Plant Molecular Biology **26**: 1139–1153, 1994.

Molecular analysis of the myosin gene family in *Arabidopsis thaliana*

Mark Kinkema, Haiyang Wang and John Schiefelbein *
Department of Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48 109, USA (author for correspondence)*

Received 9 June 1994; accepted in revised form 1 September 1994

Key words: *Arabidopsis thaliana*, actin-binding protein, cytoplasmic streaming, intracellular motility, IQ motif, molecular motor, myosin genes

Abstract

Myosin is believed to act as the molecular motor for many actin-based motility processes in eukaryotes. It is becoming apparent that a single species may possess multiple myosin isoforms, and at least seven distinct classes of myosin have been identified from studies of animals, fungi, and protozoans. The complexity of the myosin heavy-chain gene family in higher plants was investigated by isolating and characterizing myosin genomic and cDNA clones from *Arabidopsis thaliana*. Six myosin-like genes were identified from three polymerase chain reaction (PCR) products (PCR1, PCR11, PCR43) and three cDNA clones (ATM2, MYA2, MYA3). Sequence comparisons of the deduced head domains suggest that these myosins are members of two major classes. Analysis of the overall structure of the ATM2 and MYA2 myosins shows that they are similar to the previously-identified ATM1 and MYA1 myosins, respectively. The MYA3 appears to possess a novel tail domain, with five IQ repeats, a six-member imperfect repeat, and a segment of unique sequence. Northern blot analyses indicate that some of the *Arabidopsis* myosin genes are preferentially expressed in different plant organs. Combined with previous studies, these results show that the *Arabidopsis* genome contains at least eight myosin-like genes representing two distinct classes.

Introduction

Plant cells exhibit a diverse array of actin-based motility processes [50]. These include cytoplasmic streaming (intracellular movement of endoplasmic reticulum along actin filaments) [10, 20, 29, 49], directed secretory vesicle transport [38, 48, 51], organelle movements [31, 34], and cell contraction [30]. Myosin is believed to act as the molecular motor responsible for the actin-based motility observed in eukaryotic cells [13, 22, 43, 56]. Biochemical and immunological studies have shown that plants possess myosin-like proteins that range in size from 100 to 205 kDa [14, 33, 41, 44, 53, 55, 57]. Evidence for a myosin-like activity in pollen tubes has been obtained through the use of actin-based motility assays [25–27, 57].

The primary structures of several forms of

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers Z34295 (PCR1), Z34296 (PCR11), Z34297 (PCR43), Z34292 (ATM2), Z34293 (MYA2), and Z34294 (MYA3).

myosin heavy chains have been determined from animals, fungi, and protozoans [28, 40, 43, 56]. All myosins identified to date possess a conserved NH_2-terminal 'head' (or motor) domain of approximately 80 kDa with ATP- and actin-binding properties. The COOH-terminal 'tail' domain, on the other hand, is variable in length and structure in different myosins. Based on the primary sequence of the head domain, Cheney *et al.* [4] have divided all known myosins into at least seven different classes. The conventional double-headed myosins initially discovered in muscle cells [56] make up the myosin-II class. The tails of these proteins are approximately 140 kDa and display a strong α-helical coiled-coil structure which leads to their dimerization and association into filaments. The other classes of myosins (generally referred to as unconventional myosins) possess structurally diverse tail domains, which may serve to define the cellular component with which each myosin interacts [43]. Within a single species, the number of myosin isoforms can be very large, due to multiple myosin-encoding genes and alternative RNA processing [40, 43, 56].

Relatively little is known about the types or structures of myosins in plants, compared to our understanding of myosins in other eukaryotes. The sequences of two cDNAs encoding myosin-like polypeptides from *Arabidopsis thaliana* have recently been reported, and these have been designated ATM1 [24] and MYA1 [23]. Sequence comparisons suggest that the products of these genes represent members of two distinct classes of myosin. The ATM1 myosin has a predicted size of 131 kDa, possesses an unusual NH_2-terminal extension and a tail segment with a unique structure, and is proposed to be a class VIII myosin [24]. The MYA1 myosin has a predicted size of 173 kDa with an overall structure similar to the class V myosins [23], which includes the yeast MYO2 [18], mouse dilute [35], and chicken p190 [7].

Because many myosin isoforms exist in other organisms, it is likely that higher-plant species possess families of myosin genes that may exhibit developmental or functional specificities. To examine the myosin gene family in *Arabidopsis*, we have identified polymerase chain reaction (PCR) and cDNA clones that represent sequences from six new myosin-like genes. Transcripts from some of these genes preferentially accumulate in specific organs of the plant. The products of these six genes appear to be new members of two major classes of plant myosins.

Materials and methods

Plant materials and growth conditions

Arabidopsis thaliana (ecotype Columbia) plants were used for all DNA and RNA extractions. DNA was isolated from the leaves of 4-week-old plants grown at 22 °C under 16 h light/8 h dark conditions. RNA was extracted from the following tissues: leaves from 3-week-old plants (prior to the bolting stage) grown at 22 °C under 16 h light/8 h dark, stems (excluding axillary leaves) from 4–6-week–old plants grown at 22 °C under 16 h light/8 h dark, flowers at all stages of development (shoot apex, flower buds, and mature flowers) from 4–6-week-old plants grown at 22 °C under 16 h light/8 h dark, and roots (ca. 2 cm long) from 5-day-old seedlings grown on agarose-solidified media under axenic conditions at 22 °C under continuous light, as previously described [47].

Cloning and sequencing of the myosin PCR clones

Degenerate oligonucleotide primers, including restriction enzyme sites, were synthesized to the amino acid sequences ERNFHIFY (5′-acactgcaGARMGNAAYTWYCAYATHTTYTA-3′) and LDIYGFE (5′-caagaaTTCRAANCCRWADATRTCNA-3′) and used in a polymerase chain reaction (PCR). *Arabidopsis* genomic DNA (0.6 μg) was mixed with 1 μg of each oligonucleotide primer and amplified with *Taq* DNA Polymerase (Perkin Elmer-Cetus) during 30 cycles with: 94 °C, 90 s for denaturing; 50 °C, 120 s for annealing; and 72 °C, 120 s for synthesis. The PCR products were excised from

a 1.0% agarose gel, purified by electroelution, digested with *Pst* I and *Eco* RI (Boehringer Mannheim Biochemicals), and cloned into a pBluescript SK+ vector (Stratagene) digested with the same restriction enzymes. The plasmid DNA inserts were subjected to dideoxy sequencing [46] using T7 and T3 promoter primers and synthetic oligonucleotides.

The cDNA library (a kind gift from J. Ecker, University of Pennsylvania) was constructed in the λZapII vector (Stratagene) from poly(A)$^+$ RNA of etiolated, 3-day-old *Arabidopsis* seedlings (ecotype Columbia) size selected for 3–6 kb RNAs. Most of the cDNAs in this library lacked poly(A) tails. A total of 400000 plaques were screened by the method of Benton and Davis [1] using the PCR-generated clones as probes. The recombinant pBluescript plasmids were excised from the lambda clones and both DNA strands were sequenced by the dideoxy method [46]. Oligonucleotide primers were synthesized (U.M. Molecular Genetics Facility) at each step for the ordered sequencing of the clones.

Sequence analyses were performed using the software package by the University of Wisconsin Genetics Computer Group [6] and the MultAlin program (Cherwell Scientific, Oxford) [5]. The program of Lupas *et al.* [32] was used to define protein domains of coiled-coil structure. Searches of the GenBank database were performed with the FASTA program [42].

Southern and northern blot analyses

After restriction enzyme digests, the DNA was separated in 0.7% agarose gels and transferred to nylon membranes, essentially as described [45]. Prehybridization and hybridization for Southern analyses at the standard (high) stringency were carried out in 7% SDS, 0.25 M sodium phosphate, 1 mM EDTA, and 1% casein at 68 °C, with the final wash in 0.1× SSC (1× SSC is 150 mM NaCl, 15 mM sodium citrate pH 7.0) and 0.1% SDS at 60 °C for 30 min. For reduced stringency Southern analysis, prehybridization and hybridization was in 6× SSPE, 0.5% SDS, 6× Denhardt's solution (1× Denhardt's solution is 0.02% PVP, 0.02% bovine serum albumin, 0.02% Ficoll), 100 μg/ml denatured DNA, and 1 mM EDTA at 55 °C, with the final wash in 0.1× SSC and 0.1% SDS at 50 °C for 20 min.

