A Millennial Myosin Census

Jonathan S. Berg, Bradford C. Powell, and Richard E. Cheney*

Department of Cell and Molecular Physiology, CB#7545, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Monitoring Editor: Thomas D. Pollard

The past decade has seen a remarkable explosion in our knowledge of the size and diversity of the myosin superfamily. Since these actin-based motors are candidates to provide the molecular basis for many cellular movements, it is essential that motility researchers be aware of the complete set of myosins in a given organism. The availability of cDNA and/or draft genomic sequences from humans, Drosophila melanogaster, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Dictyostelium discoideum has allowed us to tentatively define and compare the sets of myosin genes in these organisms. This analysis has also led to the identification of several putative myosin genes that may be of general interest. In humans, for example, we find a total of 40 known or predicted myosin genes including two new myosins-I, three new class II (conventional) myosins, a second member of the class III/ninaC myosins, a gene similar to the class XV deafness myosin, and a novel myosin sharing at most 33% identity with other members of the superfamily. These myosins are in addition to the recently discovered class XVI myosin with N-terminal ankyrin repeats and two human genes with similarity to the class XVIII PDZ-myosin from mouse. We briefly describe these newly recognized myosins and extend our previous phylogenetic analysis of the myosin superfamily to include a comparison of the complete or nearly complete inventories of myosin genes from several experimentally important organisms.

INTRODUCTION

Myosins are actin-based motors known or hypothesized to play fundamental roles in many forms of eukaryotic motility such as cell crawling, cytokinesis, phagocytosis, growth cone extension, maintenance of cell shape, and organelle/particle trafficking. Although actin polymerization alone can drive some forms of motility, myosins appear to power an assortment of movements and are important in processes such as signal transduction (Bahrer, 2000) and establishment of polarity (Yin et al., 2000). Recent evidence even implicates myosins in the polymerization of actin (Evangelista et al., 2000; Lechler et al., 2000; Lee et al., 2000). To understand the molecular basis of actin-based motility, it is thus critical to identify the pool of candidate motor proteins.

Members of the myosin superfamily are defined by the presence of a heavy chain with a conserved ~80 kDa catalytic domain. In most myosins, the catalytic domain is followed by an α-helical light chain-binding region consisting of one or more IQ motifs. Most myosins also have a C-terminal tail and/or an N-terminal extension thought to endow class-specific properties such as membrane binding or kinase activity. Class II myosins are familiar from studies of muscle contraction, but the myosin superfamily also contains a large number of other myosins with quite different tail domains. Although the conventional-unconventional dichotomy is clearly artificial in terms of structure and evolution, it is operationally useful because of the historical emphasis on conventional myosins. The importance of unconventional myosins is stressed by the fact that they constitute 4 of 5 myosin genes in Saccharomyces cerevisiae, 11 of 13 myosins in Drosophila, and approximately two-thirds of the myosin genes in humans. In addition, current evidence indicates that typical nonmuscle cells appear to express only 1 or 2 conventional myosin genes but upward of 10 or more unconventional myosins (Bement et al., 1994).

To delineate the extent of myosin diversity, we used computer-based search algorithms to identify and predict new myosin heavy chain genes from several eukaryotic organisms. Large-scale genomic sequencing efforts have generated complete or nearly complete draft genomes from humans, the fruit fly Drosophila melanogaster (Adams et al., 2000), the nematode (The C. elegans Sequencing Consortium, 1998), the budding yeast S. cerevisiae (Goffeau et al., 1996), and the vascular plant Arabidopsis (The Arabidopsis Genome Initiative, 2000). In addition, the genome of the fission yeast Schizosaccharomyces pombe is virtually complete, and many myosin cDNAs have been identified in the slime mold Dictyostelium (Schwarz et al., 1999). Although complete analysis and annotation of genomic sequences will require many

* Corresponding author. E-mail: cheneyr@med.unc.edu.
Abbreviations used: aa, amino acid; acc. no., accession number; BAC, bacterial artificial chromosome; EST, expressed sequence tag; GAP, GTPase-activating protein.
years, we provide here an initial census and comparison of the myosins present in these organisms. For more detailed information concerning the structures, functions, and physiology of myosins, the reader should see several recent reviews (Baker and Titus, 1998; Merrall et al., 1998; Oliver et al., 1999; Sellers, 1999; Sokac and Bement, 2000; Wu et al., 2000).

GENOMIC SEQUENCE SEARCHES AND PHYLOGENETIC TREES

To identify previously unknown myosin genes, we used known myosin head sequences to search genomic clones from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) using TBLASTN (Altschul et al., 1997). Matches from this screen were used in semiautomated queries of the GenBank “nr” (nonredundant) database to eliminate clones corresponding to known myosin genes. We used the Genscan (Burge and Karlin, 1997) gene prediction algorithm (http://genes.mit.edu/GENSCAN.html) to identify the putative coding sequences of novel myosin genes and CLUSTALX (Thompson et al., 1997) to align myosin head domain sequences and generate both organism-specific and more comprehensive phylogenetic trees (Cheney et al., 1993; Goodson and Spudich, 1993; Hodge and Cope, 2000; Korn, 2000). Although it is unlikely that the trees shown here perfectly reflect myosin's evolutionary history, they do provide a number of useful insights as well as a simple method to graphically portray sequence relationships. The sequences predicted from genomic clones and the accession numbers (acc. nos.) of sequences used to make these trees are available upon request.

