

Research article

Open Access

Reconstructing the phylogeny of 21 completely sequenced arthropod species based on their motor proteins

Florian Odronitz, Sebastian Becker and Martin Kollmar*

Address: Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Goettingen, Germany

Email: Florian Odronitz - flod@nmr.mpibpc.mpg.de; Sebastian Becker - sebe@nmr.mpibpc.mpg.de; Martin Kollmar* - mako@nmr.mpibpc.mpg.de

* Corresponding author

Published: 21 April 2009

Received: 6 March 2008

BMC Genomics 2009, 10:173 doi:10.1186/1471-2164-10-173

Accepted: 21 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/173>

© 2009 Odronitz et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Motor proteins have extensively been studied in the past and consist of large superfamilies. They are involved in diverse processes like cell division, cellular transport, neuronal transport processes, or muscle contraction, to name a few. Vertebrates contain up to 60 myosins and about the same number of kinesins that are spread over more than a dozen distinct classes.

Results: Here, we present the comparative genomic analysis of the motor protein repertoire of 21 completely sequenced arthropod species using the owl limpet *Lottia gigantea* as outgroup. Arthropods contain up to 17 myosins grouped into 13 classes. The myosins are in almost all cases clear paralogs, and thus the evolution of the arthropod myosin inventory is mainly determined by gene losses. Arthropod species contain up to 29 kinesins spread over 13 classes. In contrast to the myosins, the evolution of the arthropod kinesin inventory is not only determined by gene losses but also by many subtaxon-specific and species-specific gene duplications. All arthropods contain each of the subunits of the cytoplasmic dynein/dynactin complex. Except for the dynein light chains and the p150 dynactin subunit they contain single gene copies of the other subunits. Especially the roadblock light chain repertoire is very species-specific.

Conclusion: All 21 completely sequenced arthropods, including the twelve sequenced *Drosophila* species, contain a species-specific set of motor proteins. The phylogenetic analysis of all genes as well as the protein repertoire placed *Daphnia pulex* closest to the root of the Arthropoda. The louse *Pediculus humanus corporis* is the closest relative to *Daphnia* followed by the group of the honeybee *Apis mellifera* and the jewel wasp *Nasonia vitripennis*. After this group the rust-red flour beetle *Tribolium castaneum* and the silkworm *Bombyx mori* diverged very closely from the lineage leading to the *Drosophila* species.

Background

Nearly each single cell in eukaryotes hosts particular proteins, which are responsible for intracellular transport. These molecular motor molecules are highly conserved among the different species of eukaryotes and evolved

slowly over time [1,2]. This property grants them the role of an appropriate candidate to carry out evolutionary studies. The three superfamilies of transporting motor proteins are the myosins, kinesins, and dyneins. Attached to the cytoskeletal networks (microtubules and actin) they



transport all kinds of organelles and vesicles [3], and organize and remodel the cytoskeleton and developmental processes in eukaryotes [4]. The energy for their unidirectional cargo transport on one of the filamentous cytoskeletal tracks is derived from ATP hydrolysis [5]. Out of the three superfamilies only the members of the kinesin superfamily are found in all eukaryotes, whereas members of the dynein [6] and myosin [7] superfamilies are lacking in particular eukaryotic lineages.

The members of the actin-based myosin family have their origin early in eukaryotic evolution. Based on the latest analysis, the myosins are grouped into 35 classes [7]. Myosins consist of three regions, the motor (or head) domain, a neck domain, and the tail, which comprises all C-terminal domains as well as domains N-terminal to the motor domain. The motor domain is highly conserved and contains both the ATP and actin binding site, where the force generation resides. This energy-transducing motor domain is coupled to a regulatory neck region (helical region), which is able to bind calmodulin or calmodulin-like light chains. Linked to the neck region most myosins have tail domains. Contrary to the head domains the tail domains show high variability in sequence and length, reflecting their functional diversity. The functions range from cytokinesis, organellar transport, cell polarization to signal transduction [8-10]. Some of the myosin classes also contain large domains at the N-terminus of the motor domains [7].

The second molecular motor protein family is kinesin (members also known as KRPs, KLPs, or KIFs) [11]. The members of this superfamily are microtubule-based and facilitate movement in both directions (either plus or minus end-directed) [12]. For their movement along the microtubules they utilize ATP similarly to the other motor proteins. The classical kinesin forms a tetramer with two kinesin heavy chains (KHCs) and two kinesin light chains (KLCs). Like in myosins the head domain is well conserved and responsible for the movement, whereas the stalk and tail domains play fundamental roles in the interaction with other subunits of the holoenzyme or with cargo molecules such as proteins, lipids or nucleic acids [13]. The region between the head and the stalk is family-specific and determines the direction of movement [14]. Kinesins bind a variety of cargoes and perform tasks such as vesicle and organelle transport, spindle formation and elongation, chromosome segregation, and microtubule organization [15,16].

The members of the dynein superfamily are minus end-directed motor proteins [17]. Thus they are responsible for the retrograde transport of cargoes along microtubules. They are involved in many processes like spindle formation, chromosome segregation, and the transport of

a variety of cargoes like viruses, RNAs, signaling molecules, and organelles [18]. Dyneins are multi-subunit protein complexes with two or three heavy chains (DHCs), light chains, light intermediate, and intermediate chains [19]. Supported by an activator protein called dynactin, which consists of 11 subunits, dynein is able to move and bind to membranes or further cargoes [20-22].

The genome of *Drosophila melanogaster* was the third eukaryotic genome to be completely sequenced [23]. Since then, the number of sequenced organisms has increased rapidly. Of the Arthropoda phylum, the genomes of the mosquitos *Anopheles gambiae* [24] and *Aedes aegyptii* [25], the silkworm *Bombyx mori* [26,27], the beetle *Tribolium castaneum* [28], the waterflea *Daphnia pulex* (this special series in BMC journals), and eleven of the *Drosophila* species group [29,30] have been published. The draft genome sequences of *Culex pipiens quinquefasciatus*, *Nasonia vitripennis*, and *Pediculus humanus corporis* have been finished recently. The phylogenetic relationship of the twelve sequenced *Drosophila* species has been described in detail [29].

Here, we present the analysis of the phylogenetic relationship of 21 completely sequenced arthropods based on the sequences and inventory of their motor proteins.

Results

Identification and annotation of the motor proteins

The arthropod motor protein genes were identified by TBLASTN searches against the corresponding genome data of the different species. Species, that missed certain orthologs in the first instance, were searched again with supposed-to-be orthologs of the other species. In this iterative process all motor proteins have been identified or their absence in certain species have been confirmed. The species analyzed were the mosquitos *Aedes aegyptii* (*Aea*), *Culex pipiens quinquefasciatus* (*Cpq*), and *Anopheles gambiae* (*Ang*), the silkworm *Bombyx mori* (*Bm_b*), the honeybee *Apis mellifera* (*Am*), the jewel wasp *Nasonia vitripennis* (*Nav*), the waterflea *Daphnia pulex* (*Dap*), the rust-red flour beetle *Tribolium castaneum* (*Tic*), the body louse *Pediculus humanus corporis* (*Pdc*), twelve *Drosophila* species (*Drosophila ananassae* (*Da*), *Drosophila erecta* (*Der*), *Drosophila grimshawi* (*Dg*), *Drosophila melanogaster* (*Dm*), *Drosophila mojavensis* (*Dmo*), *Drosophila persimilis* (*Dp*), *Drosophila pseudoobscura* (*Drp*), *Drosophila sechellia* (*Dse*), *Drosophila simulans* (*Dss_a*), *Drosophila virilis* (*Dv*), *Drosophila willistoni* (*Dw*) and *Drosophila yakuba* (*Dy*)), and the mollusc *Lottia gigantea* (*Lg*), which we used as outgroup. The sequences were assigned by manual inspection of the genomic DNA sequences. Exons have been confirmed by the identification of flanking consensus intron-exon splice junction donor and acceptor sequences [31]. The genomic sequences of *Drosophila virilis*, *Apis mellifera*, and

especially *Bombyx mori* contain several gaps. Many of the gaps have been filled by analyzing EST data.

Analysis of the arthropod myosins

All myosins have been classified based on the phylogenetic analysis of their motor domains together with the motor domains of the already grouped myosins [7] (Figure 1). All myosins belong to previously defined classes except one myosin from *Nasonia* that has a very similar domain organization to the class-V myosins but a considerably different motor domain. Except for class-XXI, all myosin classes are shared between arthropods and mam-

mals suggesting that their common ancestor already contained these classes [7]. *Daphnia*, which roots the insect phylogeny, possesses the largest repertoire of myosins. Although the taxon sampling is very limited in this study, it is likely that the evolution of the arthropods was accompanied by taxon- and species-specific losses of certain myosin classes. *Daphnia* still contains a class-XIX myosin, that all other analyzed arthropods have lost, and four class-I myosins. Class-XIX myosins have also been found in Deuterostomia and Cnidaria. Also, all other arthropods have lost at least one of the class-I myosins. However, the remaining variants differ between the analyzed species,

