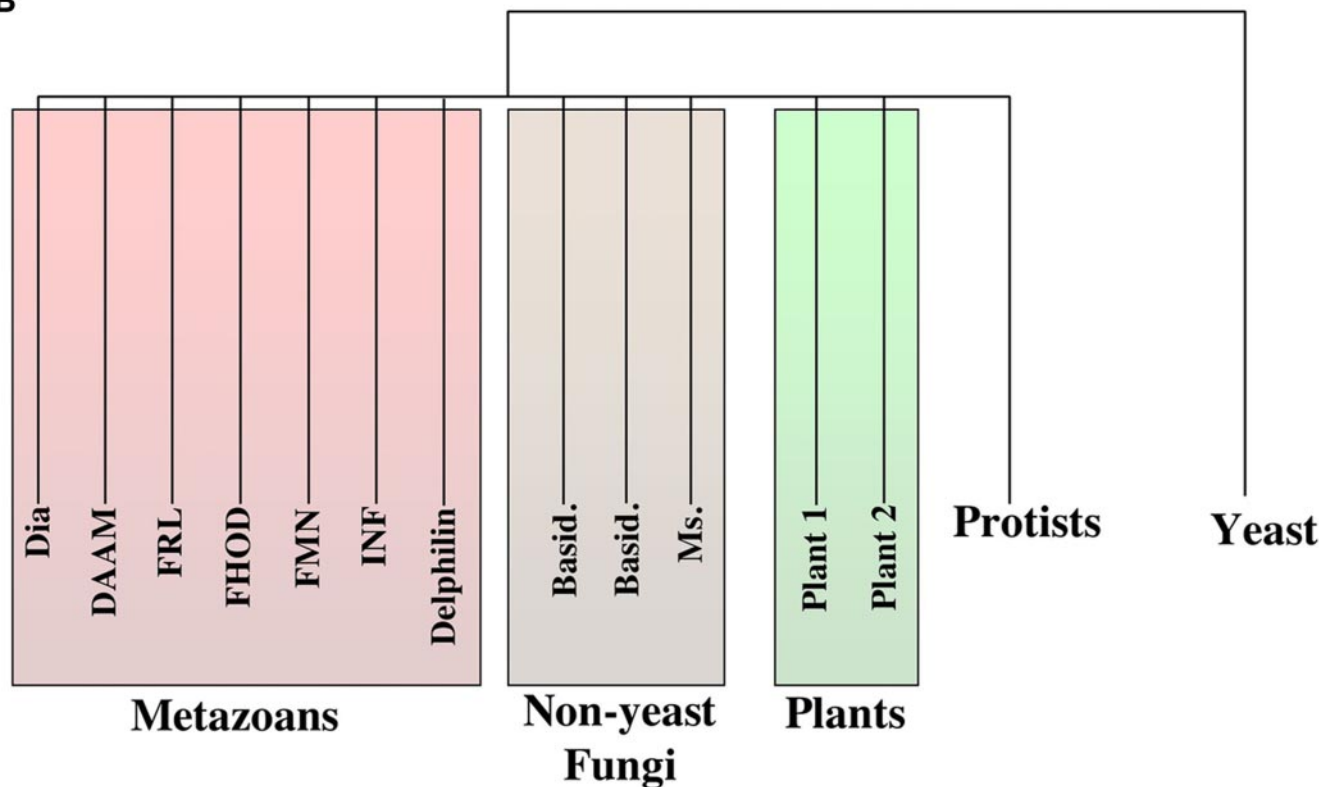
B

Figure 2b. (B) Simplified cladogram, from data in A, to show the main relationships. “Basid.” refers to Basidiomycota yeast, whereas “Ms.” refers to Microsporidia.

Materials and Methods). We used sequence identity instead of chemical similarity, due to the case-by-case nature of similarity between amino acids (e.g., arginine and lysine are both basic, but their hydrogen bonding capabilities are very different). We did not take into account putative domains identified by others for individual formins (GBD, FH3, and DAD), to analyze the sequences in an unbiased manner. Our purpose was not to identify functional domains per se but to assess similarity.

By these criteria, the Dia, FRL, DAAM, dephilin, and FHOD groups display several regions of similarity outside of the FH2 domain (Table 2). Conversely, the INF and FMN groups display no similar regions by these criteria. Even when stringency is reduced or gaps are allowed, no significant similarity regions are found for the FMN or INF groups outside of their FH2 domains.

