The evolution of eukaryotic cilia

Matthew Edmiston Hodges

The Queen's College

University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2011

Abstract

Eukaryotic cilia are complex, highly conserved microtubule-based organelles with a broad phylogenetic distribution. Cilia were present in the last eukaryotic common ancestor and many proteins involved in cilia function have been conserved through eukaryotic diversification. The evolution of these ciliary functions may be inferred from the distribution of the molecular components from which these organelles are composed. By linking protein distribution in 45 diverse eukaryotes with organismal biology, I define an ancestral ciliary inventory. Analysis of these core proteins allows the inference that the cenancestor of the eukaryotes possessed a cilium for motility and sensory function. I show that the centriolar basal body function is ancestral, whereas the centrosome is specific to the Holozoa, and I use this information to predict a number of roles for proteins based on their phylogenetic profile. I also show that while remarkably conserved, significant divergence in ciliary protein composition has occurred in many lineages, such as the unusual centriole of *Caenorhabditis elegans* and the transitional changes throughout the land plants. I exemplify this divergence through ultrastructural studies of the fern *Ceratopteris richardii* and the liverwort *Marchantia polymorpha* both of which have cilia that exhibit a number of distinctive morphological features, the most conspicuous of which is a general breakdown of canonical microtubule arrangements. Cilia have also been lost multiple times in different lineages: at least twice within the land plants. During these evolutionary transitions proteins with ancestral ciliary functions may be lost or co-opted into different functions. I have interrogated genomic data to identify proteins that I predict had an ancestral ciliary role, but which have been maintained in non-ciliated land plants. I demonstrate that several of these proteins have a flagellar localisation in protozoan trypanosomes and I use expression data correlation to predict potential non-ciliary plant roles.

List of abbreviations

A-tubule – “A” microtubule

B-tubule – “B” microtubule

BB – Basal body

BP – Base pairs

BBS - Bardet-Biedl Syndrome

BLAST – Basic Local Alignment Search Tool

BLASTp – Protein BLAST

BUG – Basal body up-regulated

CAP – Centriolar associated protein

CCP – Conserved in ciliated and plants

cDNA – Complementary DNA

CDS – Coding sequence

DAPI - 4',6-diamidino-2-phenylindole

DHC1B – Dynein heavy chain 1B

DNA – Deoxyribonucleic acid

EB1 – Microtubule plus-end associated protein

E-value – Expected value

EM – Electron microscopy

FBB17 – Flagellar basal body protein 17

FCS – Foetal calf serum

FOP – Fibroblast growth factor receptor 1 oncogene partner

IFT – Intraflagellar transport

KAN – Kanamycin

KAP – Kinesin-II associated protein

kDA – kilo Dalton

LECA – Last common eukaryotic ancestor

MCMCMC – Metropolis-coupled Markov chain Monte Carlo

MKS - Meckel-Gruber Syndrome

MLS – Multi-layered structure

MTOC – Microtubule organising centre

mRNA – Messenger RNA

NPHP – Nephronophthisis

ODA/IDA – Outer/inner arm dynein

ORF – Open reading frame

PBS – Phosphate buffered saline

PCM – Pericentriolar material

PCR – Polymerase chain reaction

PFR – Paraflagellar rod

PKD1/2 - Polycystin 1/2

PLK – Polo-like kinase

Puro – Puromycin antibiotic resistance

RBB – Reciprocal best BLAST

RNA – Ribonucleic acid

SAS4/5/6 – Spindle assembly abnormal

SEM – Scanning electron microscopy

ST – Stellate array

SZY-20 – Suppressor of ZYG-1

TEM – Transmission electron microscopy

TY – BB2 monoclonal epitope

TZ – Transition zone

WAG – Wheland and Goldman Model

WDR16 – WD40-repeat containing protein 16

XMAP215 – Microtubule-associated protein 215kDA

YFP – Yellow fluorescent protein

α/β/δ/ε/γ-tubulin – Alpha/ beta/ delta/ gamma/ epsilon-tubulin

Acknowledgements

Firstly I would like to thank Jane Langdale for being my supervisor for seven years running: you have been an amazing mentor during my undergraduate and DPhil. I really cannot put into words how great it has been learning from you (or how many red pens it has needed). I would also like to thank my joint supervisor Keith Gull for putting up with the 'plants guy' in your lab. You have both been incredibly supportive and I've really enjoyed the wonderful freedom in both labs.

To Bill Wickstead – stop denying it, you were my supervisor. I couldn't have done it without you, and I am looking forward to reading all the papers from your new lab. Mike Shaw – thanks for all the help on the EM, and in the KA - I enjoyed one more than the other. It has also been great working with Nicole Scheumann on numerous seemingly haphazard projects. Thanks to Steve Kelly for answering all my random questions and to Yuki for being so nice. Obviously I want to say a huge thank you to everyone else in both the Langdale and Gull labs for providing an amazing working environment and lifelong friends. MTW and Armando are brothers, alongside old granny Jules, lab-dad Peng, Heather, Denis, Rita, and Robert Scotland for invaluable discussions (on the various merits of his smartphones).

I would also like to thank all my friends in Oxford who have made my time so enjoyable: Tommy, undergrad and PhD - looking forward to the next stage; Michiel and Helen for surviving with me and all the PEs; Jim for sitting behind me, and the sport enthusiasts Whee and Pickles, you have all made Oxford really fun. Keep trying. And of course the Darwin's crew - lunch will never be the same again, but I am glad the mantle of the tea rota has passed onto a worthy set.

A huge thank you must go to my parents and brother Rick for being so supportive and also to my wonderful girlfriend Nina, for pretending I will be a real doctor.

Finally I would like to thank BBSRC, Gatsby, BSCB, Wellcome Trust, Genetics Society and The Queen's College for funding and support during my studies, and everyone else!

Contents

ABSTRACT ..... **II**

LIST OF ABBREVIATIONS ..... **III**

ACKNOWLEDGEMENTS ..... **V**

FIGURE LEGEND ..... **XI**

1 THE EVOLUTION OF LAND PLANT CILIA **1**

1.1 Preface **1**

1.2 Abstract **2**

1.3 Introduction **3**

1.3.1 Diversity of ciliary structure 5

1.4 Plant cilia: a case study for cilia evolution and development **10**

1.4.1 Structural differences in cilia morphology across the land plants 13

1.4.2 Ciliary proteins lost on transition to land 23

1.4.3 Ciliary proteins conserved in non-ciliated plants 24

1.5 Future perspectives: *Ceratopteris* and *Arabidopsis* as ciliary models **26**

1.6 Principles of electron microscopy **27**

1.7 Phylogeny estimation using Maximum Likelihood and Bayesian inference **29**

1.8 Thesis aims **31**

2 RECONSTRUCTING THE EVOLUTIONARY HISTORY OF THE CENTRIOLE FROM PROTEIN COMPONENTS 32

2.1 Preface 32

2.2 Abstract 33

2.3 Introduction 34

2.4 Results and Discussion 35

2.4.1 Identification of centriolar proteins in diverse eukaryotes 35

2.4.2 Phylogenetic distribution of δ- and ε-tubulin 36

2.4.3 Ancestral centriolar core components 38

2.4.4 Centriole-associated kinases: origin and diversification 41

2.4.5 Basal body function is ancestral, but centrosomal components are not 43

2.4.6 A sensory role for the ancestral cilium 46

2.4.7 Correlation between ciliary eye evolution and ciliopathy-associated proteins 48

2.5 Conclusions 49

2.6 Materials and Methods 49

2.6.1 Definition of datasets 49

2.6.2 Identification of homologous sets 50

2.7 Acknowledgements 51

2.8 Additional Information 53

2.8.1 Table S1: Source and version of the genomes used in the analysis 53

2.8.2 Table S2: Information about proteins used in this analysis 55

2.8.3 Table S3: IDs of proteins within a putative homologue cluster 58

3 UNIQUE CILIA MORPHOLOGY IN FERNS 65

3.1 Preface **65**

3.2 Abstract **66**

3.3 Introduction **67**

3.4 Results **69**

3.4.1 Novel anchoring structures are present at the cilium base 69

3.4.2 Fibres extend throughout the basal body to the transition zone 69

3.4.3 Basal bodies have a complex structure. 72

3.4.4 Contiguous stellate arrays are a unique feature of Ceratopteris richardii cilia 72

3.4.5 Staggered central pair nucleation in the axoneme 74

3.4.6 Central pair microtubules twist in relation to the outer doublets 74

3.4.7 Cilia tips are tapered 74

3.5 Discussion **77**

3.6 Materials and Methods **81**

3.6.1 Plant material 81

3.6.2 Electron microscopy 81

3.6.3 Electron tomography 82

3.7 Additional Information **83**

3.7.1 Stellate arrays of Marchantia polymorpha 83

4 CONSERVATION OF CILIARY PROTEINS IN PLANTS WITH NO CILIA 85

4.1 Preface **85**

4.2 Abstract **86**

4.3 Introduction **87**

4.4 Results **88**

4.4.1 Identification of ciliary proteins by phylogenetic distribution....88
4.4.2 Ciliary profile proteins localise to the flagellum of trypanosomes....90
4.4.3 Identification of ciliary proteins in non-ciliated land plants and fungi....92
4.4.4 CCP proteins localise to the trypanosome flagellum....94
4.4.5 Evolutionary history of CCP proteins....95
4.4.6 CCP proteins are highly expressed in male floral organs and mature pollen....95

4.5 Discussion....97

4.5.1 Flowering plants have retained CCP proteins despite loss of ability to produce cilia....97

4.6 Conclusions....99

4.7 Material and Methods....99

4.7.1 Bioinformatic definition of protein sets....99
4.7.2 Phylogenetic inference....100
4.7.3 Endogenous-locus tagging....100
4.7.4 Fluorescence microscopy....101
4.7.5 Western blot analysis....101

4.8 Acknowledgements....102

4.9 Additional information....103

4.9.1 Table 4-S1....103
4.9.2 Table 4-S2....103
4.9.3 Table 4-S3....103
4.9.4 Table 4-S4....103
4.9.5 Table 4-S5....104
4.9.6 Figure 4-S6....105
4.9.7 Primer list....106

5 GENERAL DISCUSSION....107

5.1 Summary 107

5.2 LECA ciliary components 108

5.2.1 A base to build from: the role of centrioles in the eukaryotic cenancestor 108

5.2.2 An ancestral centriolar inventory 109

5.2.3 Expansion to the axoneme: conserved ciliary proteins 109

5.3 Divergence of cilia 110

5.3.1 Recurrent themes of diversity in ciliary structure during evolution 110

5.3.2 Species specific variation in ciliary components 112

5.4 Loss of ciliary components 113

5.4.1 Cilia evolution driven by loss 113

5.4.2 Land plant cilia: stepwise evolution to loss 113

5.5 Cryptic secondary roles and co-option of plant ancestral ciliary proteins 115

5.5.1 CCP proteins 115

5.6 Impact of current work 117

REFERENCES 120

APPENDIX 137

Figure Legend

FIGURE 1-1: PRESENCE OF CILIA ACROSS EUKARYOTES AND GENERAL CILIUM STRUCTURE 4

FIGURE 1-2: DISTRIBUTION OF CILIARY GENES AND CENTRIOLE DUPLICATION PATHWAYS 8

FIGURE 1-3: LAND PLANT PHYLOGENY AND THE LOSS OF CILIA 11

FIGURE 1-4: SPERM CELL RELEASE FROM *CERATOPTERIS RICHARDII* GAMETOPHYTES 12

FIGURE 1-5: TEM SHOWING CILIUM STRUCTURE IN CHLOROPHYTA 16

FIGURE 1-6: CILIARY STRUCTURES IN LAND PLANTS 18

FIGURE 1-7: EVOLUTIONARY MODIFICATION OF CILIARY STRUCTURE ACROSS THE LAND PLANTS 25

FIGURE 2-1: TUBULIN PHYLOGENY 37

FIGURE 2-2: DISTRIBUTION OF CENTRIOLAR/CENTROSOMAL PROTEINS ACROSS EUKARYOTES 39

FIGURE 2-3: CONSTRAINED PHYLOGENY OF SAS-4 AND SAS-6 42

FIGURE 2-4: PLK SUPERFAMILY 44

FIGURE 2-5: DISTRIBUTION OF CILIOPATHY-ASSOCIATED PROTEINS AMONG EUKARYOTES 47

FIGURE 3-1: OVERALL CILIA MORPHOLOGY AND ATTACHMENT TO THE SPERM CELL BODY 70

FIGURE 3-2: BASAL BODY STRUCTURE 71

FIGURE 3-3: STELLATE ARRAY STRUCTURE 73

FIGURE 3-4: CENTRAL PAIR ROTATION 75

FIGURE 3-5: CILIA HAVE TAPERED ENDS 76

FIGURE 3-6: SUMMARY OF CILIA STRUCTURE IN *CERATOPTERIS RICHARDII* 78

FIGURE 3-7: ULTRASTRUCTURE OF *MARCHANTIA POLYMORPHA* CILIA 84

FIGURE 4-1: CILIARY PROFILE PROTEINS 91

FIGURE 4-2: CCP PROTEINS 93

FIGURE 4-3: EVOLUTION AND FUNCTION OF CCP PROTEINS 96

FIGURE 5-1: POSSIBLE ORGANISATION OF CILIA IN LECA SHOWING ROLES OF KEY PROTEINS. 111

FIGURE 5-2: EVOLUTIONARY DISTRIBUTION OF KEY PROTEINS SETS 114

FIGURE 5-3: AUTOGENOUS THEORY OF ORIGIN OF CILIUM 118

1 The evolution of land plant cilia

1.1 Preface

This chapter is in the form of a review of ciliary structure with emphasis on phylogenetic diversity in the land plant lineage. It has been accepted for publication as a Tansley Review in New Phytologist. A review on this subject is timely in that there has been a recent explosion of interest in cilia, yet with the exception of *Chlamydomonas reinhardtii,* plant models have been largely ignored. In light of genomic, proteomic and bioinformatic research the large amount of older ultrastructural data on plant sperm cell morphology needs to be readdressed. The aim of this review is to position the land plant lineage as a model for studying the evolutionary changes, and eventual loss of cilia. The focus of the review is to introduce the evolutionary pattern of cilia presence in Eukaryotes, demonstrate that diverse morphologies exist in disparate species, and relate these morphological differences to protein components. The land plant lineage provides the basis for these discussions. Particular highlights include: 1) conservation of ciliary structure and components across eukaryotes and the biological statements which can be made from these phylogenetic distributions; 2) how plant cilia differ uniquely from the canonical form and why such modifications occurred; 3) the apparent presence of ciliated proteins in the non-ciliated angiosperms, with discussions of their possible roles and; 4) the introduction of the fern *Ceratopteris richardii* and surprisingly the angiosperm *Arabidopsis thaliana* as ciliary models. This review was written by Matthew E. Hodges, with discussions and editing contributions from Jane A. Langdale, Bill Wickstead, and Keith Gull.

1.2 Abstract

Eukaryotic cilia/flagella are membrane-bound microtubule organelles involved in sensory and motility functions. Cilia are composed of over 300 proteins and display significant ultrastructural conservation where present across the eukaryotic phylogeny. However, variation within phylogenetic lineages exists, and the ability to produce cilia has been independently lost multiple times. The land plants provide a model clade for investigating cilia evolution and loss across a broad phylogeny, in that basal land plants produce cilia, whereas the majority of seed plants do not. This review aims to highlight the spectrum of changes that have occurred in cilia across the land plants, both in terms of ultrastructural morphology and protein composition. We draw attention to a number of inferred ancestral ciliary proteins which are present in the non-ciliated seed plants. These proteins have presumably undergone either subfunctionalisation or co-option to novel roles. Finally, we propose the multi-ciliated fern *Ceratopteris richardii* as a model organism for studying ciliary function, particularly with respect to the complex mechanisms of ciliary number control, co-ordination of basal body migration and placement, and beat pattern regulation.

Key words: Seed plants, *Ceratopteris*, Flagella, Sperm, Basal Body, Stellate Array, Pollen

Abbreviations: LECA – last common eukaryotic ancestor, IFT – intraflagellar transport, PFR – paraflagellar rod.

1.3 Introduction

Eukaryotic cilia/flagella are membrane-bounded, microtubule-containing projections from the cell. Cilia/flagella (termed cilia within this review) are evolutionarily independent from their prokaryotic namesake, sharing no homology (see Silverman and Simon, 1974). The last eukaryotic common ancestor (LECA) possessed a cilium utilised for motility and sensory function (see Figure 1-1A; Carvalho-Santos et al., 2011b; Hodges et al., 2010; Mitchell, 2007). Examples of cilia formation remain in all of the major eukaryotic lineages (Figure 1-1B) and these complex organelles maintain a high level of conservation in overall morphology (Satir and Christensen, 2007). However, many lineage specific modifications have been shown to occur both in terms of additional structures and loss of canonical components. Moreover, the ability to produce cilia has been independently lost multiple times, such as in fungi and land plants (Liu et al., 2006; Renzaglia and Garbary, 2001).

The structure of a typical cilium consists of an axoneme which extends from the basal body (Figure 1-1A; Satir and Christensen, 2007). The canonical basal body can be partitioned into two structurally distinct regions. The proximal end of the basal body is formed from nine symmetrically-arranged microtubule triplets, which is homologous in structure to the mitotic centriole (once a centriole has docked with a cell membrane it is known as a basal body). The distal region of the basal body is termed the transition zone and marks the region where elongation of A and B microtubules from the triplets forms doublet microtubules. Extending from the basal body is the axoneme. The axoneme typically has a ring of nine outer microtubule doublets (called a 9+0 axoneme) and in many cases a central microtubule pair (known as a 9+2 axoneme). The microtubule doublets of the axoneme are connected via nexin links and specialised inner and outer arm dynein motor proteins (Gibbons and Gibbons, 1973). The dynein arms are attached to the A microtubule of each doublet, and positioned in such a manner that the dynein motor head domains are in close proximity to the B microtubule of the neighbouring

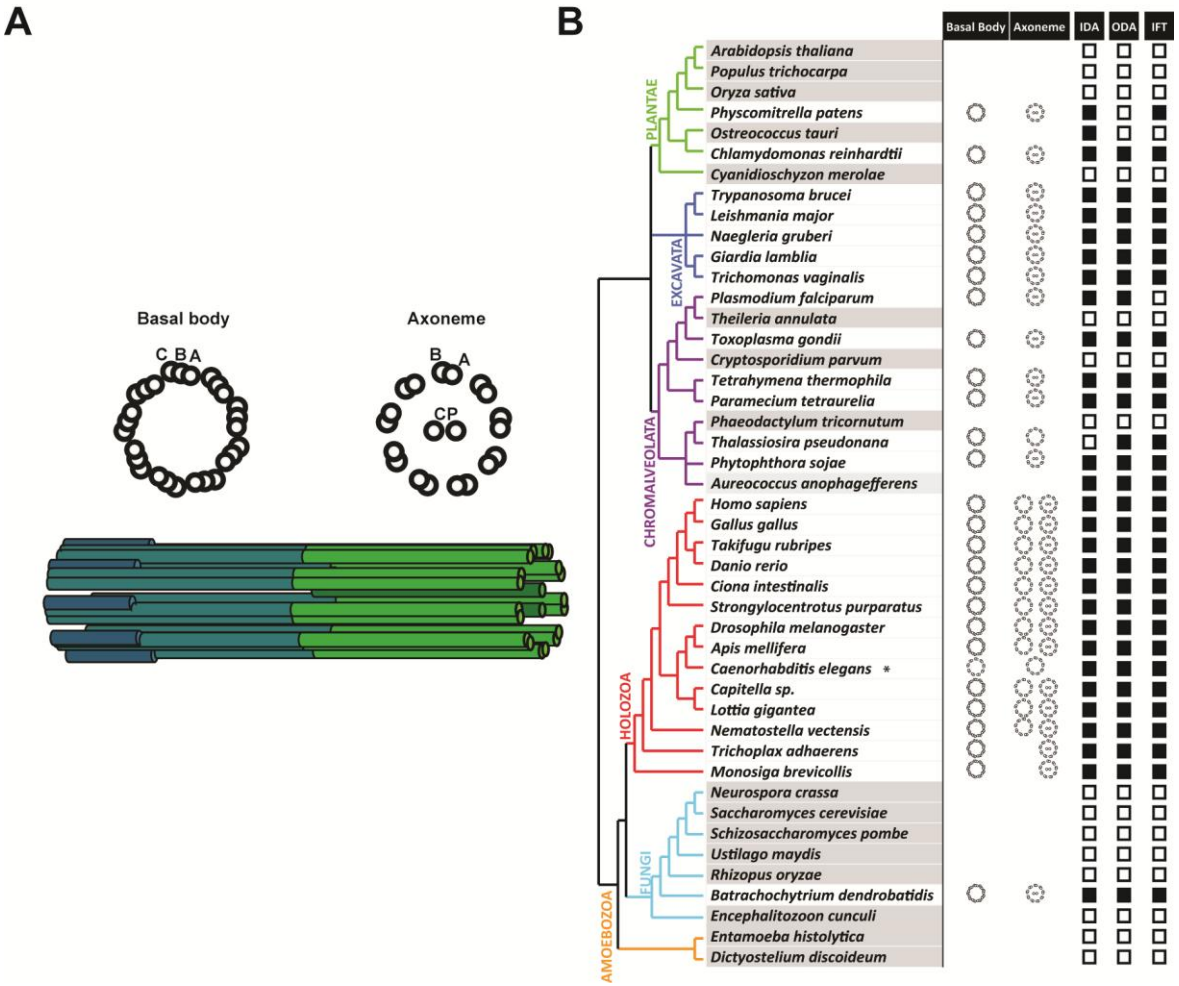


Figure 1-1: Presence of cilia across eukaryotes and general cilium structure

A) Diagrammatic illustration of the canonical structure of microtubule arrangement in a cilium, the basal body (left) and axoneme (right). Basal bodies are in general composed of nine triplet microtubules (A, B, C). Axonemes are composed of nine doublet microtubules (A, B) and may also possess a microtubule central pair (CP). B) Phylogenetic representation of 45 eukaryotic species. Grey panels indicate species which do not produce a centriole or cilium at any point in the life-cycle. Basal body and axonemal structure is shown, in addition to presence of Inner Dynein Arms (IDA), Outer Dynein Arms (ODA) and Intraflagellar Transport (IFT). *C. elegans* denoted by asterix, possesses highly divergent cilia.

doublet. When present, radial spokes attach the central pair microtubules to the surrounding doublets. These additional structures reinforce the axoneme resistance to bending, allowing microtubule sliding and thus bend propagation to occur (Mitchell, 2005). In most species bi-directional intraflagellar transport (IFT) carries cargo up the axoneme for maintenance and assembly, and down for recycling. Regulation of the equilibrium of these two IFT pathways determines ciliary length (for review see Rosenbaum and Witman, 2002b).

Multiple proteomic studies have catalogued the protein components of cilia in a number of evolutionary distant species (Keller et al., 2005; Keller et al., 2006; Kilburn et al., 2007; Pazour et al., 2005). Bioinformatic analyses have identified orthologues of a cohort of these ciliary proteins across eukaryotes (such as Avidor-Reiss et al., 2004). The phylogenetic distribution of ciliary proteins allows the inference that these proteins were present in the LECA. Interrogation in extant species of the LECA ciliary proteins shows that these proteins possess a spectrum of function, from those which also play important cellular functions outside of a ciliary role (e.g. γ-tubulin, which is found in all eukaryotic species irrespective of ability to produce a cilium; Oakley and Oakley, 1989) to those with a highly specialised role (e.g. inner and outer arm dyneins; Piperno et al., 1977). A database of ciliary components can be found at http://www.cilaproteome.org/. Modifications within this spectrum of proteins may therefore correlate with the changes in ciliary structure that are seen in specific lineages.

1.3.1 Diversity of ciliary structure

While the homology of many ciliary structures and protein components is ancient, cilia also display a vast array of modifications. This variation falls into two categories: within- and between-species variation. The distinction between primary and motile cilia falls within both of these categories. Primary cilia are immotile, composed of an elongated transition zone lacking in dynein motors and the central pair of microtubules. Primary cilia are also generally relatively short, projecting 1-2 μm from the cell membrane, and act in a sensory capacity (reviewed in Singla and

Reiter, 2006). Conversely motile cilia possess motor proteins (dynein arms) to generate sliding force, and generally contain a central pair. Motile cilia are also longer (in some cases up to 15 µm; reviewed in Satir and Christensen, 2007). The widespread phylogenetic distribution of motile cilia is indicative that the LECA possessed a motile cilium. On the other hand, primary cilia are an evolutionary innovation confined to metazoans. The presence of immotile cilia in other lineages is suggestive of independent evolutionary events, such as in the non-motile cilia of some centric diatoms (Jensen et al., 2003). Examining the life cycle biology of the species which demonstrate these examples of evolutionary homoplasy may shed light on the conditions driving the loss of motility in cilia. Furthermore, it would be interesting to determine whether the independent switch to loss in ciliary motility is achieved through loss of similar or different protein components in evolutionary distant species (i.e. comparing the ciliome of diatoms to primary cilia of metazoans).

Overlaid on the broad distinction between primary and motile ciliary forms are various morphological differences that may be reflective of differences in biological context between species. The canonical, well studied 9+2 microtubule axonemal arrangement is inferred to have been present in the LECA and remains visible in lineages as divergent as Metazoa and Plantae (Cavalier-Smith, 1987; Ringo, 1967; Silflow and Lefebvre, 2001). However, many examples of lineage specific extra-axonemal structures exist, and the presence of these structures has been linked to variation in protein components. Examples include the addition of an array of proteins in the excavate *Trypanosoma brucei* that result in the production a paraflagellar rod (PFR) structure built alongside the axoneme (Vaughan and Gull, 2003); and the diversification or loss of protein components in the nematode *Caenorhabditis elegans* that results in an immotile sensory cilium that lacks both a canonical triplet basal body structure and many typical ciliary proteins (Figure 1-1B; Hodges et al., 2010; Pereira-Leal et al., 2010; Perkins et al., 1986a). Lineage specific morphological variation is thus reflective of variation in protein composition, be it innovation or loss.

Diversity of ciliary structure is also reflected in diversity of ciliary assembly processes. Basal body assembly can occur through two pathways, both of which can happen in a single cell. In the canonical template pathway (centriolar) a new basal body is formed alongside the mature pre-existing basal body (see Figure 1-2A). Alternatively, in the *de novo* pathway (acentriolar), multiple basal bodies are formed in a cell which previously did not contain a basal body (Figure 1-2B; Beisson and Wright, 2003; Bettencourt-Dias and Glover, 2007). *De novo* assembly can be found in many eukaryotic species, such as in deuterosomes of metazoan terminally differentiated epithelia cells (Dirksen, 1971) and the blepharoplast of protozoans and land plants (reviewed in Dawe et al., 2007a). However, the most parsimonious explanation for the trait map of *de novo* synthesis in eukaryotes is that it has evolved independently multiple times, again suggestive of diversity in ciliary evolution.

Once a mature basal body has formed, extension of the axoneme can occur (Sorokin, 1962). In the canonical extracellular ciliogenesis pathway, inferred to have been present in the LECA, the basal body docks to the plasma membrane and then extends a microtubule axoneme. Conversely, during intracellular ciliogenesis the basal body initially docks to a vesicle membrane and the axoneme begins to be assembled within the vesicle. The vesicle then proceeds to dock and fuse with the plasma membrane. These two mechanisms of axoneme extension overlap in many areas, but differ in the site of membrane docking. Further differences are inherently linked to the timing and type of cilium assembly, in that cell cycle progression is often coupled to centriole/basal body positioning. For example, the excavates and some chromalveolates can form a new cilium without disassembly of the old (Gull et al., 2010), whereas most to species generally undergo disassembly of cilia before mitosis (Cavalier-Smith, 1974a).

Assembly and maintenance of cilia in general requires intraflagellar transport (IFT), with a few well noted exceptions (Rosenbaum, 2002). IFT involves the bidirectional transport of ciliary

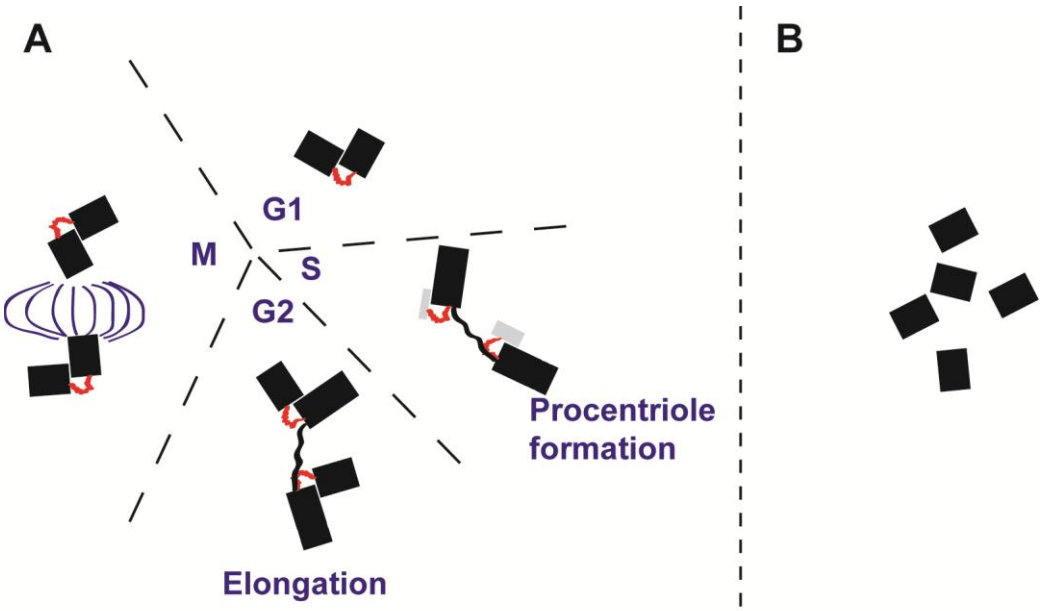


Figure 1-2: Distribution of ciliary genes and centriole duplication pathways

A) Centriole template duplication pathway through the cell cycle. Black cylinders represent mature centriole(s), grey immature centriole(s), red line centriole connector, and blue microtubule spindle fibres. Cells entering a new cycle possess one centrosome composed of a mother and a procentriole centriole. In G1, the mother centriole has the ability to become a basal body, and extend an axoneme. The centrioles duplicate in S phase. At the G2/M transition, the cilium is reabsorbed and the centrioles are used in mitotic spindle assembly. B) *De novo* centriole formation within a biocentriole/blepharoplast.

component cargo along the outer doublet microtubules of the axoneme (Parker and Katsanis, 2011). The role of motor proteins is vital in this transport of cargo: kinesin-2 mediates anterograde movement while cytoplasmic dynein 2 (DHC1B in *Chlamydomonas*) performs retrograde-IFT (Cole et al., 1998; Krock et al., 2009). However, even this highly conserved transport mechanism demonstrates plasticity within the eukaryotic lineage. For example, *Plasmodium*, *Cryptosporidium* and two *Theileria* species have lost the entire complex of IFT genes in favour of a different system that has yet to be fully characterised (Briggs et al., 2004).

As mentioned above, morphological differences in cilia structure are accompanied by the phylogenetic presence or absence of key inferred ancestral protein components. Even core structural proteins such as SAS-4 and SAS-6 (Dammermann et al., 2008; Delattre et al., 2006; Nakazawa et al., 2007) are not ubiquitously conserved in all ciliated species (Hodges et al., 2010; Pereira-Leal et al., 2010), suggesting that plasticity is a key driving force within ciliary evolution. Evolutionary plasticity is also apparent in the well-studied dynein superfamily. Phylogenetic analysis has shown that nine major classes of dynein heavy chains were present in the LECA (Wickstead and Gull, 2007), two of which encode cytoplasmic dyneins, and seven of which are built into the axoneme of motile cilia (forming either outer or inner-dynein arms). Subsequent loss has then occurred in different lineages. For example the land plants have lost all components of the outer-arm dyneins, but retained the inner-arm dyneins. This genomic study reinforces previous work which showed morphological absence of outer-arm dyneins in land plant sperm cells (Hyams and Campbell, 1985).

The general theme seen so far is that evolutionary morphological variation in many ways reflects protein composition diversity. However, the study of ciliopathies (diseases caused by defects in cilia) in various metazoan species has shown that mutations in a gene can differentially affect cilia at a tissue, and even cellular level (Fliegauf et al., 2007). This is the case in polycystic kidney disease, where mutations in genes encoding proteins Polycystin -1/-2 (PKD-1/-2) perturb ciliary

function in kidneys and some other tissue types (such as pancreas), but cilia of the trachea and sperm are largely unaffected (Yoder et al., 2002). The difference in ciliary phenotypes between cell-types represents either variation in the role of ciliary proteins in different cell types, or the role of cell context in ciliary function (i.e. some cell types may be more resilient to mutations in ciliary component genes than others).

In combination, the examples discussed thus far show that cilia evolution is marked both by conservation of form and function, and by diversification. In both cases, however, morphological variation reflects variation in protein composition. At the most dramatic end of the modification spectrum, loss of cilia has occurred independently multiple times. For example, loss is seen in both the fungal and land plant lineages (Liu et al., 2006; Renzaglia and Garbary, 2001). The land plant lineage is a good model for studying cilia evolution in that it is an ancient monophyletic group possessing examples of unique structural innovations, and at least two complete loss events.

1.4 Plant cilia: a case study for cilia evolution and development

The most parsimonious explanation for the phylogenetic distribution of cilia in the land plants (Figure 1-3A) is that cilia loss occurred twice; once in gymnosperms at the split of conifers and gnetales from cycads and *Ginkgo*, and once near the base of the angiosperms (Bremer et al., 1987; Finet et al., 2010). In land plants species with cilia, production is restricted to specialised sperm cells (Figure 1-4). In these cells, cilia are formed *de novo* from a structure termed the biocentriole in bryophytes and lycophytes, and from the larger blepharoplast in monilophytes and gymnosperms (Graham, 1978a, 1982; Graham and Mcbride, 1979; Graham and Wedemayer, 1984; reviewed in Renzaglia and Garbary, 2001).

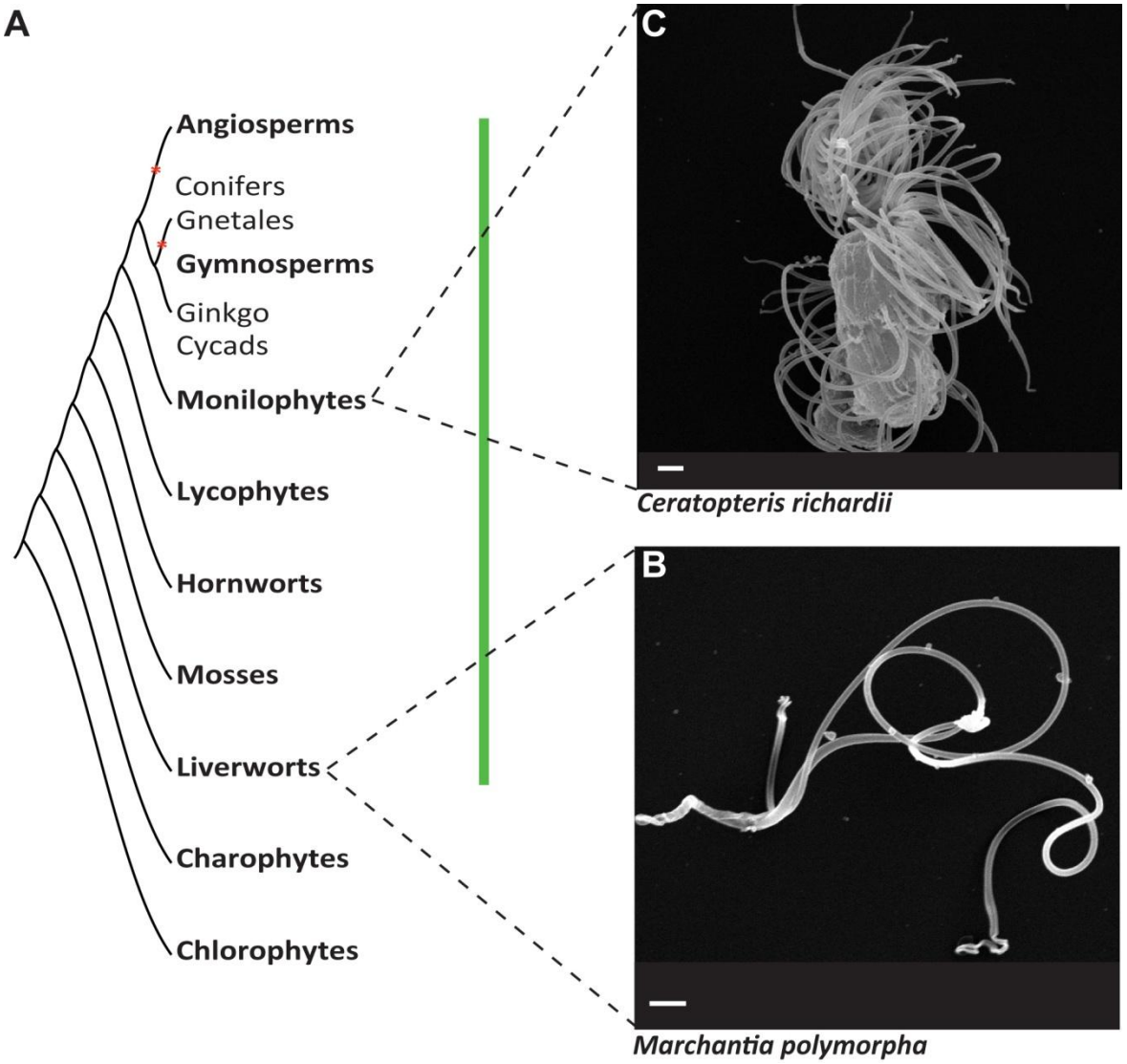


Figure 1-3: Land plant phylogeny and the loss of cilia

A) Cladogram showing relationships between the land plants. The most parsimonious explanation for the pattern of ciliary loss in land plants is two loss events (represented by red stars), once at the base of the angiosperms, and once in the gymnosperms where *Ginkgo* and the cycads diverge from the non-ciliated conifers and Gnetales. B) Scanning electron micrograph (SEM) of a sperm cell from the liverwort *Marchantia polymorpha*. C) SEM of a sperm cell from the fern *Ceratopteris richardii*. Scale bars 1 µm.

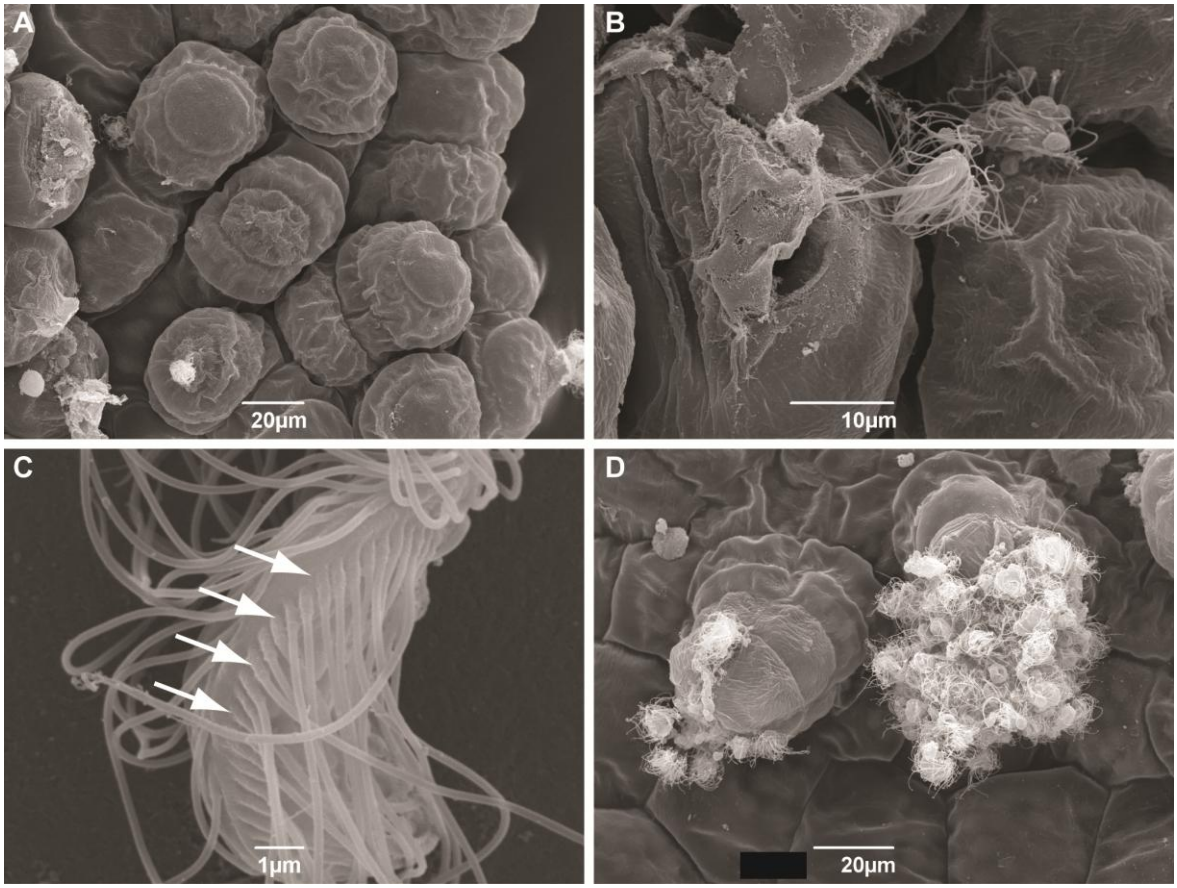


Figure 1-4: Sperm cell release from *Ceratopteris richardii* gametophytes

Scanning electron micrographs (SEM) of gametophytes and sperm cells from the fern *Ceratopteris richardii*. A, B) Male gametophyte showing abundance of antheridia from which sperm are released on addition of water (B). C) Free swimming sperm cell showing prominent ridge (white arrows) along the cell body from which multiple cilia extend. D) Multiple sperm cells clustered on female archegonia.