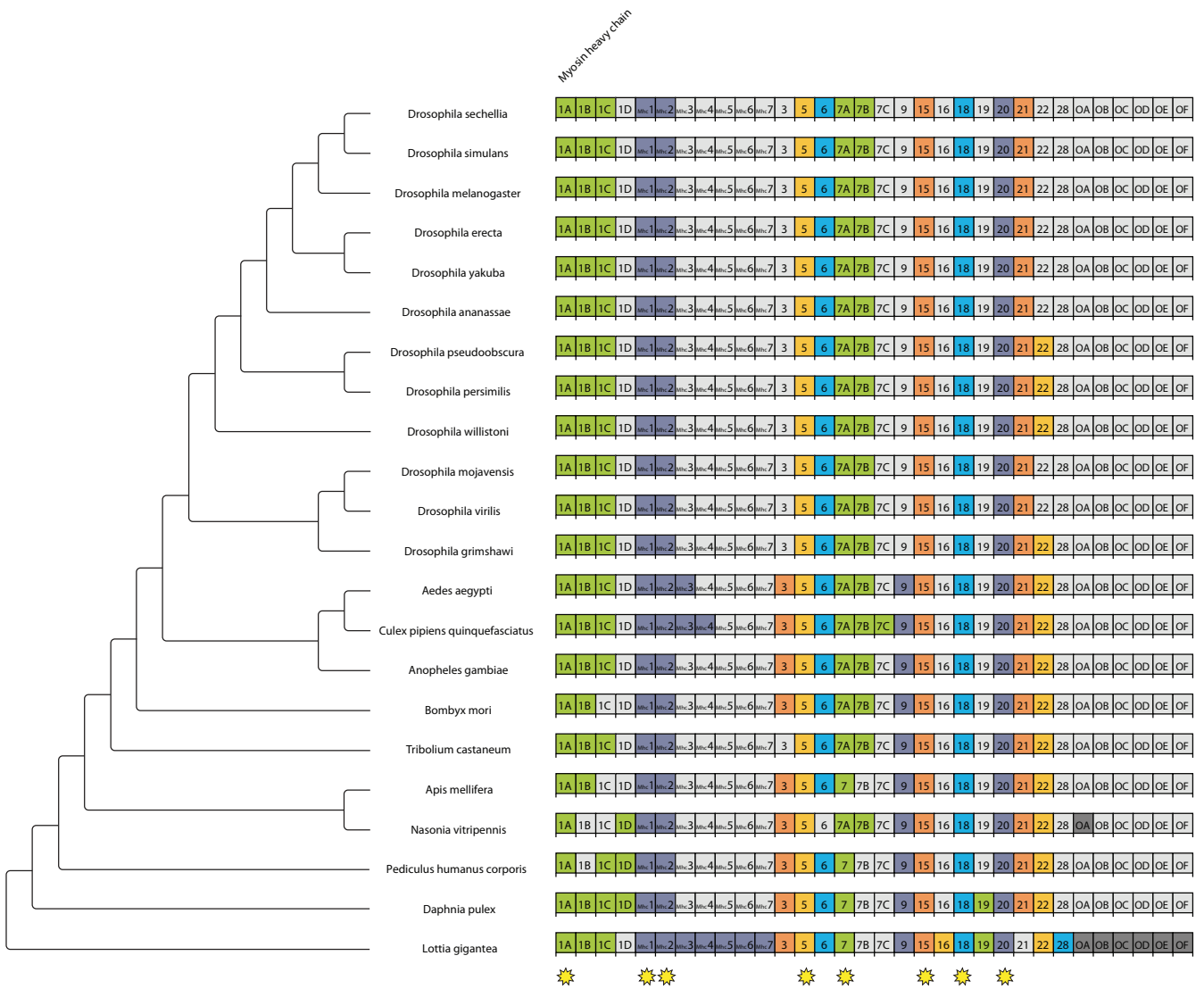


Figure 1 Myosin repertoire of the arthropods. This chart shows the myosin repertoire for all species in the analysis. To the left is a schematic phylogenetic tree, depicting the relationships (no scale). The identifiers in the boxes indicate protein classes/variants. "O" means orphan class. Colored boxes mean the class/variant exists in this species. Grey boxes mean the class/variant was not found. Columns marked with stars were included in the phylogenomic analysis.

which means that they lost the class-I variants after separating from the next closest species. For example, *Apis* and *Nasonia* both have lost the class-I myosin variant C, that their closest relative *Pediculus* still has. *Pediculus*, however, specifically lost the variants B and D, respectively. All arthropods contain a non-muscle as well as a muscle myosin heavy chain gene (class-II myosins). The alternatively spliced muscle myosin heavy chain genes have been described elsewhere [32]. The *Drosophila* species and *Tribolium* have lost the class-3 myosin. The *Drosophila melanogaster* NinaC protein has previously been classified as a class-III myosin. Based on the analysis of more than 2000 myosins the NinaC protein does not group to the vertebrate class-III myosins and all arthropod homologs of NinaC have been grouped into a new class, class-XXI [7]. Surprisingly, *Nasonia* does not contain a class-VI myosin, that all other Metazoa contain, that have been analyzed so far. The lack of the class-VI myosin might be a specific characteristic of *Nasonia vitripennis*, or due to sequencing and assembly problems, which are, however, unlikely given the high coverage of the *Nasonia* genome sequence. Finishing of the sequencing of the other two *Nasonia* genomes, which is in progress at the Baylor College of Medicine, will either confirm the lineage specific loss of the class-VI myosin or reveal sequencing problems of the *Nasonia vitripennis* genome. *Daphnia*, *Pediculus*, and *Apis* have lost the variant B of the class-VII myosin. The class-VII myosin, which they contain, is a clear homolog of the class-VII variant A myosins of the other arthropods. Another scenario would be, that the ancestor of *Apis*, *Nasonia*, and the clade containing the mosquito and *Drosophila* species has gained the class-VII myosin variant B via gene duplication of the variant A myosin. In this case, *Apis* specifically lost its class-VII myosin variant B. The *Drosophila* lineage has also completely lost the class-IX myosin. All arthropod genomes contain a class-XV, a class-XVIII, a class-XX, and a class-XXI myosin. The class-XXII myosin has independently been lost by several sub-lineages of the *Drosophila* species. The *Drosophila* species, that have been marked as having lost their class-XXII myosin, all still contain some of the exons of the ancient class-XXII myosin but spread over several hundred thousands of base pairs so that it is highly improbable that these pieces might belong to still functional genes.

The domain organizations of the arthropod myosins are identical to those found for other members of the respective classes [7]. Figure 2 shows diagrams of the *Daphnia* myosins that have the largest diversity of the arthropod myosins. The class-XXI and the class-III myosins have an identical domain organization, although the phylogenetic analysis of their motor domains reveals two distinct classes. It is highly probable, that the class-XXI myosins are the result of an arthropod specific gene duplication of the ancient class-III myosin followed by the divergence of

the new duplicate. The class-XXII myosins and the class-VII myosins have similar domain organizations. In contrast to the class-VII myosins, the class-XXII myosins lack the N-terminal SH3-like domain, they contain three instead of five IQ-motifs for the binding of calmodulin-like light chains, they have a longer coiled-coil regions containing domain till the first MyTH4 domain, and they lack the SH3 domain of the C-terminal tail.

Analysis of the arthropod kinesins

For their classification, the kinesin motor domains have been used in a phylogenetic analysis together with the motor domains of the human kinesins [11,33]. The sequences have been named according to the standardized kinesin nomenclature [34] leaving some kinesins unclassified (Figure 3). Orphan kinesins, that are clear homologs, got the same variant designation to allow for better comparison. In general, all analyzed species contain species-specific sets of kinesins. Except for *Drosophila pseudoobscura* and *Drosophila persimilis*, which have identical sets of kinesins, even closely related species like the twelve analyzed *Drosophila* species have different kinesin inventories. Thus, it is likely that the evolution of the kinesin inventories of the analyzed arthropods is strongly determined by species-specific gene duplications and gene losses. Given the limited taxon and species sampling it is impossible to identify lineage-specific duplication and loss events. Some gene duplications and gene losses are especially interesting. In this respect, we will not consider the kinesin inventory of *Bombyx mori* because the genome has not been sequenced with high coverage and is highly fragmented. The *Drosophila ananassae* genome does not contain a kinesin-2C that all other arthropods have. *Drosophila willistoni* does not contain a kinesin-4A, but two class-VI kinesins and two species-specific kinesins that have not been classified yet, kinesin-D and kinesin-E. While most arthropods contain only one kinesin-5, *Tribolium* contains a set of four class-V kinesins. The *Pediculus* genome does not encode a class-VII kinesin, but encodes a kinesin-9 that is otherwise only found in *Apis*. None of the analyzed arthropods contains a kinesin-10. *Nasonia* does not contain a kinesin-12 that all other arthropods have. The set of class-XIII kinesins in the arthropods ranges from one to four homologs. *Tribolium*, *Apis*, *Nasonia*, *Pediculus*, and *Daphnia* contain one or two additional kinesins that could not be grouped to any of the known classes.

The *Daphnia* kinesins mainly consist of the kinesin motor domain and long coiled-coil regions in the tail (Figure 4). Only the class-III kinesins contain further domains that have been characterised and named. A characteristic of almost all class-III kinesins is an FHA domain following C-terminal to the motor domain. The class-III variant A

