Analysis of Cytoskeletal and Motility Proteins in the Sea Urchin Genome Assembly

Morris RL[1,*,†], **Hoffman MP**[2,†], **Obar RA**[3], **McCafferty SS**[1], **Gibbons IR**[4], **Leone AD**[2], **Cool J**[1], **Allgood EL**[1], **Musante AM**[1], **Judkins KM**[1], **Rossetti BJ**[1], **Rawson AP**[1], and **Burgess DR**[2,*]

1 *Department of Biology, Wheaton College, Norton, Massachusetts 02766*

2 *Department of Biology, Boston College, Chestnut Hill, MA 0246*

3 *Tethys Research, LLC, 53 Downing Road, Bangor, Maine 04401; Present Address: Massachusetts General Hospital Cancer Center, Charlestown, MA 02129*

4 *Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720*

Abstract

The sea urchin embryo is a classical model system for studying the role of the cytoskeleton in such events as fertilization, mitosis, cleavage, cell migration and gastrulation. We have conducted an analysis of gene models derived from the *Strongylocentrotus purpuratus* genome assembly and have gathered strong evidence for the existence of multiple gene families encoding cytoskeletal proteins and their regulators in sea urchin. While many cytoskeletal genes have been cloned from sea urchin with sequences already existing in public databases, genome analysis reveals a significantly higher degree of diversity within certain gene families. Furthermore, genes are described corresponding to homologs of cytoskeletal proteins not previously documented in sea urchins. To illustrate the varying degree of sequence diversity that exists within cytoskeletal gene families, we conducted an analysis of genes encoding actins, specific actin-binding proteins, myosins, tubulins, kinesins, dyneins, specific microtubule-associated proteins, and intermediate filaments. We conducted ontological analysis of select genes to better understand the relatedness of urchin cytoskeletal genes to those of other deuterostomes. We analyzed developmental expression (EST) data to confirm the existence of select gene models and to understand their differential expression during various stages of early development.

INTRODUCTION

The cytoskeleton is involved in all aspects of cell migration, mitosis and cytokinesis, organelle movements, ciliary and flagellar movements, and other cell shape changes. In developing sea urchin embryos, the cytoskeleton has long been studied by developmental and cell biologists, biochemists and biophysicists. In this early work, researchers used cells of sea urchins to make important and path breaking observations on aspects of cell division (Harvey, 1956; Hiramoto, 1968; Salmon, 1975). Sea urchins have been excellent models for the study of motility of sperm (Brokaw and Gibbons, 1973; Yokota et al., 1987), the proteins responsible for actin based

movements (Begg and Rebhun, 1979; Bryan et al., 1993; Edds, 1980; Fishkind et al., 1987; Kane, 1986; Mabuchi, 1986; Mabuchi and Spudich, 1980; Schroeder, 1968; Tilney and Gibbins, 1969; Tilney et al., 1973), and the proteins of the microtubule based mitotic apparatus (Cohen and Rebhun, 1970; Kane, 1967; Scholey et al., 1985). In fact, the first dynein cloned and sequenced came from sea urchins (Gibbons, et al., 1991; Ogawa, 1991). Moreover, the sea urchin zygote has served as one of the best model systems for the biophysical and biochemical investigation of cytokinesis and mitosis (Mabuchi, 1986; Rappaport, 1996; Salmon, 1975; Sluder, 1979). For example, antibody microinjection has been used to analyze the functions of myosin, kinesin-related, and dynein motor proteins in mitosis and cytokinesis in the early sea urchin embryo (Mabuchi and Okuno, 1977; Wright et al, 1993; Strickland et al., 2005). Cells from sea urchins remain just as valuable today as models for the study of mitosis, cytokinesis, ciliary assembly and beat, and cell shape transformation.

Because cells of sea urchins serve as such exquisite models for study of the cytoskeleton, a number of key cytoskeletal proteins were identified first in these cells using biochemical methods and were subsequently cloned and sequenced. The rich diversity of cloned (some partially cloned) and sequenced sea urchin cytoskeletal proteins ranges from both actin and microtubule based motors, (including myosins I, II, V, and X, kinesin-1, -2, -5, and 14-B, and dyneins), actin and associated actin-binding proteins (such as fascin, scruin, spectrin, actin-binding protein, moesin, villin, and dystrophin,) and tubulin and various microtubule-binding proteins (such as MAPs). Such a wide variety of cloned cytoskeletal genes provides a rich established database that allows for significant assistance in annotating the sea urchin genome.

In addition, due to the fact that cytoskeletal proteins are so conserved evolutionarily, structurally and functionally, there is ample opportunity for detailed comparative analysis of the genome. Because the cytoskeleton is involved in fundamental cellular functions such as mitosis, cytokinesis, membrane translocations, and cellular motility, the genes encoding many cytoskeletal proteins are amenable to critical evolutionary analysis.

MATERIALS AND METHODS

Lists of known major cytoskeletal proteins were generated for major classes of functional proteins. Using the common names for cytoskeletal and motility proteins, the category and number of proteins annotated in the sea urchin assembly were: actin (5), actin-binding proteins (36), myosins (12 classes, 29 proteins), intermediate filament proteins (5), dyneins (15 heavy chains, 10 light chains, 4 light-intermediate chains, 4 intermediate chains), kinesins (14 families, 45 motors, 2 associated proteins), tubulin (15), microtubule-binding proteins (32), and regulatory proteins (16). Because of the abundance of cytoskeletal and motility proteins, we have focused this analysis on only the commonly known proteins and those previously known to be present in the sea urchin.

Determination of homologs from gene predictions

In order to derive and annotate *Strongylocentrotus purpuratus* cytoskeletal and motility genes, lists of proteins of interest (called 'Indexing Gene Products') were compiled from a combination of literature and database sources. Amino acid sequences corresponding to homologs of these proteins were collected from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), EST libraries (Poustka, et al., 2003; available at A. Poustka's website, www.molgen.mpg.de/~ag_seaurchin/; and Zhu, et al., 2001), the Sea Urchin Genome Project website at the California Institute of Technology (BAC-end sequence: http://sugp.caltech.edu/), and the *Chlamydomonas* Flagellar Proteome Project web page (Pazour et al, 2005: http://labs.umassmed.edu/chlamyfp/index.php). These lists were then used to identify candidate orthologs by interrogating the list of predicted gene products (or 'GLEANs') derived from processing the April 2005 release (build 1.1) of the whole-genome

shotgun sequence assembly of the *S. purpuratus* genome using the GLEAN3 algorithm (Sea Urchin Genome Sequencing Consortium, *Science*, in press). In brief, the list of predicted genes was queried with the individual protein sequences from the Indexing Gene Products List using the *S. purpuratus* Genome Search website at the National Institute of Dental and Craniofacial Research Developmental Mechanisms Unit (http://urchin.nidcr.nih.gov/blast/index.html) or the Sea Urchin Project website at the Human Genome Sequencing Center at Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/) using the TBLASTN algorithm (Altschul et al., 1993). In most cases, TBLASTN returned a table of results consisting of one or more GLEAN gene models that represented 'strong hits' with a length and sequence consistent with the protein query. However, in the cases of myosin, kinesin, and dynein heavy chains, multiple sequence alignments (ClustalW, http://www.ebi.ac.uk/clustalw/) often showed that the strongest hits corresponded to no more than 25–60% of the full-length of the query sequence. In these cases, multiple GLEAN gene models identified by TBLASTN were used to manually extend the sequence of the predicted gene. Whenever possible, gene models from the same scaffold were used to construct full-length gene models and the GLEAN gene identity assigned to the most N-terminal sequence was used to annotate the complete concatenated sequence.

Strategies to fully annotate each of the superfamilies in this study (actins, myosins, tubulins, kinesins, and dyneins) varied somewhat. For myosins, in addition to genes identified by BLAST score alone, a list of GLEANs grouped by SMART and PFAM conserved domains was generated by Hynes and Whittaker which was subsequently made available to our group (http://luria.mit.edu/urchin/) (see this issue). We derived sequences encoding predicted myosin motor domains from this list to corroborate results derived from BLAST searches described above and for conducting phylogenetic analysis described below. For kinesins, a database of 147 candidate kinesin genes was generated by keyword search of NCBI GNOMON predictions and by kinesin motor domain PFAM search of GLEAN3 predictions. Amino acid sequences of all 147 candidates were assembled into a phylogenetic tree alone or with example homologs from human, mouse, *Danio*, *Drosophila*, and *Caenorhabditis elegans*, to narrow down the 147 kinesin-related proteins to a non-redundant set. Kinesin sequences were deemed redundant if they shared 100% identity with another candidate for 95 amino acids or more. GLEAN3 predictions were selected over redundant NCBI predictions if the GLEAN3 prediction was greater than half the length of a partially identical NCBI prediction. For any two GLEAN3 predictions deemed redundant, the sequence with the most complete motor domain was retained. In cases where phylogenetic trees did not unambiguously assign a gene prediction to a kinesin family, genes encoding putative kinesin motor domains were assigned to the most similar kinesin family as determined by BLAST score. For the dyneins, the chain of exons corresponding to each of the 15 heavy chain genes was obtained by mapping the full length peptide sequence of the corresponding Indexing Gene Product onto the March 2005 scaffolds of the *S. purpuratus* genome by using the BLAT server (Kent, 2002) at the University of California at Santa Cruz Genome Bioinformatics world wide web site (http://genome.ucsc.edu). Approximately 10% of the dynein exons were either too divergent or too short to be obtained in this way and these were obtained by using one of the above Blast servers to perform *in silico* PCR between the neighboring, previously identified exons on each side. The identity of each sea urchin gene model obtained in this way was confirmed by using the BLAT server to map its translated peptide sequence back onto the human genome and verifying that its chromosomal location corresponded to that of the correct indexing gene. Some regions of the dynein genes are highly conserved and particular care was necessary to avoid crossing between genes in these regions. Because of the exceptionally large size of the dynein heavy chain genes, with each gene having 12–14 kb of coding sequence that spanned ~70 kb of genome, the exon chains for the different genes involved from 2 to 5 scaffolds each.

Phylogenetic analysis to infer gene homologies

We used a phylogenetic approach to verify the predicted *Strongylocentrotus* amino acid sequences for the actin, CLIP170, dynein, fascin, gelsolin, kinesin, lamin B, myosin, tubulin, and profilin gene and gene families. For each gene and gene family, known or predicted homologs were recovered from GenBank using a number of strategies. For CLIP170, fascin, gelsolin, lamin B, and profilin, sequences for the following taxa were downloaded from GenBank when genomic data was available: human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), frog (*Xenopus laevis/gilli*), zebrafish (*Danio rario*), nematode (*Caenorhabditis elegans*), fruit-fly (*Drosophila melanogaster*), tunicate (*Ciona intestinalis*), and sea anemone (*Nematostella vectensis*). When appropriate, yeast (*Saccharomyces cerevisiae*) sequences were included as an outgroup. Actin sequences were derived from Carlini *et al*. (2000) and from GenBank. Myosin sequences were from Hodge and Cope (2000). Kinesin sequences were obtained from Miki *et al*. (2005). Dynein gene sequences were mostly obtained from GenBank. Additional dynein gene models for Chlamydomonas (Pazour et al., 2006) were obtained from the Chlamydomonas Genome version 3 produced by the US Department of Energy Joint Genome Institute, (http://www.jgi.doe.gov/) and are provided for use in this publication only. Additional dynein gene models for Tetrahymena (Asai and Wilkes, 2004) were obtained from the Tetrahymena Genome Database (http://www.ciliate.org) and are based upon sequence determined by the Institute for Genome Research. Within each gene family, efforts were made to include representatives of all subfamily and family members to assure homology of the urchin sequences.

A multiple sequence alignment was performed using the program MAFFT version 5.743 (Katoh *et al*., 2005) for the CLIP170, fascin, gelsolin, lamin B, and profilin sequences. The program MUSCLE (Edgar, 2004) was used to align the sequences for actin, kinesin, myosin, and tubulin gene families due to the size of the datasets. Full-length sequences of dynein heavy chains were aligned with the program ClustalX (Thompsom et al., 1997)). In all cases, the resulting multiple sequence alignments were visually inspected and refined by eye using the multiple sequence editor in MacVector (www.accelrys.com/products/macvector/). All phylogenetic analyses were performed using the program package PHYLIP 3.65 (Felsenstein, 2005). The phylogenetic relationships among the aligned sequences within each gene/gene family were estimated by the Neighbor-Joining method (Saitou and Nei, 1987) with the pairwise divergence among sequences estimated using the JTT (Jones *et al*. 1992) method. Confidence levels for the branching patterns for actin, dynein, kinesin, myosin, and tubulin were assessed by nonparametric bootstrapping (Felsenstein, 1985). Predicted urchin sequences that deviated significantly from other family members for each gene or gene family were re-assessed as to the validity of the results.

Expression analysis of kinesins

Expression profiles of select kinesins were downloaded as EST counts from NCBI's transcript-based EST libraries available for *S. purpuratus*: UniGene Build #9 described at http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=7668. Expression profiles suggested by analysis of EST counts were generated automatically from egg, cleavage, blastula, gastrula, and larva libraries (dbEST Library IDs.13749, 13750, 13752, 13751, and 13748, respectively, submitted by A.J. Poustka), supplemented with additional blastula data (dbEST Library ID.16781, submitted by J. Coffman), and primary mesenchyme cell data (dbEST Library ID.17157, submitted by C.A. Ettensohn). UniGene IDs for kinesins or kinesin-associated proteins in *S. purpuratus* were identified by keyword (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=7668), and expression profiles were generated for UniGene entries that matched a previously cloned *S. purpuratus* CDS or consistently matched the same kinesin family in all organisms. These UniGene entries included UniGene Spu.93: KRP180 (SPU_021317), UniGene Spu.964: Kinesin-C (SPU_009400),

UniGene Spu.13111: KIF1b subfamily, UniGene Spu.11982: KIF1A subfamily, UniGene Spu. 15: KRP85 (SPU_018378), and UniGene Spu.198: SpKAP115 (SPU_010954).

