ARTICLES

Myosin domain evolution and the primary divergence of eukaryotes

Thomas A. Richards[1,2]† & Thomas Cavalier-Smith[2]

Eukaryotic cells have two contrasting cytoskeletal and ciliary organizations. The simplest involves a single cilium-bearing centriole, nucleating a cone of individual microtubules (probably ancestral for unikonts: animals, fungi, Choanozoa and Amoebozoa). In contrast, bikonts (plants, chromists and all other protozoa) were ancestrally biciliate with a younger anterior cilium, converted every cell cycle into a dissimilar posterior cilium and multiple ciliary roots of microtubule bands. Here we show by comparative genomic analysis that this fundamental cellular dichotomy also involves different myosin molecular motors. We found 37 different protein domain combinations, often lineage-specific, and many previously unidentified. The sequence phylogeny and taxonomic distribution of myosin domain combinations identified five innovations that strongly support unikont monophyly and the primary bikont/unikont bifurcation. We conclude that the eukaryotic cenancestor (last common ancestor) had a cilium, mitochondria, pseudopodia, and myosins with three contrasting domain combinations and putative functions.

Myosins bind to actin, hydrolysing ATP to produce physical force, and are fundamental in eukaryotic cytokinesis, organellar transport, cell polarization, intracellular transport and signal transduction[1,2]. They evolved, like microtubules, during the origin of eukaryotes[3]. Their head domains, containing the ATPase and actin-binding activities, are connected to a range of amino-terminal and carboxy-terminal domains[4], corresponding to the variety of molecular cargos that myosins bind and move. Sequence phylogeny and protein domain combinations have previously been used to establish 18 myosin 'classes'[4–7], although additional myosin types have also been reported[8]. The function of many myosin 'classes' has been characterized and is distinct, but full functional properties are unknown[4] for others or for currently unclassified myosins. Myosin and the related kinesin[9] gene families along with protein-synthesis elongation factors form the TRAFAC class of the P-loop GTPases that originated by the deletion of strands 6 and 7 in the GTPase core and the addition of two N-terminal strands[10].

Studying the diversification of eukaryote-specific molecular motors that interact closely with the cytoskeleton may be particularly fruitful for understanding phylogenetic patterns, the cellular apparatus and the functional attributes of early eukaryotes. Gene families with numerous paralogues (genes related by duplication but with non-identical functions), such as myosins, are often considered unhelpful for reconstructing ancient evolutionary relationships because of their very complexity. However, with sufficient taxon sampling and reliable sequence phylogenies, patterns of sequence synapomorphies (derived character states shared by two or more taxa) and paralogue distribution can be used to map ancient evolution. No myosin has been found in prokaryotes; thus, an innovatory shift in nucleotide-binding specificity (GTP to ATP) occurred to form the myosin–kinesin ancestor at the very origin of eukaryotes—in which actin and tubulin cytoskeletons had a central role[3]. Because both myosins and kinesins underwent marked diversification and domain rearrangements, the comparative study of these molecular motors offers great potential for disentangling early eukaryote evolution. Here we show that there are more than twice as many myosin types as previously described, all possessing unique domain structures and/or arrangements. This diversity can be divided into a limited number of subfamilies, of which three were present in the eukaryote cenancestor. Several features of myosin diversification strongly support a primary eukaryotic unikont/bikont bifurcation[3,11–13].

Immense diversity of myosin types

Our survey of the myosin gene family revealed 37 myosin types with different combinations of protein domains and scattered taxonomic distribution (Fig. 1a). The diversity of myosin paralogues encoded by each eukaryote varies considerably; for example, *Phytophthora ramorum* has 25 myosin genes encoding 13 different types, and humans have 12 (Fig. 1a), 6 of which are also present in *Dictyostelium*. The myosin types in *Phytophthora* and humans (Fig. 1a) represent independent peaks in evolutionary diversity of this gene family. In contrast, no myosin head domains could be identified from the flagellates *Giardia intestinalis*, *Trichomonas vaginalis* or the red alga *Cyanidioschyzon merolae*[14] with either BLASTp or PSI-BLAST[15]. The diversity of myosin types reported here indicates that many more might await discovery. Thirty myosin types were specific to narrow evolutionary lineages; for example, type 32 with multiple N-terminal WD40 domains was found only in Apicomplexa.

Multiple domain losses and gains

Myosin phylogeny reveals many instances of domain loss by deletion or divergence; for example, type 27 myosins (Fig. 1a) with no identified tail domains are clearly nested on the tree within clades comprising molecules with distinctive tail domains (Fig. 2), indicating a polyphyletic origin. Even paralogues not thus positioned have restricted taxonomic distribution, indicating that they might have arisen by recent tail loss. It is therefore possible that all ancestral myosins had long tails and that myosins with no identifiable tail domain (type 27) arose secondarily by multiple independent simplifications, making type 27 an artificial category.

An example of domain loss is type 27 of *Chlamydomonas*, which

[1]Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. [2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. †Present address: School of Biological and Chemical Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK.

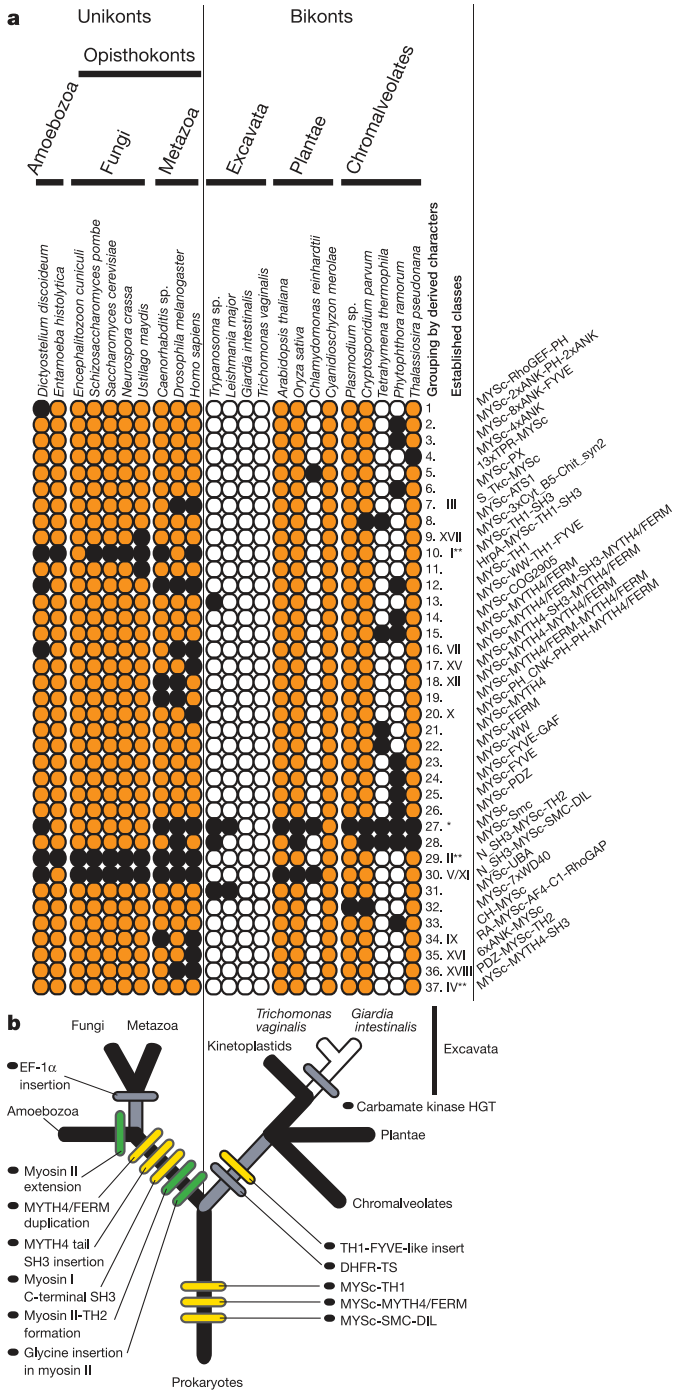


Figure 1 | Taxonomic distribution and evolutionary history of myosin paralogues. a, Comparative genomic survey of myosin paralogues in 23 key eukaryotic taxa. Taxa are shown on the X-axis and myosin domain order and classification on the Y-axis. Black dots indicate detection; open dots indicate no data available; orange dots indicate absence from a completed genome project. All myosins identified are listed with accession numbers in Supplementary Table 1. The type numbers should not be confused with previous class designations (for example ref. 4); equivalents are given in the right-hand column and in the text where appropriate. To conserve space, classes VI/VIII/XIV/XIII are indicated with an asterisk. Double asterisks, also detected in *Acanthamoeba castellanii*. **b**, Schematic tree showing myosin-derived synapomorphies (green bars) and three previously published shared characters (grey bars)[11,40,41]. Yellow bars indicate synapomorphy plus secondary loss. MYSc, myosin head domain.

strongly groups within type 30 (class XI) myosins and therefore is likely to have lost the C-terminal part of the molecule containing the dilute (DIL) domain found in all other members of the clade to which it belongs (Fig. 2). Although we detected no coiled-coil (Smc) domain in this molecule, it has a peptide tail that in its closest relatives carries an Smc domain, indicating that this region might have diverged beyond recognition. It is also clear that much variety in myosin domain organization arose by secondary domain losses, either by partial deletion or by divergence: current bioinformatic methods cannot distinguish absence from extreme divergence. The tree reveals several clear examples of novel domain gains, for example chitin synthetase to generate fungal class XVII proteins (type 9) and COG2905 yielding a *Phytophthora* myosin (type 14).