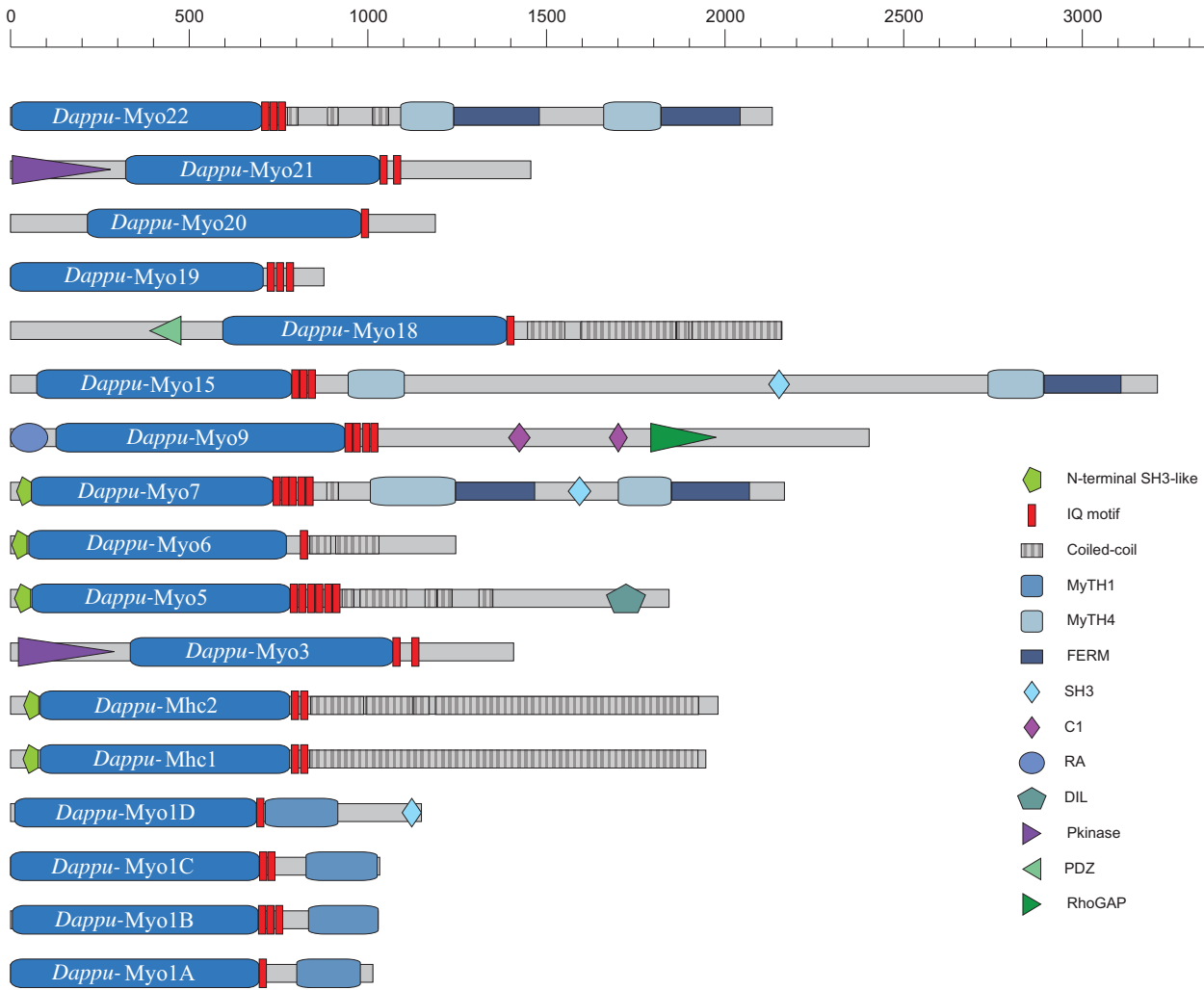


Figure 2
Domain organisation of the *Daphnia pulex* myosins. The sequence name is given in the motor domain of the respective myosin. A colour key to the domain names and symbols is given on the right except for the myosin domain that is coloured in blue. The abbreviations for the domains are: C1, Protein kinase C conserved region 1; DIL, *dilute*; FERM, band 4.1, ezrin, radixin, and moesin; IQ motif, isoleucine-glutamine motif; MyTH1, myosin tail homology 1; MyTH4, myosin tail homology 4; PDZ, PDZ domain; Pkinase, Protein kinase domain; RA, Ras association (RalGDS/AF-6) domain; RhoGAP, Rho GTPase-activating protein; SH3, src homology 3.

kinesins also contain a CAP-Gly domain at the C-terminus, while the variant B kinesins contain a PH domain.

The dynein/dynactin motor protein complex of the arthropods

All arthropods contain members of all the cytoplasmic dynein subunits (Figure 5). The dynein heavy chain proteins belong to the longest proteins in eukaryotes having lengths of 3,500 to 5,000 amino acids. The genes of the dynein heavy chains have not been analysed and classified yet because their large size in combination with the high degree of fragmentation of many genomes render their complete identification and assembly impossible.

All arthropods contain one intermediate chain. Except for *Tribolium*, the arthropods contain two light-intermediate chains. In addition, *Drosophila pseudoobscura* and *Drosophila persimilis* both contain a third light-intermediate chain. The sets of dynein light chains are very divergent in all analysed arthropods, although one of each of the LC8, Roadblock, and TcTex light chains is common to all species. In addition to these common light chains, all species have different numbers and variants of dynein light chains. It is remarkable that the *Drosophila* species have the largest number and most divergent set of light chains. Especially, they have five to eight additional light chains of the Roadblock family. The list of TcTex light chains also

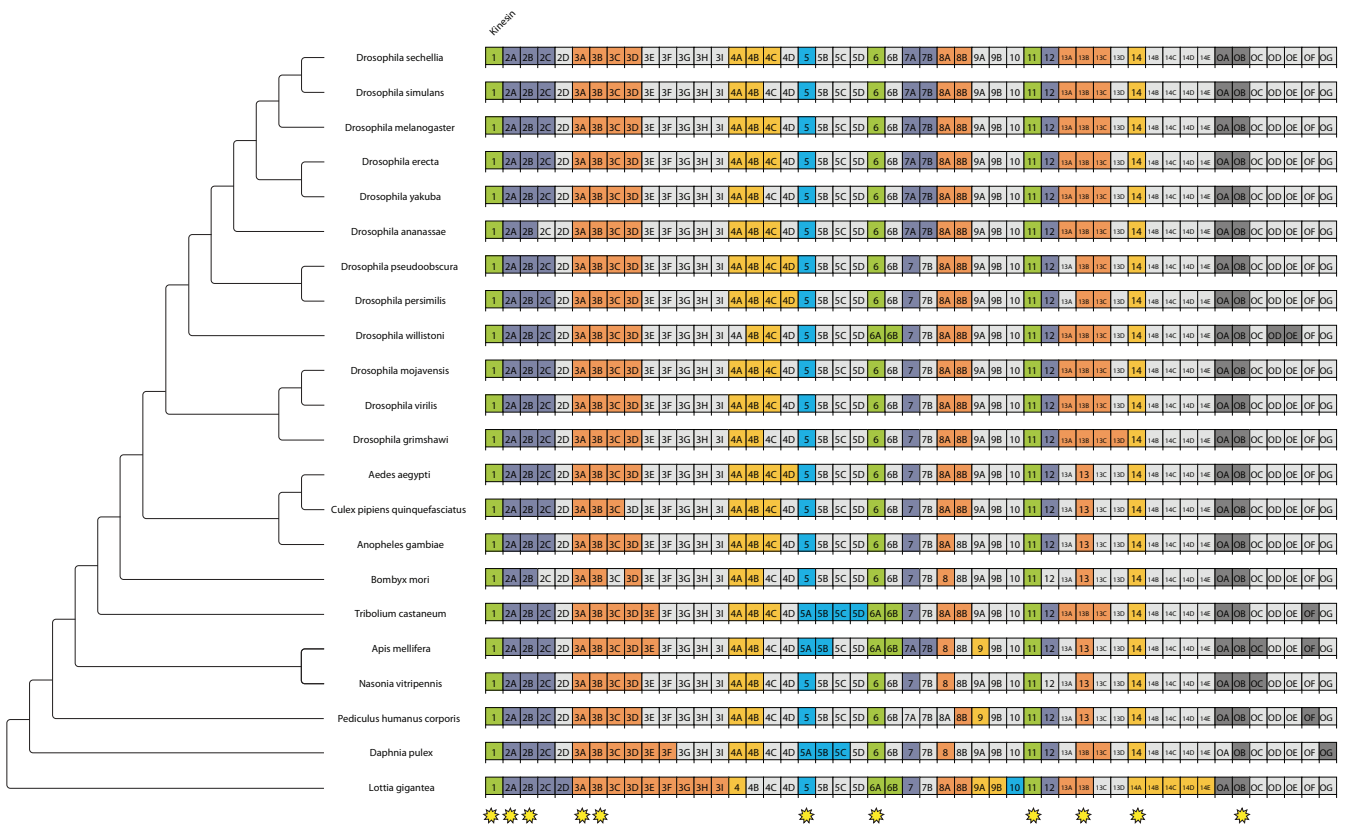


Figure 3
Kinesin repertoire of the arthropods. This chart shows the kinesin repertoire for all species in the analysis as in Figure 1.

includes the ones that are associated with the axonemal dynein heavy chain. Because of their diversity it is not possible to specify, which of the TcTex light chains are associated with the cytoplasmic dynein heavy chain. Therefore, all TcTex homologs are listed.

Similar to the mammals, the arthropods contain one gene of each of the subunits of the dynactin complex (Figure 6). Only the genomes of the *Drosophila* species encode another version of the dynactin p150 subunit. These genes are close homologs to the well described Glued (dynactin p150) gene in *Drosophila melanogaster* [35] but have not been identified previously. We did not find any splice variants of any of the dynactin transcripts of the arthropods, although different splice forms exist for all of the mammalian dynactin subunits.

Arthropod phylogeny

First, we calculated the phylogenetic tree of each of the protein families. When inspecting the phylogenetic tree of each protein family, it can be stated that three clades and their internal topologies are constant: The *Drosophila* clade, a clade of *Apis mellifera* and *Nasonia vitripennis*, and the clade of *Aedes aegypti*, *Culex pipiens quinquefasciatus*,

and *Anopheles gambiae*. Only in the tree of the LC8 proteins (see Additional File 1), the clade of *Anopheles*, *Aedes* and *Culex* is placed within the *Drosophila* clade. All other species were placed at varying branches. The discrepancy among the phylogenetic trees based on the dynein and dynactin subunits was higher when compared to the ones based on myosins and kinesins (see Additional File 1). The trees calculated from myosins and kinesins only disagree in the positions of *Bombyx mori*, *Tribolium castaneum* and *Pediculus humanus corporis*.

