Stepwise evolution of the centriole-assembly pathway

Zita Carvalho-Santos, Pedro Machado, Pedro Branco, Filipe Tavares-Cadete, Ana Rodrigues-Martins, José B. Pereira-Leal* and Mónica Bettencourt-Dias*

Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal
*Authors for correspondence (jleal@igc.gulbenkian.pt; mdias@igc.gulbenkian.pt)

Summary

The centriole and basal body (CBB) structure nucleates cilia and flagella, and is an essential component of the centrosome, underlying eukaryotic microtubule-based motility, cell division and polarity. In recent years, components of the CBB-assembly machinery have been identified, but little is known about their regulation and evolution. Given the diversity of cellular contexts encountered in eukaryotes, but the remarkable conservation of CBB morphology, we asked whether general mechanistic principles could explain CBB assembly. We analysed the distribution of each component of the human CBB-assembly machinery across eukaryotes as a strategy to generate testable hypotheses. We found an evolutionarily cohesive and ancestral module, which we term UNIMOD and is defined by three components (SAS6, SAS4/CPAP and BLD10/CEP135), that correlates with the occurrence of CBBs. Unexpectedly, other players (SAK/PLK4, SPD2/CEP192 and CP110) emerged in a taxon-specific manner. We report that gene duplication plays an important role in the evolution of CBB components and show that, in the case of BLD10/CEP135, this is a source of tissue specificity in CBB and flagella biogenesis. Moreover, we observe extreme protein divergence amongst CBB components and show experimentally that there is loss of cross-species complementation among SAK/PLK4 family members, suggesting species-specific adaptations in CBB assembly. We propose that the UNIMOD theory explains the conservation of CBB architecture and that taxon- and tissue-specific molecular innovations, gained through emergence, duplication and divergence, play important roles in coordinating CBB biogenesis and function in different cellular contexts.

Key words: Centriole, Basal body, Evolution, *Drosophila*, Flagella, Comparative genomics

Introduction

The structure of the centriole and the basal body (CBB) is remarkably conserved, comprising microtubule triplets arranged in a ninefold symmetrical configuration (Fig. 1). CBBs are found in all crown eukaryotic groups (Fig. 2A,B; supplementary material Table S1), as a centriole, within the context of a centrosome, and/or as a basal body, tethered to the membrane. This suggests that they were present in the last eukaryotic common ancestor (LECA) (Azimzadeh and Bornens, 2004; Cavalier-Smith, 2002) and that secondary loss occurred in specific branches, such as yeasts and higher plants (Fig. 2A,B; supplementary material Table S1). The conservation of CBB architecture and its structural assembly intermediates (Fig. 1) suggests the existence of common molecular assembly machinery, already present in the LECA. On the other hand, CBBs are assembled in a multiplicity of contexts, such as different cell-cycle phases or cellular locations, suggesting the need for tailored assembly pathways. Moreover, CBBs can have a wide range of functions (Beisson and Wright, 2003; Bettencourt-Dias and Glover, 2007; Delattre and Gonczy, 2004): in humans, they assemble centrosomes, and motile and sensory cilia; in *Caenorhabditis elegans*, they never form motile cilia; and in green algae, such as *Chlamydomonas*, they only form motile cilia. The conservation of the structure contrasts with the diversity of assembly contexts and functions, raising an interesting paradox.

To investigate CBB assembly in eukaryotes, we focused on the evolution of the molecular mechanisms that control this process. We used comparative genomics, a strategy that brought major insights into the origin and evolution of the assembly of cellular structures such as the nuclear pore complex (Devos et al., 2004; Mans et al., 2004), the peroxisome (Gabaldon et al., 2006) and cilia (Avidor-Reiss et al., 2004; Li et al., 2004; Wickstead and Gull, 2007). We focused on six proteins shown to be required for CBB biogenesis in humans (Fig. 1): SPD2/CEP192, SAK/PLK4, SAS6, SAS4/CPAP, BLD10/CEP135 and CP110 (Cunha-Ferreira et al., 2009a; Kleylein-Sohn et al., 2007). Orthologs of some of these proteins have been functionally described in other species (Fig. 1).

Results

A molecular toolkit to detect the CBB-assembly machinery

CBB proteins have eluded automatic comparative genomics screens for novel ciliary components (Avidor-Reiss et al., 2004; Baron et al., 2007; Li et al., 2004). They generally contain several coiled-coil domains (Fig. 3; supplementary material Fig. S1), which carry little phylogenetic signal (Rose et al., 2005). Our detailed bioinformatics analysis of each protein family revealed new conserved regions, other than coiled-coil regions (Fig. 3; supplementary material Figs S1-S6), that characterize each protein with previously untapped phylogenetic depth and breath. Our detailed approach also included the characterization of the phylogenetic distribution of known domains within specific taxonomical groups (e.g. the polo boxes of PLKs).