RNA was isolated by a guanidium thiocyanate procedure and used in northern blotting essentially as described [45]. Prehybridization, hybridization, and washing of membranes were carried out under the standard (high stringency) conditions described above for Southern analysis. A

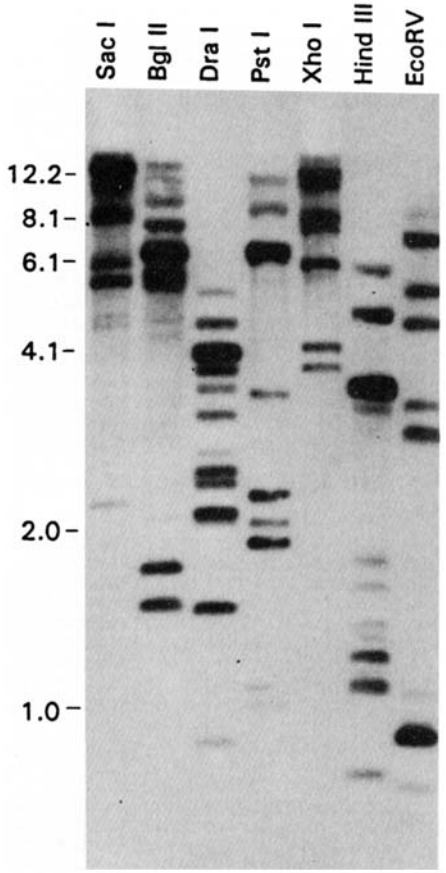
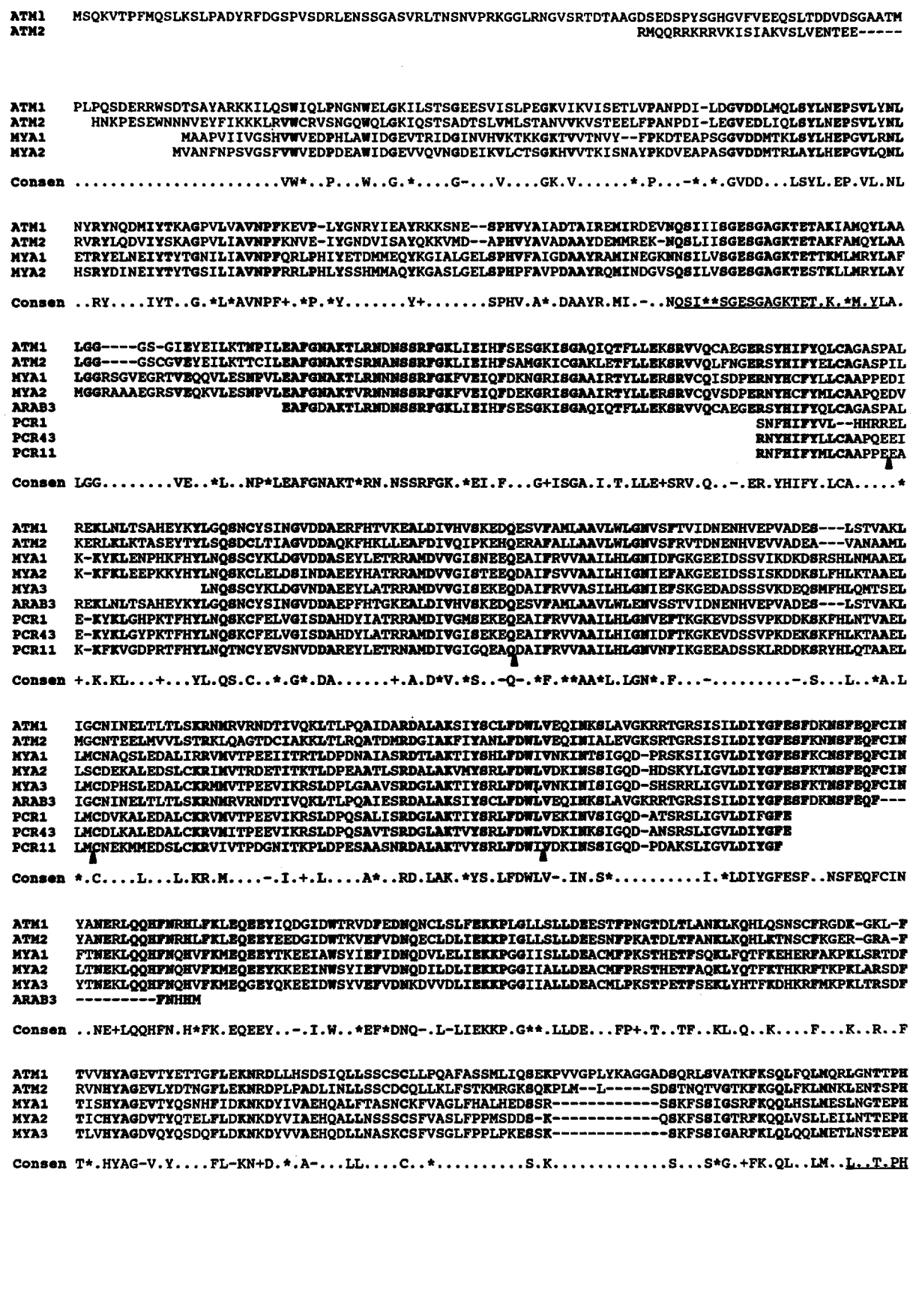


Fig. 1. Southern (DNA) blot analysis of *Arabidopsis* genomic DNA. Five μg of DNA was digested with the indicated restriction enzymes, separated in a 0.7% agarose gel, and blotted onto nylon membrane. The filter was hybridized under conditions of reduced stringency (see Materials and methods) with a random-primed ^{32}P-labeled 616 bp *Eco* RV fragment from the MYA1 cDNA clone. The fragment extends from nucleotide position 828 to 1443 [23] (corresponding to amino acids 227 to 431). Positions of the DNA molecular weight markers are indicated in kilobases.



```
ATM1    FIRCIKPNNIQSPGVYEQGLVLQQLRCCGVLEVVRISRSGFPTRMSHQKFSRRYGFLLVENIADRDPLSVSVAILHQFNILPEMYQVGYTKLF
ATM2    FIRCIKPNSKQLPRVYEEDLVLQQLRCCGVLEVVRISRSGYPTRLTHQEFAGRYGFLLSDKKVAQDPLSVSIAVLKQYDVHPEMYQVGYTKLY
MYA1    YIRCIKPNNVLKPGIFENFNVIHQLRCGGVLEAIRISCAGYPTRLAFYDFLDRFGLLAPEVLEGNYDDKVACQMILDKKSLTD-YQIGKTKIF
MYA2    YIRCIKPNNLLKPGIFENENILQQLRCGGVMEAIRISCAGYPTRKHFDEFLARFGILAPEVLVKNSDDPAACKKLLDKVGLEG-YQIGKTKVF
MYA3    YIRCVKPNNLLQPTVFDNANVLHQLRSGGVLEAIRVKCAGYPTNRTFIEFLNRFLILAPEILKGEYEAEVACKWILEKKGLTG-YQIGKSKVF

Consen  .IRCIKPNN*..P.*.E...VL.QLRC.GVLE**RIS..GYPTR....-F*.R.G.L*.E.*.......V....*.....L...YQ*G.TK*F

ATM1    FRTGQIGVLEDTRNR
ATM2    LRTGQIGIFEDRRKK
MYA1    LRAGQMAELDARRAE
MYA2    LRAGQMADLDTRRTE
MYA3    LRAGQMAELDAHRTR

Consen  LR.GQ...L-.+R..
```

Fig. 2. Comparison of the deduced amino acid sequences from the head (motor) domain of myosin gene fragments isolated from *Arabidopsis thaliana*. The sequences were aligned with the MultAlin program [5], with gaps represented by dashes lines. The consensus sequence indicates amino acids present in at least 75% of the sequences. In the consensus, * represents A, I, L, or V; – represents D or E; + represents H, K, or R; and a • represents a position with no clear consensus residue. The putative ATP-binding sequence and the actin-binding sequence are underlined in that order [43]. The positions of the four introns in the PCR1, PCR11, and PCR43 genomic sequences are indicated by triangles. Sequences: ATM1 (from [24]); ATM2, MYA2, MYA3, PCR1, PCR11, and PCR43 (this study); MYA1 (from [23]); ARAB3 (from [36]). The sequence of ARAB3, a PCR-amplified partial *Arabidopsis* myosin-like gene [36], is nearly identical to ATM1 (97.5% amino acid identity and 98.5% nucleotide identity) and may represent a different allele of ATM1; thus, it is not included in the other comparisons.

soybean 18S rRNA gene fragment (a kind gift from R. Meagher, University of Georgia) was used as a loading control probe.