As has been found, when analysing the phylogeny of several homologs in a set of species, each homolog might result in a different phylogeny. This might result from the different rates of evolutionary change for different genes, from long-branch-attraction artifacts, or from sampling unrecognized paralogs [36]. Concerning unrecognized paralogs, we are confident that we were able to distinguish paralogs and orthologs since we have used very large datasets of protein sequences for the classification of the motor proteins with a wide taxonomic sampling ([7]; Kollmar, unpublished data). In order to compensate for this asynchronous evolution, a phylogenomics approach was used to infer the phylogeny of the 21 arthropods. For

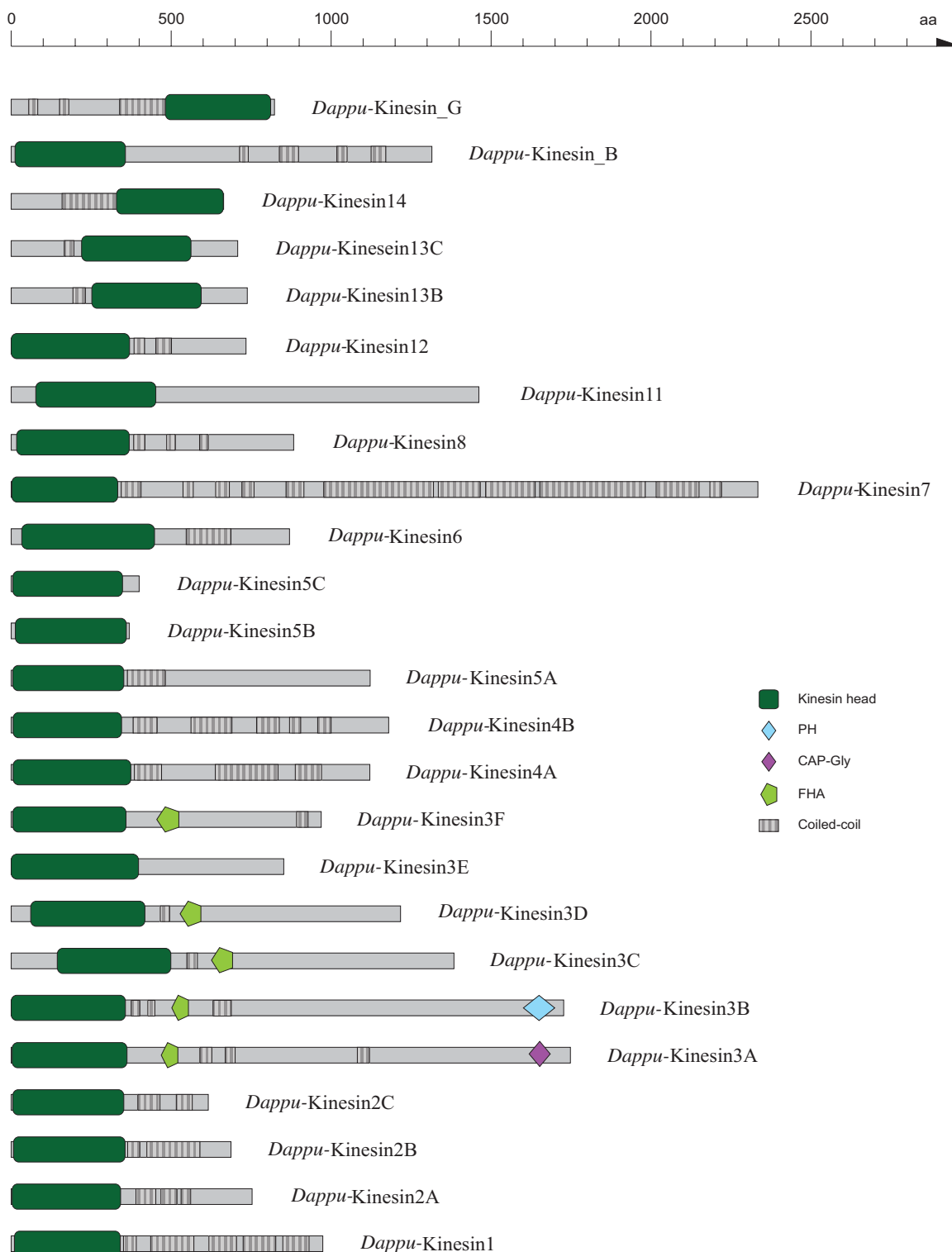


Figure 4
Domain organisation of the *Daphnia pulex* kinesins. The sequence name is given next to the respective kinesin. A colour key to the domain names and symbols is given on the right except for the kinesin domains that are coloured in dark-green. The abbreviations for the domains are: CAP-Gly, Cytoskeleton-associated protein-Gly; FHA, forkhead homology associated; PH, pleckstrin homology.

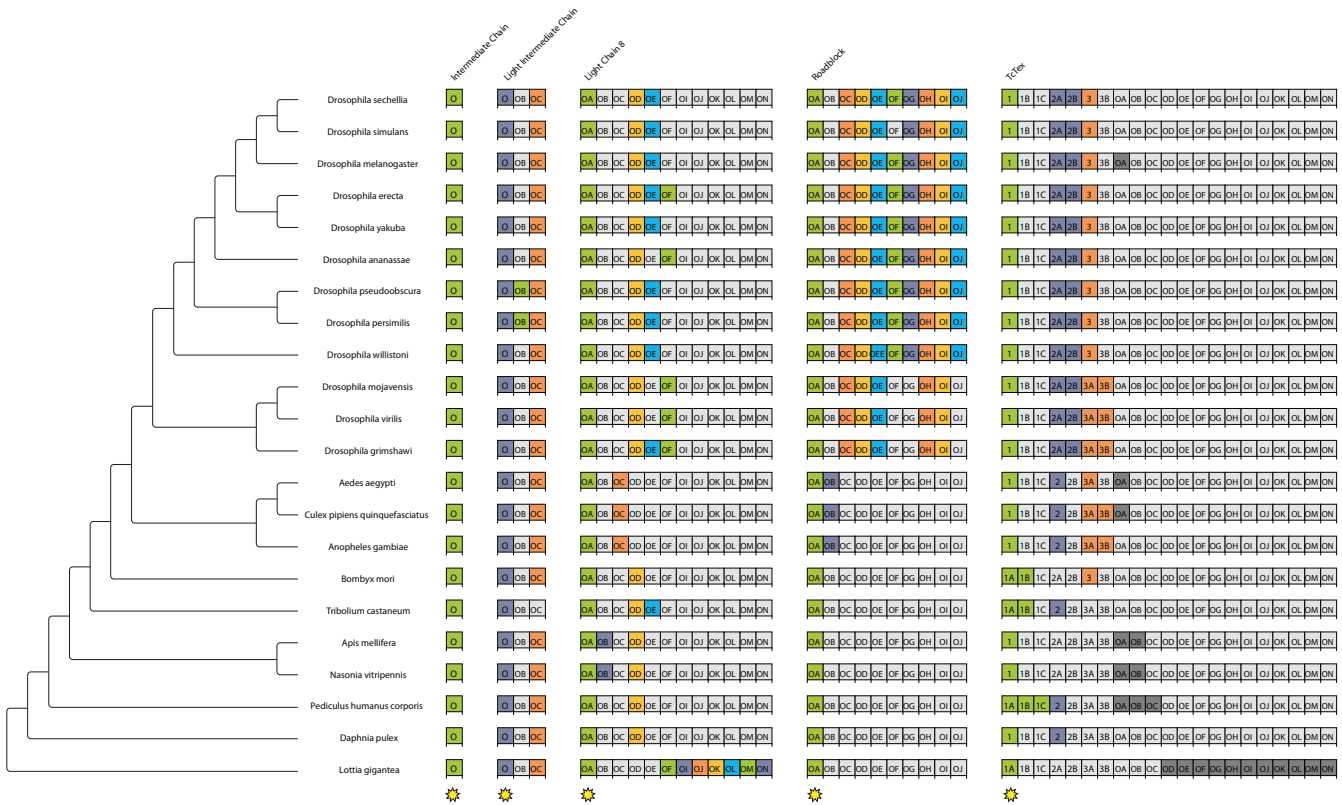


Figure 5
Dynein repertoire of the arthropods. This chart shows the dynein repertoire for all species in the analysis as in Figure 1. *

each protein family, the classes/variants, for which a homolog exists in every species, were concatenated resulting in more representative sequences by averaging out different rates of evolutionary change. For the dynein, dynactin and ARP (actin-related protein) proteins, only one of the homologs was found in all species, whereas eight of the myosins and ten of the kinesins are shared by all analysed arthropods (marked with stars in Figures 1, 3, 5, and 6). Thus, for each of the 22 species, 31 homologs were used, amounting to 682 motor protein sequences. The resulting trees are shown in Figure 7. Except for the placement of *Tribolium*, all four phylogenomics trees show identical phylogenies. All branches are supported by very high bootstrap values and are therefore reliable within the limits of the method. The placement of *Pediculus* depends on the method used. In the trees generated with neighbour joining (see Additional File 2), *Pediculus* forms a clade with *Nasonia* and *Apis*, whereas with maximum likelihood, only *Nasonia* and *Apis* are monophyletic and *Pediculus* is more closely related to *Daphnia*.

The phylogenetic tree inferred from the occurrence of classes/variants has a limited resolution and agrees only in some respects with the maximum likelihood tree: *Dro-*

sophila form a clade, *Drosophila pseudoobscura* and *Drosophila persimilis* are monophyletic, *Drosophila virilis*, *Drosophila mojavensis* and *Drosophila grimshawi* are monophyletic and *Culex*, *Aedes* and *Anopheles* are monophyletic.