To probe for similarities between metazoan groups, we conducted pairwise alignments using all sequences in each group. Members of the Dia, FRL, and DAAM groups seem similar in two regions N-terminal to the FH1 domain, named regions N1 and N2 (Figures 3 and 4). N2 forms part of a larger region of similarity, extending ~130 residues C-terminal to the start of N2 (Figure 3). Members of the other four metazoan groups do not contain regions with significant similarity to N1 or N2.

A region of similarity between the Dia, DAAM, FRL, and FHOD groups is also found C-terminal to the FH2 domain (Figures 3 and 5). This region (C1) corresponds to the previously identified DAD, whose binding to an N-terminal region of Dia formins results in autoinhibition (Alberts, 2001). This sequence seems absent in the delphilin, INF, and FMN groups.

Others have noticed that some formins contain a potential coiled-coil sequence N-terminal to the FH1 domain (Wallar and Alberts, 2003). We examined this possibility for all metazoan formins by using a well-established algorithm (Lupas *et al.*, 1991), with a minimum score of 0.9 in a 28-residue frame being our cut-off. Members of the Dia, FRL, DAAM, and FMN groups contain predicted coiled-coil regions N-terminal to the FH1 (Supplementary Table 1 and Figure 3).

The other metazoan groups (Delphilin, INF, and FHOD) do not possess clear coiled-coil regions in similar positions. Most have extremely low probabilities of coiled coils in all sequence outside the FH2 domain (Supplementary Table 1). A possible exception is the FHOD group. Mouse FHOD1 possesses a weakly predicted 21 residue coiled-coil, and FHOD2 contains a more strongly predicted 28 residue coiled-coil, almost 400 residues N-terminal to its FH1. FHOD proteins from puffer fish and *Drosophila* (but not *C. elegans*) have strong predicted 28 residue coils (Supplementary Table 1).

We also aligned non-FH2 regions from Ascomycota sequences, by using 12 of the 14 sequences (the *Candida* sequences seem incomplete). Six of these sequences, Bni1p, Nc, An, Scast, Kw, and Ag2 align strongly in several regions N-terminal to the FH1 domain, including two corresponding to N1 and N2 of the Dia-DAAM-FRL alignment (Figure 4) and one for which the metazoans show no apparent similarity (Figure 6A). In the latter region (N3), the motif MGSLVGAS is 100% identical. This similarity is compatible with the FH2 domain-derived tree (Figure 2), because Bni1p, Scast, Kw, and Ag2 are monophyletic in the tree. We refer to these sequences as being part of the Bni1 group.

Table 2. Non-FH2 regions of similarity in mammalian formin groups

Sequence	% Identity*
Dia group	
Six sequences aligned, numbering for mouse mDia1	
82-118	57
152-186	63
301-338	55
395-446	67
1171-1196	73
FRL group	
Eight sequences aligned, numbering for mouse FRL1.	
46-71	58
242-296	65
315-414	57
DAAM group	
Five sequences aligned, numbering for mouse DAAM1	
38-88	53
137-348	67
349-392	57
1031-1056	77
Delphilin group	
Five sequences aligned, numbering for mouse delphilin	
88-119	72
139-166	89
181-251	73
252-279	68
FHOD group	
Six sequences aligned, numbering for mouse FHOD1	
113-304	50

INF group: seven sequences aligned, no regions of similarity identified; FMN group, four sequences aligned, no regions of similarity identified.

* Percentage of identity defined as percentage of positions in region in which n-1 of the aligned sequences are identical. No gaps are present in any of the regions listed.

The three *S. pombe* sequences, Cdc12, Fus1, and For3, do not easily fit into the Bni1 group. These proteins have sequences similar to N1, but not N2 (Figure 4). In N3, the *S. pombe* sequences contain identical residues at the N and C termini, but they are completely dissimilar in the highly conserved MGSLVGAS sequence (Figure 6A).

The final three Ascomycota sequences, Bnr1p, Ag1, and Ag3, do not possess recognizable N1 or N2 regions. Similar to the *S. pombe* sequences, these proteins show some similarity in N3, but they do not have the MGSLVGAS sequence. These three sequences do share a separate region of similarity in the N terminus, absent in the other Ascomycota (Figure 6B). We refer to these sequences as the Bnr1 group.

All of the Bni1 group members, and two of the *S. pombe* sequences, possess recognizable C1 sequences (Figure 5). C1 sequences for the Bnr1 group are less apparent, but plausible. The *S. pombe* formin For3 does not have a discernable C1 sequence.