Genomic analysis of this type has certain caveats. First, although it can be relatively straightforward to recognize myosin-like head domain sequences in genomic DNA, it is more difficult to predict and assemble with 100% accuracy all of the dozens of exons that may be present in a full-length myosin transcript. This is especially true if the exons are very small, constitute novel C-terminal or N-terminal extensions, or are spread over more than one bacterial artificial chromosome (BAC). Second, it is possible that some of the putative myosin genes are pseudogenes. Because pseudogenes typically lack introns and are generally not transcribed into mRNA, it is thus important to determine whether transcripts for predicted genes are actually present in cDNA or expressed sequence tag (EST) databases. Third, some BACs currently contain unordered contigs and some areas of the “completed” genomes have not been fully sequenced. Thus, it is possible that these regions contain additional myosin genes that remain to be discovered. Finally, given that myosins and kinesins share strikingly similar catalytic cores despite little or no detectable sequence similarity (Kull et al., 1996), it is possible that additional myosin genes exist but are so divergent as to be unrecognizable by the search algorithms used here.

THE HUMAN GENOME INCLUDES ~40 MYOSIN GENES

Our analysis of the draft human genome sequence indicates that the human genome contains ~40 myosin genes that can be divided into ~12 classes based on analysis of their head and tail domain structures. Figure 1 shows an unrooted phylogenetic tree consisting of all known or predicted myosin genes in humans. The bar diagrams in Figure 2 illustrate the expected structures of a number of the recently discovered myosins.

Three New Conventional Myosins in Humans

The conventional class II myosins in humans consist of 15 genes, including the known cluster of 6 skeletal muscle myosin heavy chains on chromosome 17p (Weiss et al., 1999), 2 cardiac myosin heavy chains, the smooth-muscle myosin heavy chain, and 2 nonmuscle myosin-II heavy chains. It should be noted that this tally includes an apparent pseudogene on chromosome 7 (Schachat and Briggs, 1999). This pseudogene is most similar to the “superfast” myosin of cat jaw muscles and is presumably in the process of being lost from the human genome. As described below, our current tally includes three new conventional myosins predicted from genomic sequence and supported by partial cDNA evidence. The discovery of three additional conventional myosins is extremely interesting given the extensive previous research and fundamental importance of this class in cell and organismal physiology.

Genscan predictions indicate the presence of a novel conventional myosin gene from overlapping BACs on chromosome 19 (acc. nos. AC020906 and AC010515). Based on the predicted protein sequence, the conserved motor domain is ~75% identical to nonmuscle myosin-IIB and segregates with smooth muscle and nonmuscle myosins-II in the phylogenetic tree. As expected for a myosin-II, the tail domain is predicted to consist almost entirely of coiled coil. The existence of a new member of the nonmuscle/smooth muscle myosin-II group in vertebrates has important implications for studies in many areas including smooth muscle contraction, cell motility, and cytokinesis.

A recently cloned partial cDNA (KIAA1512, acc. no. AB040945) from the Kazusa DNA Research Institute (Japan) contains a small part of a myosin motor domain, two IQ motifs, and a tail domain consisting almost entirely of coiled-coil. Our searches identified a BAC (acc. no. AL132825) corresponding to this sequence that includes additional head domain exons. The predicted protein matches the cDNA almost perfectly and the conserved motor domain is ~70% identical to human skeletal muscle myosins-II. Based on its position in the phylogenetic tree we predict this gene to be a divergent member of the striated muscle myosin group, although it is not yet known what muscles or tissues it is expressed in. The gene for KIAA1512 maps to chromosome 20q11.21-12 and is thus not part of the chromosome 17 muscle myosin cluster.

A second partial cDNA (KIAA1100, acc. no. AB023217) from the Kazusa group encodes an 891-amino acid (aa) fragment predicted to form a coiled-coil structure. This sequence is 58% identical to the tail domain of human cardiac myosin-IIa. We found several BACs from chromosome 3 containing the gene coding for this putative myosin (acc. nos. AC069499 and AC020731). Although these BACs are highly fragmentary, Genscan identified several exons encoding part of a myosin motor domain with ~58% identity to human cardiac myosin-IIa. Again, the tissue distribution of this putative striated muscle myosin is not known.
Figure 1. An unrooted phylogenetic tree of the myosin superfamily in humans. In this tree, solid lines indicate myosins known from cDNA sequences and dashed lines indicate putative myosins predicted from genomic sequence. Black dots represent nodes validated by bootstrapping values of >90%. Official HUGO human gene names (http://www.gene.ucl.ac.uk/nomenclature/) are indicated in bold, and synonymous names from human or related organisms such as rat (Myr) are indicated in plain type. Recently discovered or newly predicted myosin genes are italicized, and potential gene names suggested for myosins predicted from genomic sequence are indicated in brackets. Although all the myosin genes depicted in this figure are known to be present in humans, in a few cases sequence homologues from closely related organisms such as rat (myr1, myr4, myr6, and myr8) or mouse (Myo1f) were used to generate the tree. Sequences assembled from overlapping genomic clones or cDNAs are indicated in square brackets. This and the other trees illustrated here were generated from myosin head domain aa sequences essentially as described by Cheney et al. (1993). PH, pleckstrin homology; MyTH4, myosin tail homology 4; FERM, band four point one/ezrin/radixin/moesin. Note that the conventional myosin pseudogene is not included in the phylogenetic analysis due to incomplete head domain sequence. See text for specific references.
Figure 2. Bar diagrams of newly discovered or newly predicted myosins from humans and Drosophila. Dashed lines and light shading indicate expected structure of myosins predicted from partial genomic sequence. Dotted lines indicate uncertain N-terminal extensions predicted by Genscan. Glob., AF-6/dilute-class globular tail; Pro., proline-rich region. See text for details.
Eleven New Unconventional Myosin Genes in Humans

The unconventional myosins, which make up all other myosin classes, now include ~25 genes in humans, many of which were known previously. The most recently discovered unconventional myosins include new members of classes previously associated with deafness, blindness, and organelle transport, as well as several myosins that represent novel classes (Figures 1–3).

