Article

# New insights into myosin evolution and classification

FOTH, Bernardo Javier, GOEDECKE, Marc Christian, SOLDATI-FAVRE, Dominique

**Abstract**

Myosins are eukaryotic actin-dependent molecular motors important for a broad range of functions like muscle contraction, vision, hearing, cell motility, and host cell invasion of apicomplexan parasites. Myosin heavy chains consist of distinct head, neck, and tail domains and have previously been categorized into 18 different classes based on phylogenetic analysis of their conserved heads. Here we describe a comprehensive phylogenetic examination of many previously unclassified myosins, with particular emphasis on sequences from apicomplexan and other chromalveolate protists including the model organism Toxoplasma, the malaria parasite Plasmodium, and the ciliate Tetrahymena. Using different phylogenetic inference methods and taking protein domain architectures, specific amino acid polymorphisms, and organismal distribution into account, we demonstrate a hitherto unrecognized common origin for ciliate and apicomplexan class XIV myosins. Our data also suggest common origins for some apicomplexan myosins and class VI, for classes II and XVIII, for classes XII and XV, and for some microsporidian myosins and class V, thereby [...]

Reference

FOTH, Bernardo Javier, GOEDECKE, Marc Christian, SOLDATI-FAVRE, Dominique. New insights into myosin evolution and classification. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, vol. 103, no. 10, p. 3681-3686

DOI : 10.1073/pnas.0506307103
PMID : 16505385

# New insights into myosin evolution and classification

**Bernardo J. Foth*, Marc C. Goedecke, and Dominique Soldati**

Department of Microbiology and Molecular Medicine, Centre Médical Universitaire, University of Geneva, 1 Rue Michel-Servet, 1211 Geneva, Switzerland

Edited by Edward D. Korn, National Institutes of Health, Bethesda, MD, and approved December 21, 2005 (received for review July 24, 2005)

**Myosins are eukaryotic actin-dependent molecular motors important for a broad range of functions like muscle contraction, vision, hearing, cell motility, and host cell invasion of apicomplexan parasites. Myosin heavy chains consist of distinct head, neck, and tail domains and have previously been categorized into 18 different classes based on phylogenetic analysis of their conserved heads. Here we describe a comprehensive phylogenetic examination of many previously unclassified myosins, with particular emphasis on sequences from apicomplexan and other chromalveolate protists including the model organism *Toxoplasma*, the malaria parasite *Plasmodium*, and the ciliate *Tetrahymena*. Using different phylogenetic inference methods and taking protein domain architectures, specific amino acid polymorphisms, and organismal distribution into account, we demonstrate a hitherto unrecognized common origin for ciliate and apicomplexan class XIV myosins. Our data also suggest common origins for some apicomplexan myosins and class VI, for classes II and XVIII, for classes XII and XV, and for some microsporidian myosins and class V, thereby reconciling evolutionary history and myosin structure in several cases and corroborating the common coevolution of myosin head, neck, and tail domains. Six novel myosin classes are established to accommodate sequences from chordate metazoans (class XIX), insects (class XX), kinetoplastids (class XXI), and apicomplexans and diatom algae (classes XXII, XXIII, and XXIV). These myosin (sub)classes include sequences with protein domains (FYVE, WW, UBA, ATS1-like, and WD40) previously unknown to be associated with myosin motors. Regarding the apicomplexan ''myosome,'' we significantly update class XIV classification, propose a systematic naming convention, and discuss possible functions in these parasites.**

Apicomplexa | Chromalveolata

Myosins are molecular motors that diversified very early during eukaryotic evolution (1, 2). Most eukaryotes rely on myosins, and only a few taxonomic groups, e.g., red algae and diplomonad protists (3), appear to live without them. Mammals boast up to 40 different myosin genes, and myosin mutations are linked to serious pathologies like myopathies, blindness, and hearing loss. Myosin heavy chains move along tracks of filamentous actin and are involved in various cellular functions such as organellar transport, mitosis, and cell locomotion. They commonly comprise three domains: a conserved head responsible for actin binding, ATPase activity, and generation of movement; a short neck that usually interacts with myosin light chains; and a variable tail that commonly binds the motor "cargo" and determines the functional specificity of the motor.

Myosin heavy chains have been categorized into 18 classes, mostly based on comparisons and phylogenetic analysis of the conserved motor domain. One simple and effective rule for the delineation of myosin classes has been to consider the first branches emanating from the center of an unrooted myosin phylogeny that receive >90% bootstrap (BS) support as separate classes (4, 5). Most frequently, phylogenetic analysis of myosins has used distance matrix-based methods as implemented, e.g., in PAUP (6) or CLUSTAL (7), i.e., without the benefits of more sophisticated algorithms that take amino acid substitution models into account [e.g., programs of the PHYLIP package (8)].

The phylum Apicomplexa comprises unicellular eukaryotes that live as obligate intracellular parasites and includes important pathogens such as *Plasmodium* and *Toxoplasma gondii*, the etiological agents of malaria and toxoplasmosis. Others result in huge economic losses through infections of livestock (e.g., *Eimeria* and *Babesia*). Together with ciliates and dinoflagellate algae, apicomplexans constitute the Alveolata (9), whereas alveolates and the Chromista probably belong to a common ancient evolutionary lineage, the chromalveolates (10–12). The Chromista represent a heterogeneous assemblage of algal groups and oomycetes. See the recent work by Adl *et al.* (13) for an overview of eukaryote classification.

Here, we present a phylogenetic analysis of myosins incorporating many previously unclassified sequences, with particular emphasis on myosins from parasitic and other protists. Phylogenies and a conserved amino acid polymorphism show that class XIV encompasses myosins of both apicomplexans and ciliates, with some sequences of both groups sharing the same protein domain architecture. Our data also suggest so-far-unrecognized common origins for other groups of myosins, and we establish six new myosin classes to accommodate previously unclassified myosins from chordate metazoans (class XIX), insects (XX), trypanosomatid protozoa (XXI), and apicomplexan parasites (classes XXII, XXIII, and XXIV). The proposed myosin (sub)classes include sequences with myosin tail domain architectures whose possible functional relevance for apicomplexan parasites is discussed.