1.4.1 Structural differences in cilia morphology across the land plants

Plant cilia function as both sensory and motility organelles (Bremer et al., 1987). This functional plasticity is mirrored in morphological diversity and novelty, illustrated by unique features such as the nine-pointed stellate pattern first described in *Polytoma* algae (Lang, 1963). Ultrastructural studies of alga and land plant sperm cells (Table 1-1) allow for a detailed comparison of morphology and for a perspective on cilia evolution and loss to be gained. Analysis of morphological structures present in evolutionarily distant extant plant species further allows the inference of key morphological traits that are likely to have been present in the most recent common plant ancestor and that are maintained in ciliary-building Chlorophyta such as the algae *Chlamydomonas* (Geimer and Melkonian, 2004; Lang, 1963; Mitchell and Nakatsugawa, 2004; O'Toole et al., 2003; Silflow and Lefebvre, 2001), *Mesostigma (*Manton, 1965; Melkonian, 1989*)* and *Trentophilia* (Graham and Mcbride, 1975; Figure 1-5). Such inferences allow for the hypothesis that the cenancestor of the land plants possessed two cilia where: 1) the basal body is composed of nine triplet microtubules and is duplicated through the canonical template pathway (Figure 1-5A); 2) a cartwheel structure is visible at the very proximal base of the mature basal body (Figure 1-5B, arrowhead); 3) cytoskeletal fibres surround the base of cilia (Figure 1-5B, arrow); 4) the transitional zone contains a stellate pattern (Figure 1-5C); 5) the axonemes possess both inner and outer arm dyneins (Figure 1-5D, arrows); and 6) axoneme structure is composed of nine-fold doublet microtubule symmetry (Figure 1-5D and 1-7) and a central pair (Figure 1-5G). From this baseline of canonical features, multiple innovations have occurred throughout the land plant lineage (discussed below and summarised in Figure 1-7):

1.4.1.1 Increase in cilia number

An obvious difference between sperm cell morphology in the different land plant groups is a variation in cilia number. The number increases from two in bryophytes, such as liverworts, mosses and hornworts, to 40-80 in ferns and to >1000 in the gymnosperms; until cilia are lost independently within the gymnosperms and angiosperms (Figure 1-3A). This increase in cilia

Table 1-1: Key features of ciliary structure in Viridiplantae

Mature cilium structures for a number of studied Viridiplantae species are summarised. All cilia examined possess a radially symmetrical nine triplet basal body and 9+2 axonemal structure. The proximal end is either organised (O) or lacking symmetry (D), and within this region either fibres (F), amorphous density associated with the basal body (A) or extensive dense structure (E) are present. The stellate pattern is either restricted to the doublets of the transition zone (D), within doublets and triplets (T) or has been shown to be continuous throughout these regions (C). Uncertainty/ lack of information is represented by "-". For references see text.

	Species	De novo formation	Basal body number	Basal body structure	Approx. basal body length (nm)	Proximal end	Proximal density around basal body	Cartwheel	Stellate pattern	Cilia number	Axoneme Structure	ODA	Central pair rotation	Tip
Chlorophyta	*Chlamydomonas*	No	2	9 triplet	300	Organised	Fibres	Yes	Doublet	2	9+2	Yes	Yes	9+2
	Mesostigma	No	2	9 triplet	600	Organised	Fibres	Yes	Doublet	2	9+2	Yes	-	9+2
	Trentepohlia	No	2	9 triplet	200	Organised	Fibres	Yes	Doublet	2	9+2	Yes	-	
Charophyta	*Coleochaete*	Yes	2	9 triplet		Organised	Fibres	Yes	Doublet	2	9+2	No	-	9+2
	Chara	Yes	2	9 triplet	250	Organised	Fibres	Yes	Triplet	2	9+2	No	-	9+2
Bryophyta	*Marchantia*	Yes	2	9 triplet	PBB > ABB	Disorganised	Amorphous closely associated with BB	Absent in mature sperm	Doublet	2	9+2	No	-	-
	Aulacomnium	Yes	2	9 triplet	240	Disorganised	Amorphous closely associated with BB	Absent in mature sperm	Doublet	2	9+2	No	-	-
	Polytrichum	Yes	2	9 triplet	200	Disorganised	Amorphous closely associated with BB	Absent in mature sperm	Doublet	2	9+2	No	-	-
Lycophyta	*Lycopodium*	Yes	2	9 triplet	400	Disorganised	Amorphous associated with BB	Absent in mature sperm	Doublet, triplet	2	9+2	No	-	Tapers (SEM)
	Selaginella	Yes	2	9 triplet	500	Disorganised	Amorphous associated with BB	Absent in mature sperm	Doublet, triplet	2	9+2	No	-	Tapers (SEM)
	Isoetes	Yes	11	9 triplet	Equal length	Disorganised	Amorphous associated with BB	Absent in mature sperm	Doublet, triplet	11	9+2	No	-	Tapers (SEM)
Fern	*Ceratopteris*	Yes	36-100	9 triplet	350-450	Disorganised	Extensive amorphous within/around	Absent in mature sperm	Doublet, triplet, continuous	36-100	9+2	No	Yes	Disorganised (TEM)
	Marsilea	Yes	50-100	9 triplet	-	Disorganised	Extensive amorphous within/around	Absent in mature sperm	Doublet, triplet	50-100	9+2	No	-	Tapers (SEM)
	Lygodium	Yes	30-40	9 triplet	-	Disorganised	Extensive amorphous within/around	Absent in mature sperm	Doublet, triplet	30-40	9+2	No	-	Tapers (SEM)
Gymnosperm	*Ginkgo*	Yes	>1,000	9 triplet	850	Disorganised	Extensive amorphous within/around	Absent in mature sperm	Doublets, triplets	>1000	9+2	No	-	Tapers (SEM)
	Zamia	Yes	>10,000	9 triplet	-	Disorganised	Extensive amorphous within/around	Absent in mature sperm	Doublets, triplets	>10,000	9+2	No	-	-

number correlates with an increase in both sperm cell body size and nuclear content (Renzaglia and Duckett, 1991; Renzaglia and Garbary, 2001).

1.4.1.2 De novo *formation of basal bodies*

Centrioles/basal bodies act as microtubule organising centres (MTOC) and microtubule nucleation sites. Centrioles are called basal bodies when they dock to cell membranes and extend an axoneme. Centrioles also play key roles in nuclear spindle and cell division control by organising centrosomes. In many species, centrioles perform both functions by migrating around the cell body from the cell membrane at orchestrated times during the cell cycle (reviewed in Duncan and Wakefield, 2011). In these cases, centrioles or basal bodies are typically present in every cell during the whole cell cycle, duplicating before cell division (Figure 1-2A). Interrogation of the cell cycle of chlorophyte species, such as *Chlamydomonas*, shows that centrioles persist throughout the cell cycle in this lineage. However, examination of charophyte and land plant species shows that these lineages do not persistently maintain centrioles for mitosis throughout the cell cycle. Rather, it can be inferred that a novel pathway of basal body assembly evolved. This pathway assembles basal bodies de novo within structures termed biocentrioles or blepharoplasts (Figure 1-2B, 1-6A-B; Graham, 1982; Graham and Wedemayer, 1984). Loss of the ability to produce these structures coincides with the loss of ability to produce cilia within the conifers, gnetales and angiosperms.

1.4.1.3 Asymmetric microtubule arrangements at the proximal base of the basal body

A canonical cartwheel structure is found at the proximal base of basal bodies from many different species, and as such is likely to have been present in the LECA. This cartwheel structure is suggested to act as a scaffold and stabilising factor for the nine triplet microtubules (Hiraki et al., 2005; Hiraki et al., 2007). In the majority of ciliated eukaryotic species, the proximal base of the triplet basal body, in many instances containing the cartwheel, marks a defined end to the triplet microtubules. However, examination of the proximal end of basal bodies which have extended



Figure 1-5: TEM showing cilium structure in Chlorophyta

A-D) Transverse thin sections through cilia of *Trentepohlia* showing triplet microtubules of the basal body with the typical nine-fold triplet microtubule (A). At the proximal base of the basal body a cartwheel structure (white arrowhead) is seen along with connecting fibres (arrow) (B). The transition zone contains a discrete structure termed the stellate array, which is unique to ciliated Viridiplantae (arrow) (C). The canonical 9+2 arrangement of doublet microtubules is present in the axoneme, and both inner and outer dynein arms are formed (arrows) (D). Reproduced with permission from Graham and Mcbride, 1975. E) Longitudinal section through the base of a cilium in *Mesostigma*. Of note is the distinctive basal body (bb), a small stellate array in the transition zone (st), and the central pair of the axoneme (ax). F) Transverse section showing axoneme structure of both cilia from a sperm cell of *Mesostigma*. E-F reproduced with permission from Melkonian, 1989. G) Longitudinal thin section of *Chlamydomonas* flagellum showing changes in central pair orientation with respect to outer doublet microtubules (see tubular models alongside cilium). Modified from Mitchell, 2003. Scale bars A-F) 50 nm, G) 0.2 µm.

axonemes in land plants shows that the cartwheel is missing or highly divergent. Furthermore, the structure of the very base of these cilia (Figure 1-6C-H) is typified by a loss in the nine triplet microtubule symmetry such that transverse sections illustrate doublet and even singlet microtubules (examples include Bilderback et al., 1974; Carothers and Kreitner, 1968; Duckett and Carothers, 1982; Graham, 1978a; Manton, 1965; Norstog, 1974; Renzaglia et al., 1999; Vaughn and Renzaglia, 2006). This ultrastructural feature may result either from bidirectional nucleation of microtubules at basal bodies, or from degradation of microtubule structure at the proximal base after sperm cell maturity; temporal evidence has yet to be produced to distinguish between these two hypotheses.

1.4.1.4 Elongation of the basal body module

Basal body length in land plant cilia varies across the phylogeny (see elongated bryophyte sperm basal bodies in Figure 1-6H). Among the basal land plants (liverworts, mosses and hornworts) assembled mature basal bodies have an overall length of between 200-250 nm (Bernhard and Renzaglia, 1995; Carothers and Kreitner, 1968; Paolillo et al., 1968; Paolillo, 1981; Renzaglia and Duckett, 1991), similar to those of the sister charophyte lineage (Geimer and Melkonian, 2004; Manton, 1965; Melkonian, 1989). Within lycophyte and fern sperm cells, basal body length is between 350-550 nm (Bilderback et al., 1974; Carothers et al., 1975a; Duckett et al., 1979; Hoffman et al., 1994; Maden et al., 1997; Renzaglia et al., 2002; Renzaglia et al., 2001; Renzaglia et al., 2004) whereas the motile sperm producing seed plants, *Ginkgo* and cycads, possess even longer basal bodies, of between 800-900 nm in length (Norstog, 1974; Vaughn and Renzaglia, 2006). While it is clear that much of this increase in length results both from elongation of the transition zone, and from lack of symmetry at the proximal base, further analysis of serial sections is required to fully understand changes in the triplet microtubule region of the basal body. Notably, elongation of basal body length coincides with increases in overall cilia number on each sperm cell across the land plants. This observation suggests that a larger anchoring structure may be required to ensure cell membrane integrity under greater mechanical stress.

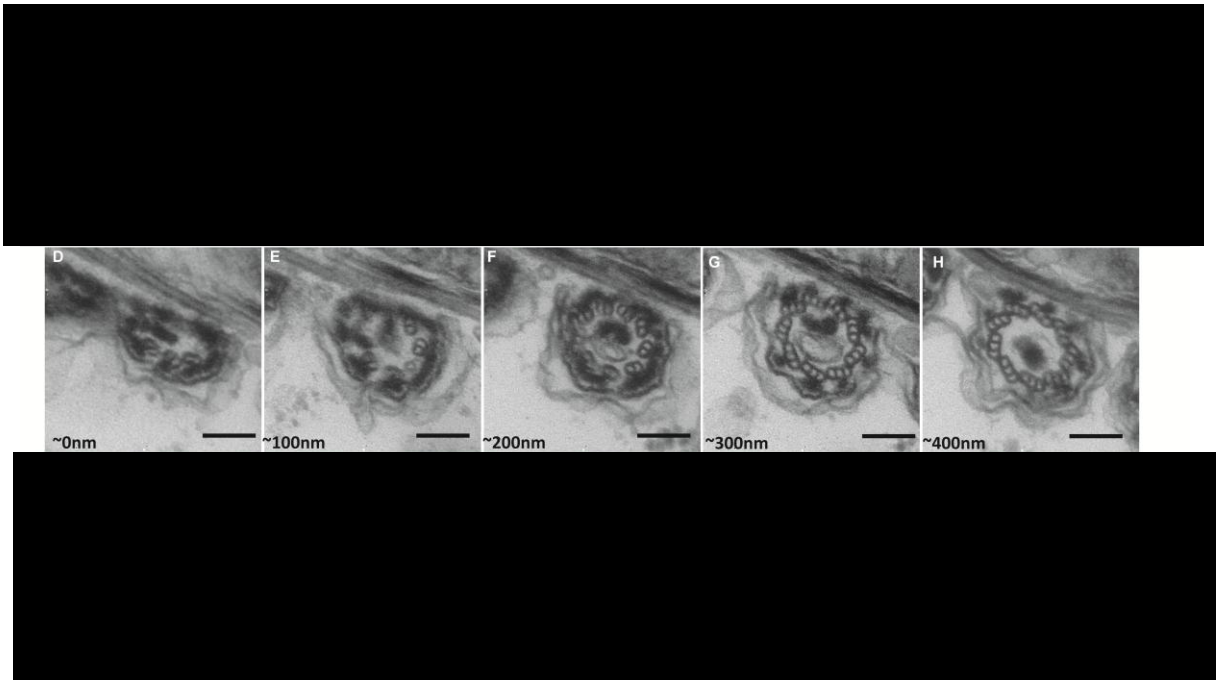


Figure 1-6: Ciliary structures in land plants

Transmission electron micrographs of cilia from land plant species. A) Typical fern blepharoplast structure which acts as the site of *de novo* formation of centrioles in sperm cells, and has a characteristic electron dense appearance. Adapted from Hoffman et al., 1994 (A) and from Norstog, 1974; Vaughn and Renzaglia, 2006 (B). C) Example of the basal body found within lycophytes. Image adapted from Carothers et al., 1975a. D-H) Serial sections showing breakdown of nine fold microtubule symmetry at the proximal base of cilia in the fern *Ceratopteris*, with the different lengths of microtubules resulting in singlet and doublet microtubules and a dense amorphous structure. I) Amorphous density seen in the sperm cells of the cycad *Zamia*, which may act to aid position and docking of basal bodies to the cell membrane. Image adapted from Norstog, 1974; Vaughn and Renzaglia, 2006. J) Examples of stellate arrays in fern sperm cilia, which are contiguous from within the triplet basal body to the doublet microtubules in the transition zone. Image adapted from Hoffman et al., 1994. Scale bars A,B), 200 nm, C-G) 50 nm, H-I) 200 nm, J) 50 nm.

Basal body elongation occurs through two mechanisms which progress one after the other. Initially the A-tubule is added to the top of the cartwheel structure, and subsequently the B- and C-tubules are formed alongside this pre-existing structure (Cavalier-Smith, 1974a). Tomographic studies in *C. elegans* (Pelletier et al., 2006) support this mechanism for assembly in that immature basal body assembly begins with the formation and elongation of a central tube followed by the peripheral assembly of the nine singlet microtubules. Once the triplet microtubules of the basal body are formed, the elongation of basal bodies can proceed with a capping structure, as is the case in the growth of kinetochore microtubules. During this process, proteins travel to the distal tips of elongating microtubules (Nigg and Raff, 2009).

1.4.1.5 Increased density at the proximal base of land plant cilia

The proximal base of ciliated chlorophyte basal bodies, such as those in *Chlamydomonas reinhardtii*, typically contains structures such as rootlets and striated, and concentric fibres (Sleigh, 1979). These features are also present in cilia of the charophyte algae *Chara* and *Coleochete* (Graham and Mcbride, 1979; Turner, 1968). However, in basal land plants (liverworts, mosses and hornworts), this same region is characterised by the appearance of dense amorphous regions alongside the canonical basal apparatus (Carothers and Kreitner, 1968). Within the vascular plants the number and area of amorphous patches increases, to the extent that in the most recently diverged ciliated gymnosperms, all of the cell body around the basal bodies stains densely (Figure 1-6I; Norstog, 1974; Vaughn and Renzaglia, 2006). In well studied species such as the fern *Ceratopteris richardii*, novel connective fibres are also seen within this proximal region, linking the subtending microtubules of the cytoskeleton to the C-tubule of the basal body triplets (see Chapter 3). While the role and protein composition of such amorphous patches is unknown, studies in the green alga *Chlorogonium elongatum* suggested that dense regions at the base of cilia are not required for waveform changes or beat patterns (Hoops and Witman, 1985). Instead it may be that this electron-dense region is required for positioning and docking of the basal body within the membrane. This hypothesis is supported by the observed increase in the size of the

electron-dense region in cilia within the land plant phylogeny – as cilia number increases in more recently derived species, so too might the need for increased coordination of basal body patterning. The cilia-packed membranes of gymnosperm sperm cells (>1000 cilia per cell), would thus be expected to exhibit the most dense staining at the base of cilia and this is indeed the observed case (Figure 1-6A).

1.4.1.6 Stellate domain expansion

The stellate array of ciliated Viridiplantae, located in the transition zone of the basal body, consists of a number of filaments that connect the A-tubules of every second outer doublet to each other, to form a nine-pointed star structure in transverse section. Although the composition of the stellate array is not known, ultrastructural immunolocalisation studies in *Chlamydomonas* (Geimer and Melkonian, 2005), *Chara* (Jin and Hasenstein, 2009; Vouilloud et al., 2005) and *Ginkgo* (Vaughn and Renzaglia, 2006) have shown that centrin is localised to the same area. While centrin may indeed be a component of the stellate array, the protein also localises to many other regions of the basal body. As such it is likely that other, as yet unidentified, proteins act as the principle components of the array. In the basal land plants, such as the liverwort *Marchantia* (Carothers and Kreitner, 1968), and the mosses *Aulacomnium* (Bernhard and Renzaglia, 1995) and *Polytrichum* (Paolillo et al., 1968) the stellate array spans a longer domain than seen in the Charophyta, but is still only present in the doublet microtubule region of the transition zone. Interestingly, the stellate array is absent from the cilia of hornworts - an example of independent loss of this structure within the land plants (Carothers and Duckett, 1980). Comparative studies between ciliary protein preparations from mosses, liverworts and hornworts may thus elucidate potential candidates for stellate array components.

In comparison to the majority of bryophytes, all ciliated vascular plants (typified by the lycophytes *Lycopodium*, *Selaginella* and *Isoetes* (Carothers et al., 1975a), the ferns *Ceratopteris*, *Marsilea* and *Lygodium* (Bilderback et al., 1974; Duckett et al., 1979; Hoffman et al., 1994; Renzaglia et al.,

2002; Renzaglia et al., 2004)) and the gymnosperms *Ginkgo* and *Zamia* (Norstog, 1974; Vaughn and Renzaglia, 2006), possess a longer stellate array that spans the transition zone from the triplet to doublet microtubule regions (Figure 1-6J). In *Ceratopteris*, the stellate array has been shown by electron tomography to have a scaffold-like nature (see Chapter 3). The increase in stellate domain length may again be related to the increased number of cilia present on the sperm cells of vascular plants. While it has been hypothesised that the array plays a role in coordinating cilia beat (Lang, 1963), this is unlikely to be the case as force generated through pendular motion is insufficient to explain complex beat patterns. An alternative hypothesis is that the structure acts to strengthen the base of motile cilia. As cilia number increased through evolution, this domain may also have increased in length to counter the increased forces acting on the sperm cell membrane. However, it must be noted that the basal body has also increased in overall size during land plant evolution, and thus the observed increase in stellate domain length may simply reflect a more general increase in cilia size.

1.4.1.7 Axonemal tips

The distal tips of eukaryotic cilia typically exhibit a discrete end point that is coincident with the termination of the central pair microtubules in the axoneme and is associated with a cap structure that is positioned at the tips of the microtubules. As a result the axonemal 9+2 organisation of microtubules is maintained from the appearance of the central pair in the axoneme just distal to the transition zone, through to the tip of the cilium (Dentler, 1980a, b). In most metazoan cilia, the outer doublet microtubules are capped, however, cap structures can vary. For example, in *Chlamydomonas* the central pair microtubules are also capped. These capping structures are hypothesised to regulate microtubule disassembly and growth, with the association between the cap and ciliary membrane appearing to limit the amount of microtubule displacement at the axonemal tip (Rosenbaum and Child, 1967). This suggestion is supported by the lack of caps in mammalian sperm cells (Woolley and Nickels, 1985) where the combination of explosive growth of the axoneme and the relatively short time span for which the cells have to be

motile for, results in a limited need for mechanisms which would normally stabilise the tip of the axoneme.

The land plants display a breakdown in the 9+2 microtubule axonemal arrangement at the tips of sperm cell cilia. Transverse ciliary sections in the tip region show that the doublet microtubules terminate at different positions, resulting in a lack of nine-fold symmetry and disorganisation. For example, in the fern *Ceratopteris* (see Chapter 3; Duckett et al., 1979) central pair microtubules are lost before the outer doublets. The outer doublet microtubules are then lost in no consistently observed order, and finally the B-tubule of the doublet ends proximally to the end of the A-tubule. The result is a breakdown of microtubule order at the end of the axoneme.

1.4.1.8 Rotation of the central pair

Rotation of central pair microtubules in relation to the nine outer doublets of the axoneme during cilia beating was first reported in *Paramecium*, and has since been shown in a number of algal species, including *Chlamydomonas* (Figure 1-5G; Mitchell, 2003; Mitchell and Nakatsugawa, 2004; Omoto et al., 1999; Omoto and Kung, 1980; Omoto and Witman, 1980, 1981). The presence of this feature within many lineages closely related to the land plants allows the inference that the last common land plant ancestor may have possessed the ability to rotate central pair microtubules. Serial sections of cilia in the sperm cells of the fern *Ceratopteris* (see Chapter 3) shows that the central pair in this species can twist in relation to the outer doublet microtubules. While rotation has not yet been shown, these data suggest that land plants may retain the ability to rotate ciliary central pair microtubules. It has been proposed that central pair twisting allows complex ciliary waveform changes, such as changes in beat direction (Mitchell, 2005). As land plants appear to have retained this feature, land plant sperm cells are thus likely to be able to undergo changes in beat direction.

1.4.2 Ciliary proteins lost on transition to land

Whilst the number of cilia on sperm cells increases during the evolution of land plants, ultimately cilia were lost. This may in part have been driven by sperm competition favouring a decreased number of larger sperm to optimise overall energy expenditure on gamete production (Miller and Pitnick, 2002). However, structural changes across the land plant phylogeny, such as the loss of nine-fold microtubule symmetry at both the proximal and distal ends of the cilium, are suggestive that loss of cilia may have occurred in a stepwise fashion.

In support of the suggestion of a stepwise loss, bioinformatic analysis has shown that key ciliary protein components have been lost within the land plants (Bettencourt-Dias and Glover, 2007; Hodges et al., 2010). Over a third of ciliary proteins that are highly conserved in ciliated eukaryotic species, but not in non-ciliated species, are absent from ciliated land plant genomes (see Chapter 4). Examples of those with a known ciliary role that are absent from land plants include: 1) Bardet-Biedl Syndrome (BBS) complex (BBS 1/2/5/7/8/9). These proteins play important roles in trafficking cargo to the basal body and axoneme (Blacque and Leroux, 2006; Fan et al., 2004; Mykytyn et al., 2004); 2) Kinesin-II Associated Protein (KAP) that functions in anterograde intraflagellar transport and sensory transduction (Pan and Snell, 2002); 3) Rab-related GTPases (such as RABL2A and RBJL1) which play a role in the regulation of vesicle transport from the Golgi body to the cilium base (Li and Hu, 2011); 4) Outer arm dynein and tektin, which are built into motile axonemes (Kamiya and Okamoto, 1985; Pirner and Linck, 1994; Yano and Mikinoumura, 1981); and 5) Certain radial spoke proteins (e.g. PF1) which are involved in cilia movement (Diener et al., 1993; Hyams and Campbell, 1985; Wickstead and Gull, 2007). The absence of these proteins elucidates a number of biological features of land plant cilia. For example, compared to the inferred ancestral eukaryotic cilium, the land plant cilium most likely have a reduced sensory role because of the absence of BBS proteins. Motility is also likely to be altered due to lack of outer dyneins, tektin and certain radial spoke proteins. Potentially different Golgi-to-cilia trafficking would result from a reduction in Rab-related GTPases and loss of proteins

such as Ca^{2+} ion channels may influence the co-ordination of ciliary beat. In combination, loss of these proteins suggests a regulatory shift during the colonisation of land.

1.4.3 Ciliary proteins conserved in non-ciliated plants

The inability to produce cilia in the majority of seed plants has an interesting parallel in the evolution of ciliary structure in dipteran *Drosophila* lineages, where some fly species produce giant, relatively immotile sperm cells (Bjork and Pitnick, 2006; Lopez-Fernandez et al., 2007; Phillips, 1970). Cilia of insect sperm cells and of fern and ciliated seed plant sperm cells possess altered ciliary ultrastructure when compared to the inferred state of their respective ancestors. For example, insect sperm cell cilia develop additional outer singlet microtubules within the axoneme, while land plant sperm cells possess many novel innovations (see Chapter 3). The result of such an evolutionary trend in land plants was the loss in ability to produce cilia, which correlated in a shift away from water-borne reproductive mechanisms on innovation of the seed. The insect lineage may be analogous in this respect, in that a possible evolutionary trajectory in insects may be the loss of cilia.

It could be assumed that the loss of cilia in the majority of land plants would be accompanied by a similar loss in ciliary function genes. While it is indeed the case that the majority of ciliary function proteins are lost during this transition, bioinformatic analysis of the genomes of non-ciliated flowering plants shows that certain ciliary function proteins are retained (see Chapter 4). This suggests an evolutionary process of co-option, whereby ciliary proteins perform novel roles in a non-ciliary context. The expression pattern of these retained genes in the flowering plant *Arabidopsis thaliana* shows that they are highly expressed in mature pollen (Schmid et al., 2005). As the pollen germ cell divides to produce sperm cells, this pattern is reminiscent of reflects the restricted production of cilia in sperm cells of ciliated land plants.



Figure 1-7: Evolutionary modification of ciliary structure across the land plants

Diagrammatic representations of ciliary ultrastructure. A) Ciliary structure representative of that in the algae *Chlamydomonas*. Cilium possess a triplet basal body with a proximal cartwheel and distal stellate array within the transition zone, and a 9+2 canonical microtubule arrangement throughout the axoneme. B) Land plant cilium possessing altered microtubule structures at base and tip, and an elongated stellate array spanning from the triplet microtubules of the basal body to the doublets of the transition zone. Basal body (bb), transition zone (tz) and axoneme (ax) are shown.

1.5 Future perspectives: *Ceratopteris* and *Arabidopsis* as ciliary models

Ciliated land plant species provide a system to study ciliary change over a large evolutionary period that is complementary to ciliary studies in model species such as *Chlamydomonas*. Previous ultrastructural studies of sperm cell morphology within the land plants have shown that numerous changes to cilia structures have occurred within the group (simplified in Figure 7). There is thus is a great scope for increasing our knowledge of such changes in phylogenetically important clades. When compared to other ciliary models, the land plants are particularly valuable for investigating the absence of centrioles in the cell cycle and the restriction of cilia to an abundant, easily isolated discrete cell type (sperm cells). Furthermore, given that many plant species can be propagated vegetatively, mutations can be studied that are lethal in many other eukaryotes – in effect because the cell cycle and ciliary roles of centrioles/basal bodies are uncoupled.

Ceratopteris richardii is a particularly attractive model species for ciliary studies because it has a relatively short life cycle compared to other land plants, and it produces many easily isolated multi-ciliated sperm (Hickok et al., 1995). As such, technically demanding questions involving *de novo* basal body formation, spatial and temporal migration and docking of basal bodies to the cell membrane, and ciliary beat coordination, can be investigated. The consistent positioning of basal bodies and axonemes within *Ceratopteris* sperm cells is suggestive that precise placement is required for effective swimming. Genes which affect migration and placement might thus be identified through mutant screens that select for altered swimming speed phenotypes. The current status of *Ceratopteris* as a model organism is such that mutagenesis and forward genetic screens are fairly straightforward (for a classic example see descriptions of the cilia *sleepy* mutant; Duckett et al., 1979). For this species to become a model system for cilia studies, however, an investment is needed in genome sequencing and transformation technology. The

large genome size (~39 chromosomes, $5x10^9$ base pairs; Hickok et al., 1995), makes this a non-trivial problem, however, the potential benefits for ciliary research are immense.

Perhaps counter-intuitively, the non-ciliated flowering plants, such as *Arabidopsis thaliana,* also present an opportunity for ciliary research, as they retain many genes with a known ciliary function in ciliated species. These plant species provide a system in which to study the process of co-option of ciliary function genes. Research in this area may well shed light on mutant phenotypes seen in animal models and potentially could provide insight into ciliopathies in humans.

1.6 Principles of electron microscopy

A transmission electron microscope (TEM) can be separated conceptually into three sections under a vacuum system to ensure low pressure (reviewed in Egerton, 2005). The first section is the illumination system, which includes the electron gun (which uses a tungsten filament to produce a high-voltage electron beam by thermionic emission) and condenser lenses to focus the electrons onto the specimen (primary beam formation). Generally, TEMs use an accelerating voltage of between 100-300 keV, as this increases the electron energy to such an extent that the de Broglie wavelength of electrons is decreased, lowering the diffraction limit to spatial resolution. The second section is the specimen stage (either held stationary or able to tilt), and whose mechanical stability is important in determining the spatial resolution (Ross et al., 2003). The third section of an electron microscope is the imaging system, which includes at least three lenses (objective and projector) that together produce a magnified image (or diffraction pattern) of the specimen on a fluorescent screen, photographic film, or by means of a fibre optic light-guide to the sensor of a CCD (charge-coupled device) camera. The objective lenses focus the beam that comes through the sample itself, while the projector lenses expand the beam onto the imaging device. Additional quad or hexipole lenses allow for the correction of asymmetrical beam

distortions (astigmatism) or spherical aberration, while energy filtering may be used to correct electron chromatic aberration (Ross et al., 2003).

The whole process occurs at low pressure to ensure that the potential difference between the cathode and anode does not generate an arc, and also to reduce the collision frequency of electrons with gas atoms to negligible levels (mean free path). The electron beam is focused by electrostatic and electromagnetic lenses, which rely on the fact that electrons behave as negatively charged particles and are therefore deflected by electric or magnetic fields, and transmitted through the fixed and stained specimen. The major disadvantages of TEM is the need for extremely thin specimen sections, and the fact that specimens are generally required to be chemically fixed, dehydrated and embedded in a polymer resin (see Chapter 3) to stabilize them sufficiently to allow ultrathin sectioning (Porter et al., 1953). Sections also need to be stained with heavy atom labels in order to achieve the required image contrast.

The scanning electron microscope (SEM) was invented shortly after the TEM (reviewed in Egerton, 2005). Within an SEM, electrons are generated as in TEM through thermionic emission from a tungsten filament (although in comparison to a TEM the electron gun is smaller as the maximum accelerating voltage is much lower at ~30 keV). Above the specimen stage are positioned the condenser and objective lenses, which determine the spatial resolution of the instrument. During raster scanning, the electron beam of an SEM is scanned horizontally by sawtooth-wave generators across the specimen in two perpendicular directions (x and y). Magnification is achieved by reducing the size of the raster on the specimen. The outputs of the two scan generators are applied to a display device, such as a synchronised cathode-ray tube, on which the distribution map of the intensity of the signal being emitted from the scanned area of the specimen will appear. For SEM analysis, a specimen requires chemical fixation, dehydration and critical point drying. The dry specimen is then mounted on a stub and splutter-coated with gold (which has a high atomic number and thus provides a high tomographic contrast and resolution) before examination in the microscope.

1.7 Phylogeny estimation using Maximum Likelihood and Bayesian inference

Likelihood-based methods of phylogenetic inference, including maximum-likelihood and Bayesian approaches, have been shown to be accurate in a number of phylogenetic simulation studies, particularly when compared to other methods such as distance or parsimony (reviewed Lewis, 2001). In general, a standard phylogenetic model assumes a tree with a discrete topology parameter and a set of independent branch lengths. Molecular evolution is typically modelled at the nucleotide or amino acid level using models.

The maximum likelihood (ML) principle (reviewed in Huelsenbeck, 1997; Guindon et al, 2010) maximises the probability of the data, D, given the parameter values, θ ($Pr(D|\theta)$). Within this calculation in a phylogenetic context, the data are generally an alignment of sequences, with each site having a likelihood which differs depending on the evolutionary model and the tree. The total likelihood is the product of the site likelihoods – the maximum likelihood tree is the tree topology that gives the highest likelihood under the given model (which takes into account the tendency of one base to change to another, composition of the sequence, and site-to-site rate variation). Models differ in their free parameters, and in general the more free parameters, the better the fit (higher the likelihood). However, increasing free parameters can result in "overfitting" of the model, and also increases the variance and reduces the power to discriminate among competing hypotheses. Nonparametric methods such as bootstrapping are used to determine confidence in ML inferences. The bootstrapping technique involves creating new alignment sets by resampling randomly (and with replacement) from the original characters, and assumes that the original characters were independently sampled. These calculated bootstrap data sets are of the same size as the original. A desired quantity of inferred trees (such as 100) are computed and used to calculate a consensus tree of branch topology.

An alternative approach to the maximum likelihood method is the Bayesian approach (reviewed in Ronquist & Deans, 2005). Bayesian approaches involve making inferences using the posterior distribution (the posterior) of hypotheses or parameters of interest, ($Pr(\theta|D)$). Prior beliefs in the parameter values need to be specified, and these are then combined with the likelihood to give the posterior. Posteriors are approximated using Markov Chain Monte Carlo (MCMC) algorithm. In this, a Markov chain is run that takes the form of a correlated random 'walk' through the parameter space, with the probability distribution being approximated by sampling values periodically. Running the Markov chain for a sufficient number of steps can make the approximation arbitrarily accurate, where each step involves a random modification of the tree topology, a branch length or a parameter in the substitution model (such as the substitution rate ratio). The proposed step is taken if the posterior that is computed for a proposed step is larger than that of the current tree topology and parameter values. Proposed steps that result in downhill moves on the posterior surface will only be taken dependent on the ratio of the new and current posteriors. Using these rules, the Markov chain visits regions of tree space in proportion to their posterior. The difficulty with MCMC sampling is to determine when to stop; while a correctly constructed MCMC will eventually converge onto the posterior distribution, this process may be very slow, particularly in multimodal posterior (known as poor mixing). Mixing can often be improved by using Metropolis coupling, in which chains capable of different magnitude changes are used to allow escape from local optima. Further, convergence of samples from independent runs can be compared as they should have identical samples of the posterior. This convergence can be measured using the potential scale reduction factor, which is expected to approach 1 as runs converge. Furthermore, tree samples can be compared using the standard deviation of split frequencies, which should approach 0 as runs converge. A consensus tree can be caluclated from these sampled trees, and Bayesian posterior probabilities of individual clades can be used as clade credibility values.

1.8 Thesis aims

The objectives of the work presented in this thesis were threefold. First, to identify a core set of key ciliary function genes that is conserved in the majority of ciliated species across the eukaryotic phylogeny (Chapter 2). Second, to investigate evolutionary changes that occurred throughout the evolution of land plant cilia: in terms of both ultrastructural morphology and protein composition (Chapter 3). Finally, to illustrate how the process of loss of such a conserved organelle in the seed plants affects gene loss and co-option (Chapter 4). A synthesis of the impact of the work on our understanding of ciliary form and function is presented in Chapter 5. Chapters 1-4 are presented as papers that are either published, submitted, or in preparation. For the purpose of this thesis, all references have been gathered in a single style, and figures have been renumbered and formatted to relate to each chapter. Information that was published as supplemental, or is supplementary, is included in a section called additional information at the end of each chapter.

2 Reconstructing the evolutionary history of the centriole from protein components

2.1 Preface

This chapter reprints a paper published in the *Journal of Cell Science* in 2010 (Hodges et al., 2010). The work arose as a method of testing a whole genome approach for linking organismal biology with phylogenetic profiling. This technique is utilised in much of the subsequent bioinformatic analysis in this thesis. There are five authors on the paper and the contributions to the work were as follows: Matthew E. Hodges designed and performed and analysed the bioinformatic experiments, and co-wrote the paper with Nicole Scheumann; Bill Wickstead, Jane A. Langdale and Keith Gull all provided contributions to data analysis and reviewed the manuscript.

2.2 Abstract

Centrioles are highly conserved structures fulfilling important cellular functions such as nucleation of cilia/flagella (basal body function) and organisation of pericentriolar material to form the centrosome. The evolution of these functions may be inferred from the distribution of the molecular components of extant centrioles and centrosomes. Here, we undertake an evolutionary analysis of 53 proteins known either for centriolar association or for involvement in cilia-associated pathologies. By linking protein distribution in 45 diverse eukaryotes with organism biology, we provide molecular evidence to show that basal body function is ancestral, whereas the centrosome is specific to the Holozoa. We define an ancestral centriolar inventory of 14 core proteins, Polo-like-kinase, and proteins associated with Bardet-Biedl Syndrome (BBS) and Meckel-Gruber Syndrome. We show that the BBSome is absent from organisms that produce cilia only for motility, predicting a dominant and ancient role for this complex in sensory function. We also show that the unusual centriole of *Caenorhabditis elegans* is highly divergent in both protein composition and sequence. Finally, we demonstrate a correlation between the presence of specific centriolar proteins and eye evolution. This correlation is used to predict proteins with functions in the development of ciliary, but not rhabdomeric, eyes.

Keywords: Centriole, Basal body, Ciliopathy, Cilia, Flagella

Abbreviations: BBS - Bardet-Biedl Syndrome; MKS - Meckel-Gruber Syndrome; NPHP – Nephronophthisis; BLAST – Basic Local Alignment Search Tool.

2.3 Introduction

Centrioles are highly conserved eukaryotic organelles consisting of nine microtubule triplets (or more rarely doublets or singlets) arranged in a radially symmetrical array. These structures are involved in a number of cellular functions, which include nucleating cilia/flagella – a context in which they are termed basal bodies (reviewed in Dawe et al., 2007a) – and organising pericentriolar material to form the centrosome (which is a microtubule organising centre in interphase cells and in some eukaryotes forms part of the spindle poles during cell division; see Delattre and Gonczy, 2004).

In recent years, experimental work on centrioles, centrosomes and basal bodies has elucidated a list of proteins that contribute to the formation of these structures (at least in the case of the individual model system studied). In addition to these proteins, several genes have been implicated in basal body or ciliary function by their association with human pathologies caused by defects in cilia, collectively named ciliopathies (reviewed in Badano et al., 2006). Despite these insights, it remains difficult to determine whether an individual protein has a role in centriolar, centrosomal or basal body function, and furthermore it cannot be assumed that a particular role in one model system extends to other model systems.

Here, we use a bioinformatic approach to determine the evolutionary history of the centriole from the perspective of its constituent proteins. We look at the conservation of multiple proteins that have either been physically localised to centrioles, centrosomes or basal bodies, or are genetically linked with basal body or ciliary function by association with ciliopathies. By ascertaining the distribution of these proteins in a wide cross-section of eukaryotes we are able to define a set of components that was present in the ancestral centriole. Linking this distribution to known organismal biology, we are also able to relate the loss of components to unusual centriolar morphologies and to predict a role for particular proteins in eye development.

2.4 Results and Discussion

2.4.1 Identification of centriolar proteins in diverse eukaryotes

To gain a deeper understanding of the biological role of known centriole- and ciliopathy-associated proteins, we analysed the phylogenetic presence and absence of these proteins across eukaryotes and linked the distribution pattern obtained to information about the protein and organism. We selected 45 organisms, each with complete or near-complete genome sequence, which represent a wide evolutionary spread across six major groups of eukaryotes (namely: Plantae, Excavata, Chromalveolata, Holozoa, Fungi and Amoebozoa). The selected organisms exhibit a rich diversity of microtubule biology and represent species that either produce cilia/flagella during their life-cycle or do not (29 ciliated species and 16 non-ciliated species; see section 2.9 Table 2-S1). We hypothesised that proteins which just have a centriolar and/or ciliary function would be present in cilia-forming species and absent in non-ciliated species.

We selected a set of 53 proteins that are known either for their centriolar, centrosomal or basal body localisation, or for their involvement in cilia-associated pathologies that present with retinitis pigmentosa (Table 2-S2, section 2.9). We used these protein sequences to interrogate the 45 predicted proteomes for orthologous sequences. We found that reciprocal best BLAST approaches - in which only proteins which are the best match in their respective genomes are considered orthologues - were generally too conservative in identifying orthologues, in part due to extensive gene duplications in particular lineages. In contrast simple BLAST searches (identifying all similar sequences) were too noisy and unable to discriminate between paralogues. To overcome these problems, we used an approach based on clustering of proteins by BLASTp score (Wickstead and Gull, 2007), with support from phylogenetic inference where necessary, as outlined in Materials and Methods.

To validate our approach, we searched for 3 proteins (γ-tubulin, EB1/Bim1p and XMAP215/ch-TOG) involved in microtubule dynamics that are expected to be widely present across the

eukaryotes. As expected, γ-tubulin was found in all organisms analysed. Importantly, we were not only able to identify tubulin homologues, but could unambiguously distinguish γ-tubulin from paralogous sequences (Fig. 2-1). The approach also performed well in identifying EB1 and XMAP215 orthologues; the lack of EB1 in *Leishmania major* has been noted before (Berriman et al., 2005) and as in a previous study we were able to validate the lack of XMAP215 using an iterative-HMM based approach (Devaux et al., 2007). These data indicate that our approach can identify homologous sequences in multiple organisms, and can also differentiate paralogous families.

