

Phylogenetic Analysis of New Plant Myosin Sequences

Magdalena Bezanilla, Amy C. Horton, Heather C. Sevenser, Ralph S. Quatrano

Department of Biology, Campus Box 1229, Washington University, St. Louis, MO 63130, USA

Received: 12 July 2002 / Accepted: 11 March 2003

Abstract. We have sampled a large number of plant taxa, ranging from brown algae to angiosperms, for the presence of myosin sequences. Using phylogenetic analysis, we show that all but two of the new plant myosin sequences fall into two of three preexisting myosin classes. We identified two outlying sequences, which do not fall into any preexisting myosin class. Additionally, all genomic sequences encoding class XI myosins contain an intron in the region studied, suggesting that this genomic region has been conserved over at least 1 billion years of plant evolution. With these data, we can rapidly and consistently classify partial myosin sequences from plants. Our data show that plant myosins do not have clear orthologues in other kingdoms, providing interesting insights into the diversification of myosins.

Key words: Myosin — Plants — Phylogeny — Conserved intron — Degenerate PCR

Introduction

Motor molecules are highly conserved proteins, which evolve slowly over time. Myosins are a superfamily of molecular motors, which in most cases use the energy from ATP hydrolysis to generate movement along actin filaments. Myosins are believed to be expressed in all eukaryotes. Most myosin molecules have two major regions: the N-terminal actin-

binding catalytic domain and a C-terminal tail domain with a variety of functions. Here we study the phylogenetic relationship of myosin sequences from a variety of different plant species.

Myosins are typically classified based on phylogenetic (neighbor-joining or maximum parsimony) analyses of the catalytic domain of the protein, which in most cases involves the first 800 amino acids. Classes traditionally have been defined by deep nodes with a 90% bootstrap support (Sellers 2000; Berg et al. 2001; Liu et al. 2001; Reddy and Day 2001). Phylogenetic analyses of the entire molecular sequence, or of the tail sequence alone, agree with the catalytic domain trees (Soldati et al. 1999). With the large-scale sequencing of various genomes, including representatives from the fungal, animal, and plant kingdoms, we now have a representative sampling across taxa sufficient to determine with less error how many classes of myosins exist.

Recent phylogenetic analysis (Hodge and Cope 2000) demonstrates that there are at least 17 myosin classes. Interestingly, of these 17 classes, 3 (classes VIII, XI, and XIII) contain sequences from a few members of the plant kingdom (Hodge and Cope 2000; Reddy and Day 2001, No. 78; Berg et al. 2001; Liu et al. 2001; Vugrek and Moepps 2002). This is surprising because it was hypothesized that all eukaryotic lineages possessed a similar basic complement of myosins, as is the case for kinesins, the microtubule-based molecular motor (Kim and Endow 2000). Kinesins from plants are found within the same class as animal and fungal kinesins. The three most well-studied myosin classes (I, II, and V) have representatives from yeast, protozoa, and animals;

however, no plant members of these families have been found. Cases of endemic myosin families within taxa (e.g., class XIV limited to apicomplexans) exist despite genome sequencing initiatives. Our findings strongly promote the hypothesis that unique myosin genes have evolved within the plant kingdom.

The majority of plant myosin sequences available to date have come from only a few plants, including several species of green algae, *Arabidopsis thaliana* (thale cress), *Helianthus annuus* (sunflower), and *Zea mays* (maize). This is a very small taxonomic sample in comparison to the number of species of animals and fungi which have been used to analyze myosin phylogeny. In particular, of the 26 plant sequences used, 16 are from the *Arabidopsis thaliana* genome. Therefore, these data may be biased by the few species represented in the phylogenetic analysis to date. To determine if plant myosins indeed comprise evolutionarily distinct classes, we have identified partial myosin sequences from a broad spectrum of plant species.

We used a PCR approach to amplify a region of the catalytic domain, which is highly conserved among all myosin classes. We sequenced the PCR products and performed phylogenetic analyses with these products and with previously identified plant myosin sequences. We found a few myosin sequences which do not appear to belong to any of the three plant myosin classes. Most surprisingly, this small region of sequence (approximately 50 amino acids) produced classifications consistent with previous phylogenies based on more extensive sequence information (approximately 800 amino acids), indicating that this region is particularly useful for cladistic analysis of plant myosins (Liu et al. 2001; Reddy and Day 2001; Vugrek and Moepps 2002). We also discovered a conserved intron in the region analyzed which serves as an independent diagnostic for class XI myosins. Thus, we are able to obtain and to classify new plant myosin sequences rapidly and we discuss the implications of these new sequences for plant myosin evolution.

Materials and Methods

PCR Methods and Subcloning

We used degenerate primers designed to amplify myosins (Bement et al. 1994a, b). The primer sequences are as follows: forward — GGI GAR WSI GGI GCI GGI AAR AC; and reverse — GT YTT IGC RTT ICC RAA IGC YTC. We used the Nucleon Phytopure Plant and Fungal DNA Extraction Kit to purify genomic DNA from 1 g of *Physcomitrella patens* protonemal tissue following the manufacturer's recommendation (Amersham Life Science, Buckinghamshire, England). We obtained *Hordeum vulgare* and *Zea mays* genomic DNA from Dr. Tuan-Hua David Ho's laboratory. We obtained *Oryza sativa* genomic DNA (TaiPei 309) from Dr. Roger Beachy's laboratory. *Adiantum tenerum*, *Blechnum glandu-*

losum, and *Cycadus rumphii* genomic DNAs were isolated by a modified CTAB protocol (Doyle and Doyle 1987), as was DNA from all other species provided by Dr. Elizabeth A. Kellogg. We used at least 25 ng of genomic DNA as template, a 0.2 μ M concentration of each primer, 1.5 mM MgCl₂, a 0.2 mM concentration of each dNTP, and 1 μ l of Elongase enzyme mix (Invitrogen, Carlsbad, CA) in a 50- μ l reaction volume. We performed two PCR reactions. The first PCR reaction used genomic DNA as template and the second used 1 μ l of the previous PCR reaction as template. The PCR protocol had an initial denaturation step at 94°C for 1 min. During cycling, the denaturation steps were 30 s at 94°C and the extension steps at 68°C. The extension times varied between 30 s and 2 min. The annealing times were 45 s and the temperatures increased from 35 to 40°C in the following manner: first 5 cycles, 35°C; next 5 cycles, 37.5°C; and last 25 cycles, 40°C. A final 10-min elongation step at 68°C was included at the end of each PCR reaction.