Three ancient myosin subfamilies

The 23 completed or near-complete genomes surveyed belong to five higher taxonomic units, namely opisthokonts, Excavata, Plantae, chromalveolates and Amoebozoa (Fig. 1), covering five of the six known eukaryotic supergroups[13,16,17], with only Rhizaria currently unsampled. Only 7 of the 37 myosin arrangements are found in more than one supergroup; most evolved after early eukaryote diversification. If we allow for fusions, partial deletions, duplications, and losses, we can use shared derived characters to rationalize myosin diversity into five broad ancestral myosin subfamilies. On the basis of taxonomic distribution, three of these seem to have been present in the eukaryote cenancestor (Fig. 1b). Of the other two, unikont-specific myosin II (type 29) is phylogenetically well defined, whereas a large weakly resolved group of chromalveolate myosins with a range of different C-terminal domains constitutes the second non-ancestral 'subfamily' (Fig. 2).

The broad taxonomic distribution of myosins with coiled-coil and dilute domains (MYSc-SMC-DIL, classes V and XI; here called MSD subfamily; type 30; Fig. 1a) in Plantae, opisthokonts and Amoebozoa, and their grouping in two robust clades (which we cannot exclude from being a single clade; Fig. 2) indicates that this arrangement might have arisen in the ancestral eukaryote and was lost by excavates and chromalveolates. The presence of an N-terminal SH3 domain varies between members of the MSD subfamily (Fig. 2). These domains are structurally similar to other SH3 domains but have many sequence differences. SH3 domains have conserved structures[18] but very variable sequences, which can make them difficult to identify; sequence alignments indicate that many MSD and class II myosins might possess an N-terminal SH3 domain not identified by conserved domain database (CDD) searches (Supplementary Fig. 1). The simplest interpretation of the scattered phylogenetic distribution of this domain within these myosin types is a combination of secondary losses and sequence divergence from an ancestral myosin that possessed an N-terminal SH3 domain. Alternative explanations involving independent additions, although possible given the numerous incidences of recombination involving SH3 domains, are less parsimonious. Thus, the ancestral eukaryote probably had a myosin with domain structure N_SH3-MYSc-SMC-DIL (Fig. 1b).

The second putatively ancestral subfamily comprises myosins of classes IV/VII/XII/XV (types 16–18 and 37) with an MYTH4/FERM domain. They are found in animals, Amoebozoa and chromalveolates, indicating that a myosin gene with one MYTH4/FERM domain might have been present in the eukaryotic cenancestor before undergoing multiple secondary losses or gene modifications. Alternatively, these domains might instead have become associated on separate occasions; the presence of a MYTH4/FERM domain at the N terminus of plant kinesins (for example GenBank accession number CAE03597) indicates that these domains might have recombined among distantly related molecular motors at least once.

Animals and Amoebozoa alone have MYTH4/FERM plus SH3 domain tails (Supplementary Fig. 2), representing a synapomorphy for unikont holophyly (being a monophyletic group with a single

evolutionary origin and including all descendents of its cenancestor). A second potential unikont synapomorphy is the duplication of the MYTH4/FERM region, indicating that the ancestral unikont domain structure might have been MYSc-MYTH4/FERM-SH3-MYTH4/FERM (type 16—myosin VII), as this is found in metazoa and *Dictyostelium* (although not in fungi, probably because of gene loss); simpler tail structures in this subfamily can readily be derived from this by differential partial deletions and/or domain insertions (for example, PH domains in vertebrate class X myosins; type 20). MYTH4/FERM myosins are dispersed among three clades (Fig. 2), but the tree base is too weakly resolved to disprove holophyly. A single origin of MYTH4/FERM tails in the cenancestor is the most

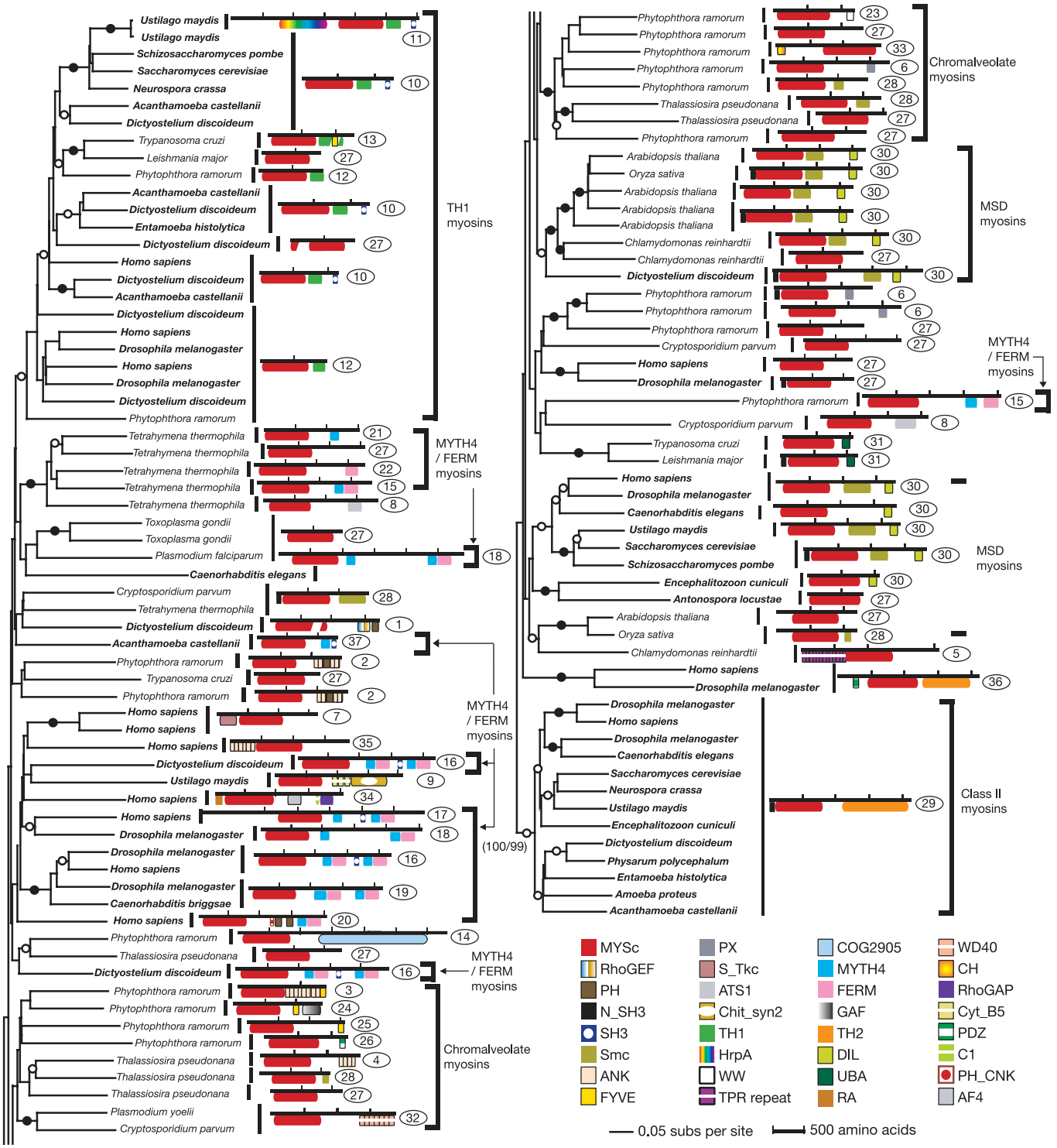


Figure 2 | Myosin head domain phylogeny (bayesian consensus: 118 myosins; 357 characters). Circled 1–37 designate domain combinations (Fig. 1a). Domains are labelled to scale (500 amino-acid residues indicated); myosin head domains (MYSc) shown in red, see key for colour-coding and Supplementary Table 1 for other abbreviations and naming. Square brackets label the five myosin subfamilies proposed here. Support values (bayesian posterior probability, 1,000 maximum-likelihood distance bootstraps or 100 Protpars bootstraps) are marked if all are more than 90% (filled circle) or more than 60% (open circle). Support values in brackets are from separate sequence-rich distance and parsimony analyses with long branches excluded. Unikonts are shown in bold. Bootstrap values over 50%, accession numbers and phylogenetic methods are given in Supplementary Fig. 4.

parsimonious interpretation, followed by differential duplication and domain losses and gains. Moreover, two properties of this subfamily—the MYTH4/FERM duplication and the SH3 insertion—support unikont holophyly (Fig. 1b).

The third probably ancestral subfamily is myosin I (types 10–12) with a membrane-binding TH1 domain tail. It is found in excavates, chromalveolates, opisthokonts and Amoebozoa but not in Plantae, implying that it was cenancestral but lost by Plantae. It is unlikely that Plantae diverged before myosin I formation, because the *dhfr-ts* fusion and their pattern of ciliary transformation support the inclusion of Plantae within bikonts[11–13]. The addition of an SH3 domain to the C terminus of this protein was detected only in unikonts, for which it may be synapomorphic (type 10; Fig. 1b and Supplementary Fig. 3). Phylogenetic analyses did not resolve this portion of the tree with significant support but did not clearly contradict this inference. In addition, bikont myosin type 12 genes contained a highly variable approximately 60 amino-acid-residue insertion within the TH1 domain. This is identified as a FYVE protein domain in trypanosomes but is unidentifiable in *Phytophthora*. Amino-acid alignments reveal that, although highly variable, this insertion contains several conserved positions, including four cysteine residues and VRV and KST motifs, indicating that it might be homologous (Supplementary Fig. 3) and a synapomorphy for bikonts or a subset of them (Fig. 1b).