2.4.2 Phylogenetic distribution of δ- and ε-tubulin

It has been demonstrated that δ- and ε-tubulin have a crucial centriolar role in *Chlamydomonas*, trypanosomes, humans and *Paramecium* (Chang and Stearns, 2000; Dupuis-Williams et al., 2002; Dutcher, 1998; Gadelha et al., 2006). These proteins are absent from non-ciliated species and also from *Drosophila melanogaster* and *C. elegans* (Chang and Stearns, 2000), both of which have cilia but build unusual centrioles (see below). Using our approach we were able to unambiguously identify δ-tubulin and ε-tubulin homologues in 25 and 26 of the 29 ciliated species, respectively. As predicted, we did not detect homologues in any non-ciliated species included in the analysis (Fig. 2-2). Consistent with previous findings (Chang and Stearns, 2000), no homologues of either centriolar tubulin were found in *D. melanogaster* or *C. elegans* (see Delattre et al., 2004). *D. melanogaster* builds centrioles based on either doublet or triplet microtubules (depending on the cell type; Callaini et al., 1997 whereas *C. elegans* centrioles are composed of singlet microtubules and lack the cartwheel structure found at the proximal end of most centrioles (Perkins et al., 1986b). Interestingly, the lack of both δ- and ε-tubulin genes in the *D. melanogaster* genome is a feature shared by two other dipteran genera – *Glossina* and *Anopheles* – and also by the silkworm *Bombyx mori*. In contrast, the honeybee *Apis mellifera* and flour beetle *Tribolium castaneum* genomes contain both tubulins (Fig. 2-2). These observations suggest that the loss of these tubulin forms is a specific feature of the Panorpoida (including

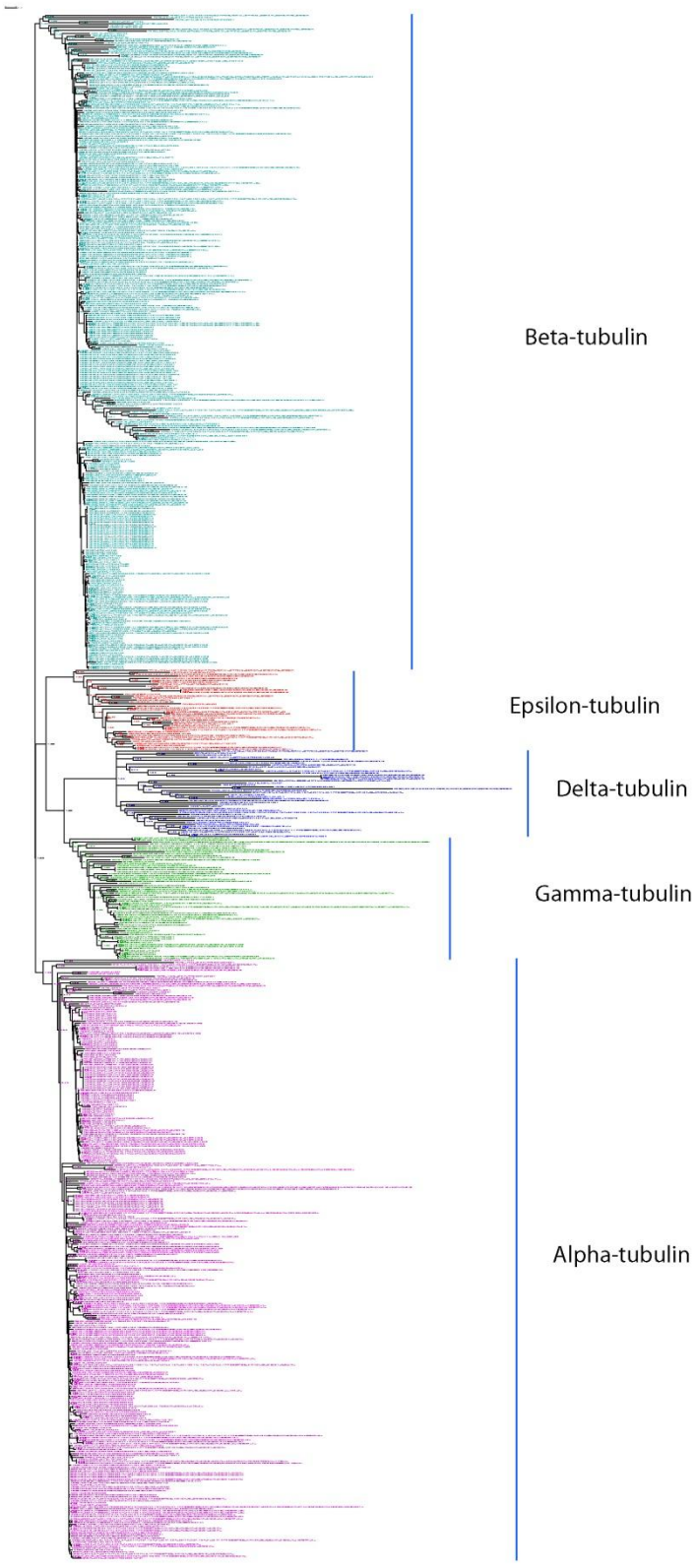


Figure 2-1: Tubulin phylogeny

An overview illustration of a BLASTp cluster depicting the different tubulin paralogues α-, β- δ-, ε-, and γ-tubulin intending to show that the cluster methodology can distinguish between these related proteins (specific protein ID is not intended to be legible).

Diptera and Lepidoptera) as opposed to being common to all insects. The distribution pattern of δ- and ε-tubulin is suggestive of a distinct functional module – in that the 2 proteins are almost always either absent or present from a given organism. The presence of ε-tubulin but no identifiable δ-tubulin in the stramenopile alga *Aureococcus anophagefferens* may be an artefact of the incomplete status of available genome sequence. However, along with *C. elegans* and Panorpoida, no evidence of either δ- or ε-tubulin was found in the ciliated diatom *Thalassiosira pseudonana*. The precise ultrastructure of the basal bodies formed in this species is unclear, however, other centric diatoms are known to produce basal bodies with only doublet microtubules (Jensen et al., 2003). These observations emphasise the association of δ- and ε-tubulin paralogues with a canonical triplet microtubule formation pathway for the centriole, as is seen from functional data on these proteins. However, the lack of δ- and ε-tubulin in dipteran genera shows that pathways for triplet formation have also evolved which are independent of these tubulins.

2.4.3 Ancestral centriolar core components

The ninefold triplet microtubule arrangement of centrioles is widely conserved across eukaryotes (reviewed in Beisson and Wright, 2003). However, it is not clear to what extent this ultrastructural consistency reflects the conservation of underlying molecular components. To investigate this question, we analysed the phylogenetic distribution of known centriolar proteins in our dataset. This analysis identified a set of 14 proteins with proven centriolar localisation that are present in at least four major eukaryotic groups (Fig. 2-2). Given the current understanding of eukaryotic evolution, the most likely explanation for this pattern is that these 14 centriolar proteins were present in the ancestor of all extant eukaryotes (the cenancestor), and that this core set performs a widely conserved function. However, no single component of the set is ubiquitously conserved in organisms that possess centrioles/basal bodies. Furthermore, extant species vary considerably in which components of the set are present – indicating a surprising plasticity in the protein composition of centrioles.

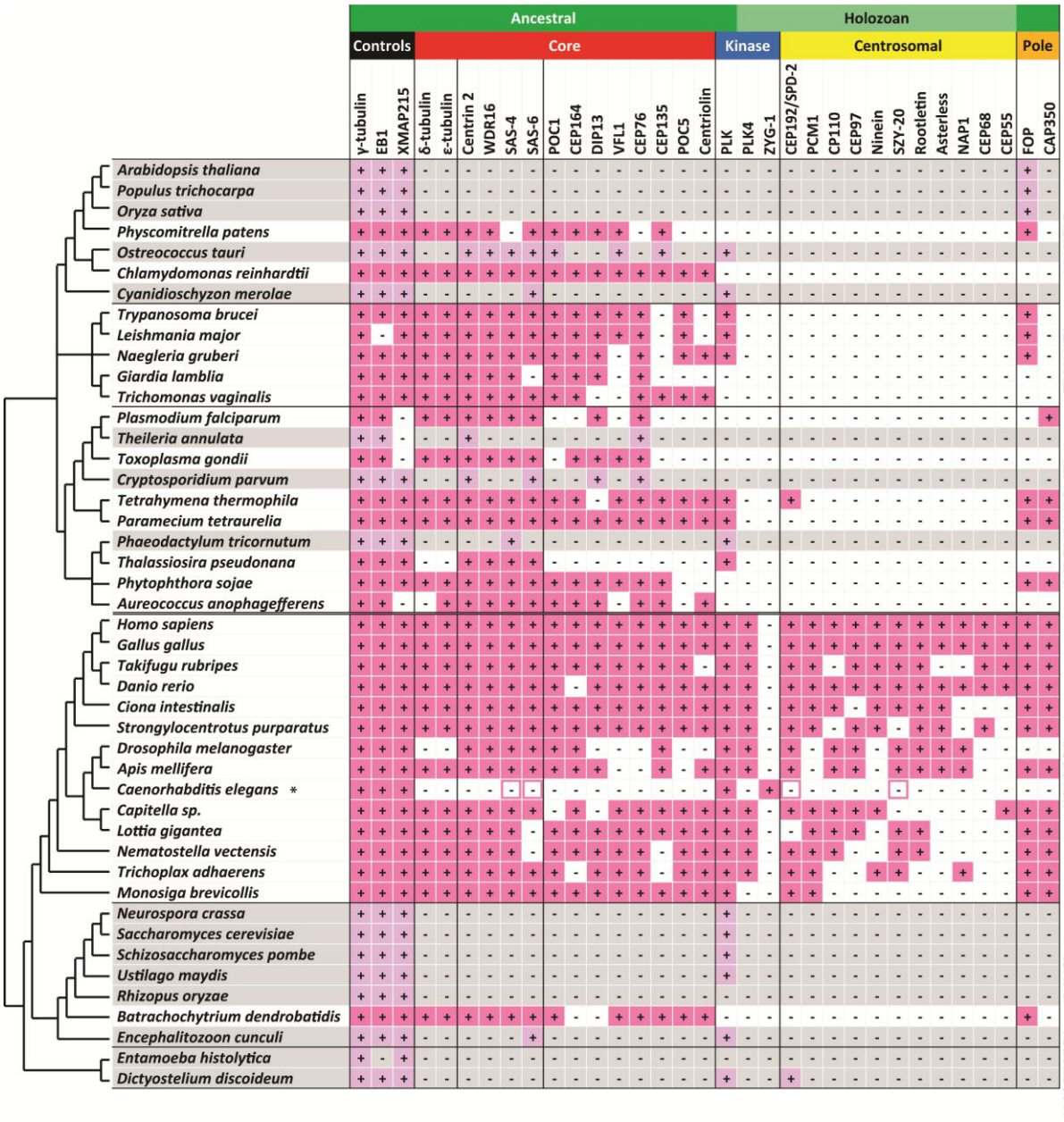


Figure 2-2: Distribution of centriolar/centrosomal proteins across eukaryotes

Protein homologues were identified in 45 eukaryotic genomes, including 29 ciliated species and 16 non-ciliated species (grey). Presence of homologue(s) is indicated by +. "Core" proteins are conserved ancestral centriolar proteins. "Centrosomal" proteins are associated with centrosomal functions. "Pole" proteins might have fulfilled a function in the ancestral spindle pole. "Controls" are proteins that are associated with general microtubule dynamics. "Ancestral" proteins are present across extant eukaryotes. "Holozoan" proteins have a restricted presence in Holozoa (Metazoa and *M. brevicollis*). Asterisk indicates sequence drift of core and centrosomal proteins in *C. elegans*; divergent homologues known in the literature but not identified by our approach are highlighted with a pink border.

In our analysis, 4 proteins (centrin 2, WDR16, SAS-4 and SAS-6) were found to be more consistently associated with ciliated species than ε-tubulin (Fig. 2-2). SAS-4 has been proposed to be involved in the attachment of microtubules to the central centriolar cylinder (Pelletier et al., 2006) and in the elongation of centriolar microtubules (Kohlmaier et al., 2009; Schmidt et al., 2009; Tang et al., 2009) and SAS-6 in cartwheel formation (Kilburn et al., 2007; Nakazawa et al., 2007; Rodrigues-Martins et al., 2007). Centrin 2 is a well characterised centriolar component which is essential for centriole duplication (Geimer and Melkonian, 2005; Koblenz et al., 2003; Salisbury et al., 2002). While the wider centrin family is involved in multiple cellular processes and is present in non-ciliated species, there is a clade termed CrCen-type/centrin 2 which has a distribution restricted to ciliated species (Bornens and Azimzadeh, 2007). In contrast, WDR16 is a protein of unknown function associated with hydrocephalus in zebrafish (Verleysdonk et al., 2007) and localisation to the basal body, pro-basal body, and to a lesser extent the flagellum of trypanosomes (H. Farr and K. Gull, unpublished). Based on the observed distribution pattern, it is likely that WDR16 also plays a key role in basal body and/or centriole function.

Interestingly, although the green alga *Ostreococcus tauri* is thought to be non-flagellate – and the whole-cell reconstruction of the vegetative stage is acentriolar (Henderson et al., 2007) – we found that 7 of the 14 core set proteins are present in this species (i.e. more than in several organisms that definitively possess centrioles). Similarly, the stramenopile alga *Aureococcus anophagefferens* is also non-flagellate and acentriolar in the life-cycle stages thus far analysed (Sieburth et al., 1988), but possesses 11 of the 14 centriolar core proteins. Due to the presence of genes encoding components of intraflagellar transport (Woodland and Fry, 2008) and flagellar apparatus (Elias and Archibald, 2009), it has previously been suggested that *A. anophagefferens* might possess a flagellate stage similar to that seen in at least some other pelagophytes. Based on our observations here, and on the presence of genes encoding inner arm dyneins (Wickstead and Gull, 2007), we predict that *O. tauri* also forms basal bodies and cilia at some point in its lifecycle – most likely in a yet to be observed gamete or zoospore.

Notably, none of the 14 core centriolar proteins could be identified in the predicted proteome of *C. elegans*. Clearly, comparative analyses cannot distinguish between proteins that have been lost entirely from an organism and those that have diverged in sequence to the extent that they are undetectable. Since SAS-4 and SAS-6 were originally identified in *C. elegans* (Dammermann et al., 2004; Kirkham et al., 2003; Leidel et al., 2005; Leidel and Gonczy, 2003b) these 2 at least are in the latter category. However, the sequences of SAS-4 and SAS-6 have changed more rapidly in *C. elegans* than in other lineages (Fig. 2-3). This suggests that the unusual centrioles of *C. elegans* evolved in concert with extensive loss or divergence of the core centriolar components. Thus, *C. elegans* appears to represent the extremes of centriole divergence, without having lost the structure completely. A similar but less extreme process appears to have occurred in *Thalassiosira pseudonana*, in which we can identify only the 4 most highly-conserved centriolar proteins (centrin 2, WDR16, SAS-4 and SAS-6). Such divergence must be considered when interpreting centriolar phenotypes or using particular organisms for comparative biology.

2.4.4 Centriole-associated kinases: origin and diversification

Kinase activity is critical for the regulation of centriole duplication. The kinases ZYG-1 in *C. elegans*, SAK in *D. melanogaster*, and PLK4 in humans are all essential for centriole duplication (Bettencourt-Dias et al., 2005; Habedanck et al., 2005; O'Connell et al., 2001). Catalytically inactive mutants of PLK4 fail to induce centriole biogenesis (Habedanck et al., 2005), and cells in SAK mutant embryos lack centrioles (Bettencourt-Dias et al., 2005). It has been suggested that PLK4, SAK and ZYG-1 are orthologous (Bettencourt-Dias et al., 2005).

To resolve the distribution and relationships between the centriole-associated kinases, we performed phylogenetic analysis of all Polo-like kinases, plus SAK and ZYG-1 homologues (Fig. 2-4). Our phylogenetic inference unambiguously identifies a well-supported Polo-like kinase clan in the majority of organisms in our analysis. This group includes the PLK1-4 isoforms, where they exist. Within the PLK family, PLK4 and SAK form a clear subgroup and are orthologous, as

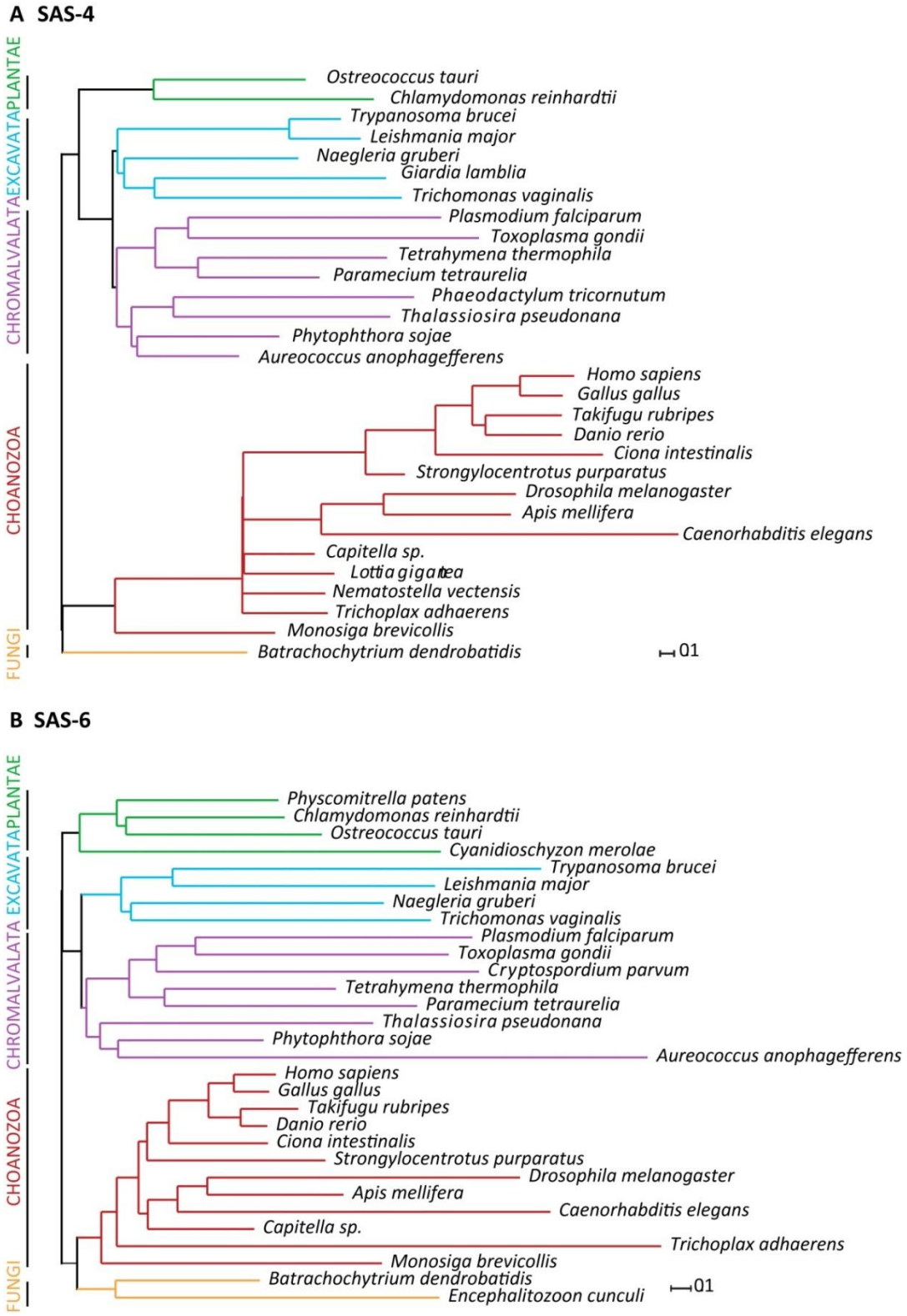


Figure 2-3: Constrained phylogeny of SAS-4 and SAS-6

Bayesian trees with fixed topology, representing the current understanding of eukaryotic evolution, were produced for identified SAS-4 (A) / SAS-6 (B) homologues including the *C. elegans* SAS-4/-6 sequences known from the literature.

expected. However, we find no evidence to support the grouping of ZYG-1 in the PLK4/SAK clan (Fig. 2-4). Moreover, despite experimental evidence suggesting that ZYG-1 has analogous function to PLK4 (Bettencourt-Dias et al., 2005; Habedanck et al., 2005; O'Connell et al., 2001), sequence analysis does not place ZYG-1 within the PLK clan (note that *C. elegans* PLK1 sequences were readily identified). Since ZYG-1 cannot be placed in any kinase family, it is unclear whether ZYG-1 is a PLK4 that has undergone rapid divergence or whether it is an entirely different kinase that has replaced PLK4 function. Regardless, the orphan nature of ZYG-1 is a further molecular demonstration of the unusual nature of *C. elegans* centrioles.

The pattern of PLK distribution shows that PLK was present in the cenancestor of eukaryotes wherein it may have had a function that included initiation of centriole duplication. However, we show that in common with much of the ancestral centriolar core, PLK has been lost multiple times. We found that neither PLK nor ZYG-1 were detectable in 9 of the 29 analysed ciliated species, suggesting either that centriolar duplication does not require kinase activity in these species, or that other kinases fulfil that role.

2.4.5 Basal body function is ancestral, but centrosomal components are not

In our analysis we identified proteins that are found only in Holozoa (represented in our analysis by the metazoa and the choanoflagellate *Monosiga brevicollis*; Fig. 2-2). In general these proteins are either associated with centrosomal functions or have been localised to the pericentriolar material within the centrosome. This restricted phylogenetic distribution has two alternative explanations: either many of the components of the animal centrosome were acquired only by the ancestor of the Holozoa; or centrosomal components, in contrast to those of the basal body, are evolving at sufficient speed to render them undetectable in distantly-related organisms. In keeping with the latter proposition the proteins SPD-2 and SZY-20 – which were originally identified in *C. elegans* (Kemp et al., 2004; Pelletier et al., 2004; Song et al., 2008) – could not be detected in *C. elegans* by our approach when using canonical homologues from other holozoa.

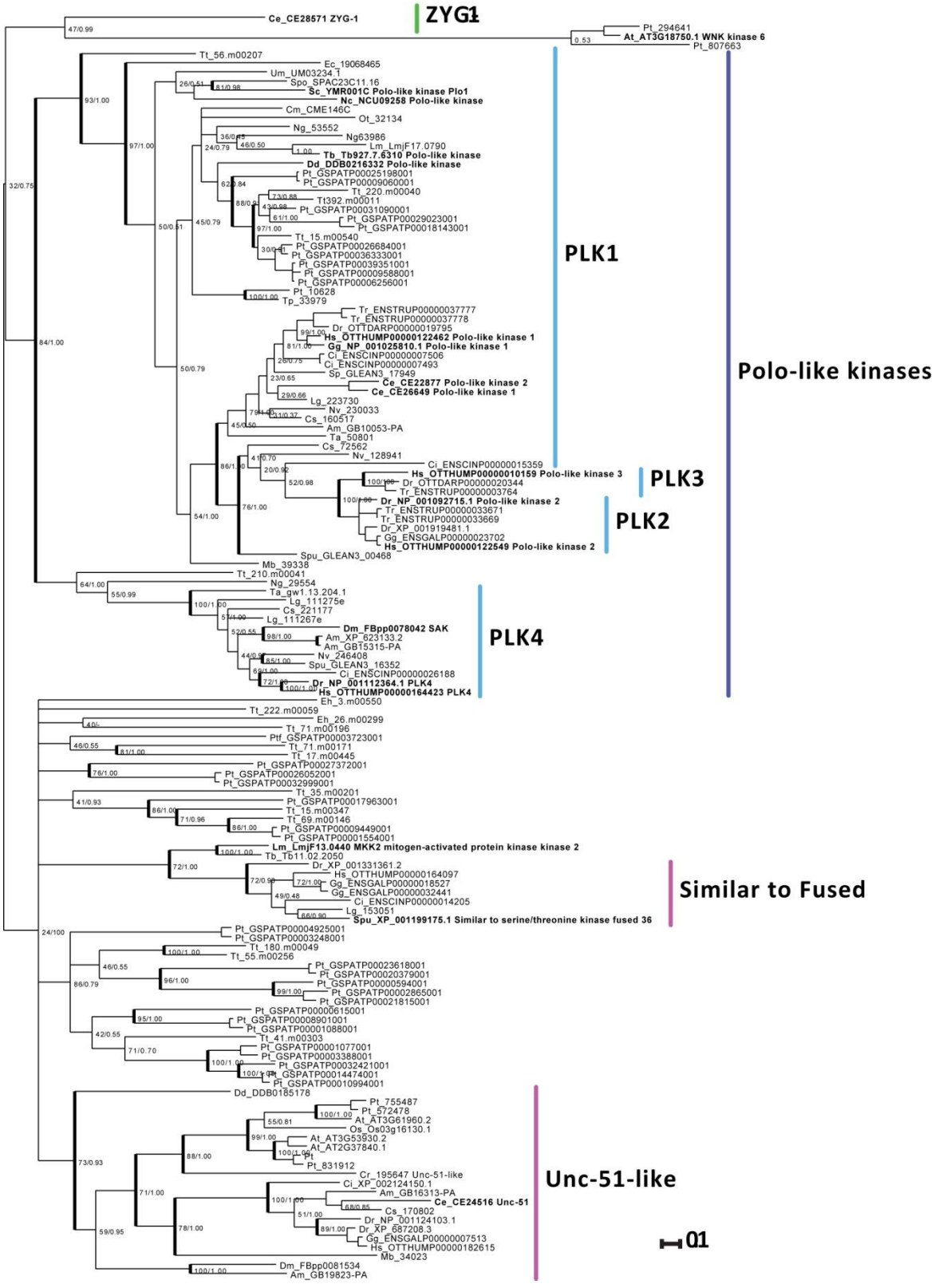


Figure 2-4: PLK superfamily

Maximum likelihood and Bayesian trees were produced for PLKs (pink) including PLK4 (blue), and ZYG-1 (green). Topology support for nodes is indicated (maximum likelihood bootstrap values/posterior probability). Bootstrap values give a conservative estimate of the confidence that a particular group of sequences are monophyletic.

Thus it appears that the extreme divergence of centriolar proteins in *C. elegans* (see above), occurred concomitantly with the divergence of centrosomal components. However, unlike centriolar components, only 2 of the 13 centrosome-associated proteins analysed (FOP and CAP350) were detectable outside of the Holozoa (Fig.2-2). The distribution of these 2 components does not correlate with organisms in which there is a centriole, which would be expected if these proteins function only in the centrosome. This implies that the animal centrosome is constructed from mainly holozoan-specific components.

The evolutionary relationship between the centriole and centrosomal material (with functions in the organisation of both cytoplasmic and spindle microtubules) is not as clear as it may at first sight appear. One possibility is that a proto-centrosome with a role in microtubule organisation existed in an ancestral cell that also possessed separate basal bodies subtending cilia. Basal bodies became centrioles when they began "piggy-backing" onto this proto-centrosome as a means of ensuring fidelity in the inheritance of cilia (Pickett-Heaps, 1971; see also Beisson and Wright, 2003; Bornens and Azimzadeh, 2007). Alternatively, centrioles with a function in cell cycle control were present before centrosomes, and material associated with microtubule organisation which can now be found in the centrosome was accrued gradually by these stably-inherited centrioles. Importantly, if the centrosome as an organelle already existed in the last common ancestor of eukaryotes, then all centrosomes are homologous. Conversely, if different material has been accumulated by centrioles in different lineages then these structures are non-homologous. With the caveat noted above regarding rates of divergence, our finding that much of the animal centrosome is specific to Holozoa supports the latter evolutionary scenario, and suggests that the animal centrosome is a holozoan innovation. This would imply that an association of centrioles with spindle poles has arisen independently in more than one lineage (minimally, in the “unikonts” and Chlorophyta; Coss, 1974).

2.4.6 A sensory role for the ancestral cilium

Defects in centrioles and cilia cause a variety of human disorders, collectively known as ciliopathies (for review see Badano et al., 2006). To assess whether proteins associated with ciliopathies are conserved across eukaryotes we investigated their phylogenetic distribution. The ciliopathy Bardet-Biedl Syndrome (BBS) is linked to mutations in *BBS1* to *14* (see Jin and Nachury, 2009) and BBS1, 2, 4, 5, 7, 8, and 9 proteins act as a biochemical complex termed the "BBSome" (Nachury et al., 2007). The BBSome complex is conserved in a modular fashion in that the 7 proteins are generally present or absent as a group (Fig. 2-5). Based on the phylogenetic distribution of this group, we propose that the cenancestor of eukaryotes possessed a BBSome.

The BBSome is proposed to play a role in transporting membrane proteins to sensory (primary) as opposed to motile cilia (reviewed in Singla and Reiter, 2006). This proposal is supported by the fact that the complex is conserved in *C. elegans*, an organism that only builds immotile sensory primary cilia. The absence of the BBSome in 4 species that produce only motile cilia in gametes/zoospores (namely *Physcomitrella patens*, *Plasmodium falciparum*, *Thalassiosira pseudonana* and *Batrachochytrium dendrobatidis;* Armbrust, 1990; Berger et al., 2005; Merchant et al., 2007a; Sinden et al., 1978) further supports the role of the BBSome in sensory function and its dispensability for motility. In combination, these observations provide molecular evidence that the cenancestral cilium served a sensory function.

Another important ciliopathy, Meckel-Gruber Syndrome (MKS), is characterised by mutations in *MKS1, 3, 4, 5* and *6* (see Dawe et al., 2007b). Our data demonstrate that as with the BBSome, MKS proteins were in the cenancestor of eukaryotes and their extant distribution correlates with the possession of cilia. However, BBSome and MKS proteins have distinct, albeit overlapping, phylogenetic distribution patterns (for example, *B. dendrobatidis* and *T. pseudonana* contain MKS but not BBS proteins). This observation suggests that these two ciliopathy-associated modules can function independently.

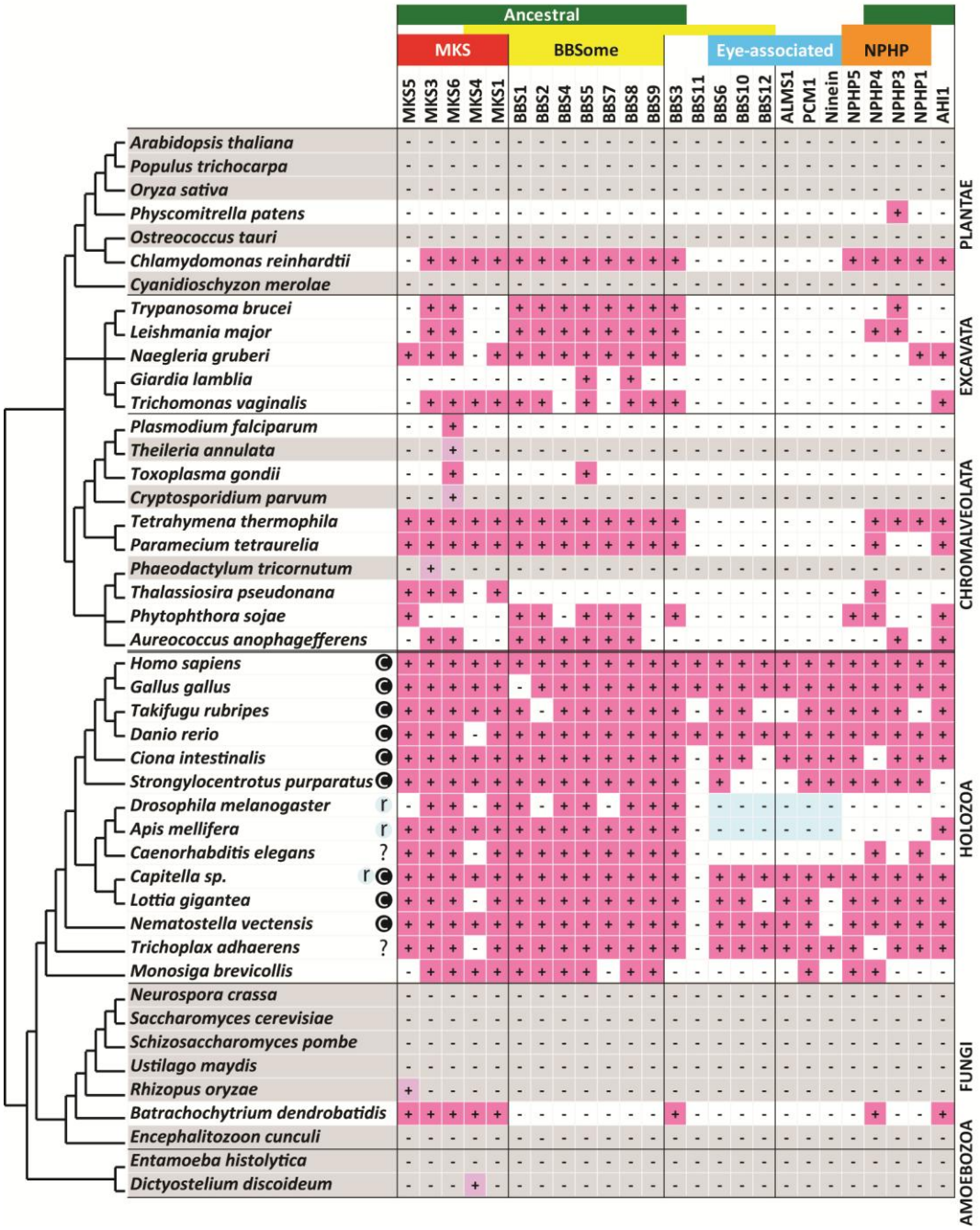


Figure 2-5: Distribution of ciliopathy-associated proteins among eukaryotes

Protein homologues were identified in 45 eukaryotic genomes, including 29 ciliated species and 16 non-ciliated species (grey). Presence of homologue(s) is indicated by +. The proteins are grouped according to the ciliopathy or/and ciliary photoreceptor association: MKS = Meckel-Gruber Syndrome; BBS = Bardet-Biedl Syndrome; NPHP = Nephronophthisis, "Eye-associated" proteins show a distribution that correlates with ciliary "c" photoreceptor (light blue shade = exclusive presence of rhabdomeric "r" photoreceptor).

2.4.7 Correlation between ciliary eye evolution and ciliopathy-associated proteins

Many patients presenting with ciliopathies manifest the eye disease retinitis pigmentosa because human eye development depends on ciliary function in photoreceptor cells. Animal photoreceptor cells can be classified as either rhabdomeric or ciliary depending on how the membranes for photopigment storage are extended. In rhabdomeric photoreceptor cells, extension is achieved through folding of the apical cell surface into actin-based microvilli, whereas in ciliary photoreceptor cells the ciliary membrane is folded (reviewed in Arendt, 2003; Eakin, 1982). The ancestral proto-photoreceptor cell likely used the ciliary mechanism and possessed ciliary-specific opsins (reviewed in Shubin et al., 2009). It has been hypothesised that the last common ancestor of the Bilateria (animals excluding *Nematostella vectensis* and *Trichoplax adherens* in our analysis) had additionally acquired rhabdomeric photoreceptor cells, and that loss of one or the other system has occurred in different lineages (see Arendt and Wittbrodt, 2001). Structural studies have shown that insects employ only rhabdomeric photoreceptor cells, whereas vertebrates and also *Ciona intestinalis*, *Strongylocentrotus purparatus*, and *Lottia gigantean* only use the ciliary system (Purschke et al., 2006). Both systems are still present in *Capitella* (Purschke et al., 2006). The phylogenetic distribution of rhabdomeric- and ciliary-specific opsins correlates with this structural analysis (Arendt et al., 2004). Proteins involved in ciliary photoreceptor development are therefore expected to have a specific pattern in our analysis: present in vertebrates, *C. intesinalis*, *S. purparatus*, *Capitella*, *L. gigantea*, *N. vectensis* and *T. adherens*, but absent in *D. melanogaster*, and *Apis mellifera*.

Four out of 23 ciliopathy-associated proteins in our analysis – BBS6, BBS10, BBS12 and ALMS-1 (Fig. 2-5) – have a profile predictive of involvement in ciliary photoreceptor cells. In contrast to the BBSome, BBS6, BBS10 and BBS12 are predicted to be chaperonin-like proteins (Stoetzel et al., 2007), and none of the 4 are ancestral. Alongside these proteins, PCM1 and Ninein also share the ciliary photoreceptor fingerprint. Interestingly, PCM1 is already functionally linked to ciliogenesis because it interacts with BBS4 (Nachury et al., 2007). Our findings now predict an additional

association between Ninein and the eye. NPHP1, 3, 4 and 5 similarly display a paraphyletic profile for eye-association in Metazoa, but these 4 proteins are also found outside this lineage, suggesting a more general function. The absence of any of these eye-associated proteins in the *C. elegans* genome also correlates with a lack of ciliary opsins (Satoh, 2006), and provides strong genomic data to support the prediction that the photoreceptor spots in this species have a rhabdomeric-like structure (McLaren, 1976).

2.5 Conclusions

Our analyses provide molecular evidence to suggest that the cenancestor of eukaryotes possessed a centriole that had basal body function, but had no association with the centrosome. Moreover, this cenancestor also possessed the BBSome complex – and hence was using flagella/cilia not only for motility but also for sensory functions. The ancestral centriolar structure contained a cohort of proteins that is still present in the centrioles of most extant ciliated eukaryotes. Interestingly, the morphology of the centriole is surprisingly insensitive to multiple losses from this cohort, in that gross changes in the ultrastructure of the centriole occur only when the majority of the cohort is lost or highly divergent, as demonstrated in the unusual centrioles of *C. elegans*. Similarly, kinases that are essential for the initiation of centriole assembly in some organisms are not ubiquitously conserved. Finally, in light of the evolutionary footprint of ciliary photoreceptor cells, we identify 6 proteins that we predict to function in ciliary eye cell development, but not in the eyes of insects. Collectively, these data provide a more coherent picture of the evolution of centriolar proteins.

2.6 Materials and Methods

2.6.1 Definition of datasets

An initial set of 53 proteins was selected on the basis of fulfilling one of two criteria: 1) known localisation to the centriole or basal body and/or centrosome; or 2) genetic evidence for an involvement in cilia-associated pathologies. Sequences of 3 proteins with an involvement in

microtubule dynamics that are predicted to be widely distributed across eukaryotes - γ-tubulin, XMAP215/ch-TOG and EB1/Bim1p - were also included as controls. In all cases, query sequences were human homologues of the proteins, with the exception of *Caenorhabditis elegans* ZYG-1 (see Results). A complete list of these proteins with references is given in Table 2-S2 (see section 2.9). These protein sequences were used to query the predicted proteomes of 45 eukaryotic organisms for which complete or near-complete genome sequence is publicly available. These organisms were chosen to represent a wide evolutionary spread of extant eukaryotes with diverse microtubule biology. A list of genomic dataset sources and versions used is provided in Table 2-S1 (see section 2.9).

2.6.2 Identification of homologous sets

During initial work, we found that reciprocal best BLAST approaches were generally too conservative when identifying orthologues, whereas simple BLAST searches (Altschul et al., 1990) were too noisy and unable to discriminate between paralogues. To overcome these problems, we used an approach based on clustering of proteins by BLASTp score (Wickstead and Gull, 2007). Briefly, a "seed" set was generated for each query from a reciprocal best BLASTp search with an e-value threshold of $<10^{-5}$. A liberal set of putative homologues was then generated on the basis of all BLASTp hits to any seed sequence with an e-value $<10^{-2}$. Results from this search were then formed into clusters using a distance matrix derived from BLASTp-scores, as described in (Wickstead and Gull, 2007), with the orthologous cluster being defined by eye (Table 2-S3, see section 2.9). In the case of ambiguous clustering, sets were aligned, trimmed and then used in Bayesian and Maximum Likelihood phylogenetic inference (Guindon and Gascuel, 2003; Ronquist and Huelsenbeck, 2003). For phylogenetic inference, amino-acid sequence alignments were produced using MAFFT, and trimmed to remove poorly aligned positions. MrBayes v3.1.2 was used to infer Bayesian trees (random start tree, 4 chains of temperature 0.2 (default setting), Whelan and Goldman (WAG) substitution matrices, 4 discrete categories of gamma distribution substitution rate, 250,000 generations). Tree likelihoods appeared to reach stationary phase at

around 60,000 generations, and the last 190,000 generations were used to construct the consensus tree. Fixed topology trees using an existing model for species relationships (see text) were achieved through setting the proposal probability of all topology moves to 0. Maximum likelihood bootstrap values were produced using PHYMLv2.4.4 (100 replicates, WAG matrix, 4 gamma categories, alpha parameter-estimation for each replicate). Bootstrap values give a conservative estimate of the confidence that a particular group of sequences are monophyletic. Generally, groups with >90% bootstrap support were considered to be well supported and those with >70% bootstrap support to have some support. Species names were abbreviated as follows: Ap = *Apis mellifera*, Ci = *Ciona intestinalis*, Ce = *Caenorhabditis elegans*, Cm = *Cyanidioschyzon merolae*, Cs = *Capitella* sp., Dd = *Dictyostelium discoideum*, Dm = *Drosophila melanogaster*, Dr = *Danio rerio*, Ec = *Encephalitozoon cunculi*, Gg = *Gallus gallus*, Hs = *Homo sapiens*, Lg = *Lottia gigantea,* Lm = *Leishmania major*, Mb = *Monosiga brevicollis*, Nc = *Neurospora crassa*, Ng = *Naegleria gruberi*, Nv = *Nematostella vectensis*, Ot = *Ostreococcus tauri*, Pt = *Paramecium tetraurelia*, Sc = *Saccharomyces cerevisiae*, Spo = *Schizosaccharomyces pombe*, Spu = *Strongylocentrotus purparatus*, Ta = *Trichoplax adhaerens*, Tb = *Trypanosoma brucei*, Th = *Tetrahymena thermophila*, Tr = *Takifugu rubripes*, Um = *Ustilago maydis.*

2.7 Acknowledgements

Thanks to Monica Bettencourt-Dias (Instituto Gulbenkian de Ciência, Portugal) for useful discussion and sharing of prepublication data. We thank Helen Farr for communicating results on WDR16 before publication. We are grateful to Steven Kelly and Neil Portman for critical reading of the manuscript and to Helen Dawe for information on ciliopathies. This work was funded by grants from the Wellcome Trust to KG and the Gatsby Charitable foundation to JAL. NS and MEH are supported by graduate studentships from the EPA Trust and BBSRC, respectively. KG is a Wellcome Trust Principal Research Fellow.