We identified two new class I myosin genes, tentatively named MYO1G and MYO1H. The putative MYO1G gene localizes to chromosome 7p11.2-p13 (acc. no. AC004847) and the predicted motor domain is ~60% identical to MYO1D/rat myr4/mouse myosin Iy. The MYO1H gene is from a chromosome 12 sequence (acc. no. NT_002188), and the predicted motor domain is ~50% identical to human MYO1C/rat myr2/mouse myosin Iβ. Based on Genscan predictions, both new myosin-I genes have a basic tail but lack an SH3 domain. Both of the predicted coding sequences match numerous ESTs, suggesting that the genes are expressed. With the addition of these new myosins-I, each of the four previously defined myosin-I subclasses (Mooseker and Cheney, 1995) now contains two genes in humans.

A human member of the ninaC class, myosin-IIIa, was recently discovered by Dose and Burnside (2000). The human MYO3A gene localizes to chromosome 10p11.1, and the protein contains an N-terminal kinase domain similar to ninaC. Myosin-IIIa is present in the retina and retinal pigmented epithelium, consistent with the possibility that ninaC it plays a role in vision (Dose and Burnside, 2000). In addition, we have identified a putative MYO3B gene from unfinished genomic BACs on chromosome 2 (acc. nos. AC012594 and AC068280). Although these clones do not include exons coding for the N-terminal kinase domain, the predicted motor domain sequence of myosin-IIIb is 60% identical to human myosin-IIIa. In addition, the predicted protein is truncated after a single IQ motif, suggesting that this myosin lacks a tail domain or that the exons coding for the tail domain are present on a different genomic clone.

A third member of the class V putative organellar transporters, myosin-Vc, has recently been cloned (acc. no. AF272390). The MYO5C gene is present on chromosome 15 immediately adjacent to the MYO5A/human dilute gene. The protein is structurally similar (~50% identical) to the two other class V myosins and appears to be a class V myosin of epithelial and glandular tissues (Rodriguez and Cheney, 1999).

A second member of the class VII myosins, myosin-VIIib, has also been cloned (Chen et al., 2001). The MYO7B gene is present on chromosome 2, and the protein is structurally similar to other class VII myosins, containing paired MYTH4 and FERM domains. It shares ~50% identity with myosin-VIIa, the gene mutated in Usher syndrome, which is characterized by deafness and progressive retinal degeneration (Weil et al., 1995; Chen et al., 1996).

Our searches identified a second class XV myosin gene in a BAC from chromosome 17 (acc. no. AC019214). The conserved motor domain of the predicted myosin is ~45% identical to myosin-XV, a ~395-kDa myosin associated with hereditary deafness in humans and mice (Wang et al., 1998). The predicted protein is structurally similar to myosin-XV, which contains an N-terminal extension and a tail domain containing MyTH4, FERM, and SH3 domains (Liang et al., 1999). Thus, we tentatively name this protein myosin-XVb (MYO15B). Although the BAC is still partly unordered, numerous ESTs indicate that this gene is likely to be expressed. The founding member of the class XVII myosins, rat myr8, was discovered by Patel et al. (1998). A partial human cDNA (KIAA0865, acc. no. AB020672) corresponds to the C-terminus of myr8. Interestingly, the full-length myr8-coding sequence (acc. no. AF209114) contains six ankyrin repeats at the N-terminus. Because these protein motifs often function as sites for protein-protein interaction, the N-terminus of this myosin is likely to interact with other proteins. Preliminary analysis of this myosin in rat indicates that it is expressed in developing brain (Patel et al., 1998) and exists in both short and long (myr8b, acc. no. AY004215) forms.

The founding member of class XVIII, “Myosin containing PDZ domain (MysPDZ),” was discovered in mouse by Furusawa et al. (2000). This myosin has an N-terminal PDZ domain, one IQ motif, and a tail of segmented coiled-coil. Furusawa et al. also found a human homologue in the GenBank database that is 94% identical overall but lacks the N-terminal PDZ domain (KIAA0216, acc. no. D86970). However, a single exon from a chromosome 17 BAC (acc. no. AC005412) encodes an ~350-aa peptide that matches with ~94% identity to the N-terminal PDZ domain of mouse MysPDZ. It is thus likely that the KIAA0216 sequence represents a partial cDNA or a splice variant of human MysPDZ that lacks the N-terminal PDZ region. Although the function of this myosin is unknown, its mRNA is apparently ubiquitous (Furusawa et al., 2000), and it may function as part of a mobile scaffolding complex through interactions with the PDZ domain. The PDZ-myoins from vertebrates and Droso- phila (see below) constitute a new class of myosins, class XVIII (Yamashita et al., 2000). Thus, the human mysPDZ gene should be named MYO18A.

Genome annotators identified part of a gene somewhat similar to the human PDZ-myoins on a BAC from human chromosome 22 (acc. no. Z98949). Although this sequence contains only the C-terminal portion of a myosin head domain and a tail consisting of segmented coiled-coil, an adjacent genomic clone (acc. no. AL080245) contains some of the remaining head domain exons. The full-length predicted protein shares ~40% protein identity with human and mouse MysPDZ and groups reliably with the other class XVIII myosins in phylogenetic trees, suggesting that the gene be named MYO18B. Despite the general similarity to the PDZ-myoins, we were not able to identify exons coding for a PDZ sequence in genomic sequence upstream of the head domain.