Ascomycota formins are somewhat heterogeneous for the putative coiled-coil region (Supplementary Table 1). The Bnr1 group possesses strong predicted coiled-coils in regions similar to those of the four metazoan groups. For the *S. pombe* formins, For3 has a strong predicted coiled-coil sequence, whereas Cdc12 has a short predicted sequence farther removed from the FH1, and Fus1 does not have a predicted coiled-coil N-terminal to the FH1. Members of the

Bni1 group have several predicted coiled-coil sequences scattered throughout their N termini, including one or more strongly predicted sequences from 300 to 500 residues N-terminal to the FH1 (Supplementary Table 1 and Figure 3).

The regions of similarity between some of the Ascomycota sequences enabled us to ask whether the non-Ascomycota fungal sequences shared these regions. Of the three Basidiomycota (Um1, Um2, and Pc) and one Microsporidia (Ec) FH2 domain-containing sequences, only those of Um1 and Um2 seemed to contain complete open reading frames (ORFs). Um2 aligns well with the Bni1 group in N1, N2, and N3, possesses a plausible C1 region, and a strong coiled-coil sequence (Figures 4, 5, and 6A). In contrast to the non-Bni1 group Ascomycota sequences, Um2 contains a nearly identical sequence to the MGSLVGAS sequence found in N3 (Figure 6A). Um1 does not align well with the Bni1 group, the *S. pombe* formins, or the other Ascomycota.

Finally, we aligned *Dictyostelium* sequences with those of metazoans and yeast. All four *Dictyostelium* sequences contain possible N1 sequences (Figure 4A) and coiled-coil regions (Supplementary Table 1), but only Dd3 contains a possible N2 sequence (Figure 4B), as well as a C1 sequence (Figure 5). Dd4 contains a possible C1, but not N2.

DISCUSSION

Our analyses supports the following conclusions: 1) metazoan FH2 domains segregate into seven groups; 2) three of these metazoan groups, Dia, DAAM, and FRL, possess some similarities outside of the FH2 domain, and the FH2 domains of FRL and DAAM group together by using both ME and ML, suggesting a monophyletic assemblage; 3) FH2 domains from other eukaryotes do not fall into any of these seven metazoan groups; 4) FH2 domains from Ascomycota (yeast) are distinct from all of the above-mentioned sequences; and 5) sequence comparisons outside the FH2 domain suggest that some Ascomycota, Basidiomycota, and *Dictyostelium* formins have similarities to the Dia, DAAM, and FRL metazoan groups.

We discuss each of these points below, but one general point must be made first. Our conclusions are drawn primarily from phylogenetic analyses and are thus genealogical in nature rather than functional. This distinction is important for two reasons. First, the fundamentally different approach from biochemical or cellular analyses provides a completely independent view of the formin family, with the potential to reveal features not yet exposed by the other approaches. Second, our analysis allows us to separate genealogy from biochemical or cellular function, genealogy being a much more sound basis for classification because functional similarity can arise from either historical continuity or evolutionary convergence (Darwin, 1859, p 420). Also, divergent sequences do not necessitate divergent function. For example, myosin motor domain sequences diverge significantly between classes and between species, but still possess actin-based motor activity.

Relationships between FH2 Domain-containing Proteins

Our analyses reveal seven distinct metazoan groups of formins, based primarily on FH2 phylogeny. These groups are strongly supported by all three phylogenetic algorithms, and the species relationships within each group give additional confidence for these groupings. In addition, alignment of non-FH2 domain sequences reveal strong group-specific similarities between members of the Dia, DAAM, FRL, del-