2.8 Additional Information

2.8.1 Table S1: Source and version of the genomes used in the analysis

Species	Source	Version	Web reference
Apis mellifera	beebase	v2 (pre)	racerx00.tamu.edu
Arabidopsis thaliana	TAIR	TAIR7	www.arabidopsis.org/
Aspergillus nidulans	Broad Institute	v1.0	www.broad.mit.edu/annotation/fgi/
Aureococcus anophagefferens	JGI	v1.0	www.jgi.doe.gov/
Batrachochytrium dendrobatidis	Broad Institute	v1.0	www.broad.mit.edu/annotation/fgi/
Caenorhabdatis elegans	WormBase	WS170	www.wormbase.org/
Capitella sp.	JGI	v1.0	www.jgi.doe.gov/
Chlamydomonas reinhardtii	JGI	v3.1	www.jgi.doe.gov/
Ciona intestinalis	JGI	v2.0	www.jgi.doe.gov/
Cryptosporidium parvum	CryptoDB	v3.4	www.cryptodb.org/
Cyanidioschyzon merolae	C.merolae genome project	v1	merolae.biol.s.u-tokyo.ac.jp/
Danio rerio	VEGA	May 2007	vega.sanger.ac.uk/
Dictyostelium discoideum	dictyBase	05-20-2007	dictybase.org/
Drosophila melanogaster	ENSEMBL	BDGP4.3, 44.43b	www.ebi.ac.uk/ensembl/
Encephalitozoon cuniculi	NCBI	1	www.ncbi.nlm.nih.gov
Entamoeba histolytica	geneDB	17102005	www.genedb.org/
Gallus gallus	ENSEMBL	WASHUC2, 44.2b	www.ebi.ac.uk/ensembl/
Giardia lamblia	GiardiaDB	v1.0	www.giardiadb.org/
Homo sapiens	VEGA	Jun-07	vega.sanger.ac.uk/
Leishmania major	geneDB	v5.2	www.genedb.org/
Lottia gigantea	JGI	v1.0	www.jgi.doe.gov/
Monosiga brevicollis	JGI	v1.0	www.jgi.doe.gov/
Naegleria gruberi	JGI	v1.0	www.jgi.doe.gov/
Nematostella vectensis	JGI	v1.0	www.jgi.doe.gov/
Neurospora crassa	Broad Institute	v7.0	www.broad.mit.edu/annotation/fgi/
Oryza sativa	TIGR	v5.0	www.tigr.org/tdb/e2k1/osa1/
Ostreococcus tauri	JGI	v2.0	www.jgi.doe.gov/
Paramecium tetraurelia	ParameciumDB	v1.04	paramecium.cgm.cnrs-gif.fr/
Phaeodactylum tricornutum	JGI	v2.0	www.jgi.doe.gov/
Physcomitrella patens	JGI	v1.1	www.jgi.doe.gov/
Phytophthora sojae	JGI	v1.1	www.jgi.doe.gov/
Plasmodium falciparum	PlasmoDB	v5.2	www.plasmodb.org/
Populus trichocarpa	JGI	v1.1	www.jgi.doe.gov/
Rhizopus oryzae	Broad Institute	v3.0	www.broad.mit.edu/annotation/fgi/
Saccharomyces cerevisiae	Saccharomyces genome database	20070713	www.yeastgenome.org/
Schistosoma mansoni	geneDB	v4.0	www.genedb.org/

Schizosaccharomyces pombe	Sanger	v19	www.sanger.ac.uk/Projects/S_pombe
Strongylocentrotus purpuratus	Baylor College of Medicine	Spur_2.1, v4	www.hgsc.bcm.tmc.edu/
Takifugu rubripes	ENSEMBL	Assembly4, 44.4e	www.ebi.ac.uk/ensembl/
Tetrahymena thermophila	TIGR	10/24/2006	www.tigr.org/tdb/e2k1/tta1/
Thalassiosira pseudonana	JGI	v3.0	www.jgi.doe.gov/
Theileria annulata	GeneDB	v4	www.genedb.org/
Toxoplasma gondii	ToxoDB	v4.1	www.toxodb.org/
Trichomonas vaginalis	TIGR	20050331	www.tigr.org/tdb/e2k1/tvg/
Trichoplax adhaerens	JGI	v1.0	www.jgi.doe.gov/
Trypanosoma brucei	geneDB	v4	www.genedb.org/
Ustilago maydis	Broad Institute	v1.0	www.broad.mit.edu/annotation/fgi/

2.8.2 Table S2: Information about proteins used in this analysis

Protein	Localisation	Additional Information	References
ALMS1	Centrosome, basal body	Mutated in Alström Syndrome	(Hearn et al., 2005)
Asterless	Centrosome	Human centrosome proteome; required for centriole duplication in *Drosophila*	(Andersen et al., 2003; Varmark et al., 2007; Blachon et al., 2008; Dobbelaere et al., 2008)
BBS1		Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; mouse model with *BBS1M309R/M309R* (most common mutation in humans) has elongated cilia with swollen distal ends	(Davis et al., 2007; Nachury et al., 2007)
BBS2	Base of cilia	Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; not required for global ciliogenesis but *BBS2-/-*mice have cilia with swollen distal ends	(Nachury et al., 2007; Shah et al., 2008)
BBS3		Mutated in Bardet-Biedl Syndrome; ARF-like GTPase; in *C. elegans*: expressed in ciliated neurons; undergoes IFT	(Fan et al., 2004)
BBS4	Base of cilia	Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; *BBS4-/-* mouse model shows BBS4 is not required for global but spermatozoa ciliogenesis; BBS-4 interacts with PCM1	(Mykytyn et al., 2004; Nachury et al., 2007)
BBS5		Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; BBS5 has 2 PH domains which interact with Phosphatidylinositol 3phosphate	(Nachury et al., 2007)
BBS6	Base of cilia (ring-shape), midbody	Mutated in Bardet-Biedl Syndrome; Type II Chaperonin-like protein	(Kim et al., 2005; Stoetzel et al., 2007)
BBS7	Base of cilia and cilium	Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; In *C. elegans*: BBS7 and BBS8 undergo IFT and are involved in stabilising IFT particle	(Blacque et al., 2004; Ou et al., 2005; Nachury et al., 2007)
BBS8	Base of cilia and cilium	Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; In *C. elegans*: BBS7 and BBS8 undergo IFT and are involved in stabilising IFT particle	(Blacque et al., 2004; Ou et al., 2005; Nachury et al., 2007)
BBS9		Mutated in Bardet-Biedl Syndrome; component of the biochemical complex BBSome which is important for ciliogenesis; BBS9 interacts with BBS8, 5, 4, 2 and 1	(Nachury et al., 2007)
BBS10	Base of cilia	Mutated in Bardet-Biedl Syndrome; involved in transient ciliogenesis in adipocytes; Type II Chaperonin-like protein	(Stoetzel et al., 2007; Marion et al., 2009)
BBS11		Mutated in Bardet-Biedl Syndrome; E3 ubiquitin ligase	(Chiang et al., 2006)
BBS12	Base of cilia	Mutated in Bardet-Biedl Syndrome; involved in transient ciliogenesis in adipocytes; Type II Chaperonin-like protein	(Stoetzel et al., 2007; Marion et al., 2009)
BBS13 /MKS1	Base of cilia	Mutated in Bardet-Biedl Syndrome and Meckel-Gruber Syndrome; depletion results in block of centriole migration to apical membrane; interacts with MKS3	(Dawe et al., 2007)
BBS14 /MKS4	Centrosome, base of cilia/centriolar satellites	Mutated in Bardet-Biedl Syndrome, Meckel-Gruber Syndrome, Joubert Syndrome, and Nephronophthisis/Senoir-Loken Syndrome; positive regulator of ciliogenesis; CEP290 depletion prevents ciliogenesis; interacts with RPGR	(Chang et al., 2006; Sayer et al., 2006; McEwen et al., 2007; Kim, J. et al., 2008; Tsang et al., 2008)
CAP350	Centrosome	Human centrosome proteome; MT anchoring centrosome with FOP and EB1 and Ninein, p150glued; FOP seems required for CAP350 localisation to the mitotic spindle	(Andersen et al., 2003; Yan et al., 2006)
Centrin	Centriole (distal end)	Required for centriole duplication in humans, Chlamydomonas, Tetrahymena; often multiple paralogues per species involved into many cellular processes	(Salisbury et al., 2002; Koblenz et al., 2003; Stemm-Wolf et al., 2005)
Centriolin	Centriole (subdistal appendages), midbody	Depletion and overexpression cause cytokinesis failure; anchoring of Exocyst and SNARE complexes at the midbody, required for secretory-vesicle-mediated abscission	(Gromley et al., 2003; Gromley et al., 2005)
CEP55	Centrosome, midbody	Human centrosome proteome; recruitment of ESCRT-III/ALIX to the midbody	(Andersen et al., 2003; Morita et al., 2007; Carlton et al., 2008)
CEP68	Centrosome	Human centrosome proteome; centrosome cohesion	(Andersen et al., 2003; Graser et al., 2007a)
CEP76	Centriole (mature and pro-centriole)	Human centrosome proteome; in *T. brucei* flagella proteome; interacts with CP110; depletion of CEP76 results in accumulation of centriolar intermediates (in some human cell lines); overexpression does not affects normal centriole duplication	(Andersen et al., 2003; Broadhead et al., 2006; Tsang et al., 2009)
CEP97	Centrosome	Human centrosome proteome; CP110 and CEP97 suppress cilia assembly; CEP97 recruits CP110 to the centrosome, and vice versa	(Andersen et al., 2003; Spektor et al., 2007)
CEP135 (BLD10)	Centriole (cartwheel)	Human centrosome proteome; *Chlamydomonas CEP135/BLD10* null mutant has no basal bodies but is viable; expression of truncated CEP135/BLD10 results in shorter cartwheel spokes (reduced centriole diameter with less MT triplets); depletion of CEP135/BLD10 in Drosophila induces shorter basal bodies leading to immotile sperm but not disturbed sensory function; CEP135/BLD10 interacts with C-NAP1 which is involved in linking the disengaged centrioles until centrosome splitting	(Andersen et al., 2003; Matsuura et al., 2004; Hiraki et al., 2007; Kim, K. et al., 2008; Mottier-Pavie and Megraw, 2009)
CEP164	Centriole (mature, distal appendages, ring-shape)	Human centrosome proteome; essential for primary cilia formation; not essential for centriole duplication in human cells	(Otto et al., 2002; Graser et al., 2007b)
CEP192/SPD -2	Centrosome	Human centrosome proteome; recruitment of PCM to the centriole; in *C. elegans* embryos: essential for centriole duplication; in *Drosophila*: not required for somatic centriole duplication, indispensible for centriole duplication in the egg	(Otto et al., 2002; Delattre et al., 2006; Pelletier et al., 2006; Dix and Raff, 2007; Gomez-Ferreria et al.,

		after fertilisation; centrosomal accumulation partially dependent on Aurora-A and Dynein, and co-dependent on SPD-5; SPD-2 required for ZYG-1 recruitment to the centrosome; overexpression results in multiple extra-centrosomal foci of gamma-tubulin and pericentrin which supports an early role of CEP192 in centrosome assembly	2007; Kleylein-Sohn et al., 2007; Zhu et al., 2008)
CP110	Centriole (distal end), not at basal body	negative regulator of cilia formation (together with CEP97); in human cells: essential for centrosome duplication; substrate for Cdk2; abundance is cell-cycle regulated (has protein degradation motif); depletion results in centriole elongation (no homogenous extension of the centriole, but elongated centrioles have frayed ends)	(Chen et al., 2002; Spektor et al., 2007; Tsang et al., 2008; Schmidt et al., 2009)
DIP13	Centriole, basal body, axoneme	In *Chlamydomonas*: depletion results in growth reduction, multi-nucleated, multi-flagellated cells	(Pfannenschmid et al., 2003)
EB1	MT, centrosome/ spindle	Associated with MT plus ends (growing ends); tubulin sheet closure (zipper)	(Sandblad et al., 2006; Vitre et al., 2008)
FOP	Centrosome	Human centrosome proteome; MT anchoring at centrosome with CAP350 and EB1; FOP interacts with CAP350 and this interaction is required for FOP's centrosomal localisation; interacts with EB1, localisation to plus-end due to this interaction	(Andersen et al., 2003; Yan et al., 2006)
MKS3	Primary cilium and plasma membrane	Mutated in Meckel-Gruber Syndrome; interacts with MKS1; Depletion results in block of centriole migration to apical membrane	(Dawe et al., 2007)
MKS5	Centrosome, basal body, cilium	Mutated in Meckel-Gruber Syndrome and Joubert Syndrome; interact with NPHP4	(Arts et al., 2007; Delous et al., 2007)
MKS6	Centrosome	Mutated in Meckel-Gruber Syndrome and Joubert Syndrome; interacts with CEP290	(Gorden et al., 2008)
C-NAP1	Centrosome	Human centrosome proteome; involved in centriole coupling/linkage after disengagement	(Mayor et al., 2002; Andersen et al., 2003; Kim, K. et al., 2008)
Ninein	Centrosome, around mother centriole appendages	Centrosomal MT anchoring; non-nucleating MT minus-end stabilising; role in organising centrosomal proteins; interacts with -tubulin	(Mogensen et al., 2000; Ou et al., 2002; Delgehyr et al., 2005)
NPHP1	Base of cilia	Mutated in Nephronopthisis/Senoir-Loken Syndrome; interacts with signalling molecules involved in cell-adhesion and actin cytoskeleton organisation	(Otto et al., 2002; Fliegauf et al., 2006)
NPHP3		Mutated in Nephronopthisis/Senoir-Loken Syndrome	(Olbrich et al., 2003)
NPHP4	Primary cilium, base of cilium	Mutated in Nephronopthisis/Senoir-Loken Syndrome	(Otto et al., 2002)
NPHP5	Primary cilium connecting cilia of photoreceptors , primary cilia of renal epithelial cells	Mutated in Nephronopthisis/Senoir-Loken Syndrome; interacts with RPGR	(Otto et al., 2005)
PCM1	Pericentriolar satellites, site of ciliogenesis	Depletion results in reduced levels of centrin, pericentrin and ninein at centrosome, and disrupted centrosomal MT organisation; involved in MT- and dynactin-dependent recruitment of proteins to the centrosome; interacts with BBS4	(Kubo et al., 1999; Dammermann and Merdes, 2002; Nachury et al., 2007)
PLK		Polo-like kinases; cell-cycle regulated kinase family	(Archambault and Glover, 2009)
PLK4	Centrosome	Member of the polo-like kinase family; required for centriole duplication in *Drosophila* and humans	(Bettencourt-Dias et al., 2005; Habedanck et al., 2005; Kleylein-Sohn et al., 2007)
POC1	Centriole (proximal end), fiber systems	Identified in *Chlamydomonas* centriole proteome and *Tetrahymena* basal body proteome; depletion in *Drosophila* is lethal; depletion in *Chlamydomonas* impairs centriole duplication and overexpression results in elongated centrioles	(Keller et al., 2005; Dietzl et al., 2007; Kilburn et al., 2007; Keller et al., 2009)
POC5	Centriole (distal end)	Identified in *Chlamydomonas* centriole proteome; centrin-interacting protein; depletion results in growth defect	(Keller et al., 2005; Azimzadeh et al., 2009)
Rootletin	Centrosome, rootlets	Interacts with C-NAP1	(Yang et al., 2006)
SAS-4	Centriole (mature and pro-centriole)	Essential for centriole duplication in *C. elegans*, *Drosophila*, and humans; involved in attachment of microtubules to the central tube/centriole; MT-interacting domain; quantity of SAS-4 affects the length of the centriole (or at least its MTs); cell-cycle dependent degradation mediated by APC/Cdh1	(Kirkham et al., 2003; Leidel and Gonczy, 2003; Hung et al., 2004; Basto et al., 2006; Pelletier et al., 2006; Kleylein-Sohn et al., 2007; Dammermann et al., 2008; Hsu et al., 2008; Kohlmaier et al., 2009; Schmidt et al., 2009; Tang et al., 2009)
SAS-6	Centriole (cartwheel)	Essential for centriole duplication in *C. elegans*, *Drosophila*, humans, *Chlamydomonas*, zebrafish, Tetrahymena; acts early during centriole biogenesis, either as component (hub) of the cartwheel or another proximal structure in C. elegans; cell-cycle dependent degradation mediated by APC/Cdh1; upstream of SAS-4 in centriolar pathway	(Dammermann et al., 2004; Leidel et al., 2005; Delattre et al., 2006; Pelletier et al., 2006; Kleylein-Sohn et al., 2007; Nakazawa et al., 2007; Peel et al., 2007; Rodrigues-Martins et al., 2007; Strnad et al., 2007; Vladar and Stearns, 2007; Yabe et al., 2007; Culver et al., 2009)
SZY-20	Centrosome	Suppressor of *C. elegans zyg-1* phenotype; negative regulator of centriole duplication	(Song et al., 2008)
VFL1	Basal body (mature and pro-basal body)	*Chlamydomonas* mutant: variable flagella length; rotational asymmetric localisation of VFL1 mostly to triplet 1 (or 2) which is the site of pro-basal body assembly	(Adams et al., 1985; Silflow et al., 2001)

WDR16	Basal body (mature, pro-basal), axoneme	Identified in *T. brucei* flagella proteome	(Broadhead et al., 2006), H. Farr K. Gull personal communication
XMAP215	Microtubles (preferentially plus ends), centrosome	Control of MT dynamics; MT polymerase	(Popov et al., 2001; Brouhard et al., 2008)
ZYG-1	Centrosome	Kinase, required for daughter centriole formation in C. elegans; requires SPD-2 for its centrosomal recruitment; upstream of SAS-6/SAS-5 and SAS-4 in the centriolar pathway	(O'Connell et al., 2001; Delattre et al., 2006; Pelletier et al., 2006)
γ-tubulin	Centrsome, MT nucleation sites	MT nucleation	(Oakley and Oakley, 1989; Raynaud-Messina and Merdes, 2007)
δ-tubulin	Centrosome (between centrioles)	In *Chlamydomonas*: uni3-1 strain has centrioles with mainly MT doublets, but also MT triplets, misplaced fiber systems; in human:, δ -tubulin localises to the centrosome, presumably between the centrioles, interacts with -tubulin; in *T. brucei*: depletion mainly basal bodies with MT duplets	(Dutcher and Trabuco, 1998; Chang and Stearns, 2000; Smrzka et al., 2000; O'Toole et al., 2003; Kato et al., 2004; Gadelha et al., 2006)
ε-tubulin	Centriole (distal appendages)	In *Chlamydomonas*: bld-2 strain has centrioles with MT singlets, null alleles for ε -tubulin not viable; in *Paramecium*: depletion results in basal bodies with aberrant MT numbers; in human: stronger association with the older centriole, localisation to the distal appendages (EM), required for centriole duplication	(Goodenough and StClair, 1975; Chang and Stearns, 2000; Dupuis-Williams et al., 2002; Dutcher et al., 2002; Chang et al., 2003)

2.8.3 Table S3: IDs of proteins within a putative homologue cluster

	ALMS1	AHI1	Asterless	BBS10	BBS11	BBS12	BBS1	BBS2	BBS3
Apis mellifera		GB12459-PA	GB11327-PA				GB15554-PA	GB16076-PA	GB13103-PA
Arabidopsis thaliana									
Aureococcus anophagefferens		72559					24826	24976	
Batrachochytrium dendrobatidis		BDEG_04818							BDEG_08460
Caenorhabditis elegans							CE25181	CE04433	CE00921
Capitella sp	229158	228034		193805		205266	108485	171644	143205 97109 202236
Chlamydomonas reinhardtii v3.1		170267					132537	126758 93657	24475
Chlamydomonas reinhardtii v2		178044					152526	159573	171782
Ciona intestinalis	ENSCINP00000013951 XP_002127193.1 ENSCINP00000021711	ENSCINP00000022494 ENSCINP00000022496 ENSCINP00000022495	XP_002122978.1 ENSCINP00000013334	XP_002125019.1			XP_002125520.1 ENSCINP00000012520	ENSCINP00000025380 ENSCINP00000013276 XP_002124766.1 ENSCINP00000025379	ENSCINP00000006473 ENSCINP00000009643
Cryptosporidium parvum									
Cyanidioschyzon merolae									
Danio rerio	XP_690375.3 OTTDARP00000008743	NP_0010710291	XP_694227.3	NP_001082932.1 OTTDARP00000015588	NP_001107066.1	XP_699172.3	NP_001091721.1	OTTDARP00000013101	OTTDARP00000019948 NP_001008733.1
Dictyostelium discoideum									
Drosophila melanogaster			FBpp0078392				FBpp0076621		FBpp0085718
Encephalitozoon cuniculi									
Entamoeba histolytica									
Gallus gallus	ENSGALP00000036306 ENSGALP00000036307 ENSGALP00000036308 XP_420887.2 ENSGALP00000025867 ENSGALP00000025793 XP_421818.2 ENSGALP00000015835	ENSGALP00000022567	XP_413819.2	ENSGALP00000016626 ENSGALP00000037679 XP_416107.2	ENSGALP00000011407	XP_420621.2 ENSGALP00000019293 ENSGALP00000036988		ENSGALP00000039092 ENSGALP00000004804 NP_001012795.1	XP_001236032.1
Giardia lamblia									
Homo sapiens	NP_055935.4 OTTHUMP00000070393 OTTHUMP00000070394 OTTHUMP00000070395 OTTHUMP00000020732 NP_001004298.2 OTTHUMP00000020731	NP_0011283041 OTTHUMP00000017263 NP_0601213 NP_0011283021 NP_0011283031 OTTHUMP00000017266 OTTHUMP00000017265	NP_055800.2 OTTHUMP00000176101	OTTHUMP00000168556 OTTHUMP00000190861	OTTHUMP00000022764 OTTHUMP00000022763	OTTHUMP00000164085 OTTHUMP00000196371	OTTHUMP00000064293 OTTHUMP00000041483 OTTHUMP00000064288 OTTHUMP00000041484 OTTHUMP00000064289 OTTHUMP00000064284 OTTHUMP00000064292 OTTHUMP00000064291 OTTHUMP00000064285 OTTHUMP00000064290	OTTHUMP00000108453 OTTHUMP00000080176 OTTHUMP00000083205 OTTHUMP00000041825 OTTHUMP00000041826	OTTHUMP00000171958 OTTHUMP00000171957 OTTHUMP00000171956 OTTHUMP00000171960
Leishmania major							LmjF35.4180	LmjF30.0590	LmjF16.1380
Lottia gigantea	231754	62094		169741			115648	214128	204362
Monosiga brevicollis							24309	22806	
Naegleria gruberi		73736					65179	71257	44202 62886
Nematostella vectensis	246718	104652		240131		241874	186039	235159	216064
Neurospora crassa									
Oryza sativa									
Ostreococcus tauri									
Paramecium tetraurelia		GSPATP00023705001 GSPATP00036655001					GSPATP00033252001	GSPATP00000964001	GSPATP00038302001 GSPATP00017502001
Phaedactylum tricornutum									
Physcomitrella patens									
Phytophthora sojae		143507					136833	134838	109614
Plasmodium falciparum									
Populus trichocarpa									
Rhizopus oryzae									
Saccharomyces cerevisiae									
Schizosacharomyces pombe									
Strongylocentrotus purparatus			GLEAN3_21838				GLEAN3_11446 XP_783764.2 XP_001176024.1	XP_001176955.1 XP_001176601.1 GLEAN3_00964 GLEAN3_06082 XP_001202157.1 XP_001187752.1 GLEAN3_27384	GLEAN3_17987 XP_796160.2 XP_001196617.1
Takifugu rubripes		ENSTRUP00000035594 ENSTRUP00000020791 ENSTRUP00000020793 ENSTRUP00000020792 ENSTRUP00000020790		ENSTRUP00000003984			ENSTRUP00000010503 ENSTRUP00000010504		ENSTRUP00000008721 ENSTRUP00000008719 ENSTRUP00000008720 ENSTRUP00000044163 ENSTRUP00000044162
Tetrahymena thermophila		54m00169					225.m00041	54.m00294	69.m00210
Thalassiosira pseudonana									
Theileria anulata									
Toxoplasma gondii									
Trichoplax adherens	60736	58114		60786		54075	26650	29585	25605
Trichomonas vaginalis		90380m00189 84684m00034 97465m00145					97190.m00036	93171.m00065	85228.m00099 89416.m00092 95645.m00121
Trypanosoma brucei							Tb09.211.2080	Tb927.6.2020	Tb927.8.5060
Ustilago maydis									

BBS4	BBS5	BBS6	BBS7	BBS8	BBS9	CEP135	Centrin
XP_624286.2 GB15760-PA	GB17090-PA		GB18131-PA XP_395119.2	XP_001123116.1 GB10554-PA	GB12962-PA	GB19598-PA XP_001121524.1 GB17790-PA	GB10266-PA XP_395906.2
1883	72072		22539	54838 72299		68641	23303
						BDEG_01407	BDEG_08567
CE37774	CE01042		CE23024	CE31071	CE32166		
94608	176376	214066	117810	163723	155802	173417	21986
129948	182299		190054	140113	101137	166062	159554
	161738		154949	171797	154402	168432	156177
ENSCINP00000019082 XP_002125730.1	ENSCINP00000005850 XP_002124803.1	XP_002129925.1	ENSCINP00000026073 ENSCINP00000008442 XP_002127808.1 ENSCINP00000008441	XP_002123464.1 ENSCINP00000019928 XP_002123334.1 XP_002123270.1 ENSCINP00000019927	XP_002125247.1	XP_002126689.1 ENSCINP00000021787 XP_002126770.1 ENSCINP00000010605	ENSCINP00000018417
							cgd3_1270
XP_001920926.1 NP_001070934.1	NP_956593.1	NP_956459.1	NP_001070613.1	NP_001035399.1	XP_001921482.1	NP_001070757.1 OTTDARP00000011178	XP_001345066.1 XP_001922913.1
FBpp0087297	FBpp0078512			FBpp0077718	NP_611580.1	FBpp0075391	FBpp0110460 FBpp0080611
XP_413706.2 ENSGALP00000002771	ENSGALP00000015990	ENSGALP00000014647	ENSGALP00000019359	ENSGALP00000017328 XP_421311.1	XP_418843.2 ENSGALP00000019874 ENSGALP00000019873	ENSGALP00000027006 XP_416892.2 ENSGALP00000036760 XP_420699.2 ENSGALP00000022332	XP_420280.1 ENSGALP00000012127 XP_420622.2 ENSGALP00000019297
	GL50803_8146			GL50803_8508			GL50803_6744
OTTHUMP00000175949 OTTHUMP00000164777	OTTHUMP00000163062	OTTHUMP00000162353 OTTHUMP00000030274	OTTHUMP00000164091 NP_060660.2	NP_938052.1 OTTHUMP00000028245 NP_938051.1 NP_653197.2	NP_001028777.1 NP_940820.1 NP_001028776.1 OTTHUMP00000158833 OTTHUMP00000024560	OTTHUMP00000161217 NP_079285.2	OTTHUMP00000071829 OTTHUMP00000162193 OTTHUMP00000026058
LmjF36.2280	LmjF12.0680		LmjF29.0710	LmjF09.0440	LmjF32.1240		LmjF07.0710
53377	207456	145572	223884	161371	222053 176459	213275 67690	218790
23427	32801			32314	33174	39071	19580
28891	34252		68114	80979	80972		56351
40410	164107	50325	148193 103100	159317 83222	10499 144317		106028
						30149	28265
GSPATP00005292001	GSPATP00036912001 GSPATP00034169001		GSPATP00026091001	GSPATP00028481001	GSPATP00027545001	GSPATP00009438001 GSPATP00019908001	GSPATP00030725001 GSPATP00026291001
						69693	128401 87691 233973
	109500		108146	109079		133301	108216
							PFA0345w
GLEAN3_18697 XP_001202895.1 XP_786510.2	GLEAN3_18028 GLEAN3_22767 XP_792212.1 XP_001202535.1	GLEAN3_24419 XP_786835.1 XP_001193589.1	XP_001203356.1 XP_001188908.1 GLEAN3_14001 XP_798575.2 XP_001191665.1 GLEAN3_05312 GLEAN3_23783 XP_790931.2 XP_001191706.1	XP_001198570.1 XP_001199780.1 GLEAN3_12560	GLEAN3_00493 XP_001190229.1 XP_001194456.1 XP_001201657.1 XP_793825.2	XP_792224.1 XP_001202035.1 XP_781904.2 XP_001189902.1	XP_794321.2 XP_001193391.1 XP_001177362.1 XP_001176455.1
ENSTRUP00000042271 ENSTRUP00000042267 ENSTRUP00000042268 ENSTRUP00000042269 ENSTRUP00000042270	ENSTRUP00000045891	ENSTRUP00000000326 ENSTRUP00000000327	ENSTRUP00000018248	ENSTRUP00000046798 ENSTRUP00000046797 ENSTRUP00000046796	ENSTRUP00000000646 ENSTRUP00000000645	ENSTRUP00000039917 ENSTRUP00000039916 ENSTRUP00000039921 ENSTRUP00000039918 ENSTRUP00000039919 ENSTRUP00000039920	ENSTRUP00000023721 ENSTRUP00000021965
214.m00049	131.m00082		96.m00149	78.m00118	64.m00248	399.m00011 260.m00025	41.m00190
							35423
							TA02500
	TGGT1_010150						TGVEG_004780 TGME49_047230 TGGT1_025770
18483	50116	53298	25051	49816	22294		25393
	90976.m00058			91962.m00087	89996.m00100	80874.m00020 82538.m00086 83472.m00460	94232.m00202 80798.m00086 81190.m00171
Tb10.6k15.3940	Tb927.1.4140		Tb927.3.3640	Tb11.01.4290	Tb11.01.6070		Tb927.8.1080

Centriolin	CEP164	CEP192	CAP350	CEP55	CEP68	CEP76	CEP97	CP110
XP_393975.3 GB18995-PA	GB16672-PA	GB11184-PA	GB13282-PA XP_393557.3				GB17708-PA XP_001122173.1	GB18645-PA
38639	61427					456		
BDEG_01294						BDEG_07240		
114721 44232	100218	219671 211819	226518 209677	223974		130762 151783	208434	217735 195460
172483	169222							
164264	156609					169826 170801		
XP_002119245.1 ENSCINP00000017781	XP_002121669.1	ENSCINP00000002349 XP_002122890.1 ENSCINP00000020451	ENSCINP00000002338 ENSCINP00000028718 XP_002123277.1 ENSCINP00000024075 ENSCINP00000002512 ENSCINP00000029651			XP_002130643.1 ENSCINP00000022633 cgd5_780 cgd6_680 cgd7_3040		XP_002121364.1 ENSCINP00000011624
XP_001919236.1		XP_693893.3 DDB0218699	XP_001343753.2 XP_001920199.1	OTTDARP00000022719 XP_692717.3	OTTDARP00000024180 NP_001074119.1 OTTDARP00000001702 XP_001919948.1	NP_001096589.1	NP_938185.1	XP_001345378.2
	FBpp0073951 FBpp0073952	FBpp0075122 FBpp0088671					FBpp0077192 FBpp0077193	FBpp0077044 FBpp0100112
XP_414814.1 ENSGALP00000007847 XP_415403.2 ENSGALP00000038078 ENSGALP00000038077 ENSGALP00000038081 ENSGALP00000038080	XP_417909.2 ENSGALP00000011817 GL50803_24423	ENSGALP00000022468	ENSGALP00000006296 XP_422260.2 ENSGALP00000002702	ENSGALP00000039000 XP_001231504.1 ENSGALP00000010710	ENSGALP00000014274 XP_419343.2 ENSGALP00000016220	XP_001231809.1 ENSGALP00000034644 ENSGALP00000022439 XP_001235558.1 GL50803_5672	XP_416617.2 ENSGALP00000024714	ENSGALP00000039138 XP_414915.2 ENSGALP00000011207
OTTHUMP00000022010 OTTHUMP00000158535 OTTHUMP00000022003 OTTHUMP00000022004	NP_055771.4 LmjF17.1140 LmjF28.2390 LmjF12.0510	OTTHUMP00000071748 OTTHUMP00000071750 OTTHUMP00000071749 NP_115518.2 OTTHUMP00000071751	OTTHUMP00000033201 OTTHUMP00000033199 OTTHUMP00000033202 OTTHUMP00000033200 OTTHUMP00000192895 OTTHUMP00000033203	NP_060601.3 NP_001120654.1 OTTHUMP00000020117 OTTHUMP00000020118	OTTHUMP00000159934 OTTHUMP00000159935 OTTHUMP00000178791 OTTHUMP00000028233	OTTHUMP00000071724 OTTHUMP00000071719 OTTHUMP00000071721 OTTHUMP00000162484 OTTHUMP00000071720 OTTHUMP00000071722 OTTHUMP00000071723 OTTHUMP00000071718 NP_115645.3 OTTHUMP00000080445 OTTHUMP00000164678 LmjF28.2090 LmjF05.1020	OTTHUMP00000172078	NP_055526.3 OTTHUMP00000162162 OTTHUMP00000080381 OTTHUMP00000080380 OTTHUMP00000162163
136351	105990		89468 135146 140328 56916			229828	210409	157573
26670 37452	13726	11927	28503			26831 32767		
77107	70383					31102		
91406	128876	225408 246496 219395	215825 212801 87242 199442			219939		222993
GSPATP00007697001 GSPATP00022750001	GSPATP00011678001 GSPATP00026805001 GSPATP00033074001 GSPATP00021670001		GSPATP00002266001 GSPATP00019107001			GSPATP00029653001 GSPATP00022693001 GSPATP00025665001 GSPATP00030898001		
	232018							
	132583 158245 143662		140128			142088		
			PFE0070w			PFF0185c PFL2110c		
XP_797673.1	XP_784792.2 XP_001189278.1	XP_001202129.1 XP_001187317.1 XP_790994.2 XP_001194286.1 GLEAN3_17452 XP_001181926.1 XP_001189923.1	GLEAN3_18326 XP_782826.2 XP_001188934.1		XP_800003.2 XP_001190247.1	XP_780096.2 XP_001184625.1	XP_001189892.1 XP_780440.2 XP_001194053.1 XP_780521.2 GLEAN3_21340	
	ENSTRUP00000017587	ENSTRUP00000031444 ENSTRUP00000031445 ENSTRUP00000031443 ENSTRUP00000031442	ENSTRUP00000024377 ENSTRUP00000024378 ENSTRUP00000024375 ENSTRUP00000024374 ENSTRUP00000024376	ENSTRUP00000027582 ENSTRUP00000027583 ENSTRUP00000027584 ENSTRUP00000027585 ENSTRUP00000027586	ENSTRUP00000001005	ENSTRUP00000030946	ENSTRUP00000031065 ENSTRUP00000031066 ENSTRUP00000031064 ENSTRUP00000031068 ENSTRUP00000031067 ENSTRUP00000031063	
54.m00206	214.m00046 94.m00127 16.m00389 7.m00305 13.m00328 57.m00243	72.m00131	57.m00227			62.m00250		
						TA11690 TA09225		
	TGME49_114360 TGVEG_099450					TGGT1_081760 TGME49_026610 TGVEG_067670 TGVEG_101860 TGME49_116730 TGME49_116720 TGGT1_094560 TGVEG_101850 TGME49_040910 TGVEG_090360 TGGT1_049850		
54619		57490 55530	53203			18537 57170		
93357.m00029	89183.m00139 94962.m00278 88914.m00133 92757.m00230 84963.m00116 90782.m00048 86088.m00555					81869.m00193 87835.m00207 90175.m00081		
	Tb927.5.2440 Tb11.01.3520 Tb927.1.3560					Tb11.01.0600		

DIP13	EB1	FOP	MKS1	MKS3	MKS4	MKS5	MKS6
GB13090-PA	XP_623370.1 GB19881-PA	XP_396617.3 GB18614-PA GB10545-PA	GB12794-PA XP_001120316.1	GB12651-PA GB19576-PA GB14376-PA	GB18714-PA XP_001122136.1	GB18181-PA XP_001121635.1 GB18807-PA	GB16247-PA
	AT5G62500.1 AT3G47690.1 AT5G67270.1	AT3G55005.1 AT3G55000.1					
55205	7035			22040 61468			72175 72596
	BDEG_07433	BDEG_02453	BDEG_00264	BDEG_06113	BDEG_01167	BDEG_03262 BDEG_03261	BDEG_06092
	CE18982 CE26217 CE26218		CE41329 CE40712	CE02726		CE25774 CE04269	CE06131
	180168 226241	151396	157534	97867 105851	228963 217708	218340 228193	155172
131284	194209		20204	168773	169091 9715		176993
159197	152507		169453	160562 161322	163597		168825 171248
XP_002131700.1 ENSCINP00000024109 XP_002119469.1 ENSCINP00000029736	ENSCINP00000028931 ENSCINP00000028930 ENSCINP00000017486 ENSCINP00000028932 ENSCINP00000017488 XP_002127374.1	XP_002125932.1	ENSCINP00000003134 XP_002121395.1 ENSCINP00000025479 ENSCINP00000025478	ENSCINP00000025044 ENSCINP00000025047 ENSCINP00000009402 XP_002126630.1 ENSCINP00000025046	XP_002121281.1 ENSCINP00000026886 ENSCINP00000014504 ENSCINP00000014503 XP_002121206.1 ENSCINP00000026885 ENSCINP00000014500 XP_002124446.1	ENSCINP00000018342 XP_002121007.1 ENSCINP00000020603 ENSCINP00000018343 ENSCINP00000018339	ENSCINP00000006696 XP_002123322.1 ENSCINP00000006692 ENSCINP00000029710 XP_002125387.1
cgd8_1690	cgd7_2370						cgd6_680 cgd7_3040
	CMJ156C						
XP_689621.2	OTTDARP00000019628 NP_571461.1 NP_001018312.1 OTTDARP00000020723 OTTDARP00000022038 OTTDARP00000022037 OTTDARP00000020725 NP_956158.1 OTTDARP00000020724 NP_001103172.1 XP_001923886.1 NP_001002170.1 OTTDARP00000007138 OTTDARP00000024111	OTTDARP00000026261	OTTDARP00000018990 NP_001070841.1	XP_700974.3		XP_001922153.1 XP_001921881.1	XP_698801.3
	DDB0191160				DDB0191488		
	FBpp0077157 FBpp0085808 FBpp0076502 FBpp0071405 FBpp0079053 FBpp0085462 FBpp0089088 FBpp0085464 FBpp0085463 FBpp0085461 FBpp0112596 FBpp0112595		FBpp0073500	FBpp0112166			FBpp0086100 FBpp0086099
	19068506 85691079						
ENSGALP00000011925 XP_418582.1 XP_001233637.1 XP_001233675.1	ENSGALP00000010754 ENSGALP00000035995 ENSGALP00000024507 NP_001026188.1	XP_419609.1 ENSGALP00000018610	XP_415705.2 ENSGALP00000000998 ENSGALP00000039895	ENSGALP00000025642 XP_418334.2	ENSGALP00000018199 XP_416130.2	ENSGALP00000012598 ENSGALP00000038922 ENSGALP00000005690 XP_001234248.1 XP_001236264.1	ENSGALP00000023382 ENSGALP00000023381 XP_420777.2 XP_001232304.1
GL50803_16263	GL50803_14048						
OTTHUMP00000022691 OTTHUMP00000171223	OTTHUMP00000030608 OTTHUMP00000163373 OTTHUMP00000073450 OTTHUMP00000073451 OTTHUMP00000122582	OTTHUMP00000017612 OTTHUMP00000017613	OTTHUMP00000182058 OTTHUMP00000165111 OTTHUMP00000182059 OTTHUMP00000165109 OTTHUMP00000165110	OTTHUMP00000177903 NP_001135773.1	OTTHUMP00000168670 NP_079390.3 OTTHUMP00000168672 OTTHUMP00000168671	NP_056087.2 OTTHUMP00000108509 OTTHUMP00000080888 OTTHUMP00000080887 OTTHUMP00000028336 NP_065099.3	OTTHUMP00000036196 OTTHUMP00000020166 NP_001073991.2 OTTHUMP00000036195 OTTHUMP00000036194 XP_001718184.1 XP_001717962.1 XP_001714415.1 XP_001725151.1 XP_001724777.1 XP_001718667.1 OTTHUMP00000020164
LmjF34.4540		LmjF33.0050		LmjF27.0020			LmjF01.0730
214610	177133	236017 84934	228399	192673		184477	215375
10940	32858	22652	30962	38408	27498		25895 12869 26368
71898	44546 65633	51775	29577	62841		61544	77673
90789	196354	233042	209121 248809	120147	248638 20626 220337 223974 105650	50175	30054 145898 224689
	NCU00243						
	Os04g54940.1 Os10g35580.1	Os11g01170.1 Os12g01170.1					
	31915 5403						
GSPATP00031715001 GSPATP00014975001	GSPATP00024775001 GSPATP00029347001 GSPATP00027278001 GSPATP00003968001 GSPATP00009957001 GSPATP00004446001 GSPATP00003747001 GSPATP00032765001 GSPATP00027670001 GSPATP00038011001 GSPATP00034868001	GSPATP00015241001 GSPATP00009206001 GSPATP00011360001 GSPATP00004410001 GSPATP00019104001 GSPATP00002270001 GSPATP00006720001	GSPATP00011904001	GSPATP00015939001 GSPATP00005677001	GSPATP00005779001	GSPATP00016244001	GSPATP00007873001 GSPATP00025665001 GSPATP00030898001 GSPATP00025666001
	54306			48099			
65272 166011 172179	52398 114249 124399 124479 66459	167541					
108625	140534 108686	143118 134177				136855	
MAL13P1.331	PFC0305w						PFF0185c PFL2110c
	661879 671317 208732 218504	563805 227781					
	RO3G_14394.1 RO3G_08998.1 RO3G_14395.1					RO3G_13248.1	
	YER016W						
	SPAC18G6.15						
XP_791762.1 XP_001178037.1	XP_001185519.1 XP_001191889.1 XP_787950.1 GLEAN3_27631	XP_787100.2 XP_001176074.1	XP_787637.1 XP_001181959.1	XP_791033.1 XP_001199054.1	XP_001200309.1 XP_796292.2 XP_001189059.1 XP_799269.2 XP_796018.1 XP_001197771.1 XP_001185617.1 XP_001192113.1 XP_781881.2 XP_001194414.1 XP_798213.2 XP_001183119.1 XP_790657.1	XP_001198524.1 XP_792785.2 XP_790646.2 XP_001179958.1	XP_001203052.1 XP_781068.2 XP_001193651.1 XP_001193792.1
ENSTRUP00000026152 ENSTRUP00000026239	ENSTRUP00000012966 ENSTRUP00000019118 ENSTRUP00000013241 ENSTRUP00000012594 ENSTRUP00000023374 ENSTRUP00000023373 ENSTRUP00000023372	ENSTRUP00000012779	ENSTRUP00000018719 ENSTRUP00000018717 ENSTRUP00000018718	ENSTRUP00000004863 ENSTRUP00000026816	ENSTRUP00000010885 ENSTRUP00000010884 ENSTRUP00000010882 ENSTRUP00000010883 ENSTRUP00000032490	ENSTRUP00000021978 ENSTRUP00000032778 ENSTRUP00000032777 ENSTRUP00000032779 ENSTRUP00000032781 ENSTRUP00000032780 ENSTRUP00000032782 ENSTRUP00000032783 ENSTRUP00000016611 ENSTRUP00000016606 ENSTRUP00000016607 ENSTRUP00000016605 ENSTRUP00000016609 ENSTRUP00000016608 ENSTRUP00000016610	ENSTRUP00000023018 ENSTRUP00000035842 ENSTRUP00000035843 ENSTRUP00000035844 ENSTRUP00000035841 ENSTRUP00000035845
	8.m00318 73.m00186 229.m00036 58.m00179 10.m00342 78.m00136 77.m00205	69.m00270 105.m00161 6.m00534 61.m00222	89.m00156	198.m00078 273.m00023	16.m00381 101.m00122	128.m00081	36.m00254 62.m00250
	36137		5008	6924		4971	6295
	TA18025						TA11690 TA09225
TGGT1_096790 TGVEG_050430 TGGT1_054490 TGME49_095450	TGVEG_068550 TGME49_027650 TGGT1_083050						TGVEG_024920 TGME49_030660 TGGT1_117190 TGGT1_049850 TGVEG_090360 TGME49_040910
24325	57922	56432	61316	55538		51343	28071
	93469.m00003 85241.m00262 89135.m00247 87955.m00265 85699.m00083		81238.m00180 85889.m00463	81244.m00185 86322.m00058 93588.m00120 90756.m00173 86498.m00447 80754.m00134	102888.m00005 93784.m00040		97465.m00138 96723.m00091 95634.m00071 88273.m00365
Tb10.61.2720	Tb09.160.1440 Tb927.8.2890	Tb10.61.2870		Tb927.5.3540			Tb09.160.3240
	UM05761.1						