Our searches predicted a novel myosin gene from a chromosome 17 genomic BAC (acc. no. AC023133). This rather divergent myosin is predicted to share at most ~35% identity with other myosin motor domains, and phylogenetic analysis indicates that it does not group with any previously known myosin classes. Genscan predictions indicate that the protein may contain an ~250-aa N-terminal extension, two IQ motifs, and a short tail domain. Although this putative myosin appears to constitute a novel class, we have postioned assignment of a class number until the predicted structure is confirmed by a full-length cDNA.
Figure 3. An unrooted phylogenetic tree of the myosin superfamily. A phylogenetic tree was assembled using myosin head domain protein sequences from all currently known or predicted myosins in humans, Drosophila, C. elegans, Dictyostelium, Arabidopsis, S. cerevisiae, and S. pombe, as well as selected myosins from other organisms. Species are identified by two or three letter abbreviations, and some gene names have been shortened to save space. Sequences predicted in full or in part from genomic clones are indicated by an asterisk. Bold type indicates orphan myosins that have not yet been assigned a class number. To minimize the introduction of gaps in the sequence alignment used to generate this tree, we increased the default CLUSTALX gap insertion penalty to 25 and the gap extension penalty to 0.5. Similar trees are available on the myosin home page (http://www.mrc-lmb.cam.ac.uk/myosin/myosin.html) and in recent papers (Soldati et al., 1999; Hodge and Cope, 2000; Sellers, 2000).
MYOSINS IN OTHER MODEL ORGANISMS

The completion or near-completion of the genomes from a variety of organisms now permits annotation of the myosin genes in these organisms as well (Figures 4 and 5). Because relatively little mouse genomic sequence is currently available, we have not performed a complete census of mouse myosins. Nevertheless, we anticipate that humans and mice will share most, if not all, myosin genes in common (see DISCUSSION for comments on mouse nomenclature). Geneticists have already identified several mouse myosin mutations that have played an important role in the understanding of human diseases and myosin functions. These include the dilute (Myo5a) mutation, which leads to coat color defects and failure of organelle transport, and the
Current myosin inventories and nomenclature for selected organisms. Classes are designated by the letter M (myosin) followed by Arabic numerals corresponding to the class number. Genes predicted in full or in part from genomic sequence are indicated by an asterisk.

**Drosophila**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr.</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>31DF</td>
<td>65573</td>
<td>Myosin-IA</td>
</tr>
<tr>
<td>M2</td>
<td>6F</td>
<td>60796</td>
<td>Myosin-II</td>
</tr>
<tr>
<td>M3</td>
<td>26A</td>
<td>50965</td>
<td>Myosin-III</td>
</tr>
<tr>
<td>M4</td>
<td>60E</td>
<td>53816</td>
<td>Myosin-IV</td>
</tr>
<tr>
<td>M5</td>
<td>26A</td>
<td>60313</td>
<td>Myosin-V</td>
</tr>
<tr>
<td>M6</td>
<td>43C</td>
<td>60382</td>
<td>Myosin-VI</td>
</tr>
<tr>
<td>M7</td>
<td>6F6</td>
<td>67007</td>
<td>Myosin-VII</td>
</tr>
<tr>
<td>M8</td>
<td>28B</td>
<td>62326</td>
<td>Myosin-VIII</td>
</tr>
<tr>
<td>M9</td>
<td>958B</td>
<td>64415</td>
<td>Myosin-IX</td>
</tr>
</tbody>
</table>

**Dictyostelium**

<table>
<thead>
<tr>
<th>Name</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>MyoA</td>
<td>P22467</td>
</tr>
<tr>
<td>M2</td>
<td>MyoB</td>
<td>P34092</td>
</tr>
<tr>
<td>M3</td>
<td>MyoC</td>
<td>33333</td>
</tr>
<tr>
<td>M4</td>
<td>MyoD</td>
<td>P34105</td>
</tr>
<tr>
<td>M5</td>
<td>MyoE</td>
<td>68080</td>
</tr>
<tr>
<td>M6</td>
<td>MyoK</td>
<td>60553</td>
</tr>
<tr>
<td>M7</td>
<td>Myo1</td>
<td>35321</td>
</tr>
<tr>
<td>M8</td>
<td>Myo2</td>
<td>35320</td>
</tr>
</tbody>
</table>

**C. elegans**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>he-1</td>
<td>AE000384</td>
</tr>
<tr>
<td>M2</td>
<td>he-2</td>
<td>AE000385</td>
</tr>
<tr>
<td>M3</td>
<td>he-3</td>
<td>AE000386</td>
</tr>
<tr>
<td>M4</td>
<td>he-4</td>
<td>AE000387</td>
</tr>
</tbody>
</table>

**Arabidopsis**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>ATM</td>
<td>X69505</td>
</tr>
<tr>
<td>M2</td>
<td>mbc</td>
<td>P22467</td>
</tr>
<tr>
<td>M3</td>
<td>mpa</td>
<td>P22467</td>
</tr>
<tr>
<td>M4</td>
<td>mpa</td>
<td>P22467</td>
</tr>
</tbody>
</table>

**S. cerevisiae**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>MYO3</td>
<td>57986</td>
</tr>
<tr>
<td>M2</td>
<td>MYO4</td>
<td>60439</td>
</tr>
</tbody>
</table>

**S. pombe**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Acc. #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>myo1</td>
<td>T39427</td>
</tr>
<tr>
<td>M2</td>
<td>myo2</td>
<td>U75357</td>
</tr>
</tbody>
</table>

![Figure 5](https://example.com/figure5.png)
Snell’s waltzer (Myo6), shaker-1 (Myo7a), and shaker-2 (Myo15a) mutations, which lead to deafness. We will focus on the remainder of the analysis on nonvertebrate model organisms.

