

Myosinome: A Database of Myosins from Select Eukaryotic Genomes to Facilitate Analysis of Sequence-Structure-Function Relationships

Divya P. Syamaladevi^{1,2}, Margaret S Sunitha¹, S. Kalaimathy¹, Chandrashekar C. Reddy¹, Mohammed Iftexhar¹, Shaik N. Pasha¹ and R. Sowdhamini¹

¹National Centre for Biological Sciences (NCBS-TIFR), GKVK Campus, Bangalore, India. ²Sugarcane Breeding Institute (SBI-ICAR), Coimbatore, India. Corresponding author email: mini@ncbs.res.in

Abstract: Myosins are one of the largest protein superfamilies with 24 classes. They have conserved structural features and catalytic domains yet show huge variation at different domains resulting in a variety of functions. Myosins are molecules driving various kinds of cellular processes and motility until the level of organisms. These are ATPases that utilize the chemical energy released by ATP hydrolysis to bring about conformational changes leading to a motor function. Myosins are important as they are involved in almost all cellular activities ranging from cell division to transcriptional regulation. They are crucial due to their involvement in many congenital diseases symptomatized by muscular malfunctions, cardiac diseases, deafness, neural and immunological dysfunction, and so on, many of which lead to death at an early age. We present Myosinome, a database of selected myosin classes (myosin II, V, and VI) from five model organisms. This knowledge base provides the sequences, phylogenetic clustering, domain architectures of myosins and molecular models, structural analyses, and relevant literature of their coiled-coil domains. In the current version of Myosinome, information about 71 myosin sequences belonging to three myosin classes (myosin II, V, and VI) in five model organisms (*Homo Sapiens*, *Mus musculus*, *D. melanogaster*, *C. elegans* and *S. cerevisiae*) identified using bioinformatics surveys are presented, and several of them are yet to be functionally characterized. As these proteins are involved in congenital diseases, such a database would be useful in short-listing candidates for gene therapy and drug development. The database can be accessed from <http://caps.ncbs.res.in/myosinome>.

Keywords: myosin, Myosinome, myosin II, myosin V, myosin VI, myosin database

Bioinformatics and Biology Insights 2012:6 247–254

doi: [10.4137/BBI.S9902](https://doi.org/10.4137/BBI.S9902)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Mobility in higher order organisms is known to be controlled by muscles. Muscles are made up of muscle fibrils, comprising thin and thick filaments, which, in turn, comprised actin and myosin oligomeric assemblies.¹ Apart from the conventional fibre-forming myosins in muscles, there are numerous non-muscle myosins localized specifically in various compartments and cytoplasmic or nuclear regions of the cell. Preliminary sequence-based phylogenetic analysis suggested myosins are subdivided into 24 different classes² and are denoted from myosin I to XXIV.

Myosins are molecular motors that bring about a movement by converting chemical energy to kinetic energy through ATP hydrolysis and conformational changes. The resultant motion of the motor is utilized in numerous mobility-associated functions inside the cell through self-assembly as well as association with various subcellular cargoes and molecules. For instance, myosin type II self-assembles into oligomeric structures at the cell plate and participates in cleavage furrow formation during cell division.^{3,4} Myosin type V forms dimers and participates in processes like melanosome and vesicle transport.⁵ Myosin V proteins are also reported to take part in growth cone formation in neuronal cells.⁶ Yet another interesting myosin, myosin VI, is the only one known to have retrograde motility towards the minus end of actin filament. These myosins act as pressure sensors at the cochlear hair cells of the inner ear. Indeed, a genetic mutation resulting in a non-functional myosin VI gives rise to congenital hearing disability.

