

Dyneins Across Eukaryotes: A Comparative Genomic Analysis

OnlineOpen: This article is available free online at www.blackwell-synergy.com

Bill Wickstead* and Keith Gull

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK

**Corresponding author: Bill Wickstead,
bill.wickstead@path.ox.ac.uk*

Dyneins are large minus-end-directed microtubule motors. Each dynein contains at least one dynein heavy chain (DHC) and a variable number of intermediate chains (IC), light intermediate chains (LIC) and light chains (LC). Here, we used genome sequence data from 24 diverse eukaryotes to assess the distribution of DHCs, ICs, LICs and LCs across Eukaryota. Phylogenetic inference identified nine DHC families (two cytoplasmic and seven axonemal) and six IC families (one cytoplasmic). We confirm that dyneins have been lost from higher plants and show that this is most likely because of a single loss of cytoplasmic dynein 1 from the ancestor of Rhodophyta and Viridiplantae, followed by lineage-specific losses of other families. Independent losses in *Entamoeba* mean that at least three extant eukaryotic lineages are entirely devoid of dyneins. Cytoplasmic dynein 2 is associated with intra-flagellar transport (IFT), but in two chromalveolate organisms, we find an IFT footprint without the retrograde motor. The distribution of one family of outer-arm dyneins accounts for 2-headed or 3-headed outer-arm ultrastructures observed in different organisms. One diatom species builds motile axonemes without any inner-arm dyneins (IAD), and the unexpected conservation of IAD I1 in non-flagellate algae and LC8 (DYNLL1/2) in all lineages reveals a surprising fluidity to dynein function.

Key words: Bayesian phylogenetics, dynein, eukaryotic evolution, IAD, IFT, OAD

Received 18 December 2006, revised and accepted for publication 24 August 2007, uncorrected manuscript published online 28 August 2007, published online 26 September 2007

Dyneins are force-generating adenosine triphosphatases (ATPases) that move along eukaryotic microtubules. One of the key cellular roles of the dynein family is in the movement of the axoneme – the highly conserved microtubule-based structure that provides the motility in all eukaryotic flagella and cilia. Most dyneins so far identified appear to

belong to this axonemal class, but there is also a highly conserved cytoplasmic class of dyneins that act as important minus-end-directed motors in processes including vesicle transport, organellar positioning, mitotic spindle organization and chromosome segregation (reviewed in 1–3).

Dynein motors are not homologous to the other eukaryotic microtubule motor superfamily, the kinesins – or to the myosin motors of the actin cytoskeleton – being instead part of the broad ATPase associated with various cellular activities (AAA⁺) superfamily. Dyneins also have both a very different structure and mode of action to that of kinesins. Whereas, the kinesin motor domain is relatively small (~400 aa) and globular; the dynein motor region is very large (~3000 aa) and forms a ring of six AAA⁺ modules, the first of which is the site of ATPase activity (4,5). Structural studies of purified dynein molecules suggest that the dynein motor acts as a winch by rotation of the dynein head, relative to the N-terminal tail (6,7).

Dynein complexes consist of 1, 2 or 3 dynein heavy chains (DHCs, alternatively often abbreviated as DYH or DYN), which are large polypeptides (~500 kDa), each containing a single motor domain. The tail region of DHCs binds a variable number of smaller subunits – light chains (LCs; 8–30 kDa), light intermediate chains (LICs; 30–60 kDa) and intermediate chains (ICs; 60–140 kDa) – which regulate dynein complex activity and aid in tethering of cargo. The motor domain of all DHCs is well conserved, but particular kinds of DHC appear to be associated with specific cellular functions. The ICs are also a family of homologous proteins (albeit, slightly less conserved in sequence than the DHC motor) and again have at least some specificity for particular dyneins. Contrastingly, dynein LCs are not a single family of proteins but several. The LCs are also apparently more promiscuous than either ICs or DHCs, with particular LCs being found in several dynein complexes (8).

Here, we have used publicly available genome sequence data from 24 diverse eukaryotes to perform the first assessment of the distribution of DHCs, ICs, LICs and LCs across five of the six proposed eukaryotic supergroups (9). We present a Bayesian phylogeny for the complete repertoire of DHCs and ICs in these 24 organisms, with extensive support from other methods, and compare the distribution of HC families to those of LICs, LCs and IFT proteins.

Re-use of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.

Results and Discussion

Bayesian DHC, IC and LC phylogenies

To assess the dynein repertoire across a broad range of eukaryotes, we selected 24 disparate organisms for which complete or near-complete genome sequences are publicly available. These organisms were: the Metazoa *Homo sapiens* (10), *Takifugu (Fugu) rubripes* (11), *Drosophila melanogaster* (12) and *Caenorhabditis elegans* (13); the yeasts *Saccharomyces cerevisiae* (14) and *Schizosaccharomyces pombe* (15); the Amoebozoa *Entamoeba histolytica* (16) and *Dictyostelium discoideum* (17); the kinetoplastids *Trypanosoma brucei* (18) and *Leishmania major* (19); the diplomonad *Giardia lamblia* (www.mbl.edu/Giardia); the ciliate *Tetrahymena thermophila* (20); the diatoms *Thalassiosira pseudonana* (21) and *Phaeodactylum tricornutum* (www.jgi.doe.gov); the oomycete *Phytophthora sojae* (22); the Apicomplexa *Plasmodium falciparum* (23), *Cryptosporidium parvum* (24) and *Toxoplasma gondii* (www.toxodb.org); the red alga *Cyanidioschyzon merolae* (25); the green algae *Chlamydomonas reinhardtii* (www.jgi.doe.gov) and *Ostreococcus lucimarinus* (www.jgi.doe.gov); and the higher plants *Arabidopsis thaliana* (26), *Populus trichocarpa* (27) and *Oryza sativa* (28). At the time of analysis, a published draft genome sequence was available for all but five of these organisms (*C. reinhardtii*, *G. lamblia*, *P. tricornutum*, *O. lucimarinus* and *T. gondii*). A list of the sources and versions of the data used is given in File S1.

Using the Pfam 'dynein heavy' domain (PF03028) model, we extracted a set of DHCs predicted to be encoded in each of the 24 genomes. Considering the size of the dynein motor domain (the most highly conserved portion of the primary sequences) is ~3000 aa, the PF03028 model is quite short (783 aa) as it covers only the C-terminal part of the motor domain. Importantly, a number of good candidates for dynein motors do not produce a hit to this model. To ensure that we had the full repertoire of DHCs for each organism, we used this initial set to define new hidden Markov models (HMMs) for specific sets of DHCs and used these to reinterrogate the sequences encoded by the 24 genomes. Figure 1 shows the results of this analysis. As expected, our new, longer models produce higher scores for all sequences (whether DHCs or not). More importantly, the Pfam model PF03028 performs poorly in the identification of particular groups of DHCs – for instance, cytoplasmic dynein 1 and 2 HCs are especially poorly predicted by PF03028 – but produce strong signals with the group-specific models. A number of protistan sequences of all DHC types are also not found by the Pfam dynein heavy model. Both these observations are the result of bias in the databases used in the construction of the Pfam model: bias towards axonemal DHCs (as there are more families than for cytoplasmic dyneins) and towards metazoan sequences (as there are more sequences available). The situation is likely to improve as more

varied species are incorporated into the databases, but currently, the model PF03028 misses several DHC sequences that can be clearly identified with our HMMs.

The DHC homologues can also be identified using BLAST similarity searches, but no small set of sequences can be used to unambiguously detect all DHCs from the complete set of 24 organisms (data not shown), making the group-specific HMMs the best current method to identify DHC sequences. We adopted a liberal score threshold (Figure 1) to extract the complete repertoire of DHCs, predicted to be encoded by the 24 eukaryotes (File S2) and used them to infer a Bayesian phylogeny, supported by partial Bayesian bootstrap replicates and full replicates using maximum-likelihood (ML), neighbour-joining (NJ) and maximum-parsimony (MP) approaches (see *Materials and Methods*). The full DHC phylogeny can be seen in Figure 2, and topology support for all nodes using all methods is supplied in File S3. Our phylogeny is consistent with the classification of DHCs into two groups: cytoplasmic and axonemal. The cytoplasmic group encompasses two families – DYNC1H1 (cytoplasmic dynein 1) and DYNC2H1 (cytoplasmic dynein 2) – of which the latter has the more restricted distribution, being found only in organisms that build axonemes at some point in their life cycle (see below). Previous works have divided the axonemal dyneins into six families – outer-arm dynein (OAD) α , OAD β , OAD γ , inner-arm dynein (IAD)-1 α , IAD-1 β and the single-headed dyneins (29). More recent work has suggested that there may be more families within the single-headed category (30). The analysis presented here utilizes the data emerging from several recent eukaryotic genome sequencing projects along with sophisticated phylogenetic methods to greatly extend these analyses. This extended analysis suggests that the OAD α , OAD β and OAD γ groups in fact only encompass two well-supported ancestral families, members of both of which are found in all organisms included in our analysis that build motile axonemes (discussed in more detail below). Our analysis also suggests that the single-headed DHCs can be classified into three distinct families. Here, we call these groups IAD-3, IAD-4 and IAD-5 to follow on from the two 2-headed IAD-1 families. The distribution of these dynein families is discussed in more detail below.

The core of most of the DHC groups is well supported by all methods used (see File S3); although, the placing of three divergent sequences (Plafa_PF14_0626, Giala_37985 and Giala_DHC; indicated in Figure 2) is not consistent between methods and shown to be treated with caution. However, all of the DHC sequences can be placed into one of the nine dynein families identified here with reasonable confidence.