Yeast Contain Five Genes from Three Classes

The complete genome of the budding yeast (S. cerevisiae) contains only five myosin genes from three classes (Brown, 1997). Two of these genes are from class I (MYO3 and MYO5), two are from class V (MYO2 and MYO4), and one is from class II (MYO1). The fission yeast (S. pombe) genome is nearly complete and also appears to contain five myosin genes from three classes. One of the genes is from class I (myo1; Lee et al., 2000), two of the genes are from class II (myo2 and myo2; Bezanilla et al., 1997; May et al., 1998), and two of the genes are from class V (myo5/myo51 and myo5/myo52; Win et al., 2001). Intriguingly, the distribution of myosin genes in these organisms differs somewhat, in that S. pombe has only a single class I myosin but a pair of conventional myosins, indicating the loss or gain of myosin genes at some point in yeast evolution. Although yeast may not be entirely representative of fungi as a kingdom, the presence of myosin genes from classes I, II, and V indicates that these classes are particularly ancient and important.

Four novel myosins from Drosophila

Several new myosins were predicted from the initial annotation of the Drosophila genomic sequence by Celera Genomics (Rockville, MD) and a consortium of researchers (Adams et al., 2000; Goldstein and Gunawardena, 2000). These myosins include a class XV myosin at chromosomal locus 10A, a gene at 89B with similarity to the PDZ-containing myosin described above, and a novel myosin gene at 29CD (Yamashita et al., 2000). In addition, we report here the identification of an additional myosin gene at 95E. The new Drosophila myosins predicted from genomic sequence (and supported by partial cDNA evidence) are depicted in bar diagram form in Figure 2B. The Drosophila myosins are highlighted against the background of the full phylogenetic tree in Figure 4A.

The myosin gene at 10A (CG2174, acc. no. AAF47980) is predicted to encode a 2422-aa myosin with ~50% identity to the catalytic domain of mouse myosin-XV. The Drosophila myosin-XV contains MyTH4 and a FERM domains in its tail but appears to lack the large N-terminal extension found in vertebrate myosins-XV (Liang et al., 1999). The fly gene also has a curious intron/exon structure, including one exon of 6915 base pairs coding for virtually the entire protein. However, the EST corresponding to CG2174 indicates that it is expressed.

The predicted myosin gene at 89B (CG10218, acc. nos. AAF55271 and AAF55272) encodes a myosin with ~40% identity to the catalytic domain of the class XVIII PDZ-myosins, two IQ motifs, and a tail of segmented coiled-coil structure. A partial cDNA encoding some of this tail domain of this myosin (acc. no. AJ132656) is present in the GenBank database. In addition, GenScan predictions for this gene suggest that an N-terminal PDZ domain may be present. Given the recent discovery of PDZ-containing myosins in vertebrates (see above), the presence of this domain and the phylogenetic analysis clearly indicate the relatedness of the fly and vertebrate class XVIII myosins (Figure 3).

The novel myosin at 29CD (CG10595, acc. no. AAF52683) is predicted to have an ~340-aa N-terminal sequence with no homology to other known proteins, although this region is enriched for glycine and serine. The catalytic domain is only ~30% identical to other myosins, and the remainder of the protein is predicted to contain a single IQ motif and a short proline-rich tail. By phylogenetic analysis, this myosin does not group tightly with any other classes, and it may represent a novel class of myosins. However, we have postponed numbering this class until the predicted structure is confirmed by a full-length cDNA.

The novel myosin identified from Drosophila genomic sequence at 95E (acc. no. AE003746) is predicted to have an ~250-aa insert at the region of loop 1 (near the nucleotide-binding site) in the catalytic domain and two IQ motifs. This myosin is not obviously related to any other classes by phylogenetic analysis of the motor domain, although the basic tail domain is similar to those of class I myosins. The initial annotation of Drosophila genomic sequence identified only the first ~30 aa of this myosin (CG5501, acc. no. AAF56246), and the motor domain was subsequently identified by Yamashita et al. (2000). An EST matching the tail region of our predicted coding sequence supports the possibility that this gene encodes a novel Drosophila myosin, but we have postponed numbering this class until a full-length cDNA is obtained.

In total, 13 myosin genes from Drosophila have now been identified (Figures 4A and 5). Of these, nine were previously known from cDNA clones (Morgan, 1995), whereas four are newly predicted from genomic sequences (Yamashita et al., 2000). We conclude that the Drosophila myosins are relatively evenly distributed among eight classes. There are two class I myosins (myosin-IA, 31DF; myosin-IB, 61F), two class II myosins (myosin heavy chain, 36A; “zipper”, 60E), and two class VII myosins (28B, “crinkled”, 35B). The remainder of the genes are distributed with one representative per class: III ("ninaC", 28A), V ("didum", 43C), VI ("jaguar/jar", 95F), XV (10A), XVIII (PDZ-myosin, 89B), and as-yet unnamed classes (29CD and 95E).

Three Additional Myosin Heavy Chains in C. elegans

Previous research in C. elegans identified four muscle myosin-II heavy chains (MHC-A, MHC-B, MHC-C, and MHC-D; Miller et al., 1986; Dibb et al., 1989) and two nonmuscle myosins (NMY-I and NMY-2). Three new conventional myosins have subsequently been predicted from genomic sequence as part of the C. elegans genome project. Each of these genes is represented by two or more ESTs but have not been widely recognized or studied.

The predicted myosin genes from cosmid F58G4 (acc. no. U50309) and cosmid F45G2 (acc. no. Z93382) are 60–63% identical to the other C. elegans muscle myosin heavy chains. The gene at F58G4.1 matches the structure of other class II myosins, whereas the predicted coding sequence for F45G2.2 may be truncated because it contains no IQ motifs and no coiled-coil, both of which would be expected in a class II myosin. Phylogenetic analysis suggests that F58G4.1 and F45G2.2 are the most divergent members of the C.
A myosin heavy chain gene from cosmid Y11D7A (acc. no. AL032632) is predicted to contain a catalytic domain with ~40% identity to other myosins, two IQ motifs, and a tail of segmented coiled-coil. By phylogenetic analysis of the predicted head domain sequence, Y11D7A.14 appears to be a highly divergent conventional myosin. It is quite surprising that this sequence branches before the fungal and amoeboid conventional myosins (Figure 3); it should be noted, however, that bootstrapping values at these nodes are too low to be conclusive and that confirmation of the predicted sequence by cDNA evidence will be crucial to determine the precise phylogenetic relationships on this part of the tree.