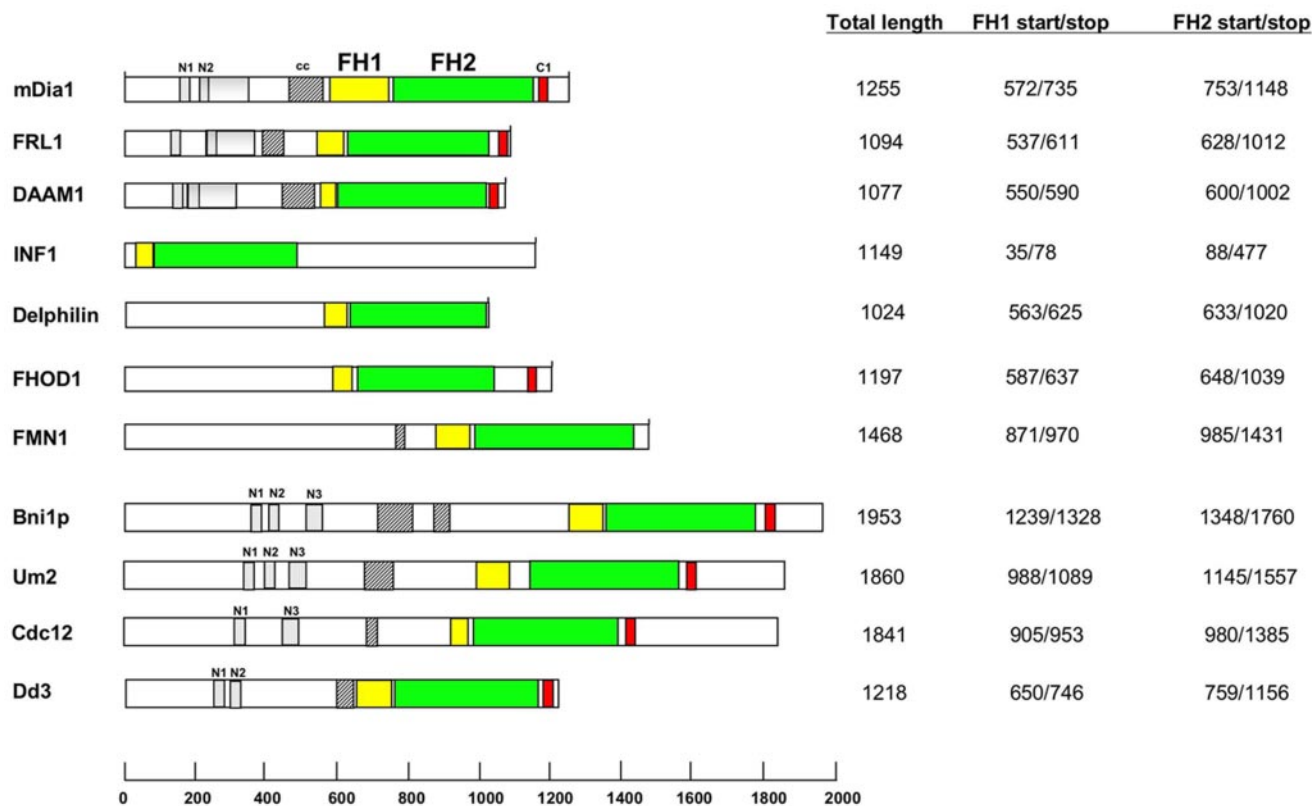


Figure 3. Bar diagram of formin proteins. One member from each metazoan group is shown (mouse protein), as well as Bni1p (budding yeast), Um2 (Basidiomycota), Cdc12 (fission yeast), and Dd3 (*Dictyostelium*). FH1 domains in yellow, FH2 domains in green, putative coiled-coil domains (cc) in diagonally lined boxes, and short region of C-terminal similarity between several groups (C1, putative DAD) in red. FH1 domains start at first proline of first poly-proline stretch, and end at last proline of last poly-proline stretch. FH2 domains determined from alignment. Dotted boxes correspond to the N1, N2, and N3 regions of similarity. The gray shaded regions for mDia1, FRL1, and DAAM1 represent the extended regions of similarity that the Dia, DAAM, and FRL groups possess C-terminal to N2.

philin, and FHOD groups, supporting the FH2-based groupings.

Although members of the INF group show no significant similarity in their non-FH2 regions, a common feature is that FH2 domain is found generally near the N terminus, whereas all other known formins possess C-terminal FH2 domains. Among the six INF group members for which the ORF seems complete (Mm INF1, Mm INF2, Dm7, Ce2, Ci6, and Tr20), five have FH2 domains in their N-terminal halves (the exception is Tr20). In separate studies, we have found that FH2 domains of mouse INF1 and INF2 are capable of nucleating actin filaments (our unpublished observations).

No members of the delphilin group in *Drosophila* or in *C. elegans* were identified. Delphilin was identified as a PDZ domain-containing binding partner of the ionotropic glutamate receptor $\delta 2$ subunit (Miyagi *et al.*, 2002). The phylogeny does not suggest a chordate-specific acquisition but instead suggests a secondary loss in the two ecdysozoan taxa, consistent with what is known about the fly and worm genomes (Copley *et al.*, 2004). Definitive testing of this hypothesis awaits full genome sequencing of nonecdysozoan protostomes (e.g., mollusks or earthworms), with the prediction that delphilin homologues would be found in these species.