Interestingly, both myosin I and MYTH4/FERM myosins have SH3 domains in their tail in unikonts but not in bikonts (Fig. 1a, b), whereas the third putatively cenancestral myosin subfamily (MSD) has an N-terminal-type SH3 domain. Because the three subfamilies must have arisen by two successive gene duplications of the first myosin gene in the cenancestor, the cenancestral myosin might have had an SH3-related domain; tandem duplication, differential deletions and divergence could have quickly generated each of the three primary myosin subfamilies. By this model the absence of SH3 domains from bikont class 1 (type 12) and MYTH4/FERM myosins would be secondary losses and thus synapomorphies for bikonts rather than for unikonts. However, the SH3 domain is in so many eukaryotic proteins that it must have been highly mobile (or have originated independently; that is, polyphyletically) during early eukaryote evolution. Consequently the insertion of SH3 domains into tails of an early unikont myosin or myosins is plausible.

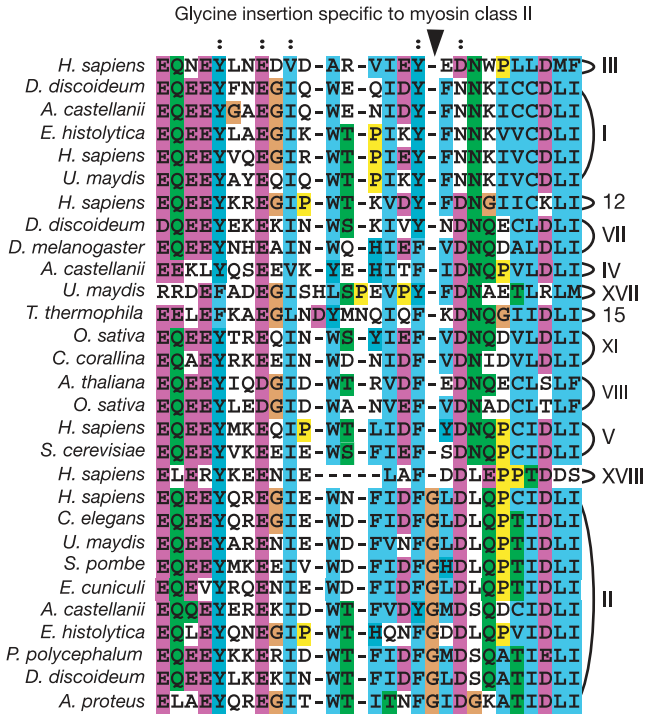


Figure 3 | Section from a myosin sequence alignment, including a representative selection of myosin types. The section illustrates a glycine insertion specific to myosin class II genes. Myosin classifications/types (Fig. 1a) are shown at the right.

Myosins and the unikont/bikont split

Class II myosins (Fig. 1) were proposed together with class I myosins to be the most ancient of all[4]; above we argued that MSD, MYTH4/FERM, and class I myosins, all occur in the widest diversity of eukaryotic supergroups and are therefore likely to be ancestral. The absence of myosin II from bikonts (Fig. 1), and the significant bootstrap support for its holophyly (Fig. 2), indicate that it might not have been a cenancestral myosin but instead a synapomorphy for unikonts only (Fig. 1a, b). A novel glycine residue inserted at position 507 (*Dictyostelium discoideum*) (Fig. 3) within all class II-derived myosins only (except myosin class XVIII—type 36) unambiguously supports the holophyly and derived nature of this paralogue. In some genes, indels can be ambiguous characters because of alignment uncertainty or evidence of multiple changes at the same site in different taxa[19]; such complications are absent in this case and the insertion is the derived state. This character, the strongly supported monophyly of myosin II (Fig. 2) and the unique myosin II coiled-coil tail (TH2) domain all make it highly improbable that the insertion occurred more than once. The less parsimonious possibility exists that myosin II was present in the first eukaryotes but was lost by the common ancestor of all sampled bikonts (by deletion or extreme divergence). Although some myosins show evidence of secondary loss and/or extreme divergence, there is no evidence of either for myosin II in the ten unikont species sampled. Characterizing myosins from Rhizaria would test our interpretation, which would be simply disproved if any have myosin II. Although the sequence phylogeny is unresolved for myosins with the MYTH4/FERM duplication and the acquisition of SH3 by class I and MYTH4/FERM myosins, all three are synapomorphies supporting unikont holophyly; thus, five independent synapomorphies give the same answer (Fig. 1b). Postulating four independent secondary losses of these paralogues is unparsimonious.

The apparent absence of myosin head domains in the two metamonad genomes (*Giardia* and *Trichomonas*) and the red alga *Cyanidioschyzon merolae* indicates that these organisms might lack myosins or that the myosins have evolved so radically that they are currently unidentifiable. Extreme sequence evolution in the common ancestor of all bikonts, and/or gene loss, could have masked the existence of bikont orthologues of the unikont myosin synapomorphies. Such hypothetical alternatives would be synapomorphies for bikonts. Independent later losses in all bikont lineages sampled would be even less parsimonious. Thus, either way, the myosin synapomorphy distribution data (Fig. 1) collectively support the partition of eukaryotes into unikonts and bikonts and are consistent with the holophyly of both groups[12]. *Giardia* and *Trichomonas* protein-encoding genes are notoriously fast-evolving compared with those of most other eukaryotes (except microsporidia[20]), which might explain why we found no myosins in their genomes. Significantly, we detected myosin class II (type 29) and class XI (type 30) in microsporidia, whose genes generally evolve even faster than those of metamonads; this ready detectability in the remarkably fast-evolving microsporidian genome makes secondary 'losses' in bikonts through rapid divergence unlikely, especially as myosin II is uniformly absent from plants and chromalveolates, which do not show unusually rapid divergence in their protein-encoding genes. It is therefore more likely that the five unikont myosin synapomorphies arose after a primary divergence of eukaryotes into unikonts and bikonts than that all five were ancestrally lost by bikonts. Together they provide the best available evidence for the holophyly of unikonts.

Our trees (if appropriately rooted: the root shown is arbitrary) weakly support the idea that class II and MSD myosins are related[4]. TH2 domains of class II myosins contain multiple sequence regions forming heptad repeats with similarity to Smc and other defined protein domains known to have a function in forming coil-coiled structures. We constructed an amino-acid alignment to investigate the potential homology between TH2 myosins and paralogues that possess Smc-type tails and/or a DIL-type tail domain (Supplementary Fig. 1). This shows that the tails are highly variable but have conserved character blocks, many present in the DIL-type tail (MSD myosins; class V and XI—type 30; Fig. 1a) or the TH2 domain (myosin II—type 29), but not both, establishing that myosin II TH2 domains are unique. Some regions do show weak homology between class II and MSD genes, consistent with a common ancestry and radical sequence divergence. The existence of TH2 domains unconnected to myosin head domains (for example *Giardia intestinalis* (GenBank accession number EAA38371) *Oryza sativa* (NP_921528) and several other eukaryotes) or connected to a kinesin domain in fungi (GenBank accession number T51930) indicates that it might have undergone illegitimate domain recombination at least once. We therefore cannot exclude the possibility that myosin II arose by gene fusion rather than from simple gross divergence of the tail from an MSD ancestor, but its support for unikont holophyly does not depend on its precise mode of origin.

Nature of the ancestral eukaryote

The presence of myosin II and its conserved amino-acid insertion in five diverse Amoebozoa (*Acanthamoeba castellanii*, *Physarum polycephalum* and *Dictyostelium discoideum*, *Amoeba proteus*—a member of Lobosea (naked aerobic amoebae with broad finger-like pseudopods but no cilia or ciliary root apparatus)—and *Entamoeba histolytica*, representing the amitochondrial Archamoebae formerly postulated to be early branching eukaryotes[21]) supports the arguments[12] that the unikont–bikont bifurcation is the oldest evolutionary diversification of known eukaryotes and shows that all these Amoebozoa are unikonts. The myosin II tree has 87%/62% bootstrap support for amoebozoan holophyly; previously Amoebozoa might have been paraphyletic, occupying a basal position to all eukaryotes, with some representatives (such as *Dictyostelium*) closer to opisthokonts, others closer to bikonts, and others diverging before the unikont/bikont bifurcation[22]. Monophyly of Amoebozoa has only weak to moderate bootstrap support in most 18S rRNA phylogenies[22–25]. Small subsets of amoebozoan taxa consistently form monophyletic clusters in sequence phylogenies of numerous nuclear or mitochondrial proteins[26–28], but taxon sampling is far too narrow to demonstrate holophyly or rooting of the Amoebozoa. Because Lobosea are entirely devoid of cilia, unlike Myxogastrea and Variosea[22], it might be argued (given the poor resolution of all sequence trees that include Lobosea) that the absence of cilia could be a primitive character and Lobosea might be the deepest branch in the eukaryotic tree, branching before cilia evolved[22]. The distribution of the myosin II synapomorphies makes this unlikely and strongly implies that ancestors of Lobosea lost cilia[3,13,22–24]. Amoebozoa are monophyletic (87%/62% bootstrap support) in the myosin II phylogeny, making paraphyly of Amoebozoa with respect to opisthokonts unlikely; the presence of a homologous approximately 130-residue extension to myosin II in all Amoebozoa analysed, but in no other eukaryotes (Supplementary Fig. 1), also supports amoebozoan holophyly. Although one key amoebozoan group (Discosea[22]) awaits sampling, there is no reason to suspect that it diverged before the fundamental unikont–bikont split or before the split between opisthokonts and the Amoebozoa sampled here. Monophyly of Mycetozoa (for example *Dictyostelium* and *Physarum*), previously unclear[22,23], was recovered with 85%/63% bootstrap support. Although the concept of a basal eukaryote bifurcation between unikonts and bikonts is relatively new[12], a recent comprehensive bayesian analysis of 18S rRNA shows a clear bipartition between unikonts and bikonts, and amoebozoan holophyly, both well supported[29]—unlike earlier distance trees[16].

