A myosin family tree

Tony Hodge¹ and M. Jamie T. V. Cope²

¹Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
(tph@mrc-lmb.cam.ac.uk)
²Drubin/Barnes Lab, 401 Barker Hall, University of California, Berkeley CA 94720, USA
(jcope@mac.com)

This unrooted phylogenetic tree of the myosin superfamily (see insert) is derived from an alignment of 139 members of the myosin superfamily. The alignment compared the core motor domains (equivalent to residues 88-780 of chicken skeletal myosin II) of each myosin, using distance matrix analysis performed with the Clustal-W package. The exceptions, shown with a dotted line, are SsVIIa, which is a partial sequence as reported in the databases, and Hs MysPDZ. The latter is reported as a complete coding sequence but has a truncated N terminus starting some 52 residues into the core motor region (i.e. residue 140 of chicken skeletal myosin II). These shorter sequences have no significant effect on the branching order of the rest of the tree.

Optimally, when one produces a phylogenetic tree from such an alignment, any positions in the alignment with gaps are excluded. Such a strategy would have excluded a large proportion of the data because of the large number of sequences. Positions with gaps were therefore included in the tree shown here. Comparison with a tree derived while excluding gaps showed some differences in branching order but only within classes. The reliability of the tree structure was tested by several methods: (1) bootstrapping (repeated redrawing of the tree structure, 1000 trials in this case) gave confidence levels for branching order (see below); (2) alignments were re-ordered randomly and alphabetically by species or by taxa, these treatments giving trees with identical branching order within the classes. (3) Analysis of the alignment using protpars from the Phylip package (a maximum parsimony method) produced a tree with a similar branching order for most classes, the main exceptions being the single sequence classes.

The tree being unrooted, the relationships
between classes as shown by the branching order at the centre of the tree is unreliable, but evolutionary information can be derived within a class. Each class is defined by the first node represented in >90% of bootstrap trials starting from the centre of the tree. The inclusion of myosin sequence data recently added to the public databases has resulted in the clustering together at such a node of some of the more disparate examples based on their motor domains (Classes III, XII, XVI and the chitin synthase containing myosins). While this would normally be taken to define a class, the low sequence similarity (long branches), lack of significance for this grouping obtained by maximum parsimony algorithms and general dissimilarity between complete molecules argue against such a classification. Analysis of the sequence identity of the two chitin synthase myosins (Pg csm1 and En csmA) shows they are orthologues; they also group together by distance matrix analysis with >90% confidence. This evidence allows us to define a new class (XVII), and the branches have been coloured accordingly.

The molecular cartoons serve to indicate possible molecular structure, especially the expected single- or double-headed nature of the myosins. Regarding the myosin XIII cartoon, the ‘?’ denotes one of the sequences (Acl myo1) having a surprisingly short tail, which may reflect a sequence truncation.

The complete alignment, the bootstrapping data, references and links for the software packages and hyperlinks to database entries for all the myosins included in the analysis, along with other myosin-related information, can be found at the Myosin Homepage: www.mrc-lmb.cam.ac.uk/myosin/

**Cell Science at a Glance on the Web**
Electronic copies of the full-size poster insert are available in the online version of this article (see www.biologists.com/jcs). Files in several formats are provided and may be downloaded for use as slides.
A Myosin Family Tree
Tony Hodge and Jamie Cope