We were unable to find support for relationships between four of the metazoan groups (INF, delphilin, FHOD, and FMN), but relationships between the Dia, FRL, and DAAM groups have some support. First, a relationship between FRL and DAAM FH2 domains is supported by ME and ML

analysis. Second, all three groups share two similar regions in their N termini (N1 and N2) and another C-terminal to the FH2 domain (C1). Third, all three contain a predicted coiled-coil region N-terminal to the FH1 domain, a feature not universally present in formins. The FHOD group possesses the C-terminal region, and some members contain weak predicted coiled-coil sequences, but no member of the FHOD group has an N-terminal region that aligns with the N1 or N2 regions of the Dia-DAAM-FRL groups.

The relationships between metazoan formins and formins from other eukaryotes remain elusive. FH2 domains from non-metazoans (with the exception of Ascomycota) are generally related to the metazoan groups, but there is no support for their inclusion into any metazoan group. This situation contrasts those of other cytoskeletal proteins, such as actin-related proteins (Goodson and Hawse, 2002) and myosins (Berg *et al.*, 2001). In contrast, non-FH2 regions of similarity linking three metazoan groups (Dia, DAAM, and FRL), *Dictyostelium*, and fungal formins can be identified. Possibly, the evolutionary pressures on the FH2 domain have been different than those acting on non-FH2 regions.

One possible exception is that two *Dictyostelium* FH2 domains, Dd1 and Dd2, group loosely with the FMN group by ME analysis (79% bootstrap, no relationship by MP). Because members of the FMN group possess no strong sequence similarity in non-FH2 regions, comparison with the *Dictyostelium* sequences in these regions is difficult. Nevertheless, no clear non-FH2 similarities between the *Dictyoste-*

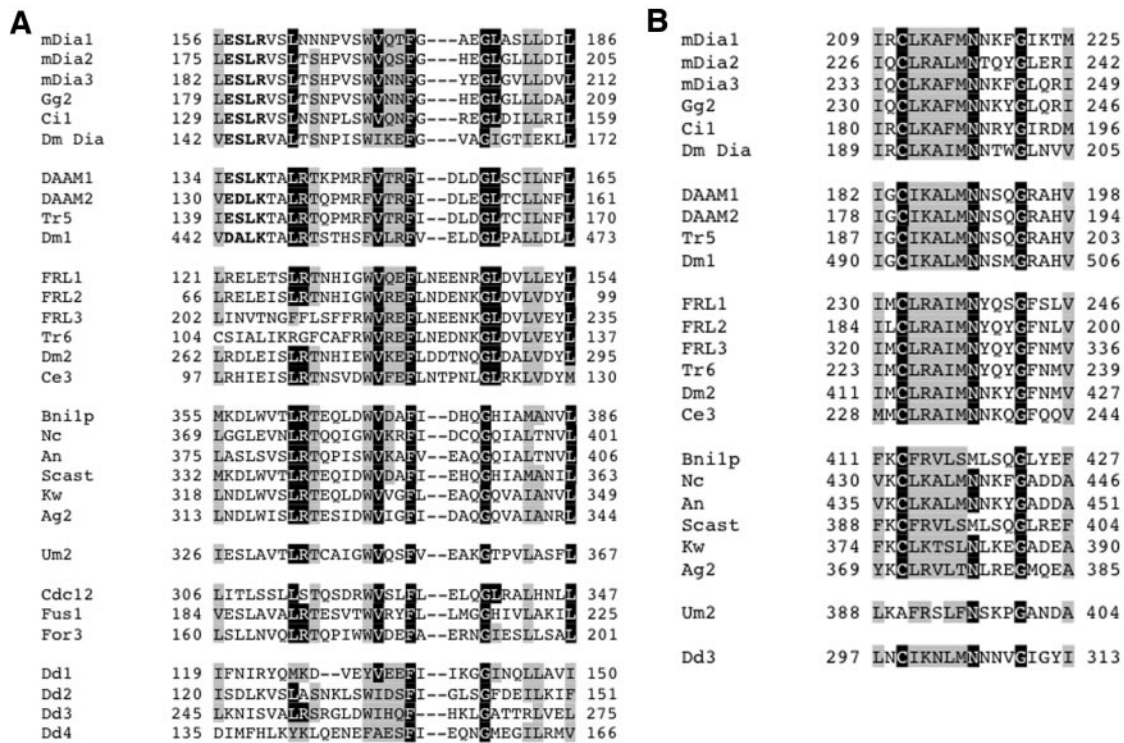


Figure 4. N-terminal sequence similarities between metazoan and nonmetazoan formins. (A) Region N1, strongly similar between the Dia and DAAM groups, with the FRL group more distantly related. Also possessing similarity are the yeast Bni1 and *pombe* groups, as well as the Basidiomycete, Um2, and the four *Dictyostelium* sequences. Letters in bold denote sequence that is strongly similar between the Dia and DAAM groups. (B) Region N2, strongly conserved between the Dia, DAAM, and FRL groups, making up the N-terminal part of a longer region of similarity (130 residues). The yeast Bni1 group, the Basidiomycete sequence Um2, and the *Dictyostelium* sequence Dd3 also possess the N2 region but not the longer region found in Dia-DAAM-FRL.

lium sequences and any individual FMN member can be detected.