PCR reactions were separated on agarose gels after the first and second PCR protocols. When using genomic DNA as a template, the products were detected only after the second PCR reaction. However, if a previously cloned myosin gene was used as template (as in positive controls; data not shown), then the PCR product was visible after the first reaction. To subclone the PCR products we used the pGEM-T Easy system (Promega, Madison, WI). For the large PCR products (> 400 bp), we purified the band from the gel before subcloning. For the smaller products we used the PCR reaction to subclone the products directly. We followed the manufacturer's recommendations for the ligation and transformation reactions. For each species studied, we sampled up to seven independent clones. An independent clone is an individual colony resulting from the ligation of the PCR mix into the pGEM vector. Generally four colonies were isolated for each species, diagnosed by restriction enzyme digestion, and clones were subsequently chosen for sequencing (Table 1). After subcloning we sequenced the products with M13Forward and M13Reverse primers using the ABI Prism Dye Terminator Cycle Sequencing Kit following the manufacturer's instructions. We checked and edited the sequences by eye in ABI's Editview 1.01. For each species, only one of multiple identical DNA sequences was retained for phylogenetic analysis. For the sequences used in the analysis in Figs. 3 and 4, we sequenced the clone on both strands. Our methods do not exhaustively sample all myosins from each species; rather they sample a few myosins (with PCR products of < 400 bp). In addition, despite strong amplification of the smallest product (~137 bp), we did not observe any species cross-contamination of PCR products as demonstrated by the data in Table 1.

Analysis of Intron Size

All sequences obtained by degenerate PCR were compared to the nonredundant nucleotide database (at the National Center for Biotechnology Information; NCBI) using the tBlastx algorithm (Altschul et al. 1990; Gish and States 1993). The majority of PCR products obtained using this protocol (Table 1) had significant sequence similarity to myosin. For those products which did not contain a continuous open reading frame, the region lacking sequence similarity to myosins was subsequently analyzed by hand. At the end of the myosin sequence similarity in the 5' region, we found an in-frame GT universal consensus 5' splice site. Just prior to the beginning of the 3' region containing sequence similarity to myosins, we found an in-frame AT universal consensus 3' splice site. Thus we determined the length of the intron to be the sequence from GT to AT. The GT and AT described here were the only sites which resulted in a continuous open reading frame with sequence similarity to myosins. We attempted to use computer programs such as GENSCAN (Burge and Karlin 1997) to verify our results. However, the sequence was too short for these programs to func-

Table 1. Cloned PCR products

Species	Clone	Sequenced	Size (bp)	Type ^a
<i>Sphagnum capilli folium</i>	Sc 1	No		
	Sc 2	No		
	Sc 3	No		
	Sc 4	Yes	139	Myosin ^b
<i>Tetraphis pellucida</i>	Tp 5	Yes	266	Myosin XI*
	Tp 6	Yes	266	Myosin XI*
	Tp 7	Yes	276	Myosin XI
	Tp 8	Yes	137	Myosin VIII
<i>Cornocladia dodanaea</i>	Cd 9	No		
	Cd 10	Yes	266	Myosin XI
	Cd 11	No		
	Cd 12	Yes	300	Myosin XI
<i>Atrichum altercristatum</i>	Aa 1	Yes	137	Myosin VIII
	Aa 2	No		
	Aa 3	Yes	137	Myosin VIII
	Aa 4	Yes	137	Myosin VIII
<i>Diphyscium peruvianum</i>	Dp 5	Yes	137	Myosin VIII*
	Dp 6	Yes	137	Myosin VIII*
	Dp 7	No		
	Dp 8	Yes	137	Myosin VIII
<i>Encalypta procera</i>	Ep 9	Yes	137	Myosin VIII
	Ep 10	Yes	296	Myosin ^b
	Ep 11	Yes	137	Myosin VIII
	Ep 12	Yes	152	Myosin ^b
<i>Clusia</i>	Clusia 13	No		
	Clusia 14	Yes	261	Myosin XI
	Clusia 15	Yes	261	Myosin XI
	Clusia 16	Yes	354	Myosin XI
<i>Sideritis romanus</i>	Sr 17	Yes	137	Myosin VIII*
	Sr 18	No		
	Sr 19	Yes	137	Myosin VIII*
	Sr 20	Yes	137	Myosin VIII
<i>Coix lachryma-jobi</i>	Cl 21	Yes	289	Myosin XI
	Cl 22	No		
	Cl 23	Yes	137	Myosin VIII
	Cl 24	Yes	257	Myosin XI
<i>Polytrichadelphus aristatus</i>	Pa 1	Yes	137	Myosin VIII
	Pa 2	Yes	137	Myosin VIII
	Pa 3	Yes	137	Myosin VIII
	Pa 4	Yes	137	Myosin VIII
<i>Ulota hutchinsiae</i>	Uh 5	Yes	137	Myosin VIII*
	Uh 6	Yes	137	Myosin VIII
	Uh 7	Yes	137	Myosin VIII*
	Uh 8	Yes	185	Myosin ^b
	Uh 15	Yes	1001	Myosin XI
<i>Diphyscium mucronifolium</i>	Dm 9	No		
	Dm 10	Yes	137	Myosin VIII
	Dm 11	Yes	137	Myosin VIII*
	Dm 12	Yes	137	Myosin VIII*
<i>Psathyrostachys juncea</i>	Pj 13	Yes	137	Myosin VIII
	Pj 14	Yes	248	Myosin XI*
	Pj 15	Yes	305	Myosin XI
	Pj 16	Yes	248	Myosin XI*
<i>Hordeum vulgare</i>	Hv 1	No		
	Hv 2	Yes	137	Myosin VIII
	Hv 3	No		
	Hv 21	Yes	253	Myosin XI*
	Hv 22	Yes	253	Myosin XI*
	Hv 23	Yes	253	Myosin XI*
	Hv 24	Yes	253	Myosin XI*
<i>Cycadus rumphii</i>	Cr 4	No		
	Cr 5	Yes	137	Myosin VIII
	Cr 6	No		