Analysis of relationship among major taxa based on cytoskeletal and motor proteins

Based on the NJ trees derived for each gene/gene family, putative homologs were determined for CLIP170, fascin, gelsolin, lamin B, myosin II, and profilin. The homologous sequences for the nine taxa listed above plus *S. purpuratus* were concatenated into a single file and the phylogenetic relationship among the ten taxa determined using maximum likelihood methods as implemented in the program PHYML (Guindon and Gascuel, 2003). The JTT+I+G model was used estimating the proportion of invariant sites (I) and the gamma distribution parameter (G) from the data. Non-parametric bootstrapping (100 replications) was used to infer levels of confidence for the resulting branching patterns. Branch support was further assessed by Bayesian analysis using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). Two heated Markov chains were run for 100,000 generations sampling every 100 generations using the JTT+G+I model. A burn-in time of 1000 generations was used to assure stationarity, and the remaining tree samples were used to generate a 50% majority rule consensus tree and to calculate the posterior probability of each clade.

RESULTS

A complete presentation of our results, showing each individual homolog identified from the GLEAN3 predictions can be found in Table 1. Overall, representatives of all major classes and families of cytoskeletal and motility proteins have been identified in the *S. purpuratus* genome. Selected representatives of major families are presented below.

Actin

Actin is the most abundant protein in most cells and both muscle and non-muscle actins have been well studied in numerous systems. Because the high sequence conservation of actin across species, likely due to the conserved nature of the abundant protein-protein interactions actin is involved in, searches for actin in the assembly are straightforward.

There are 36 different scaffolds containing genes predicted to have significant homology with actin (BLAST score > 1e-7). Phylogenetic analysis of these sequences together with a broad range of actin sequences derived from diverse species shows single predicted genes corresponding to previously characterized *S. purpuratus* cytoskeletal actins CyI (SPU_009481 groups with equal similarity to CyIa and CyIb), CyIIa, CyIIb, CyIIIb and the skeletal M actin gene. All other putative actin-encoding sequences were ruled out as outliers as they did not group into clades containing defined actin orthologs (data not shown).

Myosins

Myosins are actin-based motor proteins with a highly conserved ATPase and actin-binding globular head domain and a tail domain, allowing for tail-protein and tail-membrane interactions. Thorough analysis of the large myosin family of proteins, based on sequence analysis of the motor domain, reveals thirty-one sequences corresponding to twelve different classes of myosin. The myosin family has been well studied in sea urchins. Full-length clones and partial sequences have been previously characterized corresponding to eight different myosin classes. Initially, egg myosin II was characterized biochemically (Mabuchi, 1972; Kane, 1983), followed by the unconventional myosins I and VI (D'Andrea, 1993; Terasaki, 1997). Other myosins that have been partially sequenced include amoeboid-like myosin I, V, VI, VII, IX (Sirotkin, 2000). Of these, only myosins V and VI have been completely sequenced. Predictions of the highly conserved and myosin class-specific motor domain sequence (MYSc) has allowed the accurate prediction of homologs corresponding to myosins I, II, V, VI, IX.

Homologs pertaining to myosin classes III, IV, VII, X, XII, XV, XVI, XVII were not previously described in sea urchins and have been partially derived from whole-genome sequences To date, homologs corresponding to myosins VIII, XI, XII, XIII, and XIV have not been identified.

Actin-Binding Proteins

The ability of cells to change shape is largely due to the abundant variety of actin-binding proteins that function to regulate actin assembly into filaments, formation of actin networks, meshworks and bundles, stabilize filaments, and control disassembly of filaments and three-dimensional assemblies. From a list of common actin-binding proteins, a large variety of actin-binding proteins were identified by the search parameters in the sea urchin gene assembly. The Arp2/3 complex regulates actin branching networks and promotes actin polymerization (Kelleher, 1995) and biochemical analysis has shown that this complex exists in sea urchins (Terasaki, et al., 1997). A homolog of Arp3 has been predicted from the urchin genome. Subunits 2, 3 and 4 of the seven-subunit Arp2/3 complex have also been identified. Homologs of tropomyosin and tropomodulin, that bind actin on pointed ends and inhibit actin depolymerization, gelsolin, an actin severing protein, α-actinin, that crosslinks and bundles actin at higher concentrations (Honda, 1998) and WASP, that stimulates the actin-nucleating activity of the Arp2/3 complex, were also found. A homolog of ENA/VASP that associates with barbed ends and inhibits filament capping by CapZ (Krause M, et al., 2003) was predicted. SpCoel1, an ortholog of profilin, a protein that regulates actin polymerization by sequestering actin monomers, was previously identified in sea urchin coelomocytes (Smith LC, 1992) and this gene has been identified in the genome.

Homologs for Arp1/centractin, that polymerizes to form a complex with p150-*glued* and vesicle-specific spectrins (Holleran, 1996), Arp6 and Arp8, shown to be involved in heterochromatin organization in yeast (Shen, 2003), were also identified. Partial homologs to diaphanous Diaphanous/mDia/formin were determined. The mouse formin gene (mDia/Fmn) is located over approximately 400 kb and is encoded by at least 24 different exons (Wang et al., 1997). It is likely that the presence of multiple exons and/or large intronic regions has complicated efforts to identify a full-length formin gene from any one scaffold from the sea urchin genome assembly. A sequence encoding a putative diaphanous autoinhibitory domain (DAD) of formin was found isolated on a distinct scaffold from other predicted formin encoding exons. Other major actin-binding proteins identified include: ABP1, ABP278/filamin B, adducin, ankyrin, annexins (4,5,6,7,13), annilin, calmodulin, cofilin (1,2), ezrin/radixin/moesin, fascin, filamin (A, B, C), protocadherin, drebrin, dystonin, gelsolin, villin, spectrin (alpha, beta, G), talin, and beta-thymosin. Multiple homologs corresponding to genes encoding vertebrate skeletal muscle proteins have been predicted to exist following analysis of *S. purpuratus* genome sequences, including: tropomyosin, troponin I, troponin C, troponin T, titin, minititin, nebulin and dystrophin.

Tubulins

Because tubulin genes (like those for actins) are highly conserved, the search for tubulin-encoding genes by homology with known genes in other organisms and sea urchin ESTs is relatively straightforward. The genome of *S. purpuratus* carries multiple genes encoding alpha- and beta-tubulin (~9 and 6 genes, respectively), as well as gamma-tubulin (1 gene), delta-tubulin (1 gene) and epsilon-tubulin (2 genes). Only the alpha, beta, and gamma sequences were known previously in sea urchin. The nine expressed alpha-tubulin genes that we have identified fall into two groups based on the number and positions of their introns. Four of the genes (SPU_012679, SPU_021668, SPU_021669, and SPU_024615) contain 2 introns apiece, and five of the genes (SPU_006756, SPU_007984, SPU_016746, SPU_019990, and SPU_028221) contain 3 introns apiece. The structures of all nine identified alpha-tubulin genes share the two previously identified introns that seem to be specific to invertebrate alpha-

tubulins (Perumal, et al. 2005) at codon 1 (i.e. separating the initiator methionine codon from the second codon) and codon 75. Each of the nine genes encodes a slightly different alpha-tubulin sequence, with most of the diversity occurring near the carboxyl terminus.

There are at least six beta-tubulin genes (see Table 1), and each gene has three exons interrupted by two introns at (identical) conserved locations. Each of these genes encodes a distinct beta-tubulin sequence.

MAPs

Homologs corresponding to microtubule plus-end stabilizing protein Adenomatous Polyposis Coli (APC) and it's binding partner, End-binding protein 1 (EB1), were identified. CLIP170 binds EB1 and associates with tubulin dimers at growing microtubule ends (Folker, 2005; Strickland, et al., 2005). The dynactin complex regulates dynein interaction with vesicles and trafficking along microtubules and has been implicated in regulating the plus-end microtubule-based interaction with the cortex that leads to furrow induction and cytokinesis (Strickland et al., 2005). Putative homologs corresponding to dynactin/p150-*glued*, dynactin p25 and dynamitin have been determined.

The genome contains genes for a range of microtubule-associated proteins (MAPs), many of which were first discovered in mammalian tissues but had not been observed in sea urchin microtubules previously despite attempts in multiple laboratories. These include homologs of the heat-stable mammalian brain MAP Tau/MAP2/MAP4 family (SPU_013254, SPU_000651) which contain four of this family's hallmark SKXGSXDNXXHXPGGGXVXI microtubule-binding repeats, as well as two divergent repeats with the sequences SKCGSLGNSTHRAGGGNVKI and SKXGSXDNXXHXPSGGGXVXII; a homolog of MAP1 light chain 3 (SPU_009444); a ca. 200kDa MAP1A/MAP1B homolog (SPU_004628) that includes the carboxyl-terminal domain which encodes the MAP1 light chain and the heavy chain/light chain junction (Hammarback, et al., 1991; Langkopf, et al. 1992; Togel, et al. 1995); and TPX2 (SPU_008221). The sequence of the human genome was found to contain five homologs of the canonical sea urchin MAP, known as the 77kDa MAP (Bloom, et al. 1985) or "EMAP" (Li and Suprenant 1994), for which we have found a single gene in *S. purpuratus,* SPU_006911. By inspection of the *S. purpuratus* genome, we have also found a family of EMAP-related genes in *S. purpuratus*, including SPU_011819, SPU_026643, and SPU_001991, which may reflect the five human EMAP-like proteins termed EMLs1-5. Additional members of the well-studied MAP families and gene products that might correspond to other known sea urchin MAPs (see Vallee and Bloom 1984, Maekawa, et al. 1992, and Maekawa, et al. 1994) were not identified, and remain to be found. Agene (SPU_000096) of the MAST/orbit/CLASP family, which mediates microtubule attachment to kinetochores, was identified.

Kinesins

The kinesins represent an extensive superfamily of plus end-directed microtubule-based motor proteins. The identification of kinesins in sea urchin eggs and embryos originally exploited the ability to purify microtubules from these cells in such tremendous quantity that one could biochemically analyze the MT-associated ATPases including the kinesins (Scholey et al, 1984; Scholey et al, 1985; Collins and Vallee, 1986; Cole et al, 1992). Reverse genetics built upon this biochemical foundation to clone and sequence seven sea urchin kinesin motor polypeptides (Wright et al., 1991; Cole et al., 1993; Rashid et al., 1995; Rogers et al., 1999; Chui et al., 2000; Rogers et al., 2000) and non-motor associated polypeptides (Wedaman et al., 1993; Wedaman et al., 1996). These kinesins participate in intracellular movements such as organelle transport (Bi et al., 1997; Wright et al., 1991), mitosis (Chui et al., 2000; Rogers et al., 2000; Sharp et al., 2000) and ciliogenesis (Morris and Scholey, 1997).

From the 147 GLEAN3 or NCBI GNOMON gene predictions that included a whole or partial kinesin motor domain, 35 non-redundant kinesin genes and two kinesin-associated protein genes were identified in the *S. purpuratus* genome assembly including the 9 kinesin and kinesin-associated protein genes previously cloned from *S. purpuratus* (Chui et al., 2000; Cole et al., 1993; Rashid et al., 1995; Rogers et al., 1999; Rogers et al., 2000; Wedaman et al., 1993; Wedaman et al., 1996; Wright et al., 1991), as well as members of all 14 families of kinesins identified by Lawrence et al., (2004) and 13 of 14 identified by Wickstead and Gull (2006) (Figure 1C). Not surprisingly, a total of 35 kinesins in echinoderms is intermediate between the 45 in human and the 25 of *Drosophila* (Miki et al., 2005), but more than the 31 predicted to exist in *Ciona* (Vale, 2003). With the 28 new kinesins discovered here, *S. purpuratus* possesses the same completeness of kinesin family representation seen in vertebrates (Miki et al., 2005; Wickstead and Gull, 2006), but more than seen in *Drosophila*. Interesting findings from the *S. purpuratus* kinesin repertoire include a third kinesin-2, Sp-Osm-3, that represents the homodimeric kinesin-2 proteins, as well as an orphan kinesin Sp-KIF27-like that is most similar to the kinesin-4s but almost as equally divergent from all other families. As found by Wickstead and Gull (2006), our analysis divided the kinesin-12 family of Lawrence et al., (2005) into two clades that Wickstead and Gull name the kinesin-15 and kinesin-16 families, both of which are represented in *S. purpuratus*, but unlike Wickstead and Gull (2006) who excluded HsKIF26, our analysis supports a grouping of Kif26s together in a kinesin-11 family (Lawrence et al., 2005). Although the motor domain of SPU_015437 was KIF2-like (E value 6e-122), the full-length protein grouped with Dm CENP and Dm CENP meta which themselves usually group with the Kinesin-7s (Miki et al., 2005; Wickstead and Gull, 2006), warranting further study. Only a few sequences designated as Kinesin-14s in the analysis of Miki et al. (2005) formed clades with bootstrap values >50% in our analysis and so these sequences are neither grouped nor their family distinctly labeled in our phylogenetic tree (Figure 1C). This dispersion results from the great complexity of the kinesin-14s described before (Miki et al., 2005; Wickstead and Gull, 2006). Genes for the two known kinesin-associated non-motor polypeptides, kinesin-1-associated light chain KLC and kinesin-2-associated KAP (Wedaman et al., 1993; Wedaman et al., 1996) were also located (Table 2). Some notable absences from the urchin genome were additional kinesin-1s; Sp-KHC is more like Dm-KHC than like the three KIF5s found in mouse and human. Because of the typically long length of kinesin genes (averaging 19.9 exons per Sp gene in a sampling of 10 fully annotated sequences) some kinesin gene predictions remain incomplete and will require further analysis eventually perhaps increasing the number of kinesins found in *S. purpuratus*.

