Title
The genome and cytoskeleton of Naegleria gruberi, an amoeboflagellate

Author
Fritz-Laylin, Lillian Kathleen

Publication Date
2010

Peer reviewed|Thesis/dissertation

The genome and cytoskeleton of *Naegleria gruberi*, an amoeboflagellate

by

Lillian Kathleen Fritz-Laylin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor W. Zacheus Cande, Chair
Professor Rebecca Heald
Professor Daniel Rokhsar
Professor John Taylor

Spring 2010

The genome and cytoskeleton of *Naegleria gruberi*, an amoeboflagellate

by Lillian Kathleen Fritz-Laylin

Abstract

The genome and cytoskeleton of *Naegleria gruberi*, an amoeboflagellate

by

Lillian Kathleen Fritz-Laylin

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

W. Zacheus Cande, Chair

Naegleria gruberi is a free-living eukaryote that has been described as a unicellular "Jekyll-and-Hyde. Most of its time it can be found as a small (15 μm) amoebae, common to freshwater environments throughout the world. However, when exposed to stressful conditions the amoebae quickly and synchronously differentiate into flagellates. This dramatic change involves the formation of an entire cytoplasmic microtubule cytoskeleton, including *de novo* assembly of the centriole-like basal bodies, '9+2' flagella, and a cortical microtubule array. This 'quick-change act' offers an unprecedented opportunity to study the assembly of an entire microtubule cytoskeleton, particularly the beautifully complex structures of centrioles. However, utilization of *Naegleria* as a model organism has been frustrated by lack of sequence information and molecular tools. This dissertation describes my efforts during my graduate studies to coordinate the *Naegleria* genome project, analyze the resulting sequence data, and develop tools with which to study *Naegleria*'s amoeba-to-flagellate transition, with a focus on basal body assembly.

Although the analysis of the *Naegleria* genome revealed many interesting aspects of both *Naegleria* biology and the evolution of eukaryotes, the results presented here are limited to those that pertain to *Naegleria*'s actin and microtubule cytoskeletons. In particular, detailed manual inspection of individual *Naegleria* genes uncovered an extensive repetoire of previously characterized actin and microtubule cytoskeletal components. This indicates that despite *Naegleria*'s extremely distant relationship to animals, the transient nature of its cytoplasmic microtubule cytoskelelon, and that the

amoeboid actin cytokeleton functions independently of microtubules, *Naegleria* has the capacity to to form a canonical cytoskeleton.

Additionally, we took advantage of *Naegleria*'s distant relationships to other sequenced organisms to identify ancient genes that we predict to be involved with amoeboid or flagellar motility. To do this, we compared all the genes from a diversity of sequenced eukaryotes, and selected protein families conserved only in eukaryotes with flagellar motility (Flagellar-Motility associated genes; FMs) or amoeboid motility (Amoeboid-Motility associated genes; AMs). Along with the expected gene families, *Naegleria*'s 182 FM's include 36 novel flagellar-associated genes. The 63 AMs include genes known to be involved in amoeboid motility, membrane differentiation, and 19 novel genes. As far as we know, this is the first catalog of genes predicted to be associated specifically with amoeboid motility.

During the amoeba-to-flagellate transition, *Naegleria* synchronously assembles centrioles from scratch, providing nearly limitless amounts of material for both proteomics and microarrays, and an unparalleled oppourtunity to study how these structures assemble. Although we know that centrioles and basal bodies are composed principally of a cylinder of nine microtubule triplets, their protein composition and method of assembly remain largely mysterious. Animal centrioles usually duplicate via "templated" assembly, with the new centriole developing perpendicularly from the side of a preexisting centriole. Centrioles can also be formed "*de novo*", in cytosol devoid of preexisting centrioles in some plant and animal cells, as well as *Naegleria*. How *Naegleria* makes exactly two basal bodies *de novo* remains an open question. During my graduate studies, I have developed antibodies and used them to describe the order of expression and incorporation of three *Naegleria* centriole proteins (SAS-6, γ-tubulin, and centrin). I also used these to provide the first evidence that *Naeglearia* has templated, as well as *de novo*, basal body assembly, and suggest that having both capacities allows *Naegleria*, and other organisms (*e.g.* mouse embryos), to make the correct number of centrioles.

Finally, I have tracked the expression of *Naegleria*'s genes during differentiation to identify novel centriolar and flagellar proteins. Although about a third of *Naegleria* genes are induced and another third are repressed during differentiation, I focused on the evolutionarily conserved FM genes, and use the timing of induction to subdivide them into a subset of 55 genes enriched in known basal body proteins (induced early) and a subset of 82 genes enriched in axonemal proteins (induced late). The centrosome-enriched set includes nearly every conserved basal body component that has been previously characterized, many components required for microtubule nucleation (a process that occurs largely at centrosomes) and ten novel genes

that are conserved across eukaryotes. As a proof of principle, the human ortholog of one of the novel genes was tagged, and indeed localizes to the centrosomes of human cells.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Acknowledgements

This dissertation owes its existence to a large number of amazing mentors, collaborators, friends, and family who have helped me more than I can express. In particular I wish to thank Zac and the rest of the Cande lab. Your collective sense of humor, willingness to try insane experiments, and fellowship are a gift, and I am sad to soon be leaving such a warm group of people.

I also want to thank Chandler Fulton for his continual eagerness to help a novice Naegleriophile. Chan, your optimism and active cheerleading, particularly after several rounds of rejection of the genome paper, were greatly appreciated. Our baby would have developed so beautifully without it, and I look forward to continuing our work together.

Scott Dawson has always been there to point out the light at the end of the tunnel. Scott, you have treated me as a collaborator and your equal. Seeing myself through your eyes makes me realize that I can do more than I often give myself credit for. Thank you.

I also want to thank Simon Prochnik, who has been an invaluable friend and collaborator over the past several years. Simon, I hope by working with you, some of your ability to focus on the most rewarding project at hand, and to continually prioritize has rubbed off on me. You are a constant reminder to keep my eyes on the prize, and we wouldn't have been able to finish the genome project without you.

Finally, I want to thank my friends (particularly Meredith Carpenter and Danae Schulz), my parents, Scott Blitch, and Benjamin Cowden. I will always be indebted to you for your help and support over the past six years. Thank you.

Chapter 1: An Introduction to *Naegleria* and basal bodies

Naegleria is a complex single celled eukaryote that can live as an amoeba or a flagellate

Naegleria gruberi is a free-living heterotrophic protist commonly found in freshwater and moist oxic and suboxic soils around the world (De Jonckheere, 2002; Fulton, 1970, 1993). Its predominant form is a 15μm predatory amoeba that can reproduce every 1.6 hours. when eating bacteria (Fulton et al., 1984). Yet *Naegleria* is best known for its remarkable ability to metamorphose from amoebae into transitory biflagellates that swim a hundred times faster than amoebae crawl (Figure 1.1) (Fulton, 1993). This rapid (<1.5 hr.) change begins with the cessation of amoeboid movement and actin synthesis, followed by *de novo* assembly of an entire cytoplasmic microtubule cytoskeleton, including canonical basal bodies and 9+2 flagella (Figure 1) (Fulton, 1993). *Naegleria* also forms resting cysts, which excyst to produce amoebae (Fulton, 1970). Amoebae divide using an intranuclear spindle without centrioles (Fulton, 1993).

Although *Naegleria* is best known as a model for *de novo* basal body assembly, our recent description of *Naegleria*'s genome (Fritz-Laylin et al., 2010) suggests that it may be a good system in which to study a variety of cellular processes. In particular, *Naegleria* has many of the key features that distinguish eukaryotic cells from bacteria and archaea. These features include complete actin and microtubule cytoskeletons, extensive mitotic and meiotic machinery, calcium/calmodulin mediated regulation, transcription factors (Iyer et al., 2008), endosymbiotic organelles (mitochondria), and organelles of the membrane trafficking system Although it lacks visible Golgi, *Naegleria* contains the required genes (Dacks et al., 2003). The genome also encodes an extensive array of signaling machinery that orchestrates *Naegleria's* complex behavior. This repertoire includes entire pathways not found in parasitic protists (G-protein coupled receptor signaling and histidine kinases), as well as 265 predicted protein kinases, 32 protein phosphatases, and 182 monomeric Ras-like GTPases. Additionally, like many aerobic microbial eukaryotes, *Naegleria* oxidises glucose, various amino acids, and fatty acids via the Krebs cycle and a branched mitochondrial respiratory chain. However, the *Naegleria* genome also encodes an unexpected capacity for elaborate and sophisticated anaerobic metabolism.

In addition to the complement of eukaryotic features encoded in *Naegleria*'s genome, *Naegleria*'s unique cell biology also indicates that it is a promising model for studying basic cell biology questions. *Naegleria* can be induced in the laboratory to differentiate into three different states (amoeba, flagellate, and cyst), and can be readily mitotically synchronized (approximately 70% mitotic synchrony from using

simple temperature fluxuations, (Fulton and Guerrini, 1969)). *Naegleria* can be easily switched between axenic and xenic growth, and readily isolated from the environment. Despite these rich prospects, *Naegleria* will never mature into an experimental system without molecular tools. Although there are reports of molecular transformation (Song et al., 2006), I (and others) have not been able to successfully repeat these experiments. However, the genome does encode the necessary components for RNA mediated gene knockdown, and we are conducting ongoing experiments to adapt this technology (as well as molecular transformation) for use in *Naegleria.*

Naegleria is only distantly related to studied organisms but has many typical eukaryotic features.

Eukaryotes emerged and diversified perhaps a billion years ago (Brinkmann and Philippe, 2007; Yoon et al., 2004), radiating into new niches by taking advantage of their metabolic, cytoskeletal, and compartmental complexity. Half a dozen deeply divergent, major eukaryotic clades survive, including diverse unicellular groups along with the familiar plants, animals, and fungi (Figure 1.2). These contemporary species combine retained ancestral eukaryotic features with novelties specific to their particular lineages.
Naegleria belongs to Heterolobosea, a major eukaryotic lineage that together with Euglenoids (which includes the distantly related parasitic trypanosomes) and the Jakobids comprises the ancient and diverse clade we term "JEH" for Jakobids, Euglenoids, Heterolobosea (Figure 1.2) (Rodriguez-Ezpeleta et al., 2007). Within Heterolobosea, the genus *Naegleria* encompasses as much evolutionary diversity as the tetrapods (based on rDNA divergence (Fulton, 1993)) and includes the "brain-eating amoeba" *N. fowleri*, an opportunistic pathogen that is usually free-living in warm freshwater, but can also cause fatal meningoencephalitis in humans (Visvesvara et al., 2007).

The position of the eukaryotic root is a matter of controversy and great interest (Baldauf, 2003) with no clearly supported hypothesis at present. Although the position of the root of the eukaryotic tree remains controversial, three major hypotheses have emerged (Ciccarelli et al., 2006; Hampl et al., 2009; Stechmann and Cavalier-Smith, 2002). Two of the three main hypotheses (Figure 1.2 insets) employ different strategies for determining the most basal branches in the eukaryotic tree: the first uses archaeal sequences as an outgroup to define the deepest branches in the eukaryotic tree (Root B) (Yoon et al., 2008); in the second (Root A), the root has been inferred from a single gene fusion event (Stechmann and Cavalier-Smith, 2002). The last hypothesis (root C) relies on a monophyletic relationship between JEH and POD, forming the "excavates".
Using these rooting schemes, we can attempt to determine which genes and features were likely present in the eukaryotic ancestor. For example, the lineages leading to *Naegleria* (JEH) and humans (opisthokonts) diverged either at the root of

eukaryotes, or immediately after the separation of the "POD" clade from the eukaryotic ancestor (green highlighting in Fig. 2 inset). Therefore, genes and characteristics shared between *Naegleria* and humans were likely present in the eukaryotic ancestor (or more accurately, in the last common ancestor of non-POD eukaryotes).

Setting rooting schemes aside, genes present in all eukaryotic groups were almost certainly present in the eukaryotic ancestor. Additionally, features shared between *Naegleria* and other major eukaryotic groups likely existed in their common ancestor. Although we do not know when this ancestor lived, it was certainly before the divergence of the clades in question, and therefore relatively early in tractable eukaryotic history (Figure 1.2).

Gene loss is a process that further clouds our picture of ancestral eukaryotes. Although every major eukaryotic clade has lost a considerable number of genes (Fritz-Laylin et al., 2010), this process is particularly pronounced in obligate parasites, whose genomes are thought to be derived by gene loss and high sequence divergence (Carlton et al., 2007; Morrison et al., 2007). Therefore, absence of a gene from a parasite genome is not necessarily informative about the ancestral eukaryotic state. This includes either of the sequenced POD genomes (*Trichomonas* or *Giardia*), which were considered for a long time to be extant "primitive" eukaryotes (Simpson et al., 2006). As the first sequenced genome from a free-living member of the JEH clade, *Naegleria* has proved particularly useful for advancing our understanding of ancient eukaryotes (Fritz-Laylin et al., 2010).

Naegleria is an unparalleled organism to study centriole assembly.

Centrosomes and basal bodies are examples of MicroTubule Organizing Centers (MTOCs). Centrosomes, used by cells to organize mitotic microtubules into spindles, are cylinders of microtubules known as centrioles, plus a surrounding dense matrix known as the PeriCentriolar Material (PCM) (Figure 1.3). The beautiful and enigmatic pinwheel structures of centrioles and basal bodies are composed largely of nine microtubule triplets forming a cylinder approximately 0.2 microns in diameter and 1 micron long (Fulton and Dingle, 1971).

Although centrioles and basal bodies have been studied for well over a hundred years, their components, order of assembly, and the regulation of their number are still largely mysterious. For example, proteomic analysis of *Chlamydomonas* and *Tetrahymena* basal bodies (Keller et al., 2005; Kilburn et al., 2007) and human centrosomes (Andersen et al., 2003) suggests that these structures contain hundreds of different proteins. However, we really only know when a handful of these proteins are incorporated into the structure (*e.g.* Table 1.1). This dearth of information is largely due to technical difficulties in studying centrioles in most systems, for the following reasons:

1.) Templated centriole replication seen in most eukaryotes under most conditions does not allow proteomic isolation of developing daughters from their mothers.

2.) Templated assembly is tied to the cell cycle, rendering it difficult to distinguish basal body-specific from other induced cell cycle genes.

3.) *De novo* assembly can only be assayed in other systems in a single cell or embryo, making gathering enough material for proteomic or transcription-based approaches unfeasible.

Naegleria can undergo *de novo* basal body formation synchronously and independently from mitosis (Fulton, 1970), providing nearly limitless amounts of material for both proteomics and microarrays. Therefore, *Naegleria* is an ideal system in which to use these discovery-based approaches to learn how organisms assemble basal bodies.

Centrioles are required for forming flagella and are assembled in a step wise fashion.

Metazoan centrosomes define mitotic spindle poles, and their centrioles are called basal bodies when used to form cilia after terminal differentiation (Sorokin, 1968). Moreover, Meeves (1900) showed in a series of classical experiments that centrioles and basal bodies are inter-convertible structures (Wilson, 1928). Basal bodies of some protists can function simultaneously in both flagella and mitotic spindle poles (Ribeiro et al., 2002; Sagolla, 2005). Although centrioles have classically been thought to function in mitosis, the fact that these structures can only be found in organisms with flagella (they have been lost in all organisms which have lost the ability to form flagella, e.g. flowering plants and most fungi) strongly indicates that they are required specifically for flagellar assembly. Furthermore, fruit flies with a mutation in the centriole component SAS-4 lose centrioles by the third instar larval stage (Basto et al., 2006). Surprisingly, their cells continue to undergo mitosis, and the animals develop fairly normally. However, SAS-4 mutant flies do not produce cilia and flagella, and die quickly after reaching adulthood from complications related to this defect (Basto et al., 2006).

Centrosomes and their associated centrioles/basal bodies must replicate exactly once per cell cycle, as duplication errors can lead to problems with chromosomal segregation and cell morphology (Kramer et al., 2002). The most commonly studied method of centriole duplication, termed "templated" assembly, involves new centrioles emerging perpendicularly from a preexisting centriole (Beisson and Wright, 2003). Despite this tantalizing title, there is no evidence that the mother centriole gives any material to the developing daughter centriole during templated assembly (Marshall, 2001). However, during the development of parthenogenetic

oocytes, sperm of some plants, and several protists, basal bodies can be formed "*de novo*", in previously centriole-free cytoplasm (Marshall, 2001). Additionally, terminally differentiated fibroblasts held in S-phase can undergo *de novo* centriole assembly after removal via laser microsurgery of preexisting centrioles (Khodjakov et al., 2002).

Although we know little about the molecular pathways involved in basal body or centriole formation (Marshall, 2001), a handful of proteins have been shown to be genetically required for this process (Strnad and Gonczy, 2008) (Table 1.1 and Figure 1.3), with additional proteins localized to the developing structure. There are also a number of proteins that have been shown to be involved in centriole assembly but are found in a limited number of organisms (e.g. the protein SAS-5 used by *C. elegans* and ZYG-1/Sak/PLK4, a kinase required for centriole duplication in animals and conserved only in opisthokonts). As I am interested in conserved mechanisms of basal body assembly, I will only discuss those proteins that have identifiable homologs in at least two eukaryotic groups (Figure 2).

Electron microscopy of basal body formation in *Paramecium* has divided this process into five visibly distinct stages (also see Figure 3):

1.) Generative Disc: This electron-dense, fibrous plate-shaped object has been reported as the first recognizable step in centriole assembly in *Paramecium* and forms within the PCM of the mother centriole (Dippell, 1968). During initial centriole construction in *Chlamydomonas,* it has been suggested that a homologous amorphous ring is formed (Nakazawa et al., 2007).

2.) Cartwheel: This nine-fold symmetric star-shaped structure is found at the proximal end (nearest the nucleus) of mature *Chlamydomonas* basal bodies (Marshall, 2001), and occasionally in human cells (Alvey, 1986). In *Chlamydomonas*, assembly of the cartwheel requires the proteins Bld10 (Hiraki et al., 2007) and Bld12 (a homolog of the coiled-coil protein SAS-6 that is ubiquitious to organisms with flagella) (Nakazawa et al., 2007). Although the first structure formed during *C. elegans* centriole assembly is a hollow "tube" of coiled-coil proteins that do not contain microtubules, this structure also requires SAS-6 (Dammermann et al., 2004; Leidel et al., 2005; Pelletier et al., 2006).

It remains unclear if the cartwheel forms before, or in parallel with, the microtubules seen in the procentriole (Matsuura et al., 2004). In the green algae *Spermatozopsis similes*, the cartwheel and microtubules seem to form in parallel (Lechtreck and Grunow, 1999).

3.) Procentriole: Microtubules first appear in the single ring of 9 a-tubules (Beisson and Wright, 2003). Mutations in SAS-6 indicate that this protein is required for proper attachment of the a-microtubules into the centriole/basal body structure in

Chlamydomonas (Nakazawa et al., 2007). The pro-centriole structure continues to elongate after addition of the a-tubules, a process which requires (in humans at least) the centrin-binding centriole protein POC5 (Azimzadeh et al., 2009).

4.) Immature Centriole: The immature centriole contains all three tubule cylinders (a, b and c). Genetic studies in various organisms suggest that ε-tubulin and δ-tubulin (divergent tubulin superfamily members) are genetically required for the formation of the b and c tubules, respectively (Dutcher, 2003; Dutcher et al., 2002; Garreau de Loubresse et al., 2001; O'Toole et al., 2003).

5.) Mature Centriole: The immature centriole matures as it gathers new PCM, including the coiled-coil proteins pericentrin and ninein, and accessory structures and satellites formed by proteins such as cenexin (Dammermann and Merdes, 2002; Lange and Gull, 1995).

Although these structures are similar to those described during basal body assembly in many eukaryotes, there does seem to be some variation among organisms. The most discussed atypical structural intermediate is the central tube formed early during assembly of *C. elegans*' centrioles. This hollow tube expands and elongates, gathering microtubules to form a centriole (Pelletier et al., 2006). The centrioles of *C. elegans* are particularly interesting, because its genome is missing several widely-conserved centriolar genes, including centrin and δ-tubulin (Beisson and Wright, 2003; Bornens and Azimzadeh, 2007), and the mature centrioles only contain singlet microtubules (instead of the canonical triplet microtubule blades). However, *C. elegans* uses many canonical centriole assembly proteins, indicating that diverse centriole assembly pathways use many of the same principles and pieces, with (obviously) some modifications. As we learn more about the differences between centriole assembly routes, we can begin to understand which steps and components can change and still give rise to such a conserved and complex biological structure.

Additional proteins have been shown to be genetically required for centriole or basal body assembly (Table 1). In many cases, deletion or knockdown of these genes results in a complete block in centriole duplication, a phenotype that makes their precise function in this process difficult to tease out. In particular centrin (Kuchka and Jarvik, 1982), (Koblenz et al., 2003; Salisbury et al., 2002; Taillon et al., 1992) and gamma tubulin (Dammermann et al., 2004; Haren et al., 2006) have both been shown to be integral components of centrioles and basal bodies, and essential for their assembly. Additional proteins seem required for ensuring that centrioles grow to the proper length (although in some cases these assumptions rest upon over expression phenotypes (Table 1)).

References:

Alvey, P.L. (1986). Do adult centrioles contain cartwheels and lie at right angles to each other? Cell Biol Int Rep *10*, 589-598.

Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. Nature *426*, 570-574.

Azimzadeh, J., Hergert, P., Delouvee, A., Euteneuer, U., Formstecher, E., Khodjakov, A., and Bornens, M. (2009). hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. J Cell Biol *185*, 101-114.

Baldauf, S.L. (2003). The deep roots of eukaryotes. Science *300*, 1703-1706.
Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C.G., Khodjakov, A., and Raff, J.W. (2006). Flies without centrioles. Cell *125*, 1375-1386.

Baum, P., Furlong, C., and Byers, B. (1986). Yeast gene required for spindle pole body duplication: homology of its product with Ca2+-binding proteins. Proc Natl Acad Sci U S A *83*, 5512-5516.

Beisson, J., and Wright, M. (2003). Basal body/centriole assembly and continuity. Curr Opin Cell Biol *15*, 96-104.

Bornens, M., and Azimzadeh, J. (2007). Origin and evolution of the centrosome. Adv Exp Med Biol *607*, 119-129.

Brinkmann, H., and Philippe, H. (2007). The diversity of eukaryotes and the root of the eukaryotic tree. Adv Exp Med Biol *607*, 20-37.

Burki, F., Shalchian-Tabrizi, K., and Pawlowski, J. (2008). Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett *4*, 366.

Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R., Bidwell, S.L., Alsmark, U.C., Besteiro, S.*, et al.* (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science *315*, 207-212.

Chang, P., Giddings, T.H., Jr., Winey, M., and Stearns, T. (2003). Epsilon-tubulin is required for centriole duplication and microtubule organization. Nat Cell Biol *5*, 71-76.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. Science *311*, 1283-1287.

Dacks, J.B., Davis, L.A.M., Sjogren, A.M., Andersson, J.O., Roger, A.J., and Doolittle, W.F. (2003). Evidence for Golgi bodies in proposed 'Golgi-lacking' lineages. Proc Biol Sci *270 Suppl 2*, S168-171.

Dammermann, A., and Merdes, A. (2002). Assembly of centrosomal proteins and microtubule organization depends on PCM-1. J Cell Biol *159*, 255-266.

Dammermann, A., Muller-Reichert, T., Pelletier, L., Habermann, B., Desai, A., and Oegema, K. (2004). Centriole assembly requires both centriolar and pericentriolar material proteins. Dev Cell *7*, 815-829.

De Jonckheere, J.F. (2002). A century of research on the amoeboflagellate genus *Naegleria*. Acta Protozoologica *41*, 309-342.

Dippell, R.V. (1968). The development of basal bodies in paramecium. Proc Natl Acad Sci U S A *61*, 461-468.

Dutcher, S.K. (2003). Long-lost relatives reappear: identification of new members of the tubulin superfamily. Curr Opin Microbiol *6*, 634-640.

Dutcher, S.K., Morrissette, N.S., Preble, A.M., Rackley, C., and Stanga, J. (2002). Epsilon-tubulin is an essential component of the centriole. Mol Biol Cell *13*, 3859-3869.

Dutcher, S.K., and Trabuco, E.C. (1998). The UNI3 gene is required for assembly of basal bodies of Chlamydomonas and encodes delta-tubulin, a new member of the tubulin superfamily. Mol Biol Cell *9*, 1293-1308.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J., Carpenter, M.L., Field, M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J.*, et al.* (2010). The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. Cell *In press*.

Fulton, C. (1970). Amebo-flagellates as research partners: The laboratory biology of *Naegleria* and *Tetramitus*. Methods Cell Physiol *4*, 341-476.

Fulton, C. (1993). *Naegleria* : A research partner for cell and developmental biology. Journal of Eukaryotic Microbiology *40*, 520-532.

Fulton, C., and Dingle, A.D. (1971). Basal bodies, but not centrioles, in Naegleria. J Cell Biol *51*, 826-836.

Fulton, C., and Guerrini, A.M. (1969). Mitotic synchrony in Nalegleria amebae. Exp Cell Res *56*, 194-200.

Fulton, C., Webster, C., and Wu, J.S. (1984). Chemically defined media for cultivation of *Naegleria gruberi*. Proc Natl Acad Sci USA *81*, 2406-2410.

Garreau de Loubresse, N., Ruiz, F., Beisson, J., and Klotz, C. (2001). Role of delta-tubulin and the C-tubule in assembly of Paramecium basal bodies. BMC Cell Biol *2*, 4.

Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., and Roger, A.J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A *106*, 3859-3864.

Haren, L., Remy, M.H., Bazin, I., Callebaut, I., Wright, M., and Merdes, A. (2006). NEDD1-dependent recruitment of the gamma-tubulin ring complex to the centrosome is necessary for centriole duplication and spindle assembly. J Cell Biol *172*, 505-515.

Hiraki, M., Nakazawa, Y., Kamiya, R., and Hirono, M. (2007). Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. Curr Biol *17*, 1778-1783.

Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol *38*, 1-31.

Keller, L.C., Geimer, S., Romijn, E., Yates, J., 3rd, Zamora, I., and Marshall, W.F. (2009). Molecular architecture of the centriole proteome: the conserved WD40 domain protein POC1 is required for centriole duplication and length control. Mol Biol Cell *20*, 1150-1166.

Keller, L.C., Romijn, E.P., Zamora, I., Yates, J.R., 3rd, and Marshall, W.F. (2005). Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr Biol *15*, 1090-1098.

Khodjakov, A., Rieder, C.L., Sluder, G., Cassels, G., Sibon, O., and Wang, C.L. (2002). De novo formation of centrosomes in vertebrate cells arrested during S phase. J Cell Biol *158*, 1171-1181.

Kilburn, C.L., Pearson, C.G., Romijn, E.P., Meehl, J.B., Giddings, T.H., Jr., Culver, B.P., Yates, J.R., 3rd, and Winey, M. (2007). New Tetrahymena basal body protein components identify basal body domain structure. J Cell Biol *178*, 905-912.

Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S., and Hyman, A.A. (2003). SAS-4 is a C. elegans centriolar protein that controls centrosome size. Cell *112*, 575-587.

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.D., and Nigg, E.A. (2007). Plk4-induced centriole biogenesis in human cells. Dev Cell *13*, 190-202.

Koblenz, B., Schoppmeier, J., Grunow, A., and Lechtreck, K.F. (2003). Centrin deficiency in Chlamydomonas causes defects in basal body replication, segregation and maturation. J Cell Sci *116*, 2635-2646.

Kramer, A., Neben, K., and Ho, A.D. (2002). Centrosome replication, genomic instability and cancer. Leukemia *16*, 767-775.

Kuchka, M.R., and Jarvik, J.W. (1982). Analysis of flagellar size control using a mutant of Chlamydomonas reinhardtii with a variable number of flagella. J Cell Biol *92*, 170-175.

Lange, B.M., and Gull, K. (1995). A molecular marker for centriole maturation in the mammalian cell cycle. J Cell Biol *130*, 919-927.

Lechtreck, K.F., and Grunow, A. (1999). Evidence for a direct role of nascent basal bodies during spindle pole initiation in the green alga Spermatozopsis similis. Protist *150*, 163-181.

Leidel, S., Delattre, M., Cerutti, L., Baumer, K., and Gonczy, P. (2005). SAS-6 defines a protein family required for centrosome duplication in C. elegans and in human cells. Nat Cell Biol *7*, 115-125.

Leidel, S., and Gonczy, P. (2003). SAS-4 is essential for centrosome duplication in C elegans and is recruited to daughter centrioles once per cell cycle. Dev Cell *4*, 431-439.

Marshall, W.F. (2001). Centrioles take center stage. Curr Biol *11*, R487-496.

Matsuura, K., Lefebvre, P.A., Kamiya, R., and Hirono, M. (2004). Bld10p, a novel protein essential for basal body assembly in Chlamydomonas: localization to the cartwheel, the first ninefold symmetrical structure appearing during assembly. J Cell Biol *165*, 663-671.

Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., *et al.* (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science *317*, 1921-1926.

Nakazawa, Y., Hiraki, M., Kamiya, R., and Hirono, M. (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. Curr Biol *17*, 2169-2174.

O'Toole, E.T., Giddings, T.H., McIntosh, J.R., and Dutcher, S.K. (2003). Three-dimensional organization of basal bodies from wild-type and delta-tubulin deletion strains of Chlamydomonas reinhardtii. Mol Biol Cell *14*, 2999-3012.

Peel, N., Stevens, N.R., Basto, R., and Raff, J.W. (2007). Overexpressing centriole-replication proteins in vivo induces centriole overduplication and de novo formation. Curr Biol *17*, 834-843.

Pelletier, L., O'Toole, E., Schwager, A., Hyman, A.A., and Muller-Reichert, T. (2006). Centriole assembly in Caenorhabditis elegans. Nature *444*, 619-623.

Ribeiro, K.C., Pereira-Neves, A., and Benchimol, M. (2002). The mitotic spindle and associated membranes in the closed mitosis of trichomonads. Biol Cell *94*, 157-172.

Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G., and Glover, D.M. (2007). DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. Curr Biol *17*, 1465-1472.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A.J., Gray, M.W., Philippe, H., and Lang, B.F. (2007). Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. Curr Biol *17*, 1420-1425.

Sagolla, M.S. (2005). Analyses of microtubule arrays in Giardia lamblia and fission yeast: Implications for the evolution of the cytoskeleton. In Molecular and Cell Biology (Berkeley, UC Berkeley), pp. 203.

Salisbury, J.L., Suino, K.M., Busby, R., and Springett, M. (2002). Centrin-2 is required for centriole duplication in mammalian cells. Curr Biol *12*, 1287-1292.

Simpson, A.G., Inagaki, Y., and Roger, A.J. (2006). Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol *23*, 615-625.

Song, K.J., Jeong, S.R., Park, S., Kim, K., Kwon, M.H., Im, K.I., Pak, J.H., and Shin, H.J. (2006). Naegleria fowleri: functional expression of the Nfa1 protein in transfected Naegleria gruberi by promoter modification. Exp Parasitol *112*, 115-120.

Sorokin, S.P. (1968). Reconstructions of centriole formation and ciliogenesis in mammalian lungs. J Cell Sci *3*, 207-230.

Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. Science *297*, 89-91.

Strnad, P., and Gonczy, P. (2008). Mechanisms of procentriole formation. Trends Cell Biol *18*, 389-396.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., and Gonczy, P. (2007). Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. Dev Cell *13*, 203-213.

Taillon, B.E., Adler, S.A., Suhan, J.P., and Jarvik, J.W. (1992). Mutational analysis of centrin: an EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of Chlamydomonas. J Cell Biol *119*, 1613-1624.

Visvesvara, G.S., Moura, H., and Schuster, F.L. (2007). Pathogenic and opportunistic free-living amoebae: Acanthamoeba spp., Balamuthia mandrillaris, Naegleria fowleri, and Sappinia diploidea. FEMS Immunol Med Microbiol *50*, 1-26.

Wilson, E.B. (1928). In The Cell in Development and Heredity (London and New York, Macmillan), p. 357.

Winey, M., Goetsch, L., Baum, P., and Byers, B. (1991). MPS1 and MPS2: novel yeast genes defining distinct steps of spindle pole body duplication. J Cell Biol *114*, 745-754.

Yoon, H.S., Grant, J., Tekle, Y.I., Wu, M., Chaon, B.C., Cole, J.C., Logsdon, J.M.J., Patterson, D.J., Bhattacharya, D., and Katz, L.A. (2008). Broadly sampled multigene trees of eukaryotes. BMC Evol Biol *8*, 14.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol *21*, 809-818.

Table 1.1: Evolutionarily conserved proteins implicated in centriole or basal body formation.