Table 1. Continued

Species	Clone	Sequenced	Size (bp)	Type ^a
<i>Blechnum glandulosum</i>	Bg 10	No		
	Bg 11	No		
	Bg 12	No		
	Bg 25	Yes	276	Myosin XI*
	Bg 26	Yes	276	Myosin XI*
	Bg 27	Yes	276	Myosin XI*
	Bg 28	Yes	276	Myosin XI*
<i>Zea mays</i>	Zm 13	Yes	137	Myosin VIII
	Zm 14	Yes	175	Unknown
	Zm 15	No		
	Zm 33	Yes	300	Myosin XI
	Zm 34	No		
	Zm 35	No		
	Zm 36	Yes	355	Myosin XI
<i>Oryza sativa</i>	Os 19	Yes	249	Myosin XI
	Os 20	No		
	Os 21	Yes	137	Myosin VIII
	Os 37	Yes	296	Myosin XI
	Os 38	Yes	392	Myosin XI
	Os 39	Yes	298	Myosin XI
	Os 40	No		
	Adt 17	No		
<i>Adiantum tenerum</i>	Adt 18	Yes	206	Unknown
	Adt 19	Yes	198	Myosin XI
	Adt 20	No		

^a Within a species, asterisks are used to indicate clones which have the same sequence.

^b Clone contains sequence similarity to myosins but is not able to be placed in a specific family.

tion properly. The *Arabidopsis* genes are for the most part predicted open reading frames from the Arabidopsis Genome Initiative (2000). We used the length of the predicted intron from this region. For the rest of the products we measured the length of the PCR product by comparison to DNA base-pair standards. We performed this analysis only on species which consistently yielded 100% of the sequenced PCR products containing sequence similarity to myosin. To determine the length of the intron, we subtracted the size of the predicted open reading frame for myosin XI (150 bp). Since we analyzed the presence of the majority of bryophyte intron lengths using genomic PCR, we expect to have noted all intron lengths less than 1.5 kb.

Phylogenetic Analysis

Homologues in additional plant species were obtained from public databases referenced at <http://www.mrc-lmb.cam.ac.uk/myosin/trees/accession.html>. The *Fucus* sequence was obtained in the Quatrano laboratory from a cDNA clone (unpublished results). Sequences were aligned using CLUSTALW v1.8 (Thompson et al. 1994) with the Gonnet series of substitution matrices. The resulting alignment was manually adjusted. Unalignable regions and primer sequences were excluded from the analysis. Identical sequences were pruned to include only one representative from each species. Phylogenetic trees were constructed using the neighbor-joining method of Saitou and Nei (1987) as implemented in CLUSTALW v1.8 and in PAUP* (Swofford 2002), with 10,000 bootstrap replications. For the supplemental figure [<http://www.link.springer.com/link/service/journals/00239/index.htm>], we performed the analysis with 1000 bootstrap replications due to the large number of sequences used. We also used maximum parsimony with the heuristic search strategy of PAUP* (bootstrap replicates for this number of sequences exceeded our computational resources), achieving similar results. In addition, we used maximum likelihood

as implemented in TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1997). For maximum likelihood, amino acid sequences using available built-in models of protein evolution (JTT, BLOSUM62, VT, WAG) gave weak support for the class XI clade and gave weak or no support for the class VIII clade (Henikoff and Henikoff 1992; Jones et al. 1992; Goldman and Whelan 2000; Muller and Vingron 2000). However, DNA sequences aligned on the basis of the amino acid alignment did provide support using the TN and HKY models of evolution with maximum likelihood (Hasegawa et al. 1985; Tamura and Nei 1993).

Results and Discussion

To obtain new myosin sequences from an evolutionarily divergent group of plants, we amplified a region of the catalytic domain originally used to clone a large variety of myosin classes from vertebrates (Bement et al. 1994a). This approach should be easily applicable to cloning almost any myosin class. This fragment codes for approximately 50 amino acids between the peptides GESGAGK and LEAFGNAK (Fig. 1A), the length of which varies among different classes of myosins. As a template for PCR, we used genomic DNA from plant species ranging from bryophytes to angiosperms (Fig. 2A; boldface family names). We were particularly interested in myosins from the mosses, for comparison with data from *Physcomitrella patens*. Thus, we have a relatively large sample from this taxon. The PCR reactions yielded a variety of different-sized bands from