To complement the DHC analysis, we also inferred phylogenies for the dynein IC superfamily and the multigene Tctex1/Tctex2/LC9/LC19 (DYNLT1/2) LC family (Figure 3

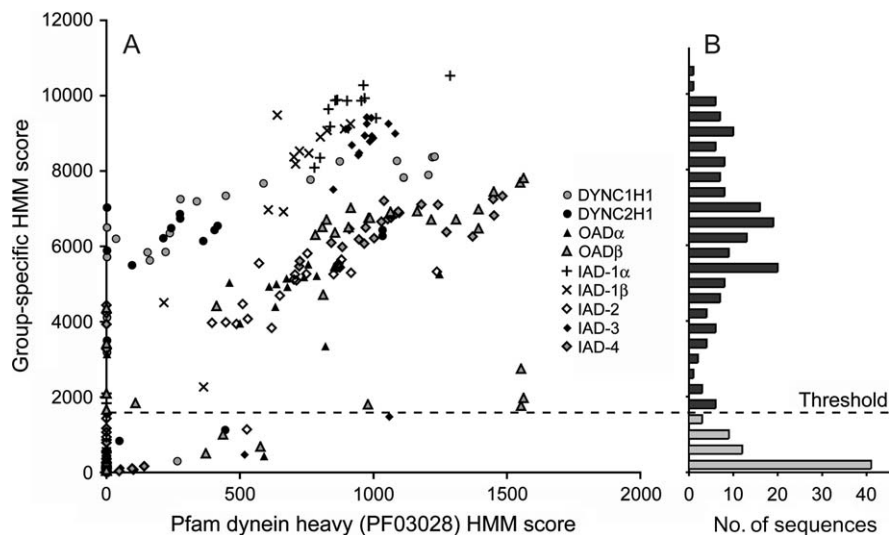


Figure 1: Use of HMMs to identify DHCs encoded in the genomes of 24 diverse eukaryotes. A) Performance of the Pfam dynein heavy HMM, PF03028.5, against group-specific HMMs. The y-axis shows score for group-specific HMM giving highest-scoring match. All predicted polypeptides in the 24 genomes with a score >0 on either axis are shown. B) Histogram of the distribution of matches to group-specific HMMs used to define the dynein heavy dataset (score > threshold). On the basis of the distribution, a liberal threshold was chosen to encompass the vast majority of dynein-like sequences without the inclusion of excessive numbers of false positives and extremely divergent sequences that accumulate in the low-score tail.

and Files S4 and S5). From the same dataset of 24 eukaryotes, we used an iterative HMM searching approach (see *Materials and Methods*) to identify 67 non-redundant ICs and 51 non-redundant DYNLT1/2 LCs from the predicted proteomes of these organisms and used them to infer Bayesian phylogenies (again supported by ML and NJ approaches). On the basis of these phylogenies, ICs can be divided into six well-supported groups – DYNC111/2 [dynein intermediate chain1 family (DIC1)], IC70 (IC2/ODA6), IC78 (IC1/ODA9), IC140 (IDA7), IC138 (BOP5) and a new IC family WDRD34 (named for the human WD-repeat domain 34-containing protein it contains). As for DHCs, there is an apparent division between cytoplasmic and axonemal clades. The DYNC111/2 family contains sequences known to be components of the cytoplasmic dynein 1 (DYNC1) complex. In contrast, it has been shown that in *Chlamydomonas*, IC70 and IC78 are components of the OAD complex (31–35), whereas IC138 and IC140 associate with IAD I1 (36–38). The newly identified clade WDRD34 contains sequences that are, as far as the authors are aware, of unknown function.

The Tctex1/Tctex2/LC9/LC19 phylogeny encompasses just two well-supported clades – Tctex1/LC9 (DYNLT1) and Tctex2/LC19 (DYNLT2). Several organisms encode more than one example of each family, but generally in the analysis presented here, they do not fall into well-supported subgroups in contrast to the situation suggested by previous more limited phylogenies (39). Only the Tctex1/LC9 (DYNLT1) family contains non-flagellate organisms on the basis of this analysis.

No dyneins in higher plants, red algae or *Entamoeba*

Lawrence et al. (40) have previously noted that there are no DHCs in the genome sequence available for *Arabidopsis*, leading to the suggestion that higher plants have dispensed with the dynein motor. This was subsequently called into question by the identification of four non-redundant sequences from DHC genes (encoding axonemal dynein family members) in the shotgun sequence generated as part of the rice (*O. sativa*) genome project, (41) raising the possibility that the situation may be more complicated than the *Arabidopsis* sequence analysis might imply. However, as the sequencing of the rice genome has progressed, these sequences have failed to assemble into any of the large contigs – the version 5 assembly, released January 2007 (<http://www.tigr.org/tdb/e2k1/osa1/>), covers 372 Mb of the estimated 430 Mb euchromatic genome – and no additional reads with good similarity to DHCs have emerged (data not shown). This makes it highly likely that the potential DHC gene sequences found during the rice genome sequencing effort are from contaminating DNA and not the plant itself.

We find no evidence in the assembled *Arabidopsis*, *Populus* or rice genomes for genes encoding DHCs. This is true, both for the predicted protein dataset and when using tBLASTn searches of the DNA itself (data not shown). Moreover, we find no good evidence for ICs, LICs or any LCs, except LC8 (DYNLL1/2). The presence of LC8 has been previously noted in plants, and this LC is notable in being highly conserved in all 24 eukaryotic genomes analyzed here, regardless of DHC complement (Figure 4B).

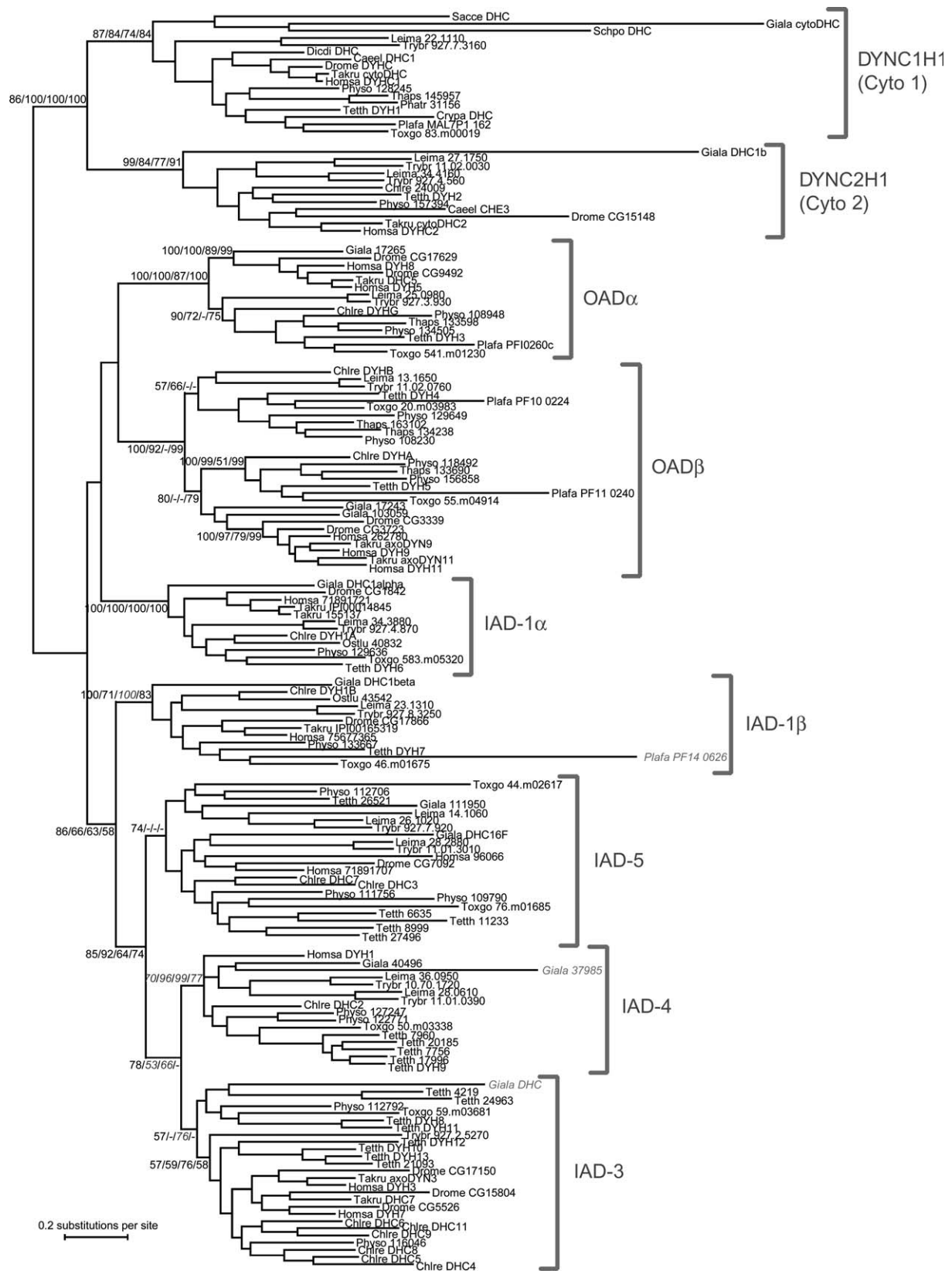


Figure 2: Legend on next page.