NAP1	Ninein	NPHP1	NPHP3	NPHP4	NPHP5	PCM1	PLK4	PLK
GB30086-PA XP_624402.1 GB10082-PA							XP_623133.2 GB15315-PA	GB10053-PA
			72389					
				BDEG_06365				
		CE16278		CE38409				CE22877 CE26649 CE30602
	201605	203909 141498 121714 120958	219276 186605	223925	168052	228845	221177	72562 160517
		191079	194130	32880				
		157551		164333	155819			
	XP_002120842.1	XP_002127999.1	ENSCINP00000029018 XP_002123449.1 ENSCINP00000009831 ENSCINP00000029015 NP_001071784.1 ENSCINP00000007133		ENSCINP00000024112 ENSCINP00000014179 XP_002131862.1	XP_002127807.1	ENSCINP00000026188 XP_002125497.1 ENSCINP00000026189 ENSCINP00000004628	ENSCINP00000015359 XP_002121371.1 ENSCINP00000007493 ENSCINP00000007506 ENSCINP00000007522
								CME146C
XP_692550.3	XP_700610.3 XP_001922227.1 XP_001335233.2 XP_001332720.1 XP_690418.3	NP_001070638.1	XP_691073.2	OTTDARP00000011238 XP_001921573.1 OTTDARP00000017142 NP_001124247.1 OTTDARP00000017097 OTTDARP00000017143	NP_001008622.1 NP_001104719.1	OTTDARP00000019790 NP_001108595.1	NP_001112364.1	OTTDARP00000020344 NP_958465.1 NP_001092715.1 XP_001919481.1 OTTDARP00000019795 NP_001003890.3
								DDB0216332
FBpp0083929							FBpp0078042	
								19068465 85690997
XP_425750.2 ENSGALP00000002726 XP_417323.2	XP_426482.2 ENSGALP00000020163	ENSGALP00000013332 XP_419298.2	XP_418790.2 ENSGALP00000019105	ENSGALP00000001428	ENSGALP00000019025 ENSGALP00000024116	ENSGALP00000031473 ENSGALP00000022090 ENSGALP00000036784	ENSGALP00000016553	ENSGALP00000023702 XP_424739.2 NP_001025810.1
OTTHUMP00000196679 OTTHUMP00000061639 OTTHUMP00000061640 OTTHUMP00000196095 OTTHUMP00000030757 NP_001030595.1	NP_891989.2 OTTHUMP00000195019 NP_065972.3 NP_891991.1 NP_057434.3 OTTHUMP00000195018 OTTHUMP00000195017 OTTHUMP00000195016 OTTHUMP00000195510 OTTHUMP00000195014 OTTHUMP00000043419	NP_001121651.1 NP_001121650.1 OTTHUMP00000161841 OTTHUMP00000161840	OTTHUMP00000172836 OTTHUMP00000172842 OTTHUMP00000172838 OTTHUMP00000172839 OTTHUMP00000172837 OTTHUMP00000172835	OTTHUMP00000000745 NP_055917.1	OTTHUMP00000172423 OTTHUMP00000158861 OTTHUMP00000172420 OTTHUMP00000172425 OTTHUMP00000172422 OTTHUMP00000172421 OTTHUMP00000172424	OTTHUMP00000177082 OTTHUMP00000177087 OTTHUMP00000177083 OTTHUMP00000096525 NP_006188.3 OTTHUMP00000177086 OTTHUMP00000177084 OTTHUMP00000177085	OTTHUMP00000164423	OTTHUMP00000010159 OTTHUMP00000122549 OTTHUMP00000122462 OTTHUMP00000081380
			LmjF34.3490	LmjF25.2000				LmjF17.0790
		232470	126945	125375	185823	153121 176986	111275 111267	223730
				8015	37188	33155		39338
		71176						63986 53552
		217835	244876	242529	158806	202784 228836	246408	128941 230033
								NCU09258
								32134
				GSPATP00008603001 GSPATP00000785001 GSPATP00010871001 GSPATP00005471001				GSPATP00029023001 GSPATP00018143001 GSPATP00025198001 GSPATP00009060001 GSPATP00031090001 GSPATP00028443001 GSPATP00026684001 GSPATP00036333001 GSPATP00022475001 GSPATP00009588001 GSPATP00006256001 GSPATP00039351001
								10628
			135076					
				140465	158462			
								YMR001C
								SPAC23C11.16
	XP_788604.2 XP_001179087.1	XP_782288.2 XP_001185182.1	XP_001203493.1 XP_795144.2 GLEAN3_22587	XP_797126.1 XP_001203768.1 XP_793289.1 XP_001195422.1	XP_001192289.1 XP_782682.2 XP_001199176.1 XP_782804.2	XP_786223.2 XP_001188717.1	GLEAN3_16352 XP_001184023.1 XP_001192451.1	GLEAN3_00468 GLEAN3_17949
	ENSTRUP00000021428 ENSTRUP00000021430 ENSTRUP00000021429 ENSTRUP00000021427 ENSTRUP00000014916 ENSTRUP00000014915 ENSTRUP00000014914		ENSTRUP00000045676 ENSTRUP00000045677 ENSTRUP00000045678	ENSTRUP00000045307	ENSTRUP00000023183 ENSTRUP00000023182	NP_001027776.1 ENSTRUP00000040906 ENSTRUP00000040912	ENSTRUP00000018584 ENSTRUP00000018583	ENSTRUP00000003764 ENSTRUP00000003763 ENSTRUP00000033671 ENSTRUP00000033670 ENSTRUP00000033668 ENSTRUP00000033669 ENSTRUP00000037778 ENSTRUP00000037777
		96.m00096	7.m00547 31.m00371	95.m00130 12.m00312				56.m00207 220.m00040 392.m00011 15.m00540
				3508				33979
55968	52869	54526	22397 54460		53628	62202	30306	50801
			Tb927.4.1280					Tb927.7.6310
								UM03234.1

POC1	POC5	Rootletin	SAS4	SAS6	SZY-20	Delta-tubulin	Epsilon-tubulin
GB15587-PA XP_624420.1		GB30086-PA GB10082-PA XP_624402.1	GB11770-PA XP_396122.3	GB15158-PA XP_395972.3	GB11242-PA	GB16674-PA XP_001122683.1	XP_394700.3 GB15835-PA
27325			15226	72351			2189
BDEG_01586	BDEG_05569 BDEG_06763 BDEG_03888		BDEG_06323	BDEG_07637		BDEG_02728	BDEG_05632 BDEG_08433
	224814		148844 226664	198072		136680 159439	218781 89314
112249	144089 9798		142043	136677 176336		136082	188195
162499	155108		152791 162786 171974	158310 157385		164724	165702
XP_002127877.1 ENSCINP00000018966	XP_002130350.1 ENSCINP00000006751	ENSCINP00000016483 ENSCINP00000016482 XP_002119303.1 ENSCINP00000016484	XP_002123980.1 ENSCINP00000000162	ENSCINP00000018623 XP_002128064.1 cgd4_2610 cgd8_490 CMB012C CMX010C	XP_002119900.1 ENSCINP00000021131	XP_002121850.1 NP_001027643.1	ENSCINP00000002530 XP_002128990.1
OTTDARP00000022008 NP_956412.1 NP_998214.1	OTTDARP00000018901	XP_692550.3 XP_688293.3 XP_001922586.1	OTTDARP00000026300 XP_697409.3	XP_001335692.2 NP_998603.1 OTTDARP00000009325	OTTDARP00000022968	NP_001002093.1 OTTDARP00000024251	OTTDARP00000003871 NP_957279.1 OTTDARP00000003844
FBpp0075597		FBpp0083929 FBpp0083930	FBpp0081088	FBpp0084912 19068525 85691117	FBpp0079057 FBpp0079058		
NP_001006232.1 ENSGALP00000018266 XP_414244.1 ENSGALP00000002194	XP_424797.2 ENSGALP00000024081	XP_425750.2 ENSGALP00000005940 ENSGALP00000002726 XP_417323.2	ENSGALP00000027641 ENSGALP00000018491 XP_419608.2 XP_417152.2 ENSGALP00000027640	NP_001026444.1 ENSGALP00000008489	ENSGALP00000005915	ENSGALP00000008290	ENSGALP00000024172
GL50803_33762			GL50803_13352			GL50803_5462	GL50803_6336
OTTHUMP00000168680 OTTHUMP00000168681 OTTHUMP00000171441 OTTHUMP00000171439 NP_056241.2	NP_001092741.1 NP_689621.2	OTTHUMP00000196095 NP_001030595.1 OTTHUMP00000030757 OTTHUMP00000196679 OTTHUMP00000061639 OTTHUMP00000002222 OTTHUMP00000002223 XP_001732878.1	OTTHUMP00000018137 OTTHUMP00000018138 OTTHUMP00000174371 OTTHUMP00000174372 OTTHUMP00000067368 NP_004601.3 OTTHUMP00000017628	OTTHUMP00000012455	OTTHUMP00000002235 NP_056424.1 XP_932280.1 XP_943492.1 XP_001714740.1	OTTHUMP00000182090 OTTHUMP00000182091 OTTHUMP00000182092	OTTHUMP00000017033 OTTHUMP00000017035 OTTHUMP00000017034
LmjF34.0470	LmjF36.2690		LmjF13.1590	LmjF35.4280		LmjF36.4990	LmjF21.1010
223819	174270	169115 133011	103874 174618		234409	126843 225368	224456
14637	37735		27468	29189		33187	29416
33676 3161	73050		61107	68996		65724 69007	44774
162437	246342 246343	232574 165346	104524 26007 90077		244736	82063 209136	89613
3915			11326	30207			
GSPATP00016878001 GSPATP00013588001 GSPATP00005111001 GSPATP00027105001	GSPATP00030131001 GSPATP00029936001		GSPATP00017531001 GSPATP00004155001 GSPATP00010124001 GSPATP00002681001 GSPATP00004862001 GSPATP00003188001	GSPATP00032784001 GSPATP00027645001 GSPATP00020097001 GSPATP00019787001 GSPATP00009285001 GSPATP00024268001 GSPATP00005603001 GSPATP00008149001 GSPATP00011824001		GSPATP00003638001 GSPATP00005791001	GSPATP00038073001
			45511				
133767				66367		66250	99422
139367			129996 108440	128318 143809		158567	118214
			PF14_0558	PFF0375c MAL13P1.85		PFI1635w	PF14_0725
XP_797493.2	XP_797483.2 XP_001186713.1	XP_001193842.1 XP_001199206.1 XP_001200030.1 XP_001180033.1 XP_001187086.1 XP_786339.2 XP_001199159.1 XP_794353.2	XP_001201907.1 XP_796778.2 XP_001201804.1 XP_796764.2	XP_791931.2 XP_001176207.1 XP_784418.2 XP_001185553.1		GLEAN3_04266	GLEAN3_13393 GLEAN3_02663 XP_781720.2 XP_001182983.1 GLEAN3_12141 XP_783098.2 XP_001183033.1
ENSTRUP00000043776 ENSTRUP00000043775 ENSTRUP00000043770 ENSTRUP00000043774 ENSTRUP00000043772 ENSTRUP00000043771 ENSTRUP00000043773 ENSTRUP00000019669 ENSTRUP00000019672 ENSTRUP00000019671 ENSTRUP00000019670 ENSTRUP00000019675 ENSTRUP00000019668 ENSTRUP00000019673 ENSTRUP00000019667 ENSTRUP00000019674	ENSTRUP00000017168 ENSTRUP00000017169 ENSTRUP00000017170	ENSTRUP00000026159 ENSTRUP00000026160	ENSTRUP00000039772 ENSTRUP00000039773 ENSTRUP00000039774	ENSTRUP00000015629 ENSTRUP00000015630	ENSTRUP00000027083 ENSTRUP00000027084 ENSTRUP00000015723 ENSTRUP00000015724 ENSTRUP00000015725	ENSTRUP00000012851 ENSTRUP00000012850	ENSTRUP00000012199
318.m00023	5.m00513		16.m00410 40.m00264	15.m00309 101.m00122 113.m00089 10.m00435 41.m00222		24.m00232 34.m00277	8.m00446
			10156	23096			
			TGME49_058710 TGVEG_075750 TGGT1_011480	TGGT1_040920 TGVEG_107890 TGME49_106430 TGME49_101420 TGGT1_059860		TGME49_007600 TGVEG_040680 TGGT1_020490	TGGT1_000470 TGME49_075870 TGVEG_055390
51125	60303		28606 55968	51343	57681	54235 28772	50572
82686.m00037 84958.m00099 93446.m00217	87526.m00089		83057.m00080 83007.m00259 92349.m00199 85746.m00238 80829.m00091 86485.m00546	94652.m00132 86581.m00019 136524.m00028		87205.m00083	82471.m00124
Tb10.70.4780	Tb10.6k15.3050		Tb11.02.0810	Tb09.211.1930		Tb11.01.2695	Tb10.70.6950 Tb11.0530

Gamma-tubulin	VFL1	WDR16	XMAP215	ZYG1
XP_394981.3 GB18755-PA		GB13234-PA	GB30364-PA GB10660-PA XP_392874.3	
AT3G61650.1 AT5G05620.1			AT2G35630.1	
23795		21121 28848		
BDEG_04849	BDEG_07084	BDEG_06826	BDEG_04589	
CE00224			CE41405 CE20707 CE20858 CE28252	CE28571
174186	212152	218269 229244	154948	
188933		150613 128114	175143	
158831	155034	159155 168135	156582	
XP_002129233.1	XP_002122133.1 ENSCINP00000025492	ENSCINP00000011660 XP_002124411.1 ENSCINP00000011659	ENSCINP00000015586 ENSCINP00000026759 ENSCINP00000015576 ENSCINP00000015580	
cgd7_1980			cgd5_1680	
CMN304C			CMO196C	
NP_957202.1	XP_690381.2	OTTDARP00000015128 NP_001018475.1	NP_001032756.1	
DDB0185068			DDB0168274 DDB0232951	
FBpp0077326 FBpp0080769		FBpp0076646	FBpp0082678	
19173393 19170982 51701890			19069715 19074990	
55.m00198 307.m00057			62.m00181	
XP_418146.2 ENSGALP00000004979	ENSGALP00000025453 XP_418316.2	ENSGALP00000001912	ENSGALP00000013536 XP_001233233.1	
GL50803_114218		GL50803_15218	GL50803_96399 GL50803_91480	
OTTHUMP00000181220 OTTHUMP00000181218	OTTHUMP00000177815 NP_001070969.1	NP_001074025.1 OTTHUMP00000182860 OTTHUMP00000196460 OTTHUMP00000182859	OTTHUMP00000081795 OTTHUMP00000081794 OTTHUMP00000081793	
LmjF25.0960	LmjF25.0880	LmjF28.1110	LmjF30.1760	
237817	65544	105926	228176	
21602	22748	17788	32952	
56069		59637	62456 49612	
173802	123151	84810	183698	
NCU03954			NCU04535	
Os05g06450.2 Os05g06450.1			Os01g60050.1 Os01g60040.2 Os01g60040.1	
25029	37183	18472	20299	
GSPATP00017668001 GSPATP00001868001	GSPATP00020493001	GSPATP00027925001 GSPATP00027924001 GSPATP00031525001 GSPATP00013985001 GSPATP00025338001 GSPATP00039722001 GSPATP00027926001	GSPATP00035152001 GSPATP00032050001 GSPATP00034778001 GSPATP00032198001 GSPATP00003603001 GSPATP00000281001 GSPATP00008659001 GSPATP00000836001	
44225			47936	
170153 170860	80573	102190	166936 106216	
109415	136445	137588	156611	
PF08_0125		MAL13P1.245		
755367 572355			727336 550419	
RO3G_11100.1			RO3G_09527.1 RO3G_06225.1	
YLR212C			YLR045C	
SPBC32F12.04			SPCC736.14 SPCC895.07	
NP_999657.1 GLEAN3_20943	XP_001185066.1 XP_793924.1	XP_787783.1 XP_001176293.1 XP_001199653.1 XP_001183767.1	XP_001199992.1 XP_798563.2 GLEAN3_15952 XP_001182362.1 XP_001183804.1 XP_001199963.1 XP_001188972.1 XP_790495.2 XP_001188266.1	
ENSTRUP00000037747	ENSTRUP00000007528 ENSTRUP00000007527 ENSTRUP00000007529	ENSTRUP00000033492	ENSTRUP00000039530 ENSTRUP00000039528 ENSTRUP00000039529	
5.m00548	83.m00211	231.m00034	85.m00176 1.m00993 49.m00230	
29237		260819	7217	
TA03470				
TGVEG_067890 TGME49_026870 TGGT1_082260	TGVEG_028740 TGGT1_104540 TGME49_078010	TGGT1_011670 TGVEG_075940 TGME49_058530		
49610	52611	53754	51833	
83992.m00196 93106.m00014		81295.m00071	105221.m00002 88808.m00117 83992.m00172	
Tb927.3.910	Tb11.03.0280	Tb11.02.5550	Tb927.6.3090	
UM03803.1			UM06328.1	

3 Unique cilia morphology in ferns

3.1 Preface

This chapter consists of a paper that is about to be submitted for publication. The work examines ciliary ultrastructure in the fern *Ceratopteris richardii* and discusses that structure in the context of ciliary diversity in the land plants. Particular emphasis is placed on examining the regulation of basal body number, migration, positioning and structure. An analysis of *Marchantia polymorpha* cilia not included in the manuscript is included as additional information in section 3.9. The paper has 5 authors who contributed as follows: Matthew E. Hodges designed, performed and analysed the experiments, and wrote the manuscript; Michael K. Shaw provided invaluable training and expertise on EM; and Bill Wickstead, Keith Gull and Jane A. Langdale contributed to the manuscript.

3.2 Abstract

• ***Premise of study:*** **In land plants, cilia are either produced in a restricted phase of the life-cycle (sperm cells) or are completely absent (most seed-plants). The land plant phylogeny thus provides an opportunity to investigate mechanisms of developmental regulation and of evolutionary loss. However, such studies require a comprehensive understanding of cilia formation in a range of land plants. To this end, we have characterized cilia in the model fern *Ceratopteris richardii*.**

• ***Methods:*** **Scanning and transmission electron microscopy, along with electron tomography, have been used to examine the ultrastructure of sperm cell cilia.**

• ***Key results: C. richardii*** **cilia are anchored to the cell body by a novel fibrous connection between the cytoskeleton and the 'C' microtubules of the basal body. Distinct fibres also extend from the basal body through the transition zone, adjacent to each 'A' and 'B' microtubule. The basal body similarly displays distinctive characteristics in that the microtubules that form the nine triplets are different lengths both within and between triplet groups. In contrast to other cilia, a contiguous scaffold-like stellate array spans the entire triplet to doublet transition zone, and the microtubules of the central pair nucleate at different positions along the proximo-distal axis. Along this axis, the central pair twists in relation to the outer doublets of the axoneme. At the tip, organization of both the outer and central doublets breaks down to give tapered cilia ends.**

• ***Conclusions: C. richardii*** **cilia exhibit a number of distinctive morphological features, the most conspicuous of which is a general breakdown of canonical microtubule arrangements.**

Key words: Cilia; Land plants; Stellate array; Basal Body; Flagella.

Abbreviations: TEM – Transmission Electron Microscopy; SEM – Scanning Electron Microscopy; MLS – Multi-Layered Structure.

3.3 Introduction

Eukaryotic cilia/flagella are membrane-bound organelles that contain core microtubules and associated proteins, with the basic structure comprising a proximal basal body that tethers the cilium to the underlying cell and a distal axoneme that protrudes from the cell (reviewed in Dawe et al., 2007a). Although cilia structure and protein components are broadly conserved across the eukaryotic lineage, morphological variants have been observed in different biological contexts. For example, the canonical 9+2 microtubule arrangement is widespread from human (Satir and Christensen, 2007) to *Chlamydomonas* (Silflow and Lefebvre, 2001) whereas the presence of an additional paraflagellar rod structure alongside the axoneme is specific to the excavata *Trypanosoma brucei* (Vaughan and Gull, 2003). Cilia structure in the nematode *Caenorhabditis elegans* is particularly notable in that the immotile sensory cilia possess a unique 9 singlet microtubule arrangement and lack many typical cilia proteins (Inglis et al., 2007; Perkins et al., 1986c). Cilia formation in *C. elegans* is thus a classic example of divergent evolution.

The ancestor of all eukaryotes possessed cilia which functioned in both sensory (for primary cilia review see Fliegauf et al., 2007; Rosenbaum and Witman, 2002a; Satir et al., 2010) and motility (reviewed in Satir and Christensen, 2008) roles, and cilia can be found in all extant eukaryotic lineages. However, cilia loss has also occurred, with significant examples being found in both fungal and land plant lineages (Bhattacharya and Medlin, 1998; Liu et al., 2006; Wickstead and Gull, 2007). The land plants provide a good system for investigating evolutionary change in ciliary structure because both changes in complexity and outright losses have occurred. Land plant cilia are confined to sperm cells where they provide a mechanism for water-borne motility prior to fertilization (for review see Renzaglia and Garbary, 2001). Sperm cells of the earliest diverging land plants, such as those of the liverworts *Marchantia polymorpha* and *Conocephalum conicum,* are characterized by a small cell body that subtends two long free cilia (Carothers and Kreitner, 1968; Renzaglia and Garbary, 2001). In contrast, more recently diverged land plant lineages

typically have a larger number of cilia on their sperm cells. For example, the fern *Ceratopteris richardii* has sperm with 40-100 cilia (Lopez-Smith and Renzaglia, 2008) whereas the gymnosperm *Ginkgo biloba* builds up to 1000 cilia on its motile sperm cells (Vaughn and Renzaglia, 2006). In other gymnosperms and in angiosperms, however, the ability to produce cilia has been lost. The most parsimonious explanation for this pattern is one loss within the ancestor of conifers and cycads and a second independent loss at the base of the angiosperms. These losses correlate with the move away from water-borne reproduction and the evolution of the seed.

A number of excellent ultrastructural studies focusing on sperm cell structure in relation to swimming behaviour and gamete architecture in land plants have been published (Bilderback et al., 1974; Carothers et al., 1975b; Duckett and Carothers, 1982; Muccifora and Gori, 2005; Myles and Bell, 1975; Paolillo et al., 1968; Renzaglia et al., 2002; Renzaglia et al., 2001; Robbins and Carothers, 1978; Sakaushi et al., 2003). These reports have demonstrated the influence of sperm cell structure on fertilization (Lopez-Smith and Renzaglia, 2008), the role of spermatogenous microtubule-organizing centres (Hoffman et al., 1994), details of the motility apparatus (Vaughn and Renzaglia, 2006) and the relationship between male gametogenesis and cell differentiation and motility (reviewed in Renzaglia and Garbary, 2001). Further studies have highlighted the presence of novel features in Viridiplantae, such as the stellate array in the transition zone of algal and land plant cilia (first described by Lang, 1963). However, to develop a model of cilia evolution within the land plants, further in-depth studies at key phylogenetic nodes are required. Here, we contribute towards this goal by investigating cilia structure in sperm cells of the fern *C. richardii.* Using electron-tomography, scanning electron microscopy (SEM) and transmission electron microscopy (TEM) we show that cilia of *C. richardii* have several unique features.

3.4 Results

3.4.1 Novel anchoring structures are present at the cilium base

To analyze cilia structure in sperm cells of the fern *C. richardii*, scanning and transmission electron micrographs were examined. SEMs showed multiple basal bodies positioned along a distinctive ridge (Fig. 3-1A) and TEMs revealed that these basal bodies are closely associated with microtubules of the multi-layered structure (MLS; Fig. 3-1B). The orientation of basal bodies in relation to the position of underlying microtubules differed between samples, with some micrographs showing near parallel or approximately orthogonal relationships between the two structures. When the orientation of both basal body and cytoskeleton microtubules is approximately parallel, multiple independent tomographic reconstructions (e.g. Fig. 3-1C, D; Fig. 3-3B-D) showed the presence of fibres connecting the C microtubules of the triplets to the adjacent microtubules of the cytoskeleton. Distal to the basal body, cilia exhibited a transition zone containing a stellate array (Fig. 3-1E, F) and an axoneme with a canonical 9 + 2 arrangement (Fig. 3-1E, G).

3.4.2 Fibres extend throughout the basal body to the transition zone

Eukaryotic cilia contain a number of fibres, from rootlet structures to transitional fibres. To assess the structure of fibres in *C. richardii* cilia, serial sections (n=30) and electron tomographs (n=2) of the basal body (Fig. 3-2, 3) and transition zone (Fig. 3-3) were analyzed. At the most proximal end of the basal body the microtubules are surrounded by a dense amorphous regions and a number of vesicles (Fig. 3-2A, F; Fig 3-3A-D). Distal to this region, dense circular fibres with a radius of ~20 nm extend throughout the basal body to the end of the transition zone (Fig. 3-2B-J). These fibres are located in a consistent position along the proximo-distal axis, sited adjacent to the outer microtubules of every group of triplets in the basal body (Fig. 3-2D, I; white arrows) and of every doublet in the transition zone (e.g. Fig. 3-3I; black arrows). Fibres are not apparent once the transition zone has ended and the central pair has nucleated (Fig. 3-3L). In *C. richardii* cilia,

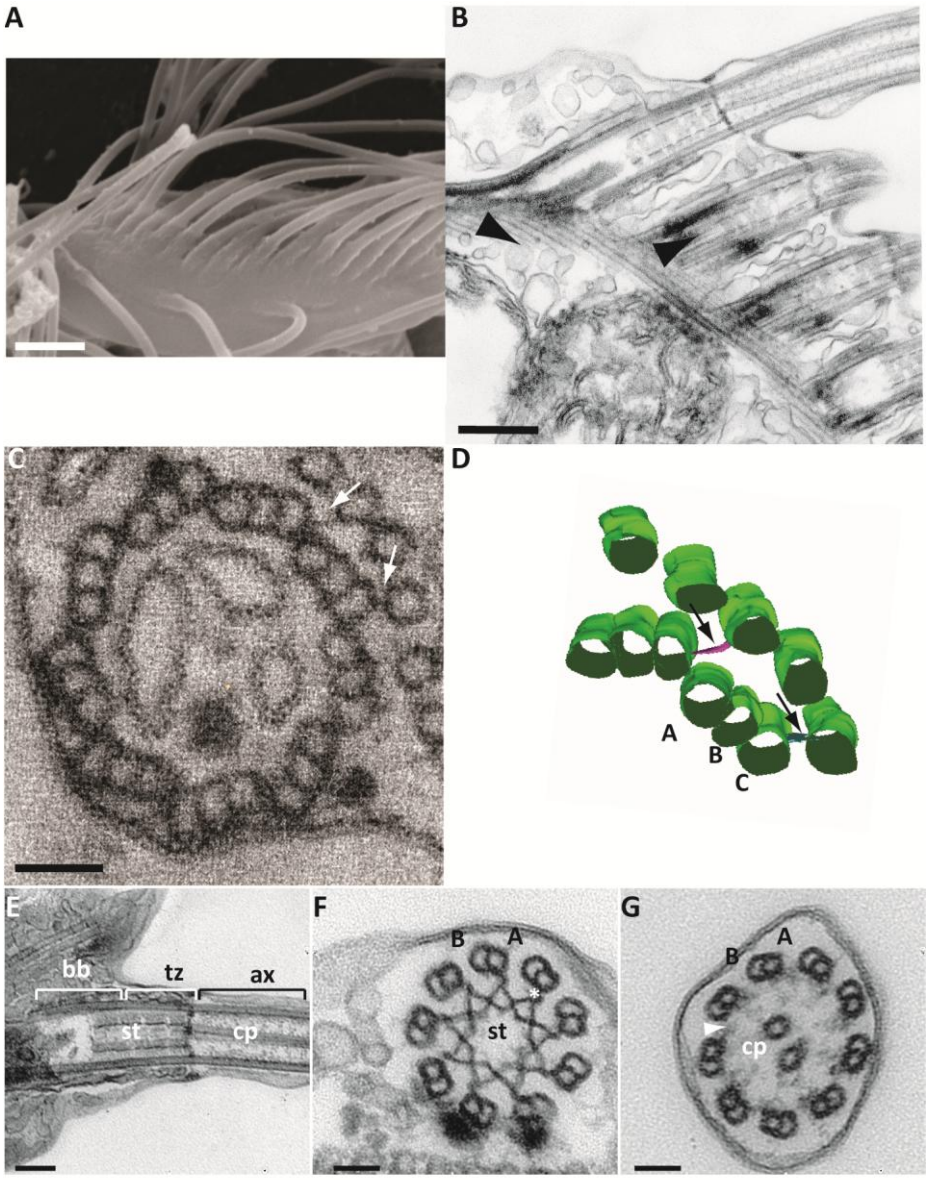


Figure 3-1: Overall cilia morphology and attachment to the sperm cell body

A) SEM of *C. richardii* sperm cell. B) TEM thin section of multiple basal bodies on a sperm cell. Cytoskeletal microtubule arrays lie in multiple directions (black arrowheads). C) TEM tomographic volume. White arrows point to fibres between 'C' microtubules of triplets and the adjacent cytoskeleton microtubules of the spline. D) Model of tomographic volume in (C) showing 'A', 'B' and 'C' microtubules. Black arrows positioned at fibres as in (C). E) Longitudinal TEM section showing the basal body (bb) and transition zone (tz) which both contain the dense parallel lines of the stellate array (st); and the axoneme (ax) which contains the central pair microtubules (cp). F, G) Transverse sections of the transition zone (F) and axoneme (G) showing the stellate array (st) and canonical microtubule 9+2 arrangement of outer doublet pairs ('A' and 'B' microtubules) and inner central pair (cp). White asterisk in (F) indicates point of contact between the stellate array and 'A' microtubules. White arrowhead in (G) points to inner dynein arms. Scale bars: A = 1 µm; B = 200 nm; C, E-G = 50 nm.

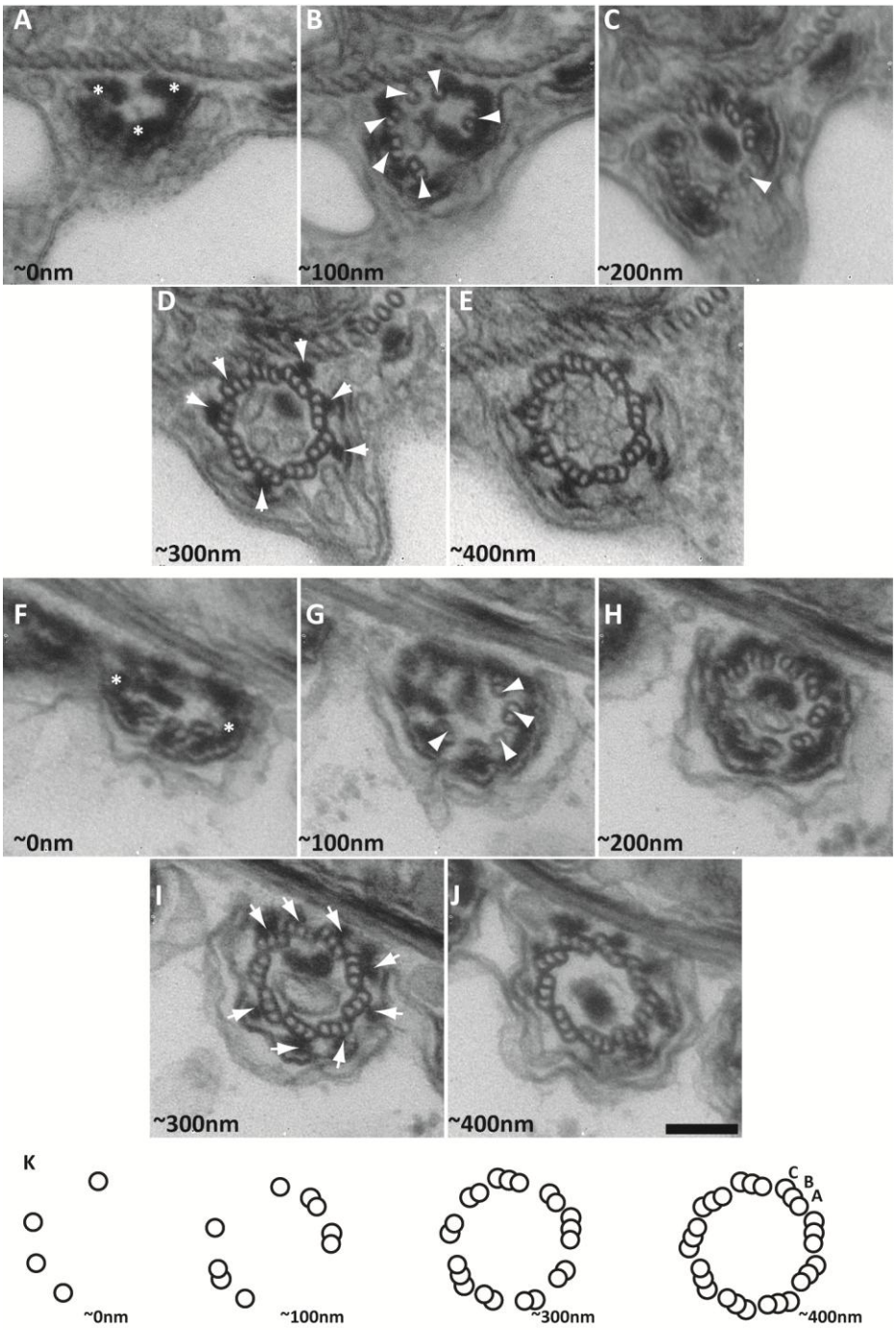


Figure 3-2: Basal body structure

A-J) Serial TEM sections of two cilia (A-E) and (F-J) showing basal body structure from proximal (0 nm) to distal (400 nm) positions. The most proximal region contains a large amount of amorphous material (A, F; white asterisks). The proximal region is composed of both singlet and doublet microtubules (B, C, G; white arrowheads), the mid-region of doublet 9-fold symmetry (H) and the most distal region of triplets (D, E, I, J). Electron dense fibres are seen adjacent to the 'A'/'B' microtubules of each triplet (D, I; white arrows). Scale bar = 100 nm. K) Diagrammatic representation of images in (A-J) showing microtubule arrangements in the basal body of *C. richardii* cilia.

fibres are thus components of the basal body and transition zone but are not present in the axoneme.

3.4.3 Basal bodies have a complex structure.

To investigate basal body organisation the proximal end of the cilium was examined by serial TEM. The axonemes of *C. richardii* cilia are subtended by canonical triplet basal bodies, possessing a symmetrical arrangement of nine triplet microtubules (Fig. 3-2D, E, I, J). However, whereas most eukaryotic basal bodies maintain this triplet arrangement to the proximal base, in *C. richardii* the proximal region contains singlet and doublet microtubules rather than triplets (Fig. 3-2B, C, G: arrowheads). Serial TEM sections (n=40) show that the transition from the 9-fold symmetry of the triplets to the region of singlets and doublets is not ordered in that 'A' and 'B' microtubules terminate at different positions (summarized in Fig. 3-2K). This leads to individual sections with different numbers of singlets and doublets (Fig. 3-2B, C, G). In combination, these observations suggest that microtubule organization degenerates at the cilium base.

3.4.4 Contiguous stellate arrays are a unique feature of Ceratopteris richardii cilia

The stellate arrays of *C. richardii* cilia display novel characteristics. Serial TEMs and electron tomography from the basal body to the axoneme (Fig. 3-3A-L) illustrate that the stellate array is between 150-200 nm in length, and that it forms a contiguous structure that stretches from within the triplet basal body (Fig. 3-3E) to the basal plate (Fig. 3-3I), without interruption by amorphous disks. Tomographic representations show that the stellate array is formed by a scaffold-like interlacing of electron-dense fibres between the 'A' microtubules of the outer doublets and triplets (Fig. 3-3E-I; M). The overall shape of the stellate array in transverse sections does not change, even when 'C' microtubules are sequentially lost from each triplet of the basal body (Fig. 3-3F-H, white arrowheads). Loss of the stellate array precedes the appearance of the central pair microtubules (Fig. 3-3I-L).

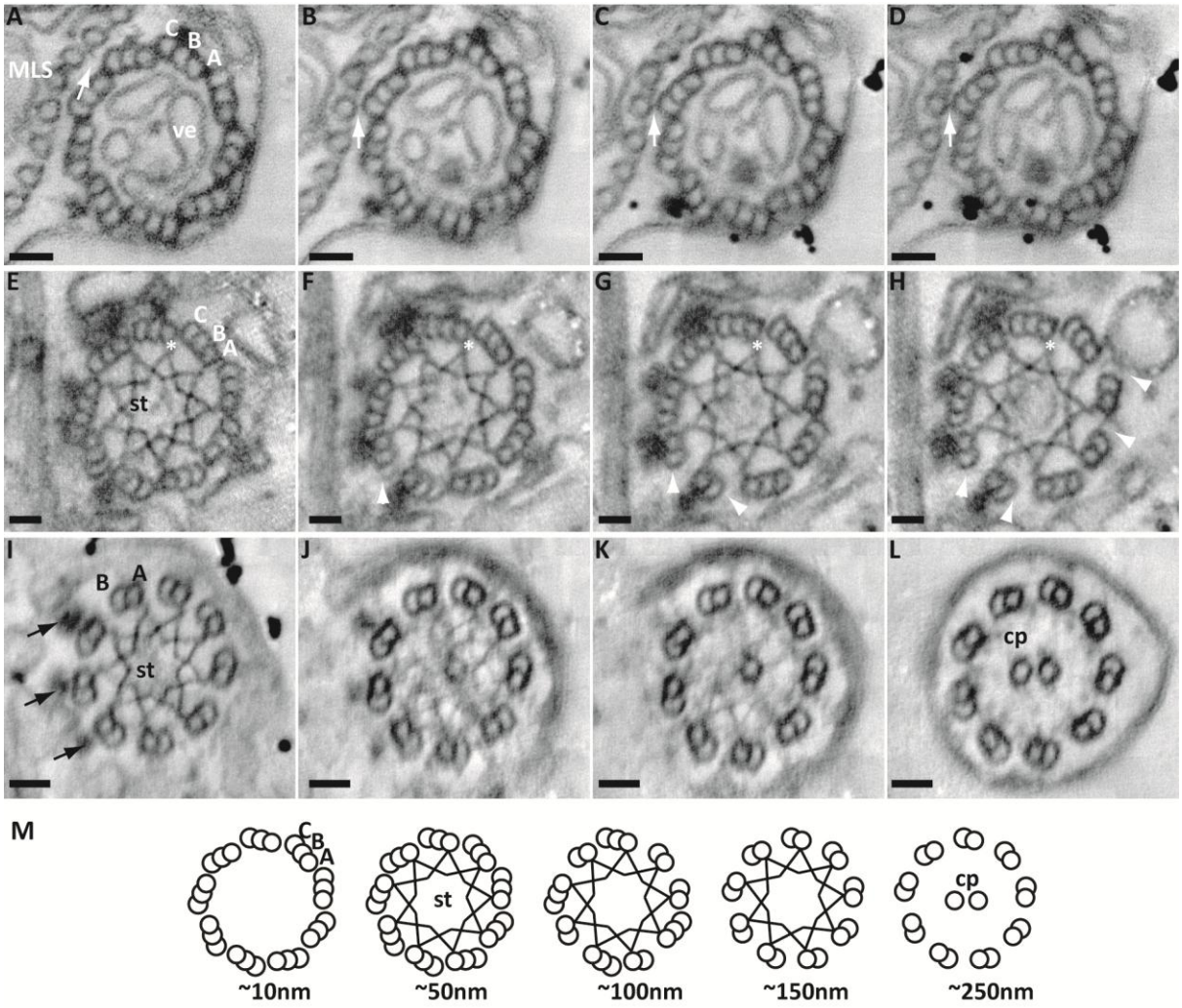


Figure 3-3: Stellate array structure

A-L) Tomographic representations of cilia from the proximal basal body (A) to the distal axoneme (L). The basal body contains 9 triplet microtubules ('A', 'B' and 'C') with prominent vesicles (ve). Fibres connect the 'C' microtubules with the multi-layered structure (MLS) of the surrounding cytoskeleton (A-D; white arrows). The stellate array (st) extends from the triplet microtubules of the basal body through the transition zone, where 'C' microtubules progressively disappear (E-H; white arrowheads). The points of the stellate array are attached to the 'A' microtubule of each triplet/doublet (E-I; white asterisks). As the stellate array ends, the microtubules of the central pair (cp) are nucleated in succession (J-L). Electron dense fibres extend adjacent to the outer triplets/doublets (e.g. I; black arrows). Scale bars = 50 nm. M) Diagrammatic representation of microtubule organization from the basal body to the axoneme.

3.4.5 Staggered central pair nucleation in the axoneme

Formation of the central pair microtubules is a core characteristic of the axoneme. However, in most eukaryotic cilia central pair microtubules are nucleated simultaneously, whereas tomographic reconstructions and serial TEMs suggest that central pair nucleation is staggered in *C. richardii*. Fig. 3-3I-L shows a representative tomographic volume spanning the distal end of the stellate array, (characterized by 9 doublet microtubules surrounding the nine-pointed star), and the axoneme, (which possesses a typical 9+2 microtubule arrangement). Immediately distal to the stellate array is a short region of axoneme without either microtubule of the central pair. The first microtubule end appears adjacent to the basal plate (Fig. 3-3J), whereas the second microtubule end is found ~20 nm distally within the axoneme (Fig. 3-3L). This difference in minus end location was found in both tomographic reconstructions (n=2) and serial sections (n=7) of the basal plate area. Central pair formation in *C. richardii* thus has distinct dynamics.

3.4.6 Central pair microtubules twist in relation to the outer doublets

To determine whether central pair positioning in *C. richardii* is fixed or variable in relation to the outer doublets, TEM serial sections of sperm cells were collected from multiple plants. Fig. 3-4A-E shows a representative series of transverse sections through a single cilium. The trajectory of the central pair (summarized in Fig. 3-4F) shows that the microtubules twist in a clockwise direction from the proximal to distal end of the cilium. All of the captured series showed the same direction of central pair twist, with an approximately even pitch of 4.5 µm (n=19). This twist in central pair configuration was also seen in tomographic reconstructions (n=4, exemplified in Fig. 3-4G).

3.4.7 Cilia tips are tapered

Ciliary tip ultrastructure was investigated by TEM of negatively-stained whole-mount cytoskeleton preparations of sperm cell populations. Negative stain micrographs show that the microtubule arrangement at the distal cilia tip differs from the normal axonemal 9+2 pattern in that the end tapers to a narrow point (Fig. 3-5A, B; seen in all 155 visible tips from 30 sperm cells).

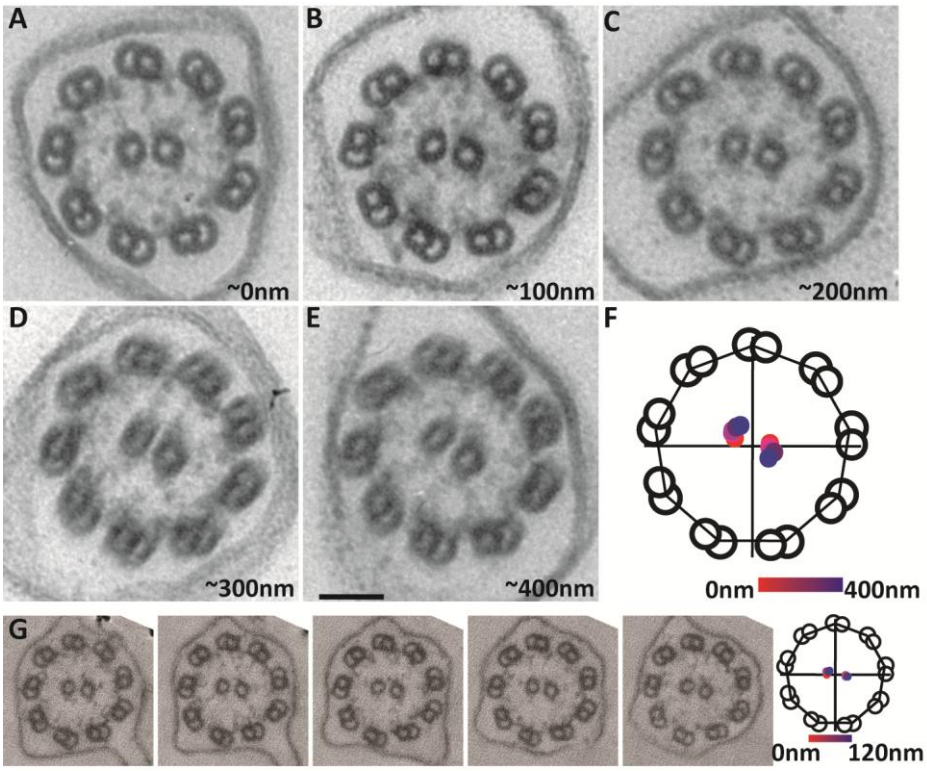


Figure 3-4: Central pair rotation

A-E) Representative TEM serial sections showing the degree of central pair movement with the relative distance separating sections indicated. Scale bar = 50 nm. F) Diagrammatic representation of serial sections in (A-E) depicting rotation of the central pair. Colours depict movement through the section. G) Representative tomographic slices through an axoneme showing clockwise twisting of central pair in relation to outer doublets.