Homologs of centriole/basal body proteins in different species are grouped by shading. Organisms are indicated in parentheses after the protein name, abbreviated as follows: Ce; *Caenorhabditis elegans*, Dm; *Drosophila melanogaster*, Cr; *Chlamydomonas reinhardtii*, Hs; *Homo sapiens*, Sc; *Saccharomyces cerevisiae*, Xl; *Xenopus laevis*, Pt; *Paramecium tetraurelia*. Methods abbreviated as follows: EM, electron microscopy; IF; immunofluorescence. Table modified from (Strnad and Gonczy, 2008)

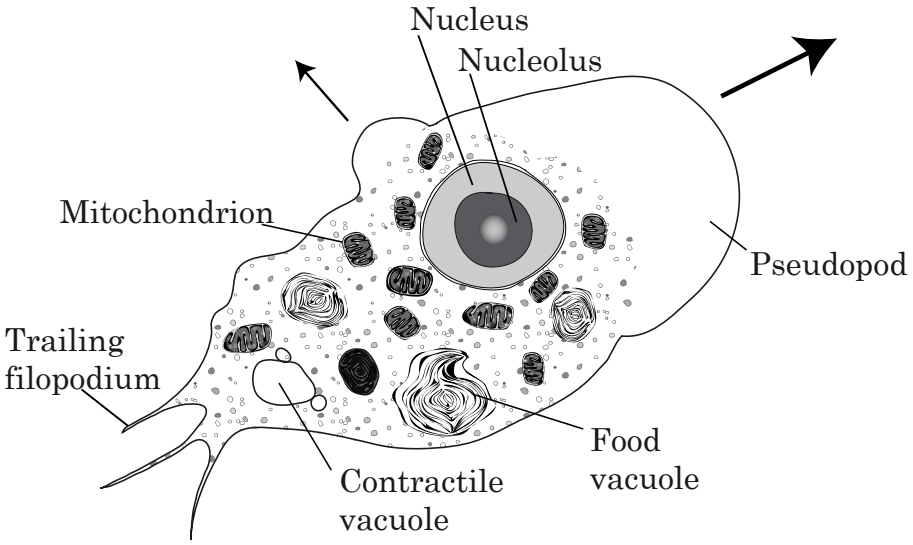
Protein	Experiment	Phenotype	References
SAS-6 (Ce)	RNAi	IF+EM; no procentriole in mitosis; no central tube	(Pelletier et al., 2006), (Leidel et al., 2005), and (Dammermann et al., 2004)
DSas-6 (Dm)	Mutation	IF+EM; reduced number of centrioles	(Rodrigues-Martins et al., 2007) and (Peel et al., 2007)
Bld12p (Cr)	Mutation	EM; no flagella in most cells, structurally defective basal bodies	(Nakazawa et al., 2007)
HsSAS-6 (Hs)	RNAi	IF+EM; no procentriole in mitosis	(Leidel et al., 2005) and (Strnad et al., 2007)
SAS-4 (Ce)	RNAi	IF+EM; no procentriole in mitosis; central tube still forms	(Kirkham et al., 2003), (Pelletier et al., 2006), and (Leidel and Gonczy, 2003)
DSas-4 (Dm)	Mutation	EM; no centrioles in mitosis	(Basto et al., 2006)
CPAP (Hs)	RNAi	IF; no centriole over-duplication upon Plk4 overexpression	(Kleylein-Sohn et al., 2007)
TBG-1 (Ce)	RNAi	IF; delayed procentriole formation	(Dammermann et al., 2004)
γ-Tubulin (Hs)	RNAi	IF; no procentriole in mitosis	(Haren et al., 2006)
Bld10p (Cr)	Mutation	EM; absence of basal bodies	(Hiraki et al., 2007)
Cep135 (Hs)	RNAi	IF; no centriole over-duplication upon Plk4 overexpression	(Kleylein-Sohn et al., 2007)
Cdc31p (Sc)	Mutation	EM; no spindle-pole body duplication	(Baum et al., 1986) and (Winey et al., 1991)
Centri	RNAi;	DIC+EM; variable	(Kuchka and Jarvik, 1982),

n (Cr)	mutation	flagellar number	(Taillon et al., 1992) and (Koblenz et al., 2003)
Centrin-2 (Hs)	RNAi	IF+EM; single centriole in each spindle pole	(Salisbury et al., 2002)
POC1 (Hs)	RNAi	IF; no centriole over-duplication in U2OS cells held in S-phase	(Keller et al., 2009)
POC5 (Hs)	RNAi	Distal centriole assembly: depleted cells accumulate short procentrioles	(Azimzadeh et al., 2009)
δ-Tubulin (Cr)	Mutation	EM; basal bodies with doublet rather than singlet microtubules	(Dutcher and Trabuco, 1998; O'Toole et al., 2003)
δ-Tubulin (Pt)	RNAi	EM; basal bodies with doublet rather than singlet microtubules	(Garreau de Loubresse et al., 2001)
ε-Tubulin (Cr)	Mutation	EM; shortened centrioles with singlet microtubules.	(Dutcher et al., 2002)
η-Tubulin (Xl)	Antibody depletion	Block in centriole duplication. IF localization to distal appendages	(Chang et al., 2003)

Figure 1.1. Schematic of *Naegleria* amoeba and flagellate forms.

Naegleria amoebae move along a surface with a large blunt pseudopod. Changing direction (arrows) follows the eruption of a new, usually anterior, pseudopod. *Naegleria* maintains fluid balance using a contractile vacuole. The nucleus contains a large nucleolus. The cytoplasm has many mitochondria and food vacuoles which are excluded from pseudopods. Flagellates also contain canonical basal bodies and flagella (insets). Basal bodies are connected to the nuclear envelope via a single striated rootlet.

Amoeboid form



Flagellate form

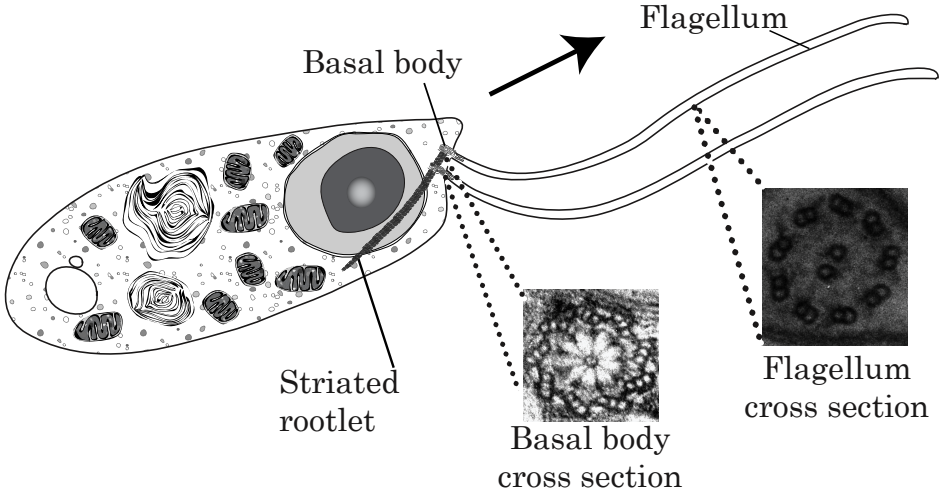
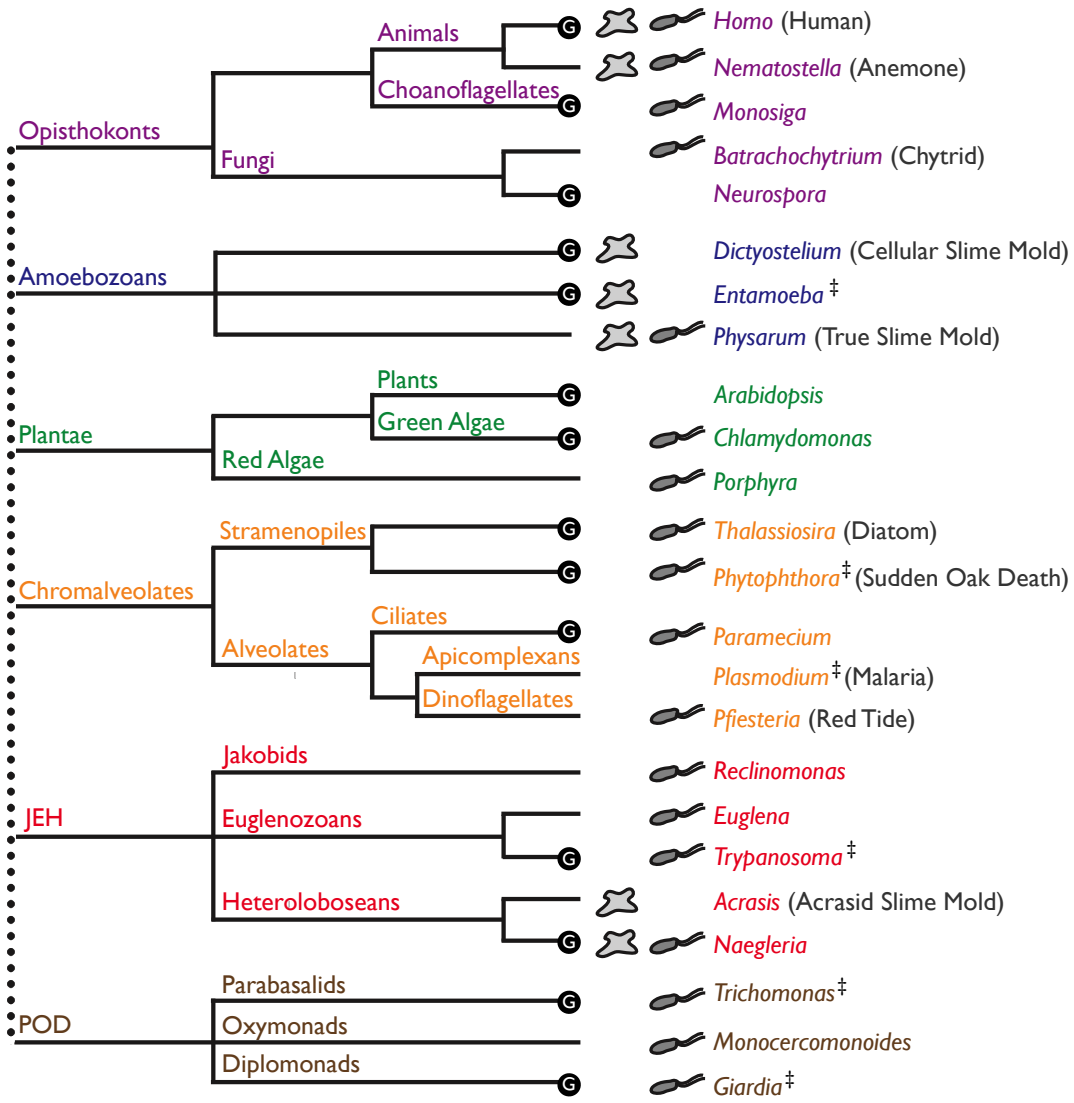


Figure 1.2. Evolutionary relationships between extant eukaryotes.

Consensus cladogram of selected eukaryotes, highlighting six major groups with widespread support in diverse molecular phylogenies (Burki et al., 2008; Rodriguez-Ezpeleta et al., 2007; Yoon et al., 2008). The dotted polytomy indicates uncertainty regarding the order of early branching events. Representative taxa are shown on the right, with glyphs indicating flagellar and/or actin-based amoeboid movement. Although commonly referred to as "amoeboid", *Trichomonas* does not undergo amoeboid locomotion. The inset depicts three contending hypotheses for the root. Root A: early divergence of unikonts and bikonts (Stechmann and Cavalier-Smith, 2002). Root B: the largely parasitic POD lineage branching first, followed by JEH (including *Naegleria*) (Ciccarelli et al., 2006). Root C: POD and JEH (together known as the "excavates") branching first (Supplemental Data). Highlighted in green are the branches connecting *Naegleria* to humans, with a black dot indicating their last common ancestor.



Amoeboid locomotion

Flagellar apparatus

G Genome compared

‡ Parasitic

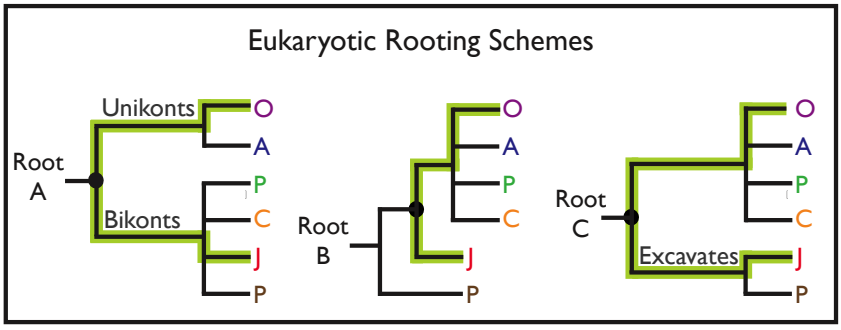


Figure 1.3. Overview of centriole assembly.

This schematic represents the events that occur during templated centriole duplication in human and *Paramecium* cells. These events are similar to those in other systems as well. Refer to Table 1.1 for further details on the proteins listed on the left-hand column.

Centriole Duplication

Cell Cycle Stage

G1

S

G2

M

Proteins required

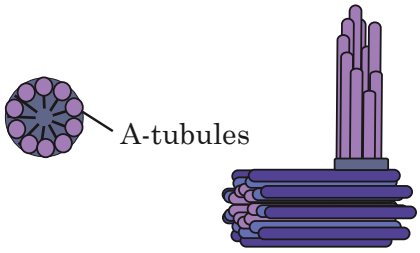
Generative Disc
Emerges from the side of the mother centriole

?

Cartwheel

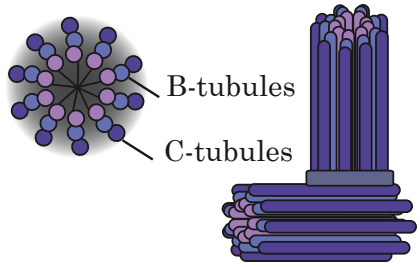
SAS-6
Bld10

Procentriole
A-tubules form



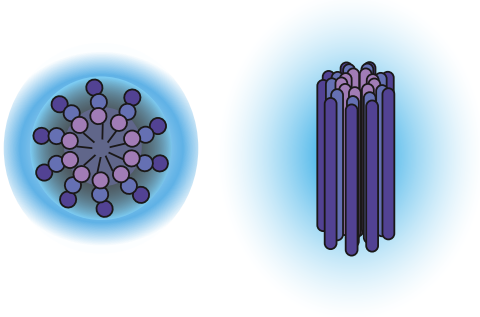
γ-tubulin?
α- and β-tubulin (modifications?)
POC5

Immature centriole
B and C tubules form



γ-tubulin?
α- and β-tubulin (modifications?)
ε-tubulin
δ-tubulin

Mature centriole
Daughter centriole disengages, and acquires pericentriolar material (PCM)



Pericentrin
Ninein
Cenexin

Chapter 2: Naegleria's Genome and Cytoskeletal Components

Abstract

Flagellar and amoeboid motility are the two most common forms of locomotion used by eukaryotes. *Naegleria gruberi* is known for its remarkable ability to metamorphose from amoebae into flagellates. By identifying conserved cytoskeletal genes in the newly sequenced *Naegleria* genome, we predict that *Naegleria* has a canonical actin and microtubule cytoskeletal repetoire. Further, we use phylogenetic profiling to identify protein families conserved only in eukaryotes with flagellar motility (Flagellar-Motility associated genes; FMs) or amoeboid motility (Amoeboid-Motility associated genes; AMs). *Naegleria*'s 182 FM gene families are consistent with typical eukaryotic flagellar function and structure, and also include 36 novel flagellar-associated genes. The 63 AMs include genes known to be involved in amoeboid motility, membrane differentiation, and 19 novel genes.

Introduction

To date, many eukaryotic genome sequencing efforts have focused on opisthokonts (a monophyletic group including animals and fungi) and plants, as well as obligate parasitic protists (which tend to be genomically streamlined), although an increasing number of free living protists (e.g., *Dictyostelium* (Eichinger et al., 2005), *Thalassiosira* (Armbrust et al., 2004), *Tetrahyemena* (Eisen et al., 2006), *Paramecium* (Aury et al., 2006), *Chlamydomonas* (Merchant et al., 2007)) are being sequenced.

The genome sequence of *Naegleria gruberi*, the first from a free-living member of a major eukaryotic group best known for its parasitic members (the Trypanosomatids), significantly broadens the phylogenetic coverage of eukaryotic genomes. Through comparison with other eukaryotic sequences we can infer features of ancestral eukaryotes as well as better understand *Naegleria*'s remarkable versatility. The published analysis of this genome (Fritz-Laylin et al., 2010) substantially extends the idea that early eukaryotes possessed complex trafficking, cytoskeletal, sexual, metabolic, signaling, and regulatory modules (Dacks and Field, 2007; Eichinger et al., 2005; Merchant et al., 2007). It also emphasizes subsequent losses, particularly in parasitic lineages. In this study we also identified a set of genes associated with amoeboid motility, and a surprising capacity for both aerobic and anaerobic metabolism. This manuscript was the result of a large collaboration (involving twenty-four scientists). As the co-lead author, I was heavily involved in writing and supervising the entire project. In

addition to this managerial role, focused largely on *Naegleria*'s cytoskeletal gene repertoire. As the genome project covered many areas of biology (including metabolism, signaling cascades, and eukaryotic evolution) I will focus on our analysis of *Naegleria*'s cytoskeletal gene repetoire as it has implications for the remainder of this dissertation.

Results

Summary of *Naegleria* gene models:

We assembled the 41 million base pair *N. gruberi* genome from ~8-fold redundant coverage of random paired-end shotgun sequence using genomic DNA prepared from an axenic, asexual culture of the NEG-M strain (ATCC 30224) (Fulton, 1974). *Naegleria* has at least twelve chromosomes (Figure 2.1). In addition to the nuclear genome, NEG-M has ~4,000 copies of a sequenced extrachromosomal plasmid that encodes rDNA (Clark and Cross, 1987; Maruyama and Nozaki, 2007), and a 50 kb mitochondrial genome (GenBank AF288092).

We predicted 15,727 protein coding genes spanning 57.8% of the genome by combining *ab initio* and homology-based methods with 32,811 EST sequences (Table 1). The assembly accounts for over 99% of the ESTs, affirming its near completeness. Nearly two-thirds (10,095) of the predicted genes are supported by EST, homology, and/or Pfam evidence (Figure 2.2). The remaining 5,632 genes may be novel, diverged, poorly-predicted, or have low expression.

Identification of *Naegleria*'s cytoskeletal gene homologs:

Naegleria contains two potentially autonomous microtubule cytoskeletons (mitotic and flagellar (Fulton, 1970)), as well as an extensive actin cytoskeleton. To determine if these structures are likely formed from canonical proteins, known microtubule and actin cytoskeleton genes were identified in the *Naegleria* genome by manual searches. This analysis revealed that *Naegleria*'s genome contains almost all well-conserved actin and microtubule components (Tables 2.2 and 2.3, respectively).

Of particular interest was the classification of *Naegleria*'s thirty tubulin genes. The phylogenetic classification of subfamilies (alpha through eta) is based on previously published annotations for non-*Naegleria* sequences, and supported by bi-directional BLAST searches for *Naegleria* sequences (Figure 2.3). As expected and based on the wide phylogenetic distribution of these proteins in flagellate organisms, the *Naegleria* genome contains homologs of alpha, beta, gamma, delta, and epsilon tubulin. *Naegleria* does not appear to have a homolog of zeta tubulin (Vaughan et al., 2000), suggesting that this tubulin family member is unique to the Trypanosomatids. However, based on bi-directional BLAST searches *Naegleria* has

a homolog of eta tubulin, which has been shown to be involved in basal body assembly (Ruiz et al., 2000) and is also found in *Chlamydomonas reinhardtii*, *Paramecium tetraurelia,* and possibly *Xenopus laevis* (its "cryptic tubulin" clusters with this group) (Dutcher, 2001; McKean et al., 2001). *Naegleria* also seems to have divergent clades of alpha and beta tubulins (two clades each). One such cluster also contains the *Naegleria* mitotic tubulin (Chung et al., 2002).

Prediction of flagellar motility (FM) and amoeboid motility (AM) gene catalogs:

The presence of flagellar motility and actin-based amoeboid locomotion in lineages spanning likely eukaryotic roots suggests that the eukaryotic ancestor had both capacities (Cavalier-Smith, 2002; Fulton, 1970). Flagellar (and ciliary) motility is generated by interactions between microtubules and molecular motors, whereas in amoeboid locomotion, the growth of actin filaments "pushes" the cell membrane forward. By searching for genes present in organisms that possess each type of motility and missing from organisms that do not, we identified two sets of genes: Flagellar-Motility associated genes (FMs) and Amoeboid-Motility associated genes (AMs) (Figure 2.4).

FMs include orthologs of all categories of flagellar and 36 novel flagellar-associated genes (Table 2.4). Two complexes of Intraflagellar transport (IFT) proteins move components within the flagella. All conserved IFT components were present, apart from the two retrograde and anterograde motors kinesin-2, and DHC1b (dynein heavy chain) respectively. Although present in the genome, these proteins were likely missed by the FM analysis as is often impossible to assign orthologs within large gene families such as kinesins and dyneins. The proteins required for flagellar beating, the outer and inner dynein arms, and the radial spoke proteins (RSP3 and RSP4), are included in the FM gene set as are cannonical basal body proteins and axonemal proteins such as RIB72, RIB43a, MBO2 and DIP13. Our analysis of the FM proteins is therefore consistent with *Naegleria* having typical eukaryotic flagella proteins, as well as typical structure.

The actin cytoskeleton enables amoeboid motility and diverse cellular processes including cytokinesis, endocytosis, and maintenance of cell morphology and polarity. We identified 63 gene families (AMs) found only in organisms with cells capable of amoeboid locomotion (Table 2.5). By definition, AMs do not include actin, Arp2/3 (which nucleates actin filaments) or other general actin cytoskeletal components, since these are found across eukaryotes regardless of their capacity for amoeboid locomotion. Nineteen AMs have unknown function, but are strongly implicated in actin-based motility (Table 2.5).

The AMs include several genes thought to keep pseudopod actin filaments dense, highly branched, and properly positioned. For example, the Arp2/3 activator WASH (AM5) is proposed to activate actin filament formation in pseudopodia

(Linardopoulou et al., 2007). The actin binding protein twinfilin (AM4) affects the relative sizes of functionally distinct pseudopodial subcompartments (Iwasa and Mullins, 2007). Filamin (AM3) stabilizes the three-dimensional actin networks necessary for amoeboid locomotion (Flanagan et al., 2001). Drebrin/ABP1 (AM2) aids in membrane attachment of actin filaments during endocytosis in yeast (Toret and Drubin, 2006), and could also function in cell migration (Peitsch et al., 2006; Song et al., 2008). The inclusion of both twinfilin and drebrin/ABP1 in the AMs argues that the actin patches formed during yeast endocytosis could have evolutionary origins in amoeboid motility.

Our analysis also suggests a role for the lipid sphingomyelin in amoeboid motility. AMs include a sphingomyelin-synthase-related protein (AM16) and Saposin-B-like proteins (AM17) that activate sphingomyelinase. (Sphingomyelinase is not an AM because it is found in the non-amoeboid *Paramecium* (Figure 2.4).) Sphingomyelin itself is enriched in pseudopodia (Jandak et al., 1990) and thus may contribute to motility via structural differentiation of the membrane, or as a second messenger in signalling pathways, as seen in human cells.

Discussion

Using both manual searches and phylogenetic profiling, we found that *Naegleria*'s repertoire of microtubule components is consistent with a canonical eukaryotic microtubule cytoskeleton. In particular, *Naegleria* seems to have the genes required for a typical flagellar and basal body structures, despite making these motile structures very quickly (within 90 minutes) and not maintaining them for more than two cell cycles (*Naegleria* return to amoeboid motility within three hours (Fulton, 1970)). The finding of seemingly *Naegleria*-specific alpha and beta tubulin subtypes suggests that *Naegleria* may have unique microtubule structures (e.g. the mitotic spindle may be composed of divergent tubulin filaments) and warrants further study.

The 63 AMs make up the first catalog of genes suggested to modulate a cell's actin cytoskeleton to achieve amoeboid motility. Included in this set are two hallmark amoeboid motility genes (twinfilin and profilin). The inclusion of these genes in the AMs indicates the other genes (particularly the 19 completely novel genes) are likely a rich source of unexplored biology relating to amoeboid motility.

Materials and methods

Strains

High quality genomic DNA was prepared from an axenic culture of amoebae of *Naegleria gruberi* strain NEG-M (ATCC 30224) (Fulton, 1974), which was derived from clonal strain NEG (Fulton, 1970) as a clone able to grow in simplified axenic

media. The amoebae were grown axenically in suspension in M7 medium (Fulton, 1974) from frozen stocks, and DNA was prepared from cells using Qiagen Genomic DNA Kit (Qiagen, USA).

Whole genome shotgun sequencing and sequence assembly

The initial sequence data set was generated from whole-genome shotgun sequencing (Weber and Myers, 1997) of four libraries. We used one library with an insert size of 2-3 kb (BCCH), one with an insert size of 6-8 kb (BCCI) and two fosmid libraries with insert sizes of 35-40 kb (BCCN, BGAG). We obtained reads as follows: 220,222 reads from the 2-3 kb insert libraries comprising 245 Mb of raw sequence, 261,984 reads from the 6-8 kb insert libraries comprising 263 Mb of raw sequence, and 52,608 reads from the 35-40 kb insert libraries comprising 54 Mb of raw sequence. The reads were screened for vector sequence using Cross_match (Ewing et al., 1998) and trimmed for vector and low quality sequences. Reads shorter than 100 bases after trimming were excluded from the assembly. This reduced the data set to 182,658 reads from the 2-3 kb insert libraries comprising 132 Mb of raw sequence, 245,457 reads from the 6-8 kb insert libraries comprising 193 Mb of raw sequence, and 43,514 reads from the 35-40 kb insert libraries comprising 26 Mb of raw sequence.

The trimmed read sequences were assembled using release 2.9 of JAZZ (Aparicio et al., 2002). A word size of 13 was used for seeding alignments between reads, with a minimum of 10 shared words required before an alignment between two reads would be attempted. The unhashability threshold was set to 50, preventing words present in the data set in more than 50 copies from being used to seed alignments. A mismatch penalty of -30.0 was used, which will tend to assemble together sequences that are more than about 97% identical. The genome size and sequence depth were initially estimated to be 35 Mb and 8.0 ×, respectively. The initial assembly contained 44.8 Mb of scaffold sequence, of which 5.9 Mb (13.1%) was gaps. There were 2,868 scaffolds, with a scaffold N/L50 of 38/384.3 Kb, and a contig N/L50 of 77/148.6 Kb. The assembly was then filtered to remove scaffolds < 1kb long as well as redundant scaffolds, where redundancy was defined as those scaffolds shorter than 5kb long with a greater than 80% identity to another scaffold greater than 5kb long.

After excluding redundant and short scaffolds, 41.1 Mb remained, of which 4.7 Mb (11.5%) was gaps. The filtered assembly contained 813 scaffolds, with a scaffold N/L50 of 33/401.6 kb, and a contig N/L50 of 69/157.7 kb. The sequence depth derived from the assembly was 8.6 ± 0.1.

To estimate the completeness of the assembly, the consensus sequences from clustering a set of 28,768 ESTs were BLAT-aligned (with default parameters) to the unassembled trimmed data set, as well as the assembly itself. 28,486 ESTs (99.0%)

were more than 80% covered by the unassembled data and 28,502 ESTs (99.1%) had hits to the assembly.

Mitochondrial genome sequence (GenBank AF288092) was used to identify the 18 scaffolds belonging to the organelle genome; this sequence is available from the JGI *Naegleria* Genome Portal (http://www.jgi.doe.gov/naegleria/).

cDNA library construction and EST sequencing

EST sequences were made from two samples: 1) asynchronous cells where some were differentiating into flagellates and others back into amoebae and 2) confluent amoeba grown in tissue culture flasks. Poly-A+ RNA was isolated from total RNA for each sample using the Absolutely mRNA Purification kit and manufacturer's instructions (Stratagene, La Jolla, CA). cDNA synthesis and cloning was a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen). 1-2 μg of poly A+ RNA, reverse transcriptase SuperScript II (Invitrogen) and oligo dT-NotI primer:
5'- GACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT -3'
were used to synthesize first strand cDNA. Second strand synthesis was performed with E. coli DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. A SalI adaptor (5'- TCGACCCACGCGTCCG and 5'- CGGACGCGTGGG) was ligated to the cDNA, digested with NotI (NEB), and subsequently size selected by gel electrophoresis (using 1.1% agarose). Two size ranges of cDNA (0.6 - 2.0 kb.p. and > 2 kb.p.) were cut out of the gel for the amoeba sample and one size range (0.6 -2.0 kb.p.) for the flagellate sample. They were directionally ligated into the SalI and NotI digested vector pMCL200_cDNA. The ligation product was transformed into ElectroMAX T1 DH10B cells (Invitrogen).

Library quality was first assessed by randomly selecting 24 clones and PCR amplifying the cDNA inserts with the primers M13-F (GTAAAACGACGGCCAGT) and M13-R (AGGAAACAGCTATGACCAT). The number of clones without inserts was determined and 384 clones for each library were picked, inoculated into 384 well plates (Nunc) and grown for 18 hours at 37°C. Each clone was amplified using RCA then the 5' and 3' ends of each insert was sequenced using vector specific primers (forward (FW): 5'- ATTTAGGTGACACTATAGAA and reverse (RV) 5' – TAATACGACTCACTATAGGG) and Big Dye chemistry (Applied Biosystems). 44,544 EST reads were attempted from the 2 samples.

The JGI EST Pipeline begins with the cleanup of DNA sequences derived from the 5' and 3' end reads from a library of cDNA clones. The Phred software (Ewing and Green, 1998; Ewing et al., 1998) is used to call the bases and generate quality scores. Vector, linker, adapter, poly-A/T, and other artifact sequences are removed using Cross_match (Ewing and Green, 1998; Ewing et al., 1998), and an internally

developed short pattern finder. Low quality regions of the read are identified using internally developed software, which masks regions with a combined quality score of less than 15. The longest high quality region of each read is used as the EST. ESTs shorter than 150 bp were removed from the data set. ESTs containing common contaminants such as *E. coli*, common vectors, and sequencing standards were also removed from the data set. There were 38,211 EST sequences left after filtering.

EST clustering was performed on 38,282 trimmed, high-quality ESTs (the 38,211 filtered and trimmed JGI EST sequences combined with the JGI ESTs combined with 71 EST sequences downloaded from GenBank (Benson et al., 2009) by making all-by-all pairwise alignments with MALIGN (Sobel and Martinez, 1986). ESTs sharing an alignment of at least 98% identity, and 150 bp overlap are assigned to the same cluster. These are relatively strict clustering cutoffs, and are intended to avoid placing divergent members of gene families in the same cluster. However, this could also have the effect of separating splice variants into different clusters. Optionally, ESTs that do not share alignments are assigned to the same cluster, if they are derived from the same cDNA clone. We made 4,873 EST clusters.

EST cluster consensus sequences were generated by running Phrap (Ewing and Green, 1998) on the ESTs comprising each cluster. All alignments generated by MALIGN {Sobel, 1986 #351 are restricted such that they will always extend to within a few bases of the ends of both ESTs. Therefore, each cluster looks more like a 'tiling path' across the gene, which matches well with the genome based assumptions underlying the Phrap algorithm. Additional improvements were made to the phrap assemblies by using the 'forcelevel 4' option, which decreases the chances of generating multiple consensi for a single cluster, where the consensi differ only by sequencing errors.

Generation of gene models and annotation

The genome assembly was annotated using the JGI Annotation Pipeline. First the 784 *N. gruberi* v.1 scaffolds were masked using RepeatMasker {Smit, 1996-20a04 #289} and a custom repeat library of 123 putative transposable element-like sequences. Next, the EST and full-length cDNAs were clustered into 4,873 consensus sequences (see above) and aligned to the scaffolds with BLAT (Kent, 2002). Gene models were predicted using the following methods: i) *ab initio* (FGENESH (Salamov and Solovyev, 2000); ii) homology-based (FGENESH+ (Salamov and Solovyev, 2000) and Genewise (Birney et al., 2004), with both of these tools seeded by Blastx (Altschul et al., 1990) alignments of sequences from the 'nr' database from the National Center for Biotechnology Information (NCBI, Genbank) (Benson et al., 2009) to the *Naegleria* genome); and iii) mapping *N. gruberi* EST cluster consensus sequences to the genome (EST_map; http://www.softberry.com/).

Truncated Genewise models were extended where possible to start and stop codons in the surounding genome sequence. EST clusters, mapped to the genome with BLAT (Kent, 2002) were used to extend, verify, and complete the predicted gene models. The resulting set of models was then filtered, based on a scoring scheme which maximises completeness, length, EST support, and homology support, to produce a single gene model at each locus, and predicting a total of 15,753 models.

Only 13% of these gene models were seeded by sequence alignments with proteins in the nr database at NCBI (Benson et al., 2009) or *N. gruberi* EST cluster consensus sequences, while 86% were *ab initio* predictions. Complete models with start and stop codons comprise 93% of the predicted genes. 30% are consistent with ESTs and 74% align with proteins in the nr database at GenBank (Benson et al., 2009).

Protein function predictions were made for all predicted gene models using the following collection of software tools: SignalP (http://www.cbs.dtu.dk/services/SignalP/), TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), InterProScan (http://www.ebi.ac.uk/interpro/ (Quevillon et al., 2005)), and hardware-accelerated double-affine Smith-Waterman alignments (http://www.timelogic.com/decypher_sw.html) against SwissProt (http://www.expasy.org/sprot/), KEGG (http://www.genome.jp/kegg/), and KOG (http://www.ncbi.nlm.nih.gov/COG/). Finally, KEGG hits were used to map EC numbers (http://www.expasy.org/enzyme/), and Interpro and SwissProt hits were used to map GO terms (http://www.geneontology.org/).
Nearly half (45%) of the gene models have Pfam (Finn et al., 2008) domain annotations (Table S13). The average gene length is 1.65 kb.p. The average protein length is 492 aa. We predicted that 3,514 proteins (22%) possess a leader peptide, 3,439 proteins (22%) possess at least one transmembrane domain, and 2060 (13%) possess both.

Web-based interactive editing tools available through the JGI genome portal (http://www.jgi.doe.gov/naegleria/) were used to manually curate the automated annotations in three ways: i) to assess and if necessary correct, predicted gene structures. ii) to assign gene functions and report supporting evidence, and iii) to create, if necessary, new gene structures.

On 19 July 2007, the manually-annotated gene set was frozen to make a catalog. This set of 15,776 transcripts encoded by 15,727 genetic loci was used for all analyses in this paper. In a few cases, as noted in the main text, manual improvements to gene models were needed before detailed analysis was possible. As of May 15, 2008, 4,016 genes (25%) have been manually curated. All annotations, may be viewed at a JGI portal (http://www.jgi.doe.gov/naegleria/).

Construction of protein families

As a pre-requisite to comparing the protein-coding potential of *Naegleria* to other organisms at the whole-genome scale, we constructed families of homologous proteins from all sequences that are found in both *Naegleria* and at least one other species from a wide a range of eukaryotes. Errors in gene prediction and large-scale species-specific gene losses can cause problems building protein families and drawing phylogenetic inferences from the families. To mitigate this, we chose our range of organisms to ensure that at least two species from every major eukaryotic group with genome sequence were included. Where several closely-related genome sequences were available, we chose manually- or well-annotated species to represent clades of interest. We also included a representative photosynthetic prokaryote, *Prochlorococcus marinus*.

Our goal was to generate families of protein sequences such that there is one family for each protein in the common ancestor of all the species which have proteins in the family, and that all the extant proteins descended from the ancestral protein are in the family. The shared ancestry (homology) of family members should enable us to infer shared function, allowing functional annotations to be transferred among family members.