Although maintaining some non-FH2 relationships with other eukaryotes, Ascomycota have diverged significantly in their FH2 domain sequences. The divergence of Ascomycota FH2 domains from those of other fungi (Basidiomycota and Microsporida) is particularly striking, because fungi are clearly monophyletic (Taylor *et al.*, 2004). In addition to the overall sequence dissimilarity to other eukaryotic FH2 domains, as indicated by the phylogenetic algorithms, Ascomycota uniformly possess extended knob loops. The function of the knob loop is unknown, but the large loop in yeast FH2 domains significantly changes the architecture in this region. In biochemical studies by us and by others (Li and Higgs, 2003; Harris *et al.*, 2004; Moseley *et al.*, 2004; Shimada *et al.*, 2004; Higgs laboratory, unpublished data), FH2 domains from Ascomycota and mammals are consistently dimeric, and all formins tested compete with capping protein for actin filament barbed ends. Thus, the divergent knob loop, and the FH2 domain divergence in general, does not affect these functions. One possibility is that the extended knob loop mediates an Ascomycota-specific interaction with another molecule.

Functional Significance of Non-FH2 Similarity Regions

The two regions of N-terminal similarity (N1 and N2) between the Dia, DAAM, and FRL metazoan groups, as well as some Ascomycota and *Dictyostelium* sequences, suggest common function. One possibility is that one or both regions

might be part of a binding site for Rho family GTPases, because the N termini of several formins bind these proteins (Wallar and Alberts, 2003). However, FHOD1 interacts specifically with Rac (Westendorf, 2001), but the FHOD group does not possess regions similar to N1 or N2.

Another possibility is that N1 and/or N2 might make up part of an autoinhibitory interaction region that binds to the FH2 domain or to the C-terminal DAD. Four formins, mDia1, mDia2, FHOD1, and Bni1p, have been proposed to be autoinhibited by interactions between an N-terminal sequence and DAD (Evangelista *et al.*, 1997; Alberts, 2001; Westendorf, 2001). A recent study from our laboratory shows that a protease-resistant region in mDia1 from residues 129–369 is able to inhibit nucleation by mDia1's C terminus in a DAD-dependent manner, and we name this region DID, for diaphanous inhibitory domain (Li and Higgs, unpublished data). Although the DAD has been mapped (see following paragraph), the identity of the N-terminal binding site is unknown.

Based on the mapped DAD sequence for mDia2 (Alberts, 2001), several proteins possess a loose consensus in the C1 similarity region we identify. Whether all of these sequences constitute autoinhibitory binding sites remains to be tested. We must emphasize that the sequences we predict to be DADs for the DAAM, FRL, and FHOD groups, as well as for yeasts and *Dictyostelium*, are very much putative, and experiments testing the roles of these sites in regulation are needed.

Perhaps equally interesting is the fact that two yeast proteins, Bni1p and For3, and three metazoan groups, INF, delphilin, and FMN, do not possess clearly recognizable

mDia1	1180	GVMDSLLEAL
mDia2	1039	GVMDNLEAL
mDia3	1051	GVMDNLEAL
Gg2	1208	GVMDSLLEAL
Ci1	1025	GVMDSLLEAL
Dm Dia	1027	GVMDSLLEAL
DAAM1	1030	GEFDDLVSAI
DAAM2	1068	GEFDDLVSAI
Tr5	1023	GEFDDLVSAI
Dm1	1390	GEFDDLISAI
FHOD1	1139	EIMDLLVQSV
FHOD2	1369	EIMDRIVKSA
Tr21	1177	EIMERIVRSA
Dm6	1337	EILESIVKTA
Ce1	1015	EILNGLVKAA
FRL1	1053	GAIEDIITDI
FRL2	947	GTIEDIITGF
FRL3	1108	GAIEDIITAI
Tr6	964	GAIEDIITDI
Dm2	1231	GAIEDILLGL
Ce3	1063	GDFEKLMLGV
Bni1p	1799	AVMDRLLEQL
Nc	1595	GAMDSLLEKL
An	1585	GAMDSLLEKL
Scast	1737	AMMDKLLDQL
Kw	1763	DVMDRLLEKL
Ag2	1815	DVMDRLLEKL
Um2	1583	AVLETLLGSL
Cdc12	1410	EAMDGLDKI
Fus1	1298	ATVNLISQL
Bnr1p	1312	DAVDLLISKI
Ag1	1600	DAVDLRRRI
Ag3	1203	HAVDVLLKRI
Dd3	1182	SVVDLDDTH
Dd4	1078	TFMDLMSKM