Patterns of pseudopodial shape and movement seem very different between Amoebozoa and Rhizaria[13,30], which both include numerous amoeboid lineages. Amoebozoa have flat or lobose pseudopods[22], whereas Rhizaria tend to have thread-like filopodia or anastomosing reticulopodia—which is consistent with these two groups' being unrelated[13–31]. Their membership of unikonts and bikonts, respectively, indicates that the formation of pseudopodia in eukaryotic cells was probably an attribute of the last common eukaryotic ancestor. If myosin II, which functions in cytokinesis in unikonts, was genuinely absent from the ancestral eukaryote this function must originally have been performed by a different myosin, possibly the related MSD myosins. The numerous types of myosin in chromalveolates with novel domain organizations might be related to analogous functional replacements of the absent MSD myosins.

Discussion

The five new myosin synapomorphies for unikonts pinpoint the root of the eukaryotic tree with greater confidence and precision; they mean that the finding of a triple gene fusion, originally used to support unikont holophyly[12], in a red alga[14] (a bikont) does not invalidate the concept of unikont holophyly. Establishing the root position allows us to specify several key features of the last common ancestral eukaryote cell: an endosymbiont-derived mitochondrion, a cilium and centriole (most parsimoniously a single one with a cone of root microtubules[21,22]), and the cellular machinery to form pseudopodia. The amoebozoan flagellate *Phalansterium* with all these characters may be the best extant model for the ancestral eukaryotic phenotype[3,22]. As argued above, the cenancestral eukaryote probably had three different myosins with contrasting tail domains: myosin I, MYTH4/FERM myosins and MSD myosins. How the primary functions of myosin in cytokinesis, phagocytosis, pseudopodial and vesicle movement—all central to the life of the first eukaryote cells—were partitioned between these myosins cannot be clearly inferred from present data. The recent demonstration of an essential function for MYTH4/FERM myosin in *Dictyostelium* adhesion, important in both phagocytosis[32] and motility[33], indicates that this might have been its early function; its loss in both Plantae and Fungi, which independently evolved cell walls—thus losing both phagocytosis and amoeboid motility—is consistent with this. It is therefore tempting to indicate that MSD myosins might ancestrally have been responsible for cytokinesis (a role retained by their myosin II descendants) and pseudopodial activity. Functional studies of a more phylogenetically representative set of the myosins detailed here are needed to test this and to clarify major evolutionary shifts in myosin function. Physiological and genomic studies of myosin function and diversity in bikonts is especially needed (particularly Rhizaria, excavates, chromalveolates and lower plants). Given the marked differences in pseudopodial organization in Amoebozoa and Rhizaria, it would be particularly valuable to determine which myosin paralogues are present in Rhizaria; this might reveal novel lineage-specific paralogues, test our tentative conclusion that only three myosin subfamilies were ancestral for all eukaryotes, and yield further improvements in myosin classification.

METHODS

Comparative genome analyses. BLASTp searches obtained all recognizable myosin paralogues from 23 eukaryotic genome projects (up to April 2005; listed in Fig. 1a) by using GenBank eukaryote genome and non-redundant (nr) databases, dictybase, The Institute for Genomic Research (TIGR), Department of Energy Joint Genome Resource and the *Cyanidioschyzon merolae* genome project. Each myosin was then searched against the protein conserved domain database (CDD)[34] and the Pfam HMM[35] database to identify and classify protein domains. Protein domain identification is limited by the sensitivity of the search system and the diversity of protein domains in the database. Pfam and CDD were used in combination to increase both the sensitivity and the protein diversity. Every individual myosin type (defined here as a unique combination of protein

domains: 1–37 in Fig. 1a; N-terminal SH3-like and IQ domain characteristics were judged too variable for such typing) was then used for BLASTp searches against the GenBank nr database to identify further homologues from organisms additional to the 23 main taxa. Extra species not in Fig. 1 were surveyed to check for consistency or contradiction with the phylogenetic inferences based on the 23 comprehensively surveyed genomes (Supplementary Table 1). PSI BLAST[15] was used to seek highly divergent myosin head domains in genomes of *Giardia*, *Leishmania*, *Trichomonas* and *Cyanidioschyzon* and used myosin class I and II genes as starting seeds with *Dictyostelium* and *Phytophthora* genome sampling, to inform the PSI BLAST alignment process. PSI BLAST was run for 20 iterations, but gene discovery stopped several iterations before all searches finished.

Sequence alignment and phylogenetic analyses. Amino-acid sequences of the myosin head domains were aligned by ClustalX[36] and refined manually with Se-Al. Insertions and sequence characters not alignable with confidence were removed. Alignments sampling extensive diversity were initially analysed and then pared down by removing closely related sequences, while maintaining representative taxonomic and paralogue diversity. For the final phylogenetic trees two alignments were analysed: one sampled 357 conserved amino-acid positions only (to reduce long-branch problems) and 118 taxa, representing known myosin diversity. The second increased sampling (150 sequences, 371 characters) but with some long-branch myosin classes removed (for example myosin classes XVIII and XII). The resulting topologies are generally congruent apart from positions weakly supported in all analyses. Edited alignments were analysed by three methods: first, MrBayes 3 (ref. 37); second, maximum-likelihood distance bootstrap values (from 1,000 replicates) (refs 38, 39, and http://hades.biochem.dal.ca/Rogerlab/Software/software.html#puzzleboot); and third, 100 Protpars[39] bootstrap replicates; Supplementary Fig. 4 gives details.

Received 9 March; accepted 24 June 2005.