Myosins have the general structure of a head domain followed by a lever arm (neck region) and a C-terminal tail. The N-terminal head possesses the ATPase activity and also retains actin-binding sites.⁷⁻⁹ In the actin-bound form of myosin, ATP hydrolysis effects a translation of conformational changes leading to a swinging motion of the lever arm. The head domain attached to the swinging lever arm finds the next binding site on the actin filament resulting in a stepping forward kind of motion.¹⁰ The molecular mechanism of head domain function is more or less conserved, whereas the tail regions differ considerably in their sequence, fine structure, and function.¹¹

Myosins are relatively large proteins of 1000 to 2000 base pairs long. Sequence similarity-based phylogeny of the head domain sequences not only

display different classes of myosins,¹²⁻¹⁴ the various classes are hypothesised to be derived from an ancestral myosin, which, in turn, is evolved from the P-loop NTPases.¹⁰ The overall domain architecture of myosins in many cases differs considerably at the tail region as in myosin types III, XII, XIV, and XVI, which do not possess an alpha helical region that can fold into a coiled coil. Instead, the head domain is directly connected to a globular tail domain.¹⁵ Myosinome database domain boundaries could be assigned using multiple approaches, and a consensus domain architecture is provided for each of the sequences.

The variations among myosins are not only due to differences in domain architectures. There have been incidences of insertions and deletions as well as amino acid substitution events resulting in functional variety. The unique insert domain of myosin VI neck is one such example where the insertion has resulted in reverting the direction of motion of the protein. The tail regions in most of the myosin types possess an alpha helix that may dimerize into a coiled coil structure that precedes a globular cargo-binding domain. The coiled coil region with a characteristic heptad repeat pattern where the 1st and 4th residues of the heptad (the a and d positions) are expected to be hydrophobic. This arrangement of amino acids forms a hydrophobic seam and results in the formation of a coiled coil through the knobs-into-holes arrangement of side chain packing, forming the hallmark or structural signature of a coiled coil.¹⁶ Many existing coiled coil prediction methods predict a continuous coiled coil and also are capable of reporting stutters and stammers. But experiments have proved that the predicted coiled coils in myosins are not necessarily interacting strongly to form perfect coiled coils. Instead, they may interact in parallel or even fold on to form trimers.¹⁷⁻¹⁹ Hence, the three-dimensional structure was obtained by imposing a coiled coil structure through a homology modeling tool (MODELLER)²⁰ at the predicted coiled coil regions in all sequence entries in the database and later analyzed for the hydrophobic content of the central seam. Thus, the Myosinome database was developed as a platform to integrate the knowledge base and predictions on the myosin sequence, architecture, structure, and function information from various sources.

Construction and Content

Data collection and integration

For sequence data collection, we employed a sensitive sequence search tool, position-specific iterative blast (PSI-BLAST). We set up the searches using BLAST+ package²¹ against the proteomes of five model organisms: human, mouse, *Drosophila*, *C. elegans*, and yeast. Since the Myosinome database focuses on the domain architectures and specifically on the coiled coil, we restricted our searches to myosin II, IV, V, and VI. Searches were performed using 24 myosin sequences that were obtained from an initial text search at the National Center for Biotechnology Information (NCBI). At least one sequence from each of the selected myosin classes were employed as a query in such sequence searches. We employed a relatively relaxed E-value cutoff of $1e-5$ in order to increase the coverage. A list of unique hits was pooled derived from all the searches using 24 query sequences.

As a second step in data collection, we validated the sequences obtained from PSI-BLAST.²² We adopted a three-fold validation. At first, using a string-based annotation filter script (genepeptscript1), we separated the annotated myosins from PSI-BLAST hits into myosin types II, IV, V, and VI using the GenePept data from NCBI as input. This filter also categorized the annotated myosin classes other than II, IV, V, and VI into a separate category. The sequences classified in this way were validated, in the second step, by assessing the evolutionary relationships of their head domains. For this, we used a phylogenetic clustering method, neighbor joining, in the PHYLIP package.²³ Head domains of PSI-BLAST hits were seeded into a set of sequences comprising other known myosins of 19 different classes and a few non-myosin sequences, such as helicases, ATPases, and Kinesins. These non-myosins acted as outgroups and helped in ensuring the selection of myosins that grouped with other myosin sequences alone. This sequence similarity-based clustering was useful in assigning the myosin classes to each of the hits and cross validating before we organized them into a database. As a third step of validation, we employed an integrated domain architecture definition obtained from various databases to provide a more comprehensive and complete definition of domains to the sequences obtained. The domain definitions were obtained from CDD,²⁴