To begin investigating the possibility that expression of some families of kinesins is up-regulated to play roles in early development, we investigated available EST libraries to estimate the relative abundance of different kinesins at different stages of development (Figure 2). A variety of kinesins appear to be more actively expressed during cleavage stages than other stages, while some are expressed only later in development. The kinesin-2 subunits Sp-KR85 and Sp-KAP appear to increase in expression concomitant to the onset of ciliogenesis at the late blastula/early gastrula stages. These results support microinjection experiments showing that antibodies which block the function of kinesin-2 are capable of eliminating all kinesin-2 function through gastrulation (Morris and Scholey, 1997). While QPCR analysis is necessary to corroborate this evidence, these data illustrate the value of cataloging the entire complement of kinesins in a classic developmental model.

Dyneins

The ample supplies of sperm flagella that can be obtained from sexually mature sea urchins have long made them favored material for studying the biochemistry and molecular biology of dynein, especially since the unusually high molecular mass of the dynein heavy chain (~550 kDa, corresponding to a 14 kb gene) places emphasis on quantity. The first dynein gene to be

cloned and sequenced was that encoding one of the axonemal dynein heavy chains in sea urchin sperm flagella (Gibbons, et al., 1991; Ogawa, 1991). Shortly thereafter, the genes of the motor domain for the other 13 axonemal and two cytoplasmic dyneins that together comprise the core of the dynein motor family in eukaryotes (Gibbons, et al., 1994; Rasmusson, et al., 1994; Vaughan et al., 1996). Subsequent completion and annotation of the human and other mammalian genomes has resulted in the definition of an almost complete set of full-length dynein heavy chain sequences. Building upon this background, we decided to annotate the dynein genes in *S. purpuratus* by screening its genome assembly with dynein sequences from other sea urchin species where available, supplemented with full-length sequences from mammalian data banks as necessary. To encourage greater uniformity of dynein gene nomenclature in different species, we use a modestly extended version of the internationally recognized mammalian nomenclature for dynein genes (Wain et al., 1994) to describe our results.

All 13 of the axonemal dynein heavy chains and the two cytoplasmic dynein heavy chain genes that were previously identified in the sea urchin *Tripneustes gratilla* (Gibbons, et al., 1994) are present in the *S. purpuratus* genome (Table 3). In most cases, we were able to align ~95% of the estimated full-length heavy chain gene to its indexing homolog. The greatest difficulty occurred in the N-terminal region, which is the least conserved region in all dynein heavy chains. Averaged over their full available lengths, most of the dynein heavy chains are 95% identical to the homologous gene in *T. gratilla* and 60–70% identical to their mammalian homologs. The major exception was Sp-DNAH14 which showed a substantially lower identity of only 41%. The majority of dynein heavy chain genes in *S. purpuratus* show an unambiguous one-to-one relationship to their mammalian homologs. One exception is the one-to-three relationship of Sp-DNAH9, corresponding to the beta heavy chain of axonemal outer arm dynein, in which the sequence is approximately equidistant between those of three closely similar genes DNAH9, DNA11 and DNAH17 in mammals. The only other exception is that two single genes in mammals, DNAH3 and DNAH5, each appear equivalent to a closely-related gene pair in sea urchins, with the former corresponding to Sp-DNAH3 and Sp-DNAH4 and the latter corresponding to Sp-DNAH5 and Sp-DNAH15, In scanning the S purpuratus genome, we encountered no new dynein heavy chain genes, additional to those previously identified in *T. gratilla* (Gibbons et al, 1994).

The three axonemal dynein intermediate chains previously identified in *Anthocidaris crassispina* (Ogawa, et al., 1996, Kagami, et al., 1998) match the three intermediate chains of *S. purpuratus* characterized here (Table 3 and supplementary material). We also identified three axonemal lights chains. Among the other dynein subunits known from mammals, we have identified one intermediate chain, 3 light intermediate chains and all six of the light chains known to be associated with cytoplasmic dynein in mammals. We have not, as yet, found orthologs of cytoplasmic dynein 1 intermediate chain 2 or of cytoplasmic dynein light intermediate chain 2. The *S. purpuratus* genome contains, at least 5 genes encoding a DLC-type light chain and 5 encoding an LC8-type light chain, genes encoding two roadblock-type light chains and 3 kinds of Tctex-like light chains as well as an intermediate chain and 3 types of light-intermediate chains (see above and Table 3).

Intermediate Filaments

Intermediate filament family proteins are a diverse group of proline-rich rod shaped proteins that play roles in such diverse cellular processes as nuclear envelope structure, cellular structural support including functions of desmosomal junctions, cell-substrate adhesion, and cell structural support. Key intermediate filament family proteins have been identified in the sea urchin genome assembly, including: acidic keratin, desmin, lamin B and nestin. Genetic

diversity within the lamin gene family is likely to have arisen following the diversification of vertebrates from the chordates.

Phylogenetic Analysis

Using the data from the for CLIP170, fascin, gelsolin, lamin B, myosin II, and profilin genes, we inferred the relationship among human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), frog (*Xenopus laevis/gilli*), zebrafish (*Danio rario*), nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), sea urchin (*Strongylocentrotus purpuratus*), tunicate (*Ciona intestinalis*), and sea anemone (*Nematostella vectensis*) using maximum likelihood methods with yeast (*Saccharomyces cerevisiae*) as an outgroup. Generally, the major groupings of Chordata, Deuterostomes, and Bilateria are upheld with the Cnidaria (represented by *C. intestinalis*) forming a basal lineage leading to the divergence of the Bilateria. The Protostomes (Ectozoa), as represented by *Drosophila* (Arthropoda) and *C. elegans* (Nematoda), do not form a monophyletic group. However, the relationship among nematode, fruit fly, and sea urchin, which defines the region where Protostomes and Deuterostomes diverged, is not particularly well supported in this analysis, with relatively low bootstrap and Bayesian support values in this region of the phylogeny. These results are consistent with recent studies on the relationship among Eukaryotes based on differing suits of genes (Nei *et al*. 2001; Blair *et al*., 2005a; Blair *et al*. 2005b), demonstrating the potential of cytoskeletal and motility proteins in determining deep level relationships among organisms.

DISCUSSION

Annotation of the cytoskeletal genome of *S. purpuratus* has revealed the presence of representatives of essentially all known cytoskeletal proteins. Detailed evolutionary analysis of the genes of such functionally conserved proteins reveals an overall consistent trend of placing the genes most closely to those of the vertebrates. We conducted a detailed phylogenetic analysis of a broad spectrum of key cytoskeletal and motor proteins derived from the genome assembly and confirm what is becoming an accepted fact: *S. purpuratus*, being a deuterostome and representing echinoderms generally, is more closely related to chordates and vertebrates than are arthropods (i.e. *D. melanogaster*), or nematodes (i.e. *C. elegans*).

Perhaps more notable is the apparent absence of homologs for several vertebrate cytoskeletal and motor proteins. For instance phylogenetic analysis of the myosin family revealed several homologs that encode the motor domain of smooth and non-muscle myosin II, whereas no clear matches were identified for the motor domain of skeletal muscle myosin II. Sea urchins move using tube feet that are controlled by striated muscles and homologs of essentially all skeletal sarcomeric proteins were identified. It is thus likely that skeletal myosin II exists and the annotation has failed to discriminate its presence from other closely related myosin IIs. We were surprised to find two predicted myosin motor domains that corresponded to a new class of myosin, XVI, recently identified in rat brain development (Patel et al., 2001). The absence of a predicted myosin XIV and III, may indicate either a failed annotation or that *S. purpuratus* lacks these genes. The identification of such a wide variety of myosin classes makes possible their functional dissection in the many motile properties of cells from sea urchins.

The sea urchin actin gene family has been previously characterized in depth and the annotation has identified a full complement of the actin genes found in most vertebrates. Based on their highly conserved nature, comparison of actin gene sequences at the nucleotide level has been used historically as a basis for understanding evolutionary relationships between organisms. Over the past two decades, researchers have used biochemistry and cloning from cDNA libraries to reveal at least eight *S. purpuratus* actin genes, including a single skeletal muscle gene and two groups of five cytoskeletal actin genes: cytoskeletal actin (Cy) I, CyIIa, and CyIIb

are closely linked as one group that is distinct from CyIIIa and CyIIIb, which are closely linked to each other (Fang and Brandhorst, 1994; Lee et al., 1984).

In both the case of the alpha-tubulins and the case of the beta-tubulins we have identified, there is relatively little heterogeneity in either gene structure or derived protein sequence. The alpha-tubulin genes, in particular, appear to be minor variations on a single theme, and 8 of the 9 identified genes cluster closely together phylogenetically (see Figure 1b). This indicates that much of the functional diversity and selective patterns of expression of the tubulin proteins of the sea urchin (Gianguzza, et al., 1989) comes as a result of post-translational modifications of the tubulin proteins which have been widely documented in these families (see McKean, et al. 2001 for a review). The conserved intron locations and general structures of the genes is consistent with the "introns-early" hypothesis (see Perumal, et al., 2005), and the occurrence of truncated or otherwise divergent pseudogenes is also well-known in the alpha- and beta-tubulin gene families.

Identification of the entire complement of kinesins in the sea urchin will allow a thorough analysis of roles of individual kinesins and their families in development. With the discovery of a complete repertoire of 35 kinesin sequences, this study validates the use of sea urchin for studying the kinesins. With 35 gene sequences, we have identified a complex but not insurmountable set of kinesins in a model deuterostomre well-suited for investigating the expression of all individual gene products during early developmental stages and the functions of whole kinesin families for which functions are not well understood. It will also allow characterization of the entire kinesin ensemble required for ciliogenesis since all kinesin families proposed to play roles in ciliogenesis (Wickstead and Gull, 2006) are represented in this organism where the ciliogenic cycle has been so well characterized (Masuda, 1979; Stephens, 1995)

The phylogenetic tree in Figure 1D compares the predicted amino acid sequences of the dynein heavy chains genes in *S. purpuratus* to each other and to the corresponding genes in model organisms of other phyla. The largely monophyletic branching observed for the heavy chain subunits of outer arm dynein (ODA-beta and ODA-gamma), for the alpha and beta subunits of inner arm 1 dynein (IDA1-alpha and IDA1-beta) and for both cytoplasmic dynein 1 (Cyto1) and cytoplasmic dynein 2 (Cyto2) indicates that the conserved pattern of these subunits extends back to the earliest eukaryotic cells and is perhaps related to the extraordinary evolutionary stability of the $9 + 2$ axonemal structure of cilia and eukaryotic flagella. On the other hand, the putative inner arm dynein components in other branches of the dynein tree (IDA Groups 1, 2, and 3) appear to be somewhat less tightly constrained and they have radiated into distinct subfamilies in such organisms as *Chlamydomonas* and *Tetrahymena* that possibly utilize more varied patterns of axonemal movement. An apparent example of more recent gene duplication in dynein is the splitting of the single ODA-beta gene present in organisms from *Tetrahymena* to sea urchin into the triad of related genes found in mammals, a divergence that might possibly have been favored by substantial thickening of the $9 + 2$ structure of ciliary axonemes into the 9+9+2 organization present in mammalian spermatozoa.

With the *S. purpuratus* genome now available for analysis, it is interesting to note that different pre-genomic strategies to systematically explore the kinesin and dynein superfamilies in sea urchin produced qualitatively and quantitatively different outcomes. The PCR-based protocol used to identify dynein genes was particularly effective in finding most or all of the dynein heavy chain genes present in sea urchins (Gibbons et al., 1994), Drosophila (Rasmussen et al, 1994) and mammals (Vaughan et al., 1996) yet did not reveal associated proteins directly. In comparison, microtubule-affinity/pan-kinesin antibody/reverse genetic approach was particularly effective at revealing the multi-subunit makeup of kinesin holoenzymes while avoiding the complications of pseudogenes (Cole and Scholey, 1995), but it uncovered only

one-fifth of the kinesin genes present in sea urchins (Table 2). With genomic catalogs of dynein and kinesin genes as well as sequences of their associated proteins now available, it may be possible to bring to light sea urchin motor protein complexes with the completeness of the dynein heavy chain discoveries and the thoroughness of the kinesin holoenzyme discoveries.

Examples of most classes of intermediate filament proteins are present although in a very limited variety. The absence of such genes for key vertebrate cytoskeletal proteins, such as lamin-A and vimentin, are consistent with their absence in other invertebrates and new appearance in the vertebrates. The rather simple fleet of intermediate filament family proteins is also consistent with the absence of intermediate filament-linked cadherins (Whittaker et al., 2006). The duplication event splitting lamin A from lamin B1 and lamin B2 occurred sometime after the divergence of complex vertebrates from primitive chordates. There appears to be only a single lamin in *Ciona intestinalis*, so the duplication event likely occurred after the split of *Ciona* from the ancestor leading to higher vertebrates.

The initial characterization of the sea urchin cytoskeleton and motility genome provides new tools with which further analysis of cell structure and motility will be made possible. Cells and tissues of developing sea urchins are excellent models for analysis of such events as ciliogenesis, mitosis, cytokinesis, morphogenetic movements, lamellipod extension, filopod movements, and organelle translocation among others. Having the sea urchin cytoskeleton and motility genome available makes possible design of inhibitory antibodies, morpholinos, fluorescently tagged peptides and proteins, inhibitory peptides, constitutively active or inactive kinase domains and other such probes. These tools, combined with the ability to microinject into fertilized eggs at specific times during mitosis allows for functional dissection of the roles of specific cytoskeletal and motility proteins and their regulators. This experimental manipulation is unique in the study of cell division to the use of dividing echinoderm eggs. Thus, cells of sea urchins remain as critical for analysis of the roles of the cytoskeleton today as they were over one hundred years ago.