## Results

**Statistical Support in Phylogenetic Analyses.** We used four different (combinations of) phylogenetic inference programs and use the term "statistical support" to refer to BS values from distance matrix-based phylogenies generated by CLUSTAL (7) or PROTDIST (together with NEIGHBOR or FITCH) (8) and to posterior probabilities (PPs) from Bayesian analysis with MRBAYES (14). Because we found that sequence input order before generating multiple sequence alignments (with CLUSTAL) had a significant effect on the statistical support for distinct clusters within the resulting trees, regardless of the phylogenetic method used, we carried out all major phylogenetic analyses on series of 12 independent alignments generated after randomization of sequence input order. We also noted a specific negative bias on statistical support for certain clusters in trees by a few usually divergent (long branches in trees) and/or unclassified myosin sequences (highlighted in Fig. 1) that we refer to as "rogue" sequences (15). Unless stated otherwise, BS values (in %) mentioned throughout the text and indicated in Fig. 1 refer to PROTDIST/NEIGHBOR BS analysis (300 replicates) based on a representative alignment ("alignment 7") of a big dataset (267 sequences) from which the rogue sequences had been omitted,

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ATS, α-tubulin suppressor; BS, bootstrap; MyTH4, myosin tail homology 4; PP, posterior probability.

Data deposition: The data reported in this paper have been deposited in the GenBank database [accession nos. DQ131541 (TgMyoF) and DQ131540 (TgMyoG)].

See Commentary on page 3498.

*To whom correspondence should be addressed. E-mail: bernardo.foth@medecine.unige.ch.

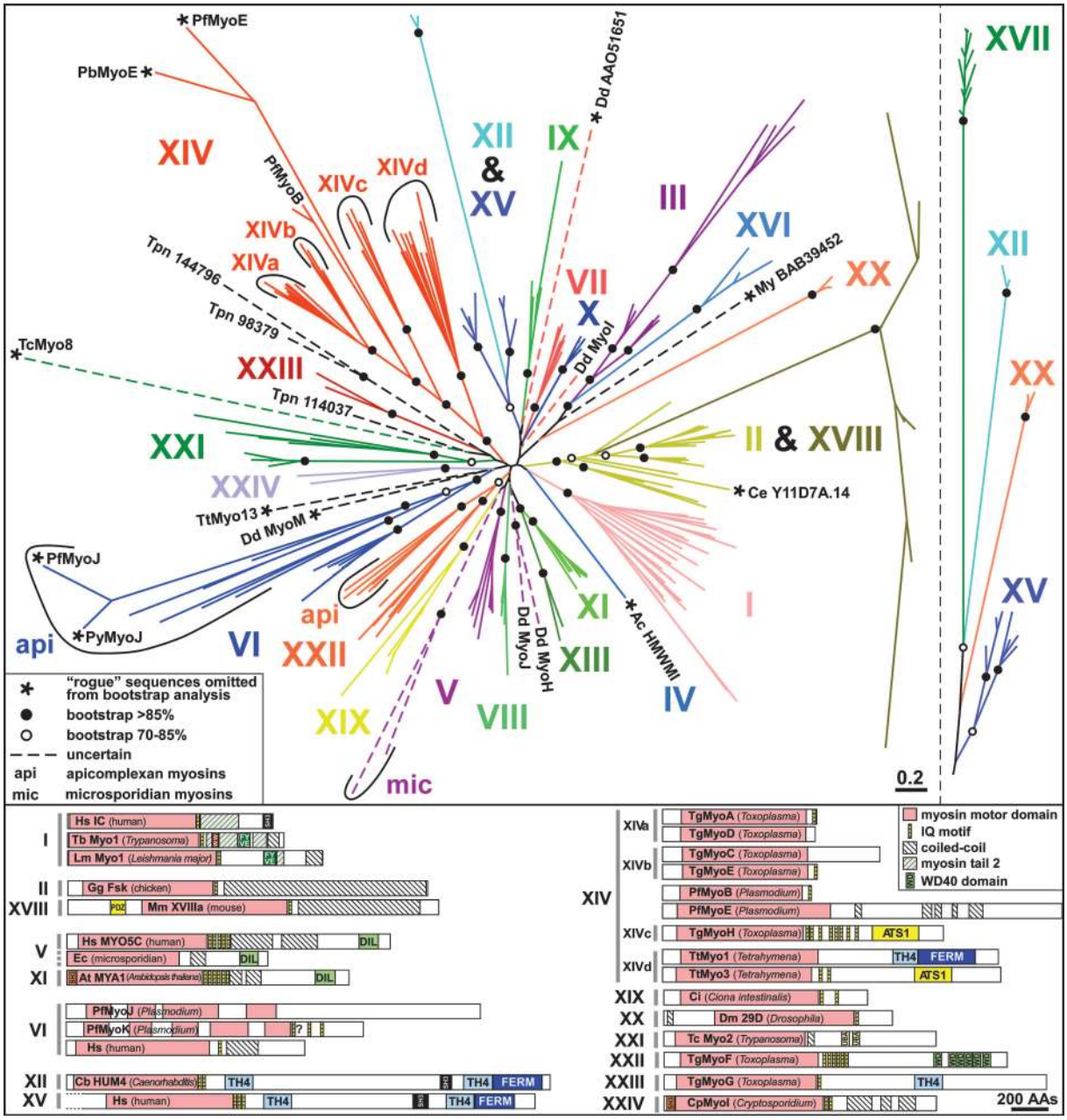


**Fig. 1.** Representative distance matrix-based phylogeny (PROTDIST/NEIGHBOR) of myosin head domains. This tree is based on alignment 7 from a series of 12 alignments and corresponds to the most representative tree, as judged by the BS for select myosin classes (see Table 2 for details). The highly divergent class XVII myosins have been omitted from the main phylogeny, whereas the *Inset* (*Upper Right*) shows the respective part of a tree in which these myosins have been included (note the long branch connecting class XVII to the rest of the tree). BS support is based on analyses excluding rogue sequences (indicated by asterisk) and is indicated by circles only for select nodes in the tree. Fig. 3, which is published as supporting information on the PNAS web site, shows the same tree including all sequence names and BS values. (*Lower*) Domain structures for select myosins. "Myosin tail 2" refers to Pfam entry 06017 (TH1). Ac, *Acanthamoeba castellanii*; Cp, *C. parvum*; Dd, *D. discoideum*; Ec, *E. cuniculi*; My, *Mizuhopecten yessoensis*; Pb, *Plasmodium berghei*; Pf, *P. falciparum*; Py, *P. yoelii*; Tc, *T. cruzi*; Tpn, *T. pseudonana*; and Tt, *T. thermophila*.