To create protein families, we first blasted [WU-BLASTP 2.0MP-WashU (Altschul et al., 1990)] each of the 15,727 protein sequences in *Naegleria* to all protein sequences in the animals human (Ensembl; Lander et al., 2001; Venter et al., 2001) and *Trichoplax adherens* (Srivastava et al., 2008); the choanoflagellate *Monosiga brevicollis* (King et al., 2008); the fungus *Neurospora crassa* (assembly v7.0; annotation v3.0, http://fungal.genome.duke.edu); the amoebae *Dictyostelium discoideum* (Eichinger et al., 2005) and *Entamoeba histolytica* (TIGR, http://www.tigr.org/tdb/e2k1/eha1/); the land plants *Arabidopsis thaliana* (Initiative, 2000) and *Physcomitrella patens* (assembly v.1) (Rensing et al., 2008); the green alga *Chlamydomonas reinhardtii* (Benson et al., 2009; Merchant et al., 2007); the oomycete *Phytophthora ramorum* (v1) (Joint Genome Institute); the diatoms *Thalassiosira pseudonana* (assembly v3.0 (Armbrust et al., 2004; Joint Genome Institute)) and *Phaeodactylum tricornutum* (assembly v2.0) (Joint Genome Institute); the alveolate *Paramecium tetraurelia* (Paramecium DB release date 28-MCH-2007; http://paramecium.cgm.cnrs-gif.fr/); the euglenozoan *Trypanosoma brucei* (v4 genome; http://www.genedb.org/genedb/tryp/); the diplomonad *Giardia lamblia* (GMOD; http://www.giardiadb.org/giardiadb/); the parabasalid *Trichomonas vaginalis* (TIGR, http://www.tigr.org/tdb/e2k1/tvg/); and the cyanobacterium *Prochlorococcus marinus* strain MIT9313 (Joint Genome Institute).

Assignment of orthology was determined by the presence of a mutual best hit between two proteins, based on score with a cutoff of E-value < 1E-10. In creating individual protein families, we first generated all possible ortholog pairs consisting

of one *Naegleria* protein and a protein from another organism. Next, paralogs that met certain criteria were added to each pair of proteins. A paralog from a given organism was added if its p-dist (defined as 1 - the fraction of identical aligning amino acids in the proteins) was less than a certain fraction of the p-dist between the two orthologs in the pair. The fractions were chosen to be 0.5 for pairs of organisms involving two eukaryotes and 0.1 for *Naegleria* and the prokaryotic cyanobacterium. Two considerations led to the choice of these values. In order to assign function correctly, we wanted to include only 'in-paralogs' (i.e. paralogs that had duplicated after speciation) (Remm et al., 2001). Secondly, we previously determined that higher (less stringent) values led to the generation of protein families with >22,000 members that could not be analyzed further (Merchant et al., 2007). As a final step, all pair-wise families of two orthologs plus paralogs were merged if they contained the same *Naegleria* protein. This created 5,115 families of homologous proteins, with 5,107 families containing proteins from *Naegleria* and at least one other eukaryote and 8 families restricted to *Naegleria* and the cyanobacterium *Prochlorococcus*. Each individual family consists of one or more *Naegleria* paralog(s), mutual best hits to proteins of other species (orthologs) and any paralogs in each of those species. The set of protein families was used in subsequent phylogenetic profiling of proteins associated with amoeboid motility (AMs) or flagellar motility (FMs) (see below). To accomplish this, we built a software tool that allowed us to search for protein families containing any desired combination of species. We call the search results a 'cut' (see below) as it represents a phylogenetic slice through the collection of protein families.

The random gene duplication, subsequent divergence and loss that accompanies the evolution of gene families means that it is challenging and sometimes impossible to precisely assign orthology and paralogy between genes. The problem gets more difficult for larger families, which are statistically more likely to undergo mutations and old families that have had longer to diverge. As a result, mutual best hit relationships between sequences may not exist, preventing family construction, or may not be between correct proteins, leading to inclusion of non-homologous proteins in families.

Generation of Flagellar Motility-associated proteins (FMs)

Genes associated with flagella function have been identified by phylogenetic profiling (Avidor-Reiss et al., 2004; Li et al., 2004; Merchant et al., 2007). We generated a list of proteins associated with flagellar function by searching the *Naegleria* protein families (see above) for those that contain proteins from organisms with flagella (*Naegleria*, *Chlamydomonas*, and human) and none from organisms lacking flagella (*Dictyostelium*, *Neurospora*, *Arabidopsis* and *Prochlorococcus*). This analysis resulted in 182 *Naegleria* proteins in 173 families (the 'FlagellateCut', Table 2.4). The proteins belonging to this 'cut' were termed FMs (Flagellar Motility associated proteins).

We compared the proteins we had identified to a hand-curated list of 101 *Chlamydomonas* flagellar proteins that had been discovered by biochemical, genetic, and bioinformatic methods (Pazour et al., 2005). Of the 182 genes, 34 are in families containing a characterized *Chlamydomonas* flagellar protein, and an additional 59 are in a family with a *Chlamydomonas* flagellar proteome protein (Pazour et al., 2005). Thus, at least 51% of the FlagellateCut genes are likely to encode proteins that localize to flagella.

Generation of Amoeboid Motility-associated proteins (AMs)

We used phylogenetic profiling to generate a catalog of proteins associated with amoeboid motility. To our knowledge, this is the first time such a catalog has been compiled. In an analogous analysis to the FlagellateCut (see above), we searched the *Naegleria* protein families (see above) for those that contain proteins from organisms that undergo amoeboid movement [*Naegleria*, human, and at least one Amoebozoan (*Dictyostelium* or *Entamoeba*)], but not in organisms that have no amoeboid movement [*Prochlorococcus*, Arabidopsis, *Physcomitrella*, Diatom, *Paramecium*, Trypanosome, *Giardia*, *Chlamydomonas*].

The search found 112 protein families containing 139 *Naegleria* proteins. 36 families contained proteins with homology (blastp E-value < 1E-10) to a protein in one or more non-amoeboid species from the list we had previously used to build the *Naegleria* protein familes, and these 36 families were excluded from the AM gene set. In addition, 13 families were removed because their members belong to very large protein families (containing ≥ 245 members) and we reasoned that difficulties in assigning correct orthology in families this large (see above) made them unlikely to be true representatives of the AmoebaCut. This left 63 AM protein families containing 67 *Naegleria* proteins (Table S10). There is no way to estimate the false positive rate for this computational analysis as no experimental catalog of AMs is available for comparison.

Although the POD member *Trichomonas* has been described as "amoeboid", it does not undergo amoeboid locomotion, and was not used to define AM protein families. However, *Trichomonas* does possess seven of the AMs (Table S10), suggesting most AMs are involved in cell locomotion, and not simply amoeboid-like morphology.

Identification of evolutionarily conserved cytoskeletal genes:

To identify potential *Naegleria* microtubule and actin cytoskeleton genes, the genome was searched (using BLASTP at the JGI genome portal (http://www.jgi.doe.gov/naegleria/) with genes from various eukaryotes (including *Chlamydomonas*, human, trypanosomes and yeast), as well as using Pfam domain annotations (Sonnhammer et al., 1998) and local BLAST (Altschul et al., 1990)

searches. To further define what classes of actin related proteins and tubulins *Naegleria*'s genome encodes, a phylogenetic tree was constructed for each protein family (data not shown). Cytoskeletal motors were also classified phylogenetically: kinesins, myosins, and dyneins (Data not shown). Information from these phylogenetic analyses were used for the classifaction of these large gene families into the types found in Table 2.2 and 2.3. Further details (including large scale phylogenies and phylogenetic methods) can be found in the *Naegleria* genome paper supplement (Fritz-Laylin et al., 2010).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A.*, et al.* (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science *297*, 1301-1310.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M.*, et al.* (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science *306*, 79-86.

Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N.*, et al.* (2006). Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature *444*, 171-178.

Avidor-Reiss, T., Maer, A.M., Koundakjian, E., Polyanovsky, A., Keil, T., Subramaniam, S., and Zuker, C.S. (2004). Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. Cell *117*, 527-539.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2009). GenBank. Nucleic Acids Res *37*, D26-31.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988-995.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol *52*, 297-354.

Chung, S., Cho, J., Cheon, H., Paik, S., and Lee, J. (2002). Cloning and characterization of a divergent alpha-tubulin that is expressed specifically in dividing amebae of Naegleria gruberi. Gene *293*, 77-86.

Clark, C.G., and Cross, G.A. (1987). rRNA genes of Naegleria gruberi are carried exclusively on a 14- kilobase-pair plasmid. Mol Cell Biol *7*, 3027-3031.

Dacks, J.B., and Field, M.C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. J Cell Sci *120*, 2977-2985.

Dutcher, S.K. (2001). The tubulin fraternity: alpha to eta. Curr Opin Cell Biol *13*, 49-54.

Eichinger, L., Pachebat, J.A., Glockner, G., Rajandream, M.A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., *et al.* (2005). The genome of the social amoeba Dictyostelium discoideum. Nature *435*, 43-57.

Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., *et al.* (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol *4*, e286.

Ensembl. http://www.ensembl.org/

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res *8*, 186-194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res *8*, 175-185.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., *et al.* (2008). The Pfam protein families database. Nucleic Acids Res *36*, D281-288.

Flanagan, L.A., Chou, J., Falet, H., Neujahr, R., Hartwig, J.H., and Stossel, T.P. (2001). Filamin A, the Arp2/3 complex, and the morphology and function of cortical actin filaments in human melanoma cells. J Cell Biol *155*, 511-517.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J., Carpenter, M.L., Field, M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J., *et al.* (2010). The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. Cell *In press*.

Fulton, C. (1970). Amebo-flagellates as research partners: The laboratory biology of *Naegleria* and *Tetramitus*. Methods Cell Physiol *4*, 341-476.

Fulton, C. (1974). Axenic cultivation of *Naegleria gruberi*. Requirement for methionine. Exp Cell Res *88*, 365-370.

Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796-815.

Iwasa, J.H., and Mullins, R.D. (2007). Spatial and temporal relationships between actin-filament nucleation, capping, and disassembly. Curr Biol *17*, 395-406.
Jandak, J., Li, X.L., Kessimian, N., and Steiner, M. (1990). Unequal distribution of membrane components between pseudopodia and cell bodies of platelets. Biochim Biophys Acta *1029*, 117-126.

Joint Genome Institute. http://www.jgi.doe.gov/

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.
King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I.*, et al.* (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature *451*, 783-788.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y., Teslovich, T.M., May-Simera, H., Li, H., Blacque, O.E., Li, L., Leitch, C.C.*, et al.* (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell *117*, 541-552.

Linardopoulou, E.V., Parghi, S.S., Friedman, C., Osborn, G.E., Parkhurst, S.M., and Trask, B.J. (2007). Human subtelomeric WASH genes encode a new subclass of the WASP family. PLoS Genet *3*, e237.

Maruyama, S., and Nozaki, H. (2007). Sequence and intranuclear location of the extrachromosomal rDNA plasmid of the amoebo-flagellate Naegleria gruberi. J Eukaryot Microbiol *54*, 333-337.

McKean, P.G., Vaughan, S., and Gull, K. (2001). The extended tubulin superfamily. J Cell Sci *114*, 2723-2733.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L.*, et al.* (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245-250.

Pazour, G.J., Agrin, N., Leszyk, J., and Witman, G.B. (2005). Proteomic analysis of a eukaryotic cilium. J Cell Biol *170*, 103-113.

Peitsch, W.K., Bulkescher, J., Spring, H., Hofmann, I., Goerdt, S., and Franke, W.W. (2006). Dynamics of the actin-binding protein drebrin in motile cells and definition of a juxtanuclear drebrin-enriched zone. Exp Cell Res *312*, 2605-2618.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res *33*, W116-120.

Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol *314*, 1041-1052.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y.*, et al.* (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64-69.

Ruiz, F., Krzywicka, A., Klotz, C., Keller, A., Cohen, J., Koll, F., Balavoine, G., and Beisson, J. (2000). The SM19 gene, required for duplication of basal bodies in Paramecium, encodes a novel tubulin, eta-tubulin. Curr Biol *10*, 1451-1454.

Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA. Genome Res *10*, 516-522.

Siripala, A.D., and Welch, M.D. (2007). SnapShot: actin regulators I. Cell *128*, 626.
Sobel, E., and Martinez, H.M. (1986). A multiple sequence alignment program. Nucleic Acids Res *14*, 363-374.

Song, M., Kojima, N., Hanamura, K., Sekino, Y., Inoue, H.K., Mikuni, M., and Shirao, T. (2008). Expression of drebrin E in migrating neuroblasts in adult rat brain: coincidence between drebrin E disappearance from cell body and cessation of migration. Neuroscience *152*, 670-682.

Sonnhammer, E.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucl Acids Res *26*, 320-322.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L.*, et al.* (2008). The Trichoplax genome and the nature of placozoans. Nature *454*, 955-960.

Toret, C.P., and Drubin, D.G. (2006). The budding yeast endocytic pathway. J Cell Sci *119*, 4585-4587.

Vaughan, S., Attwood, T., Navarro, M., Scott, V., McKean, P., and Gull, K. (2000). New tubulins in protozoal parasites. Curr Biol *10*, R258-259.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The Sequence of the Human Genome. Science *291*, 1304-1351.

Weber, J.L., and Myers, E.W. (1997). Human whole-genome shotgun sequencing. Genome Res *7*, 401-409.

Table 2.1: Genome statistics from *Naegleria gruberi* and selected species.

n.d. not determined.

Species	Genome Size (Mb)	No. chromo-somes	%GC	Protein coding loci	% coding	% genes w/ introns	Introns per gene	Median intron length (b.p.)
Naegleria	41	>=12	33	15,727	57.8	36	0.7	60
Human	2851	23	41	23,328	1.2	83	7.8	20,383
Neurospora	40	7	54	10,107	36.4	80	1.7	72
Dictyostelium	34	6	22	13,574	62.2	68	1.3	236
Arabidopsis	140.1	5	36	26,541	23.7	80	4.4	55
Chlamydomonas	121	17	64	14,516	16.3	91	7.4	174
Paramecium	72	>50	28	39,637	74.9	n.d.	n.d.	n.d.
T. brucei	26.1	>100	46	9,152	52.6	~0 (1 total)	n.d.	n.d.
Giardia	11.7	5	49	6,480	71.4	~0 (4 total)	n.d.	n.d.

Table 2.2: *Naegleria*'s actin cytoskeleton complement

To identify potential *Naegleria* actin cytoskeleton genes, the genome was searched (using BLASTP at the JGI genome portal, http://www.jgi.doe.gov/naegleria/) with genes from various eukaryotes (including *Chlamydomonas*, human, trypanosomes and yeast). "-" indicates a homolog was not found in the genome scaffolds (confirmed using tBLASTn against the scaffolds). Table is adapted, with permission, from Siripala et. al. (Siripala and Welch, 2007).

Protein Family	Biochemical Function	Naegleria Seqs	Representative Proteins H. sapiens	A. thaliana	S. cerevisiae
actin nucleators					
formins	associates with barbed ends, promotes nucleation	48740 79970 79890 79577 76205 54096 73250 62754 79644 78968 62347 59934 81446 52512	*hDia1,2,3, FHOD1,2,3,4, DAAM1,2, FMN1,2, FRL1,2,3, INF1,2, Delphilin	AtFH1-4,5,6-21	Bni1, Bnr1
Spir	nucleates unbranched filaments	-	Spir-1, Spir-2		
Arp2/3 complex	polymerizes Y-branched filament networks	50292 82653 65498	ARP2, ARP3, ARPC1A,1B,2,3,4,5,5L	ARP2, ARP3, ARPC1-5	Arp2, Arp3, Arc40,35,18,19,15
Arp2/3 complex regulators					
CARMIL	links capping protein, Arp2/3 and myosin I	-	CARMIL		
coronin	binds F-actin and Arp2/3 complex	56244	Coronin 1A,B,C; 2A,B; 7		Crn1
cortactin	activates/modulates Arp2/3 complex, binds F-actin	-	cortactin, HS1		
SCAR/WAVE family	activates Arp2/3 complex	58655 57003	SCAR1, SCAR2, SCAR3	AtSCAR 1,2,3,4	
WASP family	activates Arp2/3 complex	58361 69609	WASP*, N-WASP		Las17/Bee1
G-actin binding proteins					
CAP/Srv2	recycles actin monomers	56829 82381 62204	CAP1, CAP2	AtCAP1	Srv2
profilin	promotes actin monomer addition at barbed ends	34873 83254	profilin 1,2,3,4	PRF1-3 (vegetative), PRF4-5 (reproductive)	Pfy1
thymosin b4	sequesters G-actin	-	thymosin b4		
verprolin/WIP	binds actin monomers, WASP	-	WIP, WIRE/WICH, CR16		Vrp1
F-actin binding proteins					
Ena/VASP family	prevents capping at filament ends, recruits profilin	-	Mena, VASP, EVL		
MIM	binds both G- and F-actin, may regulate NPFs	-	MIM		
pointed end capping proteins					
tropomodulin/Tmod	caps pointed ends, binds tropomyosin	-	Tmod1,2,3,4, Lmod1,2,3		
barbed end capping proteins					
AIP1	promotes depolymerization by ADF/cofilin	80296	AIP1/WDR1	AtAIP1	Aip1

capG	caps barbed ends (calcium-dependent)	-	CapG		
capZ/capping protein/b-actinin	heterodimer caps barbed ends	34423(alpha) 80201 (beta)	CapZ (A1-3, B1-2)	AtCPA, AtCPB	Cap1,Cap2
Eps8	caps barbed ends, involved in Rac signalling	-	Eps8, Eps8L1,2,3		
actin depolymerizing/severing proteins					
ADF/cofilin	promotes actin filament disassembly	59962	cofilin1, cofilin2, ADF	AtADF1-12	Cof1
gelsolin/villin superfamily	severs actin filaments	74838 78996 57510 79010	gelsolin, adseverin/scinderin, villin, advillin, supervillin, flightless I	VLN1-5	
twinfilin	inhibits actin nucleotide exchange, severs actin	58328	twinfilin-1/A6, twinfilin-2		Twf1
actin bundling/crosslinking proteins					
a-actinin	connects and organizes actin filaments	56285	a-actinin1-4		
EPLIN	stabilizes and bundles actin filaments	-	EPLIN		
espin	forms parallel actin bundles	-	espin*		
fascin	bundles actin filaments	-	fascin 1,2,3		
filamin	orthogonally crosslinks actin filaments	80016 62518	filamin A*,B*,C		
fimbrin/plastin	bundles actin filaments	83109 65679	plastin 1-3 (I,L,T)	AtFIM1	Sac6
Endocytosis and Exocytosis					
ABP1/drebrin		82236 60013	mABP1, drebrin, drebrin-like		Abp1
EPS15		57422	EPS15	u: Q56WL2	Pan1
HIP1R		-	HIP1R		Sla2
synapsin		66720 52582	synapsin Ia,Ib,IIa,IIb,III	Synapsin	SNN-1
Plasma Membrane Association					
anillin		-	anillin		
annexins		-	annexin A1-11, 13 (actin binding: 1,2,6)	ANN1-8	
ERM proteins		64748	ezrin, radixin, moesin		
MARCKS		-	MARCKS, MRP/MACMARCKS/F52		
merlin		-	merlin/NF2*		

protein 4.1		-	4.1R,G,N,B		
spectrin		56285	a-spectrin (1-2), b-spectrin(1-4), bheavy-spectrin/karst		
Motor Proteins and Regulators					
caldesmon		-	caldesmon		
calponin		81173 76225 82066 (Calponin-homology domain containing proteins)	calponin 1-3		Scp1
myosin family		Myosin I: 59818, 64166, 78120, 33960, 77756, 67842 Myosin II: 35518, 81378, 79699 Myosin XXI: 62844, 63897	Myosin types 1,2,3,5*,6*,7*,9,10,15*,16,18, 19	Myosin types 8,11	Myosin types 1,2,5
tropomyosin		-	tropomyosin 1,2*,3*,4	u: Q8LE98	Tpm1, Tpm2
troponin		-	troponin T1*,2*,3; C1,2; I1,2,3*		
Muscle Components					
myopalladin		-	myopalladin, myotilin*		
nebulin		-	nebulin*, nebulette*, NRAP		
titin/connectin		-	titin/connectin*		

Identified Cellular Role

membrane trafficking and phagocytosis

cytokinesis

cell surface organization and dynamics

cell adhesion

multiple functions

other/unknown

u: UNIPROT database accession number

* - mutation linked to human disease

Table 2.3: *Naegleria*'s microtubule cytoskeleton complement

To identify potential *Naegleria* microtubule cytoskeleton genes, the genome was searched (using BLASTP at the JGI genome portal (http://www.jgi.doe.gov/naegleria/) with genes from various eukaryotes (including *Chlamydomonas*, human, trypanosomes and yeast). "nf" indicates a homolog was not found in the genome scaffolds (confirmed using TBLASTN against the scaffolds). Putative homologs with partial or unclear similarity are indicated with a question mark.

Tubulins	
a-tubulin (15)	60961 71268 88209 88206 72133 58607 39221 45327 56065 51830 65597 56236 88213 53284 55745
a-tubulin-like (incomplete model)	88210
b-tubulin (10)	55423 44804 53354 56391 55900 44710 44840 83350 88212 55748
b-tubulin-like (incomplete model)	88211
gamma-tubulin	56069
epsilon-tubulin/BLD2	44774
delta-tubulin/UNI3	69007
zeta-tubulin	nf
eta-tubulin/SM19	65724
kappa-tubulin	nf
Microtubule Nucleation:	
GCP2/Spc97p	79590
GCP3/Spc98p	434
GCP4	45190
GCP5	50986
GCP6	61337
Microtubule minus-end Organization	
Centrin	56351 44488
Pericentrin	nf
Spc72p	nf
Spc42p	nf
Spc29p	nf
Kar1p	nf
SUN domain protein	69660
Microtubule Capping/Severing	
EB1/Bim1p	44546 65633
CLIP170/Tip1p	nf
APC (adenomatous polyposis coli protein)	nf (only in metazoa)
Tea1p	nf
Katanin p60	63871 83220
Katanin p80	72175?
CAP-Gly domain	81169 51258
Microtubule-associated proteins	
MAP1A	nf (only in metazoa)
MAP1B/MAP5	nf (only in metazoa)
MAP2/MAP4/Tau	nf (only in metazoa)
MAP6	nf (only in metazoa)
TPX2(targeting protein for Xklp2)	45723
MAP215/Dis1 family	49612 62456
Asp	74233 48956
ORBIT/MAST/CLASP family	88225
Microtubule Capping/Severing	
EB1/Bim1p	44546 65633
CLIP170/Tip1p	nf
Motor proteins (Axonemal Dyneins Below)	
kinesin-1	33335 72809 34430
kinesin-2	31023 76756
kinesin-3	80962 31717 80350 3036 39261 68618 31878
kinesin-4	nf

kinesin-5	31750
kinesin-6	69726
kinesin-7	69788 74311 56509
kinesin-8	80478
kinesin-9	64648 31675
kinesin-10	nf
kinesin-11	nf
kinesin-12	nf
kinesin-13	31834 63167 32337
kinesin-14A	32240 73429
kinesin-14B	32757 30037 60999
kinesin-15	79561 2164
kinesin-16	71374
novel kinesins	54872 82323 29130 33324 71914 5496 78506 69050 5502 71346
Cytoplasmic Dynein motor chain	46538
Unclassified Dynein motor chain	61303
Tubulin-modifying enzymes	
Tubulin deacetylase HDAC6	nf
Tubulin tyrosine ligase-like	80835 33283 78886 29319 36744 67150
Intraflagellar Transport	
FLA10, Kinesin-II Motor Protein	31023 76756
DHC1b, Cytoplasmic Dynein Heavy Chain 1b	55825
IFT57, Intraflagellar Transport Protein 57	45002
IFT72/74, Intraflagellar Transport Protein 72 and 74	35518?
IFT20, Intraflagellar Transport Protein 20	62977
IFT52/BLD1, Intraflagellar Transport Protein 52	77715
IFT80, Intraflagellar Transport protein 80	77945 29769
IFT81, Intraflagellar Transport Protein 81	64053 79996
IFT88, Intraflagellar Transport Protein 88	63280
IFT122, Intraflagellar Transport Protein 122	71180
IFT140, Intraflagellar Transport Protein 140	48798
D1bLIC, Dynein 1b Light Intermediate Chain	80259
IFT172, Intraflagellar Transport Protein 172	63764 31150
FLA8 Kinesin II Motor Protein	63939
KAP, Kinesin II associated Protein	79669
Outer Dynein Arm	
Outer Dynein Arm Heavy Chain	55628 38988 30143 39535
Outer Dynein Arm Intermediate Chain	78637 60431 79232
Outer Dynein Arm Light Chain	74922 54720 30532
Outer Dynein Arm Docking Complex 1	nf
Outer Dynein Arm Docking Complex 2	81548
Outer Dynein Arm Docking Complex 3	nf
Inner Dynein Arms	
Inner Arm Dynein Heavy Chain	81845 78559 37059 60802 36107 83317
Inner Dynein Arm Intermediate Chain	57343 63304
Inner Dynein Arm Light Chain p28	82719
Dynein Light Chain Tctex1	29177 30532 32371
Dynein Regulatory Complex	
PF2, Dynein Regulatory Complex Protein	88226
Radial Spoke	
RSP3, Radial Spoke Protein 3	64930
Radial Spoke-Head Like Proteins (RSP4/6-like)	44954 49798 24016

RSP23, Flagellar Radial Spoke Nucleoside Diphosphate Kinase	29950
Central Pair	
KLP1, Kinesin-Like Protein 1	64648
PF16/Spag6, Central Pair Protein	30562
PF20/Spag16, Central Pair Associated WD-Repeat Protein	952
PP1, Phosphatase 1	56375
PF6/SPAG17, Central Pair Protein	nf
CPC1/KPL2/Spef2, Central Pair Complex 1	45058
Hydin	78704
BBS	
Bardet-Biedl Syndrome 1	65179
Bardet-Biedl Syndrome 2	71257
Bardet-Biedl Syndrome 3	44202
Bardet-Biedl Syndrome 4	28891
Bardet-Biedl syndrome 5	34252
Bardet-Biedl Syndrome 7	68114
Bardet-Biedl Syndrome 8	80979
Bardet-Biedl Syndrome 9	80972
Basal Body	
Sas-4	61107
Sas-5	nf
Sas-6	68996
SF-assemblin	71540
Oral-facial-digital 1	nf
Variable Flagellar Number 3 (VFL3)	nf
Basal Body Protein BLD10	77710
PACRG1	29060 82851
Axoneme	
Calmodulin	55564
DIP13, Deflagellation Inducible Protein, 13KD	71898
MBO2, Coiled-Coil Flagellar Protein	62959
RIB43a, Flagellar Protofilament Ribbon Protein	83064
RIB72, nucleoside-diphosphokinase regulatory subunit p72	81047
PP2A, Protein Phosphatase 2a	62709
Profilin	34873 83254 61478
Tektin	nf
Flagellar Length Controll	
LF3, Long Flagella 3	nf
LF4, Long Flagella Protein	65383
LF1, Long Flagella 1	nf

Table 2.4: Flagellar motility associated proteins (FMs)

Flagellar-motility associated proteins (FMs) were identified as described in Materials and Methods. Those families with characterized *Chlamydomonas* homologs include the gene name from Version 3.0 of the *Chlamydomonas* genome (http://www.jgi.doe.gov/chlamy). *ath Arabidopsis thaliana,* ppa *Physcomitrella patens*, pra *Phytophthora ramorum*, tps *Thalassiosira pseudonana*, ptr *Phaeodactylum tricornutum*, ddi *Dictyostelium discoideum,* ncr *Neurospora crassa,* hsa human, tad *Trichoplax adherens*, mbr *Monosiga brevicollis*, pte *Paramecium tetraurelia*, tbr *Trypanosoma brucei*, gla *Giardia lamblia*, ehi *Entamoeba histolytica*, tva *Trichomonas vaginalis*, cre *Chlamydomonas reinhardtii.*

Name	*Naegleria* JGI protein ID	Gene family (cluster ID)	Species with genes in family	*Chlamydomonas* homolog	Other homologs
FM1	63280	6550330	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	IFT88	IFT88
FM2	65383	6550366	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	LF4	
FM3	81047	6550418	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	RIB72	
FM4	81229	6550938	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	FAP32	
FM5	59637	6551401	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	FAP52	
FM6	77715	6551416	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	BLD1	IFT52
FM7	61993	6552659	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	FAP259	
FM8	31069	6552726	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	SEH1, MOT47	
FM9	1424	6552828	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	FAP250	
FM10	79456	6553116	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	ARL3	
FM11	82851	6553427	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	BUG21	PACRG
FM12	71898	6552987	pra,hsa,ppa,mbr,tad,tps,pte,tbr,gla,cre,ngr	DIP13	
FM13	68117	6550932	pra,hsa,ppa,mbr,tad,tps,pte,tbr,tva,cre,ngr	FAP50	
FM14	49668	6552299	pra,hsa,ppa,mbr,tad,tps,pte,tbr,cre,ngr		
FM15	63939	6550894	pra,hsa,ppa,mbr,tad,tps,gla,tva,cre,ngr	FLA2/FLA8	
FM16	66643	6550571	pra,hsa,ppa,mbr,tad,ptr,pte,cre,ngr	FAP215	
FM17	80690	6549767	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	DYF13	
FM18	78704	6549988	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	HY3	Hydin
FM19	45002	6550401	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	IFT57	
FM20	71180	6551150	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr		
FM21	30192	6551402	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP198	
FM22	64930	6551455	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	RSP3	
FM23	82719	6551498	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	IDA4	
FM24	3580	6551596	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	MOT15	
FM25	29177	6551944	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	TCTEX1	

FM26	50399	6551960	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP60	
FM27	44774	6552071	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	BLD2	Epsilon tubulin
FM28	48798	6552126	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	IFT140	
FM29	78559	6552188	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	DHC2	
FM30	2066	6552209	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP184	
FM31	54982	6552426	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP253	
FM32	32701	6552870	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP118	
FM33	30562	6552881	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	PF16	
FM34	29690	6552903	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP66	
FM35	63764	6553257	pra,hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	IFT172	
FM36	79290	6550473	pra,hsa,ppa,mbr,tad,pte,tbr,gla,cre,ngr	FAP82	
FM37	68996	6550170	pra,hsa,ppa,mbr,tad,pte,tbr,tva,cre,ngr		Sas-6
FM38	70274	6550190	pra,hsa,ppa,mbr,tad,pte,tbr,tva,cre,ngr	FAP70	
FM39	77945	6550628	pra,hsa,ppa,mbr,tad,pte,tbr,tva,cre,ngr	IFT80	
FM40	61313	6552455	pra,hsa,ppa,mbr,tad,pte,tbr,tva,cre,ngr	FAP57	
FM41	79626	6551170	pra,hsa,ppa,mbr,tad,pte,tbr,cre,ngr	FAP116	
FM42	69007	6552725	pra,hsa,ppa,mbr,tad,pte,gla,tva,cre,ngr	UNI3	Delta tubulin
FM43	29002	6552496	pra,hsa,ppa,mbr,tad,tbr,tva,cre,ngr	FAP146	
FM44	33676	6551289	pra,hsa,ppa,mbr,tad,gla,cre,ngr	POC1	
FM45	31544	6552579	pra,hsa,ppa,mbr,tad,cre,ngr	RAB23	
FM46	68950	6550567	pra,hsa,ppa,mbr,ptr,tbr,ehi,tva,cre,ngr		LAG1
FM47	74561	6551926	pra,hsa,ppa,mbr,pte,gla,tva,cre,ngr	FAP134	
FM48	33146	6553920	pra,hsa,ppa,tad,tps,pte,tbr,gla,tva,cre,ngr	FAP67	
FM49	80717	6551160	pra,hsa,ppa,tad,tps,pte,tbr,cre,ngr	MOT45	
FM50	62959	6550378	pra,hsa,ppa,tad,tps,tbr,cre,ngr	MBO2	
FM51	77902	6550398	pra,hsa,ppa,tad,ptr,cre,ngr	DAT1	

FM52	63921	6552226	pra,hsa,ppa,tad,pte,tbr, gla,tva,cre,ngr	MOT17	
FM53	62977	6551500	pra,hsa,ppa,tad,pte,tbr, tva,cre,ngr	IFT20	
FM54	380	6552004	pra,hsa,ppa,tad,pte,tbr, tva,cre,ngr	FAP59	
FM55	57343	6552331	pra,hsa,ppa,tad,pte,tbr, tva,cre,ngr	IDA7	
FM56	65518	6553128	pra,hsa,ppa,tad,pte,tbr, tva,cre,ngr	MOT16	SPATA4
FM57	33361	6552781	pra,hsa,mbr,tad,tps,ptr, tbr,cre,ngr	MOT (ECH1)	
FM58	78637	6550142	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	ODA9	
FM59	60431	6550351	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	ODA6	
FM60	79232	6550351	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	ODA6	
FM61	81548	6551027	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	ODA1	
FM62	74922	6553051	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	DLC1	
FM63	54720	6553051	pra,hsa,mbr,tad,tps,pte, tbr,gla,tva,cre,ngr	DLC1	
FM64	44967	6549754	pra,hsa,mbr,tad,tps,pte, tbr,gla,cre,ngr	FAP127	
FM65	64648	6549959	pra,hsa,mbr,tad,tps,pte, tbr,gla,cre,ngr	KLP1	
FM66	60926	6550727	pra,hsa,mbr,tad,tps,pte, tbr,tva,cre,ngr	RABL2A	
FM67	64053	6551279	pra,hsa,mbr,tad,tps,pte, tbr,tva,cre,ngr	IFT81	
FM68	52666	6552934	pra,hsa,mbr,tad,tps,pte, tbr,tva,cre,ngr	MKS1	
FM69	78645	6553047	pra,hsa,mbr,tad,tps,pte, tbr,cre,ngr	PDE14	
FM70	79669	6553456	pra,hsa,mbr,tad,tps,pte, tva,cre,ngr	FLA3	Kinesin-associated protein 3
FM71	72811	6551092	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	FBB17	
FM72	64818	6551191	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	XRP2	
FM73	34252	6551275	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	BBS5	
FM74	80979	6551366	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	BBS8	
FM75	29188	6551631	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	FAP251	
FM76	46605	6551986	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	FAP91	
FM77	34729	6552141	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	FBB9 FAP208	