Acknowledgements

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410. [PubMed: 2231712]

Asai DJ, Wilkes DE. The dynein heavy chain family. J Eukaryot Microbiol 2004;51:23–9. [PubMed: 15068262]

Begg DA, Rebhun LI. pH regulates the polymerization of actin in the sea urchin egg cortex. J Cell Biol 1979;83:241–8. [PubMed: 41844]

Bi GQ, Morris RL, Liao G, Alderton JM, Scholey JM, Steinhardt RA. Kinesin- and myosin-driven steps of vesicle recruitment for Ca2+-regulated exocytosis. Journal of Cell Biology 1997;138:999–1008. [PubMed: 9281579]

Blair JE, Hedges SB. Molecular Phylogeny and Divergence Times of Deuterostome Animals. Mol Biol Evol 2005;22:2275–2284. [PubMed: 16049193]

Blair JE, Shah P, Hedges SB. Evolutionary sequence analysis of complete eukaryote genomes. BMC Bioinformatics 2005;6:53. [PubMed: 15762985]

Bloom GS, Luca FC, Collins CA, Vallee RB. Use of multiple monoclonal antibodies to characterize the major microtubule-associated protein in sea urchin eggs. Cell Motil 1985;5:431–46. [PubMed: 2866844]

Brokaw CJ, Gibbons IR. Localized activation of bending in proximal, medial and distal regions of sea-urchin sperm flagella. J Cell Sci 1973;13:1–10. [PubMed: 4738165]

Bryan J, Edwards R, Matsudaira P, Otto J, Wulfkuhle J. Fascin, an echinoid actin-bundling protein, is a homolog of the Drosophila singed gene product. Proc Natl Acad Sci U S A 1993;90:9115–9. [PubMed: 8415664]

Carlini DB, Reece KS, Graves JE. Actin Gene Family Evolution and the Phylogeny of Coleoid Cephalopods Mollusca, Cephalopoda. Mol Biol Evol 2000;17:1353–1370. [PubMed: 10958852]

Chui KK, Rogers GC, Kashina AM, Wedaman KP, Sharp DJ, Nguyen DT, Wilt F, Scholey JM. Roles of two homotetrameric kinesins in sea urchin embryonic cell division. Journal of Biological Chemistry 2000;275:38005–38011. [PubMed: 11006281]

Cohen WD, Rebhun LI. An estimate of the amount of microtubule protein in the isolated mitotic apparatus. J Cell Sci 1970;6:159–76. [PubMed: 5461428]

Cole DG, Cande WZ, Baskin RJ, Skoufias DA, Hogan CJ, Scholey JM. Isolation of a sea urchin egg kinesin-related protein using peptide antibodies. J Cell Sci 1992;101:291–301. [PubMed: 1629246]

Cole DG, Chinn SW, Wedaman KP, Hall K, Vuong T, Scholey JM. Novel heterotrimeric kinesin-related protein purified from sea urchin eggs. Nature 1993;366:268–70. [PubMed: 8232586]

Cole DG, Scholey JM. Purification of kinesin-related protein complexes from eggs and embryos. Biophysical Journal 1995;68:158s–162s. [PubMed: 7787059]

Collins CA, Vallee RB. A microtubule-activated ATPase from sea urchin eggs, distinct from cytoplasmic dynein and kinesin. Proc Natl Acad Sci U S A 1986;83:4799–803. [PubMed: 2873571]

De Giovanni C, Lollini PL, Dolcetti R, Landuzzi L, Nicoletti G, D'Andrea E, Scotland K, Nanni P. Uncoupling of growth inhibition and differentiation in dexamethasone-treated human rhabdomyosarcoma cells. Br J Cancer 1993;67:674–9. [PubMed: 8471424]

Dehmelt L, Halpain S. The MAP2/Tau family of microtubule-associated proteins. Genome Biol 2004;6:204. [PubMed: 15642108]Epub 2004

Dutcher SK. Long-lost relatives reappear, identification of new members of the tubulin superfamily. Current Opinion in Microbiology 2003;6:634–640. [PubMed: 14662361]

Edds KT. The formation and elongation of filopodia during transformation of sea urchin coelomocytes. Cell Motility 1980;1:131–140. [PubMed: 6890871]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 2004;32:1792–97. [PubMed: 15034147]

Fang H, Brandhorst BP. Expression of the actin gene family in embryos of the sea urchin Lytechinus pictus. Dev Biol 1996;173:306–17. [PubMed: 8575631]

Felsenstein J. Confidence limits on phylogenies, an approach using the bootstrap. Evolution 1985;39:783–791.

Felsenstein, J. Distributed by the author. Department of Genome Sciences, University of Washington; Seattle: 2005. PHYLIP (Phylogeny Inference Package) version 3.6.

Fishkind DJ, Bonder EM, Begg DA. Isolation and characterization of sea urchin egg spectrin, calcium modulation of the spectrin-actin interaction. Cell Motil Cytoskeleton 1987;7:304–14. [PubMed: 3607894]

Gianguzza F, DiBernardo MG, Sollazzo M, Palla F, Ciaccio M, Carra E, Spinelli G. DNA sequence and pattern of expression of the sea urchin (Paracentrotus lividus) alpha-tubulin genes. Mol Reprod Dev 1989;1:170–81. [PubMed: 2627367]

Gibbons IR, Gibbons BH, Mocz G, Asai DJ. Multiple nucleotide-binding sites in the sequence of dynein beta heavy chain. Nature 1991;352:640–3. [PubMed: 1830927]

Gibbons BH, Asai DJ, Tang WJ, Hays TS, Gibbons IR. Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchins. Mol Biol Cell 1994;5:57–70. [PubMed: 8186465]
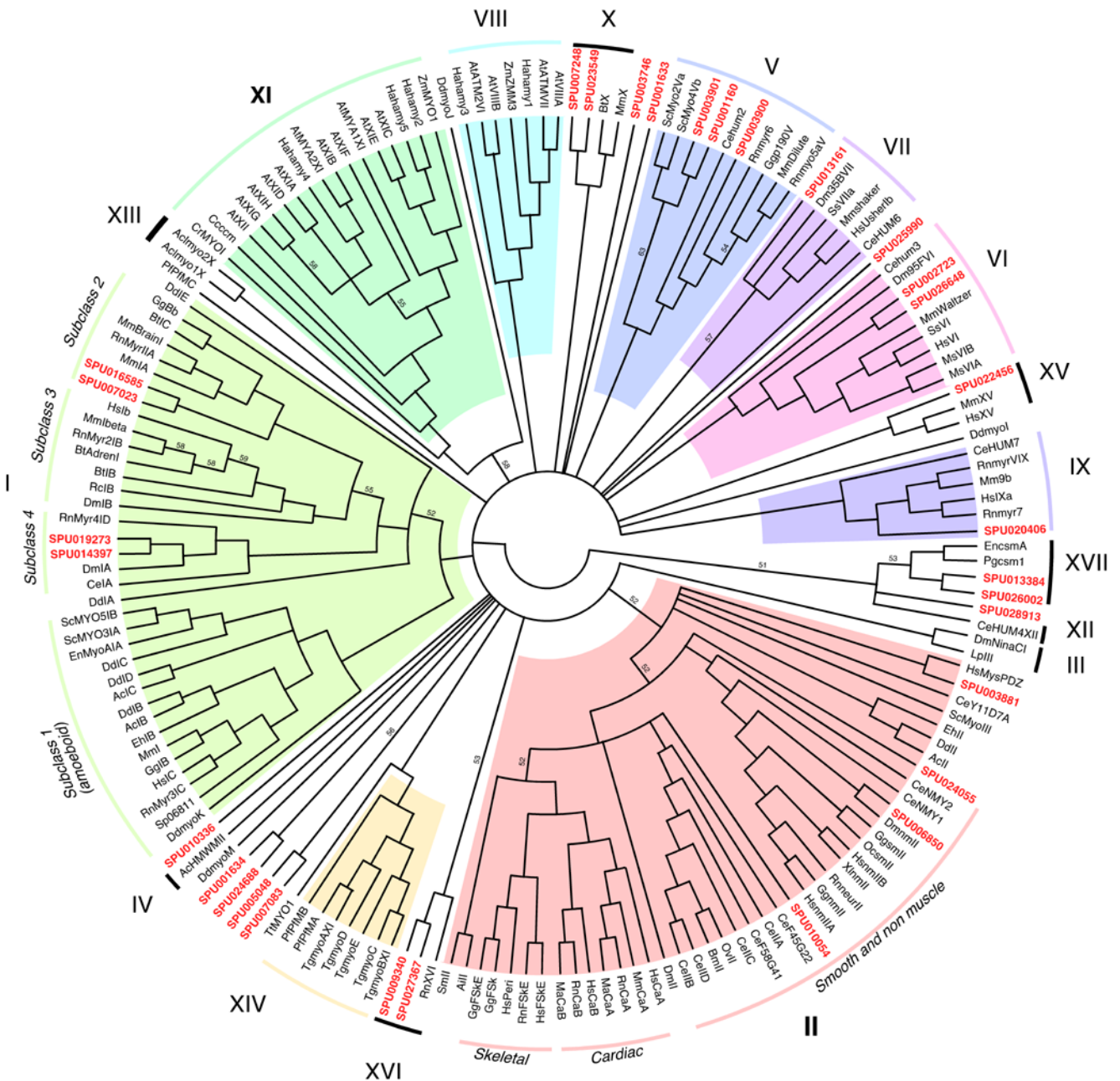
Gibbons IR. Dynein Family of Motor Proteins: Present Status and Future Questions. Cell Motil and Cytoskel 1995;32:136–144.

Gibbons IR, Rowe AJ. Dynein: A Protein with Adenosine Triphospjhatase Activity from Cilia. Science 1965;149:424–426. [PubMed: 17809406]

Gish W, States DJ. Identification of protein coding regions by database similarity search. Nature Genet 1993;3:266–272. [PubMed: 8485583]

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 2003;52:696–704. [PubMed: 14530136]

Hammarback JA, Obar RA, Hughes SM, Vallee RB. MAP1B is encoded as a polyprotein that is processed to form a complex N-terminal microtubule-binding domain. Neuron 1991;7:129–39. [PubMed: 1712602]

Harvey, EB. The American Arbacia and Other Sea Urchins. Princeton University Press; Princeton: 1956.

Hiramoto Y. The mechanics and mechanism of cleavage in the sea-urchin egg. Symp Soc Exp Biol 1968;22:311–27. [PubMed: 4880193]

Hodge, T.; Cope, MJTV. A Myosin Family Tree; Journal of Cell Science. 2000. p. 3353-3354.http://www.mrc-lmb.cam.ac.uk/myosin/trees/txalign.html

Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comp Appl Biosci (CABIOS) 1992;8:275–282.

Kagami O, Gotoh M, Makino Y, Mohri H, Kamiya R, Ogawa K. A dynein light chain of sea urchin sperm flagella is a homolog of mouse Tctex 1, which is encoded by a gene of the t complex sterility locus. Gene 1998;211:383–6. [PubMed: 9602174]

Kane R. The Mitotic Apparatus: Identification of the Major Soluble Component of the Glycerol-Isolated Mitotic Apparatus. J Cell Biol 1967;32:243–253. [PubMed: 10976219]

Kane RE. Components of the actin-based cytoskeleton. Methods Cell Biol 1986;27:229–42. [PubMed: 3517583]

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. [PubMed: 15661851]

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. [PubMed: 12136088]

Kelleher JF, Atkinson SJ, Pollard TD. Sequences, structural models, and cellular localization of the actin-related proteins Arp2 and Arp3 from Acanthamoeba. J Cell Biol 1995;131:385–97. [PubMed: 7593166]

Kent WJ. BLAT - The BLAST-Like Alignment Tool. Genome Res 2002;12:656–664. [PubMed: 11932250]

Krause M, Dent EW, Bear JE, Loureiro JJ, Gertler FB. Ena/VASP proteins, regulators of the actin cytoskeleton and cell migration. Annu Rev Cell Dev Biol 2003;19:541–64. [PubMed: 14570581]

Langkopf A, Hammarback JA, Muller R, Vallee RB, Garner CC. Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messenger RNA. J Biol Chem 1992;267:16561–6. [PubMed: 1379599]

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. [PubMed: 15479732]

Lee JJ, Shott RJ, Rose SJ3rd Thomas TL, Britten RJ, Davidson EH. Sea urchin actin gene subtypes. Gene number, linkage and evolution. J Mol Biol 1984;172:149–76. [PubMed: 6319714]

Li Q, Suprenant KA. Molecular characterization of the 77-kDa echinoderm microtubule-associated protein. Homology to the beta-transducin family. J Biol Chem 1994;269:31777–31784. [PubMed: 7989351]

Mabuchi I. Biochemical Aspects of Cytokinesis. International Review of Cytology 1986;101:175–213. [PubMed: 3516914]