Taken together, these data strongly suggest that the original observation of the lack of dyneins in higher plants (40) is correct.

Not only have flowering plants dispensed with the dynein family of motors but also our analysis suggests that the loss of cytoplasmic dynein 1 (DYNC1) predates the divergence of the Archaeplastida (i.e. the land plants, green and red algae and glaucophytes). This can be seen in the lack of genes encoding either DYNC1H1 or DYNC111/2 families in the genomes of the red alga *C. merolae* and green algae *C. reinhardtii* and *O. lucimarinus* – although *Chlamydomonas* has retained cytoplasmic dynein 2 [Figure 2 and (42)]. *Chlamydomonas* has retained the dynein families associated with the flagellum (see below), and alongside the 15 DHCs in the analysis shown in Figure 2 – plus one protein (id: 37963) that is 100% identical to Chlre_DYHG – the Joint Genome Institute (JGI) protein models for this organism (v3.0; <http://genome.jgi-psf.org/Chlre3/>) include one fragmentary sequence that was too short to be included in the analysis. We were concerned that this protein (id: 115120) might be a fragment of a DYNC1H1 protein. While only fragmentary in the assembly v2.0/v3.0 protein models (and the v3.1 gene catalog), there is a 'JAMBOREE' annotation for this *Chlamydomonas* locus that encodes a full length DHC (id: 206178). Phylogenetic inference using this sequence and a subset of dynein sequences from the larger DHC phylogeny shows clearly that this sequence does not represent a 'missing' DYNC1H1 but is a divergent member of the axonemal dyneins (File S6). It thus appears that *Chlamydomonas*, which is such an important model for axonemal dynein action, lacks the cytoplasmic dynein counterpart.

Although the flagellate *C. reinhardtii* has retained all dynein families except cytoplasmic dynein 1, the non-flagellate red alga *C. merolae* – like *Arabidopsis*, *Populus* and rice (and presumably all other angiosperms) – has no dynein motors at all. In the light of these data, we believe that the isolation by polymerase chain reaction of DHC gene fragments putatively from *Nicotiana* (43) should be treated with caution; there are no identifiable DHC sequences currently in the publicly available *Nicotiana tabacum* expressed sequence tag (EST) sequence database (www.estarray.org).

Interestingly, flowering plants and red algae are not the only lineages to have dispensed with dyneins altogether. As well as the apparent loss of DYNC1 in the ancestor of the Archaeplastida, a second, independent loss of DYNC1 has occurred in *E. histolytica* – leaving this non-flagellate organism also devoid of any dynein motors. In agreement with these data, *E. histolytica*, *C. merolae*, *A. thaliana*, *P. trichocarpa* and *O. sativa* genomes are all free of identifiable homologues of dynein ICs and all LCs except LC8 (Figure 4B). The occurrence of LC8 (DYNLL1/2) in organisms not possessing dynein motors demonstrates that although this LC is a component of DYNC1 (44,45) and also axonemal dyneins and radial spokes (8,45–48), its role is not limited to functions associated with dynein. Interestingly, the budding yeast LC8 orthologue, Dyn2p (YDR424C), is found in cytoplasmic dynein complexes but also associated with nuclear pore components (49).

As might be expected, the occurrence of the DIC1, which contains both mammalian DYNC111 and DYNC112 proteins, follows the presence of DYNC1 HC reasonably well. However, we found no good homologue of DYNC111/2 in budding yeast (*S. cerevisiae*) or in the three Excavata (*T. brucei*, *L. major* and *G. lamblia*) – each of which have apparently dispensed with this protein without losing DYNC1 in its entirety (Figure 4B). The *S. cerevisiae* protein Pac11p is a WD-repeat-containing protein (as are dynein ICs) that associates with dynein (50) and has been proposed to be the budding yeast orthologue of DYNC111/2 (51). However, Pac11p is less similar to the ICs from other organisms in our analysis than are other WD-repeat-containing proteins from yeast, which are very unlikely to be dynein subunit (File S7). Pac11p also fails to form a reciprocal-best-BLAST (RBB) match to any of the other ICs (File S8). It is thus not clear from our analysis if Pac11p is a very divergent DYNC111/2 family member or a different WD-repeat protein co-opted into the dynein complex. We also found no good homologue of the LC Tctex1/LC9 (DYNLT1) family in yeast (Figure 4B). This LC is missing from organisms that lack all DHCs (and also the green alga *Ostreococcus*, which is a special case, as described below). However, its presence in *Chlamydomonas* implies that it is not specific for DYNC1.

Figure 2: A Bayesian phylogeny for the DHC sequences from 24 diverse eukaryotes. Prefixes: Caeel, *Caenorhabditis elegans*; Chlre, *C. reinhardtii*; Crypa, *C. parvum*; Dicdi, *D. discoideum*; Drome, *D. melanogaster*; Giala, *G. lamblia*; Homsa, *H. sapiens*; Leima, *L. major*; Ostlu, *O. lucimarinus*; Phatr, *P. tricornutum*; Physo, *P. sojae*; Plafa, *P. falciparum*; Sacce, *S. cerevisiae*; Schpo, *S. pombe*; Takru, *T. rubripes*; Tetth, *T. thermophila*; Thaps, *T. pseudonana*; Toxgo, *T. gondii*; Trybr, *T. brucei*. (No DHC: *A. thaliana*, *C. merolae*, *E. histolytica*, *O. sativa* and *P. trichocarpa*). For display, the tree has been rooted by bisecting the longest internal branch, although the true position of the root is unknown. Topology support for selected nodes is indicated (Bayesian partial bootstraps/ML/NJ/MP). Bootstrap values give a conservative estimate of the confidence that a particular group of sequences are monophyletic (94). Generally, groups with >90% bootstrap support were considered to be well supported and those with >70% bootstrap support to have some support. Italicized grey values give clade support, excluding the similarly highlighted sequences. Bootstrap values for all nodes from all four inference methods are given in File S3. Cyto 1, cytoplasmic dynein 1; Cyto 2, cytoplasmic dynein 2.

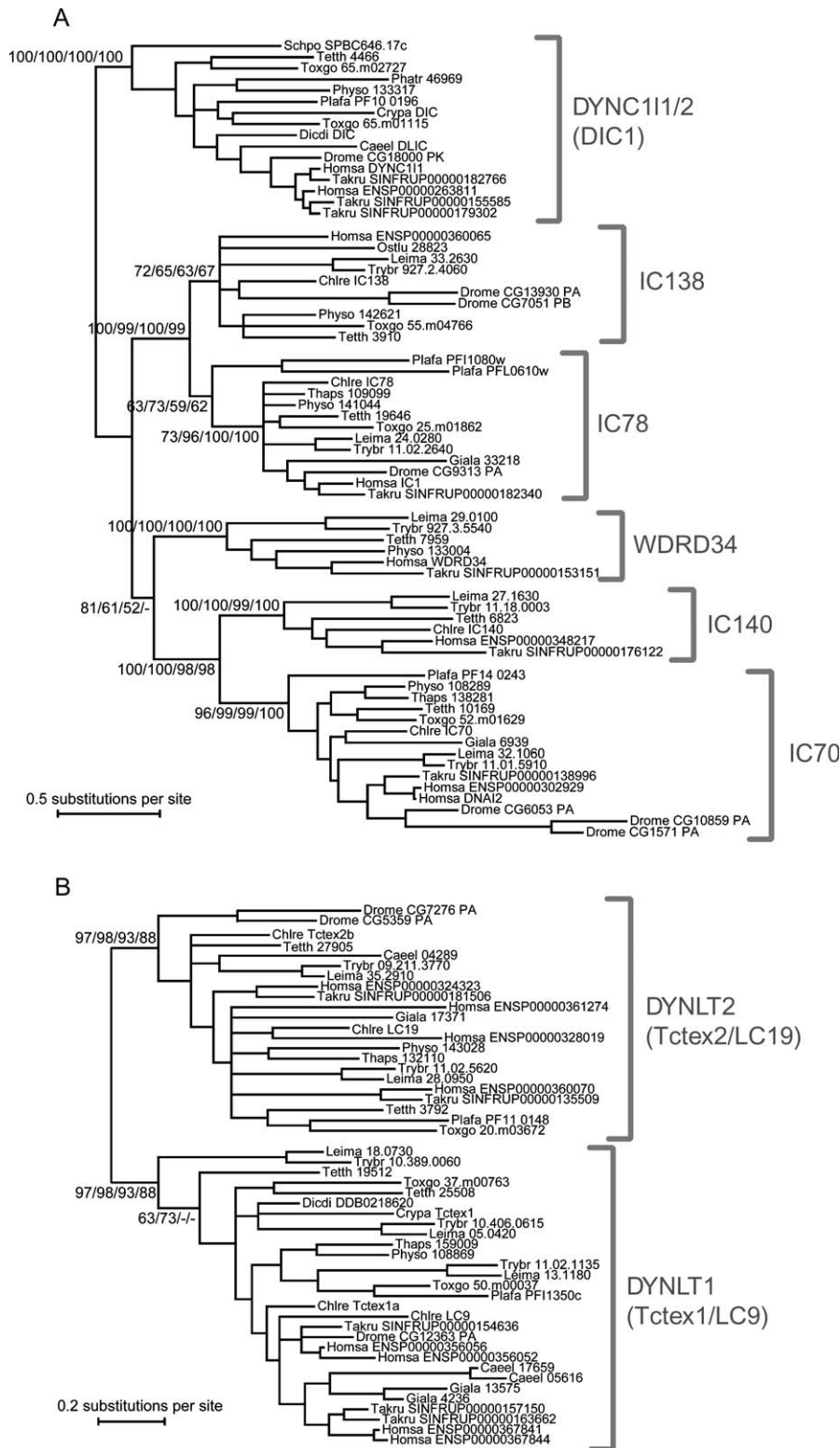


Figure 3: Bayesian phylogenies for: A) dynein IC and B) Tctex1/Tctex2 family LC sequences from 24 diverse eukaryotes. Prefixes as in legend to Figure 2. For display, trees have been rooted by bisecting the longest internal branch. Topology support for selected nodes is indicated (Bayesian partial bootstraps/ML/NJ/MP). Bootstrap values for all nodes under all four methods are given in Files S4 and S5.