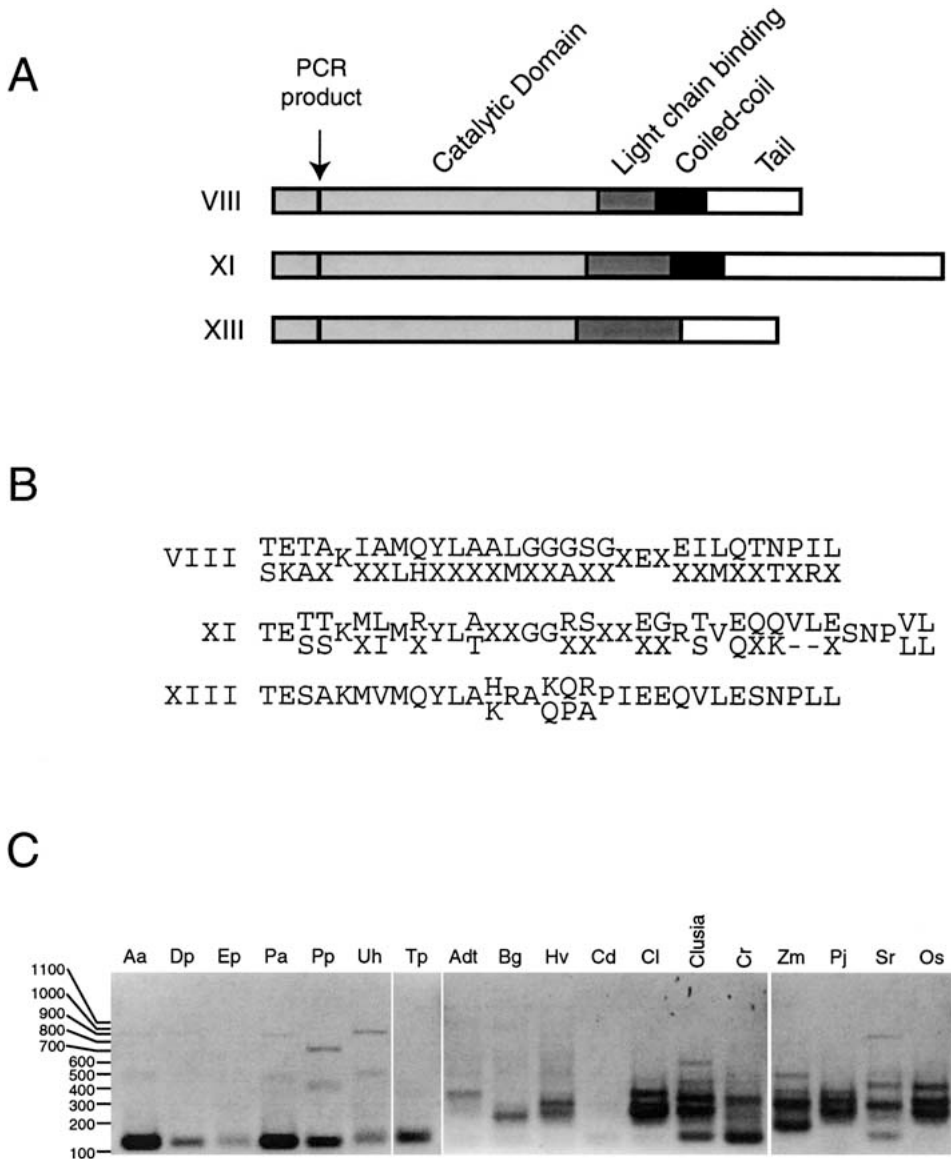


Fig. 1. **A** Schematic drawing of the three classes of plant myosins. The actin binding catalytic domain at the N terminus is denoted by a light gray box. The dark gray box denotes the light chain binding domain for each class. The number of light chain binding motifs varies from four (class VIII) to six (class XI). The black box denotes a predicted region of coiled-coil. The tail found at the C terminus is variable and denoted by the white box. The line indicated by the arrow denotes the region of the catalytic domain studied in this paper, between the highly conserved peptides GESGAGK and LEAFGNAK. **B** Consensus sequences for the region studied for each class of myosin. **C** Degenerate PCR products were separated

on a 2.5% agarose gel and stained with ethidium bromide. PCR was performed with primers described in Fig. 3. Each reaction was performed using genomic DNA from various plant species, indicated at the top of each lane: Aa — *Atrichum altercristatum*; Dp — *Diphyscium peruvianum*; Ep — *Encalypta procera*; Pa — *Polystichadelphus aristatus*; Pp — *Physcomitrella patens*; Uh — *Ulota hutchinsiae*; Tp — *Tetraphis pellucida*; Adt — *Adiantum tenerum*; Bg — *Blechnum glandulosum*; Hv — *Hordeum vulgare*; Cd — *Cornocladia dodanaea*; Cl — *Coix lachryma-jobi*; Zm — *Zea mays*; Pj — *Psathyrostachys juncea*; Sr — *Sideritis romanus*; Os — *Oryza sativa*.

each species (Fig. 1C). As we were interested in sampling a few myosins from each species, rather than obtaining all myosin sequences, we subcloned and sequenced up to seven products from each species. The majority of sequenced PCR products contained sequence similarity to myosin as described below (Table 1).

In all cases, the 137-bp product represented a continuous reading frame. However, this fragment

was not subcloned in some species sampled, due to either lack of amplification or bias in the ligation reaction. We compared the sequences against the nonredundant nucleotide database at NCBI using tBLASTx (Altschul et al. 1990; Gish and States 1993). The most closely related sequences, class VIII myosins, generated EXPECT values less than 10^{-11} .

The larger PCR products did not encode a continuous reading frame. Comparing the nucleotide