A core ancestral module defines the centriole ninefold symmetry

The universality of the CBB structure suggests the existence of an ancestral CBB-assembly mechanism. Recent studies have, in fact, suggested that several components of the flagella apparatus, such as the molecules needed to make the motile axoneme, are likely to be ancestral (Avidor-Reiss et al., 2004; Li et al., 2004; Wickstead and Gull, 2007).
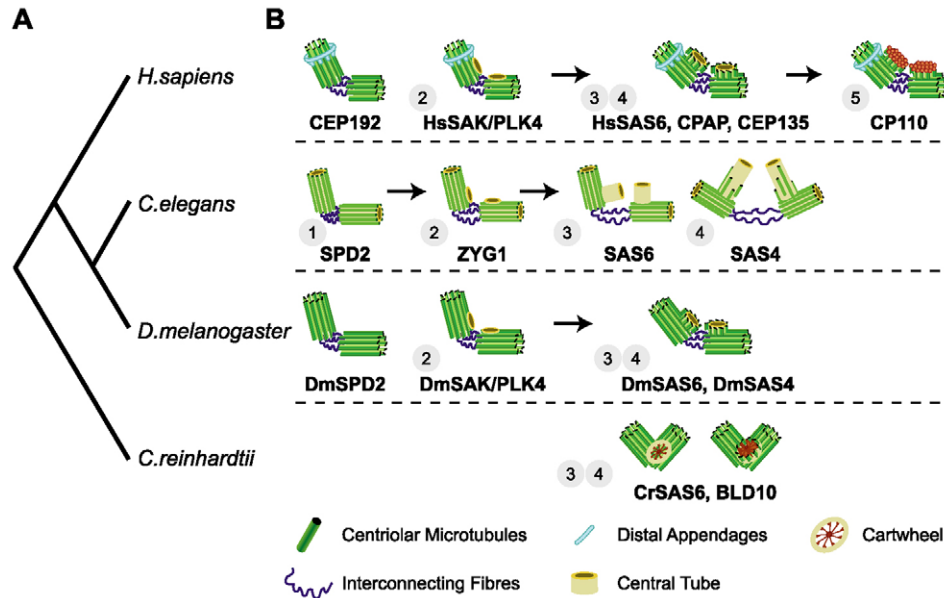
Fig. 1. Molecules and intermediate structures in CBB biogenesis. (A) Simplified taxonomic tree of organisms in which players in CBB biogenesis were identified and characterized. (**B**) Schematic representation of the stage at which each molecule is active in the process of CBB biogenesis. Only molecules that have been identified and characterized in humans are shown. (1) SPD2 has a role in recruiting PCM and ZYG1 to the centriole; (2) two protein kinases, ZYG1 in *C. elegans* and SAK/PLK4 in *D. melanogaster* and human cells, trigger centriole formation; (3) SAS6 and BLD10/CEP135 play an important role in cartwheel formation, an intermediate structure that enforces the ninefold symmetry; (4) SAS4/CPAP tethers and/or stabilizes centriolar microtubules; (5) CP110 is located in the distal part of the centriole and is necessary for centriole reduplication in human cells. Arrows indicate temporal sequence. Orthologs for which there is no clear experimental evidence of their specific role in the process are not represented. More extensive reviews on centriole biogenesis and references are available (Bettencourt-Dias and Glover, 2007; Nigg, 2007; Strnad and Gonczy, 2008).

To investigate the existence of such a universal CBB-assembly mechanism, we searched for homologs of known CBB-assembly proteins in a set of 26 representative eukaryotic species, covering the crown eukaryotic groups and representing the diversity of function and architecture (including absence) of CBBs (Fig. 2A,B; see supplementary material Tables S1 and S2). We calculated the correlation between the presence of each molecule and the presence of the CBB, using a normalized Hamming distance (Fig. 2). Given the poor annotation of the proteomes of certain species and the absence of structural information regarding the existence of a CBB in others, we arbitrarily defined that the presence of a molecule and the occurrence of the CBB structure were correlated if this occurred in at least 80% of the species (Fig. 2). To our surprise, given the conservation of the CBB structure, only a subset of CBB-assembly proteins obey the criteria above defined: SAS4/CPAP, SAS6 and BLD10/CEP135 (Fig. 2). This evolutionarily cohesive behavior suggests that these three molecules are part of the same functional ancestral module in CBB assembly, which, for simplicity, we will call UNIversal MODule (UNIMOD). Amongst the six studied families, the UNIMOD components are, in fact, the only ones required to define the CBB architecture: SAS6 and BLD10/CEP135 form the cartwheel, a structure involved in the specification and stabilization of CBB ninefold symmetry (Fig. 1) (Hiraki et al., 2007; Matsuura et al., 2004; Nakazawa et al., 2007; Rodrigues-Martins et al., 2007a), whereas SAS4/CPAP is required for assembling or stabilizing elongating centriolar microtubules (Fig. 1) (Dammermann et al., 2008; Kohlmaier et al., 2009; Pelletier et al., 2006; Schmidt et al., 2009; Tang et al., 2009). Our results suggest that the conservation of the CBB structure in the eukaryotic tree of life is achieved by the preservation of an