Labeling of DNA fragments for use as probes in hybridization analyses was performed using a random-primed labeling kit (Boehringer Mannheim Biochemicals).

Results

Reduced-stringency hybridization with a myosin gene probe

In an effort to detect myosin-like sequences in the *Arabidopsis* genome, a Southern blot was probed

Table 1. Percent amino acid sequence identity between the head domains of myosin-like polypeptides from *Arabidopsis*. Percent similarity is in parenthesis. Data were obtained using the BESTFIT program [6]. Sequences are as in Fig. 2.

MYOSIN	ATM1	ATM2	MYA1	MYA2	MYA3	PCR1	PCR11
ATM2	67.0						
	(78.2)						
MYA1	45.0	44.7					
	(64.8)	(64.2)					
MYA2	45.6	43.6	73.4				
	(66.7)	(64.3)	(83.7)				
MYA3	41.1	40.9	72.7	72.0			
	(63.2)	(61.8)	(86.0)	(82.4)			
PCR1	43.4	38.7	65.6	66.1	74.1		
	(59.9)	(55.2)	(81.2)	(80.6)	(84.2)		
PCR11	41.0	37.2	66.8	68.4	63.0	63.2	
	(60.1)	(56.3)	(81.3)	(81.3)	(79.0)	(80.5)	
PCR43	42.4	39.9	70.7	71.8	74.7	86.0	66.3
	(62.0)	(57.4)	(83.0)	(84.6)	(83.5)	(94.1)	(84.0)

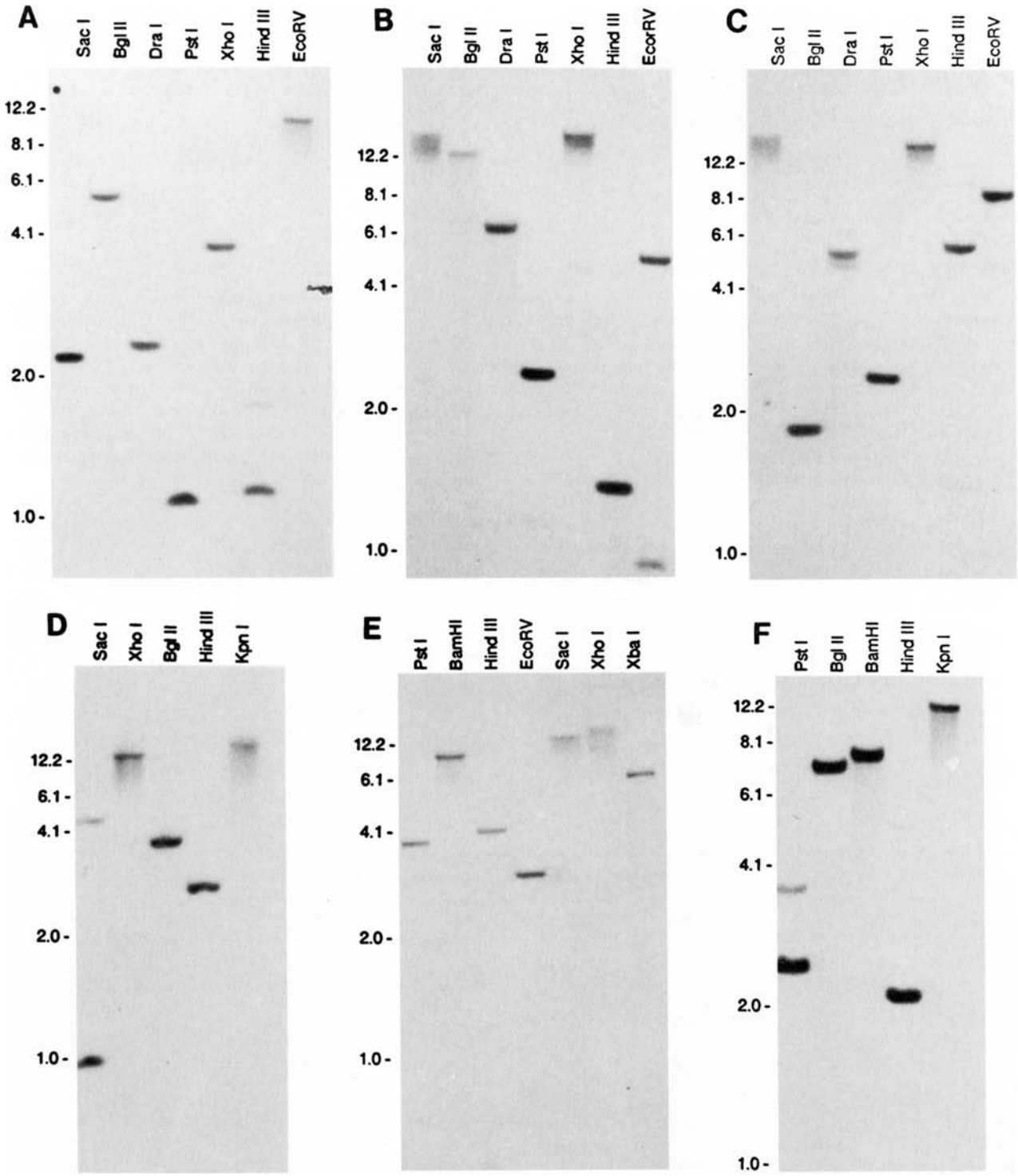


Fig. 3. Southern (DNA) blot analysis of *Arabidopsis* genomic DNA using myosin genomic (PCR) and cDNA clones. Five μg of genomic DNA was digested with the indicated restriction enzymes, separated in a 0.7% agarose gel, and blotted to nylon membrane. A. The filter was hybridized with a random-primed ^{32}P-labeled fragment from the entire PCR1 genomic clone. B. The filter was hybridized with a random-primed ^{32}P-labeled fragment from the entire PCR11 genomic clone. C. The filter was hybrid-