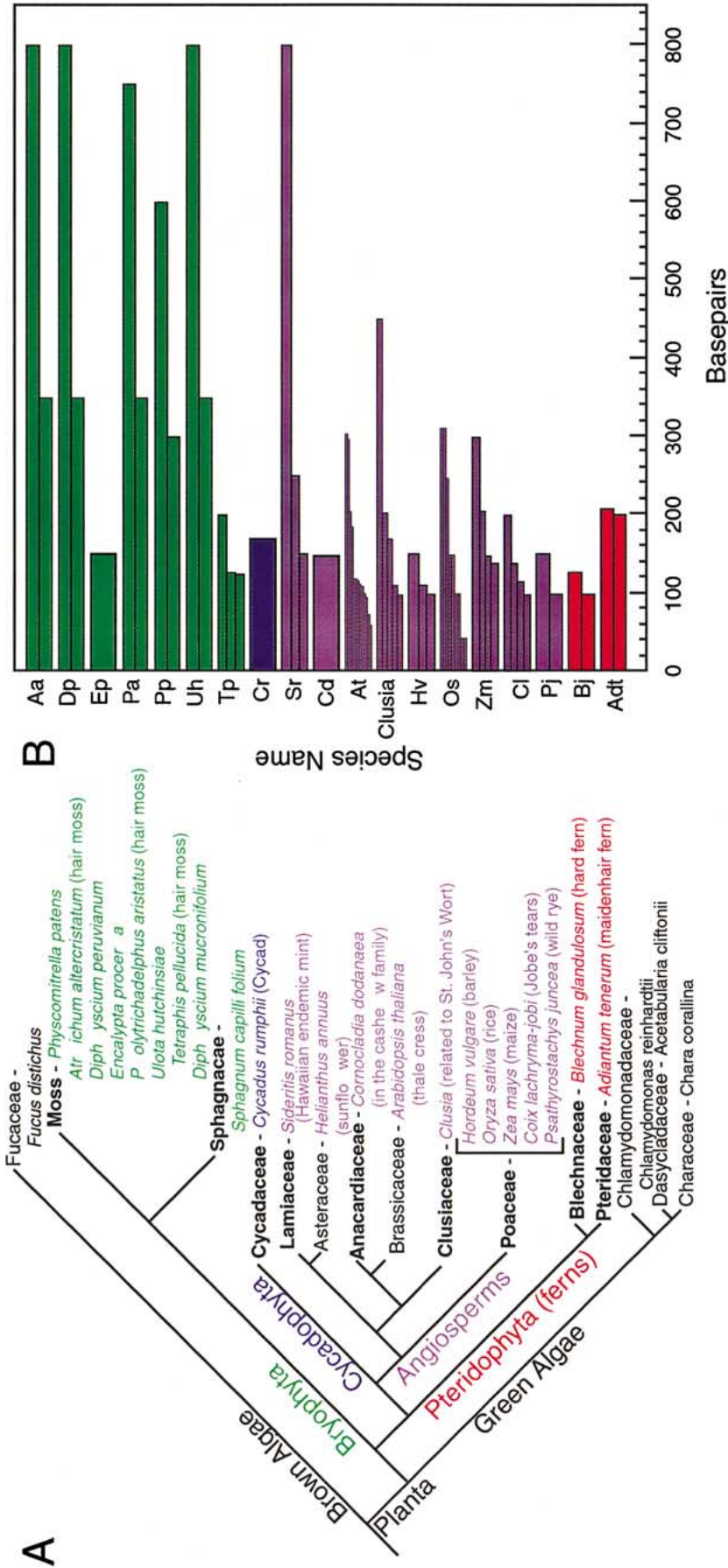


Fig. 2. A Phylogenetic relationships of plant species and families used in this study (TAP Group 1998; Kirk 1998; Pryer et al. 2001). Our degenerate PCR products came from species belonging to the orders denoted in *boldface*. **B** The intron sizes found in genomic sequences of class XI myosins. Multiple bars within a species indicate the intron sizes for individual clones. For example, each of the 13 class XI *Arabidopsis* myosins is depicted as an individual bar.