1. Sellers, J. R. *Myosins* (Oxford Univ. Press, Oxford, 1999).
2. Bahler, M. Are class III and class IX myosins motorized signalling molecules? *Biochim. Biophys. Acta* **1496,** 52–59 (2000).
3. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52,** 297–354 (2002).
4. Thompson, R. F. & Langford, G. M. Myosin superfamily evolutionary history. *Anat. Rec.* **268,** 276–289 (2002).
5. Furusawa, T., Ikawa, S., Yanai, N. & Obinata, M. Isolation of a novel PDZ-containing myosin from hematopoietic supportive bone marrow stromal cell lines. *Biochem. Biophys. Res. Commun.* **270,** 67–75 (2000).
6. Goodson, H. V. & Spudich, J. A. Molecular evolution of the myosin family: relationships derived from comparisons of amino acid sequences. *Proc. Natl Acad. Sci. USA* **90,** 659–663 (1993).
7. Hodge, T. & Cope, M. J. A myosin family tree. *J. Cell Sci.* **113,** 3353–3354 (2000).
8. Berg, J. S., Powell, B. C. & Cheney, R. E. A millennial myosin census. *Mol. Biol. Cell* **12,** 780–794 (2001).
9. Kull, F. J., Vale, R. D. & Fletterick, R. J. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *J. Muscle Res. Cell Motil.* **19,** 877–886 (1998).
10. Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317,** 41–72 (2002).
11. Stechmann, A. & Cavalier-Smith, T. Rooting the eukaryote tree by using a derived gene fusion. *Science* **297,** 89–91 (2002).
12. Stechmann, A. & Cavalier-Smith, T. The root of the eukaryote tree pinpointed. *Curr. Biol.* **13,** R665–R666 (2003).
13. Cavalier-Smith, T. Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* **39,** 338–348 (2003).
14. Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428,** 653–657 (2004).
15. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
16. Cavalier-Smith, T. Only six kingdoms of life. *Proc. R. Soc. Lond. B* **271,** 1251–1262 (2004).
17. Simpson, A. G. & Roger, A. J. The real 'kingdoms' of eukaryotes. *Curr. Biol.* **14,** R693–R696 (2004).
18. D'Aquino, J. A. & Ringe, D. Determinants of the SRC homology domain 3-like fold. *J. Bacteriol.* **185,** 4081–4086 (2003).
19. Bapteste, E. & Philippe, H. The potential value of indels as phylogenetic markers: position of trichomonads as a case study. *Mol. Biol. Evol.* **19,** 972–977 (2002).
20. Hirt, R. P. *et al.* Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96,** 580–585 (1999).
21. Cavalier-Smith, T. Archamoebae: the ancestral eukaryotes? *BioSystems* **25,** 25–38 (1991).
22. Cavalier-Smith, T., Chao, E. E. & Oates, B. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont *Phalansterium*. *Eur. J. Protistol.* **40,** 21–48 (2004).
23. Bolivar, I., Fahrni, J. F., Smirnov, A. & Pawlowski, J. SSU rRNA-based phylogenetic position of the genera *Amoeba* and *Chaos* (Lobosea, Gymnamoebia): the origin of gymnamoebae revisited. *Mol. Biol. Evol.* **18,** 2306–2314 (2001).
24. Kudryavtsev, A. A., Bernhardt, D., Schlegel, M., Chao, E. E. & Cavalier-Smith, T. 18S ribosomal RNA gene sequences of *Cochliopodium* (Himatismenida) and the phylogeny of Amoebozoa. *Protist* **156,** 215–224 (2005).
25. Milyutina, I. A., Aleshin, V. V., Mikrjukov, K. A., Kedrova, O. S. & Petrov, N. B. The unusually long small subunit ribosomal RNA gene found in amitochondriate amoeboflagellate *Pelomyxa palustris*: its rRNA predicted secondary structure and phylogenetic implication. *Gene* **272,** 131–139 (2001).
26. Dacks, J. B., Marinets, A., Doolittle, W. F., Cavalier-Smith, T. & Logsdon, J. M. Jr Analyses of RNA polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol. Biol. Evol.* **19,** 830–840 (2002).
27. Bapteste, E. *et al.* The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl Acad. Sci. USA* **99,** 1414–1419 (2002).
28. Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12,** 1773–1778 (2002).
29. Berney, C., Fahrni, J. & Pawlowski, J. How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol.* **2,** 13 (2004).
30. Bass, D. *et al.* Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist* **156,** 149–161 (2005).
31. Cavalier-Smith, T. & Chao, E. E. Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist* **154,** 341–358 (2003).
32. Titus, M. A. A class VII unconventional myosin is required for phagocytosis. *Curr. Biol.* **9,** 1297–1303 (1999).
33. Tuxworth, R. I. *et al.* A role for myosin VII in dynamic cell adhesion. *Curr. Biol.* **11,** 318–329 (2001).
34. Marchler-Bauer, A. *et al.* CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31,** 383–387 (2003).
35. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32,** D138–D141 (2004).
36. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882 (1997).
37. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
38. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502–504 (2002).
39. Felsenstein, J. *Phylip* (Department of Genetics, University of Washington, Seattle, 1995).
40. Minotto, L., Edwards, M. R. & Bagnara, A. S. *Trichomonas vaginalis*: Characterization, expression, and phylogenetic analysis of a carbamate kinase gene sequence. *Exp. Parasitol.* **95,** 54–62 (2000).
41. Baldauf, S. L. & Palmer, J. D. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl Acad. Sci. USA* **90,** 11558–11562 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Preliminary sequence data were obtained from The Institute for Genomic Research website (http://www.tigr.org) and the Department of Energy Joint Genome Institute (JGI) website (http://www.jgi.doe.gov). We thank TIGR and DOE JGI for making data publicly available, A. A. Davies for comments and assistance with data management, and D. Soanes for PSI BLAST assistance. T.A.R. was supported by a BBSRC studentship. T.C.-S. thanks NERC for research grants and NERC and the Canadian Institute for Advanced Research Evolutionary Biology Program for Fellowship support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.A.R. (thomr@nhm.ac.uk).

1. Taxonomic distribution and evolutionary history of myosin paralogues

a, Comparative genomic survey of myosin paralogues in 23 key eukaryotic taxa. Taxa are shown on the X-axis and myosin domain order and classification on the Y-axis. Black dots indicate detection; open dots indicate no data available; orange dots indicate absence from a completed genome project. All myosins identified are listed with accession numbers in Supplementary Table 1. The type numbers should not be confused with previous class designations (for example ref. 4); equivalents are given in the right-hand column and in the text where appropriate. To conserve space, classes VI/VIII/XIV/XIII are indicated with an asterisk. Double asterisks, also detected in *Acanthamoeba castellanii*. b, Schematic tree showing myosin-derived synapomorphies (green bars) and three previously published shared characters (grey bars)11,40,41. Yellow bars indicate synapomorphy plus secondary loss. MYSc, myosin head domain.

2. Myosin head domain phylogeny (bayesian consensus: 118 myosins; 357 characters)

Circled 1–37 designate domain combinations (Fig. 1a). Domains are labelled to scale (500 amino-acid residues indicated); myosin head domains (MYSc) shown in red, see key for colour-coding and Supplementary Table 1 for other abbreviations and naming. Square brackets label the five myosin subfamilies proposed here. Support values (bayesian posterior probability, 1,000 maximum-likelihood distance bootstraps or 100 Protpars bootstraps) are marked if all are more than 90% (filled circle) or more than 60% (open circle). Support values in brackets are from separate sequence-rich distance and parsimony analyses with long branches excluded. Unikonts are shown in bold. Bootstrap values over 50%, accession numbers and phylogenetic methods are given in Supplementary Fig. 4.

3. Section from a myosin sequence alignment, including a representative selection of myosin types

The section illustrates a glycine insertion specific to myosin class II genes. Myosin classifications/types (Fig. 1a) are shown at the right.