In addition to the nine class II myosins, a number of unconventional myosins have been previously identified in cDNA sequences and analysis of C. elegans genomic sequence (Baker and Titus, 1997). Many of the unconventional myosins are of unknown function, although their roles in cell motility may be hypothesized based on the functions of their homologues in other species. To summarize, C. elegans has two class I myosins (HUM-1 and HUM-5), one class V myosin (HUM-2), two class VI myosins (HUM-3 and HUM-8), a class VII myosin (HUM-6), a class IX myosin (HUM-7), and a class XII myosin (HUM-4). This brings the total number of myosin genes in C. elegans to 17, with 9 conventional myosins and 8 unconventional myosins (Figures 4B and 5).

**Dictyostelium**

Because of the use of Dictyostelium discoideum as a model system for studies of actin-based processes, there have been extensive efforts to identify myosins and their roles in this organism (Titus et al., 1994; Schwarz et al., 1999). Notably, one-half of the 12 known myosins in Dictyostelium are class I myosins (MyoA, MyoB, MyoC, MyoD, MyoE, and MyoK). This may reflect the important roles of myosins-I in processes involving membrane dynamics (Soldati et al., 1999). In contrast, Dictyostelium contains only one class II myosin (MhcA), one quite divergent class VII myosin (Myol), and one myosin similar to the structurally related classes V and XI (Myob). MyoM, a myosin with guanine nucleotide exchange activity for Rac1 small GTPases (Gessler et al., 2000; Oishi et al., 2000), may define a class unique to Dictyostelium. The myosins designated MyoF and MyoH are known only from short head domain sequences that do not allow unambiguous class assignment. MyoG and MyoL are potential myosin genes identified by low-stringency hybridization but have not yet been confirmed by sequencing. The known Dictyostelium myosins are highlighted on the phylogenetic tree in Figure 4C and discussed at length by Soldati et al. (1999).

**Plants**

Plants rely heavily on actin-based transport, and plant genomes are a particularly rich source of myosin genes. To assess the diversity of myosin expression in plants, we and others (Hodge and Cope, 2000) have scanned the now complete Arabidopsis thaliana genome. Remarkably, plants appear to lack both class I and class II myosins; instead, all of the known Arabidopsis myosins fall into just two classes, VIII and XI (Figures 3 and 4D).

The class VIII myosins (Knight and Kendrick-Jones, 1993) are thus far known only from plants and have approximately four IQ motifs, a putative coiled-coil region, and a class-specific tail of unknown function. In addition to the previously identified class VIII myosins ATM1 (acc. no. X69505) and ATM2 (acc. no. Z34292), two additional members of this class are encoded by FL14.6 (acc. no. AC007980) and M4I22.180 (acc. no. AL030978).

The class XI myosins have six IQ motifs and a tail with structural similarity to class V myosins (Kinkema and Schiefelbein, 1994). Analyses of Arabidopsis genomic sequence have dramatically inflated the representation of class XI, which previously consisted of MYA1 (acc. no. Z28389), MYA2 (acc. no. Z34293), and MYA3 (acc. no. Z34294) (Kinkema et al., 1994) but now contains at least 13 members (Figure 5). Class XI myosins are also present in Characean algae (Kashiyama et al., 2000), where they are thought to underlie the rapid cytoplasmic streaming of organelles, and in the biflagellate green alga Chlamydomonas (acc. no. AF077352).

The vascular plant Arabidopsis thus contains at least 17 myosin genes, 4 from class VIII and 13 from class XI. The absence of conventional myosins in plants may be related to the fact that during mitosis plant cells “wall-off” daughter cells via delivery and fusion of organelles rather than utilizing animal-like contractile cytokinesis (Field et al., 1999). Although plant myosins have not been intensively studied, the elaboration of class XI organelle/particle transporters in plants begs for more study.

**DISCUSSION**

The tentative inventories of myosin genes discussed here allow several important insights. First, they provide a list of candidate motors to underlie actin-based motility. Second, they help delineate the extent of myosin diversity. Third, they have led to the discovery of several novel and completely uncharacterized myosin genes. Finally, as discussed below, these inventories provide important information concerning the distribution and evolution of the myosins. The myosin diversity depicted in phylogenetic trees is summarized in Figure 6 to facilitate comparison across species.

Myosins can be more or less naturally divided into a number of classes based on the sequence relationships of their head domain sequences (Figure 3) and their tail domain structure. Each of these classes is presumably associated with a conserved set of biological functions, such as bipolar filament formation for the class II myosins and kinase activity for the class III myosins. Although class designation based on these sequence comparisons are very useful, it is not always precisely clear where class boundaries should be drawn. In these cases, consideration of additional factors such as N-terminal extensions or tail domain structure may be informative. As one example, classes VII, X, and XV (all of which contain MyTH4 and FERM domains) appear related and may form a larger superclass. Another problem that arises from inferring class relationships based solely on head domain sequences is that many classes ex-
tend via long branches almost to the center of the phyloge-netic tree and thus appear to have evolved relatively early. It should be noted that the precise pattern of gene and organismal evolution at the center of the tree is highly uncertain and that there is still considerable debate over the precise divergence patterns of crown eukaryotes, such as plants, fungi, cellular slime molds, and metazoans (Baldauf and Doolittle, 1997; Philippe and Adoutte, 1998).