whereas PPs refer to Bayesian analysis of 12 alignments based on a small dataset (74 sequences). For further details, see *Supporting Text* and Table 2, which are published as supporting information on the PNAS web site.

**Classes II and XVIII Share Common Origins.** One remarkable difference between simple neighbor-joining trees generated by CLUSTAL or PAUP (6) and trees that have been calculated by taking an amino acid substitution model (e.g., Jones–Taylor–Thornton) into account is that the latter consistently position class XVIII myosins within class II with excellent statistical support (Fig. 1 and Table 2; 95% BS, median PP 1.0). The inference that class XVIII may actually represent divergent class II myosins that have acquired an N-terminal PDZ domain is further supported by their similar neck and tail domain architectures (a single IQ motif followed by an extensive coiled-coil forming domain; Fig. 1) and is consistent with the organismal distribution of these myosins, which are both (also) found in metazoans. An overview of the known taxonomic distribution of all myosin classes is given in Table 3, which is published as

supporting information on the PNAS web site. For more information on the function of myosins, please refer to the current literature or to an appropriate review (5, 16).

**Class V and Microsporidian Myosins.** Microsporidia are unicellular parasites with a wide host range and, despite obvious differences in lifestyle and morphology, they are most closely related to fungi (17). Although myosins of other fungi belong to classes I, II, and V, the two distantly related taxa *Encephalitozoon cuniculi* and *Antonospora locustae* reveal only two types of microsporidian myosins: one class II and one previously unclassified myosin. In our Bayesian analysis, the unclassified sequences fall into myosin class V with good statistical support (median PP 0.84; Table 2). Although we have observed this monophyletic association only in Bayesian analysis [and maximum likelihood trees generated by PROML (8); data not shown], we note that these microsporidian myosins are often positioned in close proximity to class V even in distance matrix-based analyses (Fig. 1), albeit without statistical support (Table 2). The assumption that these microsporidian sequences represent divergent class V myosins is parsimonious, because class V myosins are also found in other fungi, and is supported by their similar tail domain architectures [coiled-coil and C-terminal dilute (DIL) domains; Fig. 1] and by phylogenetic analysis of DIL domain sequences that weakly support a common origin of these domains from microsporidian and other fungal class V myosins (data not shown).

**Classes XII and XV Share Common Origins.** Similar to the case of classes II and XVIII, PROTDIST-based phylogenies also suggest a common origin for classes XII and XV (Fig. 1). Although the monophyly of classes XII and XV does not receive strong statistical support in PROTDIST-based trees (70% BS) and is not or barely recovered in CLUSTAL and MRBAYES phylogenies, respectively (Table 2), it is strongly corroborated by independent evidence: the tail domain architectures of class XII and XV myosins are practically identical [two myosin tail homology 4 (MyTH4), one FERM, and one SH3 domain; Fig. 1], and MyTH4 domain phylogenies strongly suggest a common origin of these class XII/XV tail domains (BS $>$ 97%, PP $>$ 0.98; Fig. 4, which is published as supporting information on the PNAS web site) (18). Also, that class XV myosins are found in major metazoan groups but not in nematodes, whereas class XII myosins are known only from nematodes, is at least consistent with a common origin of class XII/XV myosins. In contrast, the frequent clustering of class XII with other divergent myosins [like those of classes III, XVI, XVII, XVIII, and/or XX; see Fig. 1 (4, 5, 19)] is probably artifactual because of long-branch attraction, a notion supported by the increase of median statistical support for a monophyletic class XII/XV grouping from 74% to 87% (BS) and from 0 to 0.99 (PP) when class XX myosins (in addition to other rogue sequences) are excluded from the analysis (small dataset; see Table 2). For further details on the likely monophyly of classes XII and XV, see *Supporting Text*.

**Classes XIX, XX, and XXI: Three New Classes of Chordate, Insect, and Kinetoplastid Myosins.** The human genome encodes $>$40 myosin genes belonging to 11 classes. In addition, one predicted myosin sequence had already been noted as probably constituting a 19th myosin class (5). We show that this sequence and homologs from chordates such as the ascidian *Ciona intestinalis*, a fish, chicken, dog, and mouse, comprise a new myosin class XIX (100% BS), where two IQ motifs per sequence are the only neck/tail protein domains recognizable so far (Fig. 1). The fruit fly *Drosophila melanogaster* features 13 myosins, of which 12 belong to eight previously described classes, whereas one myosin (Myo29D, GenBank accession no. AAF52683) has so far been unclassified (20). Our analyses show that this myosin belongs to a so far insect-specific myosin class XX that is well supported (100% BS, Fig. 1). It currently comprises sequences from *D. melanogaster*, *Drosophila pseudoobscura*, and the malaria parasite vector *Anopheles gambiae* that feature one well conserved IQ motif in the neck domain (Fig. 1). The genomes of parasitic kinetoplastids like *Trypanosoma* and *Leishmania* now also reveal previously unrecognized myosins (21). *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major* all contain one class I myosin (Myo1) and one myosin (Myo2) belonging to a novel class XXI (21) (78% BS, median PP 1.0; Fig. 1 and Table 2). Six additional myosins are found only in *T. cruzi* and do not have direct homologs in the other kinetoplastids. Five of these fall into class XXI, whereas TcMyo8 is currently unclassified (21) (Fig. 1). Interestingly, the kinetoplastid myosin tails contain protein domains previously unknown to be associated with myosins, e.g., FYVE, WW, or UBA-like domains (Fig. 1). For a more detailed description of kinetoplastid myosins, see *Supporting Text*.