Cytoplasmic dynein 2 and intraflagellar transport

Heavy chains of cytoplasmic dynein 2, known as DHC1b in *Chlamydomonas* and DYNC2H1 in the new mammalian nomenclature (52), are more similar at the level of primary

sequence to the DYNC1H1 family than to axonemal HCs (data not shown) and form a single clade with the DYNC1H1 family on an unrooted tree (Figure 1). However, the DYNC2H1 family is monophyletic and distinct from that

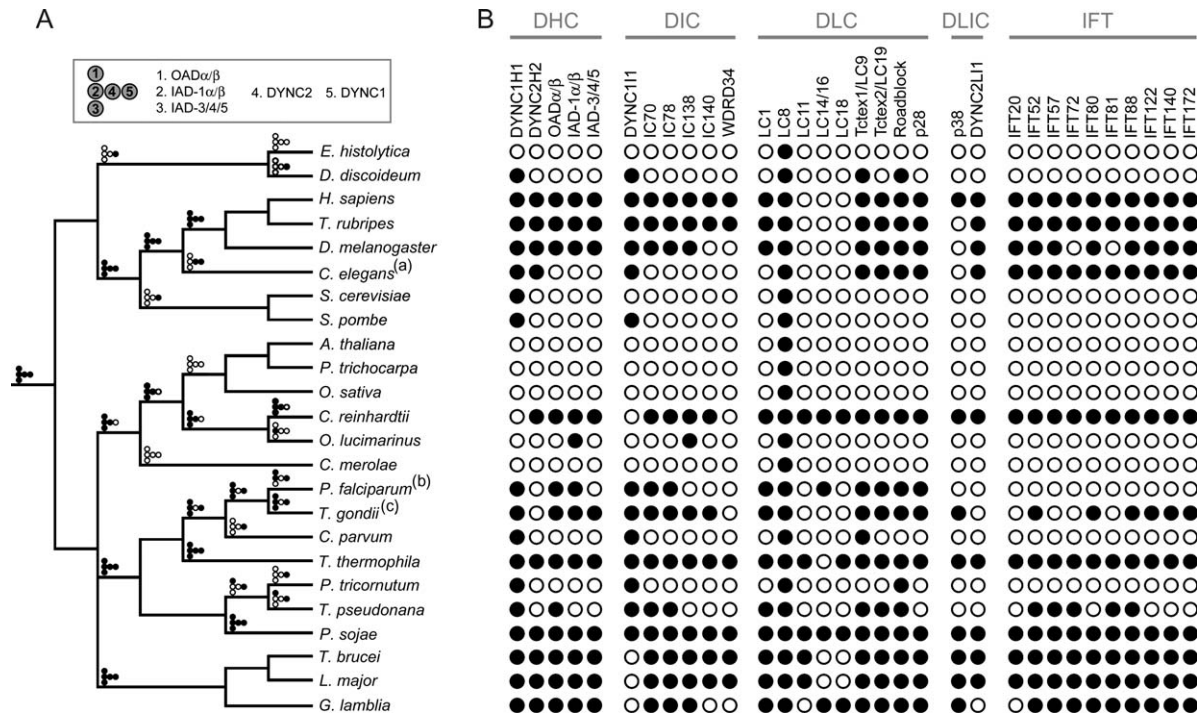


Figure 4: The distribution of dynein and IFT components across 24 diverse eukaryotes. A) Cladogram showing the likely evolutionary relationships of the organisms analysed and the inferred DHC repertoire in ancestral organisms (changes to the repertoire are shown). B) Presence (dot) or absence (circle) of identifiable orthologues of five DHC classes (DHC); six IC groups (DIC); nine LC groups (DLC); one LIC (DLIC); and 10 components of the IFT system (IFT). The names LC11, LC14, LC16, LC18 and LC19 refer to the LCs identified as 'M_t = 11 000', 'M_t = 14 000', etc., isolated from *Chlamydomonas* flagella (47,74,95). Orthologues were identified by RBB analysis (File S7) and, where necessary, iterative-HMM searches followed by phylogenetic inference (Figure 3). Grey dots indicate sequences failing the iterative-HMM cutoff but giving reciprocating BLAST-hits. Organisms building flagella/cilia are shown in bold. Notes: (a) *Caenorhabditis elegans* cilia are immotile; (b) *Plasmodium falciparum* builds flagella by an IFT-independent mechanism; (c) Briggs et al. (56) suggest that there may be cryptic orthologues of IFT57 and IFT72 encoded in the *T. gondii* genome that are not found in the predicted protein datasets used here.

of DYNC1H1. In spite of its name, the primary role of cytoplasmic dynein 2 is in building and maintaining cilia/flagella as part of the intraflagellar transport (IFT) machinery – providing the essential retrograde motor to complement the anterograde movement of motors of the kinesin-2 family (reviewed in 53). The dominance of this role is evident in the distribution of DYNC2 HCs: absent from all organisms that don't build cilia/flagella at some stage in their life cycle. However, our phylogeny reveals that dyneins of this type are also absent from some flagellum-building organisms: *P. falciparum*, *T. gondii* and *T. pseudonana*. This finding does not appear to be a result of erroneous gene models or genome sequence gaps because these three organisms also lack the LIC, D1bLIC (DYNC2LI1; Figure 4B), which is specific for DYNC2 (54). Given the most likely evolutionary relationships between the organisms (Figure 4A), the most parsimonious explanation for the observed distribution of DYNC2 would be a single loss in the common ancestor of *P. falciparum* and *T. gondii* and a second independent loss in *T. pseudonana*.

The microgametes of the malaria parasite *P. falciparum* build their flagella in the cytoplasm (55), and Briggs et al.

(56) have previously demonstrated that this occurs through an IFT-independent mechanism and that *P. falciparum* lacks IFT proteins (including kinesin-2 motors). Therefore, the finding here of an absence of DYNC2 in *P. falciparum* is predictable based on the known biology of this organism and was expected. However, the lack of DYNC2 in *T. gondii* and *T. pseudonana* was entirely unexpected as both organisms possess at least some of the central components of IFT, including the kinesin-2 motor [Figure 4B and (56,57)]. As this would strongly suggest that IFT is still active in these organisms, it raises the obvious question as to what in these organisms is providing the retrograde motor function? Alternatively, is there something about the biology of these organisms – which build flagella only as part of gametogenesis – that relieves the need for a retrograde motor?

Axonemal dyneins

Most dyneins belong to the axonemal class. In contrast to the mere two families of cytoplasmic DHCs, we find seven identifiable well-supported families of axonemal DHCs. Unsurprisingly, these axonemal families are mostly restricted to organisms that build motile flagella or cilia at

some stage of their life cycle (see above for discussion of reports of axonemal dyneins in higher plants and below for discussion of *Osteococcus*). There have been multiple losses of axonemes during eukaryotic evolution, and there have been multiple concomitant losses of axonemal dyneins. This is in contrast to the limited number of losses of cytoplasmic dynein 1. However, from the distribution of DHC families in extant eukaryotes (Figure 4), it is likely that the common ancestor of all eukaryotes had a functioning axoneme and possessed all the DHC families described in this work. Specific lineages have then lost, rather than gained, particular families and functions. The nematode *C. elegans* is an interesting case as it builds cilia in certain, differentiated cell types, but they are all immotile. In keeping with this biology, the *C. elegans* genome encodes homologues of both cytoplasmic dynein families, and also a footprint for the IFT system, but no axonemal DHCs.

Axonemal DHCs can be grouped into three classes: HCs of the axonemal OADs, 2-headed IADs and single-headed IADs. The OADs have been previously described as encompassing three DHC families – OAD α , OAD β and OAD γ (29). Rather confusingly, the nomenclature for the OADs is not consistent between species (for example, *Chlamydomonas* OAD α is the orthologue of *Tetrahymena* OAD γ , and *Chlamydomonas* OAD γ is the equivalent of *Tetrahymena* OAD α). Our more extensive phylogeny provides strong support for only two ancestral groups: (i) the OAD α family – encompassing the innermost of the OADs from all species – including *Tetrahymena* OAD α (Tetth_DYH3), metazoan OAD α and *Chlamydomonas* OAD γ ; and (ii) the OAD β family – encompassing both *Tetrahymena* OAD β (Tetth_DYH4) and OAD γ (Tetth_DYH5; orthologous to *Chlamydomonas* OAD α).