Mabuchi I, Okuno M. The effect of myosin antibody on the division of starfish blastomeres. J Cell Biol 1977;74:251–63. [PubMed: 141455]
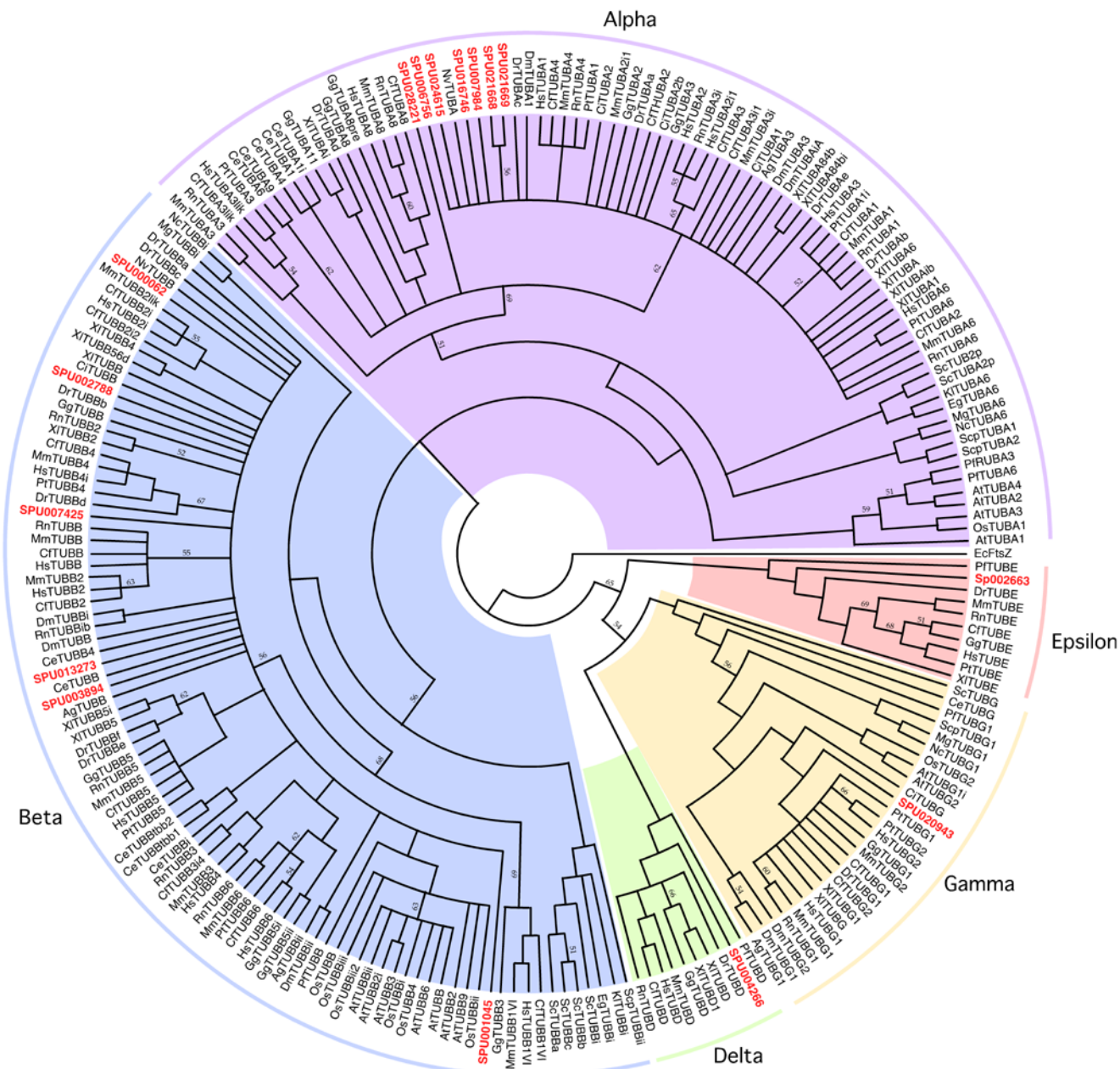
Mabuchi I, Spudich JA. Purification and properties of soluble actin from sea urchin eggs. J Biochem (Tokyo) 1980;87:785–802. [PubMed: 6893042]

Maekawa S, Mishima M, Toriyama M, Sakai H. Purification of a low molecular weight microtubule binding protein from sea urchin eggs. Biochim Biophys Acta 1994;1207:194–200. [PubMed: 7915541]

Maekawa S, Toriyama M, Sakai H. A novel 24-kDa microtubule-associated protein purified from sea urchin eggs. Eur J Biochem 1992;205:1195–200. [PubMed: 1349526]

Masuda M. Species specific pattern of ciliogenesis in developing sea urchin embryos. Develop Growth Differ 1979;21:545–552.

Miki H, Okada Y, Hirokawa N. Analysis of the kinesin superfamily: insights into structure and function. Trends Cell Biol 2005;15:467–76. [PubMed: 16084724]

Milisav I. Dynein and dynein-related genes. Cell Motil Cytoskeleton 1998;39:261–72. [PubMed: 9556328]

Mocz G, Helms MK, Jameson DM, Gibbons IR. Probing the nucleotide binding sites of axonemal dynein with the fluorescent nucleotide analogue 2′3′-O--N-Methylanthraniloyl-adenosine 5′-triphosphate. Biochemistry 1998;47:9862–9. [PubMed: 9657700]

Morris RL, Scholey JM. Heterotrimeric kinesin-II is required for the assembly of motile 9+2 ciliary axonemes on sea urchin embryos. Journal of Cell Biology 1997;138:1009–22. [PubMed: 9281580]

Nei, M.; Kumar, S. Molecular Evolution and Phylogenetics. Oxford University Press; NY: 2000. p. 333

Nei M, Xu P, Glazko G. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. PNAS 2001;98:2497–2502. [PubMed: 11226267]

Ogawa K. Four ATP-binding sites in the midregion of the beta heavy chain of dynein. Nature 1991;352:643–5. [PubMed: 1830928]

Ogawa K, Takai H, Ogiwara A, Yokota E, Shimizu T, Inaba K, Mohri H. Is outer arm dynein intermediate chain 1 multifunctional? Mol Biol Cell 1996;7:1895–1907. [PubMed: 8970153]

Patel KG, Liu C, Cameron PL, Cameron RS. Myr 8, a novel unconventional myosin expressed during brain development associates with the protein phosphatase catalytic subunits 1alpha and 1gamma1. J Neurosci 2001;21:7954–68. [PubMed: 11588169]

Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. J Cell Biol 2005;170:103–13. [PubMed: 15998802]

Pazour GJ, Agrin N, Walker BL, Witman GB. Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. J Med Genet 2006;43:62–73. [PubMed: 15937072]

Perumal BS, Sakharkar KR, Chow VT, Pandjassarame K, Sakharkar MK. Intron position conservation across eukaryotic lineages in tubulin genes. Front Biosci 2005;10:2412–2419. [PubMed: 15970504]

Pfister KK, Fisher EM, Gibbons IR, Hays TS, Holzbaur EL, McIntosh JR, Porter ME, Schroer TA, Vaughan KT, Witman GB, King SM, Vallee RB. Cytoplasmic dynein nomenclature. J Cell Biol 2005;171:411–413. [PubMed: 16260502]

Poustka AJ, Groth D, Hennig S, Thamm S, Cameron A, Beck A, Reinhardt R, Herwig R, Panopoulou G, Lehrach H. Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters. Genome Res 2003;13:2736–46. [PubMed: 14656975]

Rashid DJ, Wedaman KP, Scholey JM. Heterodimerization of the two motor subunits of the heterotrimeric kinesin, KRP85/95. J Mol Biol 1995;252:157–62. [PubMed: 7674298]

Rappaport, R. Cytokinesis in Animal Cells. Cambridge University Press; Cambridge: 1996.

Rasmusson K, Serr M, Gepner J, Gibbons I Hays TS. A family of dynein genes in Drosophila melanogaster. Mol Biol Cell 1994;5:45–55. [PubMed: 8186464]

Rogers GC, Chui KK, Lee EW, Wedaman KP, Sharp DJ, Holland G, Morris RL, Scholey JM. A kinesin-related protein, KRP(180), positions prometaphase spindle poles during early sea urchin embryonic cell division. Journal of Cell Biology 2000;150:499–512. [PubMed: 10931863]

Rogers GC, Hart CL, Wedaman KP, Scholey JM. Identification of kinesin-C, a calmodulin-binding carboxy-terminal kinesin in animal (Strongylocentrotus purpuratus) cells. J Mol Biol 1999;294:1–8. [PubMed: 10556023]
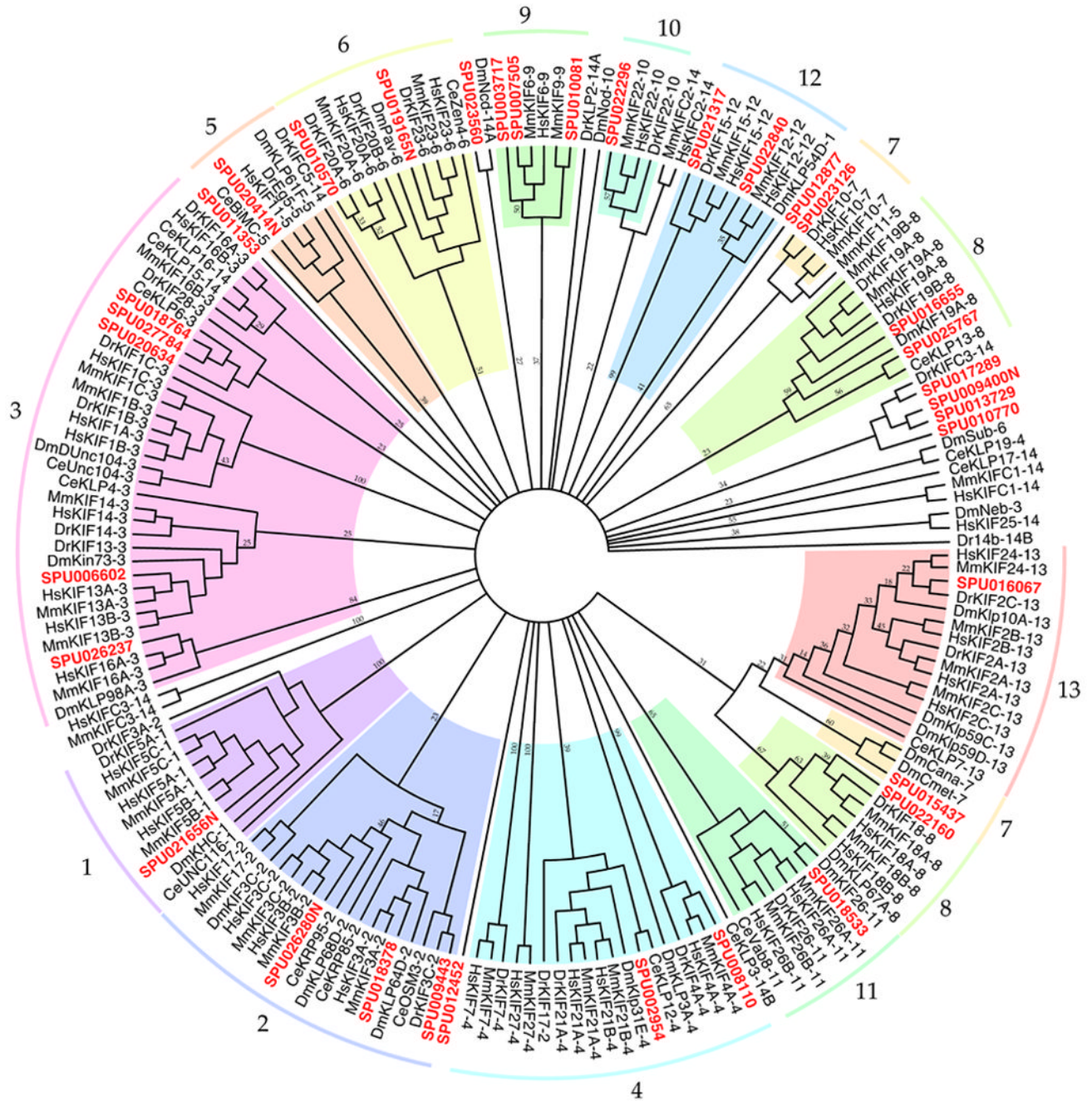
Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572–1574. [PubMed: 12912839]

Saitou N, Nei M. The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–425. [PubMed: 3447015]

Salmon ED. Spindle microtubules: thermodynamics of in vivo assembly and role in chromosome movement. Ann N Y Acad Sci 1975;253:383–406. [PubMed: 1096721]

Scholey JM, Neighbors B, McIntosh JR, Salmon ED. Isolation of microtubules and a dynein-like MgATPase from unfertilized sea urchin eggs. J Biol Chem 1984;259:6516–6525. [PubMed: 6144678]

Scholey JM, Porter ME, Grissom PM, McIntosh JR. Identification of kinesin in sea urchin eggs, and evidence for its localization in the mitotic spindle. Nature 1985;318:483–6. [PubMed: 2933590]

Schroeder TE. Cytokinesis: filaments in the cleavage furrow. Exp Cell Res 1968;53:272–276. [PubMed: 4387145]

Sharp DJ, Rogers GC, Scholey JM. Microtubule motors in mitosis. Nature 2000;407:41–7. [PubMed: 10993066]

Shen X, Ranallo R, Choi E, Wu C. Involvement of actin-related proteins in ATP-dependent chromatin remodeling. Mol Cell 2003;12:147–55. [PubMed: 12887900]

Sirotkin V, Seipel S, Krendel M, Bonder EM. Characterization of sea urchin unconventional myosins and analysis of their patterns of expression during early embryogenesis. Mol Reprod Dev Oct 2000;57:111–26.