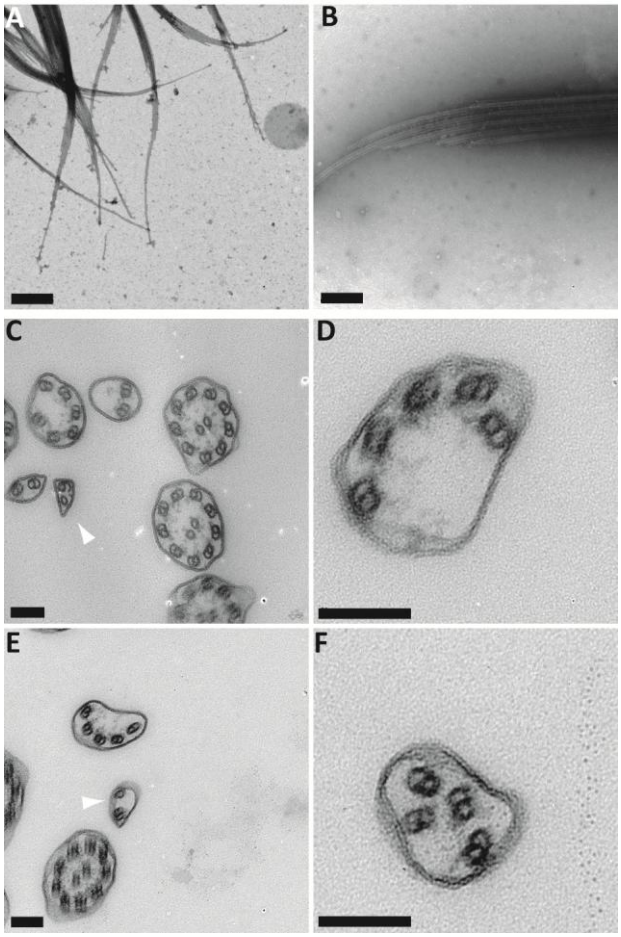


Figure 3-5: Cilia have tapered ends

A, B) Negative stain TEM of sperm cells showing that tips are tapered. C-F) Representative TEM thin sections showing various degrees of tapering at the tip. The 9+2 symmetry breaks down, with central pair microtubules lost first, and then subsequent microtubules terminating in a random order. The most distal tip of the axoneme is composed of singlet microtubules (C, E; white arrowheads). Scale bars: A = 1 µm; B = 200 nm; C-F = 100 nm.

Thin-section TEMs further show that microtubule arrangements become disorganized at the distal tip (Fig. 3-5C-F). The presence of singlet microtubules inside intact cell membranes (Fig. 3-5C, E; white arrowheads) suggests that doublet microtubules do not stop abruptly, but instead the 'B' microtubule terminates proximal to the 'A' microtubule. Central pair microtubules terminate before those of the outer doublets. While the order in which microtubules terminate cannot be defined from the micrographs, termination tends to occur in groups such that one side of the axoneme extends further than the other (Fig. 3-5D).

3.5 Discussion

Using electron microscopy and tomography we have characterized sperm cell cilia in the fern *C. richardii*. In comparison to more basal land plants such as the liverworts, *C. richardii* produces many more cilia per sperm cell and the increase in cilia number is accompanied by the elaboration of distinct structures that may be required for mechanical integrity and stability. These features include novel connecting fibres between the cilium base and a complex multi-layered structure (MLS) in the underlying cytoskeleton, fibres adjacent to the outer doublets/triplets that extend from the basal body through the transition zone, and an expanded contiguous stellate array that traverses this same basal body/transition zone region (Figs. 3-1, 2, 3). Paradoxically, these putative stabilizing properties are accompanied by the breakdown of microtubule organization at both the base and tip of the cilium (Fig. 3-2, 5). In combination these properties confer a unique cilia phenotype in *C. richardii* (summarized in Fig. 3-6).

Given that there is a general increase in cilia number and sperm cell size between green algae and gymnosperms (Carothers and Kreitner, 1968; Norstog, 1974; Vaughn and Renzaglia, 2006) selection presumably progressively favoured large multi-ciliate sperm during the first 300 million years of land plant evolution. This is consistent with the suggestion that sperm competition favours a decreased number of larger sperm (Bjork and Pitnick, 2006; Miller and Pitnick, 2002). The development of more complex structures generally necessitates more complex regulatory

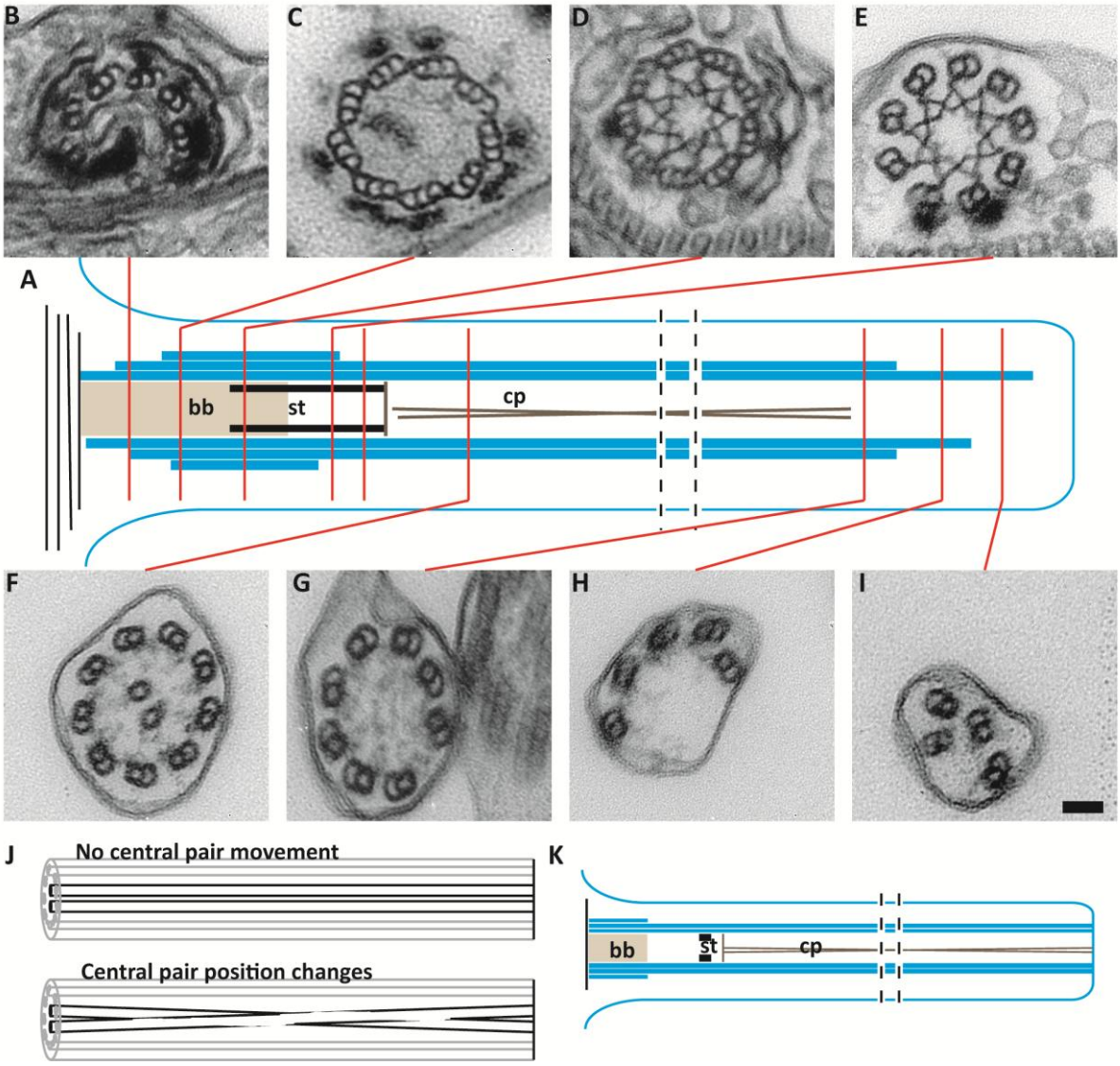


Figure 3-6: Summary of cilia structure in *Ceratopteris richardii*

A) Diagrammatic representation of a *C. richardii* cilium. Positions of the basal body (bb; shaded box), stellate array (st; horizontal thick black lines) and central pair microtubules (cp; horizontal thin grey lines) are shown. Outer microtubules (horizontal thick grey lines) are extended differentially at both the base and tip of the cilium, resulting in asymmetric patterning of triplets, doublets and singlets. Solid vertical lines represent the underlying cytoskeleton; dashed vertical lines indicate the omission of an extended section of axoneme. B-I) Representative TEM thin sections showing cilium structure at sequential positions along the proximo-distal axis. Triplet microtubules of a typical basal body (C) extend proximally in a random fashion, terminating in an amorphous base (B). Distally, the extended stellate array spans the triplet region (D), the transition from triplet to doublet region and the doublet region (E) of the basal body to axoneme transition. A canonical 9+2 axoneme extends from the transition zone throughout the majority of the cilium (F). At the tip, microtubules of the central pair terminate before the outer doublets (G), and progressive termination of the remaining microtubules then occurs (H-I). Scale bar = 50nm. J) Illustration of central pair twist in relation to outer doublets along the axoneme. K) Diagrammatic representation of *Chlamydomonas reinhardtii* cilium showing the smaller basal body organised in symmetry to the proximal base, two discrete stellate array regions, and termination of all microtubules in a cap at the distal tip. Labels as in (A).

mechanisms, however, and in the case of multi-ciliate sperm cells there is a requirement to regulate both the co-ordinated placement of cilia and the synchronization of beat patterns. The elucidation of cilia structure in *C. richardii* provides a foundation for speculating how these features evolved as sperm cell complexity increased in land plants.

In terms of cilia placement, the cytoskeletal (MLS) may provide a scaffold for the positioning and anchoring of cilia along the curved ridge of the sperm cell body (Fig. 3-1A). Furthermore, the varied alignments of the basal body and cytoskeleton (Fig. 3-1B) could generate a scaffold that was flexible enough to accommodate the beat patterns of multiple cilia. Together with the abundant connecting fibres within the basal body region (Fig. 3-2, 3), which may play a similar role to the transitional fibres in species such as the alga *Chlamydomonas reinhardtii* (Sleigh, 1979), the MLS may thus act to reinforce the periphery of *C. richardii* sperm cells, and aid placement and coordination between basal bodies.

Whereas the MLS and fibres are candidate regulators of cilia placement, the contiguous stellate array may have developed to co-ordinate beat patterns. This suggestion is consistent with the proposal that the stellate array conducts impulses that allow the pendular movement of cilia (Lang, 1963). Central pair rotation has further been suggested to facilitate changes in cilia beat on the basis that the central pair microtubules of *C. reinhardtii* cilia twist with a pitch of ~2-5 µm during cilia bending (Chasey, 1974; Mitchell and Nakatsugawa, 2004; Mitchell, 2005). Given that *C. richardii* cilia have an extended stellate array (Fig. 3-1, 3, 6) in comparison to algae and bryophytes (Cavalier-Smith, 1974b; Geimer and Melkonian, 2004; Lang, 1963; O'Toole et al., 2003; Sanders and Salisbury, 1989), and that the central pair twists with a pitch at the upper end of that seen in *C. reinhardtii* (Fig. 3-4), it is plausible that both features were modified to facilitate the co-ordinated swimming of multiple cilia.

Whilst complexity initially increased during land plant evolution, ultimately cilia were lost. This loss could have been mediated by direct loss of a master regulator or by progressive

deterioration. *C. richardii* cilia exemplify both increased complexity and a degree of degeneration. This degeneration includes the lack of ciliary caps, which play a microtubule stabilizing role and are typically present at the tip of the axoneme (Dentler, 1980a, b), and staggered central pair minus end positions. Although similar differences in the minus end position of the central pair are also seen in species such as *Tetrahymena* and *Paramecium* (Dute and Kung, 1978; Messaoudi et al., 2006), it is more typical for the minus ends of both central pair microtubules to be embedded in the basal plate as in *C. reinhardtii* (Cavalier-Smith, 1974b; Geimer and Melkonian, 2004; O'Toole et al., 2003). The lack of microtubule symmetry at both the ends of the *C. richardii* cilium may reflect loss of regulatory protein function in these sperm cells, consistent with bioinformatic analyses that show key ciliary protein components have been lost within the land plants (Hodges et al., 2010).

The inability to produce cilia in the majority of seed plants has an interesting parallel in the evolution of ciliary structure in dipteran *Drosophila* lineages. Some fly species produce giant, relatively immotile sperm cells (Bjork and Pitnick, 2006; Lopez-Fernandez et al., 2007; Phillips, 1970) that can be seen to be analogous to the large, multi-ciliated sperm of *Ginkgo*. The cilia of these insect sperm cells develop additional outer singlet microtubules or accessory fibres that may act to rigidify the structure and decrease motility (Bjork and Pitnick, 2006; Lopez-Fernandez et al., 2007; Phillips, 1970). The trajectory suggested by our analysis of cilia in fern sperm cells suggests that a similar strengthening and reduction of flexibility also occurred during land plant evolution. The result of such an evolutionary trend may have been a decreased selective pressure for cilia retention, which in combination with the push away from water-borne reproduction, resulted in total loss. Future analyses of more land plant species will enable this hypothesis to be tested further.

3.6 Materials and Methods

3.6.1 Plant material

Gametophytes of *Ceratopteris richardii* were grown from spores using sterile culture techniques, as described by Warne et al., (1986). Spores were obtained from Carolina Biological Supply Company (Burlington, North Carolina, USA), sown on agar media as described by Klekowski, (1969) and maintained at 29^{o}C under continuous light for 2 weeks until they reached sexual maturity. Male gametophytes were placed in a microfuge tube and 0.5 ml distilled water was added to stimulate sperm release. Electron microscopy was performed on sperm samples collected from more than 10 plants that were grown over a period of months.

3.6.2 Electron microscopy

For thin-section electron microscopy, swimming sperm cells were first fixed at room temperature (2.5% glutaraldehyde, 2% paraformaldehyde in 100 mM cacodylate buffer (pH 7.0) with 1 mM ethylene glycol tetraacetic acid (EGTA) and 2 mM $MgCl_2$ and picric acid) for 10 mins. The fixative was replaced with fresh fixative and left for 16-24h at 4^{o}C. After the primary fixation the samples were washed with cacodylate buffer (0.2 M, pH 7.4) and post-fixed with 1% osmium tetroxide in 100 mM cacodylate buffer (pH 7.0) for 1h at 4^{o}C followed by several rinses in distilled water. En bloc stain with 0.5% aqueous uranyl acetate for 16 h at 4 ^{o}C in dark. The sample was then moved through a dehydration gradient (30%, 50%, 70% and 90% EtOH) for 15 mins each step and then 100% EtOH for 3x 30 mins. After dehydration the embryo is passed through 2x15 mins changes of Propylene oxide and then embedded in resin (Agar 100 resin) through a series of steps. Infiltration of the resin begins with a 2:1 mix of propylene oxide:resin, then a 1:1 mix of propylene oxide:resin then a 1:2 mix of propylene oxide:resin which all stand for 1 hr. The sample was left overnight in 100% resin and then changed into fresh resin for 1 hr. The fresh resin was transferred to an embedding mould for 12-24 hrs at 60-70^{o}C. Resin containing specimen was sectioned using a diamond knife on an ultra-microtome. 50-70nm sections were cut, placed on a grid, dried and

then covered in 0.2% lead citrate for 30 secs before they were viewed using a transmission electron microscope (FEI Technai).

For scanning electron microscopy (SEM), samples were fixed at room temperature (2.5% glutaraldehyde, 2% paraformaldehyde in 100 mM cacodylate buffer (pH 7.0) with 1 mM ethylene glycol tetraacetic acid (EGTA) and 2 mM MgCl2, 0.01% picric acid and 100mM sucrose) and left in the fixative for 1-2 days at 4°C. After the primary fixation the samples were washed in cacodylate buffer (0.2 M, pH 7.4) and post-fixed with 1% osmium tetroxide in 100 mM cacodylate buffer (pH 7.0) followed by several rinses in cacodylate buffer. The samples were washed with distilled water to remove fixative/buffer. The samples were dehydrated with ethanol to 100% EtOH, critically point-dried for 10 mins, and then mounted on aluminium SEM stubs using double-sided carbon tabs. After sputter-coating with gold the samples were examined using a scanning electron microscope (JOEL JSM 6390).

3.6.3 Electron tomography

For cellular electron tomography, sperm cells were fixed and embedded as for thin-section microscopy. Ribbons of serial ~250 nm sections were collected on formvar coated copper-rhodium slot grids and post-stained with aqueous uranyl acetate and Reynold's lead citrate (Reynolds, 1963). Colloidal gold particles (10 nm) were deposited on both surfaces for use as fiducial markers during image alignment. Sections were viewed on a FEI Tecnai-T12 electron microscope operating at 120 KeV and recorded digitally. Tilt series were recorded as the sample was serially tilted by 1° increments over a range of 130° (±65°) about two orthogonal axes using the microscope control program SerialEM (Mastronarde, 2005). Tomograms were calculated using the IMOD software package and models made in 3Dmod (Kremer et al., 1996). The eTomo software suite enables course image alignment, fiducial model generation for fine alignment and tomogram combination.

3.7 Additional Information

3.7.1 Stellate arrays of Marchantia polymorpha

Negative stain and scanning electron microscopy demonstrate that the liverwort *Marchantia polymorpha* produces a sperm cell possessing two long free cilia (Fig. 3-7A, B; white arrows indicate sperm cell body). Examination of the ciliary structure by transmission electron microscopy shows that the axonemes of *Marchantia* cilia possess an ultrastructure composed of nine radially symmetrical microtubule doublets, and a central pair (Fig. 3-7E).

Outer dynein arms are absent from the axonemal structure, as is the case in all land plants. However, when compared to the ultrastructure of other bryophytes, or indeed the fern *Ceratopteris*, the radial spokes in *Marchantia* cilia consistently stain more electron-dense (compare Fig. 3-3G with Fig.3- 7E). This suggests that *Marchantia* radial spokes may be biochemically distinct.

The stellate array is maintained in both *Marchantia* (Fig. 3-7D) and *Ceratopteris* (Fig. 3-3F) and as in *Chlamydomonas reinhardtii*, each point of the array is linked to the A microtubule of each outer microtubule pair. However, longitudinal sections (Fig. 3-7C) show differences in length and positioning of the stellate array between *Chlamydomonas* and *Marchantia*, and between *Marchantia* and *Ceratopteris.* These data show that the stellate array in *Marchantia* is restricted to the doublet microtubules of the transition zone, as is the case in *Chlamydomonas.* Whereas in *Chlamydomonas* the stellate pattern forms in two discrete positions (O'Toole et al., 2003), however, in *Marchantia* the array is contiguous. When compared to the contiguous stellate array of ferns, this expanded stellate array appears to be an intermediate form in the stepwise progression of land plant cilia evolution.

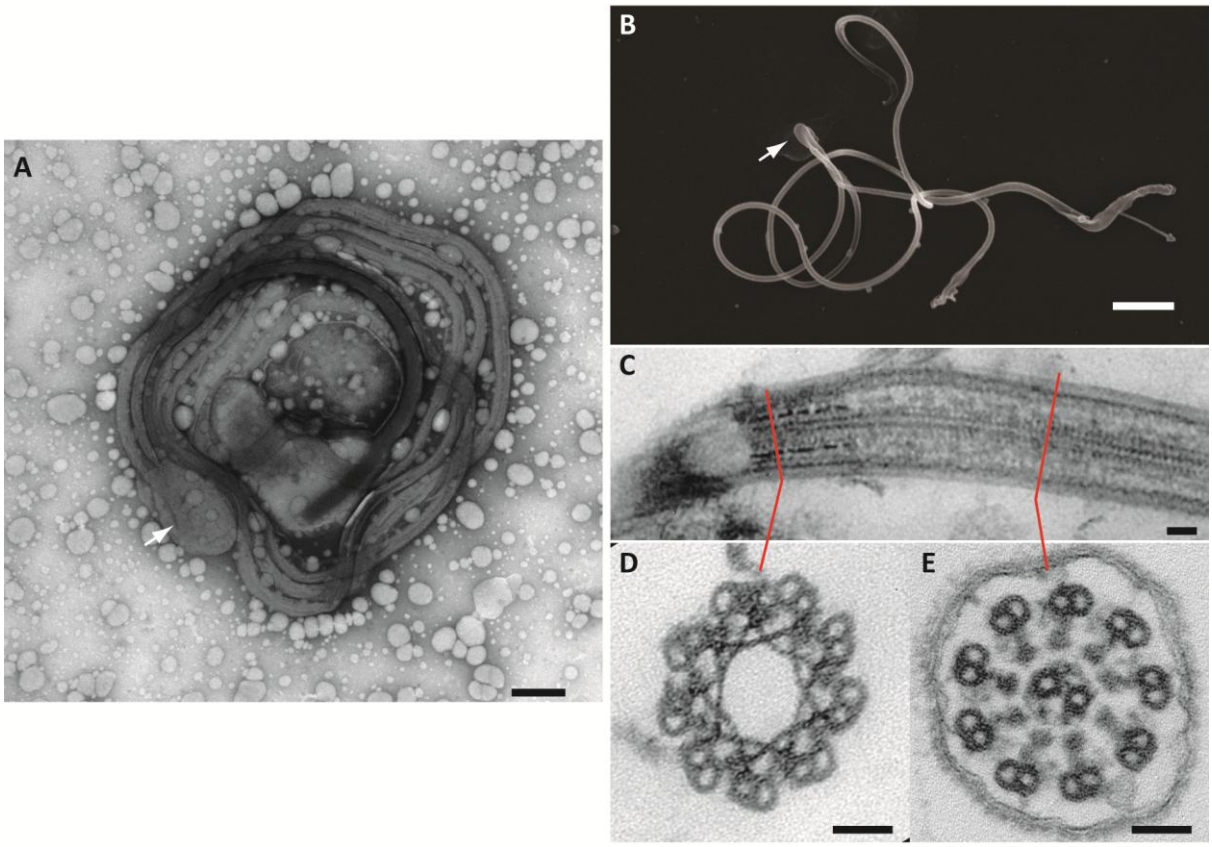


Figure 3-7: Ultrastructure of *Marchantia polymorpha* cilia

A) Negative stain of *Marchantia* sperm cell showing two long free cilia and the sperm body (white arrow). B) SEM micrograph showing free swimming sperm cell. C) Longitudinal TEM micrograph showing ciliary structure. The stellate array is formed in the transition zone (D). The axoneme has a typical 9+2 microtubule arrangement, and radial spokes are densely stained. Outer dynein arms are absent. Scale bars: A) 1 µm; B) 2 µm; C-E) 50 nm.

4 Conservation of ciliary proteins in plants with no cilia

4.1 Preface

The data presented in this chapter are in the form of a paper that has been submitted for publication in BMC Plant Biology. The work aims to provide insight into how organelle protein composition is affected during the evolutionary transitions preceding and resulting in the loss of an organelle. The paper has four authors, who contributed as follows: Matthew E. Hodges designed, performed and analysed the experiments, and wrote the manuscript; Bill Wickstead, Keith Gull and Jane A. Langdale contributed to/reviewed the manuscript.

4.2 Abstract

Background: Eukaryotic cilia are complex, highly conserved microtubule-based organelles with a broad phylogenetic distribution. Cilia were present in the last eukaryotic common ancestor and many proteins involved in cilia function have been conserved through eukaryotic diversification. However, cilia have also been lost multiple times in different lineages, with at least two losses occurring within the land plants. Whereas all non-seed plants produce cilia for motility of male gametes, some gymnosperms and all angiosperms lack cilia. During these evolutionary losses, proteins with ancestral ciliary functions may be lost or co-opted into different functions.

Results: Here we identify a core set of proteins with an inferred ciliary function that are conserved in ciliated eukaryotic species. We interrogate this genomic dataset to identify proteins with a predicted ancestral ciliary role that have been maintained in non-ciliated land plants. In support of our prediction, we demonstrate that several of these proteins have a flagellar localisation in protozoan trypanosomes. The phylogenetic distribution of these genes within the land plants indicates evolutionary scenarios of either sub- or neo-functionalisation. Expression data analysis shows that these genes are highly expressed in *Arabidopsis thaliana* pollen cells.

Conclusions: A large number of proteins possess a phylogenetic ciliary profile indicative of ciliary function. Remarkably, upon loss of cilia, many genes with an ancestral ciliary role are maintained in non-ciliated land plants. These proteins have been co-opted to perform novel functions, some of which appear related to the formation of the male gametes.

Key words: Evolution, Cilia, Flagella, Basal Body, Centriole, Land Plants

Abbreviations: LECA – Last Eukaryotic Common Ancestor; BLAST – Basic Local Alignment Search Tool; RBB – Reciprocal Best BLAST; CCP – Conserved in Ciliated and Plants.

4.3 Introduction

Centrioles and cilia/flagella are microtubule-based organelles involved in cellular motility and signalling (for review see Mitchell, 2007). The ultrastructural morphology of these ancient organelles is remarkably conserved in extant eukaryotes and their phylogenetic distribution pattern suggests that the last eukaryotic common ancestor (LECA) possessed the ability to produce a centriole plus cilia with both sensory and motility functions (Cavalier-Smith, 1974a; Hodges et al., 2010; Pereira-Leal et al., 2010). Despite this widespread phylogenetic distribution, lineage specific modifications have been shown to occur and numerous instances of independent cilia loss have been reported (Broadhead et al., 2006; Cavalier-Smith, 1987; Inglis et al., 2007; Satir and Christensen, 2007; Wickstead and Gull, 2007).

In recent years, multiple high-throughput proteomic studies and bioinformatic analyses in disparate species have identified a cohort of proteins associated with centriole and ciliary functions (Hodges et al., 2010; Inglis et al., 2006; Keller et al., 2005; Keller et al., 2006; Kilburn et al., 2007; Merchant et al., 2007b; Pazour et al., 2005; Pereira-Leal et al., 2010; Woodland and Fry, 2008). These proteins include both those with ciliary-specific roles (such as intraflagellar transport proteins, outer- and inner-dynein arms and radial spoke proteins) as well as those such as tubulins that are also involved in other cellular functions (for an extensive database see Cildb; Arnaiz et al., 2009). To date, however, little is known about what happens to genes involved in ciliary function when an evolutionary transition to cilia loss occurs.

The land plant lineage is a good model for studying the transition to ciliary loss for several reasons. First, this ancient monophyletic group was ancestrally ciliated, but there have been at least two independent loss events within the group, once in gymnosperms and once at the base of the angiosperms (Bremer et al., 1987; Finet et al., 2010; Renzaglia and Garbary, 2001). Second, sufficient genomic information exists for an in-depth analysis of protein compositional changes during the process of ciliary loss. Third, the land plants are a sister lineage to the Chlorophyta, a

group that includes the well-studied ciliary model species *Chlamydomonas reinhardtii*. The Chlorophytes thus provide a good outgroup for identification and analysis of genes with ciliary function.

Within the ciliated basal land plants, cilia are produced only in specialised sperm cells (Cavalier-Smith, 1974a). Ciliogenesis in these cells occurs *de novo*, as opposed to the canonical template pathway seen in animal lineages (Beisson and Wright, 2003). It can thus be assumed that there are regulatory mechanisms that ensure correct spatial and temporal expression of genes required for ciliary function within this restricted phase of the plant life cycle. On losing the ability to produce cilia, it is unclear whether these regulatory mechanisms are also lost.

Here, we describe an approach that uses phylogenetic profiling to identify a core set of ciliary function proteins. By applying a scoring system to orthologous sets of proteins within 45 eukaryotic species we aimed to identify proteins with an ancestral ciliary function that are maintained in non-ciliated land plants. We then tested this ancestral ciliary function in the flagellated protozoan *Trypanosoma brucei*, and investigated potential plant roles through analysis of gene expression data and phylogenetic inference. We conclude that at least some of the genes identified in cilia formation in non-seed plants retain a role in male gametophyte development in seed plants.

4.4 Results

4.4.1 Identification of ciliary proteins by phylogenetic distribution

To gain a deeper understanding of the key proteins involved in ciliary function across eukaryotes, we analysed the distribution of orthologous proteins encoded in the genomes of 45 species (Table 2-S1, see section 2.9). The species were selected to represent a wide evolutionary spread across six major groups of eukaryotes (namely: Archaeplastida, Excavata, Chromalveolata, Holozoa, Fungi and Amoebozoa). For each species, a complete or near-complete genome sequence is publically available. The species analysed include organisms that produce cilia/flagella

at some point in their life cycle (which we refer to here as “ciliated species”) and those that do not (referred to here as “non-ciliated species”). Proteins which are phylogenetically conserved in ciliated, but not in non-ciliated species, are considered to possess a “ciliary profile” and are predicted to perform a ciliary function.

To identify proteins with a ciliary profile, we performed a reciprocal best BLAST (RBB) analysis using the predicted proteome of *Chlamydomonas reinhardtii* to identify putative orthologues encoded in the genomes of 44 other eukaryotic species. The green alga *Chlamydomonas reinhardtii* was chosen as the query genome due to its phylogenetic position in the sister lineage to (at the base of) the land plants. The relatively short phylogenetic distance to the land plants reduces the chance that sequence divergence would prevent accurate identification of orthologues. Furthermore, *Chlamydomonas* is a well-studied ciliary model species.

The predicted proteome of *Chlamydomonas* has 15,143 proteins. Of these, 6,982 proteins have more than one putative orthologue. To reduce the noise of false positives in the dataset, only proteins with identifiable hits in at least 5 species within the analysis were retained. This cut-off resulted in a final defined dataset of 4,802 proteins (31.7% of the proteins in the *Chlamydomonas* predicted proteome; Table 4-S2, see section 4.9). Within this dataset, 772 proteins (16.1%) have a putative orthologue in each of the major eukaryotic groups used in this analysis. The most parsimonious explanation for the phylogenetic distribution of these 772 proteins is that these proteins were present in the last eukaryotic common ancestor (LECA), and that they have subsequently been lost in certain lineages through evolutionary divergence.

We found that a simple presence or absence analysis of protein phylogenetic distribution in ciliated and non-ciliated species was too stringent to identify proteins with ciliary function because of false negatives. Therefore, we allowed for both false positives and negatives in the dataset by applying a positive score for a hit in a ciliated species, and a negative score for a hit in a non-ciliated species (see Materials and Methods). By using this scoring system we found that

out of the 4,802 proteins in the dataset, 213 (4.4%, Table 4-S3, see section 4.9) proteins possess a score indicative of a ciliary profile (examples in Figure 4-1A, B). As an indication that the method to identify ciliary proteins is accurate, 157 of these 213 (73.7%) proteins have been previously detected within cilia in proteomic analyses. This equates to 33.3% of all proteins (472) found in proteomic analyses of cilia. The remaining 56 proteins with a ciliary profile currently have no known function. We hypothesise that these represent a novel set of proteins with an as yet uncharacterised ciliary function.

4.4.2 Ciliary profile proteins localise to the flagellum of trypanosomes

Most of the ciliary profile proteins have been detected in one or more proteomic studies of cilia (Andersen et al., 2003; Inglis et al., 2006; Keller et al., 2005; Pazour et al., 2005). To test whether the 56 proteins that have not been previously identified in proteomic studies are likely to play a role in cilia function, we investigated the cellular localisation of one of these proteins in the single-celled excavate *Trypanosoma brucei*. *T. brucei* is a notable model organism for cilia studies, with many examples of transgene introduction for localisation studies. Many of the ciliary profile proteins with known function have already been localised in *T. brucei* (for review see Bastin et al., 2000).

We selected a candidate protein, FBB17, which has a putative trypanosome orthologue (herein called TbFBB17) encoded by the gene Tb10.406.0130, for endogenous locus *TY-YFP*-tagging in *T. brucei*. TbFBB17 is a putative zinc carboxypeptidase which has not been detected in any cilia proteomic study to date (Andersen et al., 2003; Arnaiz et al., 2009; Inglis et al., 2006; Keller et al., 2005; Pazour et al., 2005). The tagging strategy ensured that the YFP-TbFBB17 chimera is transcribed by the endogenous gene promoter. This strategy has previously been shown to produce protein levels similar to wild-type for other loci (Wickstead et al., 2010). The correct incorporation of the *YFP*-tag at the 5' end of the targeted Tb10.406.0130 gene was checked by

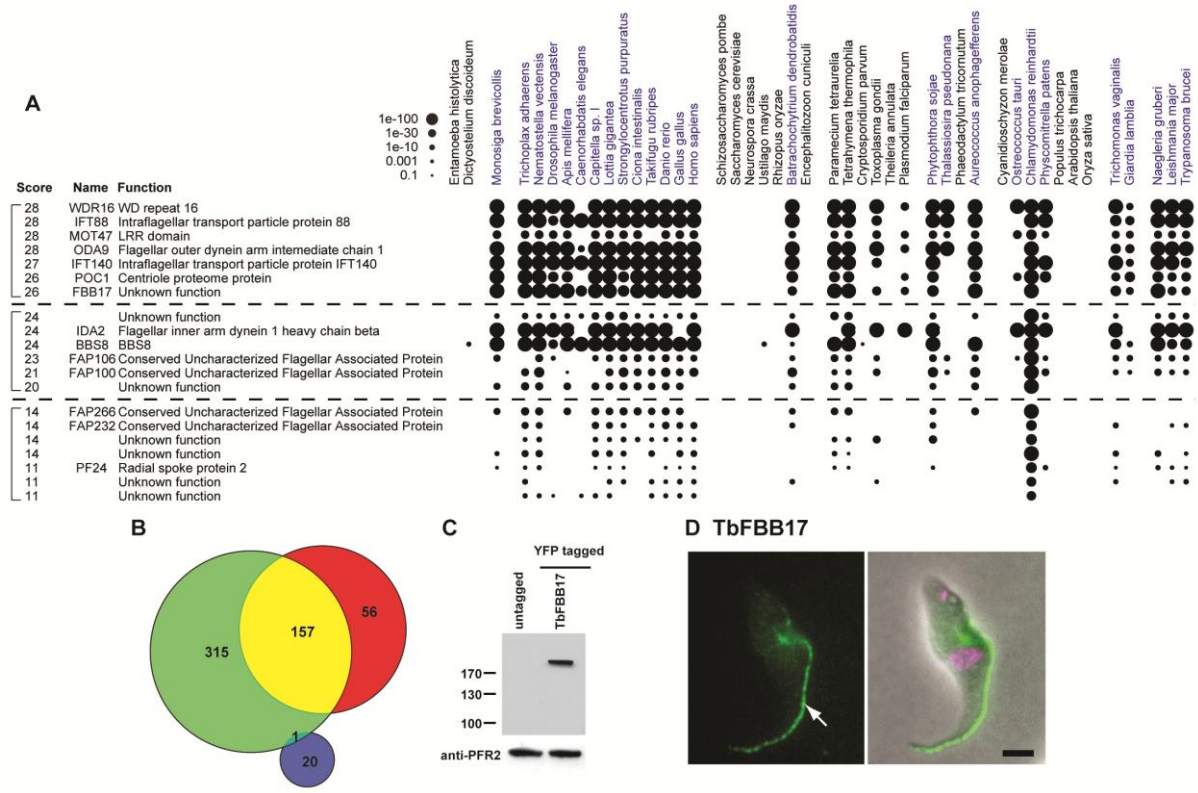


Figure 4-1: Ciliary profile proteins

A) Phylogenetic distribution of representative examples of the filtered ciliary profile proteins from the reciprocal best BLAST analysis with different scores. Ciliated species are shown in blue. B) Venn diagram depicting the results of the bioinformatic analysis. A total of 4,802 proteins have five or more identifiable orthologues in 45 queried genomes. 472 of these proteins are detected in ciliary proteomic screens (green and yellow). 213 proteins are identified with the ciliary profile (red and yellow), of which 157 are detected in proteomic screens (yellow). 21 proteins have been identified as conserved in ciliated species and in the land plants (CCP, blue), of which one has been identified in proteomic screens. C) Western blot of whole cell lysates with anti-TY BB2 antibody (above) and loading control anti-PFR2 antibody (below). D) Direct fluorescence of whole cell *Trypanosoma brucei* showing cellular localisation of TbFBB17 (Tb10.406.0130) N-terminal YPF chimera (left panel) and an overlay of signal from YFP (green), DAPI (magenta) and phase-contrast (right panel). Tb10.406.0130 localises to the flagellum (white arrow). Scale bar = 2 µm.

PCR (data not shown) and by immunoblotting of cell lysates with monoclonal antibodies directed against the short epitope TY-tag (Bastin et al., 1996; Figure 4-1C).

To determine the intracellular localisation of TbFBB17, transformed cells were examined by fluorescence microscopy. Figure 4-1D shows that TbFBB17 is found along the length of the flagellum (white arrow). This observation supports the prediction that the identified ciliary profile proteins do indeed play a role in ciliary function. To determine whether TbFBB17 is stably associated with the cytoskeleton, we assessed the localisation of the tagged protein in cells which had been detergent-extracted to remove non-cytoskeletal components. In such cells the TbFBB17 flagellum signal is removed (data not shown). This explains the lack of identification of TbFBB17 in the trypanosome cilia proteome as the proteomes are derived from cytoskeletal cell fractions (Broadhead et al., 2006). A similar explanation may apply to the other 55 proteins that have been identified in this study but not in ciliary proteomic screens.

4.4.3 Identification of ciliary proteins in non-ciliated land plants and fungi

In land plants cilia have been lost at least twice, once in the gymnosperms and once at the base of the angiosperms (Bremer et al., 1987; Finet et al., 2010; Renzaglia and Garbary, 2001). Such evolutionary transitions pose the question of what happens to ancestral ciliary proteins when the ability to produce cilia is lost. To investigate this phenomenon, a scoring system was defined that identified proteins with putative orthologues conserved in both ciliated species and also land plants, but not in other non-ciliated species (CCP, for conserved in ciliated and plants, see Materials and Methods). This analysis identified 21 CCP proteins, out of the dataset of 4,802 proteins (0.44%), that have been retained in land plants which do not possess the ability to produce a cilium (Figure 4-2A; Table 4-S4, see section 4.9). In agreement with our predictions, one of these proteins (BUG22) has previously been detected in ciliary proteomic studies (Keller et al., 2005; Laligne et al., 2010). The remaining proteins have no characterised ciliary function. A similar analysis within the non-ciliated fungal lineages (present in ciliated species and non-ciliated

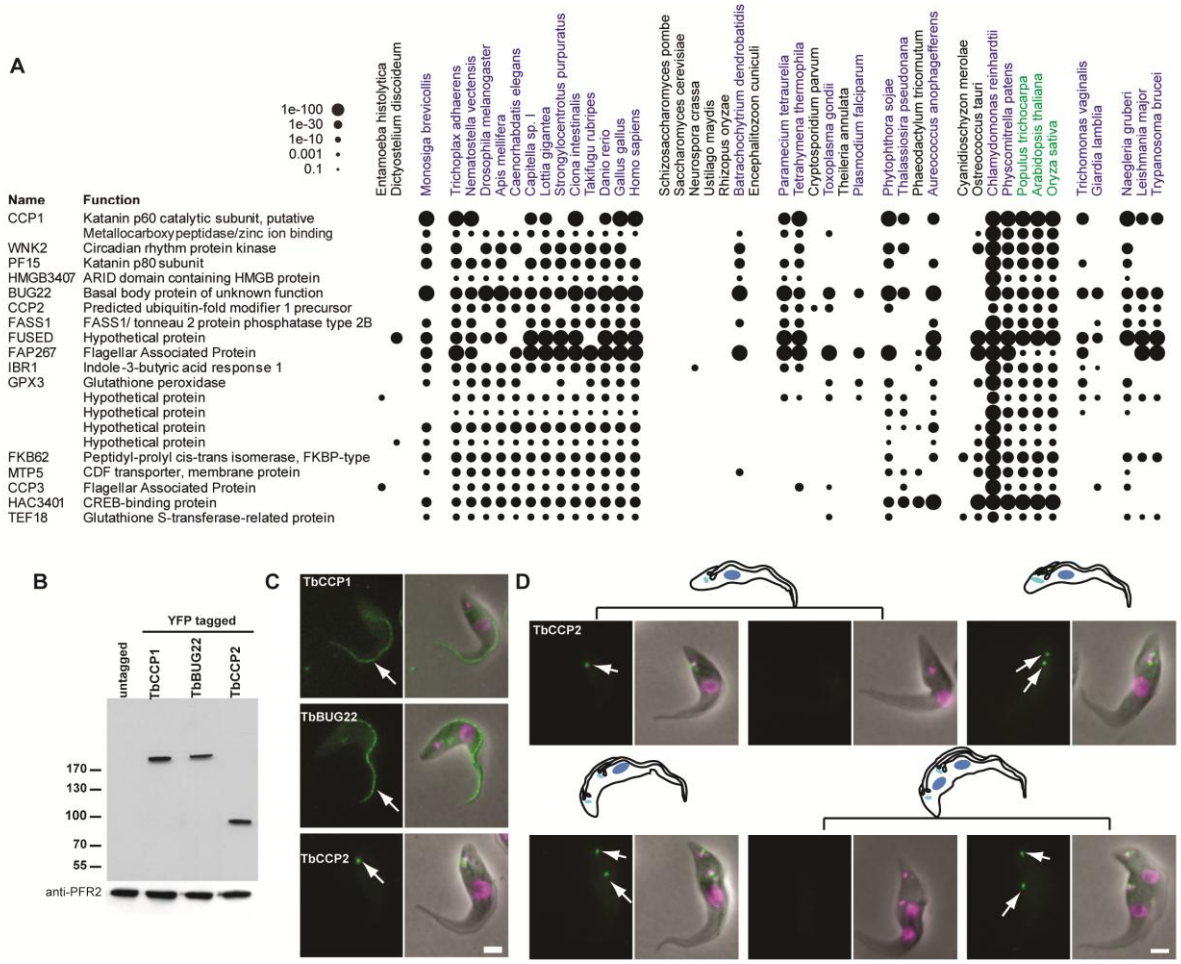


Figure 4-2: CCP proteins

A) Phylogenetic distribution of the filtered conserved in ciliated species and non-ciliated land plants, but absent in other non-ciliated species (CCP) proteins from the reciprocal best BLAST analysis. Ciliated species are shown in blue and non-ciliated land plants in green. B) Western blot of whole cell lysates with anti-TY antibody (above) and loading control anti-PFR2 antibody (below). C) Direct fluorescence of whole cell *Trypanosoma brucei* showing cellular localisation of TbCCP1 (Tb11.02.1370), TbBUG22 (Tb10.70.5560) and TbCCP2 (Tb927.8.5380). Left panel showing fluorescence signal from YFP and an overlay in right panel of signal from YFP (green), DAPI (magenta) and phase-contrast. All three possess a ciliary localisation (white arrows point to flagella (CCP1 and BUG22) and basal bodies (CCP2)). D) Direct fluorescence of whole cell *Trypanosoma brucei* showing cellular localisation of TbCCP2 (Tb927.8.5380). Expression is cell-cycle regulated (white arrows depict basal bodies). Scale bar = 2 µm.

fungi) identifies two proteins (0.04% of the 4,802 proteins in the dataset). Neither of these two proteins are present in the CCP set, and neither have been detected in any proteomic study. The lower number of proteins identified in the fungal analysis may reflect the high level of genomic streamlining in this lineage (Soanes et al., 2008).