To test the finding of only two well-supported OAD families, we created a sub-phylogeny from a new alignment of the sequences predicted to be in the OAD β family with the addition of the OAD β from sea urchin, which was the first of the family to be sequenced (58,59) and is therefore a good point of reference for the other members. By excluding other less-related DHC families, it is possible to increase the number of well-aligned sites for phylogenetic reconstruction of this part of the tree. The results from five different tree-building methods are shown in Figure 5. From the phylogenies, it can be clearly seen that, while there is a support for three subfamilies – metazoan OAD β , non-metazoan OAD β , and non-metazoan OAD γ (*Chlamydomonas* OAD α) – there is no good evidence that the metazoan and non-metazoan OAD β subfamilies are orthologous. Indeed, the Bayesian and NJ methods weakly support metazoan OAD β being an orthologue of non-metazoan OAD γ . Only a tree constructed using a distance metric derived from BLASTp scores (see *Materials and Methods*) groups metazoan and non-metazoan OAD β subfamilies, which may go some way to explain the initial annotation. The tree based on BLAST scores is fundamen-

tally different from the four others shown in Figure 5, in that, it is not based on a multiple sequence alignment and not underpinned by a model of sequence evolution. The lack of dependence on an optimized sequence alignment might be considered an advantage under certain conditions, but such methods are not generally good at reconstructing good phylogenies, and the inferred tree has several sequences in evolutionarily improbable positions. It is not clear from the phylogenies presented here whether the metazoan ancestor lost a previously present OAD γ sequence, or if the metazoan OAD β subfamily represents the ancestral state, and the occurrence of the non-metazoan OAD β and OAD γ subfamilies is the result of an ancient gene duplication in the ‘bikont’ lineage alone.

There is a difference between the dynein outer-arm composition of *Chlamydomonas*, *Tetrahymena* and *Paramecium*, which are 3 headed (60–63), and those of the sperm of several animal species, which contain only 2-headed OADs (64–70). The occurrence of the subfamilies in the OAD β family appears to account for this difference – species with 3-headed OADs possess OAD α and members of two subfamilies of OAD β , while species with 2-headed OADs possess OAD α and only one OAD β subfamily. We suggest further that the sequence data can predict the possession of 3-headed/2-headed OADs in other organisms. Hence, we expect that animal axonemes in general (not just sperm tails) will be 2 headed, whereas the axonemes of apicomplexans, diatoms and oomycetes will be 3 headed like those of *Tetrahymena* and *Chlamydomonas*. Trypanosomes and possibly *Giardia*, despite being non-metazoan organisms, possess members of only one OAD β subfamily – either through retention of the ancestral state or through secondary loss of the outermost OAD γ (Figure 5) – so are predicted to have a 2-headed OAD composition reminiscent of animal axonemes. It will be interesting to see if the prediction of this simple distinction is borne out by experiment because for most axonemes the OAD composition is still unknown.

All the organisms in our analysis that encode OAD family members (which are, unsurprisingly, all the organisms that possess motile flagella/cilia) also encode identifiable orthologues of IC70 (IC2/ODA6), IC78 (IC1/ODA9) and LC1 (Figures 3 and 4B). However, no other dynein IC or LC is found exclusively and consistently in this set of organisms – in agreement with the finding of LC8 (DYNLL1/2), Tctex2 (DYNLT2) and Roadblock (DYNLRB1/2/LC7) LCs in other dynein complexes (45–47,71–74).

The *Chlamydomonas* 2-headed IAD complex (also called I1 or *f*-dynein) contains one HC from each of the IAD-1 α and IAD-1 β families as well as the ICs IC138 and IC140 (36–38,46,75,76). The LCs associated with the 2-headed IAD complex are generally not specific to this dynein class (LC8, Roadblock, Tctex1 and Tctex2 family members). The IAD-1 α and IAD-1 β families do not appear to form a single clade in our phylogeny or in previous work (29,30). As

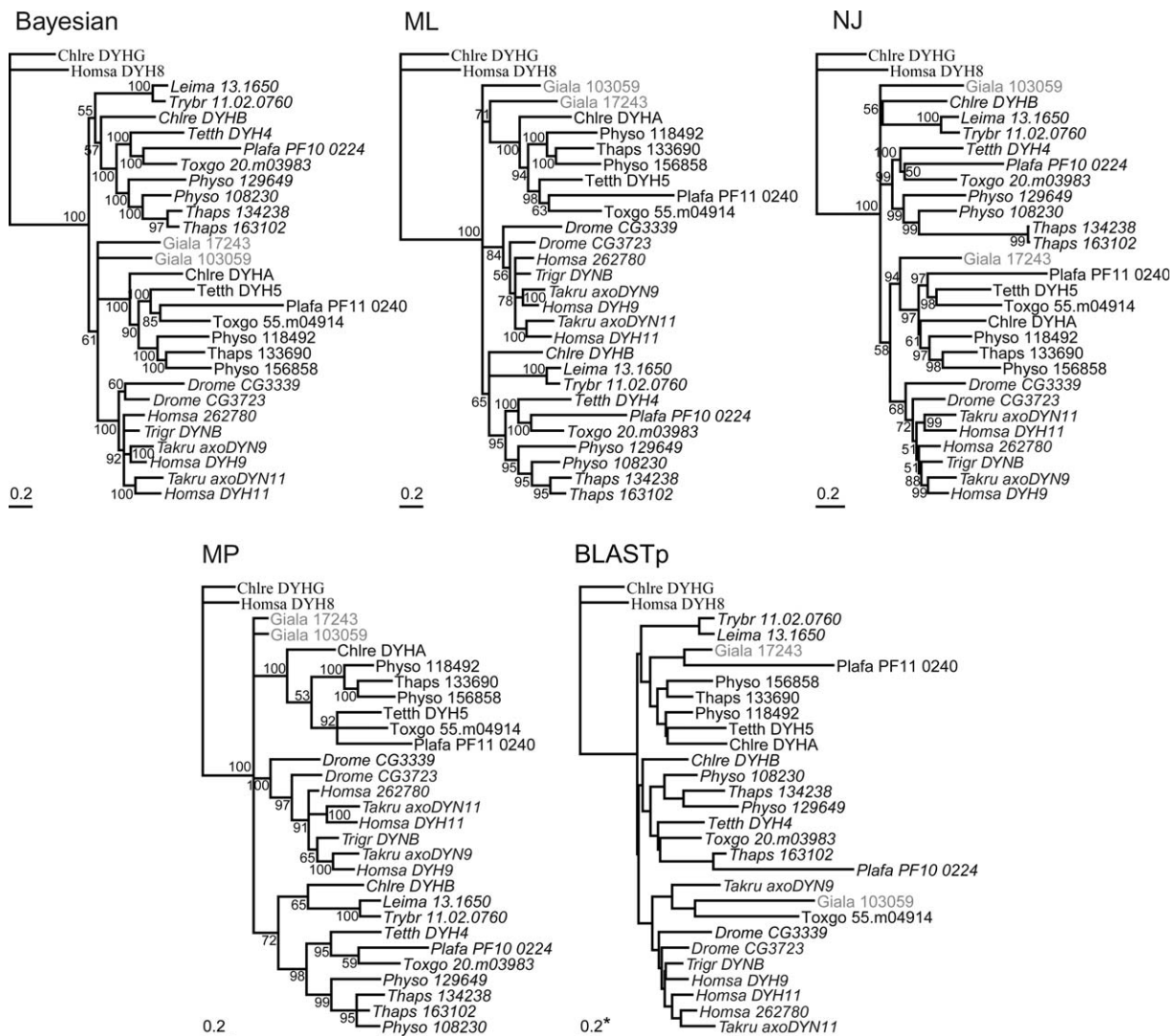


Figure 5: Relationships between OADβ family sequences (Figure 2) inferred by Bayesian, ML, NJ, MP or BLASTp means. Prefixes as in legend to Figure 2, with the addition of Trigr: *Tripneustes gratilla*. Trees are rooted using Chlre_DYHG and Homsa_DYH8 (OADα family members). Different fonts have been used to represent particular groups of sequences. Bars, substitutions per site; *except for BLASTp tree (see *Materials and Methods* for distance metric definition).

would be expected, in most organisms that build flagella/cilia, we find homologues of IAD-1α, IAD-1β, IC138 and IC140. Interestingly, in *P. falciparum*, we could only identify a member of the IAD-1β family (not IAD-1α), and we could not detect IC138 or IC140 orthologues. However, there are highly divergent DHCs predicted in the *P. falciparum* genome that were excluded from the present analysis; so, the result is ambiguous at best.

A most surprising result of our analysis was the finding of 2-headed IAD families in the alga *O. lucimarinus*. This organism is not believed to build an axoneme in any of its life cycle stages, and it does not possess cytoplasmic dynein 1 or 2 or any of the axonemal OAD or single-headed dynein families. However, *O. lucimarinus* encodes (appar-

ently canonical) orthologues of both IAD-1α and IAD-1β and also an IC138 homologue (a component of the same IAD I1 complex). We also found the same three proteins in the predicted proteome of the related alga *O. tauri* (data not shown), for which complete genome sequence is available (77). We can only speculate as to what function this dynein complex might be performing in these organisms. Presumably, at some point in its evolutionary history, the flagellate ancestor of *Osteococcus* must have attached a second function to the IAD-1α/IAD-1β dynein complex such that, when the flagellum was lost, the IAD genes remained under selective pressure.