**Apicomplexan Myosins: A Naming Convention.** Researchers commonly have named new apicomplexan myosins in the order of discovery (MyoA, MyoB, etc.), often independently for each species (22–28). This has led to the confusing situation where nonorthologous apicomplexan sequences were designated with the same letter, e.g., PfMyoC (PfM-C) and TgMyoC of *Plasmodium falciparum* and *T. gondii*, respectively (25). We therefore propose a systematic naming convention for apicomplexan myosins, such that the same letters denote homologous myosins across taxa. Because *T. gondii* has the largest myosin repertoire (11 sequences) of any apicomplexan for which significant genome data are available, we base this naming system on homology to the *T. gondii* myosins (TgMyoA-TgMyoK) and consequently rename six apicomplexan myosins: MyoC, MyoD, and MyoF of *Plasmodium* and *Babesia* are changed to MyoF, MyoJ, and MyoK, respectively, and MyoB of *Babesia* and *Theileria* is changed to MyoH. For further details, see *Supporting Text* and Table 4, which is published as supporting information on the PNAS web site.

**Class XIV Includes Apicomplexan and Ciliate Myosins.** Class XIV had previously been described as exclusively comprising apicomplexan myosins and was divided into two subclasses (25, 26). In contrast, our analyses (Fig. 2) show a well resolved group of apicomplexan myosins (MyoH) to form a new subclass XIVc and consistently place 12 myosin sequences from the ciliate *Tetrahymena thermophila* within class XIV, an unsurprising affiliation, because ciliates and apicomplexans are close relatives (both belong to the Alveolata). This extended class XIV is very well supported by all phylogenetic inference methods used (95% BS, median PP 1.0; Fig. 2 and Table 2). That this robust association had not been recognized previously (4, 19, 25, 26), with the ciliate myosins recently having been assigned to a separate class (19), is most likely because of the inclusion of rogue sequences, small dataset size, and the use of full-length myosin sequences in previous analyses. Based on our strong statistical support for this extended class XIV and on independent corroborating evidence (see below), we now include these 12 ciliate myosins in class XIV as subclass XIVd (Fig. 2 and Table 1). TtMyo1 has been shown to be involved in phagosome motility and nuclear elongation (29, 30), and the other *T. thermophila* myosins have been described elsewhere (19); one of them (TtMyo13) remains unclassified (Fig. 1). Intriguingly, four ciliate myosins are predicted (Conserved Domain Database) to contain a domain with similarity to $\alpha$-tubulin suppressor (ATS1) and the related regulator of chromosome condensation (RCC1) proteins (31–33), a character shared with apicomplexan class XIVc myosins (Fig. 2), and thus represent the first examples of ATS1-like domains associated with myosin tails. Class XIVc myosins also feature six or more IQ motifs and include the previously unclassified BbMyoH (Table 4) (25, 26).

**Class XIV: *Plasmodium* Myosins and a Class-Specific Amino Acid Polymorphism.** Proteins of *P. falciparum* are often divergent compared with homologs of other organisms because of their strong

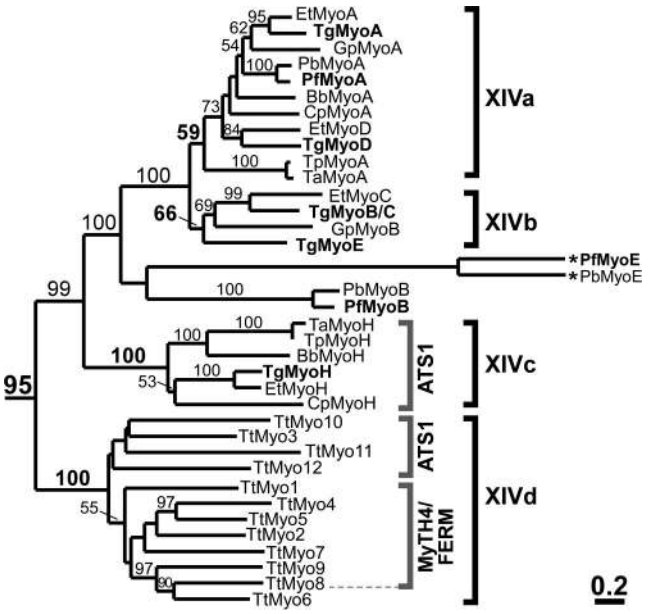


**Fig. 2.** Myosin class XIV phylogeny. This partial tree represents part of the phylogeny shown in Fig. 1. Note that sequences PfMyoE and PbMyoE had been omitted from the bootstrap analysis. ATS1 and MyTH4/FERM refer to myosin tail protein domains. Et, *E. tenella*; Gp, *G. polymorpha*; Bb, *Babesia bovis*; Ta, *Theileria annulata*; Tg, *T. gondii*; and Tp, *Theileria parva*.

amino acid composition bias (caused by the A+T rich genome) and the presence of numerous low complexity insertions (34, 35). Consequently, some *P. falciparum* myosins appear on long branches in phylogenetic trees (Fig. 1) and have been difficult to classify (25, 26), similar to the case of some *Dictyostelium discoideum* myosins. In contrast to previous classification, our analyses firmly place three *P. falciparum* myosins outside class XIV (see below), leaving only three sequences within class XIV: PfMyoA, PfMyoB, and PfMyoE (Fig. 2 and Table 1). Of these, PfMyoA is the unambiguous homolog of other apicomplexan MyoA myosins (subclass XIVa), whereas *Plasmodium* MyoB and MyoE myosins do not unambiguously cluster with any other apicomplexan sequences in particular (Fig. 2). Our analyses thus clarify the previously contradictory classification of the two subclasses XIVa and XIVb (25, 26), which are mostly confirmed as initially formulated (26). Importantly, the overall class XIV classification (Fig. 2) is strongly supported by a particular amino acid polymorphism in the otherwise highly conserved H*Y*AG sequence (position 584 in chicken skeletal muscle myosin II, GenBank accession no. AAB47555): of 230 non-class XIV sequences (including PfMyoF, PfMyoJ, PfMyoK, and Tt-Myo13), 220 (95.7%) feature a tyrosine or phenylalanine, and five sequences (2.2%) an (iso)leucine at the second position of this tetrapeptide (Table 5, which is published as supporting information on the PNAS web site). In contrast, all 37 class XIV myosins (including PfMyoB, PfMyoE, and the 12 class XIV sequences of *T. thermophila*) contain either a serine or a threonine at this position, a character shared by only two non-class XIV myosins (human and mouse class XVIII) (Table 5).