FM78	63907	6552827	pra,hsa,mbr,tad,pte,tbr, gla,tva,cre,ngr	FAP263	
FM79	32555	6551440	pra,hsa,mbr,tad,pte,tbr, tva,cre,ngr	MOT24	
FM80	68976	6552544	pra,hsa,mbr,tad,pte,tbr, tva,cre,ngr	FAP155	
FM81	71257	6552724	pra,hsa,mbr,tad,pte,tbr, tva,cre,ngr	BBS2	
FM82	80972	6553245	pra,hsa,mbr,tad,pte,tbr, tva,cre,ngr	BBS9	
FM83	56340	6551054	pra,hsa,mbr,tad,pte,tbr, cre,ngr	POC7	UNC119, HRG4
FM84	32161	6551507	pra,hsa,mbr,tad,pte,tbr, cre,ngr	FAP247	
FM85	63091	6551816	pra,hsa,mbr,tad,pte,tbr, cre,ngr	SSA4	
FM86	68114	6553892	pra,hsa,mbr,tad,pte,tbr, cre,ngr	BBS7	
FM87	62998	6553962	pra,hsa,mbr,tad,pte,tbr, cre,ngr	FAP22	
FM88	49599	6552482	pra,hsa,mbr,tad,pte,ehi, cre,ngr		
FM89	70454	6553122	pra,hsa,mbr,tad,pte,cre, ngr	POC11	
FM90	69068	6553239	pra,hsa,mbr,tad,pte,cre, ngr		
FM91	68064	6550371	pra,hsa,mbr,tad,tbr,cre, ngr	ARL13	
FM92	5673	6551616	pra,hsa,mbr,tps,ptr,pte, ehi,cre,ngr		
FM93	62358	6550664	pra,hsa,mbr,pte,cre,ngr	FAP69	
FM94	69688	6553087	pra,hsa,tad,tps,ptr,pte, gla,cre,ngr	DPY30	
FM95	67046	6554005	pra,hsa,tad,tps,ptr,cre, ngr	SSA20	
FM96	30532	6550502	pra,hsa,tad,tps,pte,tbr, gla,tva,cre,ngr	ODA12	
FM97	69956	6550107	pra,hsa,tad,tps,pte,tbr, tva,cre,ngr	FAP192	
FM98	55628	6550741	pra,hsa,tad,tps,pte,tbr, cre,ngr	ODA4	
FM99	38988	6552497	pra,hsa,tad,tps,tbr,gla, tva,cre,ngr	ODA2	
FM100	52938	6553086	pra,hsa,tad,pte,tbr,tva, cre,ngr	MOT52	
FM101	46913	6550620	pra,hsa,tad,pte,tbr,cre, ngr	FAP73	
FM102	56610	6550825	pra,hsa,tad,pte,tva,cre, ngr	MOT25	
FM103	52840	6553013	pra,hsa,tad,pte,tva,cre, ngr	MOT12	
FM104	68814	6552786	pra,hsa,tad,pte,cre,ngr	SSA3	
FM105	80259	6551480	pra,hsa,tad,tbr,cre,ngr	D1bLIC	

FM106	49289	6552258	pra,hsa,tad,cre,ngr		
FM107	71505	6552992	pra,hsa,tad,cre,ngr	GSTS1	
FM108	70195	6552992	pra,hsa,tad,cre,ngr	GSTS1	
FM109	70247	6552992	pra,hsa,tad,cre,ngr	GSTS1	
FM110	75317	6552992	pra,hsa,tad,cre,ngr	GSTS1	
FM111	56805	6553062	pra,hsa,tad,cre,ngr		
FM112	78620	6550198	pra,hsa,pte,tbr,tva,cre,ngr	FAP36	
FM113	49798	6551425	pra,hsa,pte,gla,cre,ngr	RSP4	
FM114	73137	6550379	pra,hsa,pte,cre,ngr	CAH1	
FM115	67854	6551362	hsa,ppa,mbr,tad,tps,pte,tbr,cre,ngr	FAP45	
FM116	68477	6551157	hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP65	
FM117	70051	6551331	hsa,ppa,mbr,tad,pte,tbr,gla,tva,cre,ngr		
FM118	81845	6552075	hsa,ppa,mbr,tad,tbr,cre,ngr	DHC6	
FM119	63304	6549916	hsa,ppa,mbr,tad,gla,tva,cre,ngr	BOP5	
FM120	4843	6551486	hsa,ppa,mbr,tps,ptr,cre,ngr	CYN40	
FM121	70995	6553235	hsa,ppa,mbr,tps,ptr,cre,ngr		
FM122	83269	6551165	hsa,ppa,mbr,pte,cre,ngr	AAH1	
FM123	32341	6552151	hsa,ppa,tad,tps,pte,tbr,gla,tva,cre,ngr		
FM124	29888	6554041	hsa,ppa,tad,pte,tbr,gla,tva,cre,ngr	FAP44	
FM125	50227	6549899	hsa,ppa,tad,pte,tbr,cre,ngr	FAP14	
FM126	80274	6550509	hsa,ppa,tad,ehi,cre,ngr		Sirtuin
FM127	66079	6552202	hsa,ppa,tad,cre,ngr	TRXm	
FM128	4868	6551151	hsa,ppa,ptr,cre,ngr	DNJ29	
FM129	72718	6552619	hsa,ppa,pte,tbr,tva,cre,ngr	MOT39	
FM130	4931	6553861	hsa,ppa,tbr,cre,ngr		
FM131	64631	6552548	hsa,ppa,tva,cre,ngr	FAP269	
FM132	81521	6551060	hsa,mbr,tad,tps,ptr,tva,cre,ngr		
FM133	80835	6552017	hsa,mbr,tad,tps,pte,tbr,gla,tva,cre,ngr	SSA11	
FM134	77673	6551871	hsa,mbr,tad,tps,pte,tbr,tva,cre,ngr		
FM135	62841	6552058	hsa,mbr,tad,tps,pte,tbr,tva,cre,ngr		MKS3
FM136	30379	6553447	hsa,mbr,tad,tps,pte,tbr,tva,cre,ngr		
FM137	4601	6551499	hsa,mbr,tad,pte,tbr,gla,tva,cre,ngr	FAP9	
FM138	74042	6551732	hsa,mbr,tad,pte,tbr,gla,tva,cre,ngr		

FM139	50561	6552523	hsa,mbr,tad,pte,tbr,gla, tva,cre,ngr		
FM140	61232	6552767	hsa,mbr,tad,pte,tbr,tva, cre,ngr	MOT37	
FM141	65873	6553468	hsa,mbr,tad,pte,tbr,tva, cre,ngr	FAP161	
FM142	73596	6554034	hsa,mbr,tad,pte,tbr,tva, cre,ngr	FAP61	
FM143	80404	6552775	hsa,mbr,tad,pte,ehi,cre,n gr		
FM144	62107	6551340	hsa,mbr,tad,pte,cre,ngr	POC16	
FM145	57344	6553502	hsa,mbr,tad,tbr,tva,cre,n gr		
FM146	66608	6553089	hsa,mbr,ptr,pte,cre,ngr		
FM147	68057	6552972	hsa,mbr,tbr,cre,ngr		
FM148	79419	6550542	hsa,mbr,cre,ngr	FOX1	
FM149	73977	6550542	hsa,mbr,cre,ngr	FOX1	
FM150	80346	6552423	hsa,mbr,cre,ngr		
FM151	70654	6553513	hsa,mbr,cre,ngr		
FM152	83064	6551733	hsa,tad,tps,pte,tbr,gla, tva,cre,ngr	RIB43a	
FM153	65759	6553379	hsa,tad,tps,pte,tbr,cre, ngr		TECT3
FM154	73664	6550305	hsa,tad,tps,cre,ngr		
FM155	62591	6552981	hsa,tad,ptr,tva,cre,ngr		
FM156	71996	6552728	hsa,tad,ptr,cre,ngr	MOT50	
FM157	54684	6552728	hsa,tad,ptr,cre,ngr	MOT50	
FM158	67231	6550823	hsa,tad,pte,tbr,cre,ngr	PTP1	
FM159	71676	6553723	hsa,tad,pte,tbr,cre,ngr	FAP119	
FM160	4690	6550250	hsa,tad,pte,gla,cre,ngr	FAP111	
FM161	29577	6553164	hsa,tad,pte,cre,ngr	POC12	MKS1
FM162	59473	6553478	hsa,tad,pte,cre,ngr	PSO2	
FM163	48518	6551660	hsa,tad,tbr,gla,tva,cre,		
FM164	82958	6553096	hsa,tad,tbr,tva,cre,ngr		
FM165	29126	6550596	hsa,tad,tbr,cre,ngr		
FM166	70275	6553729	hsa,tad,gla,tva,cre,ngr		
FM167	58252	6552224	hsa,tad,cre,ngr		
FM168	71452	6553949	hsa,tad,cre,ngr		
FM169	73917	6552135	hsa,tps,cre,ngr	MOT51	
FM170	67664	6552862	hsa,tps,cre,ngr		
FM171	73885	6554672	hsa,ptr,pte,cre,ngr	PKHD1-2	
FM172	80536	6549888	hsa,ptr,tva,cre,ngr		
FM173	82475	6554227	hsa,ptr,cre,ngr	GSTS3	
FM174	31511	6553580	hsa,pte,cre,ngr		
FM175	78247	6554247	hsa,pte,cre,ngr	PSK1	
FM176	78184	6553815	hsa,tbr,cre,ngr	FKB12	
FM177	59563	6553039	hsa,gla,cre,ngr		
FM178	73058	6552889	hsa,ehi,cre,ngr		
FM179	78958	6554233	hsa,tva,cre,ngr	CYG11	
FM180	68774	6554233	hsa,tva,cre,ngr	CYG11	
FM181	66783	6554233	hsa,tva,cre,ngr	CYG11	
FM182	71868	6553432	hsa,cre,ngr		

Table 2.5: Amoeboid motility associated proteins (AMs)

Amoeboid-motility associated proteins (AMs) were identified as described in Materials and Methods. Proteins encoded by multiple *Naegleria* paralogs are noted with multiple JGI ids in the second column. Red text is used to indicate AM gene families with homologs in *Trichomonas vaginalis*. Species abbreviations as in Table 2.4.

Name	Naegleria JGI protein ID(s)	Protein family (cluster ID)	species in cluster	Manual annotation of molecular function	PFAM domains (1e-10)
Actin Binding					
AM1	76225; 81173	6552646	ddi,ncr,hsa,mbr,tad,tva,ngr	Actin Binding	PF00307: Calponin homology (CH) domain|
AM2	82236	6550672	ddi,hsa,mbr,tad,ehi,ngr	Actin Binding (Drebrin/ABP-1)+1:65536	PF00018: SH3 domain|PF07653:Variant SH3 domain|
AM3	80016	6553037	hsa,tad,ehi,ngr	Actin Binding (Filamin)	PF00307: Calponin homology (CH) domain(2) | PF00630: Filamin/ABP280 repeat (4)
AM4	58328	6553194	ddi,ncr,hsa,mbr,tad,ehi,ngr	Actin Binding (twinfilin)	PF00241: Cofilin/tropomyosin-type actin-binding protein|
AM5	77687	6554206	ddi,hsa,tad,ngr	Actin Binding (Wash)	no PFAM
Signaling					
AM6	47789	6553607	hsa,ehi,ngr	Signalling	no PFAM | 7TMs predicted (TMHMM)
AM7	80282	6551531	ddi,ncr,hsa,mbr,ngr	Signalling	PF00018: SH3 domain (2) | PF07653:Variant SH3 domain (2)
AM8	67958	6554192	ddi,hsa,ehi,ngr	Signalling	PF04664: Opioid growth factor receptor (OGFr) conserved region|
AM9	71270	6554074	ddi,hsa,ngr	Signalling	PF07690: Major Facilitator Superfamily|
GAP					
AM10	80615	6553827	pra,hsa,ehi,ngr	GAP	PF00616: GTPase-activator protein for Ras-like GTPase|
AM11	81714	6549864	ddi,pra,ncr,hsa,mbr,tad,ehi,tva,ngr	GAP	PF02145: Rap/ran-GAP|
AM12	78320	6552715	ddi,ncr,hsa,mbr,tad,ehi,ngr	GAP (Nadrin)	PF00620: RhoGAP domain|
GEF					
AM13	50007; 68966	6550519	hsa,mbr,ehi,ngr	GEF	PF00618: Guanine nucleotide exchange factor for Ras-like GTPases; N-terminal motif|PF00617: RasGEF domain|
Membrane					
AM14	57266	6554111	hsa,ehi,ngr	Membrane	PF00169: PH domain (3)
AM15	4009	6551952	ddi,ncr,hsa,mbr,tad,ngr	Membrane	PF04191: Phospholipid methyltransferase|
AM16	48624	6550972	hsa,tad,ehi,ngr	Membrane (Sphyngomyelin synthase-related)	no PFAM | 6 transmembrane domains predicted by TMHMM
AM17	82816	6552197	ddi,hsa,tad,ngr	Membrane (Saposin-B)	PF05184: Saposin-like type B, region 1 (3) | PF03489: Saposin-like type B, region 2 (3)
AM18	74044	6554134	ddi,hsa,ngr	Membrane	PF00754: F5/8 type C domain|
Cytoskeletal					
AM19	68732	6551755	ddi,pra,hsa,mbr,tad,ngr	Cytoskeletal	PF04912: Dynamitin|
Vesicle					
AM20	66720	6550251	hsa,ehi,ngr	Vesicle	PF02750: Synapsin, ATP binding domain|

AM21	62049	6552718	ddi,hsa,tad,ngr	Vesicle	no PFAM
AM22	78255	6554714	ddi,hsa,ngr	Vesicle	no PFAM
Protein Trafficking					
AM23	80788	6552115	ddi,hsa,tad,ngr	Protein Trafficking	no PFAM
Protein Turnover					
AM24	58872	6553361	ddi,ncr,hsa,mbr,ngr	Protein Turnover	no PFAM
AM25	65046	6553500	ddi,pra,hsa,ngr	Protein Turnover	no PFAM
Protein Interaction					
AM26	81452	6553306	hsa,tad,ehi,ngr	Protein Interaction	PF01436: NHL repeat (5)
Cell Cycle					
AM27	29264	6553985	ddi,hsa,tad,ngr	Cell Cycle	PF04005: Hus1-like protein|
AM28	58254	6553143	ddi,pra,hsa,tad,ngr	Cell Cycle	no PFAM
Metabolism					
AM29	65213	6550836	ddi,ncr,hsa,tad,ngr	Metabolism	PF06052: 3-hydroxyanthranilic acid dioxygenase|
AM30	81411	6549768	ddi,hsa,mbr,tad,ngr	Metabolism	PF03301: Tryptophan 2,3-dioxygenase|
AM31	78567	6553649	ddi,hsa,mbr,tad,ngr	Metabolism	no PFAM
AM32	69774	6552632	ddi,hsa,mbr,ngr	Metabolism	PF03632: Glycosyl hydrolase family 65 central catalytic domain|
AM33	78233	6553977	ddi,hsa,mbr,ngr	Metabolism	PF01229: Glycosyl hydrolases family 39|
AM34	78308	6554099	ddi,hsa,mbr,ngr	Metabolism	no PFAM
AM35	54990; 33467	6554340	ddi,hsa,ngr	Metabolism	PF03747: ADP-ribosylglycohydrolase| (not found in 54990)
Nucleic Acid Metabolism					
AM36	61798	6554539	ddi,pra,hsa,mbr,ngr	Nucleic Acid Metabolism	no PFAM
AM37	71340	6553262	ddi,pra,hsa,tad,tva,ngr	Nucleic Acid Metabolism	PF04858: TH1 protein|
AM38	53469	6549994	ddi,pra,hsa,tad,ngr	Nucleic Acid Metabolism	PF02144: Repair protein Rad1/Rec1/Rad17|
AM39	61854	6551921	ddi,pra,hsa,tad,ngr	Nucleic Acid Metabolism	PF00533: BRCA1 C Terminus (BRCT) domain (6)
AM40	56696	6552937	ddi,pra,hsa,ngr	Nucleic Acid Metabolism	PF05625: PAXNEB protein|
AM41	79767	6553442	ddi,ncr,hsa,ehi,ngr	Nucleic Acid Metabolism	PF02891: MIZ zinc finger|
AM42	77967	6554210	ddi,hsa,tad,ehi,ngr	Nucleic Acid Metabolism	PF06978: Ribonucleases P/MRP protein subunit POP1|
AM43	61462	6553485	ddi,hsa,tad,ngr	Nucleic Acid Metabolism	no PFAM
AM44	67690	6554441	ddi,hsa,ngr	Nucleic Acid Metabolism	no PFAM
Unknown					
AM45	74247	6553826	ddi,pra,hsa,tad,ehi,tva,ngr	Unknown	PF07258: HCaRG protein|
AM46	79980	6551166	ddi,pra,hsa,tad,tva,ngr	Unknown	no PFAM
AM47	80574	6550569	ddi,pra,hsa,tad,ngr	Unknown	no PFAM
AM48	5651	6549995	ddi,hsa,mbr,tad,ngr	Unknown	no PFAM
AM49	69245	6553024	ddi,hsa,mbr,tad,ngr	Unknown	PF07258: HCaRG protein|
AM50	67354	6553307	ddi,hsa,mbr,tad,ngr	Unknown	no PFAM
AM51	81535	6553370	ddi,hsa,mbr,tad,ngr	Unknown	no PFAM

AM52	65831	6553163	ddi,hsa,mbr,ngr	Unknown	no PFAM
AM53	81892	6551868	ddi,hsa,tad,tva,ngr	Unknown	no PFAM
AM54	45670	6553077	ddi,hsa,tad,tva,ngr	Unknown	no PFAM
AM55	80291	6552024	ddi,hsa,tad,ngr	Unknown	no PFAM
AM56	68270	6552063	ddi,hsa,tad,ngr	Unknown	PF07258: HCaRG protein|
AM57	75398	6552274	ddi,hsa,tad,ngr	Unknown	no PFAM
AM58	63144	6552665	ddi,hsa,tad,ngr	Unknown	no PFAM
AM59	81040	6552879	ddi,hsa,tad,ngr	Unknown	PF07742: BTG family|
AM60	79421; 4350	6554251	ddi,hsa,tad,ngr	Unknown	no PFAM
AM61	5358	6551639	ddi,hsa,ngr	Unknown	no PFAM
AM62	62278	6553655	ddi,hsa,ngr	Unknown	no PFAM
AM63	69376	6554699	ddi,hsa,ngr	Unknown	no PFAM

Figure 2.1. Electrophoretic karyotype of *Naegleria gruberi*

Pulsed field electrophoresis gel of *Naegleria gruberi*, strain NEG-M (lanes 4-11), with the amount of DNA loaded increasing left to right. Lanes 1-3 contain markers with chromosome sizes indicated (*Saccharomyces cerevisiae* in lane one, and *Hansenula wingei* in lane two, and *Schizosaccharomyces pombe* in the third lane). *Naegleria* chromosome sizes are indicated, and range from ~0.7 to ~6.6 Mb.p. We estimate the total genome size to be 42 Mb.p.

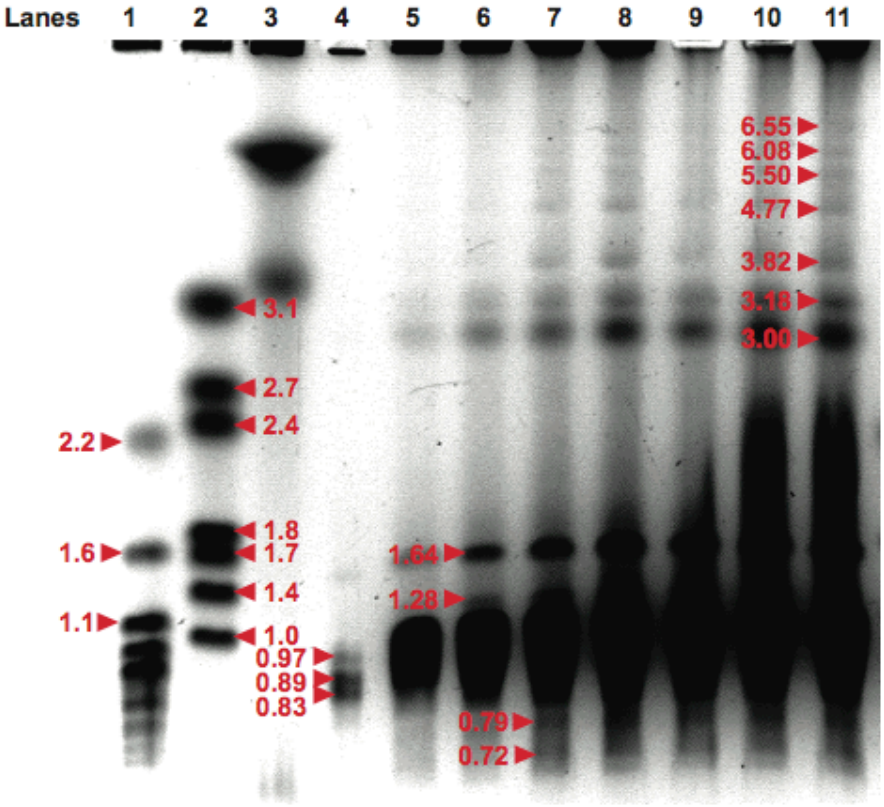


Figure 2.2. Venn diagram of evidence supporting *Naegleria* gene models.

The gene models in the *Naegleria* genome are distributed in the Venn diagram according to supporting evidence. 4,448 models have EST support; 5,938 are in a *Naegleria* protein family; 7.974 have homology in another genome we used in building protein families; and 7,042 have a Pfam annotation.

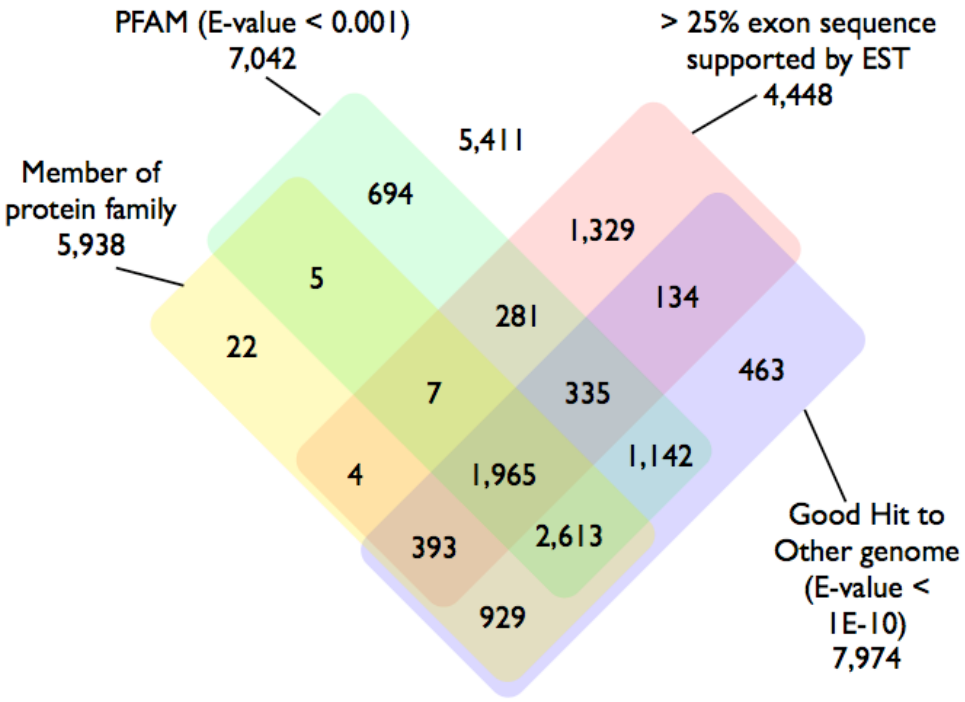
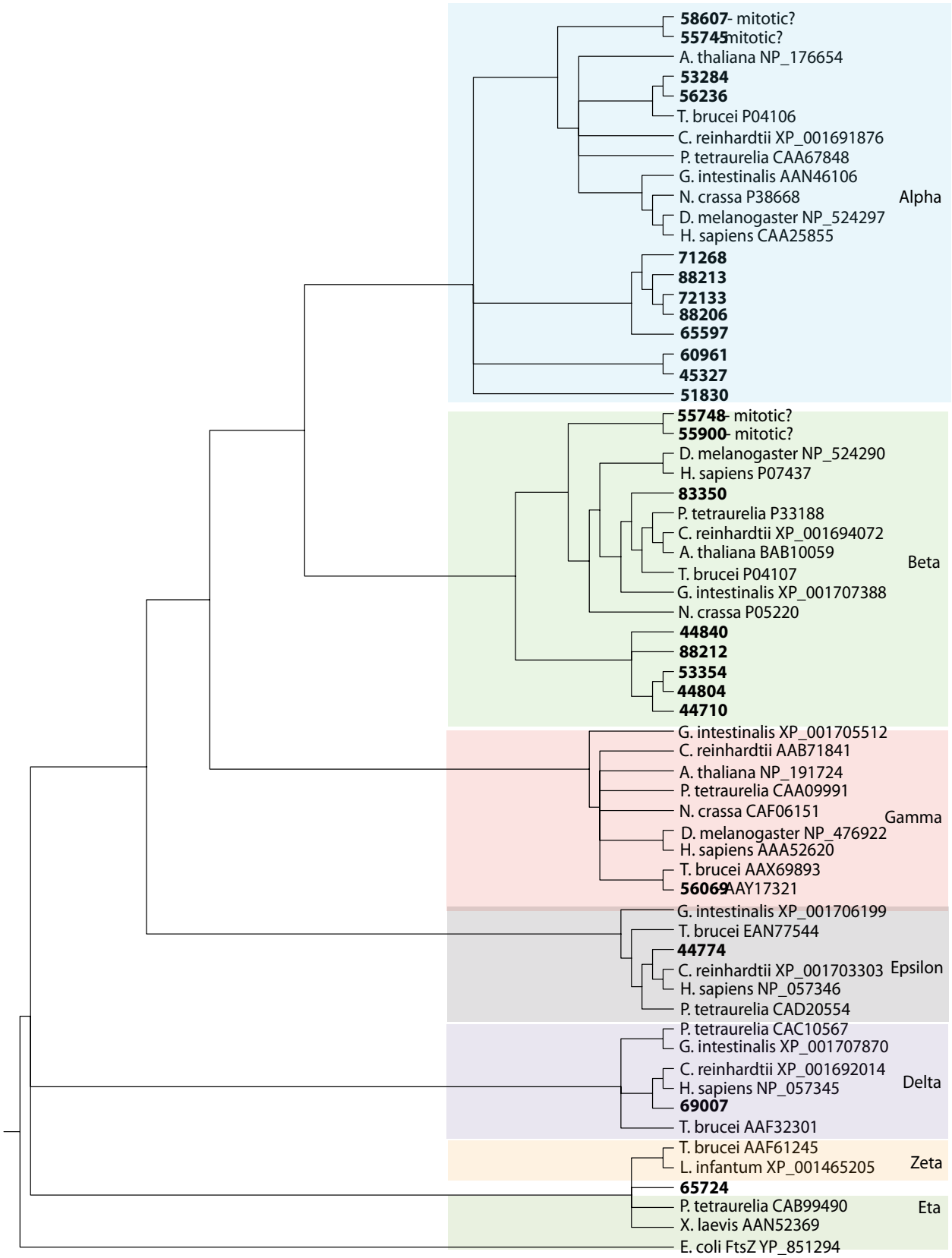


Figure 2.3. Tubulin phylogeny

A phylogenetic tree of the tubulin superfamily, including all 24 non-redundant *Naegleria* tubulin sequences with complete gene models. This maximum likelihood tree was created with RAxML using the JTT amino acid model, 1000 rapid bootstrap replicates, and *E. coli* FtsZ as the outgroup. *Naegleria* sequences are identified by their protein ID (bold), and all other sequences by the species and GenBank accession number. Bootstrap values above 50% are shown; nodes with bootstrap values below 50% were collapsed into polytomies.


58607 - mitotic?
55745 - mitotic?
A. thaliana NP_176654
53284
56236
T. brucei P04106
C. reinhardtii XP_001691876
P. tetraurelia CAA67848
G. intestinalis AAN46106
N. crassa P38668
D. melanogaster NP_524297
H. sapiens CAA25855
Alpha
71268
88213
72133
88206
65597
60961
45327
51830
55748 - mitotic?
55900 - mitotic?
D. melanogaster NP_524290
H. sapiens P07437
83350
P. tetraurelia P33188
C. reinhardtii XP_001694072
A. thaliana BAB10059
T. brucei P04107
G. intestinalis XP_001707388
N. crassa P05220
Beta
44840
88212
53354
44804
44710
G. intestinalis XP_001705512
C. reinhardtii AAB71841
A. thaliana NP_191724
P. tetraurelia CAA09991
N. crassa CAF06151
D. melanogaster NP_476922
H. sapiens AAA52620
T. brucei AAX69893
56069 AAY17321
Gamma
G. intestinalis XP_001706199
T. brucei EAN77544
44774
C. reinhardtii XP_001703303
H. sapiens NP_057346
P. tetraurelia CAD20554
Epsilon
P. tetraurelia CAC10567
G. intestinalis XP_001707870
C. reinhardtii XP_001692014
H. sapiens NP_057345
69007
T. brucei AAF32301
Delta
T. brucei AAF61245
L. infantum XP_001465205
Zeta
65724
P. tetraurelia CAB99490
X. laevis AAN52369
Eta
E. coli FtsZ YP_851294

Chapter 3: *Naegleria* uses both templated and *de novo* basal body assembly, and expresses basal body proteins in the order of their incorporation

Abstract:

Centrioles and basal bodies are composed largely of a cylinder of nine microtubule triplets. The surrounding amorphous material harbors the microtubule organizing activities of the centrosome. Metazoan centrosomes define mitotic spindle poles, and their centrioles are called basal bodies when used to form cilia. Animal centrioles usually duplicate via "templated" assembly, with the new centriole developing perpendicularly from the side of a preexisting centriole. Centrioles can also be formed "*de novo*", in cytosol devoid of preexisting centrioles in some plant and animal cells, as well as the amoeboflagellate *Naegleria*. *Naegleria* grows as an amoeba lacking a cytoplasmic microtubule cytoskeleton. When stressed, *Naegleria* rapidly and synchronously differentiates into a flagellate, forming a complete cytoplasmic cytoskeleton *de novo*, including two basal bodies and flagella. How *Naegleria* makes exactly two basal bodies *de novo* remains an open question. Here, we describe the order of expression and incorporation into basal bodies, of three *Naegleria* centriole proteins (SAS-6, γ-tubulin, and centrin). We also provide the first evidence that *Naegelria* has templated, as well as *de novo*, basal body assembly, and suggest that having both capacities allows *Naegleria*, and other organisms, such as mouse embryos, to make the correct number of centrioles.

Introduction:

The beautiful and enigmatic pinwheel structures of centrioles and basal bodies have captured the imagination of cell biologists for over 100 years. These small (~ 1 micron) organelles are composed largely of a cylinder made from nine microtubule triplets (Fulton and Dingle, 1971). The surrounding amorphous material harbors the microtubule organizing activities of the centrosome, placing centrioles in the hub of the microtubule cytoskeleton. Metazoan centrosomes define mitotic spindle poles, and their centrioles are called basal bodies when used to form cilia (Sorokin, 1968). Moreover, Meeves (1900) showed in a series of classical experiments that centrioles and basal bodies are inter-convertible structures (Wilson, 1928). Centrioles must replicate exactly once per cell cycle, as duplication errors can lead to problems with chromosomal segregation and cell morphology (Kramer et al., 2002).

Virtually all animal cells have a pair of centrosomal centrioles that duplicate via "templated" assembly, with the new centriole developing perpendicularly from the

side of a preexisting centriole (Beisson and Wright, 2003). Despite this tantalizing title, there is no evidence for transfer of material from the mother centriole to the developing daughter centriole (Marshall, 2001). Three features characterize templated centriole assembly. First, templated assembly is linked to the cell cycle, with the beginning of duplication correlated with induction of S-phase. Second, templated assembly includes a tight counting mechanism, allowing only one daughter centriole to be assembled along side each mother centriole. Finally, templated assembly involves precursor-product relationships; the mother centriole is visible as a bar-shaped structure, and the daughter emerges as a single focus that then elongates to form a second bar (Marshall, 2001).

Centrioles can also be formed "*de novo*", in cytosol completely devoid of preexisting centrioles and basal bodies. In addition to many *in vivo* examples (Marshall, 2001), terminally differentiated fibroblasts held in S-phase can assemble centrioles *de novo* after removal of preexisting centrioles by laser microsurgery (Khodjakov et al., 2002). Although structurally canonical, these centrioles are usually produced in excess (Khodjakov et al., 2002). The amoeboflagellate *Naegleria* can also make basal bodies *de novo* (Fulton and Dingle, 1971), but how *Naegleria* is able to make exactly two basal bodies *de novo* remains an open question.

Two centrosomal proteins that have been studied during *Naegleria* differentiation are centrin and γ-tubulin. Centrin is a calcium-binding phosphoprotein that is an integral component of the wall and lumen of basal bodies and the pericentriolar lattice in many organisms (Beisson and Wright, 2003; Levy et al., 1996), and localizes specifically to basal body structures throughout *Naegleria* differentiation (Levy et al., 1998). γ-tubulin is a microtubule nucleation factor that localizes to microtubule organizing centers (MTOCs) of many types. Surprisingly, *Naegleria*'s γ-tubulin homolog has been reported to localize to basal body precursor complexes, then move to the other end of the cell, and finally disappear completely (Suh et al., 2002). SAS-6 is a functionally conserved coiled-coil protein required for the formation of diverse basal body precursor structures (Culver et al., 2009; Nakazawa et al., 2007; Pelletier et al., 2006; Rodrigues-Martins et al., 2007; Strnad et al., 2007), including the "central tube" (a cylindrical basal body precursor of *C. elegans* and *Drosophila* that lacks microtubules)(Pelletier et al., 2006; Rodrigues-Martins et al., 2007) and the *Chlamydomonas* cartwheel (a flat ring nine radiating spokes) (Nakazawa et al., 2007).

Here we describe the localization of *Naegleria*'s homologs of SAS-6, γ-tubulin, and centrin during the assembly of basal bodies and rest of the cytoskeleton. We also use these antibodies to determine the order of expression and incorporation of these proteins during basal body assembly. Finally, we provide the first evidence that *Naegelria* has the capacity for templated, as well as *de novo*, basal body assembly.