The final group of axonemal dyneins is that of the 'single-headed' IADs. Of the three axonemal dynein groups, least

is known about the members of this group (it is not even clear if all the members are indeed single headed); yet, it is the largest. The recent structural work of Nicastro et al. (70) on *Chlamydomonas* flagella found five single-headed dyneins in the 96-nm repeat of the inner arms (one of which is closely associated with the '2-headed' IAD-1 α /1 β complex). Previous analyses have suggested six single-headed IADs (78,79), but the molecular identities of most of the DHCs involved are unknown. Moreover, the genome of *C. reinhardtii* is predicted to encode at least nine members of this group (Figure 1). Our analysis suggests that this group is monophyletic and consists of three DHC families – although, they are less well supported than the OAD or IAD-1 α /1 β families. Here, we have named these families IAD-3, IAD-4 and IAD-5 to follow on from the 2-headed IAD-1 α /1 β families and to reflect some overlap with the weaker-supported 'group 3', 'group 4' and 'group 5' clades, recently identified in an analysis of sea urchin motors (30). These names do not imply a relationship to the early I2 and I3 nomenclature for IADs (80), for which there is no available sequence data. Generally, there are fewer experimental data available on the functions of the proteins of the single-headed dyneins than that for OADs or 2-headed IADs. Knockout of either IAD-4 (Teth_{DYH9}) or IAD-3 (Teth_{DYH8} or Teth_{DYH12}) family members in *Tetrahymena* caused a reduction in cell motility (81), whereas in *Chlamydomonas*, DHC9 (IAD-3 family) enhances swimming under conditions of higher viscosity (82).

This classification of the single-headed IADs provides a framework for their study, but as yet, it provides little else by way of predictive biology. There are no ICs in our analysis that correlate with the possession of particular single-headed IAD families (or indeed IADs as a whole; Figure 3B). The dynein LC p28 is a component of the single-headed IADs and is apparently specific. However, although all organisms in our study that possess IAD HCs also possess p28, there is also an apparent homologue of the protein encoded by *C. elegans* – an organism that builds only immotile axonemes. Homologues of the recently identified subunit p38 are only found in organisms that possess the IAD-4 family, which encompasses the HC (Chlre_{DHC2}) with which p38 has been shown to interact in *Chlamydomonas* (83). This is suggestive of this protein being one of only two LC/LIC families that are specific to a particular dynein complex (the other being LC1 from the OAD complex as previously mentioned).

One organism in our analysis – the diatom *T. pseudonana* – lacks putative members of all five IAD families (and also IC138 and IC140). Instead, the genome of this organism encodes a canonical DYNC1H1, OAD α and the two non-metazoan subfamilies of OAD β (orthologues of *Tetrahymena* OAD β and OAD γ ; Figure 2). The *Thalassiosira* DHC repertoire can be viewed as an 'evolutionary experiment', demonstrating that motile eukaryotic flagella can be formed using OADs alone. In the light of the position,

diatoms are thought to occupy in the eukaryotic tree of life (Figure 3A); it is highly unlikely that the OADs of diatoms represent any kind of simplified ancestral state. Rather, it appears that *Thalassiosira* have secondarily pared down their DHC repertoire such that the axoneme – which is built only during gametogenesis – is constructed without the aid of DYNC2 and beats without the action of IADs. Interestingly, ultrastructural analysis of the gametes formed by the fern *Marsilea vestita* suggests that the reverse 'evolutionary experiment' has also been performed – detergent-extracted axonemes from *M. vestita* spermatozooids show no evidence of dynein outer arms (sequence data for the DHC repertoire encoded by *Marsilea* are not yet available). Similarly, mutants of *Chlamydomonas* that are unable to assemble dynein outer arms are still able to build motile axonemes (84,85). Thus, despite the high general conservation of the axonemal dyneins in organisms that build motile axonemes, particular lineages are apparently able to lose either all OADs or all IADs and still construct simplified, beating flagella.

Materials and Methods

DHC phylogeny

Sources and versions of the predicted protein datasets from the 24 genome sequencing projects used are given in File S1. From these datasets, we used HMMERv2.3.2 (<http://hmmer.wustl.edu/>) to extract all predicted proteins with a good match to the Pfam 'dynein heavy' (PF03028.5) domain (expectation value $<10^{-10}$). This created a seed dataset of 151 protein sequences, which were aligned and used to create an initial phylogeny. From this phylogeny, nine DHC families were defined and used to create family-specific alignments and new HMMs. These new HMMs were used to reinterrogate the 24 predicted protein sets, and those sequences with a score >1600 to any of the nine HMMs was extracted to give a set of 170 sequences.

It is not necessary to have full sequence information to infer relationships between sequences. However, highly truncated sequences cause problems with both alignment and phylogenetic reconstruction. For this reason, sequences <1000 aa in length were removed as fragmentary. Redundancy in the remaining sequences was reduced by excluding sequences that had $>95\%$ identity to sequences already in the analysis from the same organism. The remaining 158 protein sequences were aligned using MAFFT5.861 (86) adopting the E-INS-i strategy (87) and trimmed to well-aligned blocks (2596 characters), which cover the most conserved residues of the dynein motor domain. A list of included sequences with alternative database identifiers and descriptions is provided in File S2.

A Bayesian phylogeny was inferred from the protein alignment using Metropolis-coupled Markov chain Monte Carlo method as implemented by the program MrBAYES3.1.2 (88). The Whelan and Goldman (WAG) substitution matrix was used (89) with a gamma-distributed variation in substitution rate approximated to six discrete categories. Four Markov chains were run for 1 000 000 generations from a random starting tree sampling every 500 generations and with a 'temperature' of 0.2. Tree likelihoods appeared to reach stationary phase at around 250 000 generations, and the last 750 000 generations were used to construct the consensus tree shown in Figure 2. Support for the inferred phylogeny was produced by four methods: Bayesian partial bootstrap replicates and full bootstrap replicates under assumptions of ML, NJ and MP. For Bayesian partial replicate analysis, 10 new data matrices were created by sampling (with replacement) 1000 characters from the full alignment. These were used as above (800 000 generations), and a consensus built

from the last 400 000 generations of all 10. The ML trees (100 replicates, WAG matrix, six gamma categories, alpha parameter re-estimation for each replicate) were inferred from the 2596 character matrix using the program PHYLIP2.4.4 (90). The NJ trees [100 replicates, Jones, Taylor and Thornton (JTT) matrix] were inferred using software from the Phylip3.6 package (91): SeqBoot to generate resampled datasets, ProTDist for distance matrix generation and then Neighbor to infer the trees. The MP trees (100 replicates, heuristic search with tree bisection and reconnection) used the software PAUP4b10 (92). Consensus trees with bootstrap support for all four methods are presented in File S3. The division of phylogenies into families is subjective, but families were chosen that were (i) well supported; (ii) composed of sequence from more than one of the eukaryotic supergroups [cf. definition of kinesin superfamily (93)]; (iii) separated from the remaining sequences by a reasonably long branch; and (iv) informed by known biology (where it exists).

The OADB sub-phylogeny (Figure 5) was inferred from a fresh alignment of OADB sequences with the inclusion of sea urchin OADB (accession number P23098) and two OAD α sequences as an outgroup. The resulting trimmed alignment (3727 characters) was used in Bayesian (10 partial bootstraps of 2000 characters, 100 000 generations, burn-in 20 000), ML, NJ and MP analyses as above. For trees based on BLAST similarity searches, we derived a distance metric from BLASTp scores. The distance, d_{ij} , between sequences i and j was given by:

$$d_{ij} = -\ln\{(s_{ij} + s_{ji})/2\min(s_{ij}, s_{ji})\}$$

where s_{ij} is the BLASTp score, resulting when querying sequence i against sequence j , and $\min(s_{ij}, s_{ji})$ is a normalization factor for sequence length and composition given by the smaller of s_{ij} or s_{ji} . Note that the distance metric is symmetric ($d_{ij} = d_{ji}$) and positive ($d_{ij} = 0$ when $i = j$; $d_{ij} \geq 0$ when $i \neq j$). A matrix of distances was used to infer an NJ tree using the program Neighbor from the Phylip3.6 package (91). It is not possible to bootstrap this method.

IC and LC phylogenies

To identify all homologues of dynein ICs and Tctex1/2 (DYNLT1/2) family LCs, we used an iterative-HMM search approach. Briefly, identified homologues from *Chlamydomonas* were used to generate seed alignments. HMMER2.3.2 (<http://hmmer.wustl.edu/>) was then used to create HMMs from these alignments, which were used to interrogate a combined dataset of the 24 predicted proteomes from the chosen organisms (see File S1). Good matches to the HMM were incorporated into the alignment, and the process iterated until no further matches were identified. From the resultant datasets, redundancy was reduced by excluding sequences that had >95% identity to sequences already in the analysis from the same organism, and the remaining sequences were aligned using MAFFT5.861 (86) adopting the E-INS-i strategy (87) and trimmed to well-aligned blocks (401 and 107 characters for IC and Tctex1/2 alignments, respectively). Bayesian, ML, NJ and MP phylogenies were inferred as for DHC phylogenies (above), except that Markov chains were run for 500 000 generations (burn-in 100 000), and only four gamma categories were used. Either 10 or 50 full bootstraps of the Bayesian analyses were made for IC or Tctex1/2, respectively. Consensus trees with bootstrap support are provided in Files S4 and S5. Sequence identifiers and descriptions are provided in File S2.

Distribution of LCs, LICs and IFT proteins

In the first instance, potential orthologues of all proteins were identified by a RBB approach using identified *C. reinhardtii* and *H. sapiens* sequences against each of the predicted protein sets (e-value cutoff in both directions: 10^{-4}). Orthologues found in other organisms as RBB hits were also used to perform RBB analyses to find additional instances missed by the first analysis – any sequence *not* a RBB match to the original sequence but a RBB match to the majority of RBB matches in other organisms was considered to be a good orthologue candidate. The results of this analysis, including proteins on the OAD docking complex and radial spoke components, are presented in File S8. In the case of the gene families Tctex1/Tctex2

(DYNLT1/2; Figure 3), LC8/LC11 and LC16/LC14, additional information was gained from iterative-HMM searches and inference of phylogenies.