Discussion

Most of the myosins that we discuss here have been identified and annotated in the course of the annotation of over 2000 myosins from more than 300 organisms [7]. Since then, the genome sequences of the arthropod species *Culex pipiens quinquefasciatus* and *Pediculus humanus corporis* have been finished as well as that of the mollusc *Lottia gigantea*, which we used as outgroup. All myosins have been grouped into 35 classes. The arthropods encode members of 13 of these classes, namely members of the classes I, II, III, V, VI, VII, IX, XV, XVIII, XIX, XX, XXI, and XXII. It has been found, that the *Drosophila melanogaster* NinaC protein, which has previously been classified as class-III myosin, is part of the new class-XXI [7]. Most arthropod genomes contain a real ortholog to the mammalian class-III myosins. Although both class-III and class-XXI myosins have an N-terminal kinase domain, the phylogenetic tree of the motor domain sequences clearly shows that both classes are distinct. *Daphnia pulex* con-

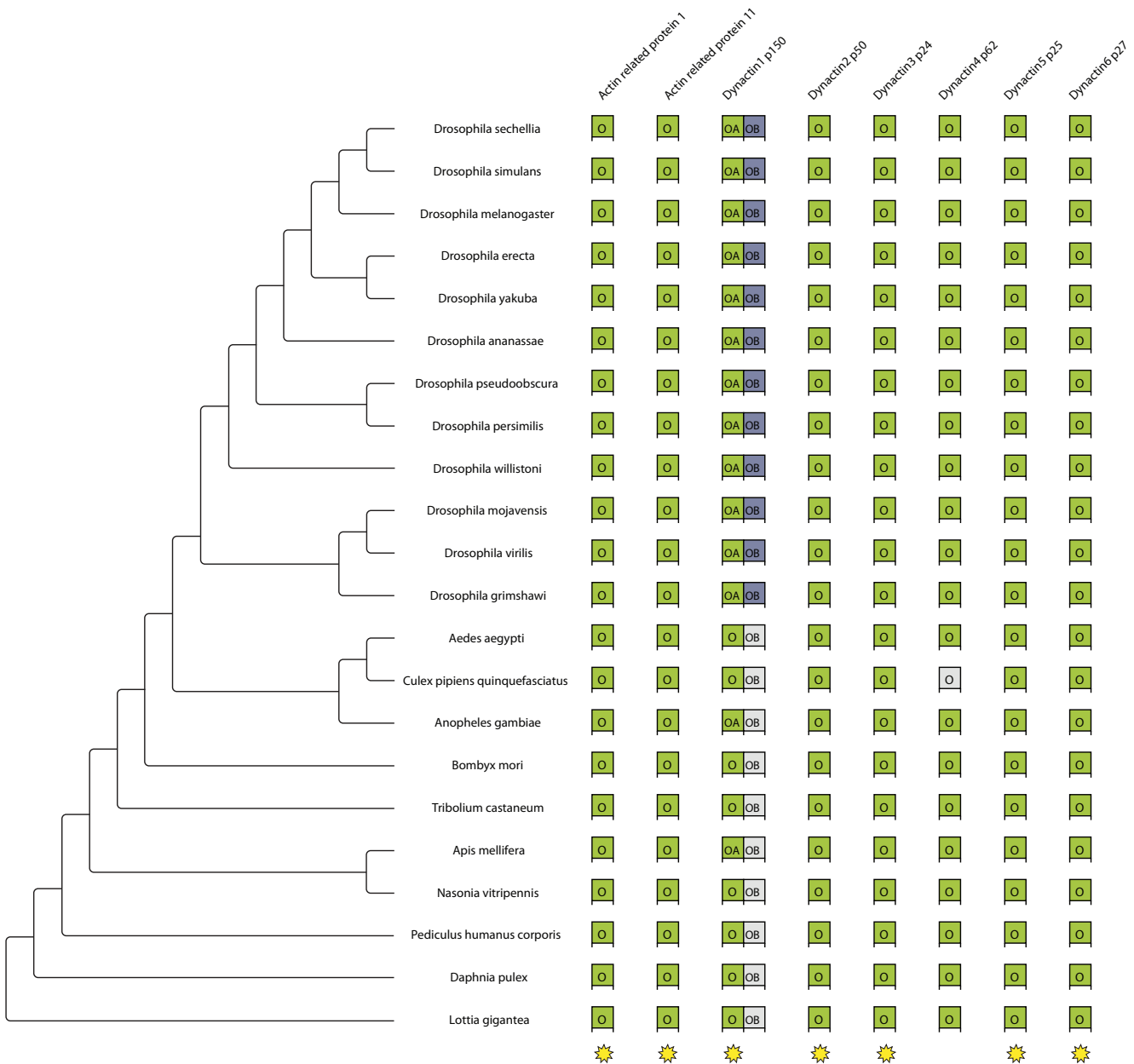
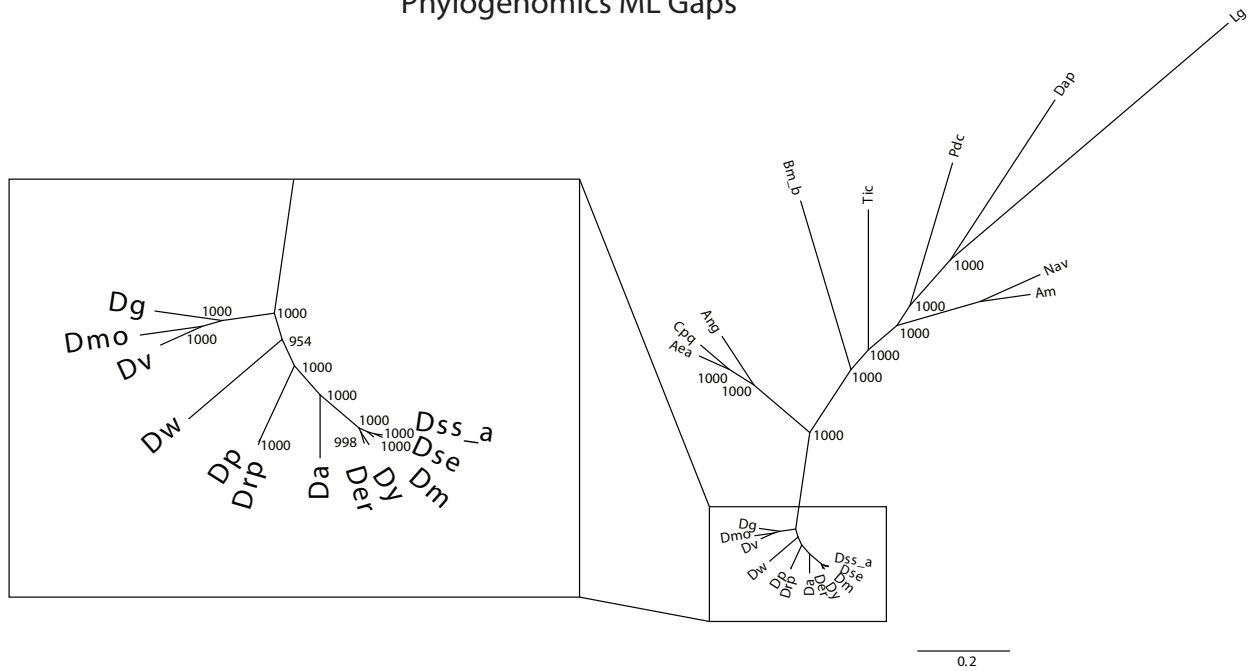


Figure 6
Arp/Dynactin repertoire of the arthropods. This chart shows the Arp/dynactin repertoire for all species in the analysis as in Figure 1.

tains the largest diversity of myosins, while the *Drosophila species* seem to have lost several classes, namely the members of class-III, class-IX, and class-XIX. Most of the *Drosophila species* have also lost their class-XXII myosin. Class-XXII myosins have two tandem repeats of MyTH4 and FERM domains like the class-VII myosin, but they miss the N-terminal SH3-like domain as well as the SH3 domain in the C-terminal tail. The specific function of a member of the class-XXII myosin has not been analyzed yet.

Of the kinesin superfamily the arthropods have members of all 14 specified classes [34] except for class-X. Class-IX kinesins have only been identified in *Apis mellifera* and *Pediculus humanus corporis*. However, the function of class-IX kinesins is not clear yet [11]. In addition to the kinesins, that could be classified, each of the analyzed arthropod species contains two or more kinesin homologs that could not be grouped to any of the known classes. Two of these orphan kinesins have been identified in all arthro-

Phylogenomics ML Gaps



Class Occurrence BI

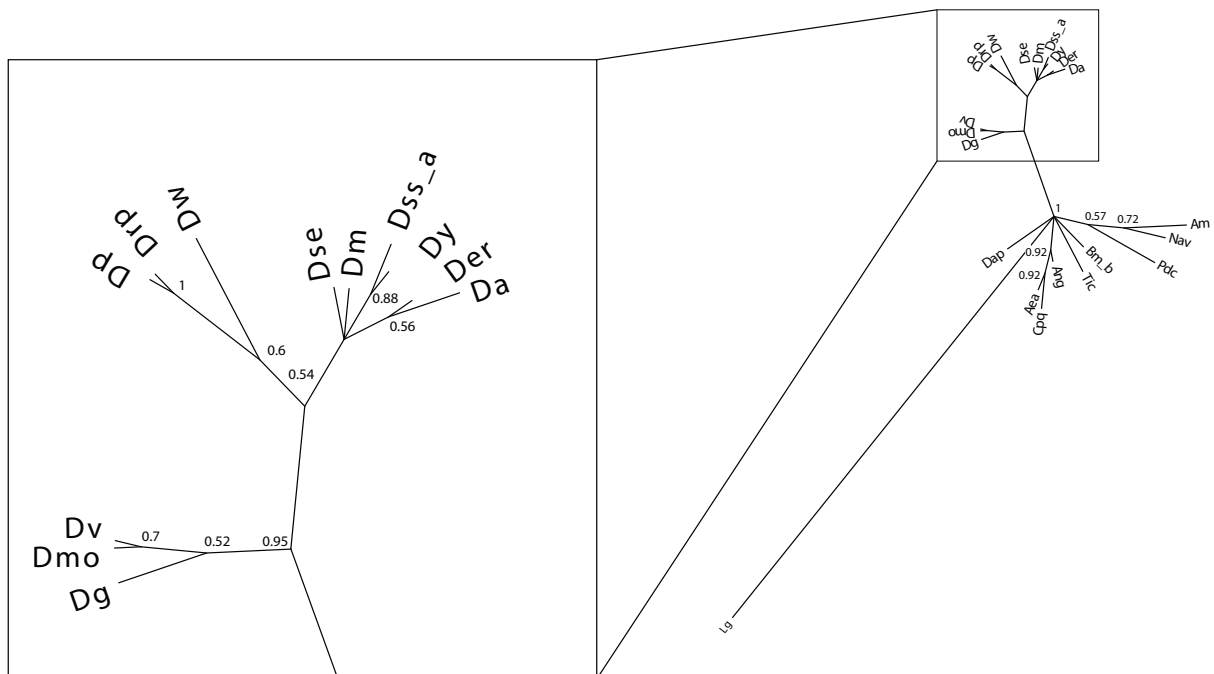


Figure 7
Phylogenomics and Class Occurrence. The trees illustrate the phylogenetic relationship between the arthropod species. The phylogenomic trees are based on a total of 682 concatenated protein sequences. Methods are indicated. The class occurrence tree was constructed using Bayesian inference based on the presence or absence of protein classes/variants as indicated in the inventory (Figures 1, 3, 5, and 6). The average standard deviation of split frequencies was 0.0087.