Are Myosin Classes Ubiquitous or Restricted to Particular Phylogenetic Lineages?

One surprise from our analysis is that few myosins—not even the canonical class I and II myosins—are found in all eukaryotes. In retrospect, this makes some sense if certain classes are associated with specific functions that evolved only in a particular group of eukaryotes. If broadly defined,
the class V/XI myosins are the only class present in slime molds, fungi, plants, and animals (Figure 6) and would thus appear to have arisen very early in eukaryotic evolution. The class V myosins of vertebrates and the class XI myosins of Dictyostelium and plants are strikingly similar in overall structure, consisting of necks with six IQ motifs and a similar conserved globular tail domain. Because the class V/XI myosins are thought to underlie actin-based organelle/particle transport, and this is still the dominant form of organelle transport in plants and certain algae (Yamamoto et al., 1999), it is likely that actin-based organelle transport also evolved very early.

Because budding yeast contain only classes I, II, and V (see above), classes I and II either evolved after the divergence of plants or were lost from the plant lineage. In the cellular slime mold Dictyostelium, there is a dramatic expansion of the myosins-I (Soldati et al., 1999). This specialization in class I myosins may be related to the increased membrane dynamics required by an amoeboid lifestyle. Additionally, a class VII-like myosin with MyTH4 and FERM domains (Myol) is required for phagocytosis in Dictyostelium (Titus, 1999). As mentioned above, the precise pattern of early eukaryotic radiation is a matter of some debate. Thus, depending on the order in which the crown eukaryotes diverged, class VII either evolved after the divergence of fungi or were lost from the yeast lineage. A much wider sampling of eukaryotic genomes will be required to determine which case is more likely.

Most other myosin classes, far from being ubiquitous, are apparently restricted to particular groups of eukaryotes. For example, a number of myosin classes appear to have originated early in the divergence of multicellular animals. These innovations resulted in more complex roles for myosins, including pointed-end or “backward” movement for the class VI myosins (Wells et al., 1999), which are found in fly, worm, and humans but not in lower eukaryotes. Several metazoan myosin classes are further restricted to subsets of organisms. The class III kinase myosins are associated specifically with the retina in flies, horseshoe crabs, and humans; C. elegans, which lacks eyes, also lacks class III myosins. Surprisingly, class IX myosins with GTPase-activating protein (GAP) activity for Rho GTPases are found in both vertebrates and C. elegans but are absent from Drosophila. The class XII myosin (HUM-4), which has a tail containing two MyTH4 domains, is thus far known only from C. elegans but may actually correspond to an otherwise missing worm myosin-XV. At least two myosin classes, class X (Berg et al., 2000) and the newly discovered class XVI (with N-terminal ankyrin repeats), appear to be vertebrate-specific, whereas the short tailed Drosophila myosin at 29E may represent a novel fly-specific myosin. Whether these lineage-specific genes are in fact novel elaborations of the myosin superfamily or were lost from other metazoan species is unknown and will require a wider sampling of metazoan genomes.

In addition to the metazoan-specific myosins, several other classes appear to have quite limited phylogenetic distributions. These include class IV (a MyTH4-containing myosin) from Acanthamoeba castellani (Horowitz and Hammer, 1990), class XIII from the green alga Acetabularia cliftonii (acc. nos. U94397 and U94398), class XIV (Heintzelman and Schwartzman, 1997) from the parasites causing toxoplasmosis (Toxoplasma gondii) and malaria (Plasmodium falciparum), and class XVII (chitin synthase myosin) from Aspergillus nidulans (Fujitwara et al., 1997). In addition, a partial myosin sequence from the ciliate Tetrahymena thermophila remains unnumbered (Garces and Gavin, 1998). It is likely that at least some of the myosins found in protozoans represent specializations for organism-specific functions such as the gliding motility of Toxoplasma. Examples such as these raise the possibility that the protozoa, which diverged early in eukaryotic evolution, have evolved myosin classes found only in those lineages. Given the vast diversity of protozoan phyla, it is likely that these organisms will provide additional novel myosins.

Lineage-Specific Specialization in Particular Myosin Classes

It is clear that certain myosin classes have undergone lineage-specific expansions (Figures 3, 4, and 6). For example, the lineage leading to plants underwent a major radiation of the class XI myosins. In contrast, metazoans express relatively few members of the structurally similar class V myosins, with Drosophila and C. elegans possessing one gene each and vertebrates containing three genes. Given the importance of actin-based organelle transport, it is perhaps surprising that fly and worm use only one class V myosin. Similarly, the slime mold Dictyostelium appears to have specialized in class I myosins, whereas Drosophila and C. elegans contain only two class I myosins each. Drosophila even appears to lack an SH3 domain-containing class I myosin. This suggests that metazoan class I and class V myosins may function as generalists, whereas the Dictyostelium class I and Arabidopsis class XI myosins were expanded for specific functions in those organisms.

Another example are the conventional (class II) myosins, which are involved in muscle contraction, stress fiber contractility, and cytokinesis. In the metazoan lineage, there is obviously an early and deep division between the nonmuscle/smooth muscle myosins and the striated muscle myosins. The Drosophila genome contains a single muscle myosin gene and a single nonmuscle myosin-II gene and creates diversity through alternate splicing to generate specific isoforms (Morgan, 1995). In contrast, the conventional myosins make up a large fraction of the C. elegans (9 of 17) and vertebrate (15 of 40) myosin genes. Importantly, the expansions of muscle myosins in C. elegans and in humans appear to be the result of independent gene radiations. Thus, vertebrates and worms appear to have developed diversity of muscle function by generating multiple myosin heavy chain genes, whereas the fruit fly maintains a single muscle myosin gene.