PFAM,²⁵ COILS,²⁶ and so on. Boundary definitions were obtained from the SMART database. A detailed schema for sequence retrieval and validation protocol is provided in Figure 1.

Coiled coil regions—the dimerization motifs—of each of the identified myosin sequences were modeled using tropomyosin as a template using MODELLER software. Whereas the length of coiled coil was much longer than the template in cases like myosin II and V, models were generated by a protocol that involves repeating the same template over and over again with an overlapping region spanning 70 residues (Margaret *et al*, oral communication; December, 2010). The models generated this way were analyzed for charged patches using an in-house script. Charged patches are defined based on the density of charged residues in every heptad. An empirical scoring scheme was applied to calculate the charge density by counting the number of charged residues within a distance cutoff of 12Å. The charge density positively correlated with the total number of charged residues. Hence, a gradient color scheme representation was adopted to demarcate dense charged patches from weak charged patches. Apart from our own models, available crystal structures, NMR structures, and cryo-electron microscopy structures were also collected from Protein Data Bank (PDB)²⁷ through keyword searches and literature-based searches.

Sequence and structural level data should be useful in the case of myosins whose functions are not yet experimentally characterized. However, numerous experiments have been performed to understand

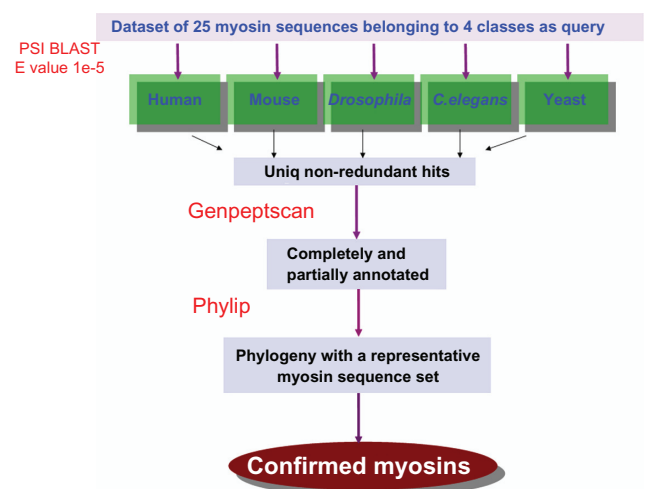


Figure 1. Schema of sequence retrieval and validation protocol.

this class of motor protein, and a consolidation of this functional information will be enlightening for finding future directions. Function of a molecule is inseparable from the pathways that it takes part in. Hence, we provide available KEGG pathway maps for different classes of myosins in this database.²⁸ Finally, a list of related literature is also presented along with the PubMed links. Literature collection involved keyword searches in PubMed. IDs of the literature relevant to individual protein accession number were obtained using automated scripts from the GenePept file.

Database implementation

All the results are delivered to the user through a web application. The database was implemented in MySQL and the graphical user interface (GUI) was designed in HTML. The communication between the GUI and the back-end database in order to retrieve information from the stored myosin data upon user query was achieved using PHP codes. The back-end BLAST parser scripts were written in Perl, and domain architecture drawings were done using HTML style sheets and mouse over functions. Phylogenetic trees, models of the coiled coils, and charge distributions are provided as images. Coiled coil models are displayed in more friendly and interactive fashion by using JMOL applets. The entire application is set up behind an Apache web server using LAMP (Linux-Apache-MySQL-PHP/Perl) (See Fig. 2 for a schema of Myosinome web applications).