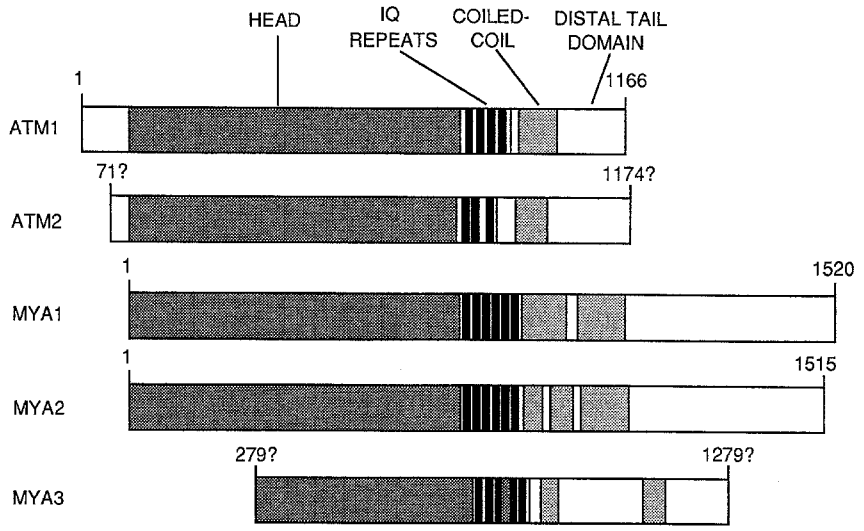


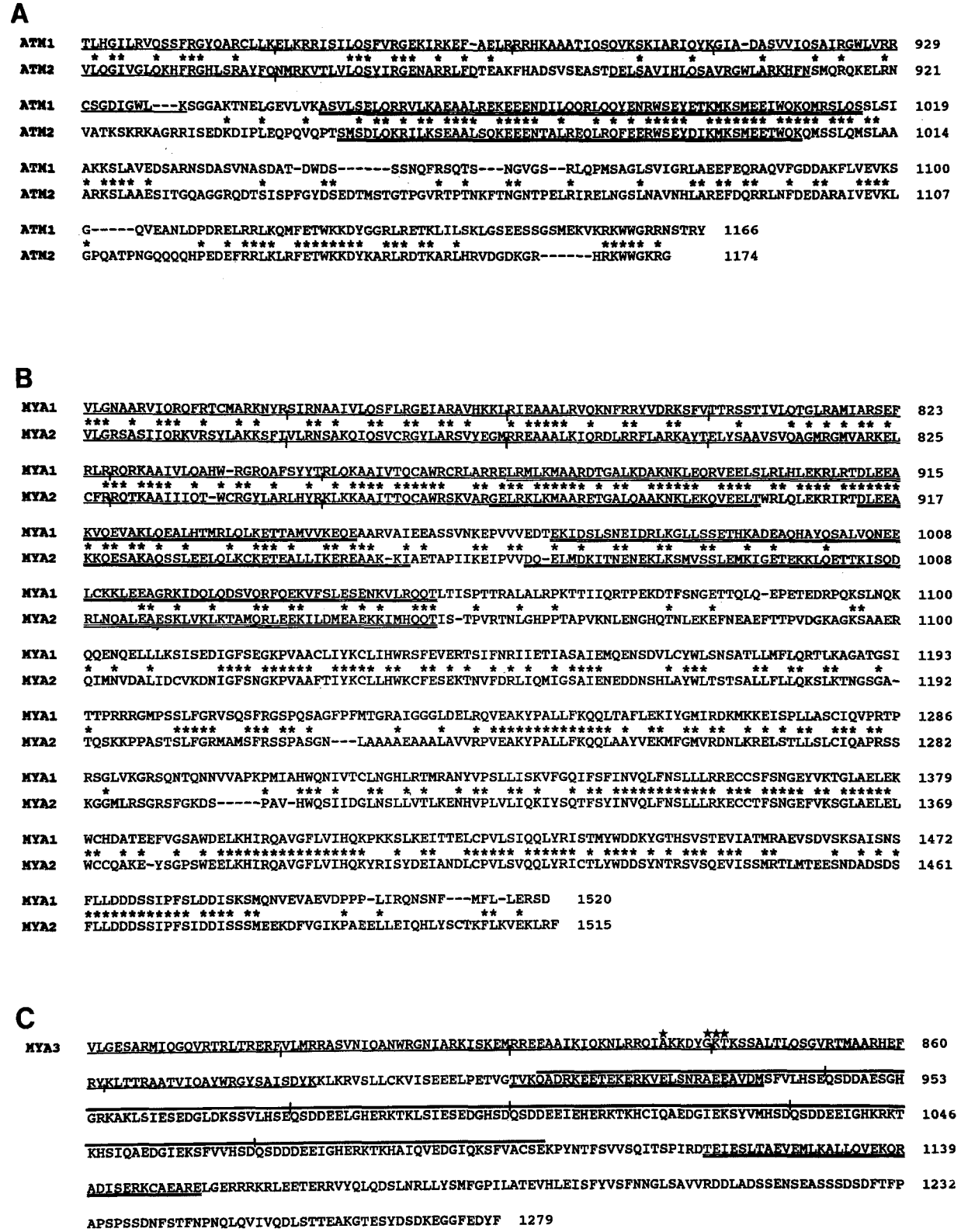
Fig. 4. Schematic comparison of the predicted domains of five myosins from *Arabidopsis thaliana*. The domains include the head (motor) domain, IQ repeats (putative calmodulin-binding sequences), coiled-coil domain, and distal tail domain. Sequences: ATM1 (131 kDa, from [24]); ATM2 (estimated 133 kDa, this study); MYA1 (173 kDa; from [23]); MYA2 (172 kDa, this study); and MYA3 (estimated 145 kDa, this study). Numbers refer to deduced amino acid sequence position. The ATM1 and ATM2 myosins possess an NH_2-terminal extension. The ATM2 and MYA3 cDNAs are truncated at the 5′ end; the numbering is based on the size of the related myosins.

at a reduced stringency (see Materials and methods) with a labeled DNA fragment corresponding to a conserved region of the myosin head. This fragment was isolated from the previously characterized *Arabidopsis* MYA1 cDNA clone [23]. Under these conditions, many hybridizing fragments were detected from each restriction enzyme digest of genomic DNA (Fig. 1). These results suggest that the *Arabidopsis* genome may contain seven or more myosin-like genes. To analyze the diversity of this large myosin gene family in greater detail, we cloned and sequenced myosin genomic and cDNA fragments.

Isolation and characterization of myosin PCR clones

A PCR was performed with *Arabidopsis* genomic DNA using degenerate oligonucleotides synthesized to two conserved amino acid motifs in the myosin head (see Materials and methods; [23]). Sequence analysis of the six major PCR products (ranging in size from 600 to 1100 bp) showed that each encode a polypeptide with some sequence similarity to known myosins. However, three of the PCR clones possess less sequence similarity and they failed to hybridize to *Arabidopsis* genomic DNA in a Southern blot analysis (data not shown); thus, these three clones were not analyzed further. The other three PCR products, designated PCR1 (888 bp), PCR11 (1067 bp), and PCR43 (888 bp), are similar to one another and to the MYA1 myosin in their deduced amino acid sequence (Fig. 2; Table 1). In addition, each of these three PCR products possesses four putative introns which vary in length but are located in identical positions (Fig. 2). On Southern blot

ized with a random-primed ^{32}P-labeled fragment from the entire PCR43 genomic clone. D. The filter was hybridized with a random-primed ^{32}P-labeled 567 bp *Pst* I/*Xhd* I fragment from the 3′ end of the ATM2 cDNA clone (corresponding to a region from amino acid 1022 to the end of the cDNA, including 126 bp of 3′-untranslated sequence). E. The filter was hybridized with a random-primed ^{32}P-labeled 789 bp *Sfu* I fragment from the MYA2 cDNA clone (corresponding to a region from amino acid 875 to 1138). F. The filter was hybridized with a random-primed ^{32}P-labeled 1148 bp *Eco* RV/*Bgl* II fragment from the MYA3 cDNA clone (corresponding to a region from amino acid 812 to 1193). Positions of the DNA molecular weight markers are indicated in kilobases.