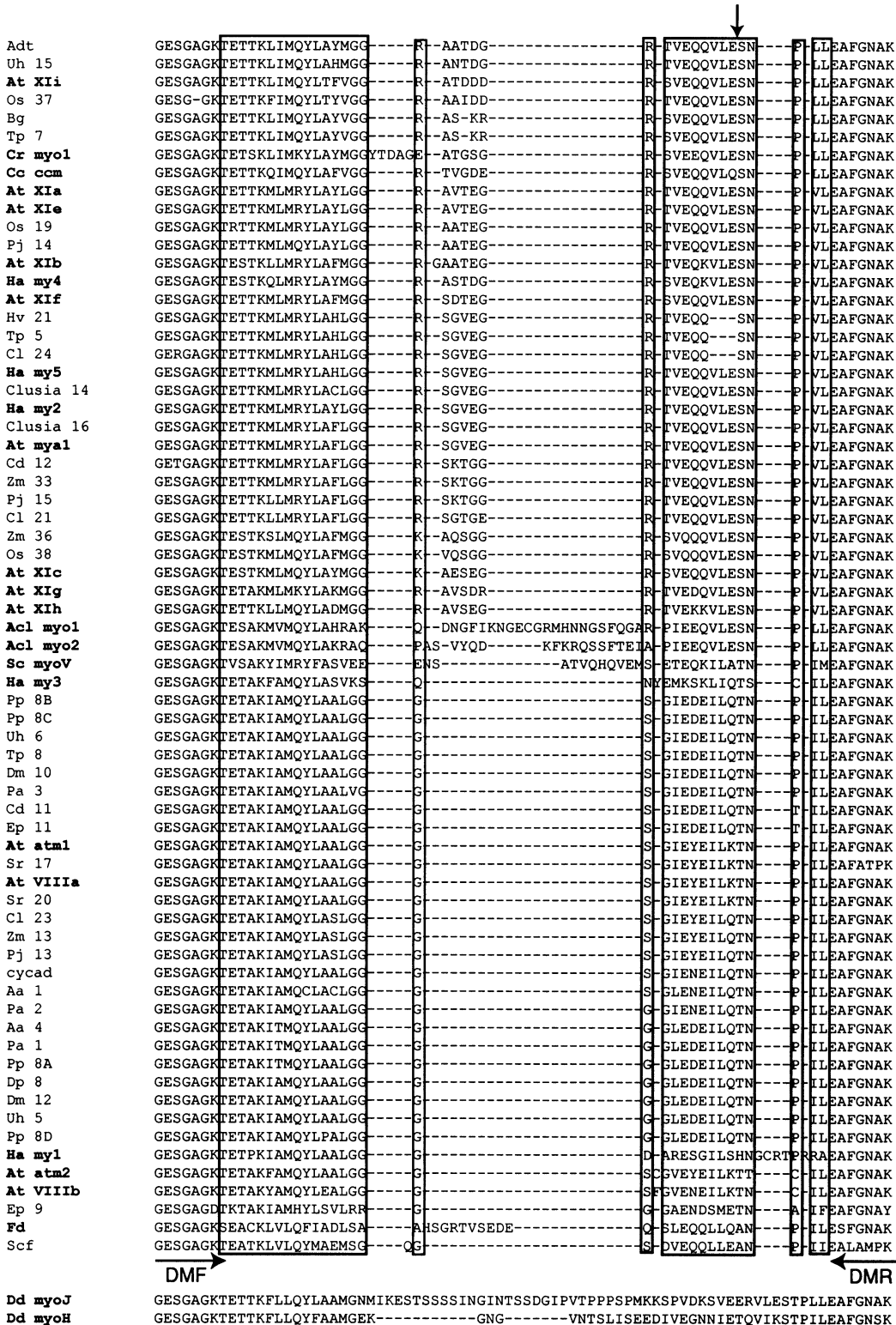


Fig. 3. Alignment of amino acid sequences used in this study. Sequences previously deposited in public databases are denoted in *boldface*. Primers used for PCR are indicated at the *bottom* of the alignment by *arrows*. The *arrow* at the *top* of the alignment indicates the position of the intron in all class XI myosins. The *boxed*

letters are the amino acids used to build the tree in Fig. 4. The two sequences at the *bottom* are putative class V myosins from *Dictyostelium discoideum*. They are shown here to demonstrate the difficulty in aligning these sequences with plant myosin sequences in this region.

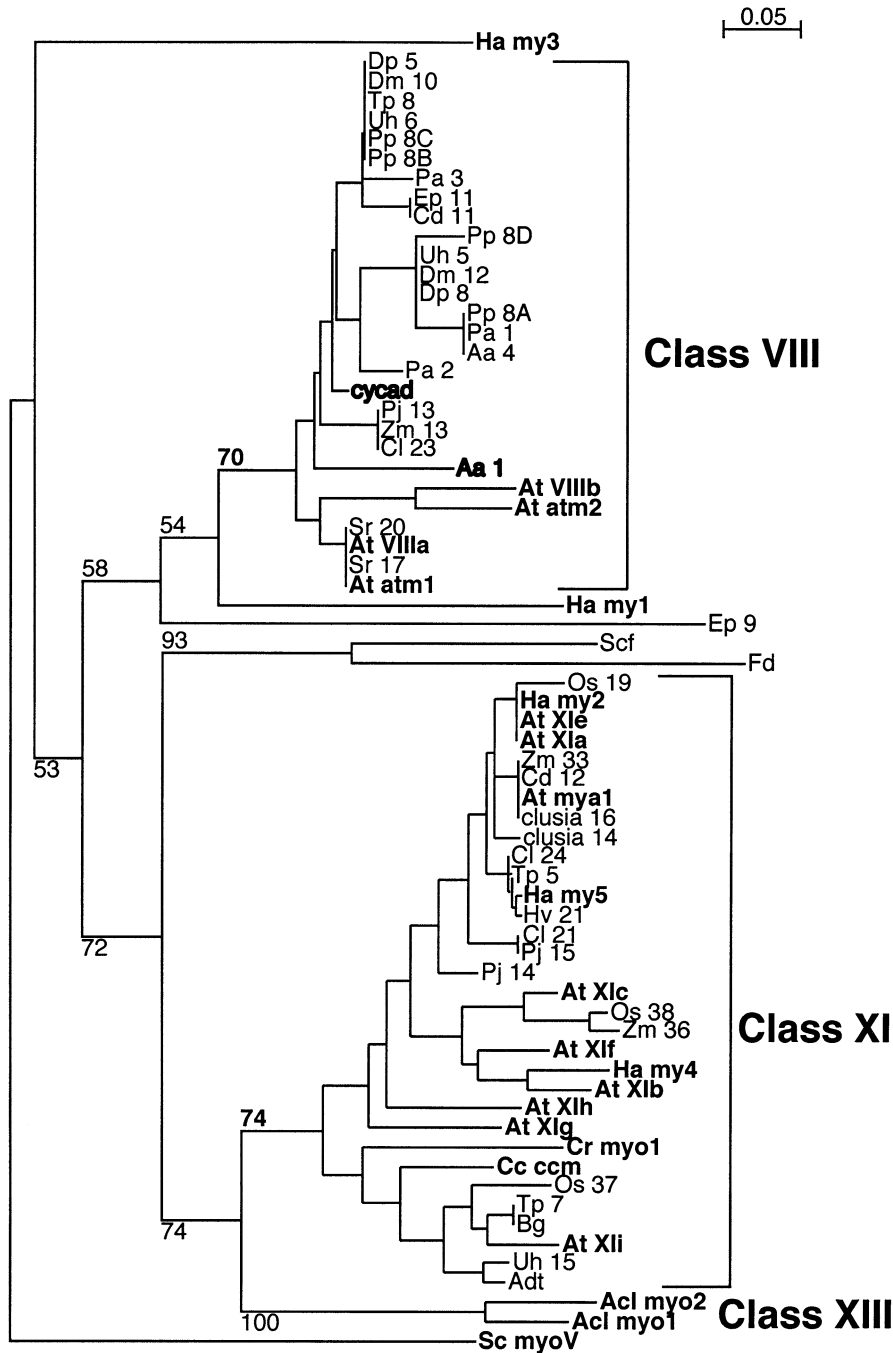


Fig. 4. Neighbor-joining tree. Numbers indicate percentage bootstrap value of 10,000 trials. Names in *boldface* are sequences obtained from public databases.

sequence using tBLASTx, there was a region of sequence similarity at the 5' end, a variable length of divergent sequence, and a region of similarity at the 3' end of each PCR product. The matching regions are closely related to class XI myosins, with EXPECT values less than 10^{-11} . Closer examination of the sequences showed a universal consensus splice site in every case which would remove the divergent region. Also, all myosin XI sequences from the *Arabidopsis thaliana* genome possess an intron at the same site. Using all publicly available annotated sequences

from the *Oryza sativa* genome, we have discovered seven myosin XI sequences and they all contain an intron at the same site (Fig. 2B). Thus, the genomic structure within this region is conserved in all known class XI myosins.