**Apicomplexan Myosins and Class VI.** The two nonclass XIV myosins of *Plasmodium* MyoJ and MyoK, as well as their apicomplexan homologs, form part of an extended class VI, whose monophyly is well supported in most analyses (Fig. 1, Table 1; 91% BS, median PP 1.0; lower BS in CLUSTAL trees and in analyses based on the small dataset; Table 2). Typically, class VI myosins feature an insert between the myosin head domain and predicted IQ motifs of 40- to 50-aa lengths that is thought to be responsible for the unusual directionality of these motors toward the minus end of actin filaments (36, 37), yet whether an equivalent insert is present in the apicomplexan class VI myosins is unclear because of the uncertainty in predicting apicomplexan IQ motifs, which may be divergent (28, 38), and the apparent lack of IQ motifs in some of these molecules.

**Classes XXII, XXIII, and XXIV: Three New Classes of Chromalveolate Myosins.** The apicomplexan MyoF sequences (Table 1), together with myosins from the diatom alga *Thalassiosira pseudonana*, form a monophyletic clade, the new class XXII, with at best only intermediate BS in PROTDIST-based analyses (79% BS; Fig. 1) but strong statistical support in CLUSTAL-generated trees and Bayesian analysis (>85% BS with CLUSTAL, median PP 1.00; Table 2). All apicomplexan class XXII sequences are predicted to feature an extended neck domain with multiple IQ motifs, a characteristic feature of myosins of class V and related classes VIII, XI, and XIII (3), classes that are often positioned in close proximity to class XXII in tree topologies (Fig. 1). In addition, apicomplexan class XXII sequences (but not the corresponding diatom myosins) for which a long-enough tail domain has been annotated or predicted (*T. gondii*, *Plasmodium* spp. *Theileria* spp.) contain four to six WD40 repeats near the C terminus of the tail, which constitutes another novel protein domain–myosin tail combination. In addition, the N terminus of the *Theileria* and *Plasmodium* MyoF tail domains is predicted to form coiled-coil structures. The putative common ancestry of several diatom sequences and the apicomplexan MyoF myosins is supported by a particular amino acid polymorphism in an otherwise highly conserved region, the LEK*S*R site (position 271 in chicken skeletal muscle myosin): in all 14 class XXII myosins included in our analysis, the fourth position in this pentapeptide is

**Table 1. Alveolate myosin (sub)classes**

| Class | Subclass | IQ motifs | Protein domains | *T. thermophila* | *T. gondii* | *E. tenella* | *C. parvum* | *P. falciparum* | *T. annulata* | *B. bovis* | *G. polymorpha* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| XIV | XIVa | 1 | No tail | — | TgMyoA | EtMyoA | CpMyoA | PfMyoA | TaMyoA | BbMyoA | GpMyoA |
| | | 0 | No tail | — | TgMyoD | EtMyoD | — | — | — | — | — |
| | XIVb | 0–1? | — | — | TgMyoB/C | EtMyoC | — | — | — | — | GpMyoB |
| | | 1 | No tail | — | TgMyoE | — | — | — | — | — | |
| | XIVc | 6–8 | ATS1 | — | TgMyoH | EtMyoH | CpMyoH | — | TaMyoH* | BbMyoH* | — |
| | — | 1 | No tail | — | — | — | — | PfMyoB | — | — | — |
| | — | 0 | Coiled-coil | — | — | — | — | PfMyoE | — | — | — |
| | XIVd | 0–3 | ATS1, MyTH4, FERM | TtMyo1-Myo12 | — | — | — | — | — | — | — |
| ? | | 0–1? | Coiled-coil | TtMyo13 | — | — | — | — | — | — | — |
| XXII | | 3–6 | WD40 | — | TgMyoF | EtMyoF | CpMyoF | PfMyoF* | TaMyoF | BbMyoF* | — |
| XXIII | | 1 | MyTH4 | — | TgMyoG | EtMyoG | — | — | — | — | — |
| XXIV | | 2 | Coiled-coil | — | TgMyoI | — | CpMyoI | — | — | — | — |
| VI | | 0 | — | — | TgMyoJ | EtMyoJ | CpMyoJ | PfMyoJ* | — | — | — |
| | | 2–3 | Coiled-coil | — | TgMyoK | — | CpMyoK | PfMyoK* | — | — | — |

*This sequence has been renamed in the present study; MyoF, MyoH, MyoJ, and MyoK were previously known as MyoC, MyoB, MyoD, and MyoF, respectively.

valine or alanine, a character shared by only 9 of the 253 non-class XXII myosins (mostly apicomplexan myosins; Table 5). In contrast, of the non-class XXII myosins, 83.4% feature a serine/threonine, and only another seven sequences (2.8%) contain any of the other aliphatic amino acids (glycine, leucine, and isoleucine) at this position (Table 5). Finally, two additional non-class XIV myosins of *T. gondii* and their apicomplexan homologs warrant establishing the two new myosin classes XXIII and XXIV. The MyoG sequences of the coccidians *T. gondii* and *Eimeria tenella* reproducibly form a very well supported clade (100% BS, median PP 1.00), class XXIII, whereas TgMyoI and a homolog from *Cryptosporidium parvum* establish the new class XXIV (100% BS, median PP 1.00). All four sequences are predicted to feature one to two IQ motifs, and the MyoG myosin tails contain a single MyTH4 domain (but no FERM domain). One diatom sequence is often associated with classes XXIII and XXIV in CLUSTAL-generated trees (80–84% BS) but not in other analyses (Table 2), so that its definitive phylogenetic position and classification remain unclear.