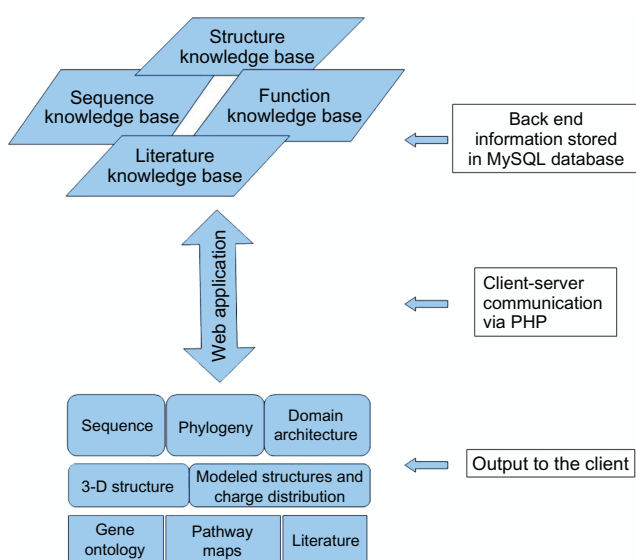


Figure 2. Schematic representation of the Myosinome web application.

Utility and Discussion

Browsing

It is possible to browse through the Myosinome database in three ways: search using keyword, browse by organism and search using sequence.

In the keyword search box, users can provide different texts viz, myosin, myosin type II, or the accession number directly. This facilitates searching when a user knows a specific sequence accession number or when a user wants to search all the sequences of a particular myosin type in all the genomes in the Myosinome database.

The home page of the database provides a facility to submit partial or complete sequences in FASTA format by the user to search for similar sequences in the database. A background BLAST search would be performed and display the related sequences available in database.

Alternately, there is a panel with the list of organisms provided, each of which, upon mouse click, will lead to the list of myosins within the whole genome along with their accession numbers. The bottom panel in the home page tabulates the general information on myosins in four columns as myosin types, organisms, PDB structures, and KEGG maps and also provides reference to representative classical articles to each type of myosin (Fig. 3).

Content

The accession numbers of individual myosins are linked to a page where detailed information associated with the sequence is provided in a tabular format (Fig. 4). Important information in this table is displayed in two main panels. The first panel covers sequence-, structure-, and function-related data in three fields, and the bottom panel contains the most relevant literature corresponding to the protein of interest. The foremost field of the first panel provides the sequence and a phylogenetic tree. The phylogeny includes classical myosins from different subtypes with the position of the query highlighted in red color font to recognize its evolutionary position amid the other myosin subtypes. The domain architecture of sequence is shown as a block diagram and a mouseover option provides the domain boundaries for each of the domains. This is followed by a structure section where the three-dimensional model of the coiled coil domain



Introduction

Myosins are a large family of motor proteins found in eukaryotic cells. The term myosin was originally used to describe a group of similar, but nonidentical, ATPases found in striated and smooth muscle cells. Later on numerous intracellular motility functions of these motor proteins were discovered. Myosins are classified in to different subclasses based on the phylogenetic analysis of N-terminal ATPase domain. Myosinome database provides information on sequence, structure and function of myosin classes II, V and VI from five different organisms.