A
ATM1 TLHGILRVQSSFRGYQARCLLKELKRRISILQSFVRGEKIRKEF-AELRRRHKAAATIQSQVKSKIARIQYKGIA-DASVVIQSAIRGWLVRR 929
ATM2 VLQGIVGLQKHFRGHLSRAYFQNMRKVTLVLQSYIRGENARRLFDTEAKFHADSVSEASTDELSAVIHLQSAVRGWLARKHFNSMQRQKELRN 921
ATM1 CSGDIGWL---KSGGAKTNELGEVLVKASVLSELQRRVLKAEAALREKEEENDILQQRLQQYENRWSEYETKMKSMEEIWQKQMRSLQSSLSI 1019
ATM2 VATKSKRKAGRRISEDKDIPLEQPQVQPTSMSDLQKRILKSEAALSQKEEENTALREQLRQFEERWSEYDIKMKSMEETWQKQMSSLQMSLAA 1014
ATM1 AKKSLAVEDSARNSDASVNASDAT-DWDS------SSNQFRSQTS---NGVGS--RLQPMSAGLSVIGRLAEEFEQRAQVFGDDAKFLVEVKS 1100
ATM2 ARKSLAAESITGQAGGRQDTSISPFGYDSEDTMSTGTPGVRTPTNKFTNGNTPELRIRELNGSLNAVNHLAREFDQRRLNFDEDARAIVEVKL 1107
ATM1 G-----QVEANLDPDRELRRLKQMFETWKKDYGGRLRETKLILSKLGSEESSGSMEKVKRKWWGRRNSTRY 1166
ATM2 GPQATPNGQQQQHPEDEFRRLKLRFETWKKDYKARLRDTKARLHRVDGDKGR------HRKWWGKRG 1174
B
MYA1 VLGNAARVIQRQFRTCMARKNYRSIRNAAIVLQSFLRGEIARAVHKKLRIEAAALRVQKNFRRYVDRKSFVTTRSSTIVLQTGLRAMIARSEF 823
MYA2 VLGRSASIIQRKVRSYLAKKSFIVLRNSAKQIQSVCRGYLARSVYEGMRREAAALKIQRDLRRFLARKAYTELYSAAVSVQAGMRGMVARKEL 825
MYA1 RLRRQRKAAIVLQAHW-RGRQAFSYYTRLQKAAIVTQCAWRCRLARRELRMLKMAARDTGALKDAKNKLEQRVEELSLRLHLEKRLRTDLEEA 915
MYA2 CFRRQTKAAIIIQT-WCRGYLARLHYRKLKKAAITTQCAWRSKVARGELRKLKMAARETGALQAAKNKLEKQVEELTWRLQLEKRIRTDLEEA 917
MYA1 KVQEVAKLQEALHTMRLQLKETTAMVVKEQEAARVAIEEASSVNKEPVVVEDTEKIDSLSNEIDRLKGLLSSETHKADEAQHAYQSALVQNEE 1008
MYA2 KKQESAKAQSSLEELQLKCKETEALLIKEREAAK-KIAETAPIIKEIPVVDQ-ELMDKITNENEKLKSMVSSLEMKIGETEKKLQETTKISQD 1008
MYA1 LCKKLEEAGRKIDQLQDSVQRFQEKVFSLESENKVLRQQTLTISPTTRALALRPKTTIIQRTPEKDTFSNGETTQLQ-EPETEDRPQKSLNQK 1100
MYA2 RLNQALEAESKLVKLKTAMQRLEEKILDMEAEKKIMHQQTIS-TPVRTNLGHPPTAPVKNLENGHQTNLEKEFNEAEFTTPVDGKAGKSAAER 1100
MYA1 QQENQELLLKSISEDIGFSEGKPVAACLIYKCLIHWRSFEVERTSIFNRIIETIASAIEMQENSDVLCYWLSNSATLLMFLQRTLKAGATGSI 1193
MYA2 QIMNVDALIDCVKDNIGFSNGKPVAAFTIYKCLLHWKCFESEKTNVFDRLIQMIGSAIENEDDNSHLAYWLTSTSALLFLLQKSLKTNGSGA- 1192
MYA1 TTPRRRGMPSSLFGRVSQSFRGSPQSAGFPFMTGRAIGGGLDELRQVEAKYPALLFKQQLTAFLEKIYGMIRDKMKKEISPLLASCIQVPRTP 1286
MYA2 TQSKKPPASTSLFGRMAMSFRSSPASGN---LAAAAEAAALAVVRPVEAKYPALLFKQQLAAYVEKMFGMVRDNLKRELSTLLSLCIQAPRSS 1282
MYA1 RSGLVKGRSQNTQNNVVAPKPMIAHWQNIVTCLNGHLRTMRANYVPSLLISKVFGQIFSFINVQLFNSLLLRRECCSFSNGEYVKTGLAELEK 1379
MYA2 KGGMLRSGRSFGKDS-----PAV-HWQSIIDGLNSLLVTLKENHVPLVLIQKIYSQTFSYINVQLFNSLLLRKECCTFSNGEFVKSGLAELEL 1369
MYA1 WCHDATEEFVGSAWDELKHIRQAVGFLVIHQKPKKSLKEITTELCPVLSIQQLYRISTMYWDDKYGTHSVSTEVIATMRAEVSDVSKSAISNS 1472
MYA2 WCCQAKE-YSGPSWEELKHIRQAVGFLVIHQKYRISYDEIANDLCPVLSVQQLYRICTLYWDDSYNTRSVSQEVISSMRTLMTEESNDADSDS 1461
MYA1 FLLDDDSSIPFSLDDISKSMQNVEVAEVDPPP-LIRQNSNF---MFL-LERSD 1520
MYA2 FLLDDDSSIPFSIDDISSSMEEKDFVGIKPAEELLEIQHLYSCTKFLKVEKLRF 1515
C
MYA3 VLGESARMIQGQVRTRLTRERFVLMRRASVNIQANWRGNIARKISKEMRREEAAIKIQKNLRRQIAKKDYGKTKSSALTLQSGVRTMAARHEF 860
RYKLTTRAATVIQAYWRGYSAISDYKKLKRVSLLCKVISEEELPETVGTVKQADRKEETEKERKVELSNRAEEAVDMSFVLHSEQSDDAESGH 953
GRKAKLSIESEDGLDKSSVLHSEQSDDEELGHERKTKLSIESEDGHSDQSDDEEIEHERKTKHCIQAEDGIEKSYVMHSDQSDDEEIGHKRKT 1046
KHSIQAEDGIEKSFVVHSDQSDDDEEIGHERKTKHAIQVEDGIQKSFVACSEKPYNTFSVVSQITSPIRDTEIESLTAEVEMLKALLQVEKQR 1139
ADISERKCAEARELGERRRKRLEETERRVYQLQDSLNRLLYSMFGPILATEVHLEISFYVSFNNGLSAVVRDDLADSSENSEASSSDSDFTFP 1232
APSPSSDNFSTFNPNQLQVIVQDLSTTEAKGTESYDSDKEGGFEDYF 1279

analyses, the PCR1, PCR11, and PCR43 clones each hybridized strongly with a single genomic fragment (with weak hybridization to one or two other fragments in some cases; Fig. 3A–C). Based on the sequence and Southern blot analyses, we conclude that PCR1, PCR11, and PCR43 represent sequences from three different *Arabidopsis* myosin genes.

Isolation and characterization of myosin cDNA clones

Using the PCR products as probes on a cDNA library from three-day-old *Arabidopsis* seedling RNA, approximately 20 cDNA clones were identified and purified. Restriction enzyme analyses indicated that these cDNA clones represent four distinct genes. The sequence of one of these cDNAs (MYA1) was recently described [23]. The longest cDNA clones representing each of the three other myosin-like genes were subsequently sequenced. Analysis of the deduced amino acid sequence of these clones suggests that one of the cDNAs (3432 bp in length) encodes a product similar to ATM1 (thus designated ATM2), while the other two cDNAs (5297 and 3218 bp) encode myosins similar to MYA1 (designated MYA2 and MYA3, respectively). The predicted sizes of the polypeptides encoded by the ATM2, MYA2, and MYA3 genes are 133, 172, and 145 kDa, respectively. Because the ATM2 and MYA3 cDNAs appear to be truncated at their 5′ ends (see below), the predicted sizes represent estimates based on the known sequences of the related ATM1 and MYA1 polypeptides.

The deduced amino acid sequence of ATM2, MYA2, and MYA3 shows that the overall structure of each of these consists of the highly conserved myosin head domain attached to a tail domain of varying size (Fig. 4). The head domains are similar to those of other known myosins, and they include the putative ATP and acting-binding regions present in all myosins (Fig. 2). Sequence comparisons indicate that the head domain of ATM2 is most similar to ATM1 (Table 1). Both ATM1 and ATM2 possess an unusual NH_2-terminal extension preceding the highly-conserved head domain (Figs 2 and 4), although the ATM2 cDNA appears to be truncated at the codon corresponding to amino acid 72 in ATM1. The MYA2 and MYA3 head domains are similar to the MYA1 head (Table 1). The MYA2 cDNA is full-length (or nearly full-length) with a 173 bp 5′ untranslated leader, while the MYA3 cDNA appears to be truncated at the codon corresponding to amino acid 279 in MYA1 (Fig. 2).

Following the highly conserved head domain, ATM2, MYA2, and MYA3 possess a series of amino acid repeats of length 22 to 26 residues known as IQ motifs (Figs 4 and 5). In previous studies, this motif has been shown to bind calmodulin [2, 3, 12, 35], and it is present in one to six copies in all known myosins [4]. The ATM2, MYA2, and MYA3 myosins possess three, six, and five copies of this repeat, respectively. The IQ motifs of MYA3 are unusual because a consensus ATP/GTP-binding site motif exists within the third and fourth IQ motifs (Fig. 5C). The presence of the IQ motifs suggests that the activity of these myosins may be regulated via Ca^{2+}/calmodulin.

The rest of the myosin tail which follows the putative calmodulin-binding repeats consists of regions that have a high probability of forming an α-helical coiled-coil [32] and distal tail domains

Fig. 5. Analysis of the deduced amino acid sequences from the COOH-terminal tail domains of *Arabidopsis* myosins ATM2, MYA2, and MYA3. The location of the IQ repeats in each sequence is indicated by a single underline. The sequences with a high probability of forming a coiled-coil are indicated by a double underline. In the comparisons, identical amino acids are indicated by an asterisk. A. Comparison of ATM1 [24] with ATM2. B. Comparison of MYA1 [23] with MYA2. C. Deduced amino acid sequence of the MYA3 tail. The region within the IQ repeats that matches the consensus ATP/GTP-binding site motif (A/GXXXXGKS/T) is indicated by stars. The six unusual repeats in MYA3 are indicated by a single overline. In each figure, numbers at the right indicate the deduced amino acid positions.

that appear to lack any major secondary structure or similarity to other (non-myosin) proteins (Fig. 4). The presence of a tail segment predicted to form a coiled-coil indicates that each of these myosins may form dimers. This portion of the ATM2 tail is most similar to that of ATM1 (45% amino acid identity); they each possess regions of similar size that are predicted to form a coiled-coil, and the distal tail domains have segments with high sequence identity (Fig. 5A). This portion of the MYA2 tail is most similar to the MYA1 tail (49.5% amino acid identity), with predicted coiled-coils of similar size and distal tail domains which share a high degree of sequence identity (Fig. 5B). The tail of MYA3 appears to be distinct from the tails of other known myosins. The predicted coiled-coil regions are relatively small (29 and 36 amino acids) and are separated by 179 amino acids (Fig. 5C). Furthermore, there is a series of six imperfect repeats of a 33 amino acid sequence in the center of the tail, consisting of the consensus sequence QSDD–EE–H–RK–K–I–EDG–S–V–HS– (Figs 5C and 6). This appears to be a unique repeated sequence, and it is not found in other myosins or other polypeptides in the sequence databases.