## Discussion

**Myosin Evolution.** All myosin classes observed today likely derive, by a process involving gene duplications and diversification, from one progenitor molecule in the distant past, and it has recently been proposed that the eukaryotic cenancestor (last common ancestor) already contained three different types of myosins: an ortholog to extant class I myosins, a second myosin containing a dilute (DIL) domain (found in extant classes V and XI), and a third myosin with MyTH4/FERM domains (in extant classes IV, VII, X, XII/XIV, XV, and XXIII) (2) (Fig. 3 and Table 3). Although our phylogenetic analyses tend to support this hypothesis, some aspects do not fit comfortably into this scenario: Several MyTH4/FERM domain containing myosins are found in classes (IV, XIV, and XXIII) that are not part of the main "MyTH4/FERM cluster" (classes VII, X, and XII/XV). Also, classes V and XI are not monophyletic in the trees but are interspersed with classes VIII, XIII, and XIX (which do not feature DIL domains). Yet these incongruities could be explained by a lack of resolution in the phylogenetic trees (statistical support for the deep-level clustering of before-mentioned classes is not significant) and/or by secondary loss(es) of characters and/or by lateral (partial) gene transfer. On the other hand, some myosin classes derive from relatively recent divergent evolution in narrow evolutionary lineages. In such cases, one may find myosins in a narrow systematic group (e.g., classes XII and XIII in *Caenorhabditis* and *Acetabularia*, respectively), whereas a closely related class (XV and XI, respectively) of more widespread distribution has apparently vanished from it. The most parsimonious assumption is that the narrowly distributed myosins represent orthologs of the "missing" myosin classes. For myosin classes XII and XV, such an assumption is corroborated by virtually identical tail domain architectures and tail domain (MyTH4) phylogenies, supporting the notion that myosin head, neck, and tail domains generally coevolve (39). Similarly, both our phylogenies and tail domain similarity argue strongly for a previously unrecognized common origin of classes II and XVIII. Although myosin evolution and classification are tightly linked, we think it would not be very helpful to change the existing classification of myosin classes II/XVIII or XII/XV, even in light of our findings. For more details on issues relating to myosin classification, see *Supporting Text*.

**Chromalveolate Myosins.** Apicomplexan myosins had been classified in class XIV, which did not encompass any nonapicomplexan sequences (3, 22–28). This study presents robust phylogenetic data that group apicomplexan myosins into several distinct classes and that classify myosins from nonapicomplexan taxa (in particular a ciliate and a diatom) in some of the same classes as apicomplexan myosins. These phylogenetic associations are not surprising, given the kinship of (chrom)alveolate organisms (9–11). So far, taxon sampling for the myosin dataset regarding heterokonts or chromistans in general has been restricted to the diatom *T. pseudonana*, making it advisable to wait with a definitive classification of some diatom myosins until more chromistan data are available. Preliminary glimpses in unpublished genome data of another heterokont organism, the oomycete *Phytophthora ramorum*, already promise an extraordinarily rich treasure trove of novel myosins: 24 myosins can readily be identified, many of which have completely novel protein domain architectures (2). Apart from a few apparent class I myosins, some of these sequences cluster with unclassified diatom sequences or appear to belong to the new chromalveolate myosin classes XXII, XXIII, or XXIV. The monophyly of chromalveolate organisms has not been easy to demonstrate (11, 12), and the phylogenetic clustering of chromalveolate myosins and/or myosin classes restricted to chromalveolate lineages would represent rare evidence from nuclear-cytoplasmic genes (as opposed to nuclear-plastid or plastid-encoded genes) supporting their monophyly. In fact, judging by the diatom and oomycete myosins, we expect more chromalveolate-specific myosin classes to be delineated in the future.

**Function of Apicomplexan Myosins.** Experimental investigations have so far focused on a narrow range of apicomplexan myosins, i.e., the class XIVa/b motors and their molecular interaction partners (reviewed in refs. 40 and 41). MyoA of *T. gondii* is known to be essential for gliding motility and host cell invasion (42), and the presence of highly conserved orthologs in all apicomplexans for which considerable genome data are available (Table 1) highlights the essential function of this motor. MyoD, a myosin very similar to MyoA, is found only in the coccidians *T. gondii* and *E. tenella* (Table 1). TgMyoD has similar properties as TgMyoA and may be involved in the cell motility of nontachyzoite life stages of *T. gondii* (28, 43). Closely related myosins of subclass XIVb are known from *T. gondii*, *E. tenella*, and *Gregarina polymorpha* (Table 1), and TgMyoB/C has been implicated in parasite cell division (44). The new subclass XIVc myosins (MyoH) feature tail domains with similarity to ATS1 and related regulator of chromosome condensation 1 (RCC1) proteins thought to act as guanine nucleotide exchange factors. RCC1 is known to bind to chromatin, and ATS1 may be involved in coordinating the microtubule state during the cell cycle in yeast (31–33), suggesting that class XIVc myosins may also play a role during mitosis, cytokinesis, or in other cell-cycle events. Class XXII myosin tails (MyoF) contain four to six WD40 repeats that are known to form $\beta$-propeller structures and have been implicated in diverse functions like signal transduction and transcriptional regulation. WD40 domains may target to centrosomes or the nucleolus or bind phosphoinositides (45, 46). MyTH4 domains are found in a range of myosin classes (Table 3) and bind, in conjunction with FERM domains, to microtubules (47), but it is unclear whether class XXIII myosins (MyoG), which possess a MyTH4 but no FERM domain, interact with the apicomplexan microtubule cytoskeleton. Class XXIV myosins (MyoI) remain most enigmatic: the CpMyoI sequence is predicted to contain coiled-coil forming regions and may represent a dimerizing myosin, whereas the function of its putative N-terminal SH3-like domain is unknown. In contrast, class VI myosins have been well studied. These motors are the only myosins known that move toward the minus end of actin filaments and have been implicated in diverse functions such as endocytosis, cell motility, and Golgi complex morphology and secretion (48). Yet function and directionality of the apicomplexan class VI myosins (MyoJ, MyoK) are speculative. Intriguingly, both the class XIV- and XXII-specific amino acid polymorphisms we have identified involve serine/threonine residues, and one may speculate that these changes are linked to unusual mechanisms of myosin regulation involving phosphorylation.