pod species except *Daphnia*, but some arthropods contain further species-specific kinesins. Notably, *Drosophila willistoni* contains two further kinesins, of which homologs have not been identified in any of the other sequenced arthropod genomes. Compared to the myosin repertoire, the kinesin inventory of the arthropods is far more varied. Although the analyzed arthropods have members of almost all classes, there are prominent differences in the subclass composition. Even the *Drosophila* species have different sets of kinesins. Thus, it is likely that the evolution of the kinesin diversity in arthropods is strongly determined by taxon- and species-specific gene losses and gene duplication events.

The arthropods contain a highly variable set of cytoplasmic dynein subunits. The dynein motor protein complex is build of dynein heavy chains, intermediate chains, light-intermediate chains, and the light chain 8, the Roadblock, and the TcTex light chains. All arthropods encode one dynein intermediate chain and a dynein light-intermediate chain. In addition, the closely related species *Drosophila pseudoobscura* and *Drosophila persimilis* contain another dynein light-intermediate chain. Of the light chains, the arthropods share one of each of the different types, the LC8, the Roadblock, and the TcTex light chains. All arthropods contain different numbers of further homologs of these light chains. Thus, they can build very specific cytoplasmic dynein complexes. For example, if all members of the Roadblock light chain family are also members of the cytoplasmic dynein complex the *Drosophila* species could build up to nine different cytoplasmic dynein complexes just by exchanging light chains of the Roadblock family. These different Roadblock light chains might bind different cargoes and by tissue specific or developmentally regulated expression of these Roadblock genes the *Drosophila* species might be able to fine tune their dynein mediated transport processes. Thus, there are far more possibilities to adjust cargo binding by combining different light chains than by using the dynein activator complex, dynactin. The arthropods contain one of each of the eleven dynactin subunits. Alternative splice forms have not been identified. Only the *Drosophila* species contain a further homolog of the p150 (Glued) subunit, that has not been identified and characterized yet.

It has been observed, given heterogeneous evolutionary rates, that the results of the maximum likelihood method are statistically more robust than the ones produced by neighbour joining [37]. Therefore we conclude that *Apis*, *Nasonia*, and *Pediculus* are not monophyletic, but that *Pediculus* is more closely related to *Daphnia*. The class occurrence tree shows that the classification system we used for the protein families does not contradict the finding of the sequence-based phylogenetic inference.

Our study suggests the following phylogeny: The *Drosophila* clade is composed of the *Drosophila simulans*/*Drosophila sechella* clade which forms a clade with *Drosophila melanogaster*. This clade together with the *Drosophila yakuba*/*Drosophila erecta* clade forms the melanogaster subgroup. This subgroup together with *Drosophila ananassae* forms the melanogaster group. The melanogaster group is most closely related to the obscura group, a clade that consists of *Drosophila pseudoobscura* and *Drosophila persimilis*. The closest relative to the obscura group is *Drosophila willistoni*. All of the before mentioned species form the subgenus Sophophora. Its sister subgenus is *Drosophila*, consisting of the clade of *Drosophila virilis*/*Drosophila mojavensis* and *Drosophila grimshawi* (taxonomy as in [29]). The phylogeny of the *Drosophila* clade is in exact agreement with what has been found in an analysis based on the complete genome sequences of the twelve species [29].

The closest relatives to the *Drosophila* clade are *Aedes aegypti* and *Culex pipiens*, forming one clade, and *Anopheles gambiae*. All these species belong to the Diptera. The placing of the remaining species, that have been analyzed here, is mainly in accordance with an analysis of 128 arthropod species that was based on 275 morphological variables as well as 18S and 28S rDNA data [38]. In accordance with this study, the Lepidoptera, to which *Bombyx mori* belongs, are the closest relatives to the Diptera forming the Mecopteroidea. Also in agreement with the morphological data, the Hymenoptera (*Nasonia vitripennis*/*Apis mellifera*) are basal to the Mecopteroidea together forming the Holometabola, and the Phthiraptera (*Pediculus humanus corporis*) are basal to the Holometabola. The main difference between our study and the analysis of the morphological data is the placement of *Tribolium castaneum*, a Coleoptera species. Our study placed *Tribolium* closer to the Mecopteroidea while the other study placed the Coleoptera outside the Hymenoptera and Mecopteroidea. *Daphnia pulex*, a Crustacea species, diverged earlier to all the other Hexapoda species.

Conclusion

In this analysis, we were able to resolve the phylogenetic relationship of 21 completely sequenced arthropod species based in their motor proteins. A large number of sequences were used that have been checked manually. We have systematically analyzed the protein inventory of all species as well as the domain composition of all members of the four protein families in *Daphnia pulex*. When inferring phylogenetic trees from the sequence data, variations in evolutionary speed were accounted for by using a phylogenomics approach. This analysis produced a phylogenetic tree that is highly resolved and that has statistically well supported branchings. Our findings are in accordance with results from studies based on whole

genome and rDNA sequences as well as morphological variables. We can conclude that from all arthropods analyzed, *Daphnia pulex* is the most basal one. *Pediculus humanus corporis* is the closest relative to *Daphnia*, followed by the clade of *Apis mellifera* and *Nasonia vitripennis*. Next, *Tribolium castaneum* and *Bombyx mori* diverged, followed by the mosquito species and the *Drosophila* clade.

Methods

Identification and annotation of the arthropod myosins, kinesins, and dynein/dynactin subunits

The genes for *Aea*, *Ang*, *Am*, *Bm*, *Cpq*, *Da*, *Der*, *Dg*, *Dm*, *Dmo*, *Drp*, *Dp*, *Dse*, *Dss*, *Dv*, *Dy*, *Dw*, *Nav*, *Pdc*, and *Tic* have been obtained by TBLASTN searches against the insects section of the NCBI wgs database [39]. The *Dap* sequences have been obtained by TBLASTN searches against the 8.7× coverage Dappu v1.1 draft genome sequence assembly (September, 2006) provided by the DOE Joint Genome Institute [40] and the *Daphnia* Genomics Consortium [41]. All hits were manually analysed at the genomic DNA level. The correct coding sequences were identified with the help of the multiple sequence alignments of the corresponding proteins. In this process, the sequence alignments of all proteins contained in our in-house version of CyMoBase have been used. As the amount of protein sequences increased (especially the number of sequences in classes with few representatives), many of the initially predicted sequences were reanalysed to correctly identify all exon borders. Where possible, EST data available from the NCBI EST database has been analysed to help in the annotation process. All sequence related data (names, corresponding species, GenBank ID's, alternative names, corresponding publications, domain predictions, and sequences) and references to genome sequencing centers are available through the CyMoBase [42,43].

Building trees

The phylogenetic trees based on protein sequences were generated using two different methods: 1. Neighbour joining using the GONNET substitution matrix with bootstrapping (1,000 replicates) using ClustalW 2.0 [44]. 2. Maximum likelihood (ML) [45] using a JTT model with estimated proportion of invariable sites and bootstrapping (1,000 replicates) using PHYML [46].

The sequence data, which was used for the analyses, were multiple sequence alignments consisting either of single homologous sequences from each species or multiple concatenated homologous sequences from each species (phylogenomics approach). For comparison, multiple sequence alignments were used including columns with gaps or with columns containing gaps removed.

The class occurrence tree was generated using Bayesian inference with a binary model using MrBayes 3.1.2 [47].

For each species the existence/non-existence of a protein class/variant was used as a binary character as depicted in Figure 7. Using this encoding, each species is represented by a series of binary characters, one for each protein class/variant. Constant rates were used whereas gamma-distributed rates gave very similar results. The tree was generated using 1.000.000 generations and a burnin of 500.000 generations since at that point the average standard deviation of split frequencies fell below 0.011.

Domain and motif prediction

Protein domains were predicted using the SMART [48,49] and Pfam [50,51] web server. The prediction of protein motifs (coiled coils, leucine zipper, etc.) is mainly based on the results of the predict-protein server [52,53]. The IQ-motifs and N-terminal domains of the myosins were predicted manually based on the homology to similar domains of other myosins included in the multiple sequence alignment of the myosins. The recognition motifs included in the SMART and Pfam databases are too restrictive, as the motifs have been created based on the small datasets available some years ago.

Authors' contributions

FO performed the data analysis of the myosins, kinesins, and dynein subunits, as well as the phylogenetic analysis of all sequences. SB assembled all dynactin sequences and performed their analysis. MK assembled all myosin, kinesin, and dynein sequences. All authors wrote and approved the final manuscript.

Additional material

Additional file 1

Phylogenetic trees of the motor proteins.