Figure 5. C1 region of similarity (“DAD” sequence). All sequences located no more than 100 residues C-terminal to FH2. The sequences for mDia2, Bni1p, and FHOD1 show evidence of participating in autoinhibitory interactions (Evangelista *et al.*, 1997; Alberts, 2001; Westendorf, 2001). The identities of all other sequences as DADs are speculative.

DAD-like sequences. Are these formins regulated by auto-inhibition and, if so, what sequences mediate these interactions? Alternately, other mechanisms might serve to regulate these formins, such as binding to inhibitory proteins.

The predicted coiled-coil regions N-terminal to the FH1 domains of many formins are currently of unknown func-

tion. Our biochemical studies on mouse mDia1 suggest that its predicted coiled-coil region mediates multimerization (unpublished observations). If this function is similar for other formins, then these proteins could multimerize by two distinct interactions: coiled-coil and FH2 domain.

The main goal of our phylogenetic analysis is to provide a unified classification system for the formin family. In addition, we suspect that the relationships we have uncovered will provide the substrate for new experimental test of function for individual formins.

ACKNOWLEDGMENTS

H.N.H. was supported by National Institutes of Health grant GM-069818, grant P20RR16437 from the Centers of Biomedical Research Excellence Program of the National Center for Research Resources, and a Pew Scholars Award. K.J.P. was supported by the National Science Foundation, NASA Ames, and Dartmouth College.

REFERENCES

Alberts, A. S. (2001). Identification of a carboxyl-terminal diaphanous-related formin homology protein autoregulatory domain. *J. Biol. Chem.* 276, 2824–2830.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280.

Berg, J. S., Powell, B. C., and Cheney, R. E. (2001). A millennial myosin census. *Mol. Biol. Cell* 12, 780–794.

Castrillon, D. H., and Wasserman, S. A. (1994). Diaphanous is required for cytokinesis in *Drosophila* and shares domains of similarity with the products of the limb deformity gene. *Development* 120, 3367–3377.

Copley, R. R., Aloy, P., Russell, R. B., and Telford, M. J. (2004). Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of *Caenorhabditis elegans*. *Evol. Dev.* 6, 164–169.

Cvrckova, F., Novotny, M., Pickova, D., and Zarsky, V. (2004). Formin homology 2 domains occur in multipe contexts in angiosperms. *B.M. C. Genomics* 5, 44.

Darwin, C. (1859). *On the Origin of Species*. John Murray (republished by Harvard University Press, 1964): London.

Emmons, S., Phan, H., Calley, J., Chen, W., James, B., and Manseau, L. (1995). Cappuccino, a *Drosophila* maternal effect gene required for polarity of the egg and embryo, is related to the vertebrate limb deformity locus. *Genes Dev.* 9, 2482–2494.

Evangelista, M., Blundell, K., Longtine, M. S., Chow, C. J., Adames, N., Pringle, J. R., Peter, M., and Boone, C. (1997). Bni1p, a yeast formin linking cdc42p and the actin cytoskeleton during polarized morphogenesis. *Science* 276, 118–122.

Goodson, H. V., and Hawse, W. F. (2002). Molecular evolution of the actin family. *J. Cell Sci.* 115, 2619–2622.



Figure 6. N-terminal sequence similarities between fungal formins. (A) Region N3, similar between all yeast formins and the Basidiomycete Um2. Letters in bold denote sequence that is strongly similar within the Bni1 group and with Um2. (B) Region of similarity between Bnr1p, Ag1, and Ag3. This region is not found in other fungal formins.