Sluder G. Role of spindle microtubules in the control of cell cycle timing. J Cell Biol 1979;80:674–91. [PubMed: 572367]

Smith LC, Britten RJ, Davidson EH. SpCoel1: a sea urchin profilin gene expressed specifically in coelomocytes in response to injury. Mol Biol Cell 1992;3:403–14. [PubMed: 1498361]

Stephens RE. Ciliogenesis in sea urchin embryos--a subroutine in the program of development. Bioessays 1995;17:331–40. [PubMed: 7741725]

Strickland LI, Wen Y, Gundersen GG, Burgess DR. Interaction between EB1 and p150glued is required for anaphase astral microtubule elongation and stimulation of cytokinesis. Curr Biol 2005;15(24): 2249–55. [PubMed: 16360686]

Suprenant KA, Tuxhorn JA, Daggett MA, Ahrens DP, Hostetler A, Palange JM, VanWinkle CE, Livingston BT. Conservation of the WD-repeat, microtubule-binding protein, EMAP, in sea urchins, humans, and the nematode C. elegans. Dev Genes Evol 2000;210:2–10. [PubMed: 10603080]

Terasaki AG, Ohnuma M, Mabuchi I. Identification of actin-binding proteins from sea urchin eggs by F-actin affinity column chromatography. J Biochem (Tokyo) 1997;122:226–36. [PubMed: 9276693]

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–82. [PubMed: 9396791]

Tilney LG, Gibbins JR. Microtubules and filaments in the filopodia of the secondary mesenchyme cells of Arbacia punctulata and Echinarachnius parma. J Cell Sci 1969;5:195–210. [PubMed: 4909565]

Tilney LG, Hatano S, Ishikawa H, Mooseker MS. The polymerization of actin, its role in the generation of the acrosomal process of certain echinoderm sperm. J Cell Biol 1973;59:109–26. [PubMed: 4356568]

Togel M, Eichinger R, Wiche G, Propst F. A 45 amino acid residue domain necessary and sufficient for proteolytic cleavage of the MAP1B polyprotein precursor. FEBS Lett 1999;451:15–8. [PubMed: 10356975]

Vale RD. The molecular motor toolbox for intracellular transport. Cell 2003;112:467–80. [PubMed: 12600311]

Vallee RB, Bloom GS. Isolation of sea urchin egg microtubules with taxol and identification of mitotic spindle microtubule-associated proteins with monoclonal antibodies. Proc Natl Acad Sci U S A 1983;80:6259–63. [PubMed: 6137821]

Vaughan KT, Mikami A, Paschal BM, Holzbaur EL, Hughes SM, Echeverri CJ, Moore KJ, Gilbert DJ, Copeland NG, Jenkins NA, Vallee RB. Multiple mouse chromosomal loci for dynein-based motility. Genomics 1996;36(1):29–38. [PubMed: 8812413]

Wain HM, Lush MJ, Ducluzeau F, Khodiyar VK, Povey S. Genew: the Human Gene Nomenclature Database, 2004 updates. Nucleic Acids Res 2004;32(Database issue):D255–7. [PubMed: 14681406]

Wang CC, Chan DC, Leder P. The mouse formin (Fmn) gene: genomic structure, novel exons, and genetic mapping. Genomics 1997;39:303–11. [PubMed: 9119367]

Wedaman KP, Knight AE, Kendrick-Jones J, Scholey JM. Sequences of sea urchin kinesin light chain isoforms. J Mol Biol 1993;231:155–158. [PubMed: 8496962]

Wedaman KP, Meyer DW, Rashid DJ, Cole DG, Scholey JM. Sequence and submolecular localization of the 115-kD accessory subunit of the heterotrimeric kinesin-II (KRP85/95) complex. Journal of Cell Biology 1996;132:371–80. [PubMed: 8636215]

Whittaker CA, Bergeron K-F, Whittle J, Brandhorst BP, Burke RD, Hynes RO. The echinoderm adhesome. Developmental Biology. (this issue)

Wickstead B, Gull K. A "holistic" kinesin phylogeny reveals new kinesin families and predicts protein functions. Mol Biol Cell 2006;17:1734–43. [PubMed: 16481395]

Wright BD, Henson JH, Wedaman KP, Willy PJ, Morand JN, Scholey JM. Subcellular localization and sequence of sea urchin kinesin heavy chain: evidence for its association with membranes in the mitotic apparatus and interphase cytoplasm. J Cell Biol 1991;113:817–833. [PubMed: 1827446]

Wright BD, Terasaki M, Scholey JM. Roles of kinesin and kinesin-like proteins in sea urchin embryonic cell division: evaluation using antibody microinjection. J Cell Biol 1993;123:681–9. [PubMed: 8227132]

Yokota E, Mabuchi I, Sato H. Activation of sea urchin sperm flagellar dynein ATPase activity by salt-extracted axonemes. J Biochem (Tokyo) 1987;102:31–41. [PubMed: 2959658]

Zhu X, Mahairas G, Illies M, Cameron RA, Davidson EH, Ettensohn CA. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. Development 2001;128:2615–27. [PubMed: 11493577]

VIII   X

XI

V

VII

XIII

VI

Subclass 2

XV

Subclass 3

I

Subclass 4

IX

XVII

Subclass 1 (amoeboid)

XII

III

IV

Smooth and non muscle

XIV

II

XVI

Skeletal   Cardiac
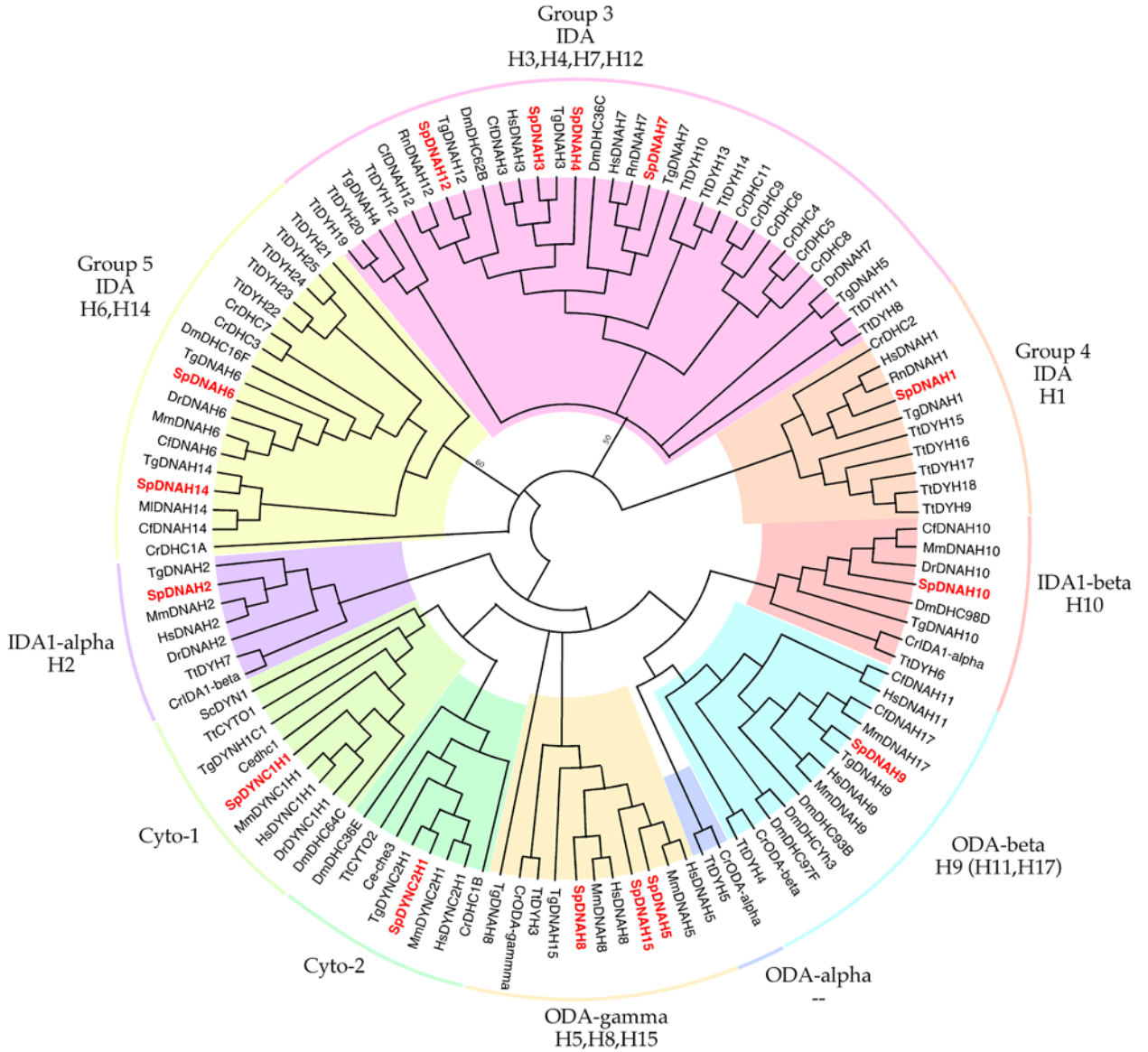
Figure 1.
Phylogenetic placement of predicted *S. purpuratus* myosin, tubulin, kinesin, and dynein sequences. (A) Myosin gene family. (B) Tubulin gene superfamily. (C) Kinesin gene superfamily. (D) Dynein gene family. All four trees are majority rule bootstrap consensus trees inferred as described in the text using full-length gene amino acid sequences wherever available. Only clades with >50% bootstrap support are shown for the myosin, tubulin, and dynein consensus trees. For the kinesin data, because we used the full-length gene sequences, the complexity of the gene product, and the method used to generate the multiple sequence alignment, the resulting 50% majority rule bootstrap consensus tree was uninformative. Therefore the bootstrap consensus tree for this tree was generated using the "extended majority rule" criteria. All clades with a bootstrap value of < 20% were collapsed. In all four trees, branches have a bootstrap support of >70% unless indicated otherwise. A list of all sequences used in the four analyses as well as the multiple sequence alignments are available from the authors (SSM). Gene family designations and abbreviated gene names follow Hodge and Cope

(2000) for the myosins, Dutcher (2003) for the tubulins, and Miki *et al*. (2005) for the kinesins. Nomenclature for dyneins follows FlyBase for *Drosophila*, Pazour et al.(2006) for *Chlamydomonas*, Asai and Wilkes (2004) for *Tetrahymena*, and the mammalian system (Wain et al., 2004) for other organisms. . The predicted *S. purpuratus* sequences are in bold and colored red. Genus species abbreviations are as follows. Ac: *Acanthamoeba castellanii*, Ag: *Anopheles gambiae*, Ai: *Aequipecten irradians*, At: *Arabidopsis thaliana*, Bm: *Brugia malayi*, Bt: *Bos taurus*, Cc: *Cara coralline*, Ce: *Caenorhabditis elegans*, Cf: *Canis familiaris*, Ci: *Ciona intestinalis*, Cr: *Chlamydomonas reinhardtii*, Dd: *Dictyostelium discoideum*, Dm: *Drosophila melanogaster*, Dr: *Danio rario*, Ec: *E. coli*, Eg: *Eremothecium gossypii*, Eh: *Entamoeba histolytica*, En: *Emiricella nidulans*, Gg: *Gallus gallus*, Ha: *Helianthus annus*, Hs: *Homo sapiens*, Kl: *Kluyveromyces lactis*, Lp: *Limulus polyphemus*, Ma: *Mesocricetus auratus*, Mg: *Magnaporthe grisea*, Ml: *Macaca mulatta*, Mm: *Mus musculus*, Ms: *Morone saxatilis*, Nc: *Neurospora crassa*, Nv: *Nematostella ventensis*, Os: *Oryza sativa* (japonica cultivar-group), Ov: *Onchocerca volvulus*, Pf: *Plasmodium falciparum*, Pg: *Pyricularia grisea*, Rc: *Rana catesbeiana*, Rn: *Rattus norvegicus*, Sc: *Saccharomyces cerevisiae*, Scp: *Schizosaccharomyces pombe*, Sm: *Schistosoma mansoni*, Sp: *Strongylocentrotus purpuratus*, Ss: *Sus scrofa*, Tg: *Toxoplasma gondii*, Tt: *Tetrahymena thermophila*, Xl: *Xenopus laevis*, Zm: *Zea mays*.
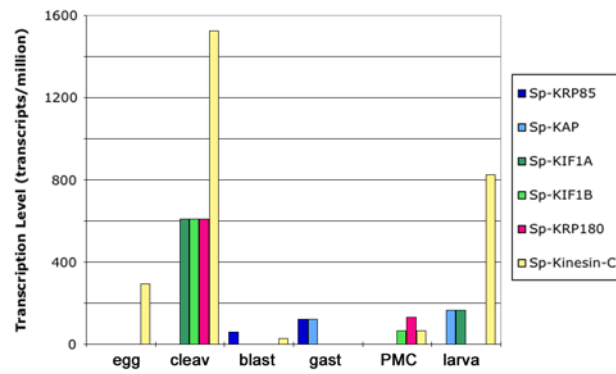
Figure 2.
Expression profiles of kinesins and associated proteins during *S. purpuratus* development. Transcript level as predicted by relative abundance of kinesin-specific ESTs is shown for a kinesin-2 (Sp-KRP85 and its associated protein Sp-KAP), two kinesin-3s (Sp-KIF1A and Sp-KIF1B), a kinesin-12 (KRP180), and a kinesin-14 (Sp-Kinesin-C) in eggs, cleavage stage (cleav), blastula stage (blast), gastrula stage (gast), specifically in primary mesenchyme cells (PMCs) and larvae. Expression levels appeared to peak during cleavage stage for several of the kinesins.
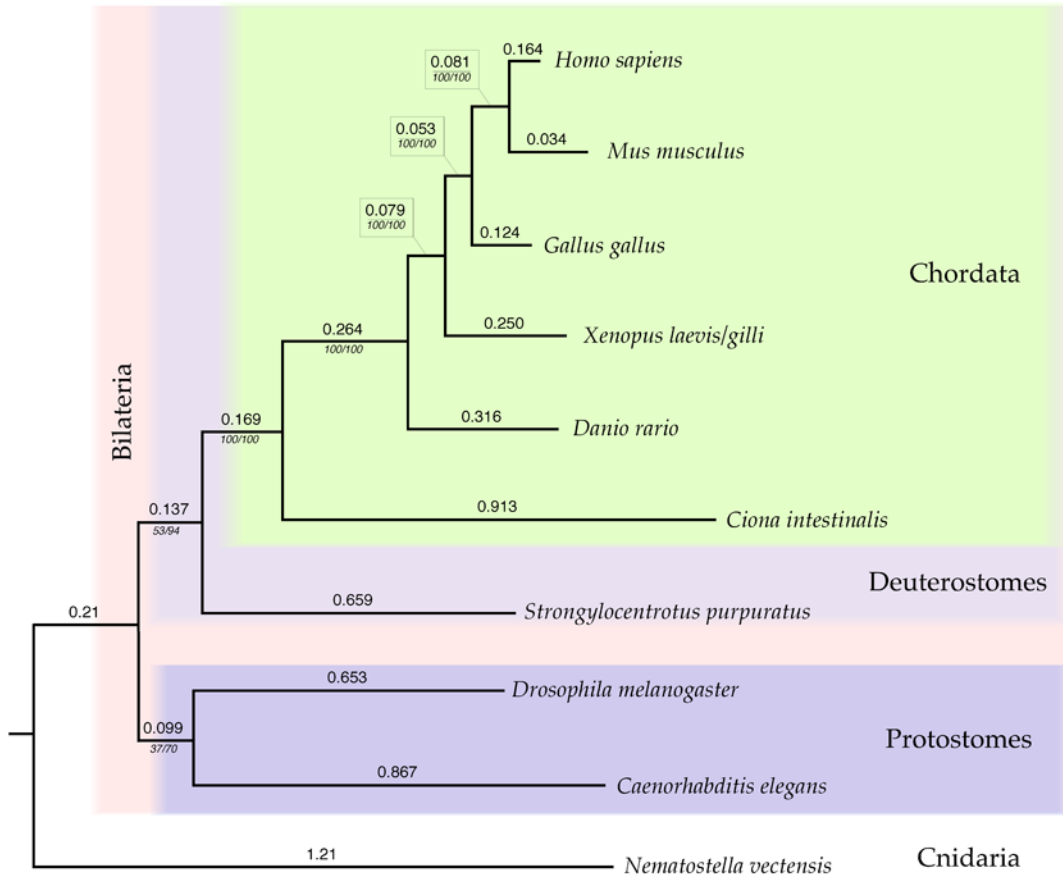
Figure 3.
Phylogenetic relationship among the ten species of Eukaryotes based on CLIP170, fascin, gelsolin, lamin, myosin II, and profilin amino acid sequences. Values above each branch are the branch lengths based on a maximum likelihood analysis as described in the text. The values below each branch are the non-parametric bootstrap confidence values from a maximum likelihood analysis and the prior probabilities from a Bayesian analysis. All branches are shown regardless of the bootstrap or Bayesian support values. The basal portion of the tree showing the relationship among *D. melanogaster*, *S. purpuratus*, and *C. elegans* is relatively unresolved.