Closer inspection of the intronic sequence reveals no obvious motifs or similarities, as would be expected for such a disparate group of species. However, we observed a trend in the length of the intron which differs between taxa (Fig. 2B). The bryophytes, on average, have significantly longer introns than the

angiosperms. The main exceptions appear to be two disparate mosses, *Tetraphis pellucida* and *Encalypta procera*, and one angiosperm, *Sideritis romanus* (Hawaiian endemic mint). Genomic information was not always available for sequences obtained from public databases since these were in most cases cDNA sequences. However, the complete genome from *Arabidopsis thaliana* and the partial genome from *Oryza sativa* show that the intron is conserved and the length on average is short compared to the bryophyte introns (Fig. 2B).

To determine better the relationships among the obtained sequences, we aligned the deduced amino acid sequences from our cloned PCR products and myosin sequences obtained from public databases using ClustalW (Thompson et al. 1994) (Fig. 3). We included a class V myosin from *S. cerevisiae* as an outgroup to root the tree. We selected the amino acids highlighted with boxes in Fig. 3 as the basis of our trees by removing the PCR primer sequences and unalignable regions in the middle, including a myosin XI specific region of five amino acids. We built phylogenetic trees using neighbor-joining, maximum-parsimony, and maximum-likelihood analyses of the resulting amino acid alignment. Although all three methods support the separation of classes VIII and XI, maximum likelihood was consistent when using amino acid sequence but gave stronger support with DNA sequence information. Our neighbor-joining analysis of the protein alignment shows that class VIII and class XI myosins form clades with bootstrap values of 70% and 74%, respectively (Fig. 4). Since the sequence used is short and well conserved, these bootstrap values show significant support for the two major groups and corroborate the groupings based on the presence or absence of the intron.

To address further the significance of this result in the larger context of myosin family assignment, we built a neighbor-joining tree containing a larger number of myosin families to determine if this region is capable of recapitulating published myosin phylogenies. We used 139 sequences representing classes I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XV, and XVII. Additional sequences are referenced at <http://www.mrc-lmb.cam.ac.uk/myosin/trees/accession.html> and in the legend to Supplemental Fig. 1. In this more complete analysis (Supplemental Fig. 1), classes VIII and XI remain supported with bootstrap values of 70 and 60%, respectively. Specifically all of the sequences in class VIII or XI shown in Fig. 4 still group together, and more importantly, no other sequences group within these clades. The plant sequences which did not group with either class XI or class VIII in Fig. 4 in the more detailed analysis, as discussed below, remain unresolved in the larger analysis; importantly, these sequences separate from all newly included families as well as from classes

VIII and XI. Additionally, the analysis of this sequence region also diagnoses class II and class V myosins with a single exception. The *C. elegans* sequence for a class II myosin (Y11D7A.14) is not included in the class II clade. We are not sure why this sequence is not resolved. The studied sequence region may also be able to diagnose class IX and class XIII since these clades have high support (bootstrap values of 97 and 99%, respectively). One caveat, however, is that these clades contain only two sequences. More sequences from classes IX and XIII must be identified to support this conclusion. Interestingly this larger analysis is unable to resolve other families, specifically classes I, VI, and VII, implying that this region of sequence is not a good diagnosis for those classes.

With our analyses we identified a few outlier sequences. The *Sphagnum* and *Fucus* sequences cannot be identified as either class VIII or class XI. Using tBlastn, the *Sphagnum* sequence is most similar to class VII myosins (EXPECT value of 10^{-11}) and the *Fucus* sequence is most similar to class II myosins (EXPECT value of 10^{-8}). However, these sequences always group together in our phylogenetic analyses. When we include several sequences from class II and class VII in a phylogenetic analysis, the support level for the clade containing *Sphagnum* and *Fucus* sequences drops precipitously, suggesting that this grouping is due to long-branch attraction (supplemental Fig. 1 and data not shown). To assign these sequences to any class reliably, longer sequences are needed.

Two previously identified class VIII myosins from *Helianthus annuus* (sunflower) do not reliably fall within our class VIII group. Ha my3 falls outside all plant sequences tested, and Ha my1, along with a new moss sequence, Ep 9, is more divergent than most class VIII myosins used in this study. Analysis of the entire sequence of Ha my3 indicates that it may not actually belong in the class VIII family (Vugrek and Moepps 2002). Ha my3 is a very short myosin, only 900 amino acids. It is predicted to contain only one light chain binding motif, compared to the four binding motifs usually found in class VIII myosins. The tail consists of 40 amino acids, with no propensity to form a coiled-coil. All myosin VIII full-length cDNA sequences, to date, have had significantly longer tails with heptad repeats. Thus, it is possible that Ha my3 is not a myosin VIII but instead a new class of myosins most closely related to class VIII (Vugrek and Moepps 2002). The moss Ep 9 is most similar to class VIII myosins, with an EXPECT value of 10^{-9} , and does not contain an intron. Therefore, it most likely belongs in the VIII family but may also be a member of a myosin family closely related to class VIII, similar to Ha my3. Strict interpretation of the unrooted tree does not allow us to place Ha my1 and Ep 9 definitively into class VIII; however, other studies placed Ha my1 well within class VIII (Hodge

and Cope 2000; Reddy and Day 2001). The placement of these three sequences indicates that the grouping of class VIII myosins is less stable in our trees than class XI myosins.