Results

Naegleria has canonical basal body genes

Using the *Naegleria* genome sequence (www.jgi.doe.gov/naegleria), we identified many conserved basal body genes (Table 3.1). These proteins include cartwheel components (Poc1, SAS-6 and BLD10), blade components (δ-tubulin and Rib43) as well as proteins previously localized to the centriole/basal body lumen (SAS-6 and γ-tubulin), and proteins that make up centriole appendages in other systems (SF-assemblin and centrin) (localizations reviewed in (Kilburn et al., 2007) and (Marshall, 2001)). Surprisingly, we do not find evidence of a homolog of pericentrin, despite reports of localization of this protein using heterologous antibodies (Suh et al., 2002). The presence of many known basal body genes indicates that, although *Naegleria*'s basal bodies are transient structures (Levy et al., 1998), they are likely canonical in composition as well as structure.

Order of centrosomal protein induction

As we were unable to reliably detect *Naegleria* γ-tubulin using heterologous antibodies for immunoflorescence, we raised and affinity-purified a polyclonal antibody to a 97 amino acid segment of *Naegleria*'s γ-tubulin gene. To investigate early basal body assembly events, we also raised a polyclonal antibody against *Naegleria*'s SAS-6 homolog. The resulting γ-tubulin and SAS-6 antibodies each recognize a single band of the appropriate size (55 and 74 kDa, respectively) on immunoblots of *Naegleria* flagellate extracts (data not shown). These antibodies, as well as published antibodies that recognize *Naegleria*'s centrin (Levy et al., 1996) and β-tubulin (Walsh, 1984) were used to follow protein levels during differentiation (Figure 3.1). Cell extracts were collected from 14 different timepoints between 0 and 90 minutes and used for Western blots, with actin protein levels serving as a loading control, as they remain constant throughout differentiation (Walsh, 1984).

The first protein to reach detectable levels was SAS-6, with detectable protein by 5 min, and reaching a stable maximum level at 25 min. This is well before basal bodies are known to assemble (visible flagella appear by 65 minutes, and basal bodies form approximately 10 minutes prior). Centrin was the next protein to appear, detectable by 25 minutes and reaching a stable maximum at 60 minutes. γ-Tubulin was not detectable until 60 minutes, when it appeared at maximum levels. Once maximum protein production was reached, all proteins remained at peak levels through the rest of the 90 min time period. In summary, basal body proteins were induced in the following order: Sas-6, Centrin, and finally γ-tubulin.

Naegleria's Sas-6 and Gamma-tubulin localize to basal bodies throughout differentiation

We also used the new γ-tubulin and SAS-6 antibodies to localize their target proteins during differentiation. As has been previously described, no microtubule or centrin containing structures were detected in interphase amoebae. Likewise, we found no γ-tubulin containing structures in interphase amoebae, or cells early in differentiation (Figure 3.2). The tubulin of the spindles of mitotic amoebae were detectable with anti-α-tubulin antibodies, but mitotic amoebae did not contain any detectable centrin- or γ-tubulin - containing structures (data not shown). γ-tubulin foci appear before flagella assemble (by 40 minutes), and remain at the base of the flagella throughout differentiation.

Like γ-tubulin, SAS-6 is absent in *Naegleria* amoebae and early-stage flagellates (Figure 3.3). However, staining with the anti-SAS-6 antibody revealed a single round focus that appeared by 40 minutes. SAS-6 was consistently located at the proximal end (relative to centrin) of both early basal body structures and mature basal bodies, with flagella emerging from the centrin-positive distal ends in fully formed flagellates (Figure 3.3). This localization pattern is similar to that seen in other systems, particularly the localization of SA6-6 to the cartwheel of *Chlamydomonas* (Nakazawa et al., 2007) and to the proximal ends of both basal bodies and animal centrioles (Culver et al., 2009; Strnad et al., 2007)

Order of incorporation of SAS-6, centrin, and γ-tubulin during assembly

To determine the order of centrin, SAS-6 and γ-tubulin incorporation into basal bodies, we performed immunoflorescence with each antibody, at 5 minute intervals during differentiation. At least 100 cells per sample were scored for localization, and times normalized to the time when 50% of cells have visible flagella. From these data it is clear that SAS-6, centrin, and γ-tubulin localize to basal bodies in the same order as protein induction. At t=30, a wave of SAS-6 foci begins, followed by centrin foci at t=35, and finally γ-tubulin localizes starting at t=40.

It should be noted that although protein localization follows the same order as induction, γ-tubulin is detectable via immunofluorescence at earlier time points than on immunoblots. This may be because the γ-tubulin in ~20% of cells at 45 minutes is concentrated enough to be detectable by immunoflorescence, but when this γ-tubulin is diluted with proteins from the remaining ~80% of cells (that do not have visible γ-tubulin localization), the total gamma tubulin protein concentration is not great enough to detect using Western blots.

Evidence for both de novo and templated assembly pathways in *Naegleria*

Templated centriole duplication produces two centrioles in very close proximity, as does *Naegleria* differentiation. *Naegleria*'s ability to "count" its basal bodies would

be easily explained if it made one basal body *de novo*, followed by templated assembly of a second. If this were true, one would expect to see production of a first basal body structure (a "bar"), followed by the initiation of a second one in close proximity (a "bar" plus a "dot"), which would grow to form two basal bodies (two "bars"). Supporting this hypothesis, close inspection of the cells fixed near the time of basal body assembly, and stained with anti-alpha and beta tubulin, anti-centrin and anti-SAS-6 antibodies revealed several types of structures (Fig. 5a). These included single small dots, slightly larger dots, ellipses, bars, bars with an associated dot, and double bars. Some cells did not contain any structures.

To determine if the relative proportions of these structures among the population of cells changes over time, counts were done on cells fixed between t=0 and t=90 minutes of differentiation. For each timepoint, at least hundred random cells were scored for the presence of the following centrin and SAS-6 containing structures: dots or single bars, bars with an associated dot, double bars, double bars with one tubulin-containing filament, and double bars with two tubulin-containing filaments. Cells without structures were scored as such. As single dots were similar in appearance to occasional specks of dust on the slide, dots were only scored if in the vicinity of a cell nucleus (within two nuclear diameters of the edge of a nucleus) and if they stained with both anti-tubulin and anti-centrin antibodies. As a single bar seen on end would resemble a dot, dots and single bars were grouped together. The progression of structures from early in differentiation to later is clear: small single dots or bars are the first structures to appear, followed by single bars with an associated dot, double bars, double bars with one tubulin-containing filament, and finally, double bars each with a tubulin-containing filament (Fig 5b). Thus, *Naegleria* seems to possess features typical of templated assembly; *Naegleria* forms one basal body first, with the second basal body emerging in close proximity as a small dot that then elongates to form a second bar.

Limited attempts to view these centrin and tubulin- containing structures via immuno-electron microscopy were not successful (Y.Y. Levy and M.B. Heintzelman, unpublished). This was probably due to the small size and scarcity of these structures in the cells. Despite the lack of ultrastructural verification of the structures observed by immunofluorescence microscopy, it is likely that the progression of these structures represents the sequence of the *de novo* formation of basal bodies in *Naegleria* and not some other cellular process. First, the antibodies used in this study were directed against known components of basal bodies, and in the case of centrin and SAS-6, components known at the ultrastructural level to be limited to basal bodies (Levy et al., 1998) (Culver et al., 2009; Nakazawa et al., 2007). The appearance and size of the two centrin and SAS-6 bars was identical to that previously seen in immunofluorescence localization of centrin to the basal bodies of mature flagellates (Levy et al., 1998), and SAS-6 in *Tetrahymena* (Culver et al., 2009). Second, the observed progression of structures occurred before the formation of flagella, which is the period in which basal body formation would be

expected. Third, while the initial appearance of a single dot or bar was unexpected, the progression of stages after the formation of the first putative basal body is essentially identical to that described using electron microscopy to study basal body duplication in numerous other organisms (Marshall, 2001).

Discussion:

Naegleria has long been recognized for its remarkable ability to form an entire cytoplasmic microtubule cytoskeleton from scratch. Although *Naegleria* makes its initial basal body *de novo*, it has canonical structure (Fulton and Dingle, 1971), and the presence of many known basal body genes argues that it has canonical composition as well. This is in contrast to well established models of centriole/basal body assembly that have modified ultrastructure or are missing otherwise well conserved genes (e.g. *C. elegans* centrioles have singlet microtubule blades (Beisson and Wright, 2003) and lack centrin (Bornens and Azimzadeh, 2007)). *Naegelria*'s *de novo* basal body assembly makes it particularly useful as a tool for studying basal body assembly, particularly as protein induction occurs in the same order as incorporation during basal body assembly (at least for the proteins studied here). In Chapter 4, I describe microarray analysis of *Naegleria* differentiation that shows that transcript induction also mirrors the assembly pathway.

The antibodies we have raised in this study to *Naegleria* SAS-6 and γ-tubulin have revealed canonical localization for these key centriolar proteins. In particular, SAS-6 localizes to the proximal ends of basal bodies, relative to centrin, as seen in other systems (Nakazawa et al., 2007; Strnad et al., 2007). Additionally, we show γ-tubulin to be consistently localized to basal bodies throughout differentiation. This consistent localization is in direct contrast to what has been previously described (Suh et al., 2002). We propose that the previous reports of γ-tubulin foci moving to the opposite end of the cell from *Naegleria*'s MTOC, after basal body assembly, were an artifact of using heterologous antibodies. This is a particular problem for *Naegleria*, as numerous antibodies seem to localize to a round structure towards the posterior of the cell (data not shown) that may perhaps represent a vacuole or a lipid storage vesicle.

We have determined an initial basal body assembly pathway for *Naegelria* by following the timing of localization centriolar proteins. These results suggest that SAS-6 localizes first, then centrin, and finally γ-tubulin. The localization of SAS-6 before centrin suggests that the *de novo* basal body assembly in *Naegleria* is similar to templated centriolar duplication. For example, SAS-6 foci form early in S-phase of human cells, before localization of centrin (Strnad et al., 2007). SAS-6 is also required for proper assembly of *Chlamydomonas*' earliest basal body precursor structure, the cartwheel (Nakazawa et al., 2007).

On the other hand, it is difficult to compare the timing of γ-tubulin localization in templated centriole duplication to that in *Naegleria*'s *de novo* assembly, as the former occurs within the centrosome, an environment already enriched in γ-tubulin, and the latter occurs during the initial stages of MTOC assembly. Previous studies have shown γ-tubulin necessary for centriole assembly (Beisson and Wright, 2003), however the mechanism of its involvement remains unclear. Our data suggest that the initial stages of centriole assembly (the incorporation of centrin and SAS-6) do not require detectable amounts of γ-tubulin. If this is true, γ-tubulin should also localize before or after centrin during *de novo* centriole assembly in human cells after laser microsurgery.

Finally, our evidence indicates that *Naegleria* makes its first basal body *de novo*, and the second using a pathway resembling templated assembly. Further, *Naegleria*'s basal body assembly is independent of the cell cycle (Fulton, 1977). Thus, we suspect that one of *Naegleria*'s ancestors acquired the ability to form a first basal body *de novo* in response to stress, and maintained the ability to form a second basal body using the canonical "templated" route. Maintenance of an evolutionarily conserved counting mechanism (templated assembly) seems more parsimonious than the evolution of a second, novel, mechanism for ensuring the formation of the correct number of basal bodies. *Naegleria*'s apparent capacity for both templated and *de novo* basal body assembly echos that of numerous other species (e.g. humans (Khodjakov et al., 2002)and *Chlamydomonas* (Marshall et al., 2001)), and suggests both abilities are likely ancestral to all eukaryotes.

Materials and methods

Identification of *Naegleria* basal body genes

To determine if *Naegleria*'s basal bodies are likely formed from canonical proteins, known components were identified in the *Naegleria* genome (www.jgi.doe.gov/naegleria) by manual searches using Pfam domain annotations (Sonnhammer et al., 1998) and BLAST (Altschul et al., 1990) searches using homologs from a variety of genomes as queries. If no homolog was found, searches were repeated using other parameters and homologs.

Production of anti-SAS-6 antibody

To produce a polyclonal antibody recognizing *SAS-6*, a conserved 1087 base-pair internal exon of the gene containing amino acids 161-516 of the protein was cloned from *Naegleria* genomic DNA using primers SAS-6 forward (5'-CATGCCATGGGTAATTCTGATCCCTTCAGTGAAAGC -3'), and SAS-6 reverse (5'-TTTATAGCGGCCGCTTACATTAGAGTATCAATCGTAAATTG -3') into the NcoI and Not1 sites of the vector pET28c (EMD Biosciences, Gibbstown, NJ) that contains an N-terminal His-tag for protein purification. The vector was transformed into *E. coli* BL21(DE3) (EMD Biosciences), and confirmed by

sequencing. Binding to Ni-NTA beads under denaturing conditions was performed according to the manufacturer's instructions (Qiagen, Valencia, CA). Two rats were inoculated with purified protein according to a 118-day immunization protocol (Covance, Denver, PA). Sera from a single rat was precleared overnight at 4 degrees C with 1% NEG amoeba acetone powder. The resulting cleared sera recognized a single band of the appropriate weight (~74 kDa) on immunoblots of *Naegleria* flagellate extracts, and was used for subsequent analyses.

Production of anti-γ-tubulin antibody

To produce a polyclonal anibody recognizing *Naegleria* γ-tubulin (Genbank accession AAY17321), a 97 amino acid portion (amino acids 346-442) was selected and used for antibody production in two rabbits using genomic antibody technology immunization protocol (Strategic Diagnostics Inc., Newark DE). The same portion of the protein was also commercially expressed in bacteria, and used for affinity purification of sera pooled from both rabbits (Strategic Diagnostics Inc., Newark DE). The resulting sera recognized a single band of the appropriate weight (~55 kDa) on immunoblots of *Naegleria* flagellate extracts, and was used for subsequent analyses.

Immunoblotting and protein quantification

1x Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, Rockford, IL) and 5 mM EDTA were added to cells from each time point, and pelleted by centrifugation. The pellet was resuspended in 1x Laemmli buffer minus dithiothreitol and bromophenol blue supplemented with 1x Halt and 5 mM EDTA, boiled for 5 minutes, and then frozen. The protein concentration of each sample was determined using the DC Protein Assay (Bio-Rad, Hercules, CA) according to the manufacturer's instructions, and all samples were diluted to 2 mg/mL with 2x Laemmli buffer. 15 μg of cell extract was loaded onto 10% SDS-PAGE gels and separated by electrophoresis, then transferred to Immobilon-P membranes (Millipore, Billerica, MA). Membranes were blocked with a 5% milk/0.5% tween-20 solution, then probed with one or more of the following antibodies: a rabbit polyclonal anti-*NgCentrin* antibody at 1:500 (Levy et al., 1996), a rabbit polyclonal anti-*Naegleria* actin antibody at 1:10,000 (Fulton et al., 1986), a mouse monoclonal anti-*Naegleria* β-tubulin antibody at 1:100 (Walsh, 1984), the rabbit anti-*Naegleria* γ-tubulin antibody described above at 1:500 or the rat anti-*Naegleria* SAS-6 antibody described above at 1:2000. After washing, membranes were then probed with the appropriate horseradish peroxidase (HRP) conjugated secondary (anti-goat and anti-rabbit from Biorad (Hercules, CA), anti-rat from Jackson ImmunoResearch Laboratories, Inc. (West Grove, PA, USA)). The blots were visualized using Amersham ECL Western Blotting Detection Reagents (GE Healthcare, Little Chalfont, Buckinghamshire, England), according to the manufacturer's instructions.

Naegleria Differentiation and visualization of the cytoskeleton

Naegleria gruberi strain NEG (Fulton, 1970) amoeba were grown on solid medium in association with *Klebsiella pneumoniae,* and differented as described (Fulton, 1970). Differentiation was tracked via the appearance of flagella stained with Lugol's Iodine (5 % Iodine, 10 % KI) and visualized using a 40X phase objective.

Cells taken at each timepoint (0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, and 90 minutes after induction of differentiation) were added to an equal volume of fixation buffer (125 mM sucrose, 50 mM sodium phosphate, pH 7.2 with 2% paraformaldehyde) and fixed in suspension for 5 minutes. Cells were then smeared onto polyethyleneimine-coated coverslips, and allowed to continue fixing for 15 minutes. The coverslips were then washed three times in PEM buffer (100 mM PIPES, 1 mM EGTA, 0.1 mM $MgSO_4$), and permeabilized with 0.1% NP-40 diluted in PEM for 10 minutes. Samples were then blocked for one hour in Detector Block (Kirkegaard & Perry Laboratories, Gaithersburg, Maryland, USA), and incubated overnight in primary antibodies at the following dilution: 1:500 *Naegleria* polyclonal centrin primary antibody, 1:10 monoclonal anti- *Naegleria* α-tubulin and 1:10 anti-β-tubulin antibodies (Walsh, 1984), 1:100 anti-γ-tubulin tubulin antibody, 1:100 anti-NgSas6 antibody. Slides were washed three times in Detector Block, and incubated with secondary antibodies (Alexa Fluor 488 anti-rabbit IgG FAB fragment, Alexa Fluor 555 anti-mouse IgG, Alexa Fluor 647 anti-mouse IgG, and/or Cy3 anti-rat IgG (Invitrogen, Carlsbad, CA)) antibody diluted 1:200 in Detector Block for 3 hours at room temperature. Samples were washed three times with Detector Block, three times with PEM supplemented with 0.01% tritonx-100, and mounted with ProLong Gold antifade reagent with DAPI (Invitrogen). Centrin, γ-tubulin, and SAS-6 foci were counted from at least 100 cells at each time point using an Olympus BX51 fluorescence microscope with an Olympus PlanApo 100, NA 1.35, oil-immersion objective.

Fluorescence deconvolution microscopy

Images were collected with SoftWorX image acquisition software (Applied Precision, Issaquah, WA) on an Olympus IX70 wide-field inverted fluorescence microscope with an Olympus PlanApo 100, NA 1.35, oil-immersion objective and Photometrics CCD CH350 camera (Roper Scientific, Tuscon, AZ). Image stacks were deconvolved with the SoftWorX deconvolution software and flattened as maximum projections (Applied Precision, Issaquah, WA).

References:

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Beisson, J., and Wright, M. (2003). Basal body/centriole assembly and continuity. Curr Opin Cell Biol *15*, 96-104.

Bornens, M., and Azimzadeh, J. (2007). Origin and evolution of the centrosome. Adv Exp Med Biol *607*, 119-129.

Culver, B.P., Meehl, J.B., Giddings, T.H., Jr., and Winey, M. (2009). The two SAS-6 homologs in Tetrahymena thermophila have distinct functions in basal body assembly. Mol Biol Cell *20*, 1865-1877.

Fulton, C. (1970). Amebo-flagellates as research partners: The laboratory biology of *Naegleria* and *Tetramitus*. Methods Cell Physiol *4*, 341-476.

Fulton, C. (1977). Cell differentiation in Naegleria gruberi. Annu Rev Microbiol *31*, 597-629.

Fulton, C., and Dingle, A.D. (1967). Appearance of the flagellate phenotype in populations of Naegleria amebae. Dev Biol *15*, 165-191.

Fulton, C., and Dingle, A.D. (1971). Basal bodies, but not centrioles, in Naegleria. J Cell Biol *51*, 826-836.

Fulton, C., Lai, E.Y., Lamoyi, E., and Sussman, D.J. (1986). Naegleria actin elicits species-specific antibodies. J Protozool *33*, 322-327.

Khodjakov, A., Rieder, C.L., Sluder, G., Cassels, G., Sibon, O., and Wang, C.L. (2002). De novo formation of centrosomes in vertebrate cells arrested during S phase. J Cell Biol *158*, 1171-1181.

Kilburn, C.L., Pearson, C.G., Romijn, E.P., Meehl, J.B., Giddings, T.H., Jr., Culver, B.P., Yates, J.R., 3rd, and Winey, M. (2007). New Tetrahymena basal body protein components identify basal body domain structure. J Cell Biol *178*, 905-912.

Kramer, A., Neben, K., and Ho, A.D. (2002). Centrosome replication, genomic instability and cancer. Leukemia *16*, 767-775.

Levy, Y.Y., Lai, E.Y., Remillard, S.P., and Fulton, C. (1998). Centrin is synthesized and assembled into basal bodies during Naegleria differentiation. Cell Motil Cytoskeleton *40*, 249-260.

Levy, Y.Y., Lai, E.Y., Remillard, S.P., Heintzelman, M.B., and Fulton, C. (1996). Centrin is a conserved protein that forms diverse associations with centrioles and MTOCs in Naegleria and other organisms. Cell Motil Cytoskeleton *33*, 298-323.
Marshall, W.F. (2001). Centrioles take center stage. Curr Biol *11*, R487-496.

Marshall, W.F., Vucica, Y., and Rosenbaum, J.L. (2001). Kinetics and regulation of de novo centriole assembly. Implications for the mechanism of centriole duplication. Curr Biol *11*, 308-317.

Nakazawa, Y., Hiraki, M., Kamiya, R., and Hirono, M. (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. Curr Biol *17*, 2169-2174.

Pelletier, L., O'Toole, E., Schwager, A., Hyman, A.A., and Muller-Reichert, T. (2006). Centriole assembly in Caenorhabditis elegans. Nature *444*, 619-623.

Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G., and Glover, D.M. (2007). DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. Curr Biol *17*, 1465-1472.

Sonnhammer, E.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res *26*, 320-322.

Sorokin, S.P. (1968). Reconstructions of centriole formation and ciliogenesis in mammalian lungs. J Cell Sci *3*, 207-230.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., and Gonczy, P. (2007). Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. Dev Cell *13*, 203-213.

Suh, M.R., Han, J.W., No, Y.R., and Lee, J. (2002). Transient concentration of a gamma-tubulin-related protein with a pericentrin-related protein in the formation of basal bodies and flagella during the differentiation of Naegleria gruberi. Cell Motil Cytoskeleton *52*, 66-81.

Walsh, C. (1984). Synthesis and assembly of the cytoskeleton of Naegleria gruberi flagellates. J Cell Biol *98*, 449-456.

Wilson, E.B. (1928). In The Cell in Development and Heredity (London and New York, Macmillan), p. 357.

Table 3.1: *Naegleria*'s basal body gene homologs.

Not found ("nf") indicates genes that could not be identified in the genome sequence. Protein sequences can be accessed via the Joint Genome Institute's Genome Portal for the *Naegleria* Genome: www.jgi.doe.gov/naegleria

Gene Name	**JGI Protein ID(s)**
CenP-J/SAS-4/CPAP	61107
SAS-5	nf
SAS-6/BLD-12	68996
SF-assemblin	71540
BLD10	77710
PACRG1	29060, 82851
RIB43	83064
Centrin	56351, 44488
Pericentrin	nf
γ-tubulin	56069
δ-tubulin/UNI3	69007
ε Tubulin	65724
POC1	33676
POC11	70454
POC16	62107
MKS1/ POC12	52666
MKS3	62841

Fig 3.1. Protein induction of basal body/MTOC genes during differentiation

Cell extracts were collected from 14 different timepoints between 0 and 90 minutes. Western blots were probed using antibodies against SAS-6, Centrin, and γ-tubulin. Actin was used as a loading control. These blots indicate the following order of protein induction to detectable levels: SAS-6, centrin, β-tubulin, and finally γ-tubulin.

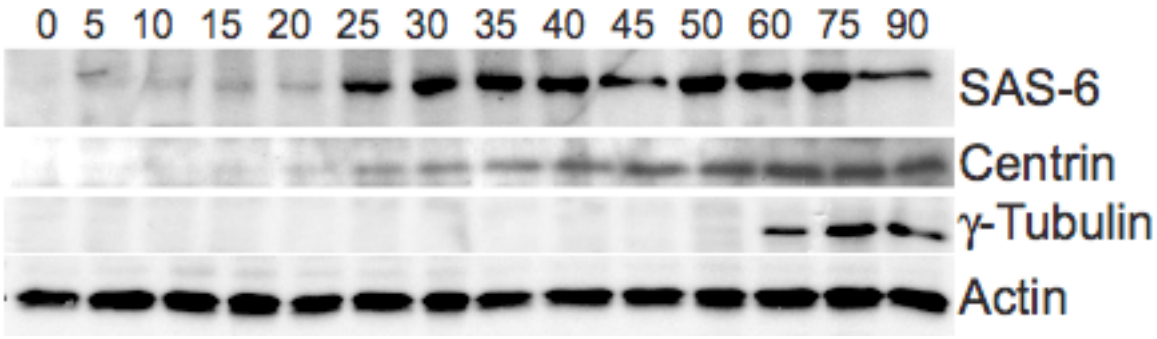


Fig 3.2. *Naegleria* γ-tubulin localization during differentiation

Naegleria cells were fixed at 30, 45, 60, and 90 minutes during differentiation and stained with antibodies to γ-tubulin (green) and α-tubulin plus β-tubulin (red), with DNA shown in blue. No appreciable localization any of thee proteins appear by 30 min. However, by 45 minutes a clear γ-tubulin-containing bar is present, developing into two bright foci by 60 minutes, which are localized to the base of the growing axonemes highlighted by tubulin staining. Basal body structures are indicated by an arrow, and enlarged in insets. Scale bar represents 10 microns.

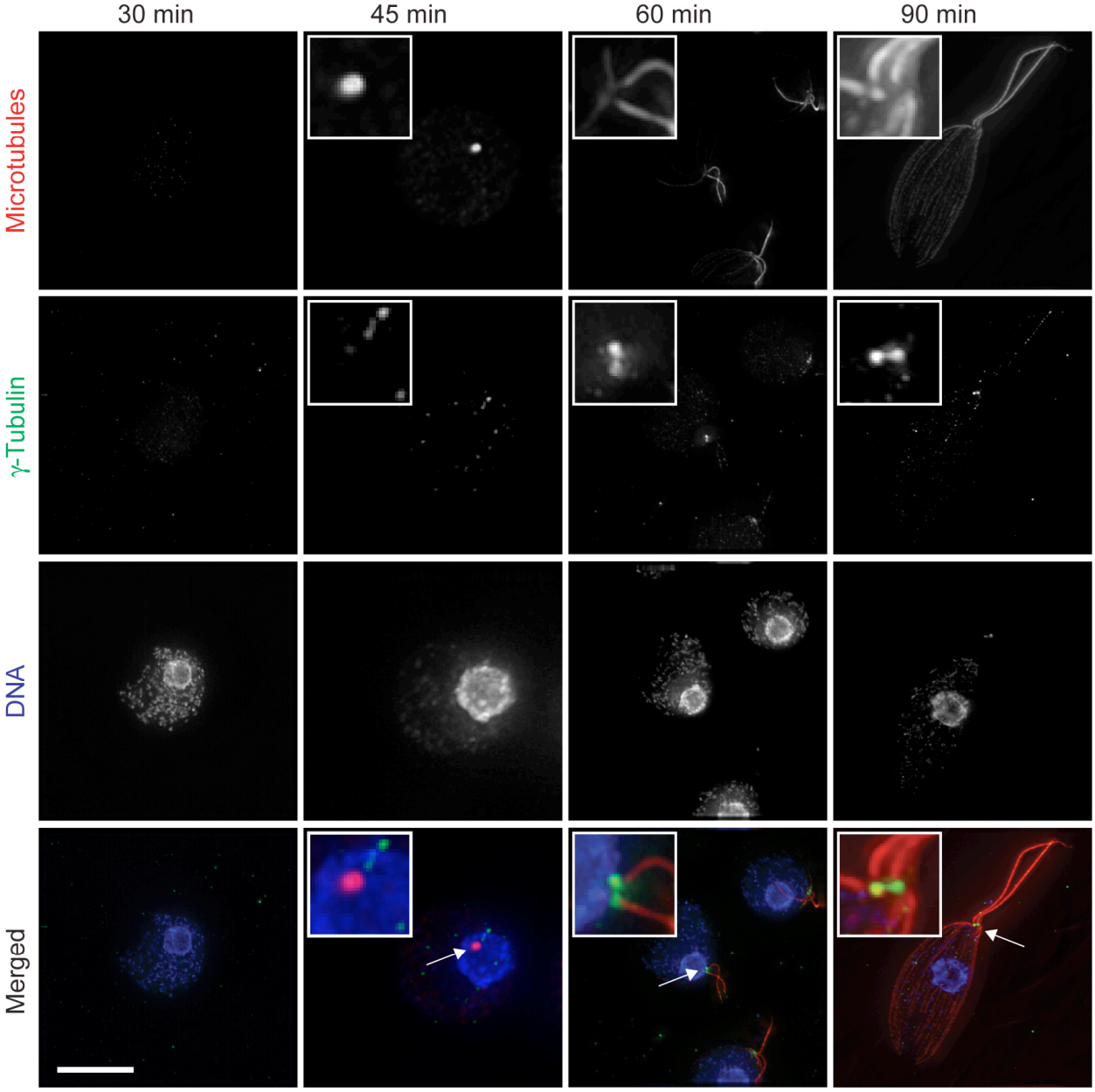


Fig 3.3. *Naegleria* SAS-6 and centrin localization during differentiation

Naegleria cells were fixed at 30, 45, 60, and 90 minutes during differentiation and stained with antibodies to SAS-6 (red), centrin (blue), and α-tubulin and β-tubulin (red), with DNA shown in gray. No appreciable localization any of thee proteins appear by 30 min. However, by 45 minutes a clear SAS-6- and centrin-containing bar is present with an associated dot, developing into two bright bar shaped foci at 60 and 90 minutes. By 45 minutes all bar shaped structures contain SAS-6 at the proximal end of the basal body with centrin localizig toward the growing axonemes highlighted by tubulin staining. Basal body structures are indicated by an arrow, and enlarged in insets. Scale bar represents 10 microns.

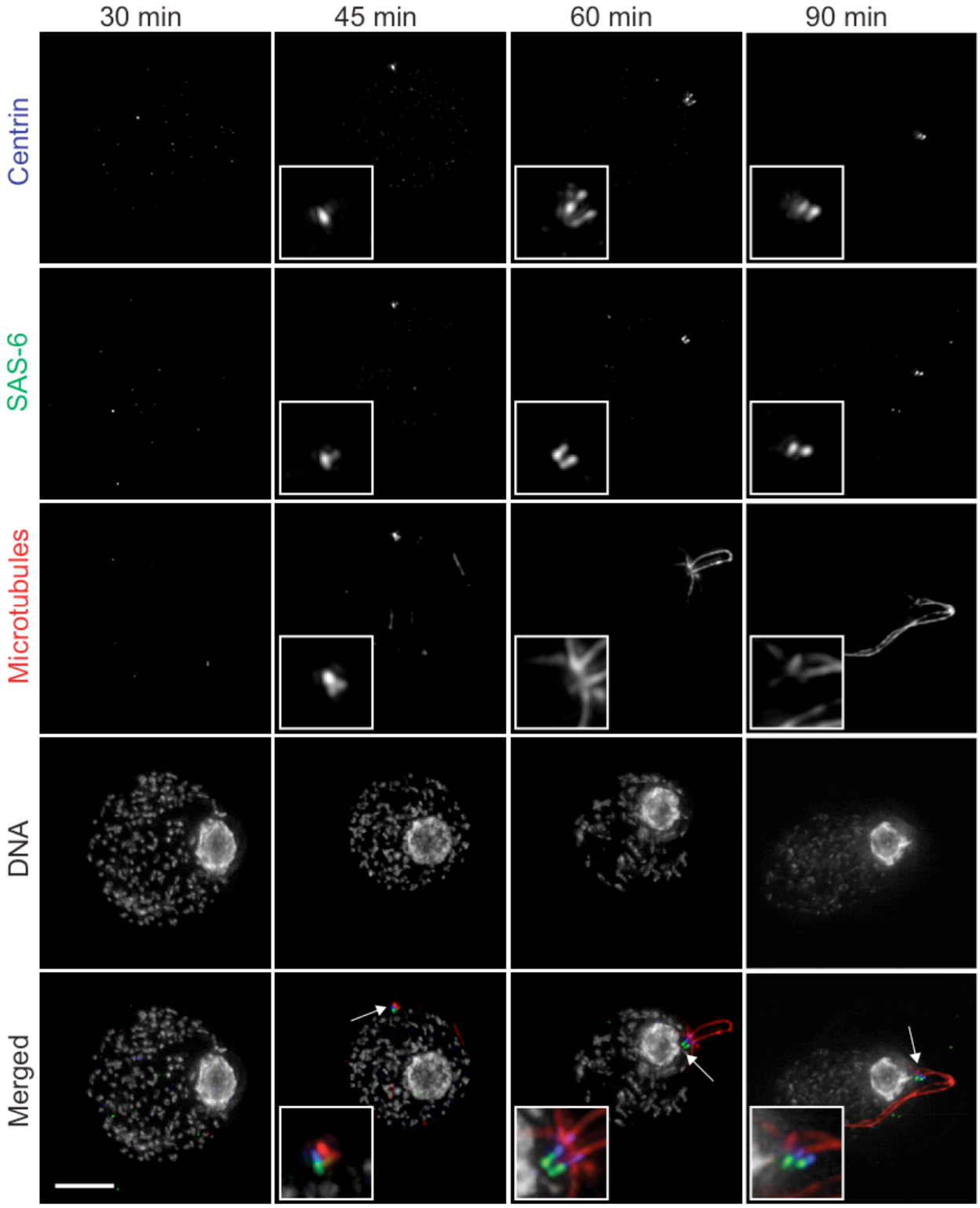

30 min
45 min
60 min
90 min
Centrin
SAS-6
Microtubules
DNA
Merged

Fig 3.4. Order of incorporation: SAS-6, then centrin, and finally γ-tubulin

At each timepoint, cells were fixed, stained with antibodies against SAS-6 (black), centrin (red), or γ-tubulin (blue), and at least 100 cells were scored for localization and times normalized to the time when 50% of cells have visible flagella. Three biological replicates are indicated by circles, squares and triangles.