Acknowledgments

Predicted protein datasets were obtained from the sources specified in File S1. We thank each of these organizations and the respective genome sequencing projects for making sequence, gene model and annotation data publicly available. We also thank Ritsu Kamiya (University of Tokyo) for useful discussion on the *Chlamydomonas* cytoplasmic dynein candidate and the referees for insightful comments on the original manuscript. This work was funded by a Wellcome Trust Programme Grant, and K. G. is a Wellcome Trust Principal Research Fellow.

Supplementary Materials

File S1: Sources and versions of genomic data used in this work.

File S2: Unique identifiers and descriptions for the repertoires of DHC, IC and Tctex1/2 sequences encoded in the genomes of 24 diverse eukaryotes.

File S3: Majority rule consensus trees with bootstrap support for DHCs, as inferred using Bayesian (partial sequence resampling), ML, NJ and MP approaches (see *Materials and Methods*).

File S4: Majority rule consensus trees with bootstrap support for dynein IC, as inferred using Bayesian, ML, NJ and MP approaches (see *Materials and Methods*).

File S5: Majority rule consensus trees with bootstrap support for Tctex1/2 sequences, as inferred using Bayesian, ML, NJ and MP approaches (see *Materials and Methods*).

File S6: Classification of the divergent *Chlamydomonas* dynein 206178. Tree shown is the consensus of 10 Bayesian tree inference runs with character resampling. Topology support from Bayesian, ML, NJ and MP approaches is shown next to selected nodes.

File S7: Histogram showing the distribution of scores for all sequences encoded 24 diverse eukaryotes, queried with a HMM built from an alignment of dynein IC sequences.

File S8: Reciprocal-best-BLAST hits from 24 diverse eukaryotes, queried with sequences identified in *Chlamydomonas* as OAD docking complex components, dynein ICs, proteins involved in IFT, dynein LCs, dynein LICs and radial spoke proteins and also components of the human cytoplasmic dynein 1 complex.

Supplemental materials are available as part of the online article at <http://www.blackwell-synergy.com>

References

1. Hirokawa N, Noda Y, Okada Y. Kinesin and dynein superfamily proteins in organelle transport and cell division. *Curr Opin Cell Biol* 1998;10: 60–73.
2. Gibbons IR. The role of dynein in microtubule-based motility. *Cell Struct Funct* 1996;21:331–342.
3. Holzbaur ELF, Vallee RB. Dyneins: molecular structure and cellular function. *Annu Rev Cell Biol* 1994;10:339–372.

4. Kon T, Nishiura M, Ohkura R, Toyoshima YY, Sutoh K. Distinct functions of nucleotide-binding/hydrolysis sites in the four AAA modules of cytoplasmic dynein. *Biochemistry* 2004;43:11266–11274.
5. Gibbons IR, LeeEiford A, Mocz G, Phillipson CA, Tang WJY, Gibbons BH. Photosensitized cleavage of dynein heavy chains: cleavage at the V1 site by irradiation at 365nm in the presence of ATP and vanadate. *J Biol Chem* 1987;262:2780–2786.
6. Oiwa K, Sakakibara H. Recent progress in dynein structure and mechanism. *Curr Opin Cell Biol* 2005;17:98–103.
7. Koonce MP, Sams M. Of rings and levers: the dynein motor comes of age. *Trends Cell Biol* 2004;14:612–619.
8. King SM. Dynein motors: structure, mechanochemistry and regulation. In: Schliwa M, editor. *Molecular Motors*. Weinheim: Wiley-VCH Verlag; 2002, pp. 45–78.
9. Simpson AGB, Roger AJ. The real 'kingdoms' of eukaryotes. *Curr Biol* 2004;14:R693–R696.
10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
11. Aparicio S, Chapman J, Stupka E, Putnam N, Chia J, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 2002;297:1301–1310.
12. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000;287:2185–2195.
13. The C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998;282:2012–2018.
14. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H et al. Life with 6000 genes. *Science* 1996;274:546–567.
15. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* 2002;415:871–880.
16. Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 2005;433:865–868.
17. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F et al. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 2005;435:43–57.
18. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohne W, Hannick L, Aslett MA, Shallom J, Marcello L et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;309:416–422.
19. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C et al. The genome of the kinetoplastid parasite *Leishmania major*. *Science* 2005;309:436–442.
20. Eisen JA, Coyne RS, Wu M, Wu DY, Thiagarajan M, Wortman JR, Badger JH, Ren QH, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ et al. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 2006;4:1620–1642.
21. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 2004;306:79–86.
22. Tyler BM, Tripathy S, Zhang XM, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CMB, Dorrance AE, Dou DL, Dickerman AW et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 2006;313:1261–1266.
23. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;419:498–511.
24. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng MQ, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA et al. Complete genome sequence of the apicomplexan *Cryptosporidium parvum*. *Science* 2004;304:441–445.
25. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 2004;428:653–657.
26. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796–815.
27. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP et al. The genome of black cottonwood, *Populus trichocarpa*. *Science* 2006;313:1596–1604.
28. Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). *Science* 2002;296:79–92.
29. Asai DJ, Wilkes DE. The dynein heavy chain family. *J Euk Microbiol* 2004;51:23–29.
30. Morris RL, Hoffman MP, Obar RA, McCafferty SS, Gibbons IR, Leone AD, Cool J, Allgood EL, Musante AM, Judkins KM, Rossetti BJ, Rawson AP, Burgess DR. Analysis of cytoskeletal and motility proteins in the sea urchin genome assembly. *Dev Biol* 2006;300:219–237.
31. Ogawa K, Kamiya R, Wilkerson CG, Witman GB. Interspecies conservation of outer arm dynein intermediate chain sequences defines two intermediate chain subclasses. *Mol Biol Cell* 1995;6:685–696.
32. Wilkerson CG, King SM, Koutoulis A, Pazour GJ, Witman GB. The 78,000 Mr intermediate chain of *Chlamydomonas* outer arm dynein is a WD-repeat protein required for arm assembly. *J Cell Biol* 1995;129:169–178.
33. Takada S, Kamiya R. Reconstitution of *Chlamydomonas* outer dynein arms from alpha-beta-subunit and gamma-subunit: requirement of a 3rd factor. *J Cell Biol* 1994;126:737–745.
34. Mitchell DR, Kang Y. Reversion analysis of dynein intermediate chain function. *J Cell Sci* 1993;105:1069–1078.
35. Mitchell DR, Kang Y. Identification of *oda6* as a *Chlamydomonas* dynein mutant by rescue with the wild-type gene. *J Cell Biol* 1991;113:835–842.
36. Hendrickson TW, Perrone CA, Griffin P, Wuichet K, Mueller J, Yang PF, Porter ME, Sale WS. IC138 is a WD-repeat dynein intermediate chain required for light chain assembly and regulation of flagellar bending. *Mol Biol Cell* 2004;15:5431–5442.
37. Perrone CA, Yang PF, O'Toole E, Sale WS, Porter ME. The *Chlamydomonas* *IDA7* locus encodes a 140-kDa dynein intermediate chain required to assemble the I1 inner arm complex. *Mol Biol Cell* 1998;9:3351–3365.