Table 1
Cytoskeletal, cytoskeletal-interacting, and motor proteins in the sea urchin genome assembly

"Gene Name" is the name we assigned to the gene in the annotation database. "Synonyms" include all names used to identify this gene or its homologs in any organism. "SPU Gene ID" is the unique and permanent identifier for this gene and it is based on identifiers originally assigned by the GLEAN3 gene identification process. Because the dynein heavy chain genes involve exons located on up to 5 genomic scaffolds per gene, they are in the annotation database as "novel genes" separate from the GLEAN system. "Accession Number of best hit" indicates an NCBI entry that will be most informative to a reader interested in studying this protein further. "BLAST score" is the BLAST score of the final GLEAN sequence deposited in the annotation database, vs. the sequence of the database entry corresponding to the Accession Number cited in the previous column. Gene models (i.e. SPUs) marked with an asterisk (*) were predicted to encode conserved Sp-myosin- motor domains (MYSc) by SMART/PFAM prior to bootstrap analysis, and BLAST scores were not derived independently for these sequences. †Smith-Waterman score.

Gene Name	Synonyms	SPU Gene ID	Accession Number of best hit	BLAST score
ACTIN				
Sp-cytoskeletal-actin-I	CyI	SPU_009481	1703128	0.0
Sp-cytoskeletal-actin-IIa	CyIIa	SPU_009482	47551037	0.0
Sp-cytoskeletal-actin-IIb	CyIIb	SPU_009483	47550921	0.0
Sp-cytoskeletal-actin-IIIb	CyIIIb	SPU_009433	47551035	0.0
Sp-muscle-actin	Muscle actin	SPU_006797	1703137	0.0
ACTIN-RELATED				
Sp-Arp1	Arp1/centractin	SPU_001366	18606465	0.0
Sp-Arp3	Arp3	SPU_010631	6681670	0.0
Sp-Arp6	Arp6	SPU_007479	72044622	1E-125
Sp-Arp8	Arp8	SPU_027580	39812115	8E-102
MYOSINS				
Sp-myosin-I-subclass-2		SPU_007023	*	*
Sp-myosin-I-subclass-4		SPU_019273	*	*
--- " ---		SPU_014397	*	*
Sp-myosin-II-smooth-and-non-muscle		SPU_024055	*	*
--- " ---		SPU_006850	*	*
--- " ---		SPU_010054	*	*
Sp-myosin-II-smooth-and-non-muscle-light-chain		SPU_001613	67083885	2E-39
Sp-Sp-myosin-III		SPU_009340	35959282	4E-97
Sp-myosin-IV		SPU_005048	*	*
--- " ---		SPU_007083	*	*
Sp-myosin-V		SPU_003901	*	*
--- " ---		SPU_001160	*	*
--- " ---		SPU_003900	*	*
Sp-myosin-VI		SPU_002723	8099610	0.0
--- " ---		SPU_002723	*	*
--- " ---		SPU_026648	*	*
Sp-myosin-VII		SPU_025990	9944237	2E-72
--- " ---		SPU_013161	9944237	0.0
--- " ---		SPU_022040	9944237	0.0
Sp-myosin-IX		SPU_020406	*	*
Sp-myosin-X		SPU_007248	50428778	0.0
--- " ---		SPU_023549	*	*
--- " ---		SPU_003746	*	*
--- " ---		SPU_001633	*	*
Sp-myosin-XV		SPU_022456	22547229	4E-177
Sp-myosin-XVI		SPU_009340	*	*
--- " ---		SPU_027367	*	*
Sp-myosin-XVII		SPU_013384	*	*
--- " ---		SPU_026002	*	*
ACTIN-BINDING/ACTIN DYNAMICS				
Sp-filamin	ABP278/filamin B	SPU_017634	32129530	0.0
Sp-ABP620	ABP620	SPU_003260	5821434	0.0
Sp-alpha-actinin	Alpha-actinin	SPU_020920	1070611	0.0
Sp-adducin/drebrin	Adducin/Drebrin	SPU_026521	28382	2E-92
Sp-calponin-1	Calponin	SPU_015595	23491586	2E-33
Sp-capz	CapZ	SPU_024912	38322686	9E-91
Sp-cofilin-1	Cofilin 1	SPU_006172	2440185	6E-07
Sp-Contactin-2	Contactin 2	SPU_021328	6981632	2E-88
Sp-coronin	Coronin	SPU_000974	11463876	0.0
Sp-drebrin	Drebrin	SPU_026521	20454881	2E-07
Sp-dystrophin-1	Dystrophin	SPU_021580	13377398	0.0
Sp-dystrophin-2	Dystrophin	SPU_021581	13377398	0.0
Sp-EWAM	EWAM	SPU_027390	551452	2E-106

Gene Name	Synonyms	SPU Gene ID	Accession Number of best hit	BLAST score
Sp-fascin	Fascin	SPU_010943	47551048	0.0
Sp-gelsolin	Gelsolin	SPU_000296	28916693	1E-20
Sp-moesin	Moesin	SPU_018705	719272	0.0
Sp-paxillin	Paxillin	SPU_008331	43415525	2E-125
Sp-profilin	Profilin	SPU_020197	400849	0.0
Sp-radixin	Radixin	SPU_007679	40804379	7E-35
Sp-scruin	Scruin	SPU_019550	633238	1E-22
Sp-alpha-spectrin	Spectrin, alpha	SPU_020836	158489	6E-174
Sp-beta-spectrin	Spectrin, beta	SPU_007720	56206997	5E-61
Sp-supervillin	Supervillin/Archvillin	SPU_007557	55957976	7E-149
Sp-talin-1	Talin	SPU_000241	45383127	0.0
Sp-tropomodulin	Tropomodulin	SPU_027778	13928838	1E-50
Sp-tropomyosin	TropoSp-myosin-	SPU_000128	38569885	5E-106
Sp-troponin-C	Troponin C	SPU_016371	72012357	1E-79
Sp-troponin-I	Troponin I	SPU_013183	18157232	6E-18
Sp-troponin-T	Troponin T	SPU_006532	48255883	2E-06
Sp-villin	Villin	SPU_000296	49257782	9E-128
Sp-vinculin	Vinculin	SPU_028781	309553	8E-12
Sp-Wiskott-Aldrich-syndrome-protein	WASP	SPU_003194	4633273	1E-17
CYTOSKELETAL REGULATORS/SIGNALING				
Sp-LIM-kinase	LIM Kinase	SPU_022207	1657756	1E-93
Sp-Calponin	Calponin/Myophillin	SPU_015595	23491586	4E-33
Sp-Cdc42	Cdc42	SPU_019494	62858789	1E-98
Sp-citron	Citron	SPU_011918	56405460	0.0
Sp-enabled	ENA (Enabled)	SPU_010763	58865872	6E-09
Sp-formin-2	Formin 2	SPU_020333	55961780	1E-36
Sp-formin-2-DAD-domain	DAD domain	SPU_020879	3834629	78†
Sp-WISH	N-WASP binding protein	SPU_020040	49188650	1E-98
Sp-Mob1	Mob1	SPU_024840	39645694	2E-106
Sp-PAK-(p21)	PAK (p21)	SPU_003534	29791476	7E-31
Sp-PAR3-1	PAR3	SPU_014037	45387877	2E-85
Sp-PAR3-2	PAR3	SPU_005813	46405829	3E-45
Sp-paxillin-1	Paxillin	SPU_008331	42415525	5E-125
Sp-profilin	profilin	SPU_020197	47551153	4E-42
Sp-RhoA	Rho1/RhoA	SPU_011032	72038732	2E-94
Sp-ROCK-1	ROCK	SPU_026779	34784414	6E-106
TUBULINS				
Sp-alpha-tubulin-1		SPU_021668	NP_006000	0.0
Sp-alpha-tubulin-2		SPU_021669	NP_006000	0.0
Sp-alpha-tubulin-3		SPU_016746	NP_006000	0.0
Sp-alpha-tubulin-4		SPU_028221	NP_006000	0.0
Sp-alpha-tubulin-5		SPU_019990	NP_006000	0.0
Sp-alpha-tubulin-6		SPU_024615	NP_006000	0.0
Sp-alpha-tubulin-7		SPU_012679	NP_006000	0.0
Sp-alpha-tubulin-8		SPU_007984	NP_006000	0.0
Sp-alpha-tubulin-9		SPU_006756	NP_006000	0.0
Sp-beta-tubulin-1		SPU_002788	NP_006079	0.0
Sp-beta-tubulin-2		SPU_001045	NP_006079	0.0
Sp-beta-tubulin-3		SPU_000062	NP_006079	0.0
Sp-beta-tubulin-4		SPU_013273	NP_006079	0.0
Sp-beta-tubulin-5		SPU_003894	NP_006079	1e-92
Sp-beta-tubulin-6		SPU_007425	NP_006079	0.0
Sp-gamma-tubulin		SPU_020943	NP_001061.2	0.0
Sp-delta-tubulin		SPU_004266	AAF09584.1	1e-108
Sp-epsilon-tubulin		SPU002663 & SPU_013900	NP_001016136.1	1e-172
MICROTUBULE-ASSOCIATED PROTEINS				
Sp-EMAP	77kDaMAP, 77kDa-Microtubule-associated Protein	SPU_006069 & SPU_003443	Q26613	0.0
Sp-EMAP-like-1	77kDaMAP-like, 77kDa-Microtubule-associated Protein	SPU_000849	Q26613	1E-138
Sp-EMAP-like-2	77kDaMAP-like, 77kDa-Microtubule-associated Protein	SPU_026640	Q26613	1E-95
Sp-EMAP-like-3	77kDaMAP-like, 77kDa-Microtubule-associated Protein	SPU_001991	Q26613	2E-94
Sp-EMAP-like-4	77kDaMAP-like, 77kDa-Microtubule-associated Protein	SPU_020821	Q26613	1E-76
Sp-EMAP-like-5	77kDaMAP-like, 77kDa-Microtubule-associated Protein	SPU_004328	Q26613	8E-73
Sp-Tektin-1-1		SPU_008777	NP_444515.1	4E-90
Sp-Tektin-1-2		SPU_019591	NP_444515.1	4E-90
Sp-Tektin-1-3		SPU_013841	NP_444515.1	1E-109
Sp-Tektin-1-4		SPU_006453	NP_444515.1	1E-109
Sp-Tektin-2		SPU_020728	NP_055281.2	1E-101
Sp-Tektin-3		SPU_023618	NP_114104.1	1E-141
Sp-HS-MAP	MAP2-like, MAP4-like, MAP215-like, Tau-like	SPU_011254	NP_002366.2	2E-32
Sp-MAP1B-like	MAP1B-like, MAP1A-like	SPU_004628	NP_005900.1	3E-81
Sp-MAP1A/1B_LC3-like-1	Microtubule-associated proteins 1A/1B light chain 3B precursor,	SPU_009444	NP_115903.1	3E-44

Gene Name	Synonyms	SPU Gene ID	Accession Number of best hit	BLAST score
Sp-MAP1A/1B_LC3-like-2	MAP1A/MAP1B LC3, MAP1A/1B light chain 3 Microtubule-associated proteins 1A/1B light chain 3B precursor, MAP1A/MAP1B LC3, MAP1A/1B light chain 3	SPU_008954	NP_115903.1	1E-36
Sp-RIB43A	flagellar protofilament ribbon protein	SPU_027579	NP_001027762.1	1E-122
Sp-TPX2-like-1		SPU_008221	NP_036244.2	2E-22
Sp-TPX2-like-2		SPU_022897	NP_036244.2	2E-22
Sp-WD-Repeat-domain-56		SPU_012598	XP_545256.2	5E-74
Sp-Adenomatous Polyposis Coli	APC	SPU_026745	6978509	5E-175
Sp-bicaudal-D	Bicaudal	SPU_017959	51093830	2E-130
Sp-diaphanous-1/3	Diaphanous 1/3	SPU_012710	55959274	3E-35
Sp-diaphanous-1	Diaphanous 1	SPU_020879	62088544	7E-80
Sp-EB1	EB1	SPU_027631	73980297	1E-25
Sp-gephyrin	Gephyrin	SPU_017877	16605466	1E-112
Sp-katanin-p60-subunit-A	Katanin	SPU_001000	3098603	0.0
Sp-MAST/orbit/CLASP	MAST/orbit/CLASP	SPU_000096	66271020	3E-75
Sp-CLIP170	CLIP170/Restin	SPU_001154	23821025	4E-61
Sp-dynactin-1	Dynactin1/p150-glued	SPU_013935	33872165	0.0
Sp-dynactin-2/dynactin-subunit-5	Dynactin 2/Dynactin subunit 5	SPU_028205	50926127/14789952	1E-74/1E-84
Sp-dynamitin-p50	Dynamitin	SPU_018500	28274444	1E-63
INTERMEDIATE FILAMENTS				
Sp-keratin-complex-1-acidic	Acidic keratin	SPU_011620	72105012	0.0
Sp-enabled	ENA (enabled)	SPU_007845	6753754	1E -33
Sp-Keratin-associated-protein	Keratin associated	SPU_019996	77157625	1E -33
Sp-nestin	Nestin	SPU_024601	72108788	1E -155
Sp-Lamin-B	Lamin B	SPU_004509	11386011	2E -66

Table 2

Kinesins in the sea urchin genome assembly

"Kinesin family" is the family, as per Lawrence et al., (2004), to which the motor domain of this gene is most closely similar. Other column headings are as in Table 1. Genus species name abbreviations are as described in Legend to Table 1.