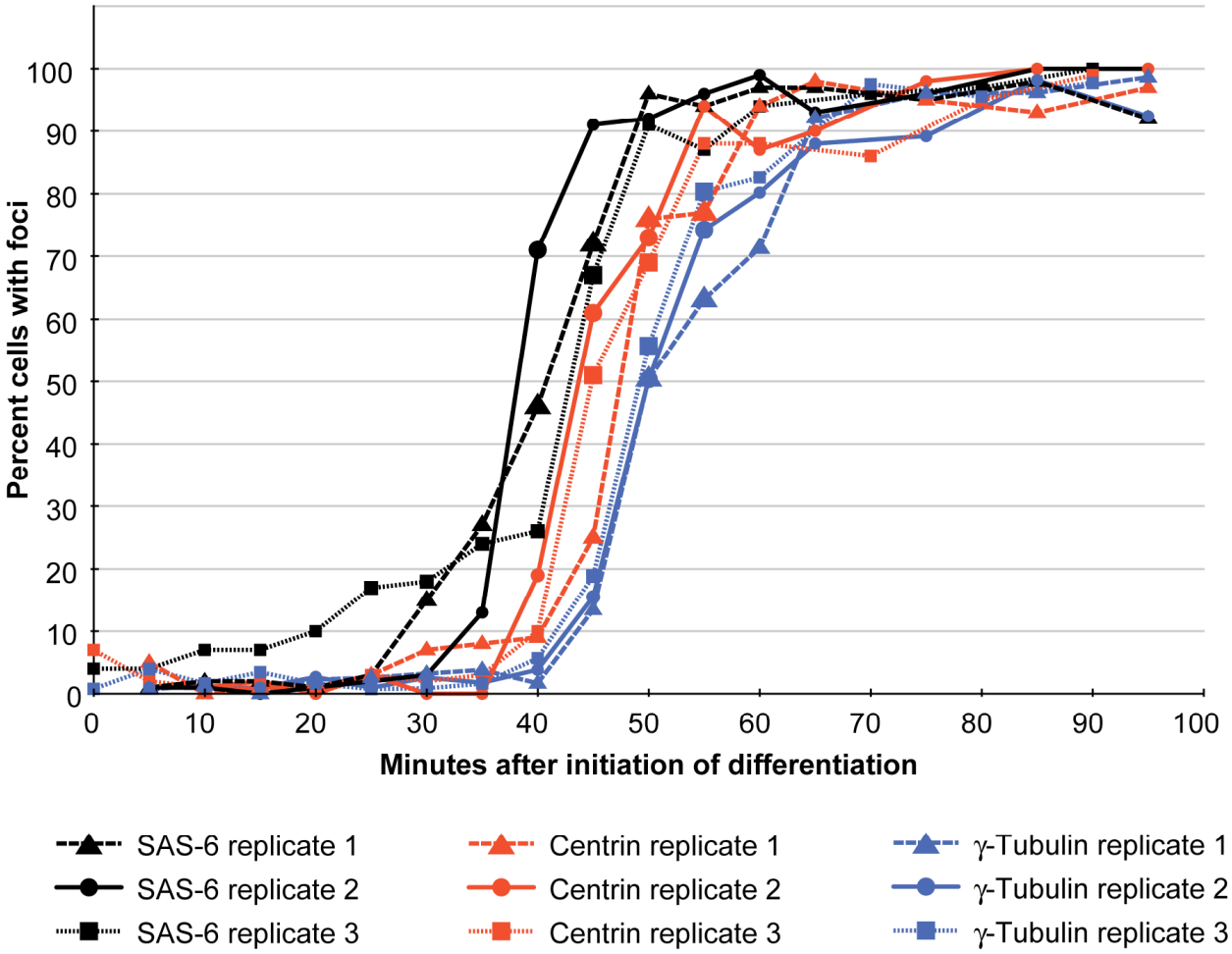
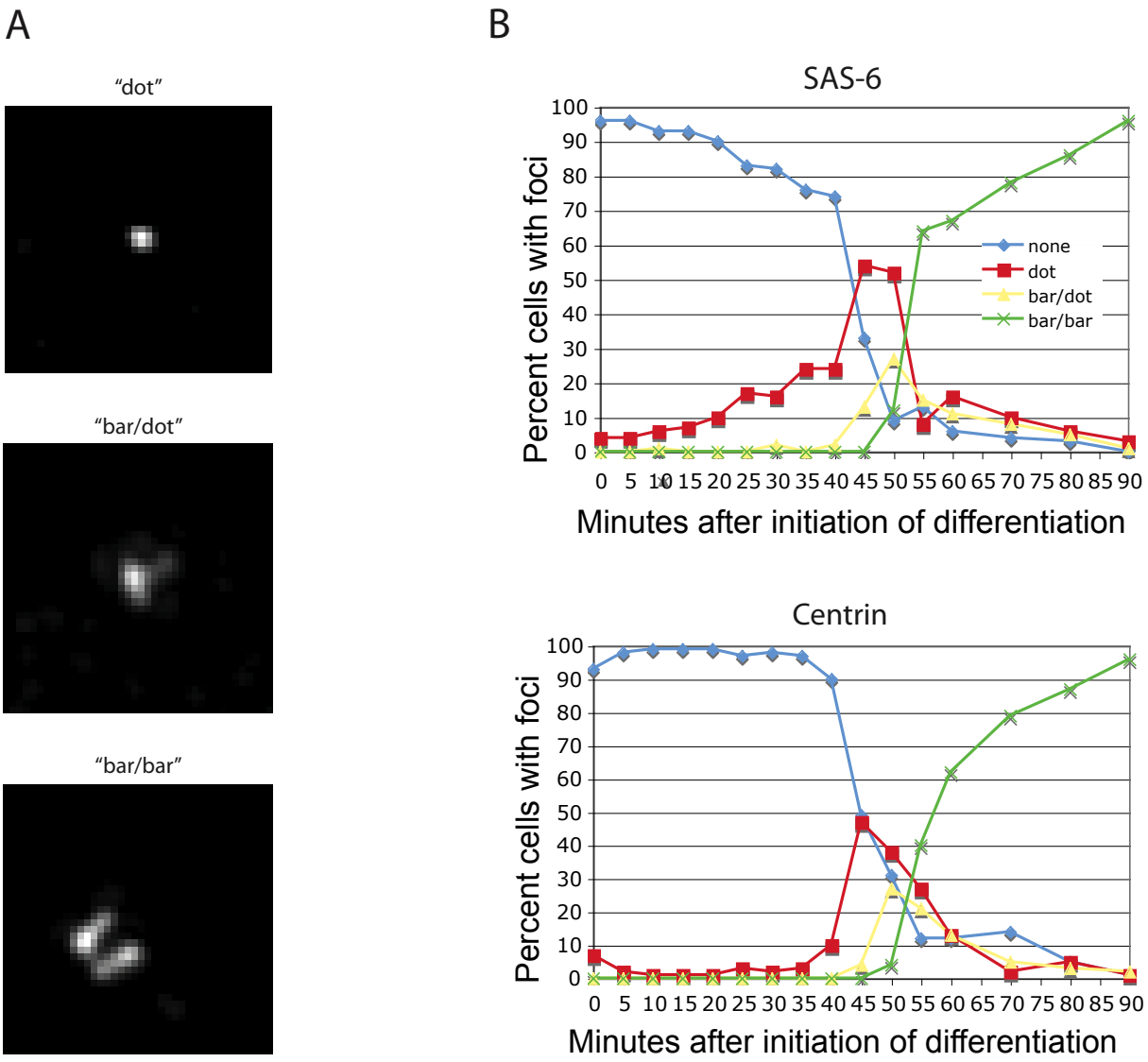


Fig 3.5. *Naegleria* has both templated and *de novo* basal body assembly

A.) Examples of "dot", "bar/dot", and "bar/bar" structures visible with either SAS-6 (shown) or centrin staining.

B.) Order of appearance of basal body protein-containing structures as visualized by both SAS-6 and centrin staining. One of three biological replicates with similar results is shown.



Chapter 4: Transcriptional analysis of *Naegleria* differentiation reveals novel ancient centrosome and flagella components

Abstract:

Flagellar and amoeboid motility are the two most common forms of locomotion used by eukaryotic cells. *Naegleria gruberi* is known for its remarkable ability to metamorphose from amoebae into streamlined biflagellates. This differentiation includes regulated synthesis of tubulin and other flagellar components, and *de novo* assembly of an entire cytoplasmic microtubule cytoskeleton, including canonical basal bodies and 9+2 flagella. Using phylogenetic profiling centered on *Naegleria* genes, we have previously identified protein families conserved only in eukaryotes with flagellar motility. These gene families include proteins required for flagellar beating, intraflagellar transport, and 36 novel flagellar-associated genes. In this study we further validated the predicted flagellar genes via microarray analysis of *Naegleria* differentiation, and use the timing of induction to subdivide them into a subset of 55 genes enriched in known basal body proteins (induced early) and a subset of 82 genes enriched in axonemal proteins (induced late). The centrosome-enriched set includes nearly every conserved basal body component that has been previously characterized, many components required for microtubule nucleation (a process that occurs largely at centrosomes) and ten novel genes that are conserved across eukaryotes. As a proof of principle, the human ortholog of one of the novel genes was tagged, and indeed localizes to the centrosomes of human cells.

Introduction:

Interphase animal cells contain numerous microtubules that emanate from the microtubule organizing centers (MTOCs) known as centrosomes. Centrosomes have two main parts: the beautiful and enigmatic pinwheel structures of centrioles, and the surrounding amorphous material that harbors the microtubule organizing activities of the centrosome. Metazoan centrosomes define mitotic spindle poles, and their centrioles are called basal bodies when used to form cilia after terminal differentiation (Sorokin, 1968). Moreover, Meeves (1900) showed in a series of classical experiments that centrioles and basal bodies are inter-convertible structures (Wilson, 1928). These small organelles are composed largely of nine microtubule triplets forming a cylinder approximately 0.2 microns in diameter 1 micron long (Fulton and Dingle, 1971).

Although centrioles are required for anchoring pericentriolar material to the centrosome (Bobinnec et al., 1998), the precise function of the complex organization of centrioles within the centrosomes of animal cells is unclear. However, centrioles (as basal bodies) do have a well-defined role in organizing cilia and flagella (Beisson

and Wright, 2003; Marshall, 2001). These whip-like eukaryotic structures propel single celled organisms and move fluids within multicellular organisms, and have received increasing attention in recent years with the discovery that ciliary defects can cause a myriad of diseases (Nigg and Raff, 2009). Many organisms also contain nonmotile cilia with many signaling receptors, and function as "antennae", gathering information about the surrounding environment for the cell (Marshall and Nonaka, 2006). For example, cells in our ears have cilia that detect changes in pressure that we interpret at sound.

The conserved function of centrioles and basal bodies is most likely their flagellar function, rather than mitosis. This is because organisms that lack cilia and flagella (e.g. plants, most fungi, and various protists) also lack centrioles, yet can undergo mitosis (Beisson and Wright, 2003). Furthermore, Basto *et al*. made flies without centrioles (via a mutation in SAS-4, a core centriolar component (Basto et al., 2006)). The cells of these flies are able to undergo mitosis successfully, but the animals fail to develop completely due to problems associated with their lack of cilia and flagella (Basto et al., 2006). It has therefore been proposed that centrioles are localized within the centrosomes to ensure proper disbursement of centrioles to daughter cells (Marshall, 2001).

Proteomic experiments indicate that centrosomes and basal body-containing MTOCs contain many of the same proteins (Andersen et al., 2003; Keller et al., 2005; Kilburn et al., 2007). A large number of the hundreds proteins identified in these proteomics experiments are thought to be components of basal bodies and centrioles (Andersen et al., 2003; Keller et al., 2005; Kilburn et al., 2007). However, only a handful of these components have been functionally characterized (Strnad and Gonczy, 2008). This is in part due the technical difficulty of studying centriole assembly in most systems: First, new centrioles usually assemble in association with a mature centriole, making proteomic characterization of intermediates impossible. Second, centriole assembly is usually tied to the cell cycle, rendering it difficult to distinguish genes specifically induced for the formation of centrioles from other induced cell cycle genes. And third, *de novo* assembly (where centrioles are built in the absence of pre-existing ones) can only be assayed in other systems in a single cell or embryo, making gathering enough material for proteomic or transcriptomic approaches unfeasible. All three of these technical roadblocks are overcome by using the single-celled eukaryote, *Naegleria gruberi*, to study centriole assembly.

The amoeboflagellate *Naegleria gruberi* grows as an amoeba that has no cytoplasmic microtubules: the amoeba's cytoplasm contains only actin-based cytoskeleton, and relies on a mitotic spindle that is always contained within the nuclear envelope and lacks centrioles (Fulton, 1970, 1977; Fulton and Dingle, 1971). However, when exposed to stressors such as temperature, osmotic, or pH changes, *Naegleria* rapidly differentiates into a flagellate, forming a complete cytoplasmic

microtubule cytoskeleton from scratch (Fulton and Dingle, 1967). This *de novo* assembly includes the formation of two basal bodies and flagella, as well as a cortical microtubule array anchored near the basal bodies (Dingle and Fulton, 1966). *Naegleria* differentiates synchronously, with approximately 90% of cells growing visible flagella in a 15 minute window (Figure 4.1). Evidence suggests that *Naegleria* expresses proteins in the same order as it incorporates them into developing basal bodies (see Chapter 3), about ten minutes before flagella assemble (Figure , 4.1 and (Fulton and Dingle, 1971)).

There are many overlapping functions between interphase mammalian centrosomes and the microtubule organizing center (MTOCs) of *Naegleria*; both contain centriole/basal bodies and the capacity to nucleate, anchor, and organize a cortical microtubule array. As *Naegleria*'s cell cycle is halted throughout the flagellate phase (Byers, 1986; Yuyama and Corff, 1978), its MTOC likely has no cell cycle functions. Therefore, *Naegleria* allows us to study specifically the conserved microtubule-based functions of an interphase centrosome, uncoupled from the other functions of mammalian centrosomes (such as signaling cascades and cell cycle functions). Furthermore, *Naegleria* forms an entire MTOC *de novo* (basal bodies and the components surrounding them which nucleate and anchor a cortical microtubule array) (Dingle and Fulton, 1966). Together, these features make *Naegleria* an important system with which to study centriole assembly, and more generally, the conserved features of the microtubule-specific interphase centrosome.

In this analysis, we follow expression of *Naegleria* genes during differentiation, and identify one set of genes likely used in the formation of centrosomes, and another set of genes likely required for flagellar assembly, many of which we confirm via proteomic profiling of *Naegleria* flagella. As centrosomes are particularly difficult to study in other systems, we focus further on the centrosome gene-enriched gene cluster. We identify ten novel genes in the centrosome-enriched gene cluster that are conserved through eukaryotic evolution. The human orthologs of at least one protein localizes to the centrosomes of human cells, providing initial validation of this analysis.

Results:

Flagella and basal body genes are transcriptionally induced with different kinetics

We isolated RNA from three biological replicates at 20 minute intervals during differentiation (at 0, 20, 40, 60 and 80 minutes, Figure 4.1, panel A). Approximately 24% of *Naegleria* genes are induced during differentiation, and about 39% are repressed during the amoeba-to-flagellate transition (4065 and 6484 genes, respectively, with at least 2-fold induction or repression, and p-values < 0.01, after correction for multiple testing). This represents a large percentage of the *Naegleria* genome and includes genes involved in regulating stress responses, metabolic changes, as well as those involved in the amoeba-to-flagellate transition.

Although over four thousand genes have increased transcription during differentiation, only a fraction of these are likely to be specific to the microtubule cytoskeleton. To aid in our search for novel and evolutionarily conserved centrosomal components, we focused our attention on those genes found in *Naegleria* and other flagellates, but missing in non-flagellated organisms (the FM gene set, see Chapter 2). This gene set includes genes required specifically for flagellar motility (i.e. components unique to basal bodies and/or flagella), but will not include components also used for other processes (such as alpha and beta tubulin, which are used for many structures in all eukaryotes). To this list we added *Naegleria* homologs of previously characterized microtubule cytoskeleton proteins (those found on Table 2.3) to allow analysis of proteins also used in non-flagellar roles (e.g. alpha-tubulin). As a way of tracking the specificity of our assay, we also followed the expression levels of the 63 genes conserved in organisms that undergo amoeboid movement, and missing in organisms that do not undergo amoeboid locomotion (the AM gene set, see Chapter 2).

Cluster analysis of the expression data for these 310 genes (the AM and FM gene sets, and *Naegleria* homologs of known cytoskeletal components) resulted in five major gene clusters (Figure 4.2). Two clusters contain mostly genes found in the FM gene set, indicating that they are enriched in genes required for flagellar assembly. The genes in these two clusters tend to have almost no expression at zero minutes and during early differentiation, which increases during differentiation, a behavior we would expect from basal body and flagellar genes. However, the two clusters differ in expression; one cluster contains genes that reach their peak expression levels by 20 minutes, and begin dropping by 40 minutes of differentiation, while the second cluster's expression largely reached peak levels by 40 minutes, and retained high expression patterns throughout differentiation. Manual inspection of these clusters revealed that the cluster with earlier expression contains genes encoding known centrosomal (centriolar and pericentriolar material) components (Table 4.1), and the cluster with later expression contains many flagellar components, as well as proteins required for flagellar assembly (Table 4.2). Based on the difference in induction of these two gene clusters, it seems *Naegleria* induces it genes in the same order as they are used, as *Naegleria* assembles its basal bodies before it assembles its flagella (T=55 and T=65 minutes, respectively).

Centrosome-enriched gene cluster

The 55 genes found in the centrosome gene cluster include *Naegleria* homologs of seven genes known to be required for centriole/basal body assembly (Table 4.1), including Epsilon- (Dutcher et al., 2002), Delta- (Dutcher and Trabuco, 1998; Garreau de Loubresse et al., 2001; O'Toole et al., 2003), and Eta-(Dutcher et al., 2002) tubulins, CenpJ/SAS-4/CPAP (Basto et al., 2006), SAS-6 (Pelletier et al., 2006), (Dammermann et al., 2004; Leidel et al., 2005) (Leidel et al., 2005; Nakazawa et al., 2007; Peel et al., 2007; Rodrigues-Martins et al., 2007; Strnad et al., 2007),

centrin (Baum et al., 1986; Kuchka and Jarvik, 1982; Winey et al., 1991; Koblenz et al., 2003; Salisbury et al., 2002; Taillon et al., 1992), and POC1 (Culver et al., 2009; Keller et al., 2009). This represents the majority of components shown to be required for centriole assembly that are conserved outside animals (the set is missing the evolutionary conserved protein POC5 which is required for centriole elongation in human cells (Azimzadeh et al., 2009), and SAS5 (Pelletier et al., 2006), which is only found in some invertebrate animals).

The centrosome-enriched gene cluster also contains homologs of microtubule nucleation factors (gamma-tubulin, GCP3, and GCP6 (Raynaud-Messina and Merdes, 2007)), as well as proteins required for the general microtubule organizational role of the centrosome (*e.g.* katanin p60, which can sever microtubules and has been localized to the centrosome (Hartman et al., 1998)). Additionally, the centrosome-enriched gene cluster contains several genes whose proteins localize to centrosomes, and ten previously uncharacterized genes which may have centrosomal function (Table 4.1).

Flagella and motile cilia contain 9 doublet microtubules surrounding an interal pair of microtubules, an arrangement known as "9+2" microtubules. Dynein motor proteins cause the outer microtubule doublets to slide past each other, producing movement. Unexpectedly, the centrosome-enriched cluster also contains nine homologs of axonemal dynein arm components, as well as a *Naegleria* homolog of kintoun/PF13 that has been shown in other organisms to be required for assembly of dynein arm complexes (Omran et al., 2008). This suggests that *Naegleria* assembles its flagellar dyneins in the cytoplasm well before flagellar assembly begins.

Flagella-enriched gene cluster

The flagella-enriched gene cluster contains 82 genes (Table S1). This set includes genes used for transporting proteins to the base of the growing flagellum (BBS components BBS1-5 and BBS7-9), and within the flagella to its growing tip (FLA3, kinesin 2, IFT20, IFT52, IFT57, IFT80, IFT88, IFT122, and IFT140), as well as structural components of the flagellum itself (*e.g.* PF20 and PF16, RSP4 and Rib72 (Pazour et al., 2005)). This gene set also includes 23 genes previously identified via proteomic analysis of *Chlamydomonas* flagella (Pazour et al., 2005), but have had no additional characterization. These genes are conserved across eukaryotic evolution, were found in proteomic analysis of *Chlamydomonas* flagella, and have an expression profile similar to other *Naegleria* flagellar components. Thus, these proteins are likely core components of eukaryotic flagella, and are prime candidates for functional analyses. Finally, the flagellar gene cluster contains 12 genes without any functional characterization related to flagella (although some have been identified in other flagella phylogenetic profiling analyses).

To further validate putative flagellar components, we conducted a proteomic analysis of *Naegleria* flagella. After deflagellating approximately 4*10^8 *Naegleria* flagellate cells, we separated the flagella from the cell bodies by low speed centrifugation, and then further purified the flagella using a sucrose step gradient. The resulting sample was composed largely of two proteins of the right size to be alpha and beta tubulin (Figure 4.2, panel B), as is typical for clean flagellar preparations (e.g. (Pazour et al., 2005)). To identify the other proteins in the sample, we did MUDPIT proteomics, which identified 415 proteins in the sample.

Of the 82 genes in the flagellar-enriched gene cluster, 23 were also identified in the proteomics analysis (Table 4.2), indicating that they are used within the flagellum itself (rather than used for flagellar function, but located within the cell body, such as the BBS proteins which traffic flagellar components to the base of the flagella). Included in this overlap are seven FAPs (proteins identified only from proteomic studies of *Chlamydomonas* flagella (Pazour et al., 2005)).

POC11 is a conserved centrosomal protein

The gene cluster enriched in centrosomal components includes ten genes that have not been previously localized or functionally characterized (Table 1), which we refer to as putative conserved centrosomal components (pCCCs). To determine if the pCCCs are likely to encode centrosomal components, we chose one to clone and localize as a proof of principle.

For this analysis we chose the human homolog of POC11, a protein well conserved across eukaryotes (Table 4.3, Figure 4.4). Although sequences from organisms from almost every branch of the eukaryotic tree align well, there are no identifiable domains in the conserved regions (data not shown). This new gene family may thus represent a novel class of centrosomal proteins. Although POC11 was originally identified in a proteomic analysis of *Chlamydomonas* basal bodies (Keller et al., 2005), it has not been shown to localize specifically to centrosomes or basal bodies, and has not been functionally analysed.

To verify that POC11 is a centrosomal component, we cloned its human cDNA into a C-terminal epitope tag vector, transiently expressed it in U2OS and HeLa cells, and stained the cells with antibodies to centrin-2 (a centriolar component). Localization of the human homolog results in heavy staining of the centrosomes of both U2OS and HeLa cells (Figure 4.5, and data not shown, respectively), suggesting that POC11 is indeed a novel, though conserved, centrosomal component. POC11 forms discrete foci near to each centrin foci (which mark the distal ends of centrioles, Figure 4.5), indicating that POC11 may be a novel component of the proximal end of centrioles.

Discussion:

As part of the *Naegleria* genome analysis (Fritz-Laylin et al., 2010), we previously identified a set of genes phylogenetically associated with flagellar motility. However, this data set, like previous phylogenetic profile analyses (Li et al., 2004; Merchant et al., 2007), contains flagella as well as centriole proteins. In this study, we use *Naegleria*'s unique ability to synchronously assemble basal bodies and then flagella, to separate proteins used to form centrioles, from proteins used to form flagella. Understanding how centrioles and flagella assemble and function requires a full inventory of components. This analysis is another step in that direction. As part of this study, we are planning to make our proteomics and microarray publicly available, and encourage others to perform additional analyses on these data. We suggest this may be a useful way of further validating putative, evolutionarily conserved, centrosomal and flagellar proteins.

Our analysis has identified a comprehensive set of centriole components, which contains virtually every known centriolar protein conserved outside the opistokhonts (a monophyletic clade containing both animals and fungi). Many additional proteins identified in this screen have been previously found in other large-scale analyses of centrioles and centriole assembly (although many have not been analysed experimentally) indicating that we may have identified the majority of evolutionarily conserved centriolar components. As we were specifically trying to identify genes conserved across evolution (and our analyses were designed to do so), it is quite likely that additional genes are used specifically for *Naegleria* basal body and flagellar assembly. Likewise, genes that are required for centrosome functions in a particular group of organisms would not be included in this data set (e.g. PLK4, a gene limited to opisthokonts, Zita Santos, personal communication).

The centrosome-enriched gene cluster also contains ten evolutionarily conserved genes that have had no functional characterization. We predict these genes have microtubule-specific centrosomal functions. As a proof-of-principle experiment, we localized the human ortholog of POC11 in interphase cells. POC11 indeed localizes near centrioles, and may represent a novel component of the proximal end of the centriole, the site of centriolar outgrowth. We are currently localizing the remaining nine novel genes, and are attempting functional experiments on POC11, as well as several other of the novel gene set.

This study represents the first broad characterization of *Naegleria* flagella, including both proteomics and transcriptional verification of protein components. In particular, this analysis suggests that one mechanism *Naegleria* uses to rapidly assemble its flagella is to assemble its flagellar dynein complexes while initial microtubule structures (e.g. a MTOC, and initial cytoplasmic microtubules) are still being built, well before flagellar outgrowth actually begins. Also included in the centrosome enriched gene cluster are several proteins whose *Chlamydomonas*

homologs were identified in a flagellar proteomic analysis (Pazour et al., 2005). Because they are expressed with the same kinetics as other flagellar dynein complex proteins, these proteins may represent additional conserved, and as yet uncharacterized, components of the flagellar dynein arm complexes.

The analysis of the flagella-enriched gene cluster also suggests that many of the proteins previously identified in proteomic profiling of *Chlamydomonas* flagella (e.g. FAPs, (Pazour et al., 2005)) are evolutionarily conserved protein components of flagella. In particular, seven otherwise uncharacterized FAPs are contained in the *Naegleria* flagella-enriched gene cluster and were found in the *Naegleria* flagellar proteome. We are currently performing localization experiments on one of these as a proof-of-principle that our analysis can confirm uncharacterized flagellar proteins.

Previous analyses have used proteomic profiling to attempt to identify a complete centriole (Keller et al., 2005; Kilburn et al., 2007) or centrosome (Andersen et al., 2003) proteome. In this analysis, we overlay a previous comparative genomics analysis with transcriptional analysis of *Naegleria* differentiation to identify proteins required specifically for flagellar assembly or centriole assembly. Each of these approaches has advantages. Proteomics analyses can, potentially, identify all proteins contained in a structure, including those also required for other biological functions. Our analysis, by definition, will identify specifically those proteins required specifically for assembly of a flagellar apparatus, including basal bodies. Using the unique biology of *Naegleria* differentiation, we can further specify which of these proteins are likely to be required for organizing microtubules into the beautiful and enigmatic structures called centrioles.

Analyses in various organisms have recovered a shared core set of basal body/centriole components (Strnad and Gonczy, 2008). However, each organism's centriole proteome also contains many more proteins that are not conserved (Andersen et al., 2003; Keller et al., 2005; Kilburn et al., 2007). The fact that each organism has a seemingly unique set of centriolar or centrosomal components is particularly striking because the structural organization of centrioles is well conserved across nearly all of eukaryotic evolution (with the exception of lineages which have lost flagella and basal bodies). Does this mean that most basal body components are *not* required for their structure, and are simply found there for various organism-specific functions? Or, does this mean that there are multiple ways of achieving the same structure, and assembly pathways have diverged through eukaryotic evolution? The latter would be particularly exciting, as it would indicate that the centriole, an extremely well-organized organelle, can be assembled in multiple, evolutionarily malleable, ways. These questions can only be answered with additional functional experiments in a variety of organisms. To fully realize their potential for understanding the evolution of centrioles, these analyses will need to compare pathways of centriole assembly from organisms spread throughout the eukaryotic tree, such as those ongoing in the ciliates (*Paramecium*

and *Tetrahymena*), algae (*Chlamydomonas*), and animals, as well as those in emerging (or rather, re-emerging) model organisms such as *Naegleria*.

Materials and Methods:

Naegleria differentiation and RNA isolation:

For hybridization with whole-genome oligoarray, a series of three independent biological replicates was obtained from differentiating *N. gruberi*, strain NEG grown on Klebsiella, using standard protocols (Fulton, 1970). Synchrony of differentiation was estimated by fixing in lugols iodine (Fulton and Dingle, 1967), and using a phase-contrast microscope with a 40X objective to count the percentage of cells with flagella after (Figure 1). Approximately 10^7 cells were harvested at 0, 20, 40, 60 and 80 minutes after initiation of differentiation. RNA was extracted using Trizol Reagent (Invitrogen), followed by RNAeasy columns (Qiagen), treated with Turbo DNAse (Ambion), and re-purified over another RNAeasy column (Qiagen), all according to the manufacturers instructions. RNA purity was verified using both gel electrophoresis (Figure 4.2, panel A) and spectrophotometry.

NimbleGen Expression Oligoarrays

The *N. gruberi* whole-genome expression oligoarray version 1.0 (NimbleGen Systems) consists of 182,813 probe sets corresponding to 15,777 gene models predicted on the *N. gruberi* genome sequence version 1.0 (Fritz-Laylin et al., 2010), and an additional 963 orfs identified in intergenic regions. For each gene, 11 unique 60-mer oligonucleotide probes were designed by NimbleGen Systems using a multi-step approach to select probes with optimal predicted hybridization characteristics. The *Naegleria* version 1.0 oligoarray will be fully described and publicly available via the Gene Expression Omnibus (GEO) at NCBI (http://www.ncbi.nlm.nih.gov/geo) in the near future.

Preparation of samples, hybridization and scanning were performed by NimbleGen Systems Inc. (Madison, WI USA), following their standard operating protocol. The raw data was subjected to RMA (Robust Multi-Array Analysis; (Irizarry et al., 2003)), quantile normalization (Bolstad et al., 2003), and background correction as implemented in the NimbleScan software package, version 2.4.27 (Roche NimbleGen, Inc.). Average expression levels were calculated for each gene from the independent probes on array and were used for further analysis. Reproducibility between biological replicates was inspected using MA and scatter plots of log intensities constructed from the above data using the R statistical package (http://www.r-project.org/) (Figure 4.2, panel C).

Expression clustering

As we were interested in following the expression of cytoskeletal genes during *Naegleria* differentiation, the 173 FMs, 63 AMs, as well as known microtubule cytoskeletal components were selected for further study. Together the log-transformed expression data for these 310 genes were subjected to gene normalization followed by hierarchical clustering, with centered correlation and complete linkage in the Cluster program (Eisen et al., 1998).

For each gene, fold expression was calculated for the average expression of each replicate, and P-values were calculated in a simple paired data comparison model with all gene probe duplicates considered independently. All p-values were corrected for multiple testing using the BH (false discovery rate controlled) procedure within the R statistical package (http://www.r-project.org/).

Proteomics of *Naegleria* Flagella

Cells in suspension culture (4x108, 110ml) were pelleted at room temperature by low speed centrifugation in a Clinical centrifuge (#6, 2 minutes). The pellets were suspended in ice cold 10 ml Detailing medium (10 mM Na acetate, pH 3.7, 2.5 mM MgCl2, 75 mM sucrose, 0.1mM EGTA, 1 mM PMSF, vortexed for 15 sec, then 0.5ml Neutralizing buffer (0.5M Tris, pH8.25)was added. Cell bodies were removed by three low speed centrifugations (500xg for 3, 4 and 4 minutes). The axonemes were then pelleted by high speed centrifugation (30,000xg for 20 min). The axonemes were pelleted by high spend centrifugation (30,000xg for 30min), and resuspended in buffer ZC (25mM Tris (pH 7.6), 3 mM MgCl2, 1mM EGTA, 0.1mM DTT and 1mM PMSF) and layered on top of a sucrose step gradient (1.2, 1.4, 1.6, 1.8 and 2.0 M sucrose in buffer ZC). The gradient and axonemes were centrifuged for 80 min at 13,000xg in a swinging bucket rotor, and the gradient was hand-fractionated. Axonemes, but not cell bodies, were observed by phase microscopy at the 1.4-1.6 M sucrose interface. The interface fraction was collected and diluted 10-fold with buffer ZC, and axonemes were pelleted at high speed centrifugation (30,000 xG for 30 minutes). Ten percent of the resulting pellet was run on a 10% SDS-PAGE gel (using standard protocols), and the gel was stained using SYPRO Ruby Protein Gel stain, according to the manufacturer's protocols (Invitrogen). Proteins were precipitated by addition of TCA to 20% and the pellets were washed 3 times in ice-cold 0.01MHCl/90% acetone and allowed to air dry. Protein samples were analyzed via MUDPIT mass spectrometry as described in (Keller et al., 2005)

Localization of POC11

POC11 cDNA was obtained from Open Biosystems (accession BC006444). The ORF was amplified using the following forward (POC11F) and reverse (POC11R) primers (containing attB sites):

POC11F:
5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGAAGCATGAACTTTACCCCA
ACACACAC3',

POC11R:
5'GGGGACCACTTTGTACAAGAAAGCTGGGTCACAGAGTCTAAGTTCATTC3'

The resulting PCR product was transferred into a Gateway donor vector (pDONR221, Invitrogen) according to the manufacturers instructions, and verified by sequencing. The POC11 ORF was then transferred into the C-terminal V5 epitope-tagged pcDNA-DEST40 Gateway Vector (Invitrogen) according to the manufacturer's protocol.

Approximately 2*10^4 cells were plated in in 0.5 mL media [DMEM (GIBCO catalog #10569) supplemented with 10% FBS, 1% nonessential amino acids, and 1% sodium pyruvate] onto coated coverslips in 24-well plates. The next day, the cells were transfected with the POC11 expression vector using Lipofectamine 2000, according to manufacturers guidelines. Cells were fixed 14 hours later with -20° methanol for three minutes, rehydrated in TBS-T (TBS with 0.1% TritonX-100) three times, for 5 minutes each. Cells were blocked overnight at 4° in Abdil (1X TBS with 2% BSA and 0.1% TritonX-100), and stained for 1 hour at room temperature in the appropriate antibodies (see below). Samples were washed three times in Abdil, and incubated with secondary antibodies for 1 hour. Samples were then washed 3X with 1 mL Abdil, 3X in TBS-T, and mounted with ProLong Gold antifade reagent with DAPI (Invitrogen).

Antibodies

Centrin protein was visualized via a mouse anti-centrin antibody (20H5, (Sanders and Salisbury, 1994)), used at 1:400. V5-tagged POC11 protein was visualized using a chicken polyclonal antibody (ab9113, Abcam) at a 1:500 dilution. Secondary antibodies were both diluted to 1:500: Cy3 conjugated anti-chicken F(ab')2 fragment (catalog number 703-166-155, Jackson ImmunoResearch), Alexa Fluor 647 anti-mouse IgG (Invitrogen, Carlsbad, CA).

Fluorescence deconvolution microscopy

Images were collected with SoftWorX image acquisition software (Applied Precision, Issaquah, WA) on an Olympus IX70 wide-field inverted fluorescence microscope with an Olympus PlanApo 100, NA 1.35, oil-immersion objective and Photometrics CCD CH350 camera (Roper Scientific, Tuscon, AZ). Image stacks were deconvolved with the SoftWorX deconvolution software and flattened as maximum projections (Applied Precision, Issaquah, WA).