4.4.4 CCP proteins localise to the trypanosome flagellum

To test the hypothesis that CCP proteins possess an ancestral ciliary function, localisation of three proteins from this subset was examined in *T. brucei* cells. These three proteins, named TbCCP1 (the putative p60 subunit of katanin), TbBUG22 (the trypanosome orthologue of BUG22, and TbCCP2 (a proteins of hypothetical function) are encoded by genes Tb11.02.1370, Tb10.70.5560 and Tb927.8.5380, respectively. Orthologues of CCP1 and CCP2 have not been detected in any previous ciliary proteomic study, whereas BUG22 has been identified in a proteomic analysis of *Chlamydomonas* cilia (Keller et al., 2005). The localisation of each protein was investigated using endogenous locus N-terminal *TY-YFP* tagging as for TbFBB17. Correct insertion of the tag was checked by PCR (data not shown) and immunoblotting of cell lysates with monoclonal antibodies directed against the TY tag (Figure 4-2B).

Fluorescence microscopy of transformed cells demonstrated that all three proteins have a localisation pattern consistent with a role in ciliary function. Both TbCCP1 and TbBUG22 localise to the flagellum of *T. brucei* throughout the cell cycle (Figure 4-2C, white arrows). Interestingly, TbCCP2 was found to localise to a position consistent with the basal body of the flagellum in a cell-cycle regulated fashion (Figure 4-2C, D, white arrows). Trypanosomes undergo morphological changes which can be used to position cells within the cell cycle (see cartoons in Figure 4-2D). Early in the cell cycle, the pro-basal body matures to form a second basal body from which a new flagellum will extend. Shortly thereafter, the kinetoplast (an organelle containing all the mitochondrial DNA) divides, coupled to the separation of the two basal bodies and extension of a new flagellum from the posterior basal body. Once the new flagellum is fully extended, nuclear

mitosis occurs and finally cytokinesis. Examination of TbCCP2 fluorescence in the light of this positional information reveals that TbCCP2 is not detected just before kinetoplast division, nor during nuclear division, but it is present on or near the basal bodies throughout the rest of the cell cycle (Figure 4-2D). In conclusion, these data demonstrate that the identified CCP proteins do indeed perform a ciliary function in a species which can produce flagella.

4.4.5 Evolutionary history of CCP proteins

To investigate possible functions of the CCP proteins in non-ciliated land plants we performed phylogenetic analysis of the identified protein sets. Understanding the evolutionary relationships of the CCP proteins provides insight into possible mechanisms of neo- or subfunctionalisation. Bayesian phylogenetic inference of the identified homologous proteins of BUG22 (Figure 4-3A) suggest that the gene duplicated at the base of the land plants because *Physcomitrella patens* a basal land plant with the ability to produce cilia, possesses four copies of the gene that resolve in 2 clades. Subfunctionalisation may then have occurred with one copy retaining the ancestral ciliary role and the other being co-opted for alternative roles. Subsequently the ancestral copy was lost within the non-ciliated land plants (represented by *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa*). In contrast, the Bayesian phylogenies of the other CCP orthologues are consistent with the hypothesis of neofunctionalisation/co-option of an ancestral ciliary protein upon the evolutionary transition to loss of ability to produce cilia because the genes resolve into a single land plant clade (Figure 4-3A).

4.4.6 CCP proteins are highly expressed in male floral organs and mature pollen

While the CCP proteins have a ciliary role in ciliated species, the question of what their function is within the non-ciliated land plants remains. To this end, we analysed microarray data mined from a public database of the non-ciliated flowering plant *Arabidopsis thaliana* (Schmid et al., 2005). The expression pattern of each of the genes at different developmental stages is shown in Figure 4-3B, (expression data in Table 4-S5, see section 4.9). Clustering of the CCP expression data

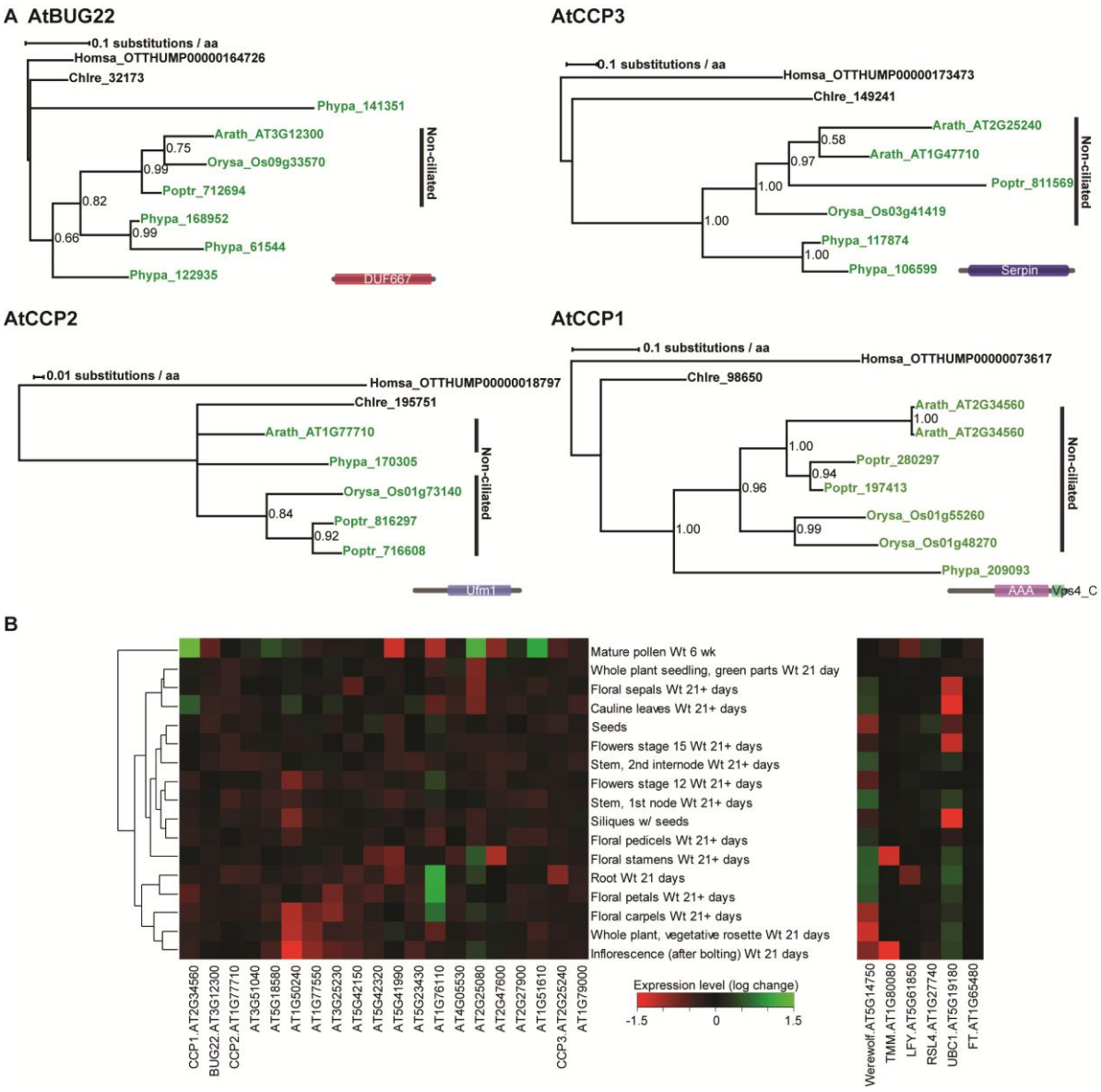


Figure 4-3: Evolution and function of CCP proteins

A) Phylogenetic Bayesian inferences for four of the CCP proteins, demonstrating different evolutionary scenarios. Land plants in green. For each protein, Pfam domain structure of human orthologue is shown. B) Heatmap illustration of microarray expression data for the CCP proteins (left) and controls (right) in different developmental stages scored by fold change from median.

by fold change shows that some of these genes are highly expressed in mature pollen. In contrast, the expression pattern of a number of control genes shows a more ubiquitous expression pattern.

4.5 Discussion

Over 300 proteins have previously been associated with ciliary function (for example see Arnaiz et al., 2009) with most being identified through proteomic analysis of isolated cilia (e.g. Inglis et al., 2006; Keller et al., 2005; Merchant et al., 2007b; Pazour et al., 2005). However, such studies fail to detect proteins which are not stably associated with the protein preparation used. Here, we have identified a novel set of 56 proteins possessing a ciliary profile, through phylogenetic distribution analysis and partial experimental validation in *T. brucei* (Figure 4-1). In addition we identified 21 proteins with a predicted ancestral ciliary role that have been maintained in non-ciliated land plants (Figures 4-2, 3).

4.5.1 Flowering plants have retained CCP proteins despite loss of ability to produce cilia

The phylogenetic distribution of cilia in land plants shows that cilia loss occurred at least twice: once in gymnosperms at the divergence of Ginkgo and the Cycads from the Conifers and Gnetals and once at the base of the flowering plants (Bremer et al., 1987; Finet et al., 2010). All other land plants produce cilia only within specialised sperm cells (Graham, 1978b, 1982, 1984; Graham and Mcbride, 1979; reviewed in Renzaglia and Garbary, 2001). During the evolutionary transition to loss of cilia, many ciliary function proteins were also lost (Table 4-S3, see section 4.9). However, we have identified 21 proteins with identifiable orthologues that are conserved in ciliated species and in the land plants (CCP proteins). At least a subset of these proteins possess a ciliary function in the flagellated *Trypanosoma brucei*, but the function of orthologues within the non-ciliated land plants has presumably diverged. Interestingly, a similar analysis for those proteins conserved during the evolutionary loss of cilia in fungal lineages identifies only two proteins, neither of which is present in the CCP set. Such low (21 and 2) numbers suggest that while loss of proteins is

widespread during evolutionary transitions, it is relatively rare for a protein to be co-opted to a different function without undergoing significant sequence divergence.

The flagella localisation of all three tested CCP proteins in trypanosomes (Figure 4-2) supports their hypothesised ancestral ciliary role and validates the scoring system that we used for CCP identification. While there is limited data on the putative roles of the 21 CCP proteins in non-ciliated plants, it is of note that where information is available they are generally associated with cytoskeletal/microtubule roles. For example, CCP1 is a putative katanin, a class of known AAA+ microtubule-severing proteins (Mcnally and Vale, 1993) and when mutated the CCP gene *tonneau2/fass1* markedly affects cell shape and arrangement in *Arabidopsis thaliana,* due to cytoskeletal abnormalities (Camilleri et al., 2002). Another CCP protein, FUSED, has been shown to be involved in the hedgehog signalling pathway in *Drosophila*, and hedgehog receptors localise to the cilium in many species (Oh et al., 2005). Together, these observations suggest that CCP proteins play regulatory roles that were ancestrally associated with cilia, but are now maintained for cytoskeletal functions in non-ciliated plants. Examination of the CCP proteins for Pfam domains (Finn et al., 2010) which include ubiquitanases, AAA+ domain containing proteins, protein kinases, tubulin-tryosine ligases and phosphatidic acid phosphatases, supports the suggestion that ancestral ciliary function proteins have been co-opted in land plants to perform cytoskeletal roles. Picket-Heaps (1971) proposed an autogenous origin of cilia from a microtubule-organising centre ancestrally involved in the regulation of cytoplasmic and mitotic microtubules (Cavalier-Smith, 1974a). If this was the case, it is equally possible that upon loss of cilia evolutionary pressures drove the co-option of cytoskeletal functions in the opposite direction.

CCP protein function can also be inferred from the observation that some CCP genes are expressed within mature pollen (Schmid et al., 2005; Figure 4-3; Table 4-S4, see section 4.9). This expression pattern is curious given the ancestral role of these proteins in ciliary formation in the

sperm cells of basal land plants and the fact that sperm cells in pollen are not ciliated. It is conceivable that genes involved in the cell-specific formation of cilia in basal land plants may have been co-opted on loss of cilia to perform similar cell-specific (possibly cytoskeletal) roles in a non-ciliary context. Such specificity would explain why only a small number of ciliary function genes were maintained or co-opted on the evolutionary transition to loss of cilia in seed plants, whereas the majority were lost or diverged.

4.6 Conclusions

The work presented here demonstrates the existence of a set of ancestral ciliary proteins within non-ciliated land plants, and suggests that the process of cilia loss can be accompanied by the co-option of ciliary function proteins into novel contexts.

4.7 Material and Methods

4.7.1 Bioinformatic definition of protein sets

Every predicted protein encoded in the *Chlamydomonas* genome was used as a query in a reciprocal best BLAST (RBB - see below) search with an e-value threshold of $<10^{-5}$ against each of the predicted proteomes of each of 45 eukaryotic organisms. During RBB, a BLAST search is performed using a query sequence against a target genome. If the top hit from the target genome identifies the query sequence in the reciprocal BLAST, the sequence identified is a RBB hit. The species used in the analysis have a broad phylogenetic distribution and include 29 organisms which produce centriole/cilia/flagella at some point in their life-cycle (termed “ciliated”) and 16 organisms which do not (termed “non-ciliated”). Importantly, each of these species possess a complete or near-complete genome sequence which is publically available. If the top hit from the RBB was the initial query sequence from *Chlamydomonas*, the identified target sequence was inferred to be an orthologue and was thus included in the dataset. Non-reciprocating sequences were discarded. The ciliary profile proteins were determined as those proteins whose evolutionary distribution of orthologues within the 45 species scored above 10 for a metric of *β* -

10α, (where *α* was the number of non-ciliated species possessing an orthologue and *β* was the number of ciliated species possessing an orthologue). The conserved in ciliated species and plants (CCP) protein set was defined as proteins with plant orthologues where *β>18* and *μ<2*, (where *μ* was the number of non-ciliated species, excluding plant species, possessing an orthologue).

4.7.2 Phylogenetic inference

For phylogenetic analysis, the sequences of each protein set were aligned using MAFFT6.24 (Katoh and Toh, 2008) adopting the L-INS-I strategy, and trimmed by eye to well aligned blocks. A Bayesian phylogeny was inferred from the protein alignment using the metropolis-coupled Markov chain Monte Carlo (MCMCMC) method as implemented in the program MrBayes3.1.2, with 2 independent runs from random start trees using the WAG substitution matrix with a gamma-distributed variation in substitution rate approximated to 4 discrete categories (Ronquist and Huelsenbeck, 2003) (random start tree, 4 chains of temperature 0.2 (default setting), Whelan and Goldman substitution matrices, 4 discrete categories of gamma distribution substitution rate, 500,000 generations). Tree likelihoods appeared to reach stationary phase at around 100,000 generations, and the last 350,000 generations were used to construct the consensus tree.

4.7.3 Endogenous-locus tagging

Endogenous-locus tagging was performed using procyclic-form *Trypanosoma brucei* Lister 427 cells (Cross, 1975; see Figure 4-S6, section 4.9). Cells were maintained as axenic cultures in SDM79 supplemented with 10% v/v foetal calf serum at 28°C, as described by Brun and Schönenberger (1979). Sequences were integrated at endogenous gene loci following the strategy outlined in Kelly et al., (2007). Briefly, genomic DNA for PCR was isolated from strain Lister 427 *T. brucei* procyclic cells as in (Chong, 2001); for primers see Table S6, section 4.9). PCR amplicons containing ~200 bp from the 5'end of the target cDNA sequence (CDS) with a linearisation site added to the 5' end of the 3' primer, and ~200 from the 3'end of the 5' intergenic region for the target CDS with the same linearization site incorporated into the 5' primer for this fragment, were

cloned together into the XbaI-BamHI sites downstream of *TY-YFP* in the pEnT6P vector, such that the N-terminal end of the CDS was in frame with *YFP*. The vector was then linearised at the site between the 5'end of the CDS and intergenic region fragments prior to transfection into procyclic-form Lister 427 *T. brucei* cells as described previously (Wirtz et al., 1999). Post-electroporation, cells were allowed to recover in normal growth media for 14 hr, after which time a stable transformant population was selected by the addition of 1 μg ml^{-1} puromycin.

4.7.4 Fluorescence microscopy

For analysis of YFP-tagged proteins in fixed whole cells, procyclic trypanosomes were harvested from mid-log phase cultures by centrifugation, washed once in phosphate buffered saline (PBS) and allowed to adhere to ethanol-washed plain glass slides for 5 min in PBS at ~$2x10^7$ cells ml^{-1}. Cells were then fixed for 5 min in 2% w/v formaldehyde in PBS, followed by permeabilisation in -20°C methanol for 10 min and rehydrated in PBS. Samples were mounted in 1% w/v 1,4-Diazabicyclo[2.2.2]octane, 90% glycerol, 50 mM sodium phosphate pH 8.0 containing 0.25 μg ml^{-1} 4',6-diamidino-2-phenylindole.

4.7.5 Western blot analysis

For whole-cell protein samples, cells were harvested, washed in PBS buffer, pelleted and immediately resuspended in boiling Laemmli buffer (2% sodium dodecyl sulphate (SDS), 10% glycerol, 400 mM β-mercaptoethanol, 50 mM Tris-HCl pH 7.2). SDS-polyacrylamide gel electrophoresis and immersion transfer to nitrocellulose membrane were performed using standard techniques described elsewhere (Chong, 2001). For immunoblotting, the membranes were blocked with 5% (w/v) skimmed milk powder in TBS-T (20 mM Tris-HCl, 500 mM NaCl, 0.05% Tween 20), labelled with anti-TY BB2 monoclonal antibody (Bastin et al., 1996) in TBS-T/1% skimmed milk powder, and developed with horseradish peroxidase-conjugated anti-mouse immunoglobulins (Sigma). As a loading control, membranes were re-probed with anti-PFR2 (L8C4) monoclonal antibody and developed as above.

4.8 Acknowledgements

We would like to thank Nicole Scheumman (Dunn School of Pathology, University of Oxford) for expert guidance on the transformation of *T. brucei*. We are grateful to Robert Scotland, Steven Kelly and Jim Fouracre (Dept. Plant Sciences, University of Oxford) for useful discussion. This work was funded by grants from the Wellcome Trust to KG and the Gatsby Charitable foundation to JAL. MEH is supported by a graduate studentship from BBSRC.

4.9 Additional information

The supplemental data in Table 4-S2-5 are such that the tables are too large for printed form. The legends for each Table have been included here, but please refer to the electronic submission for a complete dataset.

4.9.1 Table 4-S1

See section 2.9.

4.9.2 Table 4-S2

Protein information and sequence identification for orthologous sets of each protein identified in the bioinformatic screen. Database is too large for print purposes – for data please contact corresponding author or alternatively electronic journal. Protein information shown is the annotation from the JGI database for *Chlamydomonas reinhardtii*. Apime = *Apis mellifera*, Cioin = *Ciona intestinalis*, Caeel = *Caenorhabditis elegans*, Cyame = *Cyanidioschyzon merolae*, Capsp = *Capitella sp.*, Dicdi = *Dictyostelium discoideum*, Drome = *Drosophila melanogaster*, Danre = *Danio rerio*, Enccu = *Encephalitozoon cunculi*, Galga = *Gallus gallus*, Homsa = *Homo sapiens*, Lotgi = *Lottia gigantea*, Leima = *Leishmania major*, Monbr = *Monosiga brevicollis*, Neucr = *Neurospora crassa*, Naegr = *Naegleria gruberi*, Nemve = *Nematostella vectensis*, Ostta = *Ostreococcus tauri*, Parte = *Paramecium tetraurelia*, Sacce = *Saccharomyces cerevisiae*, Schpo = *Schizosaccharomyces pombe*, Strpu = *Strongylocentrotus purparatus*, Triad = *Trichoplax adhaerens*, Trybr = *Trypanosoma brucei*, Tetth = *Tetrahymena thermophila*, Takru = *Takifugu rubripes*, Ustma = *Ustilago maydis*.

4.9.3 Table 4-S3

Protein information and sequence identification for ciliary profile proteins.

4.9.4 Table 4-S4

Protein information and sequence identification for CCP proteins.

4.9.5 Table 4-S5

mRNA Expression data for *Arabidopsis thaliana* mined from and used in analysis of CCP protein functions.

4.9.6 Figure 4-S6

N-terminal endgenous locus gene tagging strategy

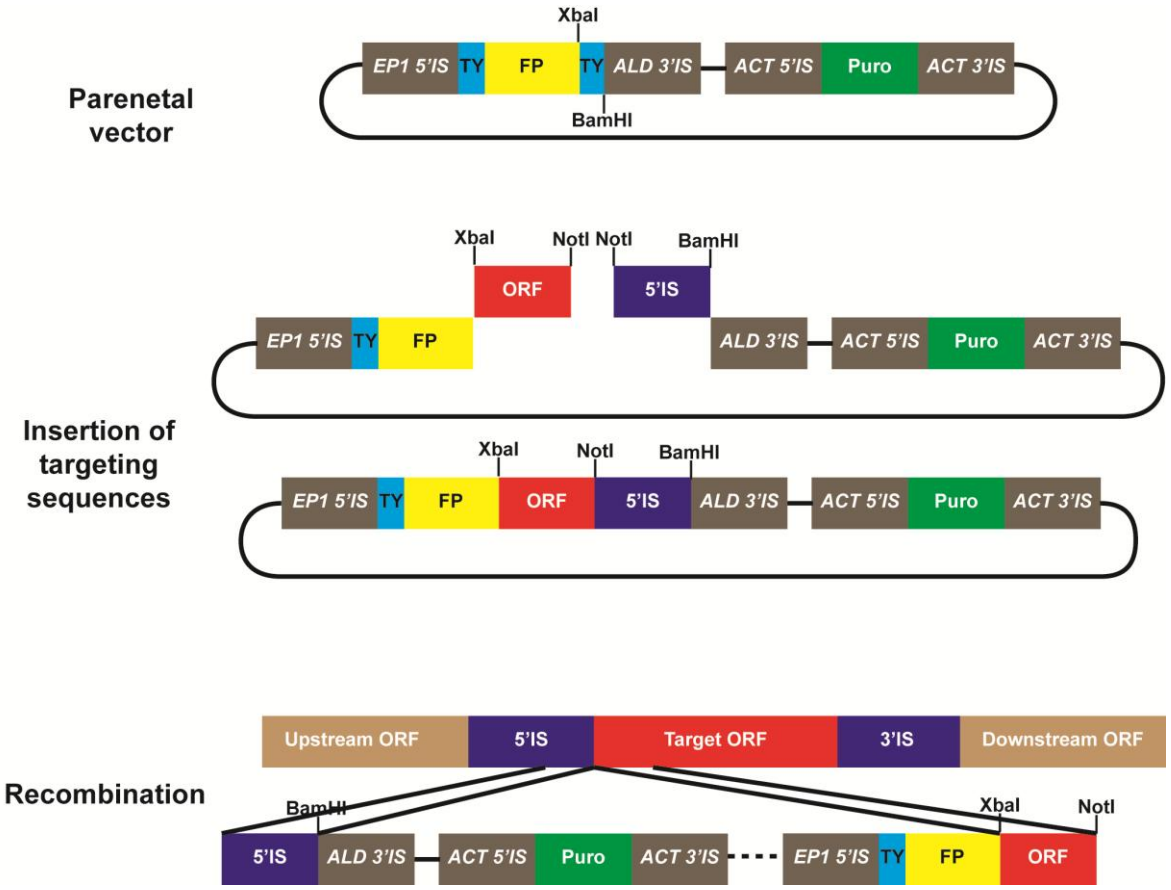


Figure 4-S6: Endogenous YFP Tagging strategy

4.9.7 Primer list

Table S6: Primer list.

Primer	Sequence (5'-3')	Comments
CCP1_1	gctctaga ATGAGTAAAAAGGGTAACATTCGC	Forward gene
CCP1_2	attgcggccgc TTATCGGCTGGTGAGAACTTTT	Reverse gene
CCP1_3	attgcggccgc GCATAGAACTGCCCTTGTATTGT	Forward intergenic
CCP1_3	attggatcc ATCCCAAAAAAAAAAACCGTCT	Reverse intergenic
CCP2_1	gctctaga ATGTCTCAAAATGAAGAAGC	Forward gene
CCP2_2	attgcggccgc GCGTGGAATTAGGCGTATCT	Reverse gene
CCP2_3	attgcggccgc ACCCCAACAGTTTGGACAAT	Forward intergenic
CCP2_4	attggatcc TATTAGCTAACAACCTCCCCTTC	Reverse intergenic
BUG22_1	gctctaga ATGTTCCACAACAGTTACCAGTCG	Forward gene
BUG22_2	attgcggccgc CCCTTGTCATCGAGAATCGT	Reverse gene
BUG22_3	attgcggccgc GCTCATTCTATTTTCCCCCTTT	Forward intergenic
BUG22_4	attggatcc ATGTTGGACCTTTTCGAGTG	Reverse intergenic
FBB17_1	gctctaga ATGGCAAGGGTTTGCAAG	Forward gene
FBB17_2	attgcggccgc GAAAAGGGTCCCGAAGAGTT	Reverse gene
FBB17_3	attgcggccgc GCACTCCTCTCCCAACTTCC	Forward intergenic
FBB17_4	attggatcc TACGAGAGCAATTTTTGCTAACAC	Reverse intergenic

5 General discussion

5.1 Summary

The cenancestor of the eukaryotes possessed a motile cilium with sensory roles. Where present, the ultrastructural morphology of extant eukaryotic cilia is remarkably conserved. However, lineage specific variation does occur, and cilia have been independently lost in multiple eukaryotic lineages. While a vast array of data exists for ciliary components in individual species; little has been done to identify core conserved proteins that play important roles in ciliary function across the broad range of the eukaryotic phylogeny. The work in this thesis has aimed to redress this imbalance. Taking advantage of the substantial amount of genomic data now publically available, bioinformatic analyses have identified key ciliary profile proteins with both known and unknown functions. In addition, morphological changes within the land plants have been associated with changes in ciliary protein profiles. Finally, studying the process of ciliary loss in land plants has allowed the identification of a group of proteins with inferred ancestral ciliary functions that have been maintained in non-ciliated plants.

Abbreviations: MTOC – Microtubule Organising Centre; PCM – Pericentriolar Material; LECA – Last Eukaryotic Common Ancestor; BBS – Bardet-Biedl Syndrome; CCP – Conserved in Ciliated and Plants.

5.2 LECA ciliary components

5.2.1 A base to build from: the role of centrioles in the eukaryotic cenancestor

Extant species utilise centrioles in a variety of functions. Acting as a basal body, this conserved microtubule structure plays a critical role as the ciliary foundation that extends the axoneme within the cell body (Satir and Christensen, 2007). Within cell cycle regulation, the specific microtubule organising centre (MTOC) properties of centrioles are utilised during mitotic division at the core of the centrosome, coordinating the protein components of the electron-dense pericentriolar material (PCM; Duncan and Wakefield, 2011). Within these two guises centrioles are associated with motility, sensory and regulatory roles; to the extent that even minor perturbations in centriole function often present as serious or lethal ciliopathies (Badano et al., 2006; Bettencourt-Dias et al., 2011).

Despite the key roles that centrioles paly, little is understood about the ancestral state of these complex organelles. Insight into the cenancestral centriole structure and composition would increase knowledge of how evolutionary pressures shape organelles by allowing comparisons to extant centrioles. Further, understanding centriole evolution would provide insight into the evolutionary importance of loss within different eukaryotic lineages. Analysing the phylogenetic profile of known centriolar proteins allowed the identification of those which were likely to be present in the last eukaryotic common ancestor (LECA). Included in this set are proteins such as the SAS-4/6 and the BBSome which have been shown in different model species to be involved in basal body formation and signalling respectively (Blacque and Leroux, 2006; Leidel et al., 2005; Leidel and Gonczy, 2003a). Provided that these conserved proteins have maintained a similar function, it can be inferred that the cenancestor of the eukaryotes possessed a cilium for both motility and sensory function. Furthermore, because protein components of the centrosome, for example PCM1 and Ninein, are restricted to the Holozoa (Hodges et al., 2010), it can be inferred that the centrosome is an evolutionary innovation specific to the Holozoa.

5.2.2 An ancestral centriolar inventory

In recent years a number of important centriolar proteins have been discovered (for database see Arnaiz et al., 2009). In many cases, however, evidence is species-specific and thus it is unclear how they operate in a phylogenetic context. By investigating the presence and absence of known centriolar proteins in diverse lineages, those which have been highly conserved through evolution have been identified. Such proteins can be inferred to play important roles in centriolar function. Linking the evolutionary distributions of centriole components to organismal biology potentially allows greater understanding of the mechanism behind centriole action.

The patterns of conservation and loss of 54 known centriolar proteins within the eukaryotes have allowed the definition of an ancestral centriolar inventory. This includes 14 core proteins, Polo-like-kinase, and proteins associated with Bardet-Biedl Syndrome (BBS) and Meckel-Gruber Syndrome. These proteins possess key centriolar functions which have been maintained in ciliated species since the divergence from the LECA. Interestingly, the morphology of the centriole is surprisingly insensitive to multiple losses from this cohort in that gross changes in the ultrastructure of the centriole occur only when the majority of the cohort is lost or highly divergent (see below).

5.2.3 Expansion to the axoneme: conserved ciliary proteins

While examining the phylogenetic distribution of proteins with known functions can reveal new insights into their potential roles and partners, such a technique relies on previous characterisation studies, and as such may miss key players and regulators. An unbiased whole genome approach circumvents such limitations, and provides the capacity for in-depth identification of novel ciliary factors.

To this end, the genomes of 45 eukaryotic species were analysed for orthologues which possess a high phylogenetic ciliary profile score (in that they are present in ciliated species, but largely absent in non-ciliated species). This analysis identified a core set of 213 proteins. The widespread

distribution of these proteins suggests that they were present in the LECA, and that they have subsequently been maintained in ciliated lineages during eukaryotic evolution (Figure 5-1). While this does not preclude the possibility of secondary functions outside of a ciliary role (co-option or divergence may have occurred in specific lineages), this conserved evolutionary profile infers an important ciliary role which is similar in all ciliated species. As such, investigation of the role of a protein in one species is likely to be applicable to others. The detection of the majority of the core proteins in previous ciliary proteomic studies (Inglis et al., 2006; Keller et al., 2005; Pazour et al., 2005) validates this ciliary profile approach. Intriguingly, however, the function of 56 of the proteins remains unknown, providing exciting avenues for future research.

5.3 Divergence of cilia

5.3.1 Recurrent themes of diversity in ciliary structure during evolution

Key ciliary structures are conserved across the eukaryotic phylogeny. Motifs such as the nine-fold radial symmetry basal body and axonemal microtubules, the change from triplet to doublet microtubules through the transition zone, and building mechanisms such as intraflagellar transport, are prevalent in ciliated species, although exceptions exist. Many key ciliary profile proteins are likewise conserved in this manner. Despite conservative maintenance of ciliary structures throughout evolution, however, multiple specific modifications have independently occurred since the divergence from the cenancestor of the eukaryotes.

The LECA is inferred to have possessed a motile cilium with a canonical nine triplet microtubule basal body and a nine doublet microtubule axoneme, a central pair of microtubules, inner and outer dynein arms and a sensory array (Carvalho-Santos et al., 2011a; Cavalier-Smith, 1974b). A distinctive modification of these features is found in the primary cilia of animals and in the independently evolved cilium of the amastigote form of *Leishmania major*. Both ciliary types are immotile (lacking dynein motors), short (1-2 μm), lack a central pair, and are hypothesised to act in a sensory rather than motility role. Independent loss of dynein arms provide another case of

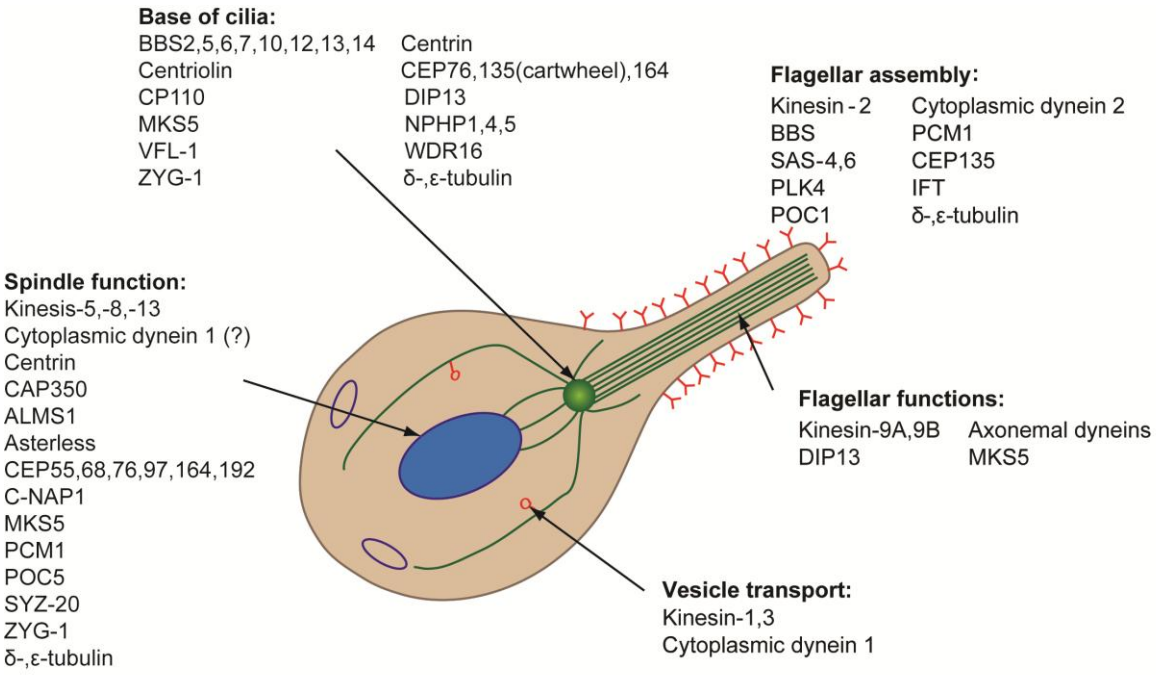


Figure 5-1: Possible organisation of cilia in LECA showing roles of key proteins.

The last common ancestor of the eukaryotes is hypothesised to have possessed a microtubule and actin based cytoskeleton (green), nuclear material surrounded by a double membrane, plastids, and a cilium utilised in motility and sensory function. The roles of the core ciliary proteins are shown within the context of this cenancestor.

modification: for example, outer arm dyneins have been lost in the land plants and teleost fish (Gibbons et al., 1985; Renzaglia and Garbary, 2001), inner arm dynein is lost in dipteran flies (Dallai, 1988), and insects have examples of loss of both (Afzelius et al., 1991). Even mechanisms for centriole biogenesis have evolved independently multiple times. The canonical template pathway, whereby the immature centriole is formed adjacent to the mature centriole, has been maintained in the majority of lineages. However, *de novo* formation of centrioles also occurs in highly divergent species, ranging from the land plants (Renzaglia and Garbary, 2001) to humans (Bettencourt-Dias and Glover, 2007).

The occurrence of multiple independent, yet structurally similar, modifications in diverse evolutionary lineages reinforces the modular nature of ciliogenesis and the inherent plasticity of this complex organelle. Alongside these frequent, in some case phylum-wide changes, species specific specialisation in ciliary structure has also occurred.

5.3.2 Species specific variation in ciliary components

In-depth analysis of eukaryotic genomes provides a number of insights into ciliary evolution. One clear message from the work presented here is that ciliary components have been modified in many lineages to play a specific role required by the particular group of organisms. For example, the unusual centriole of *Caenorhabditis elegans* is highly divergent in both protein composition and sequence, reflecting the unique role cilia play in neuronal perception in this species. The lack of key centriolar core proteins in *C. elegans,* such as δ- and ε-tubulin, shows how powerful evolutionary pressures can be in forcing high divergence and loss rates within specific lineages.

Modification in ciliary components does not only reflect divergence. In some cases whole protein complexes have been added to fulfil specific novel roles. This is the case with specific centrosomal proteins unique to the Holozoa, and in eye evolution where specific proteins are predicted to function in the development of ciliary, but not rhabdomeric, eyes. Likewise, the formation of the stellate array (Lang, 1963) in the basal bodies of plant cilia exemplifies macro modification in

ciliary structures. Intriguingly, the stellate array is subject to additional modification throughout the land plant lineage, to the extent that in ferns a contiguous scaffold-like array spans the entire triplet to doublet transition zone while the microtubules of the central pair nucleate at different positions along the proximo-distal axis. It can thus be seen that ciliary evolution is driven by divergence coupled with the addition of lineage specific components reflecting biological needs. Alongside these two processes, cilia loss also plays a vital role.

5.4 Loss of ciliary components

5.4.1 Cilia evolution driven by loss

Despite the high level of conservation of ciliary architecture across eukaryotes, the evolution of this complex organelle is largely driven by a process of loss (Figure 5-2). Interrogation of core centriolar and ciliary profile proteins (evolutionary conserved ciliary components derived from the LECA), shows that no one protein is conserved in all ciliated eukaryotic species. Even definitive core centriole proteins, such as SAS-4/6, and PLK4, are absent in species such as *Chlamydomonas reinhardtii*. The importance of loss is further highlighted by the specialised role of many ciliary proteins - if a general function is no longer required, whole complexes of proteins can be lost in a relatively short evolutionary timescale. This is the case with loss of the BBSome from organisms that produce cilia only for motility, predicting a dominant and ancient role for this complex in sensory function. Despite such phylogenetic distribution studies, however, little is known about this process of loss.

5.4.2 Land plant cilia: stepwise evolution to loss

The land plants provide a rich platform for ciliary evolution studies, in that they illustrate both diverse morphological change over a phylogenetically relevant timescale, and at least two occurrences of loss in ability to produce cilia (once at the base of the angiosperms, and once with the gymnosperms; Renzaglia and Garbary, 2001). Comparing the structure of cilia in the green algae *Chlamydomonas reinhardtii* (O'Toole et al., 2003) to that of the fern *Ceratopteris richardii*

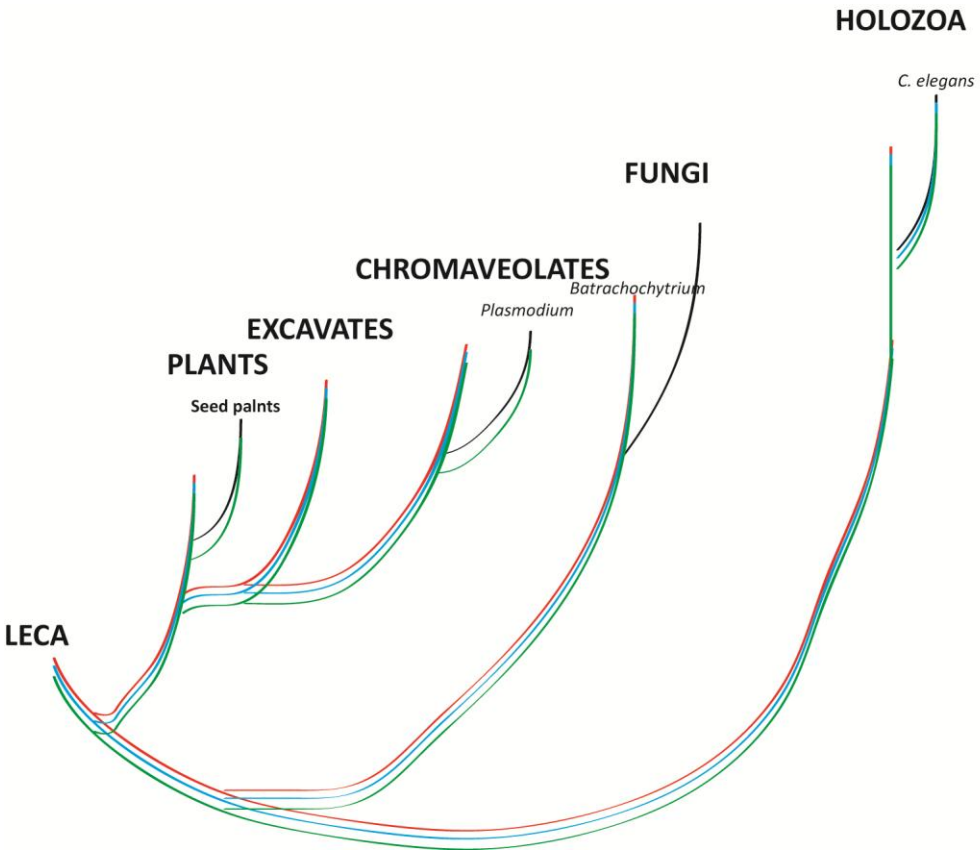


Figure 5-2: Evolutionary distribution of key proteins sets

Protein loss drives evolutionary features. This diagram shows that the core ciliary proteins (red), IFT proteins (blue) and CCP proteins (green) have each been lost respectively in different lineages. Even within these different sets of proteins, no one protein is ubiquitously conserved in all ciliated species.

(Duckett et al., 1979) shows the robust nature of ciliary structure in the plant lineage. Conserved features include basic microtubule organisation (triplet basal bodies, doublet axonemes, and central pair formation), a stellate array within the transition zone, and twisting of the central pair in relation to the outer doublets of the axoneme along the proximo-distal axis. However, multiple structural modifications have also occurred, including formation of novel fibrous connections (both anchoring the basal body to the underlying cytoskeleton and throughout the transition zone) and expansion in stellate array length to encompass the triplet microtubules of the basal body and the doublets of the transition zone. Significant structures have also been lost: the outer dynein arms within the axoneme are no longer present, while a breakdown of microtubule nine-fold symmetry occurs at both the base and tip of the fern cilium.

Changes in protein composition occur alongside the structural modifications described above. Many proteins with a conserved ciliary profile are lost from Plantae species upon the transition to land. These alterations mimic those found in other eukaryotic lineages, such as the dipteran flies, which may also eventually lose the ability to produce cilia. Parallels include the expansion in certain unique structural motifs (the stellate array in plants and the 9+9+0 microtubule arrangement in flies; Afzelius et al., 1991), coupled to a deterioration in ciliary function (the disorder of microtubule arrangement in plants and inability to swim in some dipteran sperm cells). Insight into the minimum ciliary genome may be gained through study of these transitional periods.

5.5 Cryptic secondary roles and co-option of plant ancestral ciliary proteins

5.5.1 CCP proteins

The ability to produce cilia has been lost at least twice within the land plants (Renzaglia and Garbary, 2001). During this transition, the majority of ciliary-associated proteins are lost. However, a number of proteins with a predicted ancestral ciliary role have been maintained

within non-ciliated land plants (CCP proteins). With one exception (BUG22), these proteins have not been detected in any ciliary proteomic study to date (Arnaiz et al., 2009; Laligne et al., 2010). Functional information about the proteins in either a ciliary or non-ciliary context is also lacking.