## Materials and Methods

**Data Sources.** To a previously published myosin dataset (4), we added sequences from the GenBank database (all accession nos.

are listed in Table 5) and from the following sources: ToxoDB (http://ToxoDB.org), The Institute for Genomic Research (www.tigr.org/tdb/t_gondii), the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/Eimeria/tenella, assembly_2004_07_27), and the Department of Energy Joint Genome Institute (http://genome.jgi-psf.org/ramorum1/ramorum1.home.html). For detailed acknowledgments of these institutions and individual people and for other data sources, please see *Supporting Text*.

**Cloning and Sequencing.** *T. gondii* tachyzoites (RH strain) were grown as described (28), and total RNA was prepared from extracellular parasites by using TRIzol (Invitrogen) or RNeasy Mini Kit (Qiagen, Valencia, CA). RT-PCR was performed by using the Titan One Tube RT-PCR Kit (Roche Applied Science, Basel) or SuperScript II (Invitrogen). Oligonucleotide sequences are listed in Table 6, which is published as supporting information on the PNAS web site, and PCR products were cloned into and sequenced within pGEM-T Easy (Promega). Myosin annotation from unpublished sources was checked by sequence comparisons (pairwise BLAST and CLUSTALX) and edited manually where necessary.

**Multiple Protein Sequence Alignments.** Sequence order was randomized before generating sequence alignments with CLUSTALX (Ver. 1.83) (7) by using default multiple alignment parameters and these "Fast-Approximate" pairwise alignment parameters: gap penalty = 4, K-tuple size = 1, top diagonals = 50, window size = 50. Before generating alignments for phylogenetic analysis, we identified (by inspecting CLUSTALX alignments and by extensive pairwise BLASTing) and deleted insertions that were present in very few or individual myosins from these sequences if the insertions impaired the alignment (e.g., for *Plasmodium*, *Dictyostelium*, and class XVIII). Alignment positions containing gaps in >50% of the sequences were excluded from phylogenetic analyses. Please see Datasets 1 and 2, which are published as supporting information on the PNAS web site.

**Phylogenetic Analysis.** Phylogenetic analyses were carried out by using CLUSTALX, the PHYLIP (Ver. 3.63) programs PROTDIST, NEIGHBOR, FITCH, SEQBOOT, CONSENSE, and PROML (8), and MRBAYES (Ver. 3.1) (14). The Jones–Taylor–Thornton amino acid substitution matrix was used in PROTDIST and PROML. $\gamma$ distribution parameters for four variable rate categories and the fraction of invariant sites were estimated with TREEPUZZLE (Ver. 5.2) (49). FITCH was run with the Fitch–Margoliash method and without global rearrangements. The input order of species for phylogenetic analysis was always randomized where this option is given. Bayesian phylogenies were inferred by using MRBAYES: Two independent runs with four chains each (one cold, three heated with heating parameter = 0.025) were run using the Jones fixed-rate model and a $\gamma$-shaped rate variation across sites with four rate categories plus invariable sites. Trees were sampled every 100 generations, and analyses were continued until at least 1,000 trees per run had been sampled where the average standard deviation of split frequencies (ASDoSF) between the two runs was <0.10 (convergence). Trees sampled before convergence or a minimum of 500 trees were discarded as burnin, and analyses in which the ASDoSF was still >0.10 after 200,000 generations were aborted and restarted. Phylogenetic trees were visualized with DRAWTREE (8) or TREEVIEW.

**Protein Domain Predictions.** Protein domains and IQ motifs were predicted by SMART, the Conserved Domain Database, and/or PFAM. For details, please see *Supporting Text*.

We thank Tony Hodge for his myosin alignment; many sequencing centers, people, and funding agencies for making unpublished sequence data available (in particular the *Toxoplasma* and *Eimeria* Genome Sequencing Consortia; see *Supporting Text*); Pierre Rivailler for improved *E. tenella* myosin annotation; Patrick Keeling for valuable discussions; Jennifer Pinder for helpful comments on the manuscript; Vital-IT (Swiss Institute of Bioinformatics), Lausanne, Switzerland, for access to computer clusters; and Jacques Rougemont for porting and optimizing phylogenetics software for the Vital-IT infrastructure. B.J.F. gratefully acknowledges the financial support of a European Molecular Biology Organization long-term fellowship, and D.S. gratefully acknowledges funding from the Wellcome Trust and the Swiss National Science Foundation. D.S. is an International Research Scholar of the Howard Hughes Medical Institute.