assembly mechanism based on a set of conserved structural components – the UNIMOD. Similar profiles have been assigned to axonemal proteins that are present in organisms such as green algae, humans and trypanosomes, but missing from the higher land plants (Avidor-Reiss et al., 2004; Li et al., 2004; Wickstead and Gull, 2007).

Predicting extra components of the ancestral assembly pathway: PLKs trigger CBB formation

SAK/PLK4 (a polo-like kinase) is indispensable for centriole biogenesis in human cells and *Drosophila melanogaster* (Bettencourt-Dias et al., 2005; Habedanck et al., 2005; Kleylein-Sohn et al., 2007; Rodrigues-Martins et al., 2007b). High levels of this protein lead to the appearance of supernumerary centrioles through either canonical (Bettencourt-Dias et al., 2005; Habedanck et al., 2005) or de novo (Peel et al., 2007; Rodrigues-Martins et al., 2007b) biogenesis. Because of its importance, we were surprised to observe that SAK/PLK4 is not part of the UNIMOD and is only found in opisthokonts (purple clades in Figs 2 and 4); we therefore investigated what could be triggering CBB biogenesis in other groups. Gene duplication is believed to play a major role in generating complexity of cellular mechanisms in evolution (Ohno, 1970). We tested whether other PLK family members could play a role in CBB biogenesis in other groups. We found that PLKs are present in all branches of the eukaryotic tree of life (Figs 2 and 4). The PLKs outside the opisthokonts contain a kinase domain that clusters with opisthokont Polo/PLK1 rather than SAK/PLK4 (Fig. 4), and possess two polo boxes, similar to Polo/PLK1 (supplementary material Fig. S5D). This suggests that a Polo/PLK1-like protein is the ancestral member of the family that duplicated,

giving rise to SAK/PLK4 prior to the divergence of fungi and animals (Figs 2 and 4). Our results support the scenario that an ancestral Polo/PLK1 triggered CBB biogenesis in the LECA. This is further supported by two observations: human PLK1 (Liu and Erikson, 2002; Tsou et al., 2009) and human PLK2 (Warnke et al., 2004) play a role in centriole duplication, suggesting the presence
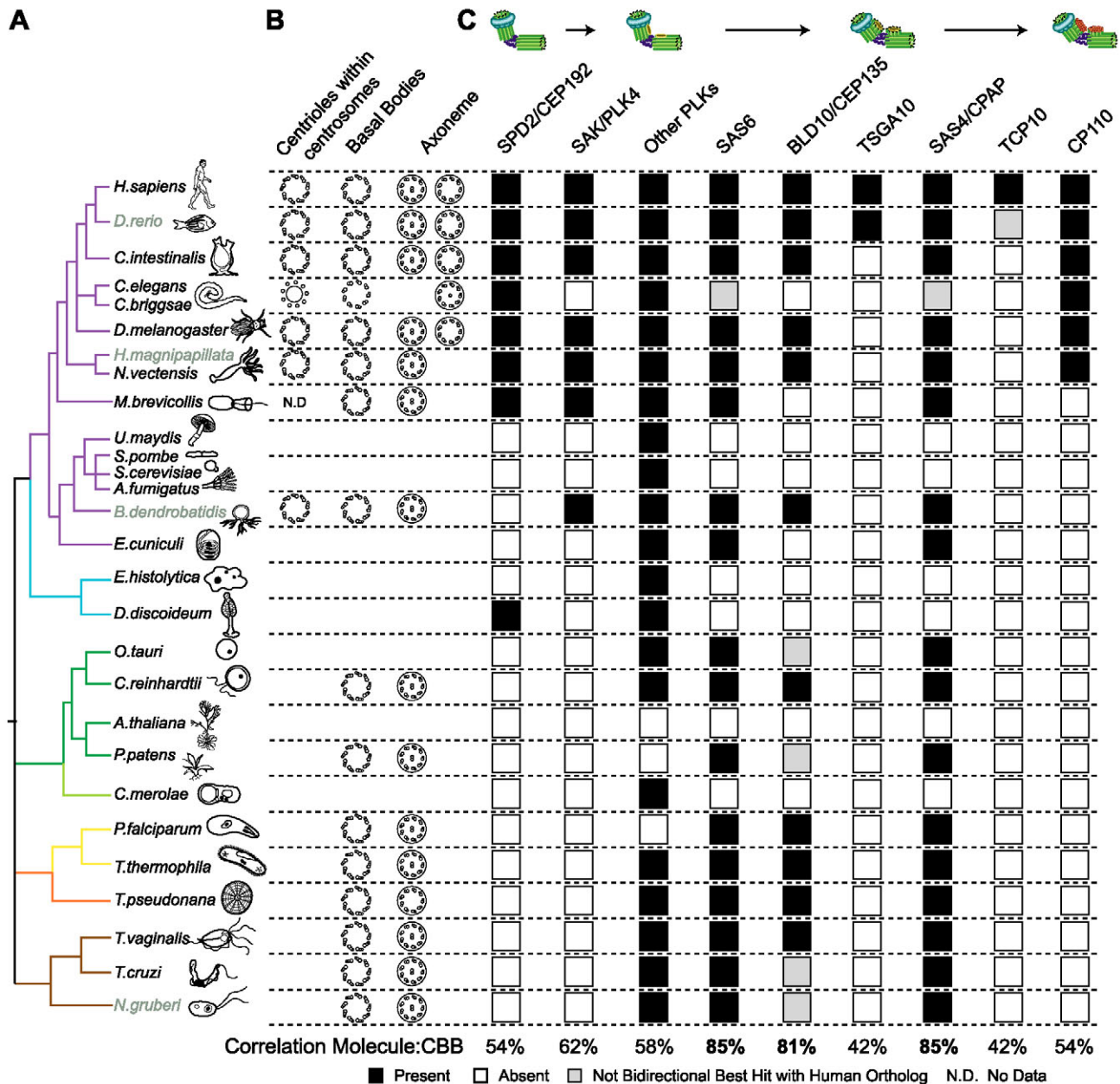
Correlation Molecule:CBB