The inferred ancestral ciliary function of these CCP proteins has been confirmed through protein localisation studies in the flagellated single-celled excavate *Trypanosoma brucei*. All three tested proteins (CCP1, CCP2 and BUG22) localise to the flagella or basal body. Given that these CCP proteins are present in a non-ciliary context, the question arises as to what their derived role is. The phylogenetic profile of the CCP proteins indicates that they have either been subfunctionalised (in that the ancestor had two gene copies each with either a ciliary ora cryptic, secondary role) or co-opted into novel roles within the non-ciliated land plants. Bayesian inference of the CCP proteins shows that they fall into both categories. For example, BUG22 possesses the phylogenetic pattern suggestive of duplication within the ciliated land plants allowing the protein to potentially perform two functions. Subsequently, the ciliary function copy was lost, correlating with the loss in ability of plants to produce cilia. Other CCP proteins, such as CCP1, CCP2 and CCP3, display a different phylogenetic profile which is indicative of a co-option of an ancestral ciliary protein to a novel role. The study of these evolutionary scenarios is of great interest.

Analysis of microarray data for these CCP proteins shows that some are expressed in mature pollen of *Arabidopsis thaliana*. This expression profile is consistent with the inferred ancestral ciliary role for CCP proteins, in that land plant cilia are only produced in sperm cells by the gametophyte (angiosperm pollen is the gametophyte). It is conceivable that regulatory controls restrict the accumulation of ancestral ciliary protein sperm cells may have been co-opted in a non-ciliary context in the same cell type.

Where evidence is present for the role of CCP proteins, it shows a strong correlation with cytoskeletal functions. Picket-Heaps (1971) proposed a autogenous origin of cilia from an MTOC

that organised cytoplasmic and mitotic microtubules (Figure 5-3). The ciliary protrusion from the cell body aided motility and sensing, and evolutionary optimisation resulted in the conserved nine-fold symmetry of the organelle (Cavalier-Smith, 1974a). It is therefore possible that just as proteins with a cytoskeletal role were co-opted to a ciliary function; evolutionary pressures could have driven co-option of cytoskeletal functions in the opposite direction. Upon loss of cilia, these CCP proteins would have been maintained for their other roles. Of the CCP proteins, Tonneau 2 in *Arabidopsis thaliana* (Camilleri et al., 2002) plays an essential role in cortical cytoskeleton regulation, while CCP1 is a katanin, a class of proteins known to be involved in microtubule severing (Mcnally and Vale, 1993). It has further been shown that the ubiquitin-ligase activity of the tumour suppressor protein BRCA1 modifies γ-tubulin on residues Lys48 and Lys344, resulting in inhibition of microtubule-nucleating activity by the centrosome (Sankaran et al., 2006). As CCP2 possesses predicted ubiquitin-fold modification domains, it can be hypothesised that it may play a similar role.

In conclusion, the gene expression and protein function data are compatible with evolutionary scenarios of both sub- and neofunctionalisation. Further investigation of these proteins in both land plants and evolutionary distant species (such as *Trypanosoma brucei*) will shed light on this complex process of organelle loss coupled with protein maintenance.

5.6 Impact of current work

The pace of investigation into ciliary function is expanding at a tremendous rate. One output of this is an ever-increasing list of proteins that are in some form associated with cilia. Much of this information is obtained through sensitive mass spectrometry detection of protein components of cilia isolated from various model species, alongside the study of the causative mechanisms in mutagenic screens. However, a black box in this data set is an understanding of how ciliary components are shaped through evolution, and in particular of which proteins are maintained across the eukaryotic lineage and which are lost or significantly divergent.

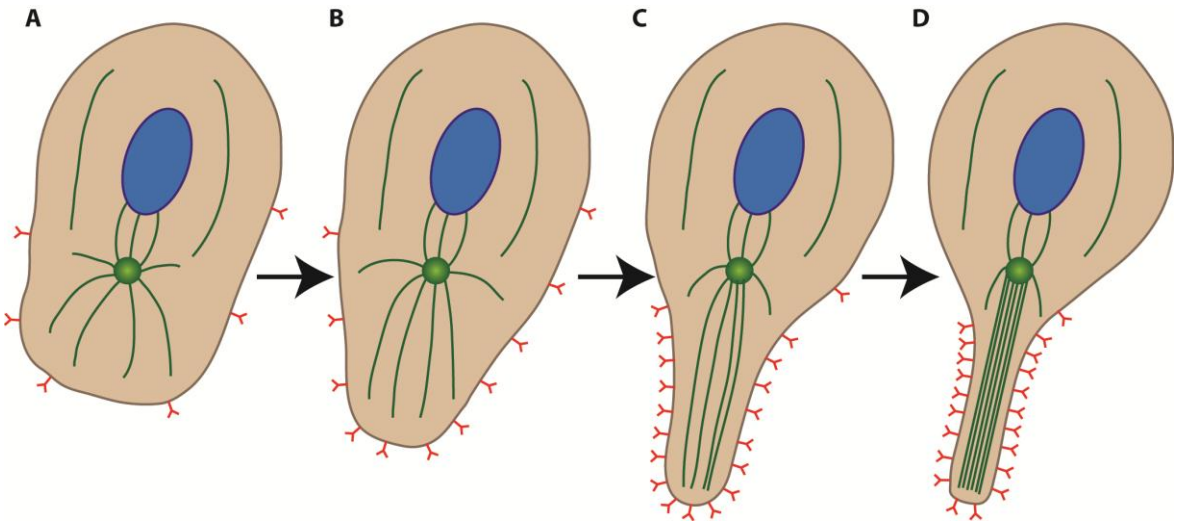


Figure 5-3: Autogenous theory of origin of cilium

A) LECA possessed a cytoskeletal composed of actin and microtubules, an MTOC, nucleus and endomembrane system. Sensory receptors were distributed throughout cell membrane (red) B) Vesicle traffic was targeted to a specific site on the cell membrane, which protruded from the cell due to forces acting from microtubules. Receptor proteins begin to be redistributed by this action. C) Specialist proto-cilium formed with specific membrane composition and maintained by diffusion barriers. This proto-cilium is capable of sensing and gliding, which further drives refinement of organelle. IFT system for formation and maintenance evolves. D) Specific arrangement of nine-fold microtubule arrangement evolves to optimise movement driven by motors. Basal body is formed to support additional stresses.

The work in this thesis identifies a core set of ciliary function proteins which are highly conserved throughout the eukaryotic lineage in ciliated species. 56 of these proteins have unknown functions and have not been previously linked to ciliary function, providing a novel set of candidates for future investigation. Through the investigation of ciliary morphology in the land plants *Ceratopteris* and *Marchantia*, this work has demonstrated ultrastructural changes across the phylogeny, and linked these changes to evolution of protein components. This has provided an understanding of protein compositional changes through the stepwise deterioration of cilia within land plants, and has shown that loss of ciliary proteins occurred on the transition to land. Finally, a cohort of ancestral ciliary proteins has been identified that is maintained in non-ciliated angiosperms, despite the fact that flowering plants have lost the ability to produce cilia. It is likely that these proteins were co-opted to perform novel functions within these species, and as such provide an attractive avenue for future research.

References

Afzelius, B.A., Bellon, P.L., Dallai, R., and Lanzavecchia, S. (1991). Diversity of microtubular doublets in insect sperm tails - a computer-aided image-analysis. Cell Motility and the Cytoskeleton *19*, 282-289.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Bioliogy *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.

Arendt, D. (2003). Evolution of eyes and photoreceptor cell types. International Journal of Developmental Biology *47*, 563-571.

Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A.W., and Wittbrodt, J. (2004). Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science *306*, 869-871.

Arendt, D., and Wittbrodt, J. (2001). Reconstructing the eyes of Urbilateria. Philosophical Transactions of Royal Society London B Biological Sciences *356*, 1545-1563.

Armbrust, E.V. (1990). The life cycle of the centric diatom *Thalassiosira weissflogii*: control of gametogenesis and cell size. Doctoral Dissertation at Woods Hole Oceanographic Institution Massachusetts Institute of Technology.

Arnaiz, O., Malinowska, A., Klotz, C., Sperling, L., Dadlez, M., Koll, F., and Cohen, J. (2009). Cildb: a knowledgebase for centrosomes and cilia. Database (Oxford) *2009*, bap022.

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.

Badano, J.L., Mitsuma, N., Beales, P.L., and Katsanis, N. (2006). The ciliopathies: an emerging class of human genetic disorders. Annual Reviews Genomics Human Genetics *7*, 125-148.

Bastin, P., Bagherzadeh, Z., Matthews, K.R., and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis *in Trypanosoma brucei*. Molecular Biochemistry of Parasitology *77*, 235-239.

Bastin, P., Pullen, T.J., Moreira-Leite, F.F., and Gull, K. (2000). Inside and outside of the trypanosome flagellum: a multifunctional organelle. Microbes and Infection *2*, 1865-1874.

Beisson, J., and Wright, M. (2003). Basal body/centriole assembly and continuity. Current Opinion in Cell Biology *15*, 96-104.

Berger, L., Hyatt, A.D., Speare, R., and Longcore, J.E. (2005). Life cycle stages of the amphibian chytrid *Batrachochytrium dendrobatidis*. Diseases of Aquatic Organisms *68*, 51-63.

Bernhard, D.L., and Renzaglia, K.S. (1995). Spermiogenesis in the moss *Aulacomnium palustre*. Bryologist *98*, 52-70.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B.*, et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei*. Science *309*, 416-422.

Bettencourt-Dias, M., and Glover, D.M. (2007). Centrosome biogenesis and function: centrosomics brings new understanding. Nature Reviews Molecular Cell Biology *8*, 451-463.

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G., and Godinho, S.A. (2011). Centrosomes and cilia in human disease. Trends in genetics : TIG *27*, 307-315.

Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M.K., Carmo, N., Balloux, F., Callaini, G., and Glover, D.M. (2005). SAK/PLK4 is required for centriole duplication and flagella development. Current Biology *15*, 2199-2207.

Bhattacharya, D., and Medlin, L. (1998). Algal phylogeny and the origin of land plants. Plant Physiology *116*, 9-15.

Bilderback, Jahn, T.L., and Fonesca, J.R.J. (1974). Locomotor behavior of *Lygodium* and *Marsilea* sperm. American Journal of Botany *61*, 34-34.

Bjork, A., and Pitnick, S.J. (2006). Intensity of sexual selection along the anisogamy-isogamy continuum. Nature *441*, 742-745.

Blacque, O.E., and Leroux, M.R. (2006). Bardet-Biedl syndrome: an emerging pathomechanism of intracellular transport. Cellular and Molecular Life Sciences *63*, 2145-2161.

Bornens, M., and Azimzadeh, J. (2007). Origin and evolution of the centrosome. In Eukaryotic membranes and cytoskeleton: Origins and evolution, G. Jèkely, ed. (Landes Bioscience), pp. 119-129.

Bremer, K., Humphries, C.J., Mishler, B.D., and Churchill, S.P. (1987). On cladistic relationships in green plants. Taxon *36*, 339-349.

Briggs, L.J., Davidge, J.A., Wickstead, B., Ginger, M.L., and Gull, K. (2004). More than one way to build a flagellum: comparative genomics of parasitic protozoa. Current Biology *14*, R611-R612.

Broadhead, R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw, M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G.*, et al.* (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. Nature *440*, 224-227.

Brun, R., and Schonenberger, M. (1979). Cultivation and Invitro Cloning of Procyclic Culture Forms of *Trypanosoma brucei* in a Semi-Defined Medium. Acta Tropica *36*, 289-292.

Callaini, G., Whitfield, W.G., and Riparbelli, M.G. (1997). Centriole and centrosome dynamics during the embryonic cell cycles that follow the formation of the cellular blastoderm in *Drosophila*. Experimental Cell Research *234*, 183-190.

Camilleri, C., Azimzadeh, J., Pastuglia, M., Bellini, C., Grandjean, O., and Bouchez, D. (2002). The *Arabidopsis* TONNEAU2 gene encodes a putative novel protein phosphatase 2A regulatory subunit essential for the control of the cortical cytoskeleton. Plant Cell *14*, 833-845.

Carothers, Z., and Kreitner, G.L. (1968). Studies of spermatogenesis in Hepaticae .2. Blepharoplast structure in spermatid of *Marchantia*. Journal of Cell Biology *36*, 603.

Carothers, Z.B., and Duckett, J.G. (1980). The bryophyte spermatozoid: a source of new phylogenetic information. Bulletin of the Torrey Botanical Club *107*, 281-297.

Carothers, Z.B., Robbins, R.R., and Haas, D.L.J. (1975a). Some ultrastructural aspects of spermatogenesis in *Lycopodium complanatum*. Protoplasma *86*, 339-350.

Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J.B., and Bettencourt-Dias, M. (2011a). Tracing the origins of centrioles, cilia, and flagella. The Journal of Cell Biology *194*, 165-175.

Cavalier-Smith, T. (1974a). Basal body and flagellar development during the vegetative cell cycle and the sexual cycle of *Chlamydomonas reinhardii*. Journal of Cell Science *16*, 529-556.

Cavalier-Smith, T. (1987). The origin of eukaryotic and archaebacterial cells. Annuals NY Academy of Science *503*, 17-54.

Chang, P., and Stearns, T. (2000). Delta-tubulin and epsilon-tubulin: two new human centrosomal tubulins reveal new aspects of centrosome structure and function. Nature Cell Biology *2*, 30-35.

Chasey, D. (1974). 3-Dimensional arrangement of radial spokes in flagella of *Chlamydomonas reinhardii*. Experimental Cell Research *84*, 374-380.

Chong, L. (2001). Molecular cloning - A laboratory manual, 3rd edition. Science *292*, 446-446.

Cole, D.G., Diener, D.R., Himelblau, A.L., Beech, P.L., Fuster, J.C., and Rosenbaum, J.L. (1998). *Chlamydomonas* kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in *Caenorhabditis elegans* sensory neurons. Journal of Cell Biology *141*, 993-1008.

Coss, R.A. (1974). Mitosis in *Chlamydomonas reinhardtii* basal bodies and mitotic apparatus. Journal of Cell Biology *63*, 325-329.

Cross, G.A.M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting surface coat of *Trypanosoma brucei.* Parasitology *71*, 393-&.

Dallai, R. (1988). The spermatozoon of *Asphondyliidi* (Diptera, Cecidomyiidae). Journal of Ultrastructure and Molecular Structure Research *101*, 98-107.

Dammermann, A., Maddox, P.S., Desai, A., and Oegema, K. (2008). SAS-4 is recruited to a dynamic structure in newly forming centrioles that is stabilized by the gamma-tubulin-mediated addition of centriolar microtubules. Journal of Cell Biology *180*, 771-785.

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

Dawe, H.R., Farr, H., and Gull, K. (2007a). Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells. Journal of Cell Science *120*, 7-15.

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.* (2007b). The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Human Molecular Genetics *16*, 173-186.

Delattre, M., Canard, C., and Gonczy, P. (2006). Sequential protein recruitment in *C. elegans* centriole formation. Current Biology *16*, 1844-1849.

Delattre, M., and Gonczy, P. (2004). The arithmetic of centrosome biogenesis. Journal of Cell Science *117*, 1619-1630.

Delattre, M., Leidel, S., Wani, K., Baumer, K., Bamat, J., Schnabel, H., Feichtinger, R., Schnabel, R., and Gonczy, P. (2004). Centriolar SAS-5 is required for centrosome duplication in *C. elegans*. Nature Cell Biology *6*, 656-664.

Dentler, W.L.J. (1980a). The central and outer doublet microtubules are attached to a cap in mammalian tracheal cilia. Journal of Cell Biology *87*, A36-A36.

Dentler, W.L.J. (1980b). Structures linking the tips of ciliary and flagellar microtubules to the membrane. Journal of Cell Science *42*, 207-220.

Devaux, S., Kelly, S., Lecordier, L., Wickstead, B., Perez-Morga, D., Pays, E., Vanhamme, L., and Gull, K. (2007). Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. Molecular Biology of the Cell *18*, 1293-1301.

Diener, D.R., Ang, L.H., and Rosenbaum, J.L. (1993). Assembly of flagellar radial spoke proteins in *Chlamydomonas* - identification of the axoneme binding domain of Radial Spoke Protein-3. Journal of Cell Biology *123*, 183-190.

Dirksen, E.R. (1971). Centriole morphogenesis in developing ciliated epithelium of mouse oviduct. Journal of Cell Biology *51*, 286.

Duckett, J.G., and Carothers, Z.B.S. (1982). The assembly of flagella during spermatogenesis in land plants. Amos, W B and J G Duckett (ed) Symposia of the society for experimental biology, Cambridge University Press: Cambridge, England; New York, NY, USA Illus, P533-562.

Duckett, J.G., Klekoski, E.J., and Hickok, L.G.J. (1979). Ultrastructural studies of mutant spermatozoids in ferns .1. mature non-motile spermatozoid of mutation 230x in *Ceratopteris thalictroides (l) brongn*. Gamete Research *2*, 317-343.

Duncan, T., and Wakefield, J.G. (2011). 50 ways to build a spindle: the complexity of microtubule generation during mitosis. Chromosome Research *19*, 321-333.

Dupuis-Williams, P., Fleury-Aubusson, A., de Loubresse, N.G., Geoffroy, H., Vayssie, L., Galvani, A., Espigat, A., and Rossier, J. (2002). Functional role of epsilon-tubulin in the assembly of the centriolar microtubule scaffold. Journal of Cell Biology *158*, 1183-1193.

Dute, R., and Kung, C.J. (1978). Ultrastructure of proximal region of somatic cilia in *Paramecium tetraurelia*. Journal of Cell Biology *78*, 451-464.

Eakin, R.M. (1982). Morphology of invertebrate photoreceptors. Methods of Enzymology *81*, 17-25.

Edgerton, R.F. (2005). The principles of electron microscopy: and introduction to TEM, SEM and AEM. Springer-Verlag gmbh & Co.

Elias, M., and Archibald, J.M. (2009). The RJL family of small GTPases is an ancient eukaryotic invention probably functionally associated with the flagellar apparatus. Gene *442*, 63-72.

Fan, Y.L., Esmail, M.A., Ansley, S.J., Blacque, O.E., Boroevich, K., Ross, A.J., Moore, S.J., Badano, J.L., May-Simera, H., Compton, D.S.*, et al.* (2004). Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. Nature Genetics *36*, 989-993.

Finet, C., Timme, R.E., Delwiche, C.F., and Marleta, F. (2010). Multigene phylogeny of the green lineage reveals the origin and diversification of Land Plants. Current Biology *20*, 2217-2222.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K.*, et al.* (2010). The Pfam protein families database. Nucleic Acids Research *38*, D211-D222.

Fliegauf, M., Benzing, T., and Omran, H. (2007). When cilia go bad: cilia defects and ciliopathies. Nature Reviews Molecular Cell Biology *8*, 880-893.

Gadelha, C., Wickstead, B., McKean, P.G., and Gull, K. (2006). Basal body and flagellum mutants reveal a rotational constraint of the central pair microtubules in the axonemes of trypanosomes. Journal of Cell Science *119*, 2405-2413.

Geimer, S., and Melkonian, M. (2004). The ultrastructure of the *Chlamydomonas reinhardtii* basal apparatus: identification of an early marker of radial asymmetry inherent in the basal body. Journal of Cell Science *117*, 2663-2674.

Geimer, S., and Melkonian, M. (2005). Centrin scaffold in *Chlamydomonas reinhardtii* revealed by immunoelectron microscopy. Eukaryotic Cell *4*, 1253-1263.

Gibbons, B.H., Baccetti, B., and Gibbons, I.R. (1985). Live and reactivated motility in the 9 + 0 flagellum of *Anguilla* sperm. Cell Motility and the Cytoskeleton *5*, 333-350.

Gibbons, B.H., and Gibbons, I.R. (1973). Effect of partial extraction of dynein arms on movement of reactivated sea-urchin sperm. Journal of Cell Science *13*, 337-357.

Graham, L.E. (1978a). Ultrastructure of spermatozoids of *Coleochaete scutata*. Journal of Phycology *14*, 30-30.

Graham, L.E. (1982). The occurrence, evolution, and phylogenetic significance of parenchyma in *Coleochaete breb* (Chlorophyta). American Journal of Botany *69*, 447-454.

Graham, L.E. (1984). *Coleochaete* and the origin of Land Plants. American Journal of Botany *71*, 603-608.

Graham, L.E., and Mcbride, G.E. (1975). Ultrastructure of multilayered structures associated with flagellar bases in motile cells of *Trentepohlia aurea*. Journal of Phycology *11*, 86-96.

Graham, L.E., and Mcbride, G.E. (1979). Occurrence and phylogenetic significance of a multilayered structure in *Coleochaete* spermatozoids. American Journal of Botany *66*, 887-894.

Graham, L.E., and Wedemayer, G.J. (1984). Spermatogenesis in *Coleochaete pulvinata* (Charophyceae) - sperm maturation. Journal of Phycology *20*, 302-309.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systems Biology *59*(3), 307-321

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. System Biology *52*, 696-704.

Gull, K., Lacomble, S., Vaughan, S., Gadelha, C., Morphew, M.K., Shaw, M.K., and McIntosh, J.R. (2010). Basal body movements orchestrate membrane organelle division and cell morphogenesis in *Trypanosoma brucei*. Journal of Cell Science *123*, 2884-2891.

Habedanck, R., Stierhof, Y.D., Wilkinson, C.J., and Nigg, E.A. (2005). The Polo kinase Plk4 functions in centriole duplication. Nature Cell Biology *7*, 1140-1146.

Henderson, G.P., Gan, L., and Jensen, G.J. (2007). 3-D ultrastructure of *O. tauri*: electron cryotomography of an entire eukaryotic cell. PLoS ONE *2*, e749.

Hickok, L.G., Warne, T.R., and Fribourg, R.S. (1995). The biology of the fern *Ceratopteris* and its use as a model system. International Journal of Plant Sciences *156*, 332-345.

Hiraki, M., Matsuura, K., Kamiya, R., and Hirono, M. (2005). Filament-forming properties of Bld10p, a coiled-coil protein that constitutes the *Chlamydomonas* basal-body cartwheels. Cell Structure and Function *30*, 57-57.

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. Current Biology *17*, 1778-1783.

Hodges, M.E., Scheumann, N., Wickstead, B., Langdale, J.A., and Gull, K. (2010). Reconstructing the evolutionary history of the centriole from protein components. Journal of Cell Science *123*, 1407-1413.

Hoffman, J.C., Vaughn, K.C., and Joshi, H.C.J. (1994). Structural and immunocytochemical characterization of microtubule-organizing centers in pteridophyte spermatogenous cells. Protoplasma *179*, 46-60.

Hoops, H.J., and Witman, G.B. (1985). Basal bodies and associated structures are not required for normal flagellar motion or phototaxis in the green-alga *Chlorogonium elongatum*. Journal of Cell Biology *100*, 297-309.

Huelsenbeck, J.P., Crandall, K.A. (1997). Phylgeny estimation and hypothesis testing using maximum likelihood. Annual Reviews of Ecology, Evolution and Systematics *28*, 437-66.

Hyams, J.S., and Campbell, C.J.J. (1985). Widespread absence of outer dynein arms in the spermatozoids of lower plants. Cell Biology Int Rep *9*, 841-848.

Inglis, P.N., Boroevich, K.A., and Leroux, M.R. (2006). Piecing together a ciliome. Trends in Genetics *22*, 491-500.

Inglis, P.N., Ou, G., Leroux, M.R., and Scholey, J.M. (2007). The sensory cilia of *Caenorhabditis elegans*. WormBook, 1-22.

Jensen, K.G., Moestrup, O., and Schmid, A.M.M. (2003). Ultrastructure of the male gametes from two centric diatoms, *Chaetoceros laciniosus* and *Coscinodiscus wailesii* (Bacillariophyceae). Phycologia *42*, 98-105.

Jin, H., and Nachury, M.V. (2009). The BBSome. Current Biology *19*, R472-473.

Jin, Q.J., and Hasenstein, K.H. (2009). Cytoskeletal control of sperm release in *Chara contraria*. Botanica Marina *52*, 162-169.

Kamiya, R., and Okamoto, M. (1985). A mutant of *Chlamydomonas* reinhardtii that lacks the flagellar outer dynein arm but can swim. Journal of Cell Science *74*, 181-191.

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics *9*, 286-298.

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

Keller, L.C., Romijn, E.P., Zamora, I.P., Yates, J.R., and Marshall, W.F. (2006). Dissecting the centriole proteome. Faseb Journal *20*, A493-A493.

Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B.*, et al.* (2007). Functional genomics in *Trypanosoma brucei*: A collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. Molecular and Biochemical Parasitology *154*, 103-109.

Kemp, C.A., Kopish, K.R., Zipperlen, P., Ahringer, J., and O'Connell, K.F. (2004). Centrosome maturation and duplication in *C. elegans* require the coiled-coil protein SPD-2. Developmental Cell *6*, 511-523.

Kilburn, C.L., Pearson, C.G., Romijn, E.P., Meehl, J.B., Giddings, T.H., Culver, B.P., Yates, J.R., and Winey, M. (2007). New *Tetrahymena* basal body protein components identify basal body domain structure (vol 178, pg 905, 2007). Journal of Cell Biology *179*, 167-167.

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.

Klekowski, J.J.R.J. (1969). Reproductive biology and evolution in the Pteridophyta. Dissertation Abstracts B Sciences and Engineering *29*, 3221-B.

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

Kohlmaier, G., Loncarek, J., Meng, X., McEwen, B.F., Mogensen, M.M., Spektor, A., Dynlacht, B.D., Khodjakov, A., and Gonczy, P. (2009). Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. Current Biology *19*, 1012-1018.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. Journal of Structural Biology *116*, 71-76.

Krock, B.L., Mills-Henry, I., and Perkins, B.D. (2009). Retrograde intraflagellar transport by cytoplasmic Dynein-2 is required for outer segment extension in vertebrate photoreceptors but not Arrestin translocation. Investigative Ophthalmology & Visual Science *50*, 5463-5471.

Laligne, C., Klotz, C., de Loubresse, N.G., Lemullois, M., Hori, M., Laurent, F.X., Papon, J.F., Louis, B., Cohen, J., and Koll, F. (2010). Bug22p, a conserved centrosomal/ciliary protein also present in higher plants, is required for an effective ciliary stroke in *Paramecium*. Eukaryotic Cell *9*, 645-655.

Lang, N.J. (1963). An additional ultrastructural component of flagella. The Journal of Cell Biology *19*, 631-634.

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. Nature Cell Biology *7*, 115-125.

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

Li, Y., and Hu, J. (2011). Small GTPases and cilia. Protein Cell *2*, 13-25.

Liu, Y.J.J., Hodson, M.C., and Hall, B.D. (2006). Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of Kingdom Fungi inferred from RNA polymerase II subunit genes. BMC Evolutionary Biology *6*, -.

Lopez-Fernandez, C., Baltanas, A., De la Torre, J., and Gosalvez, J.J. (2007). *Drosophila melanogaster* and *Eucypris virens* giant spermatozoa as visualized by cell inclusion in microgels. Journal of Experimental Zoology Part A *307A*, 140-144.

Lopez-Smith, R., and Renzaglia, K. (2008). Sperm cell architecture, insemination, and fertilization in the model fern, *Ceratopteris richardii*. Sexual Plant Reproduction *21*, 153-167.

Maden, A.R., Whittier, D.P., Garbary, D.J., and Renzaglia, K.S. (1997). Ultrastructure of the spermatozoid of *Lycopodiella lateralis* (Lycopodiaceae). Canadian Journal of Botany-Revue Canadienne De Botanique *75*, 1728-1738.

Manton, I. (1965). Observations on fine structure of 2 species of *Platymonas* with special reference to flagellar scales and mode of rigin of theca. Journal of the Marine Biological Association of the United Kingdom *45*, 743.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology *152*, 36-51.

McLaren, D.J. (1976). Nematode sense organs. Adv Parasitol *14*, 195-265.

Mcnally, F.J., and Vale, R.D. (1993). Identification of Katanin, an ATPase that severs and disassembles stable microtubules. Cell *75*, 419-429.

Melkonian, M. (1989). Flagellar apparatus ultrastructure in *Mesostigma viride* (Prasinophyceae). Plant Systematics and Evolution *164*, 93-122.

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.* (2007a). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science *318*, 245-251.

Messaoudi, C., de Loubresse, N.G., Boudier, T., Dupuis-Williams, P., and Marco, S.J. (2006). Multiple-axis tomography: applications to basal bodies from *Paramecium tetraurelia*. Biological Cell *98*, 415-425.

Miller, G.T., and Pitnick, S.J. (2002). Sperm-female coevolution in *Drosophila*. Science *298*, 1230-1233.

Mitchell, D.R. (2003). Orientation of the central pair complex during flagellar bend formation in *Chlamydomonas*. Cell Motility and the Cytoskeleton *56*, 120-129.

Mitchell, D.R. (2007). The evolution of eukaryotic cilia and flagella as motile and sensory organelles. Eukaryotic Membranes and Cytoskeleton: Origins and Evolution *607*, 130-140.

Mitchell, D.R., and Nakatsugawa, M.J. (2004). Bend propagation drives central pair rotation in *Chlamydomonas reinhardtii* flagella. Journal of Cell Biology *166*, 709-715.

Mitchell, D.R.S. (2005). Regulation of eukaryotic flagellar motility. AIP Conference Procedings *755*, 130-136.

Muccifora, S., and Gori, P.J. (2005). Ultrastructure of mature spermatozoids in the fern *Asplenium onopteris L*. Micron *36*, 539-544.

Mykytyn, K., Mullins, R.F., Andrews, M., Chiang, A.P., Swiderski, R.E., Yang, B.L., Braun, T., Casavant, T., Stone, E.M., and Sheffield, V.C. (2004). Bardet-Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly. Proceedings of the National Academy of Sciences of the United States of America *101*, 8664-8669.

Myles, D.G., and Bell, P.R.J. (1975). Ultrastructural study of spermatozoid of fern, *Marsilea vestita*. Journal of Cell Science *17*, 633-645.

Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peranen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C.*, et al.* (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell *129*, 1201-1213.

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. Current Biology *17*, 2169-2174.

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

Norstog, K. (1974). Fine-structure of spermatozoid of *Zamia vierergruppe*. American Journal of Botany *61*, 449-456.

O'Connell, K.F., Caron, C., Kopish, K.R., Hurd, D.D., Kemphues, K.J., Li, Y., and White, J.G. (2001). The *C. elegans* zyg-1 gene encodes a regulator of centrosome duplication with distinct maternal and paternal roles in the embryo. Cell *105*, 547-558.

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*. Molecular Biology of the Cell *14*, 2999-3012.

Oakley, C.E., and Oakley, B.R. (1989). Identification of gamma-tubulin, a new member of the tubulin superfamily encoded by mipA gene of *Aspergillus nidulans*. Nature *338*, 662-664.

Oh, S.A., Johnson, A., Smertenko, A., Rahman, D., Park, S.K., Hussey, P.J., and Twell, D. (2005). A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. Current Biology *15*, 2107-2111.

Omoto, C.K., Gibbons, I.R., Kamiya, R., Shingyoji, C., Takahashi, K., and Witman, G.B. (1999). Rotation of the central pair microtubules in eukaryotic flagella. Molecular Biology of the Cell *10*, 1-4.

Omoto, C.K., and Kung, C. (1980). Rotation and twist of the central-pair microtubules in the cilia of *Paramecium*. Journal of Cell Biology *87*, 33-46.

Omoto, C.K., and Witman, G.B. (1980). Rotation of the central pair of microtubules in an algal flagellum. Journal of Cell Biology *87*, A35-A35.

Omoto, C.K., and Witman, G.B. (1981). Functionally significant central-pair rotation in a primitive eukaryotic flagellum. Nature *290*, 708-710.

Pan, J.M., and Snell, W.J. (2002). Kinesin-II is required for flagellar sensory transduction during fertilization in *Chlamydomonas*. Molecular Biology of the Cell *13*, 1417-1426.

Paolillo, D.J., Kreitner, G.L., and Reighard, J.A. (1968). Spermatogenesis in *Polytrichum juniperinum* .2. Mature sperm. Planta *78*, 248-&.

Paolillo, D.J.J. (1981). The swimming sperms of land plants. Bioscience *31*, 367-372.

Parker, D.S., and Katsanis, N. (2011). Understanding cargo specificity in intraflagellar transport. EMBO Journal *30*, 2518-2519.

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

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

Pelletier, L., Ozlu, N., Hannak, E., Cowan, C., Habermann, B., Ruer, M., Muller-Reichert, T., and Hyman, A.A. (2004). The *Caenorhabditis elegans* centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. Current Biology *14*, 863-873.

Pereira-Leal, J.B., Carvalho-Santos, Z., Machado, P., Branco, P., Tavares-Cadete, F., Rodrigues-Martins, A., and Bettencourt-Dias, M. (2010). Stepwise evolution of the centriole-assembly pathway. Journal of Cell Science *123*, 1414-1426.

Perkins, L.A., Hedgecock, E.M., Thomson, J.N., and Culotti, J.G. (1986a). Mutant sensory cilia in the nematode *Caenorhabditis elegans*. Developmental Biology *117*, 456-487.

Phillips, D.M.J. (1970). Insect sperm . their structure and morphogenesis. J Cell Biol *44*, 243.

Pickett-Heaps, J. (1971). Autonomy of the centriole - fact or fallacy? Cytobios *3*.

Piperno, G., Huang, B., and Luck, D.J.L. (1977). 2-dimensional analysis of flagellar proteins from wild-type and paralyzed mutants of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences of the United States of America *74*, 1600-1604.

Pirner, M.A., and Linck, R.W. (1994). Tektins are heterodimeric polymers in flagellar microtubules with axial periodicities matching the tubulin lattice. Journal of Biological Chemistry *269*, 31800-31806.

Purschke, G., Arendt, D., Hausen, H., and Muller, M.C. (2006). Photoreceptor cells and eyes in Annelida. Arthropod Structural Development *35*, 211-230.

Renzaglia, K.S., Bernhard, D.L., and Garbary, D.J. (1999). Developmental ultrastructure of the male gamete of *Selaginella*. International Journal of Plant Sciences *160*, 14-28.

Renzaglia, K.S., Dengate, S.B., Schmitt, S.J., and Duckett, J.G. (2002). Novel features of *Equisetum arvense* spermatozoids: insights into pteridophyte evolution. New Phytologist *154*, 159-174.

Renzaglia, K.S., and Duckett, J.G. (1991). Towards an understanding of the differences between the blepharoplasts of mosses and liverworts, and comparisons with hornworts, biflagellate lycopods and charophytes - a numerical-analysis. New Phytologist *117*, 187-208.

Renzaglia, K.S., and Garbary, D.J. (2001). Motile gametes of land plants: Diversity, development, and evolution. Critical Reviews in Plant Sciences *20*, 107-213.

Renzaglia, K.S., Johnson, T.H., Gates, H.D., and Whittier, D.P. (2001). Architecture of the sperm cell of *Psilotum*. American Journal of Botany *88*, 1151-1163.

Renzaglia, K.S., Wood, K.D., Rupp, G., and Hickok, L.G. (2004). Characterization of the sleepy sperm mutant in the fern *Ceratopteris richardii*: A new model for the study of axonemal function. Canadian Journal of Botany-Revue Canadienne De Botanique *82*, 1602-1617.

Reynolds, E.S. (1963). Use of lead citrate at high pH as an electron-opaque stain in electron microscopy. Journal of Cell Biology *17*, 208-&.

Ringo, D.L. (1967). Flagellar motion and the fine structure of the flagellar apparatus in *Chlamydomonas*. The Journal of Cell Biology *33*, 543-571.

Robbins, R.R., and Carothers, Z.B.J. (1978). Spermatogenesis in lycopodium - mature spermatozoid. American Journal of Botany *65*, 433-440.

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. Current Biology *17*, 1465-1472.

Ronquist, F., and Deans, A.R. (2010). Bayesian phylogenetics and its influence on Insect systematics. Annual Review of Entomology *55,* 189-206.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572-1574.

Rosenbaum, J. (2002). Intraflagellar transport. Current Biology *12*, R125-R125.

Rosenbaum, J., and Child, F.M. (1967). Flagellar regeneration in protozoan flagellates. Journal of Cell Biology *34*, 345.

Rosenbaum, J.L., and Witman, G.B. (2002b). Intraflagellar transport. Nature Reviews Molecular Cell Biolbiology *3*, 813-825.

Sakaushi, S., Okoshi, M., Miyamura, S., and Hori, T.J. (2003). Swimming behavior and ultrastructure of sperm of *Lygodium japonicum* (Pteridophyta). Sexual Plant Reproduction *16*, 113-122.

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

Sanders, M.A., and Salisbury, J.L. (1989). Centrin-mediated microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. Journal of Cell Biology *108*, 1751-1760.

Sankaran, S., Starita, L.M., Simons, A.M., and Parvin, J.D. (2006). Identification of domains of BRCA1 critical for the ubiquitin-dependent inhibition of centrosome function. Cancer Research *66*, 4100-4107.

Satir, P., and Christensen, S.T. (2007). Overview of structure and function of mammalian cilia. Annual Review of Physiology *69*, 377-400.

Satir, P., and Christensen, S.T. (2008). Structure and function of mammalian cilia. Histochemical Cell Biology *129*, 687-693.

Satir, P., Pedersen, L.B., and Christensen, S.T. (2010). The primary cilium at a glance. Journal of Cell Science *123*, 499-503.

Satoh, G. (2006). Exploring developmental, functional, and evolutionary aspects of amphioxus sensory cells. International Journal of Biological Science *2*, 142-148.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nature Genetics *37*, 501-506.

Schmidt, T.I., Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Lavoie, S.B., Stierhof, Y.D., and Nigg, E.A. (2009). Control of centriole length by CPAP and CP110. Current Biology *19*, 1005-1011.

Sherwin, T., and Gull, K.J. (1989). The cell-division cycle of *Trypanosoma brucei brucei* - timing of event markers and cytoskeletal modulations. Philosophical Transactions of the Royal Society of London B *323*, 573.

Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. Nature *457*, 818-823.

Sieburth, J.M., Johnson, P.W., and Hargraves, P.E. (1988). Ultrastructure and ecology of *Aureococcus anophagefferens* Gen-Et-Sp-Nov (Chrysophyceae) - the dominant picoplankter during a bloom in Narragansett Bay, Rhode-Island, Summer 1985. Journal of Phycology *24*, 416-425.

Silflow, C.D., and Lefebvre, P.A. (2001). Assembly and motility of eukaryotic cilia and flagella. Lessons from *Chlamydomonas reinhardtii*. Plant Physiology *127*, 1500-1507.

Silverman, M., and Simon, M. (1974). Flagellar rotation and mechanism of bacterial motility. Nature *249*, 73-74.

Sinden, R.E., Canning, E.U., Bray, R.S., and Smalley, M.E. (1978). Gametocyte and gamete development in *Plasmodium falciparum*. Proceedings of the Royal Society of London B Biological Sciences *201*, 375-399.

Singla, V., and Reiter, J.F. (2006). The primary cilium as the cell's antenna: Signaling at a sensory organelle. Science *313*, 629-633.

Sleigh, M. (1979). Contractility of the roots of flagella and cilia. Nature *277*, 263-264.

Soanes, D.M., Alam, I., Cornell, M., Wong, H.M., Hedeler, C., Paton, N.W., Rattray, M., Hubbard, S.J., Oliver, S.G., and Talbot, N.J. (2008). Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. Plos ONE *3*, 15.

Song, M.H., Aravind, L., Muller-Reichert, T., and O'Connell, K.F. (2008). The conserved protein SZY-20 opposes the Plk4-related kinase ZYG-1 to limit centrosome size. Developmental Cell *15*, 901-912.

Sorokin, S. (1962). Centrioles and formation of rudimentary cilia by fibroblasts and smooth muscle cells. Journal of Cell Biology *15*, 363.

Stoetzel, C., Muller, J., Laurier, V., Davis, E.E., Zaghloul, N.A., Vicaire, S., Jacquelin, C., Plewniak, F., Leitch, C.C., Sarda, P.*, et al.* (2007). Identification of a novel BBS gene (BBS12) highlights the major role of a vertebrate-specific branch of chaperonin-related proteins in Bardet-Biedl syndrome. American Journal of Human Genetics *80*, 1-11.

Tang, C.J., Fu, R.H., Wu, K.S., Hsu, W.B., and Tang, T.K. (2009). CPAP is a cell-cycle regulated protein that controls centriole length. Nature Cell Biology *11*, 825-831.

Turner, F.R. (1968). An ultrastructural study of plant spermatogenesis - spermatogenesis in Nitella. Journal of Cell Biology *37*, 370.

Vaughan, S., and Gull, K. (2003). The trypanosome flagellum. Journal of Cell Science *116*, 757-759.

Vaughn, K.C., and Renzaglia, K.S. (2006). Structural and immunocytochemical characterization of the *Ginkgo biloba L*. sperm motility apparatus. Protoplasma *227*, 165-173.

Verleysdonk, S., Hirschner, W., Pogoda, H.M., Kramer, C., Thiess, U., Hamprecht, B., Wiesmuller, K.H., and Lautner, M. (2007). Biosynthesis of Wdr16, a marker protein for kinocilia-bearing cells, starts at the time of kinocilia formation in rat, and wdr16 gene knockdown causes hydrocephalus in zebrafish. Journal of Neurochemistry *101*, 274-288.

Vouilloud, A.A., Caceres, E.J., and Leonardi, P.I. (2005). Changes in the absolute configuration of the basal/flagellar apparatus and evidence of centrin during male gametogenesis in *Chara contraria* var*. nitelloides* (Charales, Charophyta). Plant Systematics and Evolution *251*, 89-105.

Warne, T.R., Walker, G.L., and Hickok, L.G.J. (1986). A novel method for surface-sterilizing and sowing fern spores. American Fern Journal *76*, 187-188.

Wickstead, B., Carrington, J.T., Gluenz, E., and Gull, K. (2010). The expanded Kinesin-13 repertoire of trypanosomes contains only one mitotic kinesin indicating multiple extra-nuclear roles. PloS ONE *5*.

Wickstead, B., and Gull, K. (2007). Dyneins across eukaryotes: A comparative genomic analysis. Traffic *8*, 1708-1721.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A.M. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Molecular and Biochemical Parasitology *99*, 89-101.

Woodland, H.R., and Fry, A.M. (2008). Pix proteins and the evolution of centrioles. PloS One *3*.

Woolley, D.M., and Nickels, S.N. (1985). Microtubule termination patterns in mammalian sperm flagella. Journal of Ultrastructure Research *90*, 221-234.

Yano, Y., and Mikinoumura, T. (1981). Recovery of sliding ability in arm-depleted flagellar axonemes after recomination with extracted dynein. Journal of Cell Science *48*, 223-239.

Yoder, B.K., Hou, X.Y., and Guay-Woodford, L.M. (2002). The polycystic kidney disease proteins, Polycystin-1, Polycystin-2, Polaris, and Cystin, are co-localized in renal cilia. Journal of the American Society of Nephrology *13*, 2508-2516.

Appendix