The *Acetabularia* class XIII myosins grouped together, closer to class XI than class VIII. We did not identify any other members of class XIII. Thus, it is possible that the green alga *Acetabularia cliftonii* has evolved a unique myosin. This is similar to the situation in the plasmodia, such as *Toxoplasma gondii*, which have their own unique myosin not found in any other eukaryotic organisms.

Previous phylogenetic studies of the catalytic domain identified a sequence from *Dictyostelium* (myoJ) believed to be a class V myosin. However, in some analyses, myoJ is more closely related to class XI myosins (Hodge and Cope 2000; Sellers 2000) than to class V myosins, although it lacks the characteristic intron (M. Titus, personal communication). This is intriguing because it would place the first nonplant sequence in this class. We included myoJ and a putatively closely related sequence, myoH (sequences are shown at the bottom of the alignment in Fig. 3), in our analysis to relate these sequences to plant sequences; the sequences are very divergent from the plant sequences and are difficult to align. In our phylogenetic analyses, if myoJ alone is included with the plant sequences, it is placed within the myosin XI clade. However, the bootstrap value for the class XI node is lowered. If both myoJ and myoH sequences are included, they form a separate clade from the class XI sequences and the bootstrap value of the class XI node is notably lowered (Supplemental Fig. 1 and data not shown). As would be expected from highly divergent sequences, the *Dictyostelium* sequences' placement is unstable. However, if we include a larger number of class V myosins as outgroups without the *Dictyostelium* sequences, we obtain bootstrap values similar to those in Fig. 4. We also obtain similar trees if we change the class of myosin sequence used as an outgroup (Supplemental Fig. 1 and data not shown). Therefore, we are confident of the analysis with the plant sequences but unable to determine the relationship of the *Dictyostelium* sequences with respect to the plant class XI sequences.

The basal positioning of these atypical *Dictyostelium* sequences between class V and class XI myosins observed by others (Hodge and Cope 2000; Sellers 2000) leads to an interesting conjecture that class XI myosins evolved from a class V/XI ancestor via early eukaryotic speciation. Other similarities supporting this hypothesis include the fact that both class V and class XI myosins have six light chain binding motifs per heavy chain, whereas class VIII myosins possess only four. Additionally, the tails of class V and class XI myosins are similar in length, while class VIII

myosin tails are significantly shorter. We attempted to resolve whether class XI and class V myosins group separately from class VIII myosins through the addition of other classes to our analysis but were unable to find any significant support for such a hypothesis. This issue may be resolved in the future by phylogenetic analysis of longer sequences of these genes, including sequences from myosins from all kingdoms, together with functional data.

Conclusion

Our study shows that plant myosins can be classified with a fair degree of certainty using two characteristics. First, phylogenetic analysis of the alignment of a small stretch of sequence found in the catalytic domain from GESGAGK to LEAFGNK divides plant myosins into two major classes. Second, genomic sequence supports the classification because the presence of an intron exclusively diagnoses a class XI myosin. Our results, based on a more diverse sample of myosin sequences from plant taxa, strengthen earlier analyses showing that plant sequences group together exclusive of sequences from other kingdoms of life. This study provides a rapid way to classify a large number of partial myosin sequences from different plant species. Further studies may focus on determining the function of class VIII and XI myosins from model plant organisms.

Acknowledgments. We would like to thank Dr. Elizabeth A. Kellogg for providing us with genomic DNA from so many diverse plant species. She also provided us with invaluable suggestions and insights regarding the manuscript. We thank the Ho lab, especially Rodolfo Zentella, for the barley and maize genomic DNA. We thank the Beachy lab, especially Dr. Shunhong Dai, for providing us with rice genomic DNA. This work was supported by a Helen Hay Whitney postdoctoral fellowship to M.B. and a grant from the National Science Foundation to R.S.Q. (IBN-0112461).

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–104
- Arabidopsis Genome Initiative project (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Bement WM, Hasson T, Wirth JA, Cheney RE, Mooseker MS (1994a) Correction: Identification and overlapping expression of multiple unconventional myosin genes in vertebrate cell types. *Proc Natl Acad Sci USA* 91:11767c
- Bement WM, Hasson T, Wirth JA, Cheney RE, Mooseker MS (1994b) Identification and overlapping expression of multiple unconventional myosin genes in vertebrate cell types. *Proc Natl Acad Sci USA* 91:6549–6553
- Berg JS, Powell BC, Cheney RE (2001) A millennial myosin census. *Mol Biol Cell* 12:780–794
- Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78–94

- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3:266–272
- Goldman N, Whelan S (2000) Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Mol Biol Evol* 17:975–978
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–74
- Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919
- Hodge T, Cope MJ (2000) A myosin family tree. *J Cell Sci* 113:3353–3354
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282
- Kim AJ, Endow SA (2000) A kinesin family tree. *J Cell Sci* 113:3681–3682
- Kirk DL (1998) *Volvox*: Molecular-genetic origins of multicellularity and cellular differentiation. Cambridge University Press, New York
- Liu L, Zhou J, Pesacreta TC (2001) Maize myosins: Diversity, localization, and function. *Cell Motil Cytoskeleton* 48:130–148
- Muller T, Vingron M (2000) Modeling amino acid replacement. *J Comput Biol* 7:761–767
- Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409:618–622
- Reddy AS, Day IS (2001) Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence. *Genome Biol* 2:RESEARCH0024
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sellers JR (2000) Myosins: A diverse superfamily. *Biochim Biophys Acta* 1496:3–22
- Soldati T, Geissler H, Schwarz EC (1999) How many is enough? Exploring the myosin repertoire in the model eukaryote *Dictyostelium discoideum*. *Cell Biochem Biophys* 30:389–411
- Strimmer K, von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci USA* 94:6815–6819
- Swofford DL (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
- TAP Group (1998) An ordinal classification for the families of flowering plants. *Ann Mo Bot Gard* 85:531–553
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Vugrek O, Moepps B (2002) Hamy3, a novel type 100 kDa myosin from sunflower. *J Exp Bot* 53:769–771