a

Unikonts

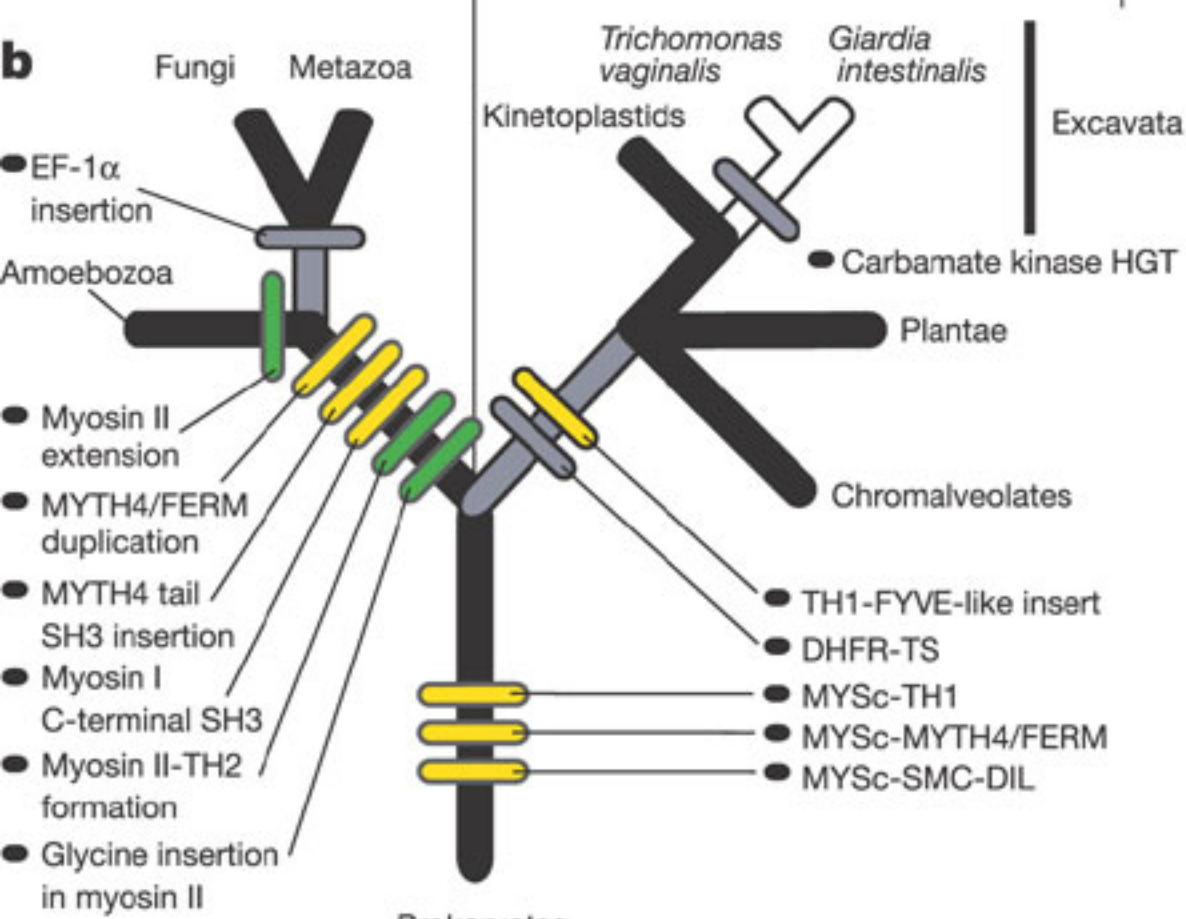
Opisthokonts

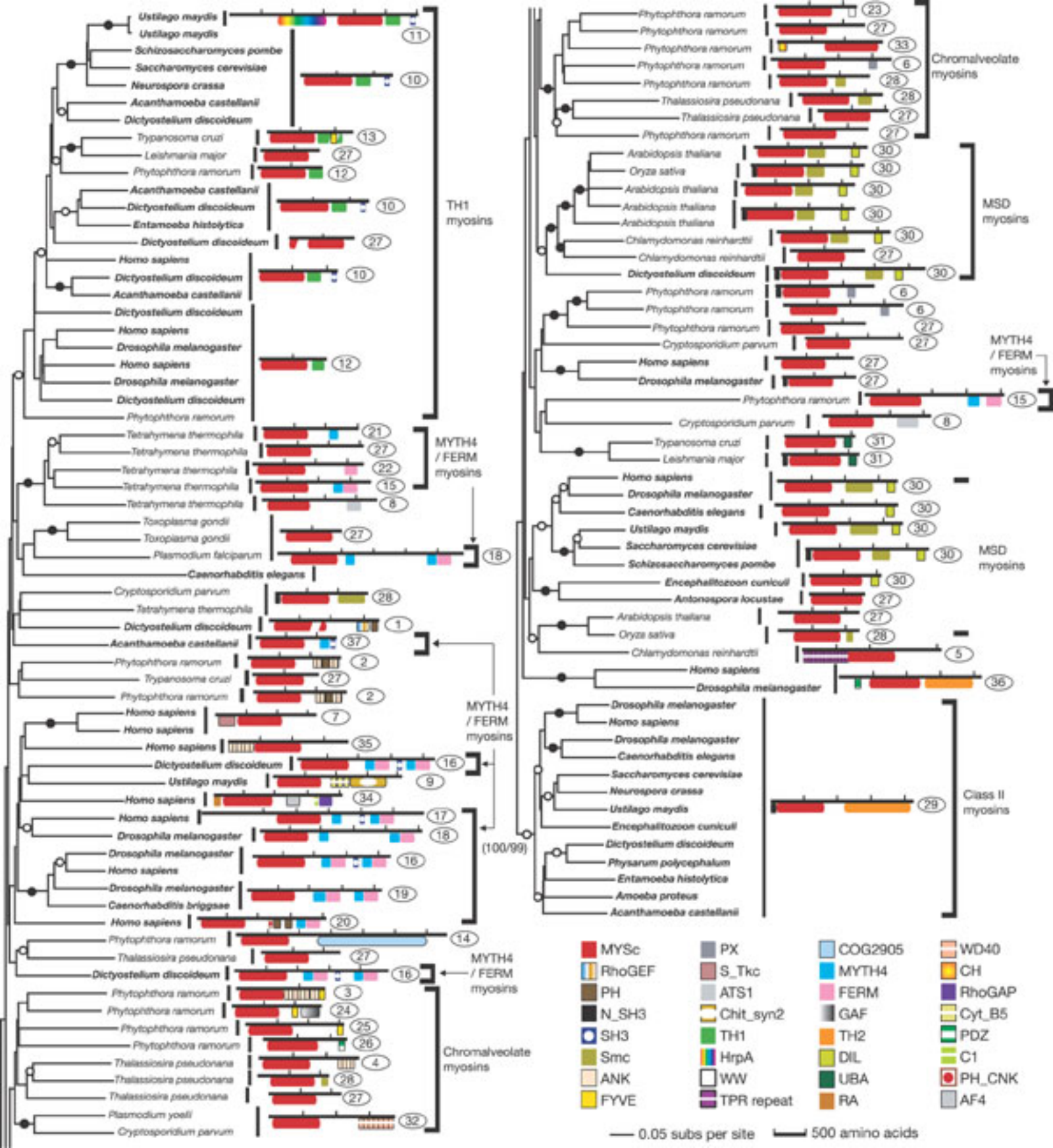
Amoebozoa | Fungi | Metazoa | Bikonts | Excavata | Plantae | Chromalveolates

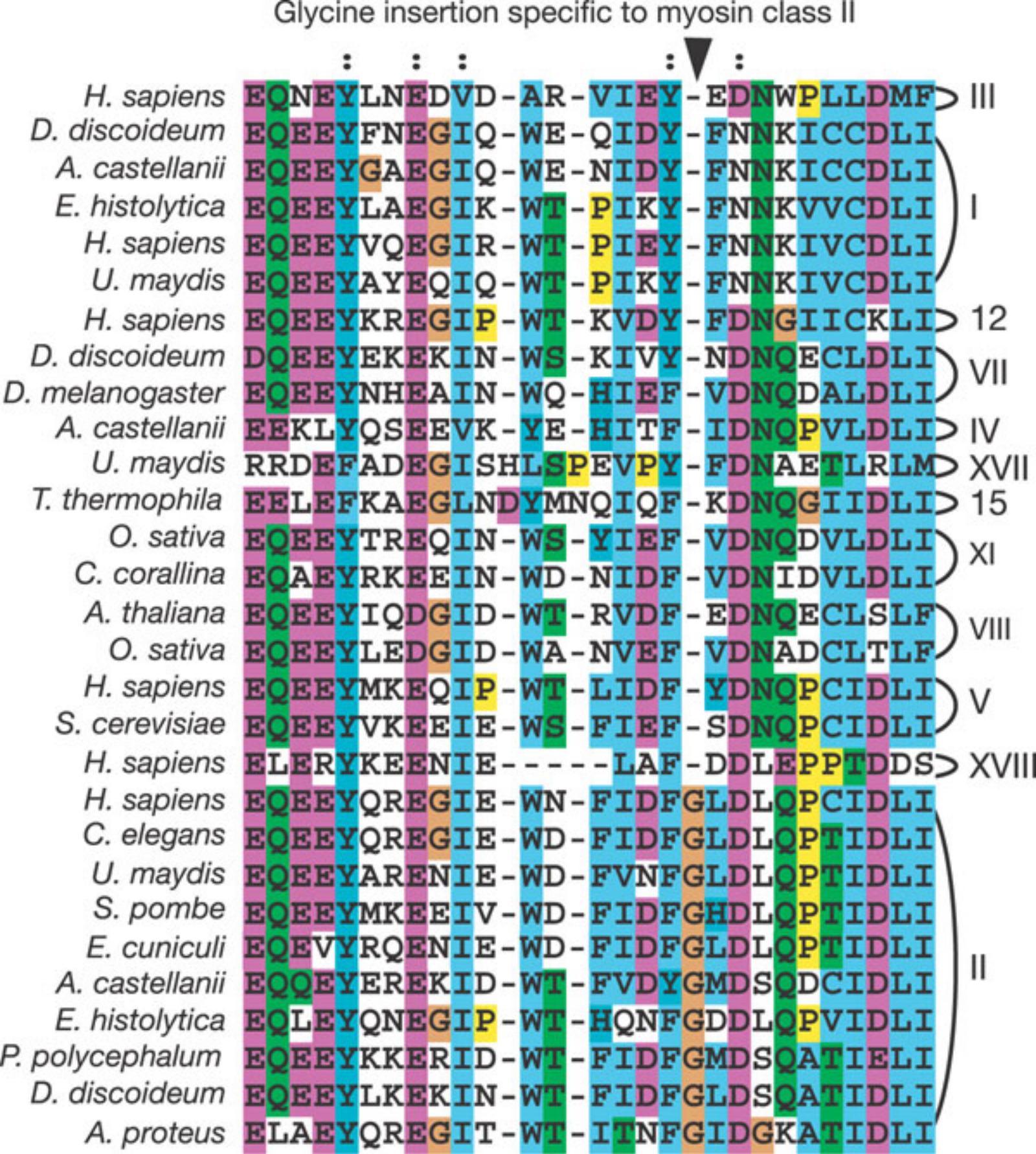
	Grouping by derived characters	Established classes
Dictyostelium discoideum	1	
Entamoeba histolytica	2.	
Encephalitozoon cuniculi	3.	
Schizosaccharomyces pombe	4.	
Saccharomyces cerevisiae	5.	
Neurospora crassa	6.	
Ustilago maydis	7.	III
Caenorhabditis sp.	8.	
Drosophila melanogaster	9.	XVII
Homo sapiens	10.	I**
Trypanosoma sp.	11.	
Leishmania major	12.	
Giardia intestinalis	13.	
Trichomonas vaginalis	14.	
Arabidopsis thaliana	15.	
Oryza sativa	16.	VII
Chlamydomonas reinhardtii	17.	XV
Cyanidioschyzon merolae	18.	XII
Plasmodium sp.	19.	
Cryptosporidium parvum	20.	X
Tetrahymena thermophila	21.	
Phytophthora ramorum	22.	
Thalassiosira pseudonana	23.	
	24.	
	25.	
	26.	
	27.	*
	28.	
	29.	II**
	30.	V/XI
	31.	
	32.	
	33.	
	34.	IX
	35.	XVI
	36.	XVIII
	37.	IV**