Genomic Southern and northern analysis

To analyze the expression of the myosin genes in *Arabidopsis*, a fragment from each clone was used as a probe in northern (RNA) blotting experiments. For the ATM2, MYA2, and MYA3 clones, the fragments were selected by their ability to detect single (gene-specific) DNA fragments in Southern blot hybridization analyses (Fig. 3D–F). In some restriction digests, two DNA fragments were detected because of a restriction site within the probe fragment, which provided further evidence that the probe is gene-specific.

The myosin gene-specific fragments were used as probes on northern blots with RNA from flowers, leaves, roots, and stems of *Arabidopsis* (Fig. 7). The ATM2 gene fragment hybridized to an RNA species (ca. 4 to 5 kb) from flower, leaf and root samples (Fig. 7A), the MYA2 gene probe recognized an RNA species (ca. 5 to 6 kb) in all four tissues (Fig. 7B), and the MYA3 gene fragment hybridized to an RNA (ca. 5 to 6 kb) from the flower and leaf samples (Fig. 7C). These results indicate that each of the three genes is expressed and transcripts from two of them preferentially accumulate in an organ-dependent manner in the *Arabidopsis* plant.

Gene-specific probes could not be obtained from the three myosin PCR clones, because these encompass a portion of the conserved head domain. Southern blotting analyses showed that each of the PCR clones hybridized preferentially to a single DNA fragment from the *Arabidopsis* genome (Fig. 3A–C). Under similar hybridization conditions, northern blotting experiments with PCR11 as a probe led to the identification of a single RNA species (ca. 6 kb) that preferentially accumulates in stem tissue (Fig. 7D). No RNA species could be detected when the PCR1 or PCR43 clones were used as probes on northern blots (data not shown).

```
          QADRKEETEKERKVELSNRAEEAVDMSFVLHSE   944
          QSDD-AESGHGRKAKLSIESEDGLDKSSVLHSE   976
          QSDD-EELGHERKTKLSIESEDG-------HSD  1001
          QSDD-EEIEHERKTKHCIQAEDGIEKSYVMHSD  1033
          QSDD-EEIGHKRKTKHSIQAEDGIEKSFVVHSD  1065
          QSDDDEEIGHERKTKHAIQVEDGIQKSFVACSE  1098

CONSENSUS QSDD.EE..H.RK.K..I..EDG...S.V.HS.
```

Fig. 6. Comparison of the amino acid sequence of the six imperfect repeats in the MYA3 tail. Gaps introduced into the repeats are indicated by a dashed line. The consensus sequence indicates amino acids present in at least five of the six repeats, with a period used to represent a residue with no clear consensus.

Discussion

In this paper, we have reported evidence for six new myosin-like genes in the higher plant *Arabidopsis thaliana*. The predicted amino acid sequence of each of these shows similarity to the highly-conserved head domain present in all myosins. When combined with results from previous studies [23, 24, 36], these data show that the *Arabidopsis* genome contains at least eight myosin-like gene sequences. It is likely that more myo-

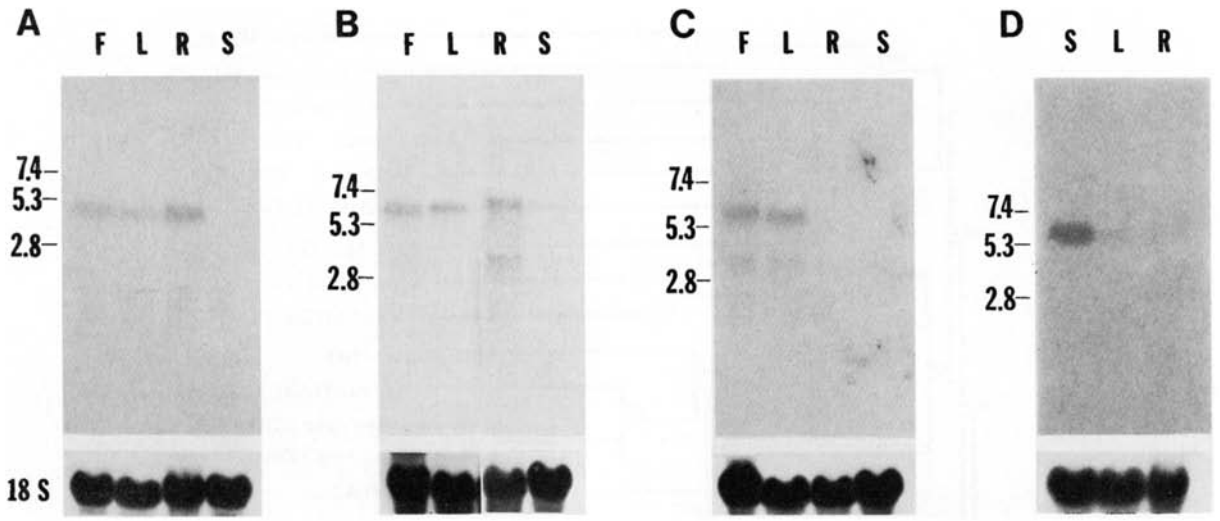


Fig. 7. Northern (RNA) blot analysis of *Arabidopsis* myosin gene fragments. Ca. 20μg of total cellular RNA was separated in a 1.0% agarose-formaldehyde gel and blotted onto nitrocellulose membrane. RNA sources: F, flowers from 4–6-week-old plants; L, leaves from 3-week-old plants; R, roots from 4-day-old seedlings; S, stems from 4–6-week-old plants. Filters were hybridized with a random-primed ^{32}P-labeled fragment from: (A) ATM2; (B) MYA2; (C) MYA3; and (D) PCR11. The probe fragments are the same as those described in legend to Fig. 2. Positions of the RNA molecular weight markers are indicated in kilobases. In some lanes, two hybridizing RNAs were detected; the smaller RNA in each lane corresponds to the position of the 28S rRNA. The bottom panels show the results of reprobing the filters with a soybean 18S rRNA gene fragment.

sin genes have yet to be identified in *Arabidopsis*, since the genes identified in these experiments were limited to those expressed in etiolated seedlings (the source of RNA for the cDNA library) or those that are able to be amplified under the conditions used in the PCR.

A classification system has recently been devised which separates all known myosins into classes based on the primary sequence of the head domain [4]. Using this system, the previously characterized *Arabidopsis* myosins, ATM1 and MYA1, are distinct from one another and do not clearly fall into any one of the previously identified classes [24] (Fig. 8). Among the six new myosins from this study, ATM2 appears to be a member of the same class as ATM1, while MYA2, MYA3, PCR1, PCR11, and PCR43 all appear to be in the MYA1 class (Table 1). A phylogenetic analysis of the *Arabidopsis* myosins with myosins from other eukaryotes shows that the ATM1/ATM2 and the MYA1/MYA2/MYA3/PCR1/PCR11/PCR43 groups appear to be distinct from the other classes of eukaryotic myosins (Fig. 8).

The only other myosin sequences reported from members of the plant kingdom are two partial sequences that were PCR-amplified from the fern *Anemia phyllitidis* [36]. From the available sequence, one of these fern myosins appears to be related to the ATM class of myosin and the other one is most similar to the MYA class (Fig. 8). This indicates that, based on the primary sequence of the head domain, all of the myosins identified in the plant kingdom to date can be placed into two major groups.