Multiple sequence alignment

POC11 homologs were collected by searching the nr database at NCBI (Benson et al., 2009) with BLAST (Altschul et al., 1990). An initial alignment was built using ClustalW (Thompson et al., 1997), and the alignment of individual amino acids manually edited in JalView (Clamp et al., 2004).

References:

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. Nature *426*, 570-574.

Azimzadeh, J., Hergert, P., Delouvee, A., Euteneuer, U., Formstecher, E., Khodjakov, A., and Bornens, M. (2009). hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. J Cell Biol *185*, 101-114.

Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C.G., Khodjakov, A., and Raff, J.W. (2006). Flies without centrioles. Cell *125*, 1375-1386.

Baum, P., Furlong, C., and Byers, B. (1986). Yeast gene required for spindle pole body duplication: homology of its product with Ca2+-binding proteins. Proc Natl Acad Sci U S A *83*, 5512-5516.

Beisson, J., and Wright, M. (2003). Basal body/centriole assembly and continuity. Curr Opin Cell Biol *15*, 96-104.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2009). GenBank. Nucleic Acids Res *37*, D26-31.

Bobinnec, Y., Khodjakov, A., Mir, L.M., Rieder, C.L., Edde, B., and Bornens, M. (1998). Centriole disassembly in vivo and its effect on centrosome structure and function in vertebrate cells. J Cell Biol *143*, 1575-1589.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Byers, T.J. (1986). Molecular biology of DNA in Acanthamoeba, Amoeba, Entamoeba, and Naegleria. Int Rev Cytol *99*, 311-341.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004). The Jalview Java alignment editor. Bioinformatics *20*, 426-427.

Dammermann, A., Muller-Reichert, T., Pelletier, L., Habermann, B., Desai, A., and Oegema, K. (2004). Centriole assembly requires both centriolar and pericentriolar material proteins. Dev Cell *7*, 815-829.

Dawe, H.R., Smith, U.M., Cullinane, A.R., Gerrelli, D., Cox, P., Badano, J.L., Blair-Reid, S., Sriram, N., Katsanis, N., Attie-Bitach, T., *et al.* (2007). The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum Mol Genet *16*, 173-186.

Dingle, A.D., and Fulton, C. (1966). Development of the flagellar apparatus of *Naegleria*. J Cell Biol *31*, 43-54.

Dutcher, S.K. (2001). The tubulin fraternity: alpha to eta. Curr Opin Cell Biol *13*, 49-54.

Dutcher, S.K., Morrissette, N.S., Preble, A.M., Rackley, C., and Stanga, J. (2002). Epsilon-tubulin is an essential component of the centriole. Mol Biol Cell *13*, 3859-3869.

Dutcher, S.K., and Trabuco, E.C. (1998). The UNI3 gene is required for assembly of basal bodies of Chlamydomonas and encodes delta-tubulin, a new member of the tubulin superfamily. Mol Biol Cell *9*, 1293-1308.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J., Carpenter, M.L., Field, M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J., *et al.* (2010). The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. Cell *In press*.

Fulton, C. (1970). Amebo-flagellates as research partners: The laboratory biology of *Naegleria* and *Tetramitus*. Methods Cell Physiol *4*, 341-476.

Fulton, C. (1977). Cell differentiation in Naegleria gruberi. Annu Rev Microbiol *31*, 597-629.

Fulton, C., and Dingle, A.D. (1967). Appearance of the flagellate phenotype in populations of Naegleria amebae. Dev Biol *15*, 165-191.

Fulton, C., and Dingle, A.D. (1971). Basal bodies, but not centrioles, in Naegleria. J Cell Biol *51*, 826-836.

Garreau de Loubresse, N., Ruiz, F., Beisson, J., and Klotz, C. (2001). Role of delta-tubulin and the C-tubule in assembly of Paramecium basal bodies. BMC Cell Biol *2*, 4.

Gorden, N.T., Arts, H.H., Parisi, M.A., Coene, K.L., Letteboer, S.J., van Beersum, S.E., Mans, D.A., Hikida, A., Eckert, M., Knutzen, D., *et al.* (2008). CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. Am J Hum Genet *83*, 559-571.

Hammond, J.W., Cai, D., and Verhey, K.J. (2008). Tubulin modifications and their cellular functions. Curr Opin Cell Biol *20*, 71-76.

Harrison, A., Olds-Clarke, P., and King, S.M. (1998). Identification of the t complex-encoded cytoplasmic dynein light chain tctex1 in inner arm I1 supports the involvement of flagellar dyneins in meiotic drive. J Cell Biol *140*, 1137-1147.

Hartman, J.J., Mahr, J., McNally, K., Okawa, K., Iwamatsu, A., Thomas, S., Cheesman, S., Heuser, J., Vale, R.D., and McNally, F.J. (1998). Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. Cell *93*, 277-287.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Kagami, O., Gotoh, M., Makino, Y., Mohri, H., Kamiya, R., and Ogawa, K. (1998). 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 *211*, 383-386.

Keller, L.C., Geimer, S., Romijn, E., Yates, J., 3rd, Zamora, I., and Marshall, W.F. (2009). Molecular architecture of the centriole proteome: the conserved WD40 domain protein POC1 is required for centriole duplication and length control. Mol Biol Cell *20*, 1150-1166.

Keller, L.C., Romijn, E.P., Zamora, I., Yates, J.R., 3rd, and Marshall, W.F. (2005). Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr Biol *15*, 1090-1098.

Kilburn, C.L., Pearson, C.G., Romijn, E.P., Meehl, J.B., Giddings, T.H., Jr., Culver, B.P., Yates, J.R., 3rd, and Winey, M. (2007). New Tetrahymena basal body protein components identify basal body domain structure. J Cell Biol *178*, 905-912.

Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S., and Hyman, A.A. (2003). SAS-4 is a C. elegans centriolar protein that controls centrosome size. Cell *112*, 575-587.

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.D., and Nigg, E.A. (2007). Plk4-induced centriole biogenesis in human cells. Dev Cell *13*, 190-202.

Koblenz, B., Schoppmeier, J., Grunow, A., and Lechtreck, K.F. (2003). Centrin deficiency in Chlamydomonas causes defects in basal body replication, segregation and maturation. J Cell Sci *116*, 2635-2646.

Kuchka, M.R., and Jarvik, J.W. (1982). Analysis of flagellar size control using a mutant of Chlamydomonas reinhardtii with a variable number of flagella. J Cell Biol *92*, 170-175.

Lachmann, M., Gelbmann, D., Kalman, E., Polgar, B., Buschle, M., Von Gabain, A., Szekeres-Bartho, J., and Nagy, E. (2004). PIBF (progesterone induced blocking factor) is overexpressed in highly proliferating cells and associated with the centrosome. Int J Cancer *112*, 51-60.

Leidel, S., Delattre, M., Cerutti, L., Baumer, K., and Gonczy, P. (2005). SAS-6 defines a protein family required for centrosome duplication in C. elegans and in human cells. Nat Cell Biol *7*, 115-125.

Leidel, S., and Gonczy, P. (2003). SAS-4 is essential for centrosome duplication in C elegans and is recruited to daughter centrioles once per cell cycle. Dev Cell *4*, 431-439.

Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y., Teslovich, T.M., May-Simera, H., Li, H., Blacque, O.E., Li, L., Leitch, C.C.*, et al.* (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell *117*, 541-552.

Marshall, W.F. (2001). Centrioles take center stage. Curr Biol *11*, R487-496.

Marshall, W.F., and Nonaka, S. (2006). Cilia: tuning in to the cell's antenna. Curr Biol *16*, R604-614.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L.*, et al.* (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245-250.

Nakazawa, Y., Hiraki, M., Kamiya, R., and Hirono, M. (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. Curr Biol *17*, 2169-2174.

Nigg, E.A., and Raff, J.W. (2009). Centrioles, centrosomes, and cilia in health and disease. Cell *139*, 663-678.

O'Toole, E.T., Giddings, T.H., McIntosh, J.R., and Dutcher, S.K. (2003). Three-dimensional organization of basal bodies from wild-type and delta-tubulin deletion strains of Chlamydomonas reinhardtii. Mol Biol Cell *14*, 2999-3012.

Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N.T., Hagiwara, H., Zhang, Q., Leblond, G., O'Toole, E., Hara, C., *et al.* (2008). Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. Nature *456*, 611-616.

Pazour, G.J., Agrin, N., Leszyk, J., and Witman, G.B. (2005). Proteomic analysis of a eukaryotic cilium. J Cell Biol *170*, 103-113.

Peel, N., Stevens, N.R., Basto, R., and Raff, J.W. (2007). Overexpressing centriole-replication proteins in vivo induces centriole overduplication and de novo formation. Curr Biol *17*, 834-843.

Pelletier, L., O'Toole, E., Schwager, A., Hyman, A.A., and Muller-Reichert, T. (2006). Centriole assembly in Caenorhabditis elegans. Nature *444*, 619-623.
Raynaud-Messina, B., and Merdes, A. (2007). Gamma-tubulin complexes and microtubule organization. Curr Opin Cell Biol *19*, 24-30.

Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G., and Glover, D.M. (2007). DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. Curr Biol *17*, 1465-1472.

Salisbury, J.L., Suino, K.M., Busby, R., and Springett, M. (2002). Centrin-2 is required for centriole duplication in mammalian cells. Curr Biol *12*, 1287-1292.
Sanders, M.A., and Salisbury, J.L. (1994). Centrin plays an essential role in microtubule severing during flagellar excision in Chlamydomonas reinhardtii. J Cell Biol *124*, 795-805.

Sorokin, S.P. (1968). Reconstructions of centriole formation and ciliogenesis in mammalian lungs. J Cell Sci *3*, 207-230.

Strnad, P., and Gonczy, P. (2008). Mechanisms of procentriole formation. Trends Cell Biol *18*, 389-396.

Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., and Gonczy, P. (2007). Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. Dev Cell *13*, 203-213.

Taillon, B.E., Adler, S.A., Suhan, J.P., and Jarvik, J.W. (1992). Mutational analysis of centrin: an EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of Chlamydomonas. J Cell Biol *119*, 1613-1624.

Tam, L.W., and Lefebvre, P.A. (2002). The Chlamydomonas MBO2 locus encodes a conserved coiled-coil protein important for flagellar waveform conversion. Cell Motil Cytoskeleton *51*, 197-212.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res *25*, 4876-4882.

Wilson, E.B. (1928). In The Cell in Development and Heredity (London and New York, Macmillan), p. 357.

Winey, M., Goetsch, L., Baum, P., and Byers, B. (1991). MPS1 and MPS2: novel yeast genes defining distinct steps of spindle pole body duplication. J Cell Biol *114*, 745-754.

Wloga, D., Rogowski, K., Sharma, N., Van Dijk, J., Janke, C., Edde, B., Bre, M.H., Levilliers, N., Redeker, V., Duan, J., *et al.* (2008). Glutamylation on alpha-tubulin is not essential but affects the assembly and functions of a subset of microtubules in Tetrahymena thermophila. Eukaryot Cell *7*, 1362-1372.

Yuyama, S., and Corff, S. (1978). Differentiation-dependent decline of DNA synthetic activities in Naegleria gruberi. J Protozool *25*, 408-415.

Table 4.1: Centrosomal enriched gene cluster

The 55 genes in the centrosome-enriched cluster are organized by function (or predicted function). The second column contains ortholog gene and/or protein names. Cases where orthology was not identifiable are indicated by a question mark. The third column contains JGI protein IDs for the *Naegleria* orthologs (sequence information is available at www.jgi.doe.gov/naegleria). Function and/or localization of orthologs is based on experiments referenced in the right-hand column.

	Name	JGI Peptide ID	Function	Reference
Basal body/centriole Assembly	Epsilon Tubulin	44774	Basal body assembly	(Dutcher et al., 2002)
	Delta Tubulin	69007	Basal body assembly	(Dutcher and Trabuco, 1998; Garreau de Loubresse et al., 2001; O'Toole et al., 2003)
	CenpJ/SAS-4/CPAP	61107	Basal body assembly	(Kirkham et al., 2003), (Leidel and Gonczy, 2003; Pelletier et al., 2006) (Basto et al., 2006; Kleylein-Sohn et al., 2007)
	SAS-6	68996	Basal body assembly	(Pelletier et al., 2006), (Dammermann et al., 2004; Leidel et al., 2005) (Leidel et al., 2005; Nakazawa et al., 2007; Peel et al., 2007; Rodrigues-Martins et al., 2007; Strnad et al., 2007)
	Centrin	44488 56351	Basal body assembly	(Baum et al., 1986) and (Kuchka and Jarvik, 1982; Winey et al., 1991), (Koblenz et al., 2003; Salisbury et al., 2002; Taillon et al., 1992)
	Eta Tubulin	65724	Basal body assembly	(Dutcher et al., 2002)
	POC1	33676	Basal body assembly	(Keller et al., 2009)
Microtubule-specific centrosome function	GCP6	61337	Microtubule nucleation	(Raynaud-Messina and Merdes, 2007)
	Gamma tubulin	56069	Microtubule nucleation	(Raynaud-Messina and Merdes, 2007)
	GCP3	434	Microtubule nucleation	(Raynaud-Messina and Merdes, 2007)
	SSA11/TTL 13	80835	Tubulin tyrosine ligase	(Hammond et al., 2008)
	Katanin P60	63871	MT Severing	(Hartman et al., 1998)
	Alpha Tubulin	71268	Microtubules	(Dutcher, 2001)
	Alpha Tubulin	51830	Microtubules	(Dutcher, 2001)
MTOC localization	POC12	29577	Unknown	(Keller et al., 2005)
	PIBF	73664	Unknown	(Lachmann et al., 2004)
	TTL1	33283	Tubulin tyrosine ligase	(Wloga et al., 2008)

Basal body-specific function	MKS6/CC2D2A	77673	Cilium/basal body function	(Gorden et al., 2008)
	MKS1	52666	BB migration/docking to cell membrane	(Dawe et al., 2007)
Flagellar function	MKS3/meckelin	62841	BB migration/docking to cell membrane	(Dawe et al., 2007)
	MBO2	62959	Maintain direction of motility	(Tam and Lefebvre, 2002)
	ODA6	60431	Dynein intermediate chain 2, axonemal	(Pazour et al., 2005)
	IDA7	57343	Dynein intermediate chain, axonemal	(Pazour et al., 2005)
	ODA6	79232	Dynein intermediate chain, axonemal	(Pazour et al., 2005)
	DLC1	74922	Dynein light chain, axonemal	(Pazour et al., 2005)
	IDA4/P28	82719	Dynein light inermediate chain, axonemal	(Pazour et al., 2005)
	IDA4/P28	63304	Dynein light inermediate chain, axonemal	(Pazour et al., 2005)
	tctex-1	29177	Dynein light chain, axonemal	(Harrison et al., 1998; Kagami et al., 1998)
	DLC1	54720	Dynein light chain, axonemal	(Pazour et al., 2005)
	MOT24	32555	Dynein light chain, axonemal	(Merchant et al., 2007)
	MOT45/kintoun/PF13	80717	Axonemal dynein complex assembly	(Omran et al., 2008)
Putative flagellar function	FAP184	2066	Unknown	(Pazour et al., 2005)
	FAP215	66643	Nucleotidase	(Pazour et al., 2005)
	FAP127	44967	Unknown	(Pazour et al., 2005)
	FAP57	61313	Unknown	(Pazour et al., 2005)
Uncharacterized	?	30379	Unknown	
	MOT52	52938	Unknown	(Merchant et al., 2007)
	POC16	62107	Unknown	(Keller et al., 2005)
	POC11	70454	Unknown	(Keller et al., 2005)
	?	49668	Unknown	
	?	68814	Unknown	
	MOT37	61232	Unknown	(Merchant et al., 2007)
	LRR6	31069	Unknown	
	TECT3	65759	Unknown	
	MOT39	72718	Unknown	(Merchant et al., 2007)
Other	?	73058	Dual-speciticiy phosphatase	
	?	29177	Nucleotide kinase	
	?	81169	High mobility group protein	
	?	82958	High mobility group protein	
	?	54684	Elastase	
	MOT50	71996	Elastase	(Merchant et al., 2007)

Table 4.2: Flagellar enriched gene cluster

Genes contained in the flagella-enriched gene cluster are classified by predicted function (left). JGI protein IDs for each gene are listed in the first column. Gene information for each protein ID can be found at: www.jgi.doe.gov/naegleria. Genes with hits in the proteomic analysis of *Naegleria* flagella are indicated in the second column with a "Yes". Predicted orthologs of each gene are listed on the far right. Predictions are as described in Chapter 2.

	JGI Peptide ID	Proteomics Hit	Predicted homolog
IFT (Intraflagellar transport)	79669		FLA3/Kinesin-associated protein 3
	63939		FLA2/FLA8 (kinesin 2 homolog)
	62977		IFT20
	77715		IFT52
	45002	Yes	IFT57
	77945		IFT80
	63280	Yes	IFT88
	71180		IFT122
	48798		IFT140
BBS (Transport to flagellar base)	65179		BBS1
	71257		BBS2
	44202		BBS3
	28891		BBS4
	34252		BBS5
	68114		BBS7
	80979		BBS8
	80972		BBS9
Structural components	39221	Yes	Alpha-tubulin
	56065		Alpha-tubulin
	56236		Alpha-tubulin
	83350		Beta-tubulin
	55423	Yes	Beta-tubulin
	56391		Beta-tubulin
	78637		ODA9 (outer arm dynein intermediate chain)
	80259	Yes	D1bLIC (dynein light intermediate chain)
	81548	Yes	ODA1 (p66 outer dynein arm docking complex protein)
	30532		ODA12
	81047	Yes	RIB72
	83064	Yes	RIB43A-domain containing protein
	49798	Yes	Radial spoke protein 4
	44954	Yes	Radial spoke-head-like
	30562	Yes	PF16
	952		PF20/SPG16
	82851	Yes	PACRG
	55564		Flagellar calmodulin (CAM1)
	64648		KLP1 (kinesin-9)
General MT	72175	Yes	Katanan P80

associated	83220		katanin p60
	62120		CLASP-like
	71898		DIP13
	64818	Yes	XRP2
	4601	Yes	FAP9
	50227		FAP14
	62998		FAP22
	81229		FAP32
	67854	Yes	FAP45
	68117		FAP50
	59637	Yes	FAP52
	380	Yes	FAP59
	50399		FAP60
	73596		FAP61
	29690		FAP66
FAPP	33146		FAP67
	62358		FAP69
	70274		FAP70
	79290		FAP82
	46605	Yes	FAP91
	79626	Yes	FAP116
	32701		FAP118.
	74561		FAP134
	65873		FAP161
	61993		FAP259
	54982		FAP253
	29188	Yes	FAP251
	32341		Kinase
	66608		Guanylate cyclase
	63921		MOT17
	56610		MOT25
	73917		MOT51
	63091	Yes	SSA4
Unknown	67046		SSA20 (chymotrypsin-like domain)
	74042		Leucine-rich repeats
	56805		Unknown
	60926		RABL2A
	79456	Yes	ARL3
	68064		ARL13
	80690		DYF13
	56340		POC7/UNC119/HRG4.
Other	29950		Nucleoside diphosphate kinase (flagellate-specific domain structure)
	69688	Yes	DPY30
	78184		FKB12
	78645	Yes	PDE14

Table 4.3: Orthologs of POC11 in various eukaryotes

Species names, NCBI accessions, and proteins lengths (in amino acids) are as indicated.

Common name	**Species**	**Accession**	**Length**
Human	*Homo sapiens*	Q9BR77	488
Mouse	*Mus musculus*	NP_080304	489
Xenopus	*Xenopus laevis*	NP_001086567	496
Zebra fish	*Danio rerio*	NP_001020649.1	427
Sea anemone	*Nematostella vectensis*	EDO34896	453
Trichoplax	*Trichoplax adherins*	XP_002108951	450
Choanoflagellate	*Monosiga brevicola*	XP_001743324	528
Naegleria	*Naegleria gruberi*	EFC41717	387
Chlamydomonas	*Chlamydomonas reinhardtii*	XP_001693122	451
Paramecium_1	*Paramecium* tetraurelia	CAK72449	502
Paramecium_2	*Paramecium* tetraurelia	XP_001438763	510
Tetrahymena	*Tetrahymena thermophila*	XP_001023528	511

Figure 4.1: *Naegleria* differentiation

Each curve represents the percentage of cells with visible flagella during the differentiation time course of one replicate used in the microarray analysis. The time points collected for RNA extraction are indicated with an asterisk. Important events during *Naegleria* basal body assembly are indicated (based on results described in Chapter 3).

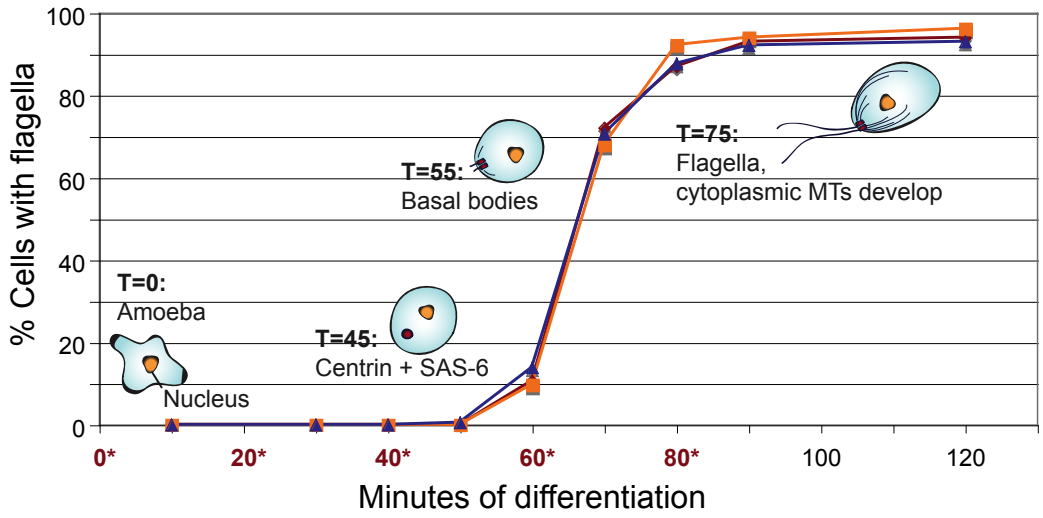


Figure 4.2: RNA and Protein samples used for large-scale analyses

A.) Electrophoretic analysis of representative sample set of differentiation timepoints of total RNA after purification, with *Naegleria* rRNA indicated to the left. Samples were run on a non-denaturing agarose gel and visualized with an RNA dye.

B.) SDS-PAGE stained with Ruby Red protein stain, with sizes indicated. Each marker represents approximately 0.1 microgram of protein, 1/20 of sample loaded.

C.) Log intensities of normalized gene expression levels compared between biological replicates. Axis are in arbitrary units of expression level. Darker regions indicate more genes with that level of expression in the compared experiments.

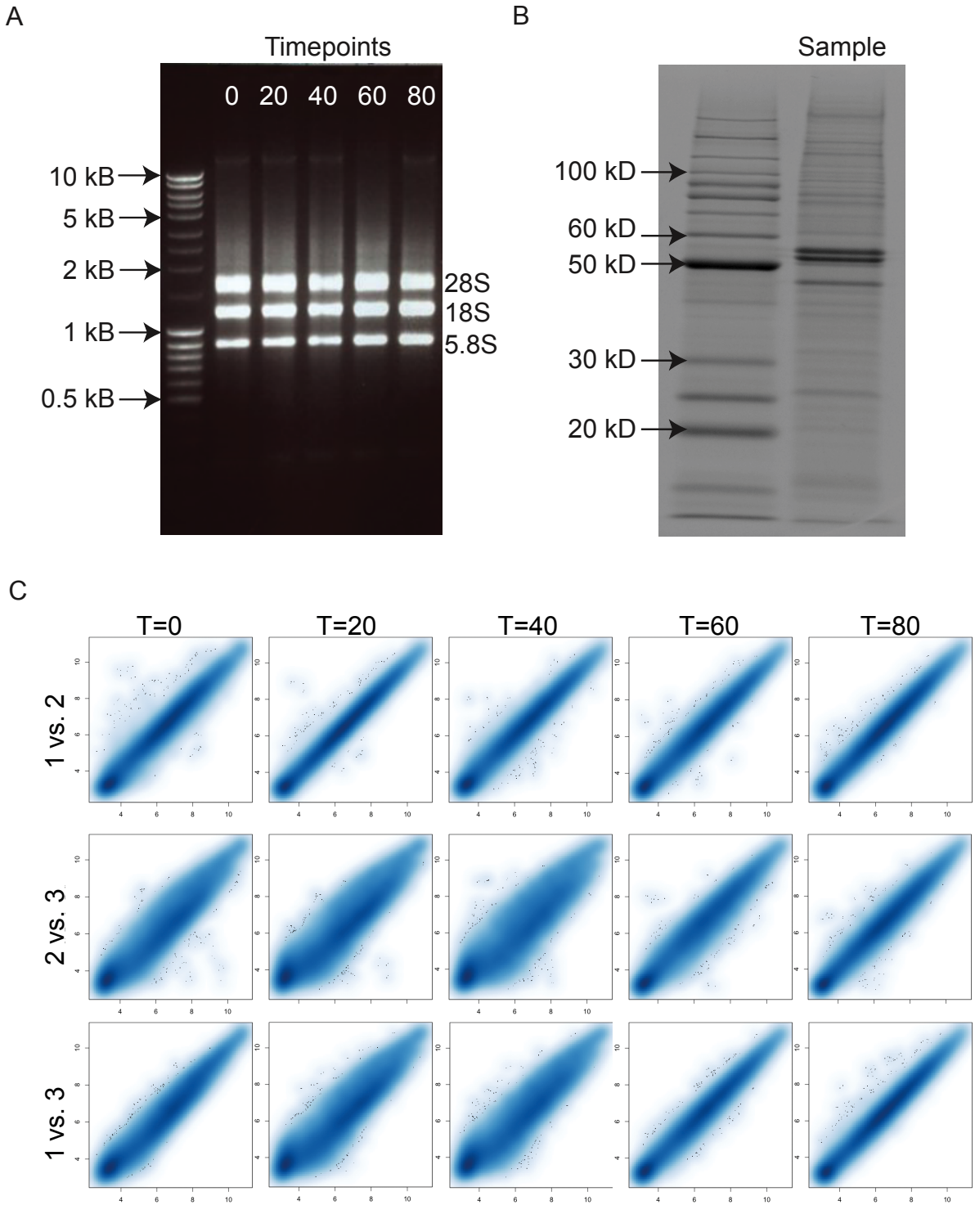

A
Timepoints
0 20 40 60 80
10 kB
5 kB
2 kB
1 kB
0.5 kB
28S
18S
5.8S
B
Sample
100 kD
60 kD
50 kD
30 kD
20 kD
C
T=0
T=20
T=40
T=60
T=80
1 vs. 2
2 vs. 3
1 vs. 3

Figure 4.3: Basal body and flagella enriched gene clusters

Each row represents a gene and each column represent the mean expression at the indicated time point, with blue indicating low gene expression, and yellow indicating high gene expression. The gene clustering cladogram is shown to the left; genes grouped close to each other have high similarity in their expression values across the five time points. On the right are red and green boxes that indicate membership in the AM (amoeboid-motility specific) and FM (flagellar-motility specific) gene sets, respectively. Clusters of genes enriched in basal body and flagellar genes are indicated.