Phylogenetic trees of the motor proteins. The file contains the phylogenetic trees of the concatenated sequences of the myosin, the kinesins, the dynein subunits, the dynactin subunits, and the ARP proteins.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-173-S1.pdf>]

Additional file 2

Phylogenetic tree of the arthropods based on the neighbor joining method.

Phylogenetic tree of the arthropods based on the neighbor joining method. The file contains the phylogenetic tree of the concatenated sequences of all motor proteins calculated using the neighbor joining method.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-173-S2.pdf>]

Acknowledgements

This work has been funded by grant I80798 of the VolkswagenStiftung and grants KO 2251/3-I and KO 2251/6-I of the Deutsche Forschungsgemeinschaft.

The sequencing and portions of the analyses were performed at the DOE Joint Genome Institute under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, Los Alamos National Laboratory under Contract No. W-7405-ENG-36 and in collaboration with the Daphnia Genomics Consortium (DGC) [41]. Additional analyses were performed by wFleaBase, developed at the Genome Informatics Lab of Indiana University with support to Don Gilbert from the National Science Foundation and the National Institutes of Health. Coordination infrastructure for the DGC is provided by The Center for Genomics and Bioinformatics at Indiana University, which is supported in part by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. Our work benefits from, and contributes to the Daphnia Genomics Consortium.

References

- Vale RD: **The molecular motor toolbox for intracellular transport.** *Cell* 2003, **112**:467-480.
- Schliwa M, Woehlke G: **Molecular motors.** *Nature* 2003, **422**:759-765.
- Mallik R, Gross SP: **Molecular motors: strategies to get along.** *Curr Biol* 2004, **14**:R971-982.
- Lakemper S, Meyhofer E: **Back on track – on the role of the microtubule for kinesin motility and cellular function.** *J Muscle Res Cell Motil* 2006, **27**:161-171.
- Vale RD, Milligan RA: **The way things move: looking under the hood of molecular motor proteins.** *Science* 2000, **288**:88-95.
- Lawrence CJ, Morris NR, Meagher RB, Dawe RK: **Dyneins have run their course in plant lineage.** *Traffic* 2001, **2**:362-363.
- Odronitz F, Kollmar M: **Drawing the tree of eukaryotic life based on the analysis of 2269 manually annotated myosins from 328 species.** *Genome Biol* 2001, **8**(9):R196.
- Desnos C, Huet S, Darchen F: **'Should I stay or should I go?': myosin V function in organelle trafficking.** *Biol Cell* 2007, **99**:411-423.
- Krendel M, Mooseker MS: **Myosins: tails (and heads) of functional diversity.** *Physiology (Bethesda)* 2005, **20**:239-251.
- Burgess DR: **Cytokinesis: new roles for myosin.** *Curr Biol* 2005, **15**:R310-311.
- Miki H, Okada Y, Hirokawa N: **Analysis of the kinesin superfamily: insights into structure and function.** *Trends Cell Biol* 2005, **15**:467-476.
- Wade RH, Kozielski F: **Structural links to kinesin directionality and movement.** *Nat Struct Biol* 2000, **7**:456-460.
- Hirokawa N, Takemura R: **Molecular motors and mechanisms of directional transport in neurons.** *Nat Rev Neurosci* 2005, **6**:201-214.
- Vale RD, Case R, Sablin E, Hart C, Fletterick R: **Searching for kinesin's mechanical amplifier.** *Philos Trans R Soc Lond B Biol Sci* 2000, **355**:449-457.
- Caviston JP, Holzbaur EL: **Microtubule motors at the intersection of trafficking and transport.** *Trends Cell Biol* 2006, **16**:530-537.
- Mazumdar M, Misteli T: **Chromokinesins: multitasking players in mitosis.** *Trends Cell Biol* 2005, **15**:349-355.
- Oiwa K, Sakakibara H: **Recent progress in dynein structure and mechanism.** *Curr Opin Cell Biol* 2005, **17**:98-103.
- Vallee RB, Williams JC, Varma D, Barnhart LE: **Dynein: An ancient motor protein involved in multiple modes of transport.** *J Neurobiol* 2004, **58**:189-200.
- Wickstead B, Gull K: **Dyneins across eukaryotes: a comparative genomic analysis.** *Traffic* 2007, **8**:1708-1721.
- Levy JR, Holzbaur EL: **Cytoplasmic dynein/dynactin function and dysfunction in motor neurons.** *Int J Dev Neurosci* 2006, **24**:103-111.
- Muresan V, Stankewich MC, Steffen W, Morrow JS, Holzbaur EL, Schnapp BJ: **Dynactin-dependent, dynein-driven vesicle transport in the absence of membrane proteins: a role for spectrin and acidic phospholipids.** *Mol Cell* 2001, **7**:173-183.
- Schroer TA, Sheetz MP: **Two activators of microtubule-based vesicle transport.** *J Cell Biol* 1991, **115**:1309-1318.
- 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, Sutton GG, Wortman JR, Yeaman MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Bereman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabriellian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidenkiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wasserman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC: **The genome sequence of Drosophila melanogaster.** *Science* 2000, **287**:2185-2195.
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Lofthus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anouhair V, Arensburg P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscut D, Barnstead M, Cai S, Center A, Chaturvedi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL: **The genome sequence of the malaria mosquito Anopheles gambiae.** *Science* 2002, **298**:129-149.
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, Ren Q, Zdobnov EM, Lobo NF, Campbell KS, Brown SE, Bonaldo MF, Zhu J, Sinkins SP, Hogenkamp DG, Amedeo P, Arensburg P, Atkinson PW, Bidwell S, Biedler J, Birney E, Bruggner RV, Costas J, Coy MR, Crabtree J, Crawford M, Debruyen B, Decaprio D, Eglmeier K, Eisenstadt E, El-Dorri H, Gelbart WM, Gomes SL, Hammond M, Hannick LI, Hogan JR, Holmes MH, Jaffe D, Johnston JS, Kennedy RC, Koo H, Kravitz S, Kriventseva EV, Kulp D, Labutti K, Lee E, Li S, Lovin DD, Mao C, Mauceli E, Menck CF, Miller JR, Montgomery P, Mori A, Nascimento AL, Naveira HF, Nusbaum C, O'Leary S, Orvis J, Pertea M, Quesneville H, Reidenbach KR, Rogers