Myosins may also be classified by comparing the sequences of the tail domains. In general, the classification of myosins based on the primary structure of the tail domain is in agreement with that based on the head domain sequence [4]. For example, all myosins-II (conventional myosins) closely resemble one another in the head domain, and they all have a rod-like α-helical coiled-coil tail of ca. 140 kDa [4, 56]. Likewise, in the present study, we find that the ATM1 and ATM2 myosins of *Arabidopsis* are more similar to one another than to any other myosins in both their head and tail domains (Table 1 and Figs 5A and 8). However, we have identified two inconsistencies

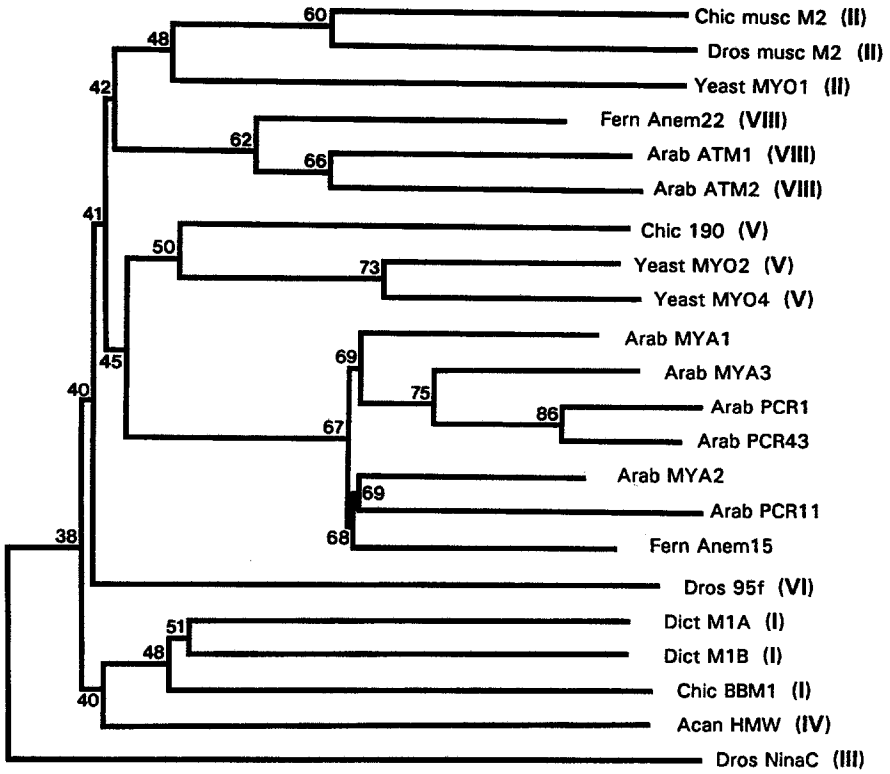


Fig. 8. Phylogenetic analysis of myosins using amino acid sequences of head domains. The analysis was performed with Clustal V [15], correcting for multiple substitutions. All of the plant myosins sequenced to date were included in this comparison as well as representative myosins from other species. The horizontal branch lengths are proportional to the differences in head-domain amino acid sequence. The roman numerals in parenthesis indicate the myosin class each member resides in, as defined by Cheney *et al.* [4]. The numbers at the branch points represent the approximate amino acid sequence identity between different members. In these analyses, myosin sequences consistently clustered together into the major groups shown here (*e.g.* class I, II, V, VIII, and the MYA1 group); however, within a major group (*e.g.* the MYA1 group) the branching pattern often varied between trials. Sequences: Chic musc M2, avian muscle myosin [37]; Dros musc M2, *Drosophila* muscle myosin [9]; Yeast MYO1, *Saccharomyces* MYO1 myosin [52]; Fern Anem22, *Anemia* PCR myosin clone [36]; Arab ATM1, *Arabidopsis* ATM1 myosin [24]; Arab ATM2, *Arabidopsis* ATM2 myosin (this study); Chic 190, chicken p190 myosin [7]; Yeast MYO2, *Saccharomyces* MYO2 myosin [18]; Yeast MYO4, *Saccharomyces* MYO4 myosin [11]; Arab MYA1, *Arabidopsis* MYA1 myosin [23]; Arab MYA3, *Arabidopsis* MYA3 myosin (this study); Arab PCR1, *Arabidopsis* PCR1 myosin (this study); Arab PCR43, *Arabidopsis* PCR43 myosin (this study); Arab MYA2, *Arabidopsis* MYA2 myosin (this study); Arab PCR11, *Arabidopsis* PCR11 myosin (this study); Fern Anem15, *Anemia* PCR myosin clone [36]; Dros 95f, *Drosophila* 95f myosin [21]; Dict M1A, *Dictyostelium* myosin-IA [54]; Dict M1B, *Dictyostelium* myosin-IB [19]; Chic BBM1, avian brush border myosin I [8]; Acan HMW, *Acanthamoeba* high-molecular-weight myosin I [16]; Dros NinaC, *Drosophila* ninaC protein [39].

in the head vs. tail classification of some of the *Arabidopsis* myosins. Based on their tail sequences, the MYA1 and MAY2 myosins clearly appear to be members of the myosin-V class, since they (like all myosins-V) possess six IQ motifs, followed by a coiled-coil region, and a distal tail domain which is most similar to a protein reported to be a glutamate decarboxylase [17, 23] (Fig. 5B). However, their head domains are not clearly similar to the class V heads, nor to any of the other non-plant myosins (Fig. 8). A second inconsistency in the classification systems involves the MYA3 sequence. Although the head domain of MYA3 is similar to MYA1 and MYA2 (Table 1 and Fig. 8), the structure of the MYA3 tail is significantly different from the tails of the MYA1 and MYA2 myosins (Fig. 4). The MYA3 tail is smaller, has only five IQ motifs (including

a putative second ATP-binding site), contains an unusual imperfect repeat sequence, and displays no apparent sequence similarity to any other myosins (Fig. 5C). The inconsistencies in the classification of these myosins using head vs. tail sequences may be the result of evolutionary constraints on the myosin molecule in plants.

The specific function of each of the members of the *Arabidopsis* myosin family is not clear. In principle, a given family member may: (1) have a unique cellular function, (2) have a redundant function, or (3) represent a pseudogene. It is likely that many functionally distinct *Arabidopsis* myosins exist, since many different types of actin-based intracellular motility are known to occur in plant cells. The myosin genes identified here do not display strong sequence identity, and some of them appear to be preferentially expressed in different organs (Fig. 7), so they may encode functionally distinct products. On the other hand, it is possible that some of these myosins are functionally equivalent. For example, the MYA1 and MYA2 are similar in overall structure (Fig. 4), and they may possess overlapping (or equivalent) functions. A possible function for these myosins is vesicle/organelle transport, since the structurally similar myosins-V have been implicated in these processes [18, 35]. Finally, it is possible that two of the myosin gene sequences identified here (PCR1 and PCR43) are not expressed, since transcripts were not detected from these in northern analysis. However, the absence of in-frame stop codons in these sequences argues against the notion that these represent pseudogenes.

An important future goal will be to assign specific functions to the myosin genes identified in this study. *Arabidopsis* should prove to be a useful organism for sorting out these structure/function relationships because of the ease with which molecular genetic studies can be carried out in this plant.

Acknowledgements

We thank Xiaopei Huang, Carrie Begin, and Sin-Hang Chan for their contributions to this research, and Richard Cheney for assistance with the phylogenetic analysis. We also thank Alex Knight, Susan Brown, Susan Lillie, and James Masucci for helpful discussions and critical comments. We acknowledge the contribution of the size-selected *Arabidopsis* cDNA library by Joe Ecker. This work was supported by U.S. Department of Energy Research Grant DE-FG02-92ER20068, National Institutes of Health grant M01 RR00042, and the University of Michigan Rackham Graduate School.