FASTA Sequence

Submit Query Reset

Myosin Type	Organism	PDB-Structures	KEGG pathways	Review Articles
Myosin type II	<i>Homo sapiens</i>	1b7l 1br1 1br2 1br4		Of mice and men: Relevance of cellular and molecular characterizations of myosin IIA to MYHB-related human disease. Cell Adh Migr. 2007 1(3):152-5
	<i>Mus musculus</i>	1d0x 1d0y 1d0z	map04270	
	<i>Drosophila melanogaster</i>	1d1a 1d1b 1d1c	map05410	
	<i>Caenorhabditis elegans</i>	1dfk 1dfi 1fmv 1fmw	map04290	
	<i>Saccharomyces cerevisiae</i>	1g8x 1l84 1jwy 1jx2	map04810	
		1kk7 1kk8 1kqm	map04530	
		1kwo 1l2o 1lxx 1lvk	map04666	
		1m8q 1mma 1mmd	map04670	
		1mmg 1mmn 1mnd	map05416	
		1mne 1mwv 1o18		
		1o19 1o1a 1o1b		
		1o1c 1o1d 1o1e		
		1o1f 1o1g 1qvl		
		1s5g 1sr6 1vom		
		1w9l 1w9j 1w9k		
	1w9l 1yv3 2aka			
	2ec6 2ekv 2ekw			
	2jhr 2jij 2mys 2os8			
	2oig 2ovk 2oy6			
	3bz7 3bz8 3bz9			
	3dtp 3dtr 3i5g 3i5h			
	3i5i			
Myosin type V	<i>Homo sapiens</i>			Myosin V from head to tail . Cellular and Molecular Life Sciences 05(9):1378-1389
	<i>Mus musculus</i>			
	<i>Drosophila melanogaster</i>	1oe9 1w7l 1w7j		
	<i>Caenorhabditis elegans</i>	1w8j 2bkh 2bki 2dts		

Figure 3. Screenshot of the Myosinome homepage indicating the browse options and links.

of the sequence is provided for download as well as for graphical display. Besides, a graphical representation of charged patches is displayed (Fig. 5; please also see Section on Construction and Content) along with the hydrophobic residue distribution at the central region at the a and d positions of the heptads. These would give user clear under-

standing of the relative stability of protein-protein interaction sites and would be of predictive value in designing point mutations for functional analysis. Links to context-dependent help pages are provided where the methodology is explained in detail allowing the user to analyze the coiled coils for designing experiments.