MYSc-RhoGEF-PH
MYSc-2xANK-PH-2xANK
MYSc-8xANK-FYVE
MYSc-4xANK
13xTPR-MYSc
MYSc-PX
S_Tkc-MYSc
MYSc-ATS1
MYSc-3xCyt_B5-Chit_syn2
MYSc-TH1-SH3
HrpA-MYSc-TH1-SH3
MYSc-TH1
MYSc-WW-TH1-FYVE
MYSc-COG2905
MYSc-MYTH4/FERM
MYSc-MYTH4/FERM-SH3-MYTH4/FERM
MYSc-MYTH4-SH3-MYTH4/FERM
MYSc-MYTH4-MYTH4/FERM
MYSc-MYTH4/FERM-MYTH4/FERM
MYSc-PH_CNK-PH-PH-MYTH4/FERM
MYSc-MYTH4
MYSc-FERM
MYSc-WW
MYSc-FYVE-GAF
MYSc-FYVE
MYSc-PDZ
MYSc
MYSc-Smc
N_SH3-MYSc-TH2
N_SH3-MYSc-SMC-DIL
MYSc-UBA
MYSc-7xWD40
CH-MYSc
RA-MYSc-AF4-C1-RhoGAP
6xANK-MYSc
PDZ-MYSc-TH2
MYSc-MYTH4-SH3

b




Ustilago maydis
Ustilago maydis
Schizosaccharomyces pombe
Saccharomyces cerevisiae
Neurospora crassa
Acanthamoeba castellanii
Dictyostelium discoideum
Trypanosoma cruzi
Leishmania major
Phytophthora ramorum
Acanthamoeba castellanii
Dictyostelium discoideum
Entamoeba histolytica
Dictyostelium discoideum
Homo sapiens
Dictyostelium discoideum
Acanthamoeba castellanii
Dictyostelium discoideum
Homo sapiens
Drosophila melanogaster
Homo sapiens
Drosophila melanogaster
Dictyostelium discoideum
Phytophthora ramorum
TH1 myosins
Tetrahymena thermophila
Tetrahymena thermophila
Tetrahymena thermophila
Tetrahymena thermophila
Tetrahymena thermophila
MYTH4 / FERM myosins
Toxoplasma gondii
Toxoplasma gondii
Plasmodium falciparum
Caenorhabditis elegans
Cryptosporidium parvum
Tetrahymena thermophila
Dictyostelium discoideum
Acanthamoeba castellanii
Phytophthora ramorum
Trypanosoma cruzi
Phytophthora ramorum
Homo sapiens
Homo sapiens
Homo sapiens
Dictyostelium discoideum
Ustilago maydis
Homo sapiens
Homo sapiens
Drosophila melanogaster
Drosophila melanogaster
Homo sapiens
Drosophila melanogaster
Caenorhabditis briggsae
Homo sapiens
MYTH4 / FERM myosins
(100/99)
Phytophthora ramorum
Thalassiosira pseudonana
Dictyostelium discoideum
MYTH4 / FERM myosins
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Thalassiosira pseudonana
Thalassiosira pseudonana
Plasmodium yoelii
Cryptosporidium parvum
Chromalveolate myosins
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Thalassiosira pseudonana
Thalassiosira pseudonana
Phytophthora ramorum
Chromalveolate myosins
Arabidopsis thaliana
Oryza sativa
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Dictyostelium discoideum
MSD myosins
Phytophthora ramorum
Phytophthora ramorum
Phytophthora ramorum
Cryptosporidium parvum
Homo sapiens
Drosophila melanogaster
MYTH4 / FERM myosins
Phytophthora ramorum
Cryptosporidium parvum
Trypanosoma cruzi
Leishmania major
Homo sapiens
Drosophila melanogaster
Caenorhabditis elegans
Ustilago maydis
Saccharomyces cerevisiae
Schizosaccharomyces pombe
Encephalitozoon cuniculi
Antonospora locustae
MSD myosins
Arabidopsis thaliana
Oryza sativa
Chlamydomonas reinhardtii
Homo sapiens
Drosophila melanogaster
Drosophila melanogaster
Homo sapiens
Drosophila melanogaster
Caenorhabditis elegans
Saccharomyces cerevisiae
Neurospora crassa
Ustilago maydis
Encephalitozoon cuniculi
Dictyostelium discoideum
Physarum polycephalum
Entamoeba histolytica
Amoeba proteus
Acanthamoeba castellanii
Class II myosins
MYSc
RhoGEF
PH
N_SH3
SH3
Smc
ANK
FYVE
PX
S_Tkc
ATS1
Chit_syn2
TH1
HrpA
WW
TPR repeat
COG2905
MYTH4
FERM
GAF
TH2
DIL
UBA
RA
WD40
CH
RhoGAP
Cyt_B5
PDZ
C1
PH_CNK
AF4
0.05 subs per site
500 amino acids


Glycine insertion specific to myosin class II
H. sapiens EQNEYLNEDVD-AR-VIEY-EDNWPLLDMF III
D. discoideum EQEEYFNEGIQ-WE-QIDY-FNNKICCDLI
A. castellanii EQEEYGAEGIQ-WE-NIDY-FNNKICCDLI
E. histolytica EQEEYLAEGIK-WT-PIKY-FNNKVVCDLI I
H. sapiens EQEEYVQEGIR-WT-PIEY-FNNKIVCDLI
U. maydis EQEEYAYEQIQ-WT-PIKY-FNNKIVCDLI
H. sapiens EQEEYKREGIP-WT-KVDY-FDNGIICKLI 12
D. discoideum DQEEYEKEKIN-WS-KIVY-NDNQECLDLI VII
D. melanogaster EQEEYNHEAIN-WQ-HIEF-VDNQDALDLI
A. castellanii EEKLYQSEEVK-YE-HITF-IDNQPVLDLI IV
U. maydis RRDEFADEGISHLSPEVPY-FDNAETLRLM XVII
T. thermophila EELEFKAEGLNDYMNQIQF-KDNQGIIDLI 15
O. sativa EQEEYTREQIN-WS-YIEF-VDNQDVLDLI XI
C. corallina EQAEYRKEEIN-WD-NIDF-VDNIDVLDLI
A. thaliana EQEEYIQDGID-WT-RVDF-EDNQECLSLF VIII
O. sativa EQEEYLEDGID-WA-NVEF-VDNADCLTLF
H. sapiens EQEEYMKEQIP-WT-LIDF-YDNQPCIDLI V
S. cerevisiae EQEEYVKEEIE-WS-FIEF-SDNQPCIDLI
H. sapiens ELERYKEENIE-----LAF-DDLEPPTDDS XVIII
H. sapiens EQEEYQREGIE-WN-FIDFGLDLQPCIDLI
C. elegans EQEEYQREGIE-WD-FIDFGLDLQPTIDLI
U. maydis EQEEYARENIE-WD-FVNFGLDLQPTIDLI
S. pombe EQEEYMKEEIV-WD-FIDFGHDLQPTIDLI
E. cuniculi EQEVYRQENIE-WD-FIDFGLDLQPTIDLI
A. castellanii EQQEYEREKID-WT-FVDYGMDSQDCIDLI II
E. histolytica EQLEYQNEGIP-WT-HQNFGDDLQPVIDLI
P. polycephalum EQEEYKKERID-WT-FIDFGMDSQATIELI
D. discoideum EQEEYLKEKIN-WT-FIDFGLDSQATIDLI
A. proteus ELAEYQREGIT-WT-ITNFGIDGKATIDLI

Phylogeny of myosin head domains; this is the same tree as in Fig. 2 of the main body of the paper, except that accession numbers replace species names and exact support values are shown. This Bayesian (consensus) tree was calculated from an alignment of 118 myosin head domains and a sampling of 357 conserved amino acid character positions. The phylogeny was analysed in three ways: 1) MRBAYES 3[1] with amino acid substitution model set to 'mixed', allowing the mcmc to search over all substitution models (reducing the assumptions prior to the analysis). Rate variation across sites was modeled using a gamma distribution with four rate categories and a proportion of invariant sites. The MCMC search was run with four chains for 1,000,000 generations, sampling every 100 generations. The first 2000 trees (200,000 generations) were discarded as "burnin" ensuring that all parameters had reached a plateau in the MCMC searches; 2) Maximum Likelihood distance bootstrap values (from 1000 replicates) were obtained using TREEPUZZLE 5.1[2] for parameter estimation (substitution model, 8 multivariant + invariant sites and in coordination with PUZZLEBOOT[3] to obtain distance matrices. Programs from the PHYLIP package[4] were used to create bootstrap datasets (Seqboot—1000 replicates), calculate distance trees (neighbor) and assembling a bootstrap consensus tree (consense). 3) 100 bootstrap replicates were also analysed using parsimony methods using the program PROTPARS with 3x global rearrangements. A consensus tree was calculated using consense. Topology support values are labelled on the Bayesian (consensus) topology tree in the order % Bayesian posterior probability/% bootstrap support from 1000 ML distance replicates/% bootstrap support from 100-protein parsimony replicates. Posterior probability is shown as a % value to reduce and standardise the characters used. Only when one bootstrap support value is in excess of 49% are the support values labelled.

1. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models.

Bioinformatics 19, 1572-1574 (2003).