38. Yang P, Sale WS. Cloning and characterization of the 140 kD intermediate chain gene of inner dynein arm I1 in *Chlamydomonas*. *Mol Biol Cell* 1996;7:3305–3305.
39. DiBella LM, Gorbatyuk O, Sakato M, Wakabayashi K, Patel-King RS, Pazour GJ, Witman GB, King SM. Differential light chain assembly influences outer arm dynein motor function. *Mol Biol Cell* 2005;16:5661–5674.
40. Lawrence CJ, Morris NR, Meagher RB, Dawe RK. Dyneins have run their course in plant lineage. *Traffic* 2001;2:362–363.
41. King SM. Dyneins motor on in plants. *Traffic* 2002;3:930–931.
42. Porter ME, Bower R, Knott JA, Byrd P, Dentler W. Cytoplasmic dynein heavy chain 1b is required for flagellar assembly in *Chlamydomonas*. *Mol Biol Cell* 1999;10:693–712.
43. Scali M, Vignani R, Moscatelli A, Jellbauer S, Cresti M. Molecular evidence for a cytoplasmic dynein heavy chain from *Nicotiana tabacum* L. *Cell Biol Int* 2003;27:261–262.
44. Wilson MJ, Salata MW, Susalka SJ, Pfister KK. Light chains of mammalian cytoplasmic dynein: identification and characterization of a family of LC8 light chains. *Cell Motil Cytoskeleton* 2001;49:229–240.
45. King SM, Barbarese E, Dillman JF, Patel-King RS, Carson JH, Pfister KK. Brain cytoplasmic and flagellar outer arm dyneins share a highly conserved M(r) 8,000 light chain. *J Biol Chem* 1996;271:19358–19366.
46. Harrison A, Olds-Clarke P, King SM. Identification of the t complex-encoded cytoplasmic dynein light chain Tctex1 in inner arm I1 supports the involvement of flagellar dyneins in meiotic drive. *J Cell Biol* 1998;140:1137–1147.
47. King SM, Patel-King RS. The M(r) = 8,000 and 11,000 outer arm dynein light-chains from *Chlamydomonas* flagella have cytoplasmic homologs. *J Biol Chem* 1995;270:11445–11452.
48. Yang PF, Diener DR, Rosenbaum JL, Sale WS. Localization of calmodulin and dynein light chain LC8 in flagellar radial spokes. *J Cell Biol* 2001;153:1315–1325.
49. Krogan NJ, Cagney G, Yu HY, Zhong GQ, Guo XH, Ignatchenko A, Li J, Pu SY, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006;440:637–643.
50. Kahana JA, Schlenstedt G, Evanchuk DM, Geiser JR, Hoyt MA, Silver PA. The yeast dynactin complex is involved in partitioning the mitotic spindle between mother and daughter cells during anaphase B. *Mol Biol Cell* 1998;9:1741–1756.
51. Geiser JR, Schott EJ, Kingsbury TJ, Cole NB, Totis LJ, Bhattacharyya G, He L, Hoyt MA. *Saccharomyces cerevisiae* genes required in the absence of the CIN8-encoded spindle motor act in functionally diverse mitotic pathways. *Mol Biol Cell* 1997;8:1035–1050.
52. Pfister KK, Fisher EMC, Gibbons IR, Hays TS, Holzbaur ELF, McIntosh JR, Porter ME, Schroer TA, Vaughan KT, Witman GB, King SM, Vallee RB. Cytoplasmic dynein nomenclature. *J Cell Biol* 2005;171:411–413.
53. Cole DG. The intraflagellar transport machinery of *Chlamydomonas reinhardtii*. *Traffic* 2003;4:435–442.
54. Hou YQ, Pazour GJ, Witman GB. A dynein light intermediate chain, D1bLIC, is required for retrograde intraflagellar transport. *Mol Biol Cell* 2004;15:4382–4394.
55. Sinden RE. Cell biology. In: Killick-Kendrick R, Peters W, editors. *Rodent Malaria*. London: Academic Press; 1978, pp. 85–168.
56. Briggs LJ, Davidge JA, Wickstead B, Ginger ML, Gull K. More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr Biol* 2004;14:R611–R612.
57. Wickstead B, Gull K. A "holistic" kinesin phylogeny reveals new kinesin families and predicts protein functions. *Mol Biol Cell* 2006;17:1734–1743.
58. Gibbons IR, Gibbons BH, Moczek G, Asai DJ. Multiple nucleotide-binding sites in the sequence of dynein beta heavy chain. *Nature* 1991;352:640–643.
59. Ogawa K. Four ATP-binding sites in the midregion of the beta heavy chain of dynein. *Nature* 1991;352:643–645.
60. Goodenough U, Heuser J. Structural comparison of purified dynein proteins with *in situ* dynein arms. *J Mol Biol* 1984;180:1083–1118.
61. Hamasaki T, Barkalow K, Richmond J, Satir P. cAMP-stimulated phosphorylation of an axonemal polypeptide that copurifies with the 22S dynein arm regulates microtubule translocation velocity and swimming speed in *Paramecium*. *Proc Natl Acad Sci U S A* 1991;88:7918–7922.
62. Johnson KA, Wall JS. Structure and molecular weight of dynein ATPase. *J Cell Biol* 1983;96:669–678.
63. Nicastro D, McIntosh JR, Baumeister W. 3D structure of eukaryotic flagella in a quiescent state revealed by cryo-electron tomography. *Proc Natl Acad Sci U S A* 2005;102:15889–15894.
64. Wada S, Okuno M, Nakamura KI, Mohri H. Dynein of sperm flagella of oyster belonging to Protostomia also has a 2-headed structure. *Biol Cell* 1992;76:311–317.
65. Mohri H, Kubo-Irie M, Irie M. Outer arm dynein of sperm flagella and cilia in the animal kingdom. *Mem Mus Natn Hist Nat* 1995;1666:15–22.
66. Mohri H, Inaba K, Kubo-Irie M, Takai H, Yano-Toyoshima Y. Characterization of outer arm dynein in sea anemone, *Anthopleura midori*. *Cell Motil Cytoskeleton* 1999;44:202–208.
67. Sale WS, Goodenough U, Heuser J. The substructure of isolated and *in situ* outer dynein arms of sea urchin sperm flagella. *J Cell Biol* 1985;101:1400–1412.
68. Ragana SPM, Gagnon C, White D, Isles MB, Johnson KA. Structure and mass analysis of 12S and 19S dynein obtained from bull sperm flagella. *Cell Motil Cytoskeleton* 1987;8:368–378.
69. Gatti JL, King SM, Moss AG, Witman GB. Outer arm dynein from trout spermatozoa: purification, polypeptide composition and enzymatic properties. *J Biol Chem* 1989;264:11450–11457.
70. Nicastro D, Schwartz C, Pierson J, Gaudette R, Porter ME, McIntosh JR. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* 2006;313:944–948.
71. Pazour GJ, Wilkerson CG, Witman GB. A dynein light chain is essential for the retrograde particle movement of intraflagellar transport (IFT). *J Cell Biol* 1998;141:979–992.
72. Pazour GJ, Koutoulis A, Benashski SE, Dickert BL, Sheng H, Patel-King RS, King SM, Witman GB. LC2, the *Chlamydomonas* homologue of the t complex-encoded protein Tctex2, is essential for outer dynein arm assembly. *Mol Biol Cell* 1999;10:3507–3520.
73. Bowman AB, Patel-King RS, Benashski SE, McCafferty JM, Goldstein LS, King SM. Drosophila roadblock and *Chlamydomonas* LC7, a conserved family of dynein-associated proteins involved in axonal transport, flagellar motility, and mitosis. *J Cell Biol* 1999;146:165–180.
74. Patel-King RS, Benashski SE, Harrison A, King SM. A *Chlamydomonas* homologue of the putative murine t complex distorter Tctex-2 is an outer arm dynein light chain. *J Cell Biol* 1997;137:1081–1090.
75. Myster SH, Knott JA, Otoole E, Porter ME. The *Chlamydomonas* Dhc1 gene encodes a dynein heavy chain subunit required for assembly of the I1 inner arm complex. *Mol Biol Cell* 1997;8:607–620.
76. Smith EF, Sale WS. Structural and functional reconstitution of outer and inner dynein arms in *Chlamydomonas* flagellar axonemes. *J Cell Biol* 1992;117:573–581.
77. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroove S, Echevnyie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O et al. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci U S A* 2006;103:11647–11652.
78. Kagami O, Kamiya R. Translocation and rotation of microtubules caused by multiple species of *Chlamydomonas* inner-arm dynein. *J Cell Sci* 1992;103:653–664.

79. Goodenough UW, Gebhart B, Mermall V, Mitchell DR, Heuser JE. High-pressure liquid-chromatography fractionation of *Chlamydomonas* dynein extracts and characterization of inner-arm dynein subunits. *J Mol Biol* 1987;194:481–494.
80. Piperno G, Ramanis Z, Smith EF, Sale WS. Three distinct inner dynein arms in *Chlamydomonas* flagella: molecular composition and location in the axoneme. *J Cell Biol* 1990;110:379–389.
81. Liu SM, Hard R, Rankin S, Hennessey T, Pennock DG. Disruption of genes encoding predicted inner arm dynein heavy chains causes motility phenotypes in *Tetrahymena*. *Cell Motil Cytoskeleton* 2004;59:201–214.
82. Yagi T, Minoura I, Fujiwara A, Saito R, Yasunaga T, Hirono M, Kamiya R. An axonemal dynein particularly important for flagellar movement at high viscosity: implications from a new *Chlamydomonas* mutant deficient in the dynein heavy chain gene DHC. *J Biol Chem* 2005;280:41412–41420.
83. Yamamoto R, Yanagisawa H, Yagi T, Kamiya R. A novel subunit of axonemal dynein conserved among lower and higher eukaryotes. *FEBS Lett* 2006;580:6357–6360.
84. Brokaw CJ, Kamiya R. Bending patterns of *Chlamydomonas* flagella: IV. Mutants with defects in inner and outer dynein arms indicate differences in dynein arm function. *Cell Motil Cytoskeleton* 1987;8:68–75.
85. Kamiya R, Okamoto M. A mutant of *Chlamydomonas reinhardtii* that lacks the flagellar outer dynein arm but can swim. *J Cell Sci* 1985;74:181–191.
86. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
87. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 2005;33:511–518.
88. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–1574.
89. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. *Mol Biol Evol* 2001;18:691–699.
90. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704.
91. Felsenstein J. PHYLIP – Phylogeny Inference Package (version 3.2). *Cladistics* 1989;5:164–166.
92. Swofford DL. *Phylogenetic Analysis Using Parsimony (*And Other Methods)*. Sunderland, MA, USA: Sinauer; 1998.
93. Lawrence CJ, Dawe RK, Christie KR, Cleveland DW, Dawson SC, Endow SA, Goldstein LSB, Goodson HV, Hirokawa N, Howard J, Malmberg RL, McIntosh JR, Miki H, Mitchison TJ, Okada Y et al. A standardized kinesin nomenclature. *J Cell Biol* 2004;167:19–22.
94. Soltis PS, Soltis DE. Applying the bootstrap in phylogeny reconstruction. *Stat Sci* 2003;18:256–267.
95. Patel-King RS, Benashski SE, Harrison A, King SM. Two functional thioredoxins containing redox-sensitive vicinal dithiols from the *Chlamydomonas* outer dynein arm. *Syst Biol* 1996;271:6283–6291.