- Harris, E. S., Li, F., and Higgs, H. N. (2004). The mouse formin, FRLa, slows actin filament barbed end elongation, competes with capping protein, accelerates polymerization from monomers, and severs filaments. *J. Biol. Chem.* 279, 20076–20087.
- Hedges, S. B. (2004). The origin and evolution of model organisms. *Nat. Rev. Genet.* 3, 838–849.
- Hodge, T., and Cope, J.T.V. (2000). A myosin family tree. *J. Cell Sci.* 113, 3353–3354.
- Kovar, D. R., Kuhn, J. R., Tichy, A. L., and Pollard, T. D. (2003). The fission yeast cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. *J. Cell Biol.* 161, 875–887.
- Li, F., and Higgs, H. N. (2003). The mouse formin mDia1 is a potent actin nucleation factor regulated by autoinhibition. *Curr. Biol.* 13, 1335–1340.
- Lupas, Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162–1164.
- Miyagi, Y., *et al.* (2002). Delphilin: a novel PDZ and formin homology domain-containing protein that synaptically colocalizes and interacts with glutamate receptor d2 subunit. *J. Neurosci.* 22, 803–814.
- Moseley, J. B., Sagot, I., Manning, A. L., Xu, Y., Eck, M. J., Pellman, D., and Goode, B. L. (2004). A conserved mechanism for Bni1- and mDia1-induced actin assembly and dual regulation of Bni1 by Bud6 and profilin. *Mol. Biol. Cell* 15, 896–907.
- Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* 30, 371–403.
- Page, R.D.M., and Holmes, E. C. (1998). *Molecular Evolution: A Phylogenetic Approach*, Blackwell Science Ltd.: Oxford.
- Petersen, J., Nielsen, O., Egel, R., and Hagan, I. M. (1998). FH3, a domain found in formins, targets the fission yeast formin Fus1 to the projection tip during conjugation. *J. Cell Biol.* 141, 1217–1228.
- Pruyne, D., Evangelista, M., Yang, C., Bi, E., Zigmond, S., Bretscher, A., and Boone, C. (2002). Role of formins in actin assembly: nucleation and barbed-end association. *Science* 297, 612–615.
- Sagot, I., Rodal, A. A., Moseley, J., Goode, B. L., and Pellman, D. (2002). An actin nucleation mechanism mediated by Bni1 and profilin. *Nat. Cell Biol.* 8, 626–631.
- Sellers, J. R. (2000). Myosins: a diverse superfamily. *Biochim. Biophys. Acta* 1496, 3–22.
- Shimada, A., Nyitrai, M., Vetter, I. R., Kuhlmann, D., Bugyi, B., Narumiya, S., Geeves, M. A., and Wittinghofer, A. (2004). The core FH2 domain of the Diaphanous-related formins is an elongated actin binding protein that inhibits polymerization. *Mol. Cell* 13, 511–522.
- Swofford, D. L. (2002). PAUP* phylogenetic analysis using parsimony (* and other methods) v. 4.0b10 for Macintosh, Sunderland, MA: Sinauer Associates, Inc.
- Taylor, J. W., Spatafora, J., O'Donnell, K., Lutzoni, F., James, T., Hibbert, D. S., Geiser, D., Bruns, T. D., and Blackwell, M. (2004). The fungi. In: *Assembling the Tree of Life*, ed. J.C.a.M.J. Donoghue, New York: Oxford University Press, 171–194.
- Waller, B. J., and Alberts, A. S. (2003). The formins: active scaffolds that remodel the cytoskeleton. *Trends Cell Biol.* 13, 435–446.
- Wasserman, S. (1998). FH proteins as cytoskeletal organizers. *Trends Cell Biol.* 8, 111–115.
- Westendorf, J. J. (2001). The dormin/diaphanous-related protein, FHOS, interacts with Rac1 and activates transcription from the serum response element. *J. Biol. Chem.* 276, 46453–46459.
- Westendorf, J. J., Mernaugh, R., and Hiebert, S. W. (1999). Identification and characterization of a protein containing formin homology (FH1/FH2) domains. *Gene* 232, 173–182.
- Woychik, R. P., Maas, R. L., Zeller, R., and Leder, P. (1990). 'Formins': proteins deduced from the alternative transcripts of the *limb deformity* gene. *Nature* 346, 850–853.
- Xu, Y., Moseley, J., Sagot, I., Poy, F., Pellman, D., Goode, B. L., and Eck, M. J. (2004). Crystal Structures of a formin homology-2 domain reveal a tethered dimer architecture. *Cell* 116, 711–723.
- Zeller, R., Haramis, A. G., Zuniga, A., McGuigan, C., Dono, R., Davidson, G., Chabanis, S., and Gibson, T. (1999). Formin defines a large family of morphoregulatory genes and functions in establishment of the polarising region. *Cell Tissue Res.* 296, 85–93.
- Zigmond, S. H. (2004). Formin-induced nucleation of actin filaments. *Curr. Opin. Cell Biol.* 16, 99–105.