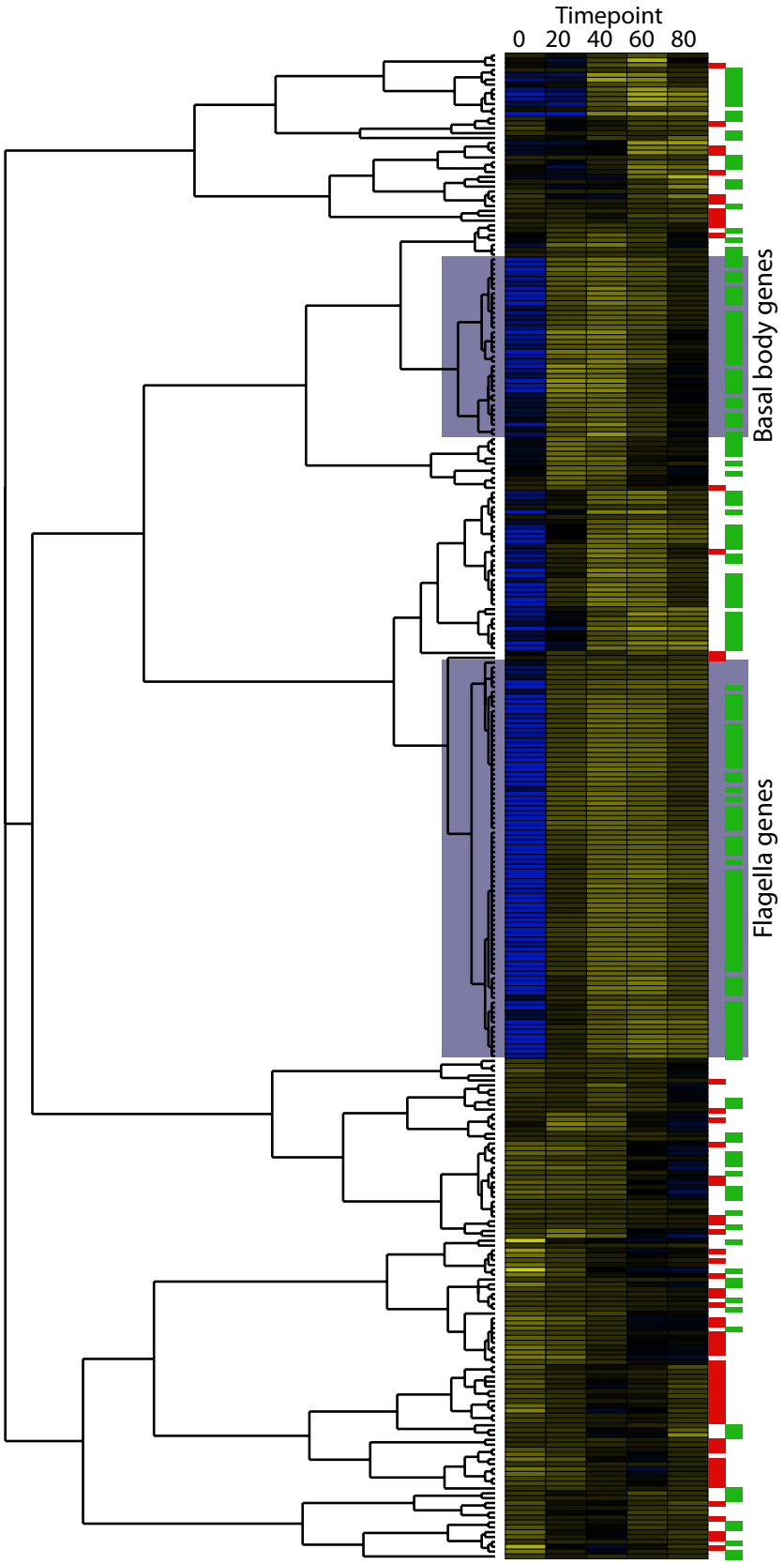
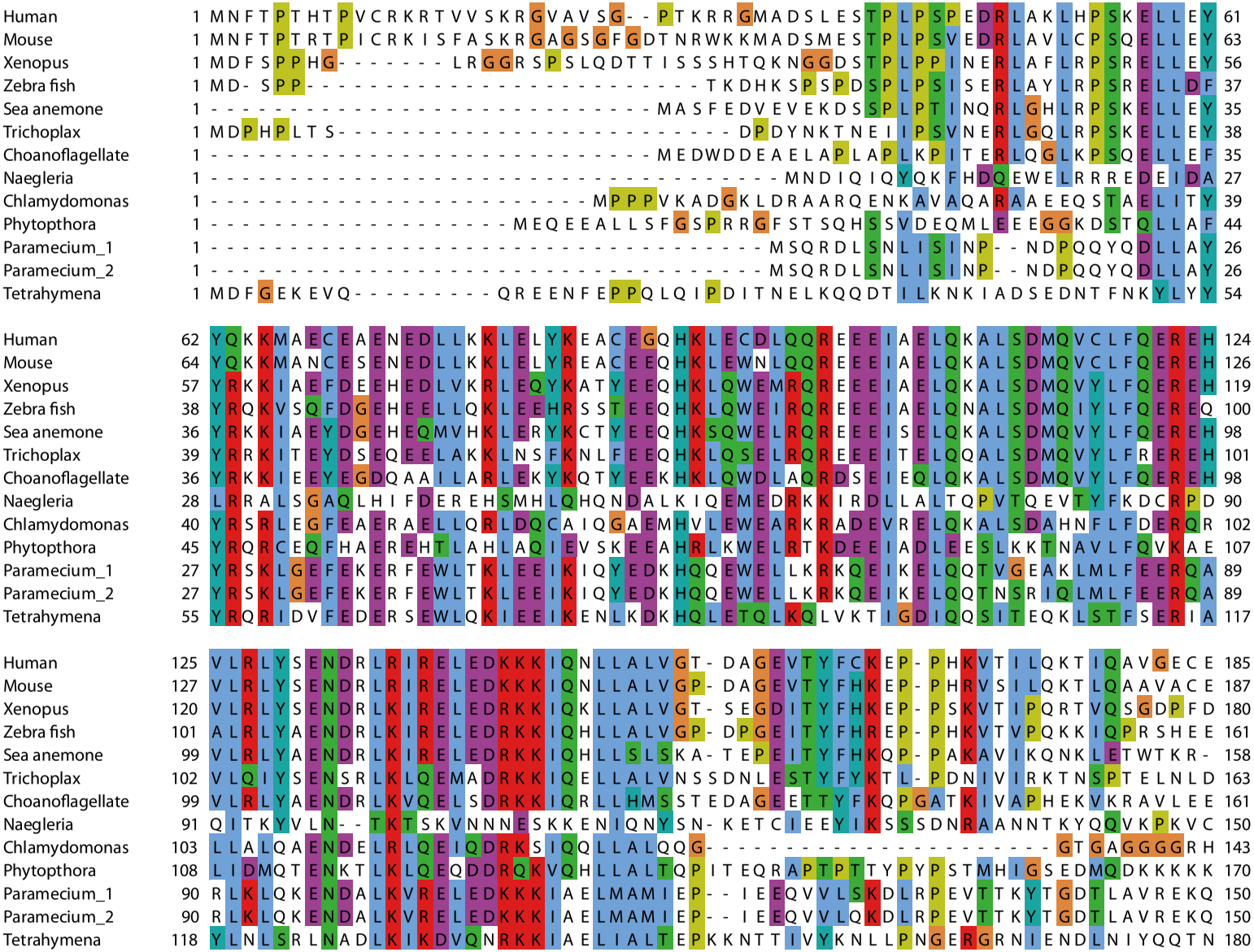
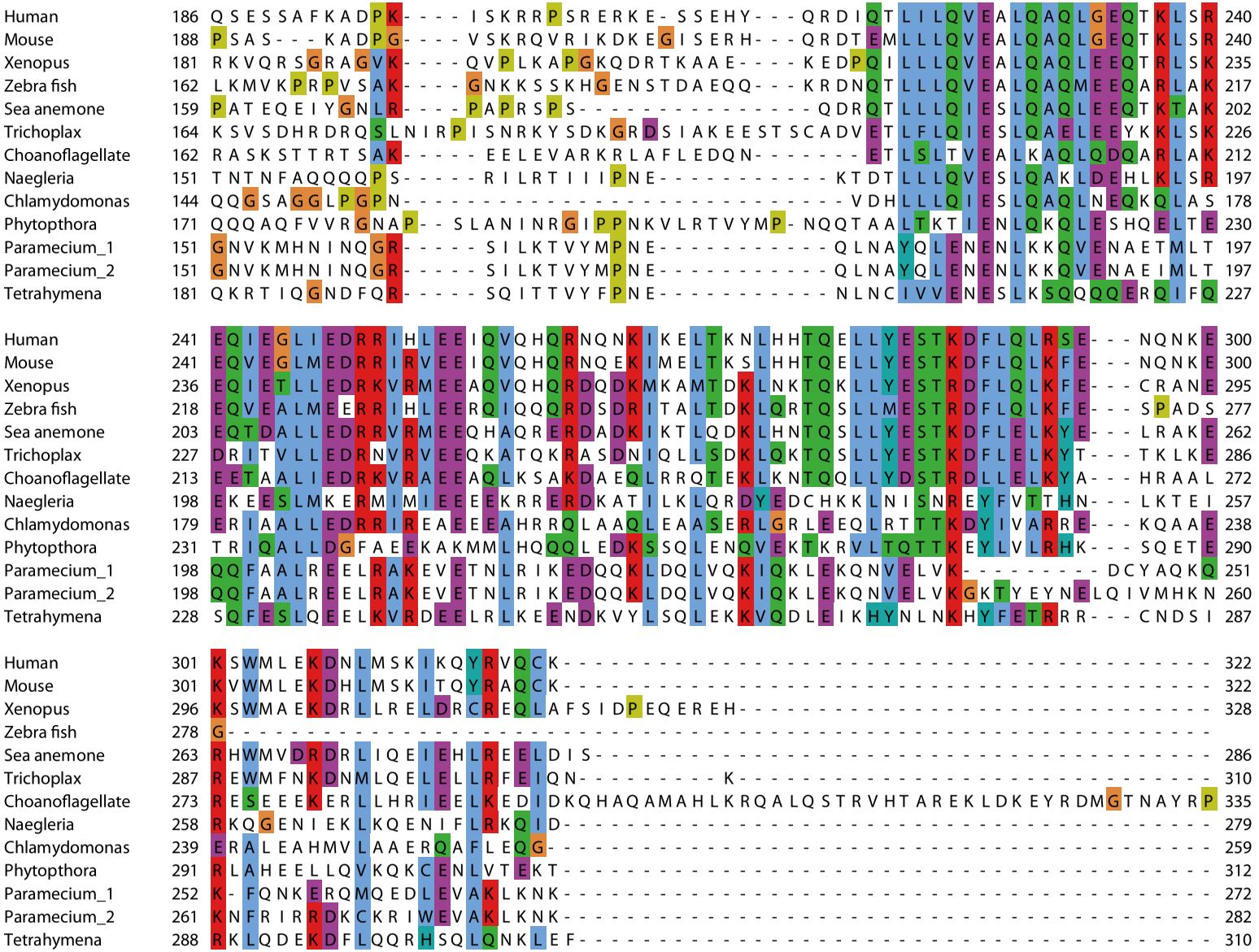
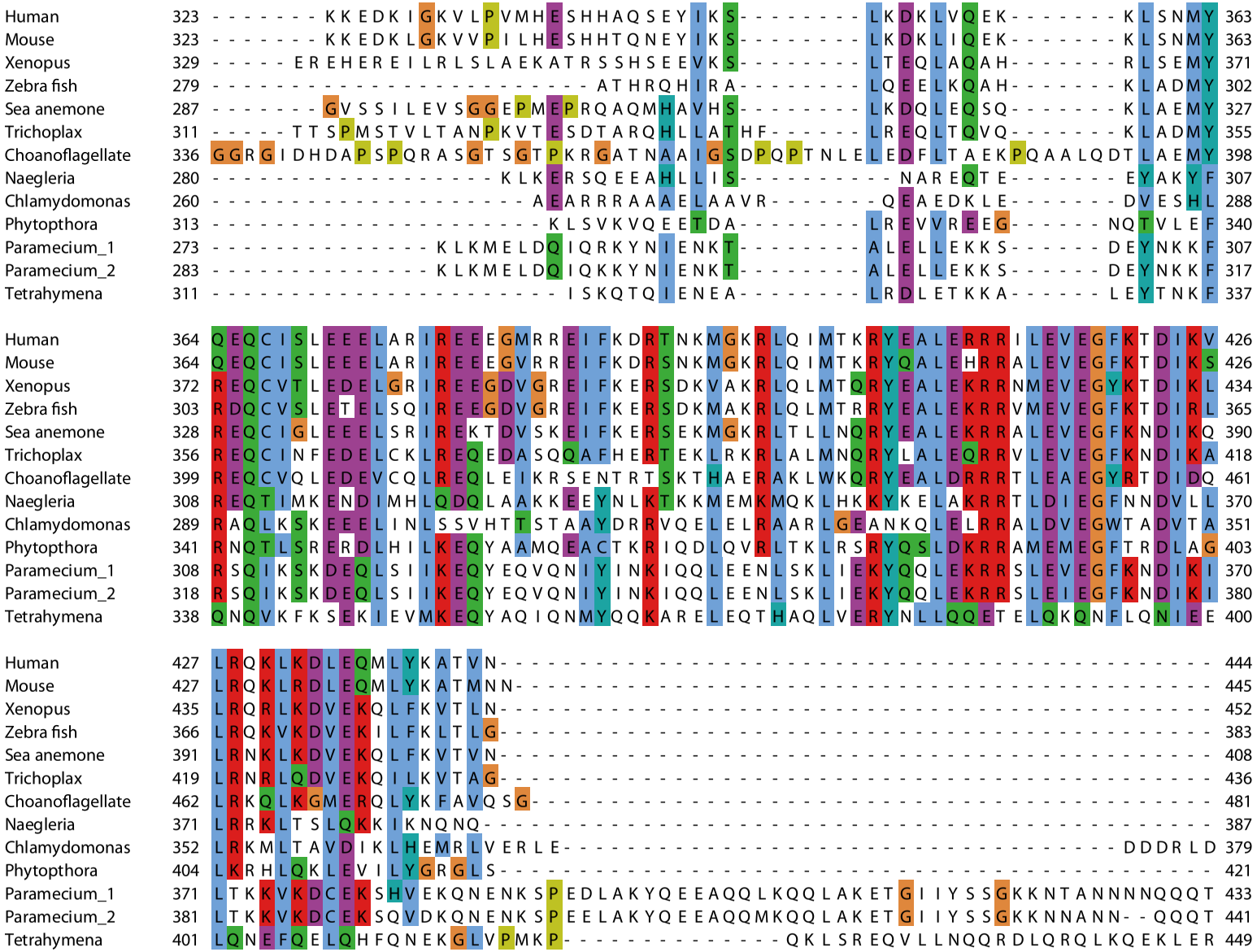

Timepoint
0 20 40 60 80
Basal body genes
Flagella genes

Figure 4.4: POC11 is conserved through eukaryotic evolution

Multiple sequence alignment (spaced over 4 pages) showing POC11 homologs in a diversity of eukaryotes. Amino acids are colored according to a combination of conservation and amino acid type.


Human 1 MNFTPTHTPVCRKRTVVSKRGVAVSG--PTKRRGMADSLESTPLPSPEDRLAKLHPSKELLEY 61
Mouse 1 MNFTPTRTPICRKISFASKRGAGSGFGDTNRWKKMADSMESTPLPSVEDRLAVLCPSQELLEY 63
Xenopus 1 MDFSPPHG-------LRGGRSPSLQDTTISSSHTQKNGGDSTPLPPINERLAFLRPSRELLEY 56
Zebra fish 1 MD-SPP----------TKDHKSPSPDSPLPSISERLAYLRPSRELLDF 37
Sea anemone 1 ----------MASFEDVEVEKDSSPLPTINQRLGHLRPSKELLEY 35
Trichoplax 1 MDPHPLTS----------DPDYNKTNEIIPSVNERLGQLRPSKELLEY 38
Choanoflagellate 1 ----------MEDWDDEAELAPLAPLKPITERLQGLKPSQELLEF 35
Naegleria 1 ----------MNDIQIQYQKFHDQEWELRRREDEIDA 27
Chlamydomonas 1 ----------MPPPVKADGKLDRAARQENKAVAQARAAEEQSTAELITY 39
Phytopthora 1 ----------MEQEEALLSFGSPRRGFSTSQHSSVDEQMLEEEGGKDSTQLLAF 44
Paramecium_1 1 ----------MSQRDLSNLISINP--NDPQQYQDLLAY 26
Paramecium_2 1 ----------MSQRDLSNLISINP--NDPQQYQDLLAY 26
Tetrahymena 1 MDFGEKEVQ----------QREENFEPPQLQIPDITNELKQQDTILKNKIADSEDNTFNKYLYY 54
Human 62 YQKKMAECEAENEDLLKKLELYKEACEGQHKLECDLQQREEEIAELQKALSDMQVCLFQERE H 124
Mouse 64 YQKKMANCESENEDLLKKLELYREACEEQHKLEWNLQQREEEIAELQKALSDMQVCLFQERE H 126
Xenopus 57 YRKKIAEFDEEHEDLVKRLEQYKATYEEQHKLQWEMRQREEEIAELQKALSDMQVYLFQERE H 119
Zebra fish 38 YRQKVSQFDGEHEELLQKLEEHRSSTEEQHKLQWEIRQREEEIAELQNALSDMQIYLFQERE Q 100
Sea anemone 36 YRKKIAEYDGEHEQMVHKLERYKCTYEEQHKSQWELRQREEEISELQKALSDMQVYLFQERE H 98
Trichoplax 39 YRRKITEYDSEQEELAKKLNSFKNLFEEQHKLQSELRQREEEITELQQALSDMQVYLFRERE H 101
Choanoflagellate 36 YRKKIEEYEGDQAAILARLEKYKQTYEEKHKLQWDLAQRDSEIEQLQKALSDMQVYLFQERE H 98
Naegleria 28 LRRALSGAQLHIFDEREHSMHLQHQNDALKIQEMEDRKKIRDLLALTQPVTQEVTYFKDCRPD 90
Chlamydomonas 40 YRSRLEGFEAERAELLQRLDQCAIQGAEMHVLEWEARKRADEVRELQKALSDAHNFLFDERQR 102
Phytopthora 45 YRQRCEQFHAEREHTLAHLAQIEVSKEEAHRLKWELRTKDEEIADLEESLKKTNAVLFQVKAE 107
Paramecium_1 27 YRSKLGEFEKERFEWLTKLEEIKIQYEDKHQQEWELLKRKQEIKELQQTVGEAKLMLFEERQA 89
Paramecium_2 27 YRSKLGEFEKERFEWLTKLEEIKIQYEDKHQQEWELLKRKQEIKELQQTNSRIQLMLFEERQA 89
Tetrahymena 55 YRQRIDVFEDERSEWLQKIEEIKENLKDKHQLETQLKQLVKTIGDIQQSITEQKLSTFSERIA 117
Human 125 VLRLYSENDRLRIRELEDKKKIQNLLALVGT-DAGEVTYFCKEP-PHKVTILQKTIQAVGECE 185
Mouse 127 VLRLYSENDRLRIRELEDKKKIQNLLALVGP-DAGEVTYFHKEP-PHRVSILQKTLQAAVACE 187
Xenopus 120 VLRLYSENDRLKIRELEDRKKIQKLLALVGT-SEGDITYFHKEP-PSKVTIPQRTVQSGDPFD 180
Zebra fish 101 ALRLYAENDRLKIRELEDRKKIQHLLALVGP-DPGEITYFHREP-PHKVTVPQKKIQPRSHEE 161
Sea anemone 99 VLRLYAENDRLKIRELEDRKKIQHLLSLSKA-TEPEITYFHKQP-PAKAVIKQNKLETWTKR- 158
Trichoplax 102 VLQIYSENSRLKLQEMADRKKIQELLALVNSSDNLESTYFYKTL-PDNIVIRKTNSPTELNLD 163
Choanoflagellate 99 VLRLYAENDRLKVQELSDRKKIQRLLHMSSTEDAGEETTYFKQPGATKIVAPHEKVKRAVLEE 161
Naegleria 91 QITKYVLN--TKTSKVNNNESKKENIQNYSN-KETCIEEYIKSSSDNRAANNTKYQQVKPKVC 150
Chlamydomonas 103 LLALQAENDELRLQEIQDRKSIQQLLALQQG----------GTGAGGGGRH 143
Phytopthora 108 LIDMQTENKTLKLQEQDDRQKVQHLLALTQPITEQRAPTPTTYPYPSTMHIGSEDMQDKKKKK 170
Paramecium_1 90 RLKLQKENDALKVRELEDKKKIAELMAMIEP--IEEQVVLSKDLRPEVTTKYTGDTLAVREKQ 150
Paramecium_2 90 RLKLQKENDALKVRELEDKKKIAELMAMIEP--IEEQVVLQKDLRPEVTTKYTGDTLAVREKQ 150
Tetrahymena 118 YLNLSRLNADLKIKDVQNRKKIAELIALTEPKKNTTIVYKNLLPNGERGRNIENDLNIYQQTN 180


Human 186 QSESSAFKADPK----ISKRRPSRERKE-SSEHY---QRDIQTLILQVEALQAQLGEQTKLSR 240
Mouse 188 PSAS---KADPG----VSKRQVRIKDKEGISERH---QRDTEMLLLQVEALQAQLGEQTKLSR 240
Xenopus 181 RKVQRSGRAGVK----QVPLKAPGKQDRTKAAE----KEDPQILLLQVEALQAQLEEQTRLSK 235
Zebra fish 162 LKMVKPRPVSAK----GNKKSSKHGENSTDAEQQ---KRDNQTLLLQVEALQAQMEEQARLAK 217
Sea anemone 159 PATEQEIYGNLR----PAPRSPS----------QDRQTLLLQIESLQAQLEEQTKTAK 202
Trichoplax 164 KSVSDHRDRQSLNIRPISNRKYSDKGRDSIAKEESTSCADVETLFLQIESLQAELEEYKKLSK 226
Choanoflagellate 162 RASKSTTRTSAK-----EELEVARKRLAFLEDQN-------ETLSLTVEALKAQLQDQARLAK 212
Naegleria 151 TNTNFAQQQQPS-----RILRTIIIPNE----------KTDTLLLQVESLQAKLDEHLKLSR 197
Chlamydomonas 144 QQGSAGGLPGPN----------VDHLLLQIESLQAQLNEQKQLAS 178
Phytopthora 171 QQQAQFVVRGNAP--SLANINRGIPPNKVLRTVYMP-NQQTAALTKTIENLQKQLESHQELTE 230
Paramecium_1 151 GNVKMHNINQGR-----SILKTVYMPNE----------QLNAYQLENENLKKQVENAETMLT 197
Paramecium_2 151 GNVKMHNINQGR-----SILKTVYMPNE----------QLNAYQLENENLKKQVENAEIMLT 197
Tetrahymena 181 QKRTIQGNDFQR-----SQITTVYFPNE----------NLNCIVVENESLKSQQQERQIFQ 227
Human 241 EQIEGLIEDRRIHLEEIQVQHQRNQNKIKELTKNLHHTQELLYESTKDFLQLRSE---NQNKE 300
Mouse 241 EQVEGLMEDRRIRVEEIQVQHQRNQEKIMELTKSLHHTQELLYESTKDFLQLKFE---NQNKE 300
Xenopus 236 EQIETLLEDRKVRMEEAQVQHQRDQDKMKAMTDKLNKTQKLLYESTRDFLQLKFE---CRANE 295
Zebra fish 218 EQVEALMEERRIHLEERQIQQQRDSDRITALTDKLQRTQSLLMESTRDFLQLKFE---SPADS 277
Sea anemone 203 EQTDALLEDRRVRMEEQHAQRERDADKIKTLQDKLHNTQSLLYESTKDFLELKYE---LRAKE 262
Trichoplax 227 DRITVLLEDRNVRVEEQKATQKRASDNIQLLSDKLQKTQSLLYESTKDFLELKYT---TKLKE 286
Choanoflagellate 213 EETAALIEDRKVRAEEAQLKSAKDAEQLRRQTEKLKNTQQLLYDSTRDLLELKYA---HRAAL 272
Naegleria 198 EKEESLMKERMIMIEEEEKRRERDKATILKLQRDYEDCHKKLNISNREYFVTTHN---LKTEI 257
Chlamydomonas 179 ERIAALLEDRRIREAEEEEAHRRQLAAQLEAASERLGRLEEQLRTTTKDYIVARRE---KQAAE 238
Phytopthora 231 TRIQALLDGFAEEKAKMMLHQQQLEDKSSQLENQVEKTKRVLTQTTKEYLVLRHK---SQETE 290
Paramecium_1 198 QQFAALREELRAKEVETNLRIKEDQQKLDQLVQKIQKLEKQNVELVK----------DCYAQKQ 251
Paramecium_2 198 QQFAALREELRAKEVETNLRIKEDQQKLDQLVQKIQKLEKQNVELVKGKTYEYNELQIVMHKN 260
Tetrahymena 228 SQFESLQEELKVRDEELRLKEENDKVYLSQLEKKVQDLEIKHYNLNKHYFETRRR---CNDSI 287
Human 301 KSWMLEKDNLMSKIKQYRVQCK---------- 322
Mouse 301 KVWMLEKDHLMSKITQYRAQCK---------- 322
Xenopus 296 KSWMAEKDRLLRELDRCREQLAFSIDPEQERE EH---------- 328
Zebra fish 278 G---------- 278
Sea anemone 263 RHWMVDRDRLIQEIEHLREELDIS---------- 286
Trichoplax 287 REWMFNKDNMLQELELLRFEIQN---------K---------- 310
Choanoflagellate 273 RESEEEKERLLHRIEELKEDIDKQHAQAMAHLKRQALQSTRVHTAREKLDKEYRDMGTNAYRP 335
Naegleria 258 RKQGENIEKLKQENIFLRKQID---------- 279
Chlamydomonas 239 ERALEAHMVLAAERQAFLEQG---------- 259
Phytopthora 291 RLAHEELLQVKQKCENLVTEKT---------- 312
Paramecium_1 252 K-FQNKERQMQEDLEVAKLKNK---------- 272
Paramecium_2 261 KNFRIRRDKCKRIWEVAKLKNK---------- 282
Tetrahymena 288 RKLQDEKDFLQQRHSQLQNKLEF---------- 310


Human 323 KKEDKIGKVLPVMHESHHAQSEYIKS LKDKLVQEK KLSNMY 363
Mouse 323 KKEDKLGKVVPILHESHHTQNEYIKS LKDKLIQEK KLSNMY 363
Xenopus 329 EREHEREILRLSLAEKATRSSHSEEVKS LTEQLAQAH RLSEMY 371
Zebra fish 279 ATHRQHIRA LQEELKQAH KLADMY 302
Sea anemone 287 GVSSILEVSGGEPMEPRQAQMHAVHS LKDQLEQSQ KLAEMY 327
Trichoplax 311 TTSPMSTVLTANPKVTESDTARQHLLATHF LREQLTQVQ KLADMY 355
Choanoflagellate 336 GGRGIDHDAPSPQRASGTSGTPKRGATNAAIGSDPQPTNLELEDFLTAEKPQAALQDT LAEMY 398
Naegleria 280 KLKERSQEEAHLLIS NAREQTE EYAKYF 307
Chlamydomonas 260 AEARRRAAAELAAVR QEAEDKLE DVESHL 288
Phytopthora 313 KLSVKVQEETDA LREVVREEG NQTVLEF 340
Paramecium_1 273 KLKMELDQIQRKYNIENKT ALELLEKKS DEYNKKF 307
Paramecium_2 283 KLKMELDQIQKKYNIENKT ALELLEKKS DEYNKKF 317
Tetrahymena 311 ISKQTQIENEA LRDLETKKA LEYTNKF 337
Human 364 QEQCISLEEELARIREEEGMRREIFKDRTNKMGKRLQIMTKRYEALERRRILEVEGFKTDIKV 426
Mouse 364 QEQCISLEEELARIREEEGVRREIFKDRSNKMGKRLQIMTKRYQALEHRRALEVEGFKTDIKS 426
Xenopus 372 REQCVTLEDELGRIREEGDVGREIFKERSDKVAKRLQLMTQRYEALEKRRNMEVEGYKTDIKL 434
Zebra fish 303 RDQCVSLETELSQIREEGDVGREIFKERSDKMAKRLQLMTRRYEALEKRRVMEVEGFKTDIRL 365
Sea anemone 328 REQCIGLEEELSRIREKTDVSKEIFKERSEKMGKRLTLLNQRYEALEKRRALEVEGFKNDIKQ 390
Trichoplax 356 REQCINFEDELCKLREQEDASQQAFHERTEKLRKRLALMNQRYLALEQRRVLEVEGFKNDIKA 418
Choanoflagellate 399 REQCVQLEDEVCQLREQLEIKRSENTRTSKTHAERAKLWKQRYEALDRRRTLEAEGYRTDIDQ 461
Naegleria 308 REQTIMKENDIMHLQDQLAAKKEEYNLKTKKMEMKMQKLHKKYKELAKRRTLDIEGFNNDVLL 370
Chlamydomonas 289 RAQLKSKEEELINLSSVHTTSTAAYDRRVQELELRAARLGEANKQLELRRALDVEGWTADVTA 351
Phytopthora 341 RNQTLSRERDLHILKEQYAAMQEACTKRIQDLQVRLTKLRSRYQSLDKRRAMEMEGFTRDLAG 403
Paramecium_1 308 RSQIKSKDEQLSIIKEQYEQVQNIYINKIQQLEENLSKLIEKYQQLEKRRSLEVEGFKNDIKI 370
Paramecium_2 318 RSQIKSKDEQLSIIKEQYEQVQNIYINKIQQLEENLSKLIEKYQQLEKRRSLEIEGFKNDIKI 380
Tetrahymena 338 QNQVKFKSEKIEVMKEQYAQIQNMYQQKARELEQTHAQLVERYNLLQQETELQKQNFLQNIEE 400
Human 427 LRQKLKDLEQMLYKATVN 444
Mouse 427 LRQKLRDLEQMLYKATMNN 445
Xenopus 435 LRQRLKDVEKQLFKVTLN 452
Zebra fish 366 LRQKVKDVEKILFKLTLG 383
Sea anemone 391 LRNKLKDVEKQLFKVTVN 408
Trichoplax 419 LRNRLQDVEKQILKVTAG 436
Choanoflagellate 462 LRKQLKGMERQLYKFAVQSG 481
Naegleria 371 LRRKLTSLQKKIKNQNQ 387
Chlamydomonas 352 LRKMLTAVDIKLHEMRLVERLE DDDRLD 379
Phytopthora 404 LKRHLQKLEVILYGRGLS 421
Paramecium_1 371 LTKKVKDCEKSHVEKQNENKSPEDLAKYQEEAQQLKQQLAKETGIIYSSGKKNTANNNNQQQT 433
Paramecium_2 381 LTKKVKDCEKSQVDKQNENKSPEELAKYQEEAQQMKQQLAKETGIIYSSGKKNNANN--QQQT 441
Tetrahymena 401 LQNEFQELQHFQNEKGLVPMKP QKLSREQVLLNQQRDLQRQLKQEKLER 449

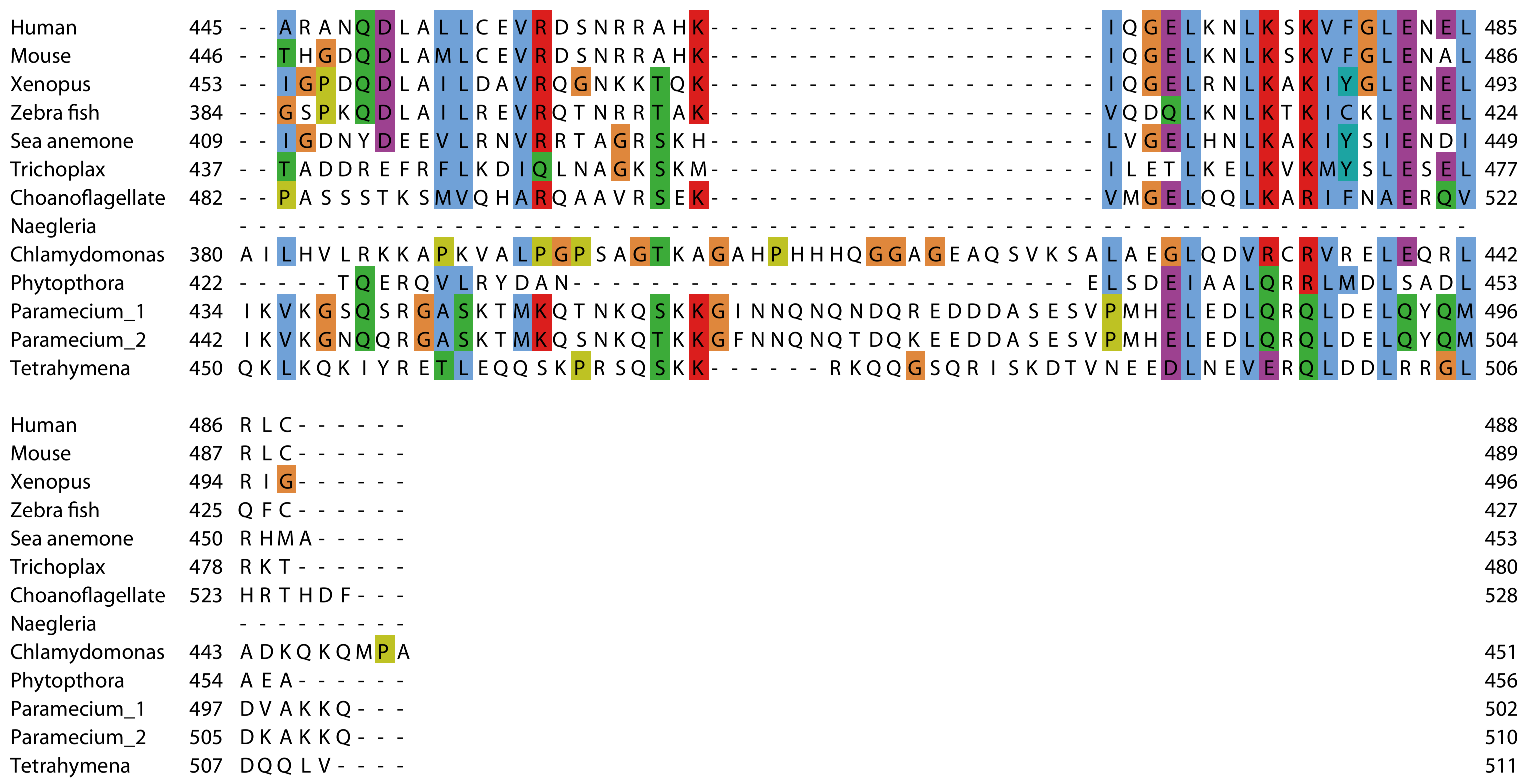
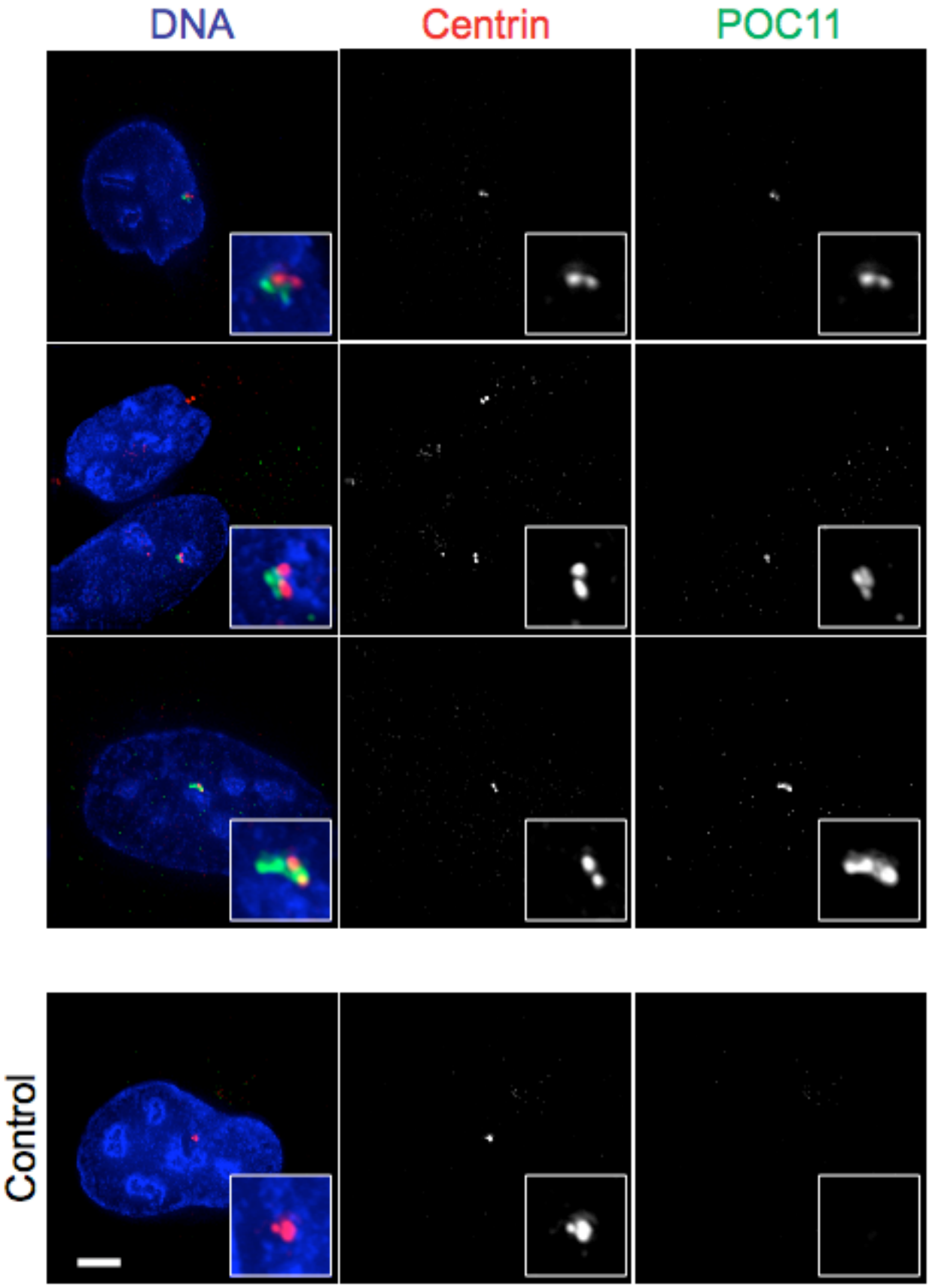

Human 445 --ARANQDLALLCEVRDSNRRAHK--------------------IQGELKNLKSKVFGLENEL 485
Mouse 446 --THGDQDLAMLCEVRDSNRRAHK--------------------IQGELKNLKSKVFGLENAL 486
Xenopus 453 --IGPDQDLAILDAVRQGNKKTQK--------------------IQGELRNLKAKIYGLENEL 493
Zebra fish 384 --GSPKQDLAILREVRQTNRRTAK--------------------VQDQLKNLKTKICKLENEL 424
Sea anemone 409 --IGDNYDEEVLRNVRRTAGRSKH--------------------LVGELHNLKAKIYSIENDI 449
Trichoplax 437 --TADDREFRFLKDIQLNAGKSKM--------------------ILETLKELKVKMYSLESEL 477
Choanoflagellate 482 --PASSSTKSMVQHARQAAVRSEK--------------------VMGELQQLKARIFNAERQV 522
Naegleria ---
Chlamydomonas 380 AILHVLRKKAPKVALPGPSAGTKAGAHPHHHQGGAGEAQSVKSALAEGLQDVRCRVRELEQRL 442
Phytopthora 422 -----TQERQVLRYDAN--------------------------ELSDEIAALQRRLMDLSADL 453
Paramecium_1 434 IKVKGSQSRGASKTMKQTNKQSKKGINNQNQNDQREDDDASESVPMHELEDLQRQLDELQYQM 496
Paramecium_2 442 IKVKGNQQRGASKTMKQSNKQTKKGFNNQNQTDQKEEDDASESVPMHELEDLQRQLDELQYQM 504
Tetrahymena 450 QKLKQKIYRETLEQQSKPRSQSKK------RKQQGSQRISKDTVNEEDLNEVERQLDDLRRGL 506
Human 486 RLC------ 488
Mouse 487 RLC------ 489
Xenopus 494 RIG------ 496
Zebra fish 425 QFC------ 427
Sea anemone 450 RHMA----- 453
Trichoplax 478 RKT------ 480
Choanoflagellate 523 HRTHDF--- 528
Naegleria ---------
Chlamydomonas 443 ADKQKQMPA 451
Phytopthora 454 AEA------ 456
Paramecium_1 497 DVAKKQ--- 502
Paramecium_2 505 DKAKKQ--- 510
Tetrahymena 507 DQQLV---- 511

Figure 4.5: Human POC11 localizes to centrosomes

Localization of GFP-tagged protein corresponding to the human ortholog of POC11 (in green) after transient transfections of U2OS cells. Distal ends of centrioles are labeled with an antibody to centrin-2 (red), and DNA is shown in blue. Scale bar represents 5 microns.


DNA
Centrin
POC11
Control