- YH, Roth CW, Schneider JR, Schatz M, Shumway M, Stanke M, Stinson EO, Tubio JM, Vanzee JP, Verjovski-Almeida S, Werner D, White O, Wyder S, Zeng Q, Zhao Q, Zhao Y, Hill CA, Raikhel AS, Soares MB, Knudson DL, Lee NH, Galagan J, Salzberg SL, Paulsen IT, Dimopoulos G, Collins FH, Birren B, Fraser-Liggett CM, Severson DW: **Genome sequence of *Aedes aegypti*, a major arbovirus vector.** *Science (New York, NY)* 2007, **316**:1718-1723.
26. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Yu J, Wang J, Li R, Shi J, Li H, Li G, Su J, Wang X, Li G, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Xu J, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li C, Li D, Sun Y, Zhang Z, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Li J, Ye J, Chen H, Zhou Y, Liu B, Wang J, Ye J, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B, Wang W, Ye C, Li S, Wang J, Wong GK, Yang H: **A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*).** *Science* 2004, **306**:1937-1940.
27. Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikummar G, Shimomura M, Nagamura Y, Shin IT, Abe H, Shimada T, Morishita S, Sasaki T: **The genome sequence of silkworm, *Bombyx mori*.** *DNA Res* 2004, **11**:27-35.
28. Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Friedrich M, Grimmeikhuijzen CJ, Klingler M, Lorenzen M, Richards S, Roth S, Schroder R, Tautz D, Zdobnov EM, Muzny D, Gibbs RA, Weinstock GM, Attaway T, Bell S, Buhay CJ, Chandrabose MN, Chavez D, Clerk-Blankenburg KP, Cree A, Dao M, Davis C, Chacko J, Dinh H, Dugan-Rocha S, Fowler G, Garner TT, Ganes J, Gnirke A, Hawes A, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Jackson L, Kovar C, Kowis A, Lee S, Lewis LR, Margolis J, Morgan M, Nazareth LV, Nguyen N, Okwuonu G, Parker D, Richards S, Ruiz SJ, Santibanez J, Savard J, Scherer SE, Schneider B, Sodergren E, Tautz D, Vattahil S, Villasana D, White CS, Wright R, Park Y, Beeman RW, Lord J, Oppert B, Lorenzen M, Brown S, Wang L, Savard J, Tautz D, Richards S, Weinstock G, Gibbs RA, Liu Y, Worley K, Weinstock G, Elsik CG, Reese JT, Elhaik E, Landan G, Graur D, Arensburger P, Atkinson P, Beeman RW, Beidler J, Brown SJ, Demuth JP, Drury DW, Du YZ, Fujiwara H, Lorenzen M, Maselli V, Osanai M, Park Y, Robertson HM, Tu Z, Wang JJ, Wang S, Richards S, Song H, Zhang L, Sodergren E, Werner D, Stanke M, Morgenstern B, Solovyev V, Kosarev P, Brown G, Chen HC, Ermolaeva O, Hlavina W, Kapustin Y, Kiryutin B, Kitts P, Maglott D, Pruitt K, Sapojnikov V, Souvorov A, Mackey AJ, Waterhouse RM, Wyder S, Zdobnov EM, Zdobnov EM, Wyder S, Kriventseva EV, Kadowaki T, Bork P, Aranda M, Bao R, Beermann A, Berns N, Bolognesi R, Bonneton F, Bopp D, Brown SJ, Bucher G, Butts T, Chaumot A, Denell RE, Ferrier DE, Friedrich M, Gordon CM, Jindra M, Klingler M, Lan Q, Lattorff HM, Laudet V, von Letvetsov C, Liu Z, Lutz R, Lynch JA, da Fonseca RN, Posnien N, Reuter R, Roth S, Savard J, Schinko JB, Schmitt C, Schoppmeier M, Schroder R, Shiply TD, Simonnet F, Marques-Souza H, Tautz D, Tomoyasu Y, Trauner J, Zee M Van der, Vervoort M, Wittkopp N, Wimmer EA, Yang X, Jones AK, Sattelle DB, Ebert PR, Nelson D, Scott JG, Beeman RW, Muthukrishnan S, Kramer KJ, Arakane Y, Beeman RW, Zhu Q, Hogenkamp D, Dixit R, Oppert B, Jiang H, Zou Z, Marshall J, Elpidina E, Vinokurov K, Oppert C, Zou Z, Evans J, Lu Z, Zhao P, Sumathipala N, Altincicek B, Vilcinskas A, Williams M, Hultmark D, Hetru C, Jiang H, Grimmeikhuijzen CJ, Hauser F, Cazzamali G, Williamson M, Park Y, Li B, Tanaka Y, Predel R, Neupert S, Schachtner J, Verleyen P, Raible F, Bork P, Friedrich M, Walden KK, Robertson HM, Angeli S, Forest S, Bucher G, Schuetz S, Maleszka R, Wimmer EA, Beeman RW, Lorenzen M, Tomoyasu Y, Miller SC, Grossmann D, Bucher G: **The genome of the model beetle and pest *Tribolium castaneum*.** *Nature* 2008, **452**:949-955.
29. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comeron JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipiski A, Findeiss S, Freyhult E, Fulton L, Fulton R, Garcia AC, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J, Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jaffe DB, Jagadeeshan S, Jeck WR, Johnson J, Jones CD, Jordan WC, Karpen GH, Kataoka E, Keightley PD, Kheradpour P, Kirkness EF, Koerich LB, Kristiansen K, Kudrna D, Kulathinal RJ, Kumar S, Kwok R, Lander E, Langley CH, Lapoint R, Lazzaro BP, Lee SJ, Levesque L, Li R, Lin CF, Lin MF, Lindblad-Toh K, Llopart A, Long M, Low L, Lozovsky E, Lu J, Luo M, Machado CA, Makalowski W, Marzo M, Matsuda M, Matzkin L, McAllister B, McBride CS, McKernan B, McKernan K, Mendez-Lago M, Minx P, Mollenhauer MU, Montooth K, Mount SM, Mu X, Myers E, Negre B, Newfeld S, Nielsen R, Noor MA, O'Grady P, Pachter L, Papacait M, Parisi MJ, Parisi M, Parts L, Pedersen JS, Pesole G, Phillippy AM, Ponting CP, Pop M, Porcelli D, Powell JR, Prohaska S, Pruitt K, Puig M, Quesneville H, Ram KR, Rand D, Rasmussen MD, Reed LK, Reenan R, Reily A, Remington KA, Rieger TT, Ritchie MG, Robin C, Rogers YH, Rohde C, Rozas J, Rubenfield MJ, Ruiz A, Russo S, Salzberg SL, Sanchez-Gracia A, Saranga DJ, Sato H, Schaeffer SW, Schatz MC, Schlenke T, Schwartz R, Segarra C, Singh RS, Sirot L, Sirot M, Sisneros NB, Smith CD, Smith TF, Spieth J, Stage DE, Stark A, Stephan W, Strausberg RL, Stempel S, Sturgill D, Sutton G, Sutton GG, Tao W, Teichmann S, Tobarri YN, Tomimura Y, Tsolas JM, Valente VL, Venter E, Venter JC, Vieira FG, Villella AJ, Villasante A, Walenz B, Wang J, Wasserman M, Watts T, Wilson D, Wilson RK, Wing RA, Wolfner MF, Wong A, Wong GK, Wu CI, Wu G, Yamamoto D, Yang HP, Yang SP, Yorke JA, Yoshida K, Zdobnov E, Zhang P, Zhang Y, Zimin AV, Baldwin J, Abdouelleil A, Abdulkadir J, Abebe A, Abera B, Abreu J, Acer SC, Aftuck L, Alexander A, An P, Anderson E, Anderson S, Arachi H, Azer M, Bachantsang P, Barry A, Bayul T, Berlin A, Bessette D, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Bourzgui I, Brown A, Cahill P, Channer S, Cheshatsang Y, Chuda L, Citroen M, Collymore A, Cooke P, Costello M, D'Acio K, Daza R, De Haan G, DeGray S, DeMaso C, Dhargay N, Dooley K, Dooley E, Doricent M, Dorje P, Dorjee K, Dupes A, Elong R, Falk J, Farina A, Faro S, Ferguson D, Fisher S, Foley CD, Franke A, Friedrich D, Gadbois L, Gearin G, Gearin CR, Giannoukos G, Goode T, Graham J, Grandbois E, Grewal S, Gyaltsen K, Hafez N, Hagos B, Hall J, Henson C, Hollinger A, Honan T, Huard MD, Hughes L, Hurhula B, Husby ME, Kamat A, Kanga B, Kashin S, Khazanovich D, Kisner P, Lance K, Lara M, Lee W, Lennon N, Letendre F, LeVine R, Lipovsky A, Liu X, Liu J, Liu S, Lokyitsang T, Lokyitsang Y, Lubonja R, Lui A, MacDonald P, Magnisalis V, Maru K, Matthews C, McCusker W, McDonough S, Mehta T, Meldrim J, Meneus L, Mihai O, Mihalev A, Mihova T, Mittelman R, Mlenga V, Montmayeur A, Mulrain L, Navidi A, Naylor J, Negash T, Nguyen T, Nguyen N, Nicol R, Norbu C, Norbu N, Novod N, O'Neill B, Osman S, Markiewicz E, Oyono OL, Patti C, Phunkhang P, Pierre F, Priest M, Raghuraman S, Rege F, Reyes R, Rise C, Rogov P, Ross K, Ryan E, Settipalli S, Shea T, Sherpa N, Shi L, Shih D, Sparrow T, Spaulding J, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Strader C, Tesfaye S, Thomson T, Thoulutsang Y, Thoulutsang D, Topham K, Topping I, Tsamila T, Vassiliev H, Vo A, Wangchuk T, Wangdi T, Weiland M, Wilkinson J, Wilson A, Yadav S, Young G, Yu Q, Zembek L, Zhong D, Zimmer A, Zvirko Z, Jaffe DB, Alvarez P, Brockman W, Butler J, Chin C, Gnerre S, Grabherr M, Kleber M, Mauceli E, MacCallum I: **Evolution of genes and genomes on the *Drosophila* phylogeny.** *Nature* 2007, **450**:203-218.
30. Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang Y, Waldron L, Verduzco D, Clerk-Blankenburg KP, Dubchak I, Noor MA, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA: **Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution.** *Genome Res* 2005, **15**:1-18.
31. Breathnach R, Chambon P: **Organization and expression of eucaryotic split genes coding for proteins.** *Annu Rev Biochem* 1981, **50**:349-383.

32. Odronitz F, Kollmar M: **Comparative genomic analysis of the arthropod muscle myosin heavy chain genes allows ancestral gene reconstruction and reveals a new type of 'partially' processed pseudogene.** *BMC Mol Biol* 2008, **9**:21.
33. Miki H, Setou M, Kaneshiro K, Hirokawa N: **All kinesin superfamily protein, KIF, genes in mouse and human.** *Proc Natl Acad Sci USA* 2001, **98**:7004-7011.
34. Lawrence CJ, Dawe RK, Christie KR, Cleveland DW, Dawson SC, Endow SA, Goldstein LS, Goodson HV, Hirokawa N, Howard J, Malmberg RL, McIntosh JR, Miki H, Mitchison TJ, Okada Y, Reddy AS, Saxton WM, Schliwa M, Scholey JM, Vale RD, Walczak CE, Wordeman L: **A standardized kinesin nomenclature.** *J Cell Biol* 2004, **167**:19-22.
35. Waterman-Storer CM, Holzbaur EL: **The product of the Drosophila gene, Glued, is the functional homologue of the p150Glued component of the vertebrate dynactin complex.** *J Biol Chem* 1996, **271**:1153-1159.
36. Doyle JJ: **Gene trees and species trees: Molecular Systematics as one-character taxonomy.** *Systematic Botany* 1992, **17**:144-163.
37. Hasegawa M, Fujiwara M: **Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny.** *Mol Phylogenet Evol* 1993, **2**:1-5.
38. Wheeler WC, Whiting M, Wheeler QD, Carpenter JM: **The Phylogeny of the Extant Hexapod Orders.** *Cladistics* 2001, **17**:113-169.
39. **NCBI BLAST with arthropoda genomes** [http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=insects]
40. **JGI: Joint Genome Institute** [<http://www.jgi.doe.gov/Daphnia>]
41. **wFleaBase: Daphnia waterflea genome database** [<http://wFleaBase.org>]
42. Odronitz F, Kollmar M: **Pfarao: a web application for protein family analysis customized for cytoskeletal and motor proteins (CyMoBase).** *BMC Genomics* 2006, **7**:300.
43. **CyMoBase** [<http://www.cymobase.org/>]
44. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**:3497-3500.
45. Felsenstein J: **Evolutionary trees from DNA sequences: a maximum likelihood approach.** *J Mol Evol* 1981, **17**:368-376.
46. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic biology* 2003, **52**:696-704.
47. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics (Oxford, England)* 2003, **19**:1572-1574.
48. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P: **SMART 5: domains in the context of genomes and networks.** *Nucleic acids research* 2006, **34**:D257-260.
49. **SMART** [<http://smart.embl-heidelberg.de/>]
50. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR: **The Pfam protein families database.** *Nucleic Acids Res* 2004, **32**:D138-141.
51. **Pfam** [<http://www.sanger.ac.uk/Software/Pfam/>]
52. Rost B, Yachdav G, Liu J: **The PredictProtein server.** *Nucleic acids research* 2004, **32**:W321-326.
53. **PredictProtein** [<http://www.predictprotein.org/>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