Myosin Diversity in Vertebrates

From inspection of the vertebrate lineage, several trends are apparent. First, at ~40 genes, the total number of myosins in humans is more than twice as large as in C. elegans (17) or Drosophila (13). Second, although the myosins-I were the first unconventional myosins to be discovered (Pollard and Korn, 1973), very little is known about one-half of the vertebrate members of this class. There are now eight class I myosins in humans with two members for each of the previously defined myosin-I subclasses. In addition, several other myosin classes (III, VII, IX, XV, and XVIII) contain two members

Vol. 12, April 2001 791
each in humans and the branch lengths separating these pairs of myosins have similar lengths (Figure 1). This raises the possibility that these pairs of genes resulted from a genome duplication in the vertebrate lineage. It is important to realize that these pairs of unconventional myosins exhibit at most 50–60% identity to one another and thus may have significantly different properties. In this regard, different class II myosins that are only ~50% identical (such as muscle myosins versus nonmuscle myosins) can have biochemical properties such as ATPase rates and sliding filament velocities differing by as much as 50-fold. Furthermore, small changes in the myosin motor domain can lead to dramatic changes in motor function, such as reversal of direction as seen with myosin-VI. Finally, although diversity among myosins once appeared to be largely related to divergent tail domain structure, several myosins are now known to have large N-terminal extensions. In addition to the class III/ninac myosins with their N-terminal kinase domain, myosin-XV has an immense ~1200-aa N-terminal extension of unknown function, the new class XVI myosin has N-terminal ankyrin repeats, and the class XVIII MysPDZ has an N-terminal PDZ domain.

Resolution of Conflicts in Nomenclature
As is the case for much of postgenome biology, the nomenclature system for myosins can be complex and confusing. This is due to both the large number of myosins and defects in nomenclature that yield many synonyms and similarly sounding names for myosins that may or may not be similar. However, the completion of genomic sequences from a variety of organisms will eventually provide an inventory of these genes and should lead to greater clarity of nomenclature. We have included additional discussion of the vertebrate myosin-I nomenclature in Supplemental Materials available online at www.molbiocell.org. When in doubt, the use of sequence accession numbers can act in proxy as unambiguous identifiers. For vertebrate myosins, we recommend use of the official HUGO human gene names (http://www.gene.ucl.ac.uk/nomenclature/). These are in the form of MYH1–13 (myosin heavy chain) for the class II myosins and MYO1–18 for unconventional myosins (with the Arabic numeral corresponding to the myosin class). The official gene names for mouse homologues are the same as in humans, except that the mouse nomenclature uses lowercase letters (Myh1–13 and Myo1–18) in gene names (http://www.informatics.jax.org/).

It may also be prudent to finish numbering the myosin classes from fully sequenced organisms and to at least temporarily consider myosins known only from sequence fragments and/or restricted to lower eukaryotes as “orphan” classes. To assist researchers in deciphering the nomenclature, we have included a list of the myosins from the organisms discussed here along with synonyms and accession numbers (Figure 5). We have intentionally left several newly discovered myosin genes as “orphan or not numbered” until complete cDNA sequences confirming their structures are obtained.

Conclusions
We report here an initial census of myosin genes in humans and other model organisms. The *C. elegans* and *Drosophila* genomic sequences are finished and relatively stable, but the human genome continues to undergo further sequencing and refinement. Additional myosin genes may exist in currently unsequenced (or highly fragmented) regions of the human genome or may escape detection by sequence comparisons if they are highly divergent. Thus, although we believe that the majority of myosin genes are represented in this analysis, it is possible that additional human myosin genes will be discovered in the future. At the most basic level, it will be important to obtain the full-length cDNA sequences for myosins predicted from genomic sequence to determine their precise structure and splice forms. It is also critical to realize that most of the myosins in this inventory have never been studied in terms of their structure, light chains, or biochemical and motor properties. Nevertheless, characterization of novel myosins has already provided us with a “reverse” myosin as well as a processive myosin, and it is likely that further surprises and insights into basic motor protein biology lie ahead. Myosins are fundamentally important in cell and organismal physiology, and defects in myosins have already been shown to underlie hereditary forms of cardiomyopathy, deafness, and blindness. With the completion of the various genome projects, we have now begun to attain a complete inventory of the candidate motors for actin-based motility.

Notes added in proof. Manuscripts describing the sequence of the human genome were recently published by the International Human Genome Sequencing Consortium (Nature 409, 860–921) and Celera Genomics (Science 291, 1301–1351). Our most recent analyses of these sequences identified no additional myosin genes. Suprisingly, many well-known myosins were absent, truncated, or fragmented into multiple predicted genes in the preliminary automated annotations of the human genome.

Preliminary analysis of myosin-XVb suggests that this gene might be a transcribed pseudogene (Erich Boger and Tom Friedman, personal communication). It is possible that other myosin genes predicted from genomic sequence are also pseudogenes, however, recent work has shown that the HA-2 minor histocompatibility antigen involved in graft-versus-host disease is derived from the novel class I myosin predicted from chromosome 7 genomic sequence, confirming that MYO1G is transcribed and translated. Although the function of “myosin-ig” is still unclear, its expression appears to be limited to hematopoietic cells (Rich Pierce and Victor H. Engelhard, personal communication).

ACKNOWLEDGMENTS
The catalogue of myosins summarized here is the result of many years of effort by our colleagues in myosin and genome research. We regret that many valuable contributions had to be left out or could only be cited via reviews and accession numbers. We especially thank David Corey (myosin-VIIb), Richard Cameron and Krishna Patel (class XVI myr8), and Meg Titus (*C. elegans* and *Dictostelium* myosins) for valuable discussions. We also thank Tom Pollard and Olga Rodriguez for helpful comments and suggestions. The authors and this work were supported by National Institutes of Health grant DC03299.
REFERENCES