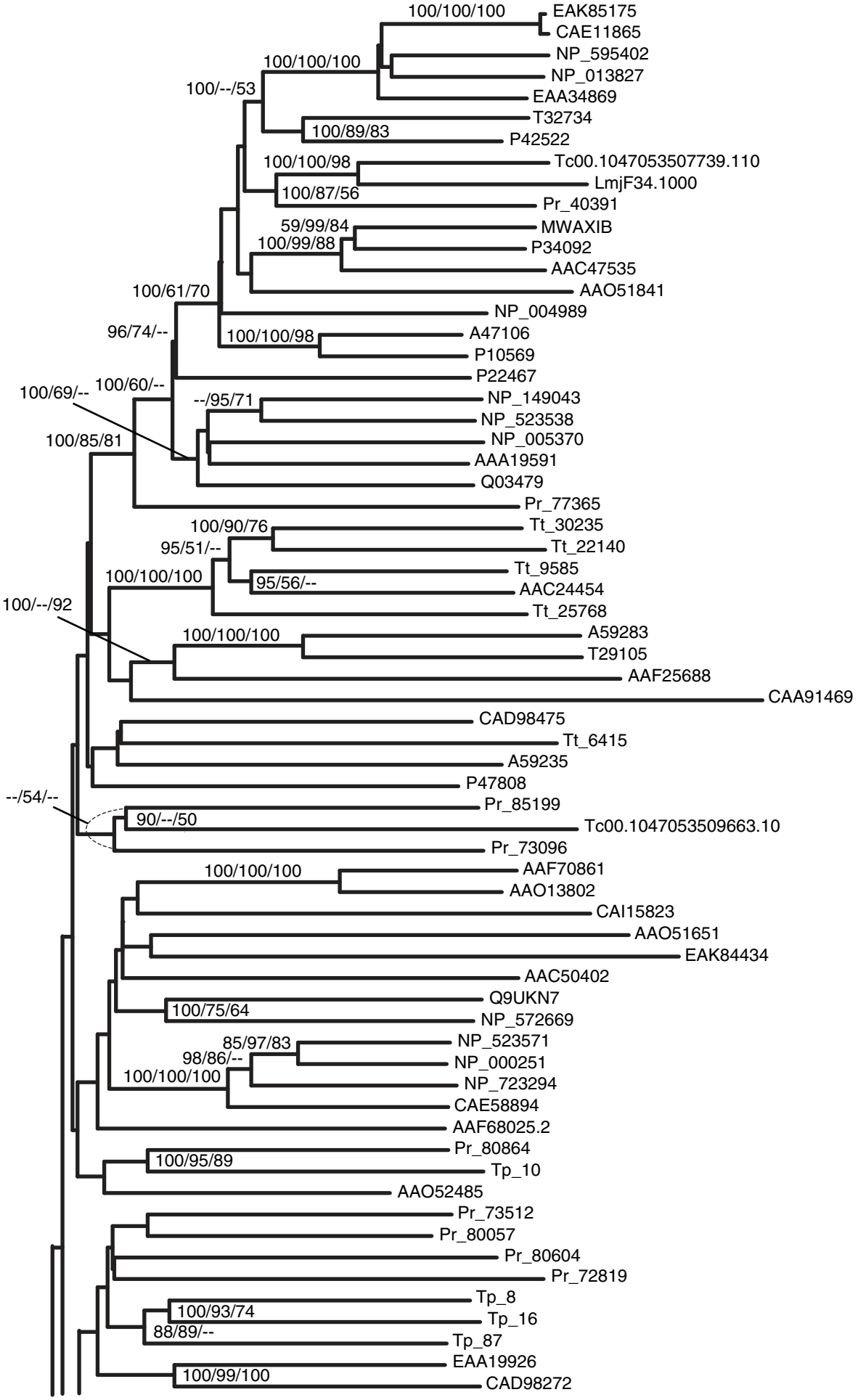
2. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood

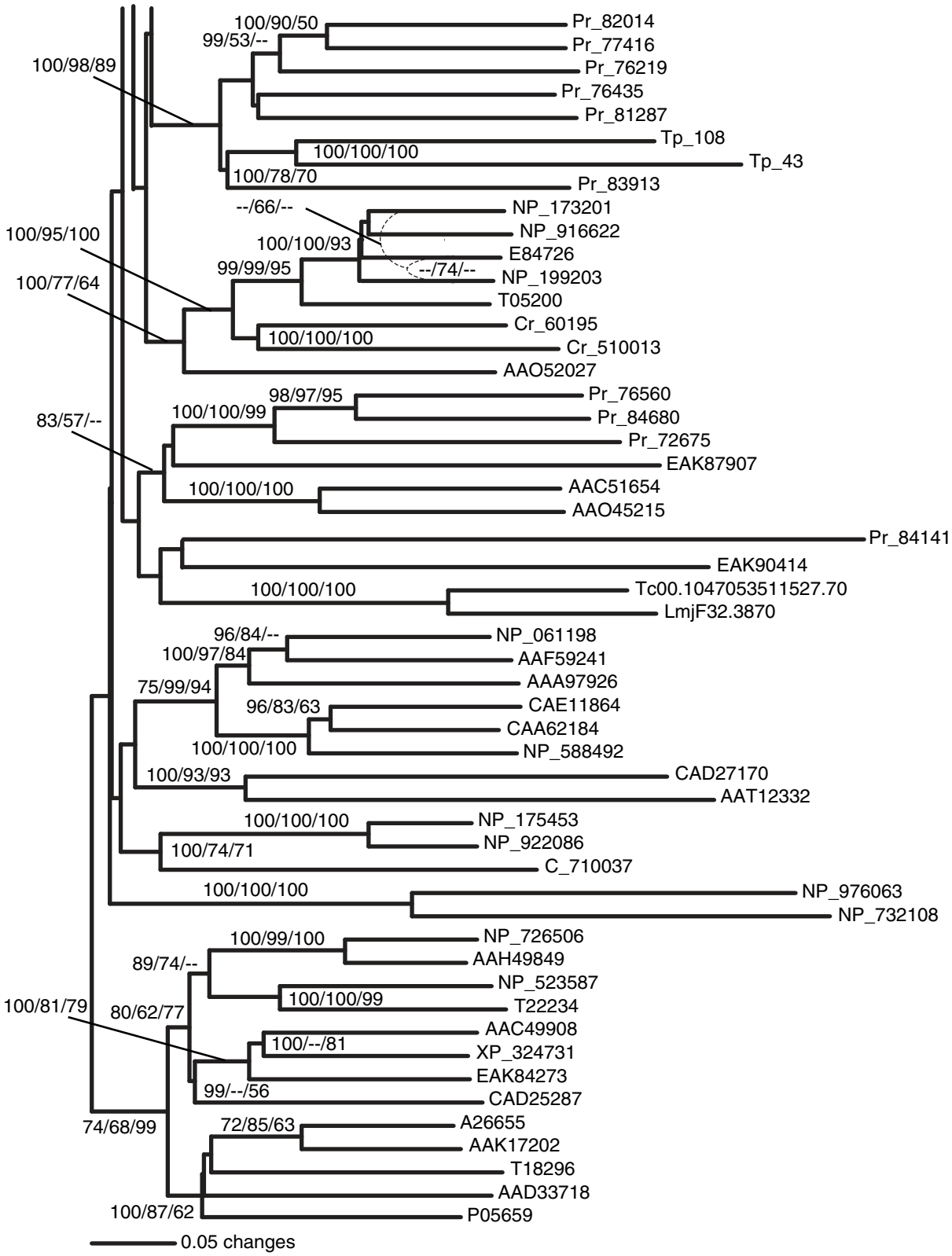
phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502-504 (2002).

3. Holder, M. & Roger, A. J. PUZZLEBOOT version 1.03.

http//hades.biochem.dal.ca/Rogerlab/Software/software.html.

4. Felsenstein, J. PHYLIP. (Department of Genetics, University of Washington, Seattle, 1995).


100/100/100
EAK85175
CAE11865
NP_595402
100/100/100
NP_013827
100/--/53
EAA34869
T32734
100/89/83
P42522
100/100/98
Tc00.1047053507739.110
LmjF34.1000
100/87/56
Pr_40391
59/99/84
MWAXIB
100/99/88
P34092
AAC47535
100/61/70
AAO51841
NP_004989
96/74/--
100/100/98
A47106
P10569
P22467
100/69/--
100/60/--
--/95/71
NP_149043
NP_523538
100/85/81
NP_005370
AAA19591
Q03479
Pr_77365
100/90/76
Tt_30235
95/51/--
Tt_22140
Tt_9585
100/100/100
95/56/--
AAC24454
100/--/92
Tt_25768
100/100/100
A59283
T29105
AAF25688
CAA91469
CAD98475
Tt_6415
A59235
P47808
--/54/--
Pr_85199
90/--/50
Tc00.1047053509663.10
Pr_73096
100/100/100
AAF70861
AAO13802
CAI15823
AAO51651
EAK84434
AAC50402
Q9UKN7
100/75/64
NP_572669
85/97/83
NP_523571
98/86/--
NP_000251
100/100/100
NP_723294
CAE58894
AAF68025.2
Pr_80864
100/95/89
Tp_10
AAO52485
Pr_73512
Pr_80057
Pr_80604
Pr_72819
Tp_8
100/93/74
Tp_16
88/89/--
Tp_87
EAA19926
100/99/100
CAD98272


100/90/50
99/53/--
100/98/89
Pr_82014
Pr_77416
Pr_76219
Pr_76435
Pr_81287
Tp_108
100/100/100
Tp_43
100/78/70
Pr_83913
--/66/--
NP_173201
NP_916622
100/95/100
100/100/93
E84726
--/74/--
99/99/95
NP_199203
100/77/64
T05200
Cr_60195
100/100/100
Cr_510013
AAO52027
98/97/95
Pr_76560
100/100/99
Pr_84680
83/57/--
Pr_72675
EAK87907
100/100/100
AAC51654
AAO45215
Pr_84141
EAK90414
100/100/100
Tc00.1047053511527.70
LmjF32.3870
96/84/--
NP_061198
100/97/84
AAF59241
AAA97926
75/99/94
96/83/63
CAE11864
CAA62184
100/100/100
NP_588492
100/93/93
CAD27170
AAT12332
100/100/100
NP_175453
NP_922086
100/74/71
C_710037
100/100/100
NP_976063
NP_732108
100/99/100
NP_726506
AAH49849
89/74/--
NP_523587
100/100/99
T22234
100/81/79
80/62/77
AAC49908
100/--/81
XP_324731
EAK84273
99/--/56
CAD25287
74/68/99
72/85/63
A26655
AAK17202
T18296
AAD33718
100/87/62
P05659
0.05 changes