1. Thompson, R. F. & Langford, G. M. (2002) *Anat. Rec.* **268,** 276–289.
2. Richards, T. A. & Cavalier-Smith, T. (2005) *Nature* **436,** 1113–1118.
3. Vale, R. D. (2003) *Cell* **112,** 467–480.
4. Hodge, T. & Cope, M. J. (2000) *J. Cell Sci.* **113,** 3353–3354.
5. Berg, J. S., Powell, B. C. & Cheney, R. E. (2001) *Mol. Biol. Cell* **12,** 780–794.
6. Swofford, D. L. (2003) PAUP (Sinauer, Sunderland, MA).
7. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003) *Nucleic Acids Res.* **31,** 3497–3500.
8. Felsenstein, J. (2005) PHYLIP (Dept. of Genome Sciences, University of Washington, Seattle).
9. Cavalier-Smith, T. (1993) *Microbiol. Rev.* **57,** 953–994.
10. Cavalier-Smith, T. (1999) *J. Eukaryotic Microbiol.* **46,** 347–366.
11. Harper, J. T., Waanders, E. & Keeling, P. J. (2005) *Int. J. Syst. Evol. Microbiol.* **55,** 487–496.
12. Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 15507–15512.
13. Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., *et al.* (2005) *J. Eukaryotic Microbiol.* **52,** 399–451.
14. Huelsenbeck, J. P. & Ronquist, F. (2001) *Bioinformatics* **17,** 754–755.
15. Sanderson, M. J. & Shaffer, H. B. (2002) *Annu. Rev. Ecol. Syst.* **33,** 49–72.
16. Sellers, J. R. (2000) *Biochim. Biophys. Acta* **1496,** 3–22.
17. Keeling, P. J. & Slamovits, C. H. (2004) *Eukaryot. Cell* **3,** 1363–1369.
18. Chen, Z. Y., Hasson, T., Zhang, D. S., Schwender, B. J., Derfler, B. H., Mooseker, M. S. & Corey, D. P. (2001) *Genomics* **72,** 285–296.
19. Williams, S. A. & Gavin, R. H. (2005) *Cell Motil. Cytoskeleton* **61,** 237–243.
20. Tzolovsky, G., Millo, H., Pathirana, S., Wood, T. & Bownes, M. (2002) *Mol. Biol. Evol.* **19,** 1041–1052.
21. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., *et al.* (2005) *Science* **309,** 409–415.
22. Heintzelman, M. B. & Schwartzman, J. D. (1997) *J. Mol. Biol.* **271,** 139–146.
23. Heintzelman, M. B. & Schwartzman, J. D. (2001) *J. Parasitol.* **87,** 429–432.
24. Heintzelman, M. B. (2004) *Cell Motil. Cytoskeleton* **58,** 83–95.
25. Chaparro-Olaya, J., Margos, G., Coles, D. J., Dluzewski, A. R., Mitchell, G. H., Wasserman, M. M. & Pinder, J. C. (2005) *Cell Motil. Cytoskeleton* **60,** 200–213.
26. Lew, A. E., Dluzewski, A. R., Johnson, A. M. & Pinder, J. C. (2002) *Cell Motil. Cytoskeleton* **52,** 202–220.
27. Matuschewski, K., Mota, M. M., Pinder, J. C., Nussenzweig, V. & Kappe, S. H. (2001) *Mol. Biochem. Parasitol.* **112,** 157–161.
28. Hettmann, C., Herm, A., Geiter, A., Frank, B., Schwarz, E., Soldati, T. & Soldati, D. (2000) *Mol. Biol. Cell* **11,** 1385–1400.
29. Hosein, R. E., Williams, S. A. & Gavin, R. H. (2005) *Cell Motil. Cytoskeleton* **61,** 49–60.
30. Williams, S. A., Hosein, R. E., Garces, J. A. & Gavin, R. H. (2000) *J. Eukaryotic Microbiol.* **47,** 561–568.
31. Shields, C. M., Taylor, R., Nazarenus, T., Cheatle, J., Hou, A., Tapprich, A., Haifley, A. & Atkin, A. L. (2003) *Curr. Genet.* **44,** 184–194.
32. Kirkpatrick, D. & Solomon, F. (1994) *Genetics* **137,** 381–392.
33. Hutchins, J. R. A., Moore, W. J., Hood, F. E., Wilson, J. S. J., Andrews, P. D., Swedlow, J. R. & Clarke, P. R. (2004) *Curr. Biol.* **14,** 1099–1104.
34. Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. (2003) *Cell* **115,** 771–785.
35. Bastien, O., Lespinats, S., Roy, S., Metayer, K., Fertil, B., Codani, J. J. & Marechal, E. (2004) *Gene* **336,** 163–173.
36. Wells, A. L., Lin, A. W., Chen, L. Q., Safer, D., Cain, S. M., Hasson, T., Carragher, B. O., Milligan, R. A. & Sweeney, H. L. (1999) *Nature* **401,** 505–508.
37. Menetrey, J., Bahloul, A., Wells, A. L., Yengo, C. M., Morris, C. A., Sweeney, H. L. & Houdusse, A. (2005) *Nature* **435,** 779–785.
38. Herm-Gotz, A., Weiss, S., Stratmann, R., Fujita-Becker, S., Ruff, C., Meyhofer, E., Soldati, T., Manstein, D. J., Geeves, M. A. & Soldati, D. (2002) *EMBO J.* **21,** 2149–2158.
39. Korn, E. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12559–12564.
40. Soldati, D. & Meissner, M. (2004) *Curr. Opin. Cell Biol.* **16,** 32–40.
41. Soldati, D., Foth, B. J. & Cowman, A. F. (2004) *Trends Parasitol.* **20,** 567–574.
42. Meissner, M., Schluter, D. & Soldati, D. (2002) *Science* **298,** 837–840.
43. Herm-Gotz, A., Delbac, F., Weiss, S., Nyitrai, M., Stratmann, R., Tomavo, S., Sibley, L. D., Geeves, M. A. & Soldati, D. (2006) *J. Muscle Res. Cell Motil.*, in press.
44. Delbac, F., Sanger, A., Neuhaus, E. M., Stratmann, R., Ajioka, J. W., Toursel, C., Herm-Gotz, A., Tomavo, S., Soldati, T. & Soldati, D. (2001) *J. Cell Biol.* **155,** 613–623.
45. Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. (1999) *Trends Biochem. Sci.* **24,** 181–185.
46. Grasberger, H. & Bell, G. I. (2005) *Int. J. Biochem. Cell Biol.* **37,** 1421–1437.
47. Weber, K. L., Sokac, A. M., Berg, J. S., Cheney, R. E. & Bement, W. M. (2004) *Nature* **431,** 325–329.
48. Buss, F., Spudich, G. & Kendrick-Jones, J. (2004) *Annu. Rev. Cell Dev. Biol.* **20,** 649–676.
49. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002) *Bioinformatics* **18,** 502–504.