References

1. Benton WD, Davis RW: Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. Science 196: 180–182 (1977).
2. Chapman ER, Au D, Alexander KA, Nicholson TA, Storm DR: Characterization of the calmodulin binding domain of neuromodulin. J Biol Chem 266: 207–213 (1991).
3. Cheney RE, Mooseker MS: Unconventional myosins. Curr Opin Cell Biol 4: 27–35 (1992).
4. Cheney RE, Riley MA, Mooseker MS: A phylogenetic analysis of the myosin superfamily. Cell Motil Cytoskel 24: 215–223 (1993).
5. Corpet F: Multiple sequence alignment with hierarchical clustering. Nucl Acids Res 16: 10881–10890 (1988).
6. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
7. Espreafico EM, Cheney RE, Matteoli M, Nascimento AAC, De Camilli PV, Larson RE, Mooseker MS: Primary structure and cellular localization of chicken brain myosin V (p190), an unconventional myosin with calmodulin light chains. J Cell Biol 119: 1541–1557 (1992).
8. Garcia AE, Coudrier E, Carboni J, Anderson J, Vandekerckhove MS: Partial deduced sequence of the 110-kDa calmodulin complex of the avian intestinal microvillus shows that this mechanoenzyme is a member of the myosin-I family. J Cell Biol 109: 2895–2903 (1989).
9. George EL, Ober MB, Emerson CP: Functional domains of the *Drosophila melanogaster* muscle myosin heavy-chain gene are encoded by alternatively spliced exons. Mol Cell Biol 9: 2957–2974 (1989).
10. Grolig F, Williamson RE, Parke J, Miller C, Anderton BH: Myosin and Ca^{2+}-sensitive streaming in the alga *Chara*: detection of two polypeptides reacting with a monoclonal anti-myosin and their localization in the streaming endoplasm. Eur J Cell Biol 47: 22–31 (1988).
11. Haarer BK, Petzold A, Lillie SH, Brown SS: Identifica-

tion of *MYO4*, a second class V myosin gene in yeast. J Cell Sci 107: 1055–1064 (1994).
12. Halsall DJ, Hammer JA: A second isoform of chicken brush border myosin I contains a 29 residue inserted sequence that binds to calmodulin. FEBS Lett 267: 126–130 (1990).
13. Harrington WF, Rodgers ME: Myosin. Annu Rev Biochem 53: 35–73 (1984).
14. Heslop-Harrison J, Heslop-Harrison Y: Myosin associated with the surface of organelles, vegetative nuclei and generative cells in angiosperm pollen grains and tubes. J Cell Sci 94: 319–325 (1989).
15. Higgins DG, Bleasby AJ, Fuchs R: Clustal V, improved software for multiple sequence alignment. Comput Appl Biosci 8: 189–191 (1992).
16. Horowitz JA, Hammer JA: A new *Acanthamoeba* myosin heavy chain. J Biol Chem 265: 20646–20652 (1990).
17. Huang WM, Reed-Fourquet L, Wu E, Wu JY: Molecular cloning and amino acid sequence of brain I-glutamate decarboxylase. Proc Natl Acad Sci USA 87: 8491–8495 (1990).
18. Johnston GC, Prendergast JA, Singer RA: The *Saccharomyces cerevisiae MYO2* gene encodes an essential myosin for vectorial transport of vesicles. J Cell Biol 113: 539–551 (1991).
19. Jung G, Saxe CL, Kimmel AR, Hammer JA: *Dictyostelium discoideum* contains a gene encoding a myosin-I heavy chain. Proc Natl Acad Sci USA 86: 6186–6190 (1989).
20. Kachar B, Reese TS: The mechanism of cytoplasmic streaming in characean algal cells: sliding of endoplasmic reticulum along actin filaments. J Cell Biol 106: 1545–1552 (1988).
21. Kellerman KA, Miller KG: An unconventional myosin heavy chain gene from *Drosophila melanogaster*. J Cell Biol 119: 823–834 (1992).
22. Kiehart DP: Molecular genetic dissection of myosin heavy chain function. Cell 60: 347–350 (1990).
23. Kinkema M, Schiefelbein J: A myosin from a higher plant has structural similarities to class V myosins. J Mol Biol 239: 591–597 (1994).
24. Knight A, Kendrick-Jones J: A myosin-like protein from a higher plant. J Mol Biol 231: 148–154 (1993).
25. Kohno T, Shimmen T: Accelerated sliding of pollen tube organelles along Characean actin bundles regulated by Ca^{2+}. J Cell Biol 106: 1539–1543 (1988).
26. Kohno T, Chaen S, Shimmen T: Characterization of the translocator associated with pollen tube organelles. Protoplasma 154: 179–183 (1990).
27. Kohno T, Okagaki T, Kohama K, Shimmen T: Pollen tube extract supports the movement of actin filaments *in vitro*. Protoplasma 161: 75–77 (1991).
28. Korn ED, Hammer JA: Myosins of nonmuscle cells. Annu Rev Biophys Biophys Chem 17: 23–45 (1988).
29. Kuroda K: Cytoplasmic streaming in plant cells. Int Rev Cytol 121: 267–307 (1990).
30. La Claire JW: Actin cytoskeleton in intact and wounded coenocytic green algae. Planta 177: 47–57 (1989).
31. Lloyd CW, Pearce KJ, Ridge RW, Rawlins DJ, Shaw PJ: Microtubules are involved in the coupled migration of the nucleus with the tip of legume root hairs, but F-actin is necessary for basipetal migration. Cell Motil Cytoskel 8: 27–36 (1987).
32. Lupas A, VanDyke M, Stock J: Predicting coiled coils from protein sequences. Science 252: 1162–1164 (1991).
33. Ma Y, Yen L: Actin and myosin in pea tendrils. Plant Physiol 89: 586–589 (1989).
34. Menzel D, Schliwa M: Motility in the siphonous green alga *Bryopsis* II. Chloroplast movement requires organized arrays of both microtubules and actin filaments. Eur J Cell Biol 40: 286–295 (1986).
35. Mercer JA, Seperack PK, Strobel MC, Copeland NG, Jenkins NA: Novel myosin heavy chain encoded by murine dilute coat colour locus. Nature 349: 709–713 (1991).
36. Moepps B, Conrad S, Schraudolf H: PCR-dependent amplification and sequence characterization of partial cDNAs encoding myosin-like proteins in *Anemia phyllitidis* (L.) Sw. and *Arabidopsis thaliana* (L.) Heynh. Plant Mol Biol 21: 1077–1083 (1993).
37. Molina MI, Kropp KE, Gulick J, Robbins J: The sequence of an embryonic myosin heavy chain gene and isolation of its corresponding cDNA. J Biol Chem 262: 6478–6488 (1987).
38. Mollenhauer HH, Moore DJ: Cytochalasin B, but not colchicine, inhibits migration of secretory vesicles in root tips of maize. Protoplasma 87: 39–48 (1976).
39. Montell C, Rubin G: The *Drosophila ninaC* locus encodes two photoreceptor cell specific proteins with domains homologous to protein kinases and the myosin heavy chain head. Cell 52: 757–762 (1988).
40. Mooseker M: A multitude of myosins. Curr Biol 3: 245–248 (1993).
41. Parke J, Miller C, Anderton BH: Higher plant myosin heavy-chain identified using a monoclonal antibody. Eur J Cell Biol 41: 9–13 (1986).
42. Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448 (1988).
43. Pollard T, Doberstein S, Zot H: Myosin-I. Annu Rev Physiol 53: 653–681 (1991).
44. Qiao L, Grolig F, Jablonsky PP, Williamson RE: Myosin heavy chains: detection by immunoblotting in higher plants and localization by immunofluorescence in the alga *Chara*. Cell Biol Int Rep 13, 107–117 (1989).
45. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
46. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
47. Schiefelbein JW, Somerville C: Genetic control of root

hair development in *Arabidopsis thaliana*. Plant Cell 2: 235–243 (1990).
48. Schnepf E: Cellular polarity. Annu Rev Plant Physiol 37: 23–47 (1986).
49. Seagull RW, Heath IB: The differential effects of cytochalasin B on microfilament populations and cytoplasmic streaming. Protoplasma 103: 231–240 (1980).
50. Staiger CJ, Schliwa M: Actin localization and function in higher plants. Protoplasma 141: 1–12 (1987).
51. Steer MW: Role of actin in tip growth. In: Heath IB (ed) Tip Growth in Plant and Fungal Cells, pp. 119–145. Academic Press, San Diego (1991).
52. Sweeney FP, Watts FZ, Pocklington MJ, Orr E: The MYO1 gene from *Saccharomyces cerevisiae*: its complete nucleotide sequence. Nucl Acids Res 18: 7147 (1990).
53. Tang X, Hepler PK, Scordilis SP: Immunochemical and immunocytochemical identification of a myosin chain polypeptide in *Nicotiana* pollen tubes. J Cell Sci 92: 569–574 (1989).
54. Titus MA, Warrick HM, Spudich JA: Multiple actin-based motor genes in *Dictyostelium*. Cell Regul 1: 55–63 (1989).
55. Vahey M, Titus M, Trautwein R, Scordilis S: Tomato actin and myosin: contractile proteins from a higher land plant. Cell Motil 2: 131–148 (1982).
56. Warrick HM, Spudich J: Myosin structure and function in cell motility. Annu Rev Cell Biol 3: 379–421 (1987).
57. Yokota E, Shimmen T: Isolation and characterization of plant myosin from pollen tubes of lily. Protoplasma 177: 153–162 (1994).