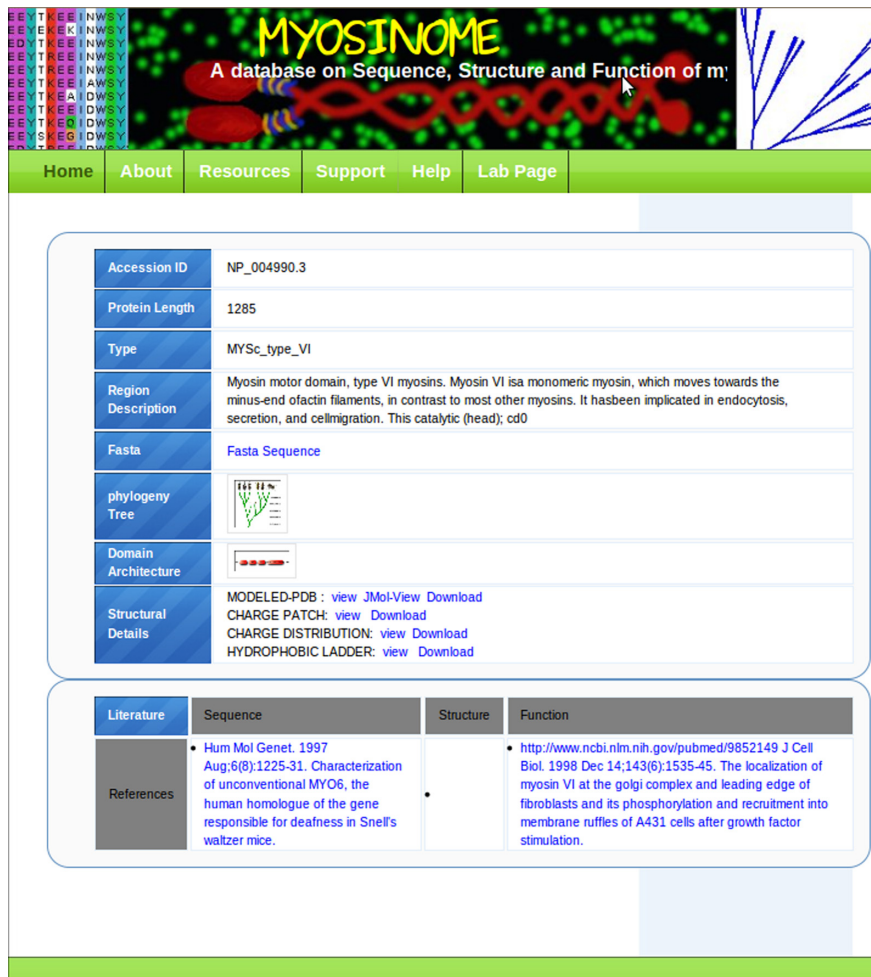


Figure 4. Screenshot of data table for a single myosin.

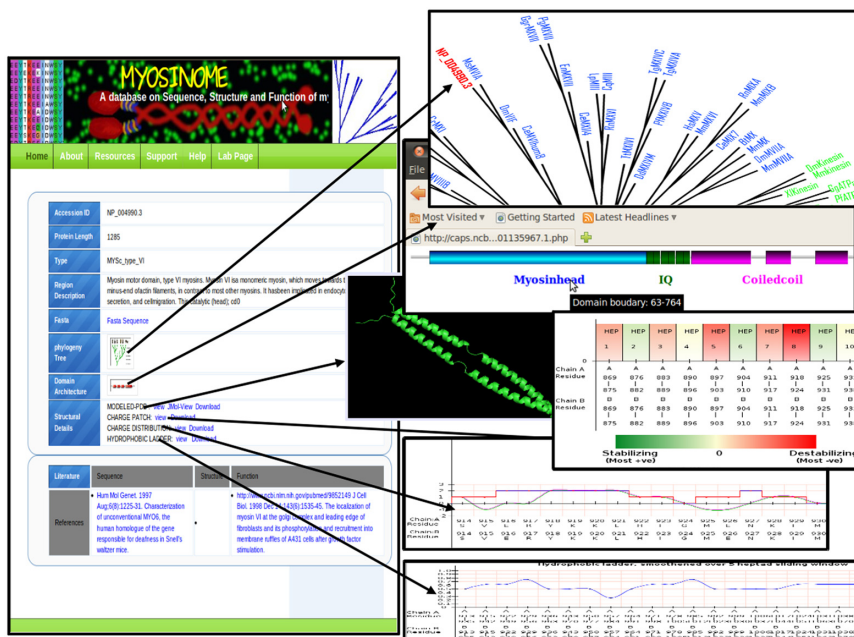


Figure 5. Data content on individual myosins.

Conclusions

Myosinome is a unique repository of sequence, structural, and function data of a limited number of myosin types from select model organisms. It provides downloadable files and in-depth information on sequence, structure, functions, pathways, and relevant literature. It focuses on aspects of variability such as sequence level differences as well as domain architectural level changes. Sequence level changes are examined through the analysis of the coiled coil, which is a crucial domain facilitating self-assembly and interaction with other proteins. The community participation portal in the Myosinome web application is a feature that allows expansion of the database. It is a fully integrated resource with external data sources for bridging maximum knowledge. The interface of Myosinome is navigation-friendly, efficient, and easy to use and, thereby, accelerates knowledge exploration and accumulation. Myosinome would be useful in performing sequence similarity searches against a database of known myosins belonging to selected classes and also in designing mutational experiments that can further expand the functional knowledge base of myosins.

Author Contributions

RS and DPS devised the framework of database. DPS, KS, NS, MS, CC and RS contributed to the analysis of sequence and structures of myosins. MI and DPS developed the back-end and web interface. RS and DPS participated in discussions and manuscript preparation.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Funding

This work was supported by Human Frontier Science Program (HFSP) Grant (Code: RGP0054/2009-C/).