Kinesin Family	Gene Name	Synonyms or Closest Homologs	SPU Gene ID	Accession Number of best hit	BLAST score
MOTOR POLYPEPTIDES					
kinesin-1	Sp-KHC	conventional Kinesin Heavy Chain, kinesin family member 5B [Hs]	SPU_021656	CAA40175	0.0
kinesin-2	Sp-KRP85	kinesin II, 85 kDa [Sp]; kinesin family member 3A [Hs]	SPU_018378	NP_999777	0.0
kinesin-2	Sp-KRP95		SPU_026280	NP_999817	0.0
kinesin-2	Sp-Osm-3	Kinesin-related protein 95 kD [Sp]; kinesin family member 3B [Hs]; Kinesin-like protein FLA10 (KHP1 protein) [Cr] kinesin family member 17 [Hs]	SPU_009443	NP_065867	5e-52
kinesin-3	Sp_KIF13A2	KINESIN-13A2 [Hs]	SPU_006602	CAC20443	0.0
kinesin-3	Sp-KIF1B-like1	kinesin family member 1B [Hs],	SPU_023560	BAE02544	8e-106
kinesin-3	Sp-KIF16-like	kinesin family member 1Bbeta isoform 1 [Hs]	SPU_026237	CAI43180	0.0
kinesin-3	Sp-KLP6-like1	SNX23_HUMAN, Kinesin-like	SPU_027784	CAI95220	2e-128
kinesin-3	Sp-KLP6-like2	motor protein C20orf23 [Hs],	SPU_018764	NP_904325	8e-109
kinesin-3	Sp-KLP6-like3	Sorting nexin 23 kinesin family member 1B [Hs] kinesin family member 1B isoform alpha [Hs] "axonal transport of synaptic vesicles" [Hs]	SPU_020634	NP_004312	3e-118
kinesin-4	Sp-KIF4	KIF4, chromokinesin [Hs]	SPU_008110	NP_036442	0.0
kinesin-4	Sp-KIF21A	kinesin family member 21A [Hs]	SPU_002954	NP_060111	8e-123
kinesin-4	Sp-KIF27	kinesin-related protein KIF27 [Mm]	SPU_012877	NP_780423	0.0
kinesin-5	KRP170	BimC, Eg5, kinesin family member 15 [Hs]	SPU_020414	AAG18583	0.0
kinesin-6	KRP110	KRP110 [Sp]	SPU_019165	AAG18582	0.0
kinesin-6	Sp-KIF20A	kinesin family member 20A [Hs]	SPU_010570	NP_005724	2e-91
kinesin-7	Sp-CENPE	CENPE_HUMAN, Centromeric	SPU_023126	NP_001804	2e-101
kinesin-7	Sp-KIF2-like	Protein E, CENPE-E protein, KIF10 kinesin heavy chain member 2 [Hs]	SPU_015437	NP_004511	6e-168
kinesin-8	Sp-KIF18	kinesin family member 18A [Hs]	SPU_022160	NP_112494	3e-120
kinesin-8	Sp-KIF19	kinesin family member 19 [Hs]	SPU_016655	AAI10990	1e-111
kinesin-8	Kinesin-8-like	kinesin family member 19 [Hs]	SPU_025767	AAI10990	2e-30
kinesin-9	Sp-KIF6	Kinesin Family Member 6 [Hs]	SPU_007505	NP_659464	0.0
kinesin-9	Sp-KIF9	Kinesin Family 9 isoform 3 [Hs]	SPU_010081	NP_878906	1e-133
kinesin-9	Sp-KIF6-like	Kinesin Family Member 6 [Hs]	SPU_003717	NP_659464	5e-72
kinesin-10	Sp-KIF22	kinesin family member 22 [Hs]	SPU_022296	AAH28155	7e-115
kinesin-11	Sp-KIF26	KIF26B protein [Hs]	SPU_018533	AAH35896	8e-147
kinesin-12	KRP180	kinesin family member 15 [Hs]	SPU_021317	NP_999656	0.0
kinesin-12	Sp-KIF12	Kinesin-like protein at 54D CG15844-PA [Dm], kinesin family member 12 [Hs]	SPU_022840	NP_524883	3e-100
kinesin-13	Sp-KIF24	kinesin family member 2C [Hs]	SPU_016067	CAI12999	3e-84
kinesin-14	Sp-Kinesin-C	kinesin-C [Sp] cloned sequence	SPU_009400	AAF04841	0.0
kinesin-14	Sp-KIFC1	kinesin family member C1 [Hs]	SPU_017289	AAH73878	2e-45
kinesin-14	Sp-KIFC3-like1	kinesin family member C3 [Hs]	SPU_013729	AAH01211	7e-77
kinesin-14	Sp-KIFC3-like2	kinesin family member C3 [Hs]	SPU_010770	AAH01211	1e-75
orphan	Sp-KIF25-like	kinesin family member 25	SPU_012452	NP_085118	1e-49
orphan	Sp-KIF27-like	isoform 1 [Hs] kinesin family member 13A [Hs]	SPU_011353	NP_071396	3e-61
KINESIN-ASSOCIATED PROTEINS					
kinesin-2 associated	Sp-KAP	kinesin II-accessory 115 kDa polypeptide [Sp]; KAP, Kinesin II associated Protein [Cr]; kinesin-associated protein 3 [Hs]	SPU_010954	NP_999823	0.0
kinesin-1 associated	Sp-KLC	Kinesin Light Chain [Sp]	SPU_018898	Q05090	0.0

Dyneins in the sea urchin genome assembly

Table 3

Column headings are as in Table 1 and as described in the text. Genus species abbreviations are as follows. Cf: *Canis familiaris*, Hs: *Homo sapiens*, Tg: *Tripneustes gratilla*.

Gene Name	Common Name	Synonyms (organism)	Predicted polypeptide Length (% identity to Indexing Gene)	SPU Gene ID	Indexing Gene ID (Accn. Number)	Homologous Genes Other homologs used (Accn. Number)
Sp-DYNC1H1	Cytoplasmic dynein 1 heavy chain homolog	DYN1 (sc), DYH1A (tg), DHC1A (cr), DHC64C (dm), DHC1B (cr)	4652 (74%)	SPU_030236	hsDYNC1H1 (Q14204)	tgDYNC1H1 (CAA79935)
Sp-DYNC2H1	Cytoplasmic dynein 2 heavy chain homolog	IFT dynein (cr), Beethoven (dm), che-3 (ce)	4312 (95%)	SPU_030235	tgDYNC2H1 (AAA63583.2)	hsDYNC2H1 (NP_001367)
Sp-DNAH1	Axonemal dynein heavy chain 1 homolog	DYH6 (tg)	4154 (64%)	SPU_030223	hsDNAH1 (NP_056327.3)	tgDNAH1 (AAA63589)
Sp-DNAH2	Axonemal dynein heavy chain 2 homolog	IDA1-alpha (cr), DYH5C (tg)	4458 (65%)	SPU_030224	hsDNAH2 (NP_065928.2)	tgDNAH2 (AAA63590)
Sp-DNAH3	Axonemal dynein heavy chain 3 homolog	DYH7B (tg)	4038 (64%)	SPU_030225	hsDNAH3 (NP_060009.1)	tgDNAH3 (AAA63593)
Sp-DNAH4	Axonemal dynein heavy chain 4 homolog	DYH7D (tg)	4208 (56%)	SPU_030222	hsDNAH3 (NP_060009.1)	tgDNAH4 (AAA63595)
Sp-DNAH5	Axonemal dynein heavy chain 5 homolog	DYH3B(tg), ODA-gamma (cr)	4532 (62%)	SPU_030226	hsDNAH5 (NP_001360.1)	tgDNAH5 (AAA63584)
Sp-DNAH6	Axonemal dynein heavy chain 6 homolog	DYH5A (tg)	4160 (64%)	SPU_030227	cfDNAH6 (XP_532984.2)	tgDNAH6 (AAA63591)
Sp-DNAH7	Axonemal dynein heavy chain 7 homolog	DYH7A (tg)	3982 (68%)	SPU_030228	hsDNAH7 (NP_061720.1)	tgDNAH7 (AAA63592)
Sp-DNAH8	Axonemal dynein heavy chain 8 homolog	DYH3C (tg)	4592 (66%)	SPU_030229	hsDNAH8 (CAI42433.1)	tgDNAH8 (AAA63586)
Sp-DNAH9	Axonemal dynein heavy chain 9 homolog	DHC-beta (tg), DYH2 (tg), OAD-beta (cr)	4466 (95%)	SPU_030230	tgDNAH9 (beta) (CAA42170)	hsDNAH9 (Q9NYC9), hsDNAH11 (CAC60121.1), cfDNAH17 (XM_533129.1)
Sp-DNAH10	Axonemal dynein heavy chain 10 homolog	DYH4 (tg)	4592 (68%)	SPU_030231	cfDNAH10 (XP_543369.2)	tgDNAH10 (AAA63587)
Sp-DNAH12	Axonemal dynein heavy chain 12 homolog	IDA1-beta (cr), DYH7C (tg)	3935 (66%)	SPU_030232	cfDNAH12 (XP_541831.2)	tgDNAH12 (AAA63594)
Sp-DNAH14	Axonemal dynein heavy chain 14 homolog	DYH5B (tg)	4337 (41%)	SPU_030233	cfDNAH14 (XM_537236)	tgDNAH14 (AAA63588)
Sp-DNAH15	Axonemal dynein heavy chain 15 homolog	DYH3A (tg)	4647 (72%)	SPU_030234	hsDNAH8 (CAI42433.1)	tgDNAH15 (AAA63586)
Sp-DYNC1I1	Cytoplasmic dynein 1 intermediate chain 1		645 (51%)	SPU_022687	hsDYNC1I1 (NP_004402.1)	
Sp-DYNC1LI1	Cytoplasmic dynein 1 light intermediate chain 1		523 (45%)	SPU_015909	hsDYNC1LI1 (NP_057225.1)	
Sp-DYNC2LI1	Cytoplasmic dynein 2 light intermediate chain 1		351 (48%)	SPU_018582	hsDYNC2LI1 (NP_057092.2)	
Sp-DYNC2LI2	Cytoplasmic dynein 2 light intermediate chain 2		201 (48%)	SPU_008154	hsDYNC2LI2 (NP_056337)	
Sp-DYNLL1	Cytoplasmic dynein light chain, LC8-type 1		89 (89%)	SPU_004009	hsDYNLL1 (AAI00290.1)	
Sp-DYNLRB1	Dynein light chain, roadblock-type 1		96 (59%)	SPU_003137	hsDYNLRB1 (NP_054902.1)	

Strongylocentrotus purpuratus Gene		Predicted polypeptide Length (% identity to Indexing Gene)	SPU Gene ID	Homologous Genes	
Gene Name	Common Name Synonyms (organism)			Indexing Gene ID(Accn. Number)	Other homologs used(Accn. Number)
Sp-DYNLRB2	Dynein light chain, roadblock-type 2	96 (60%)	SPU_008699	hsDYNLRB2 (NP_570967.1)	
Sp-DYNLT1	Dynein light chain, Tctex-type 1	113 (85%)	SPU_008471	hsDYNLT1 (CAI95303.1)	
Sp-DYNLT2	Dynein light chain, Tctex-type 2	198 (35%)	SPU_006354	hsDYNLT2 (AAN34631.1)	
Sp-DYNLT3	Dynein light chain, Tctex-type 3	116 (59%)	SPU_008471	hsDYNLT3 (NP_006511.1)	
Sp-DNAI1	Axonemal dynein intermediate chain 1 homolog	699 (60%)	SPU_019506	hsDNAI1 (NP_036276)	
Sp-DNAI2	Axonemal dynein intermediate chain 2 homolog	605 (68%)	SPU_006699	hsDNAI2 (AAG38489.1)	
Sp-Txl-2	Thioredoxin-like chain 2 homolog	330 (46%)	SPU_007092	hsTxl-2 (AAG28497.1)	
Sp-DNALI1	Axonemal dynein light intermediate polypeptide 1	280 (75%)	SPU_015320	hsDNALI1 (NP_003453.2)	
Sp- TCTEX1D1	Axonemal outer arm dynein light chain 1 homolog	179 (54%)	SPU_013200	hsTCTEX1D1 (NP_689878.1)	
Sp-DNAL1	Axonemal outer arm dynein light chain 1homolog	190 (69%)	SPU_018854	hsDNAL1 (AAQ11377.1)	
Sp-DNAL4	Axonemal dynein light polypeptide 4 homolog	105 (77%)	SPU_004377	hsDNAL4 (NP_005731.1)	