Acknowledgements

We thank NCBS for infrastructural facilities. We thank Prof. Jim Spudich for useful discussions and Mr. Anu Nair for help with scripts on block diagrams of domain architectures.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of

compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Fischman DA. Electron microscope study of myofibril formation in embryonic chick skeletal muscle. *J Cell Biol.* 1967;3:557–75.
2. Foth BJ, Goedecke MC, Soldati D. New insights into myosin evolution and classification. *Proc Natl Acad Sci U S A.* 2006;103(10):3681–6.
3. Yumura S, Fukui Y. Reversible cyclic AMP-dependent change in distribution of myosin thick filaments in Dictyostelium. *Nature.* 1985;314:194–6.
4. Zang J, Spudich JA. Myosin II localization during cytokinesis occurs by a mechanism that does not require its motor domain. *Proc Natl Acad Sci U S A.* 1998;95(23):13652–7.
5. Hasson T. Myosin VI: two distinct roles in endocytosis. *J Cell Sci.* 2003;116:3453–61.
6. Desnos C, Huet S, Darchen F. ‘Should I stay or should I go?’: myosin V function in organelle trafficking. *Biol Cell.* 2007;99:411–23.
7. Bozler E, Prince JT. The control of energy release in extracted muscle fibers. *J Gen Physiol.* 1953;37(1):53–61.
8. Emoto Y, Kawamura T, Tawada K. Characterization of the ATPase active site in myosin subfragment-1 with the use of vanadate plus ADP as a reversible “affinity-labeling” reagent: evidence for heterogeneity in the active sites. *J Biochem.* 1985;98(3):735–45.
9. Dalgarno DC, Prince HP, Levine BA, Trayer IP. Identification of a surface actin-binding site on myosin. *Biochim Biophys Acta.* 1982;707(1):81–8.
10. Purcell TJ, Morris C, Spudich JA, Sweeney HL. Role of the lever arm in the processive stepping of myosin V. *Proc Natl Acad Sci U S A.* 2002;99(22):14159–64.
11. Krendel M, Mooseker MS. Myosins: tails (and heads) of functional diversity. *Physiology (Bethesda).* 2005;20(4):239–51.
12. Thompson RF, Langford GM. Myosin superfamily evolutionary history. *Anat Rec.* 2002;268:276–89.
13. Foth JB, Goedecke MC, Soldati D. New insights into myosin evolution and classification. *Proc Natl Acad Sci U S A.* 2006;103(10):3681–6.
14. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature.* 2005;436(25):1113–8.
15. Goodson HV, Dawson SC. Multiplying myosins. *Proc Natl Acad Sci U S A.* 2006;103(10):3498–9.
16. Crick FHC. The Fourier transform of a coiled-coil. *Acta Crystallogr.* 1953;6:685–97.
17. Knight PJ, Thirumurugan K, Yu Y, et al. The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J Biol Chem.* 2005;280:34702–8.
18. Mukherjea M, Llinas P, Kim H, et al. Myosin VI dimerization triggers an unfolding of a three-helix bundle in order to extend its reach. *Mol Cell.* 2009;35(3):305–15.
19. Spudich JA, Sivaramakrishnan S. Myosin VI: an innovative motor that challenged the swinging lever arm hypothesis. *Nat Rev Mol Cell Biol.* 2010;11(2):128–37.



20. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234:779–815.
21. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
22. Syamaladevi DP, Kalaimathy S, Pasha N, Subramonian N, Sowdhamini R. A three-step validation following genome-wide data mining for myosin family members improves search efficiency. In Proceedings of IEEE ICDM 2011 Workshop on Biological Data Mining and its Applications in Healthcare; Dec 10, 2011; Vancouver, Canada.
23. Felsenstein J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics.* 1989;5:164–6.
24. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 2004;32:W327–31.
25. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2010;38:D211–22.
26. Lupas A, Dyke MV, Stock J. Predicting coiled coils from protein sequences. *Science.* 1991;252:1162–4.
27. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;8:235–42.
28. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27–30.