	Centrioles within centrosomes	Basal Bodies	Axoneme	SPD2/CEP192	SAK/PLK4	Other PLKs	SAS6	BLD10/CEP135	TSGA10	SAS4/CPAP	TCP10	CP110
	54%	62%	58%	85%	81%	42%	85%	42%	54%			

■ Present   □ Absent   ▦ Not Bidirectional Best Hit with Human Ortholog   N.D. No Data

Fig. 2. Phylogenetic distribution of families of molecular players in CBB biogenesis. (A) Simplified taxonomic tree representing crown eukaryotic groups in different colors (these groups contain a common ancestor and all its descendants) [modified from Baldauf (Baldauf, 2003) and Hedges (Hedges, 2002)]. Unikonts include eukaryotic cells that, for the most part, have a single emergent flagellum. Unikonts can be divided into opisthokonts (which propel themselves with a single posterior flagellum; animals, fungi and choanoflagellates) and amebozoa (Cavalier-Smith, 2002). Bikonts include eukaryotic organisms with two emergent flagella (Cavalier-Smith, 2002). Branch color code: purple – opisthokonts, blue – amebozoa, green – plants, yellow – alveolates, orange – heterokonts, brown – excavates and discicristates. Organisms with incomplete genomes are indicated in grey. Detailed taxonomical and genome-sequencing status information is found in supplementary material Table S2. **(B)** Distribution of the structures present in each organism: internal architecture of the centrioles (within the centrosome), basal bodies and axonemes of cilia and/or flagella present in each species (see supplementary material Table S1 for details). In some cases, because of lack of information, we used structural information from other species in the same group (see supplementary material Table S1). Note that the presence of CBBs correlates with the occurrence of a flagellated stage. The basal body was lost in species that do not assemble flagella, such as yeasts (*S. pombe, S. cerevisiae*). **(C)** Phylogenetic distribution of molecular players in CBB biogenesis. Protein families and their role in CBB biogenesis, as indicated in Fig. 1, are schematized at top. Black boxes represent the presence of homologous sequences that are bidirectional best hits to family members in humans; grey boxes represent homologous sequences that are bidirectional best hits to family members in organisms other than human; white boxes indicate no detectable ortholog.

of a residual function in this process; and in *Trypanosoma brucei*, the depletion of PLK1 leads to defects in basal-body duplication and cytokinesis (Hammarton et al., 2007).

What could be the consequences of this duplication event? In humans and *D. melanogaster*, Polo/PLK1 is known to have important roles in the cell cycle, such as entry and progression in mitosis and cytokinesis, and γ-tubulin recruitment to the centrosome (Archambault and Glover, 2009). This explains the presence of PLKs in species that do not assemble CBBs. On the other hand, since SAK/PLK4 emerged, it became strictly correlated with CBBs,
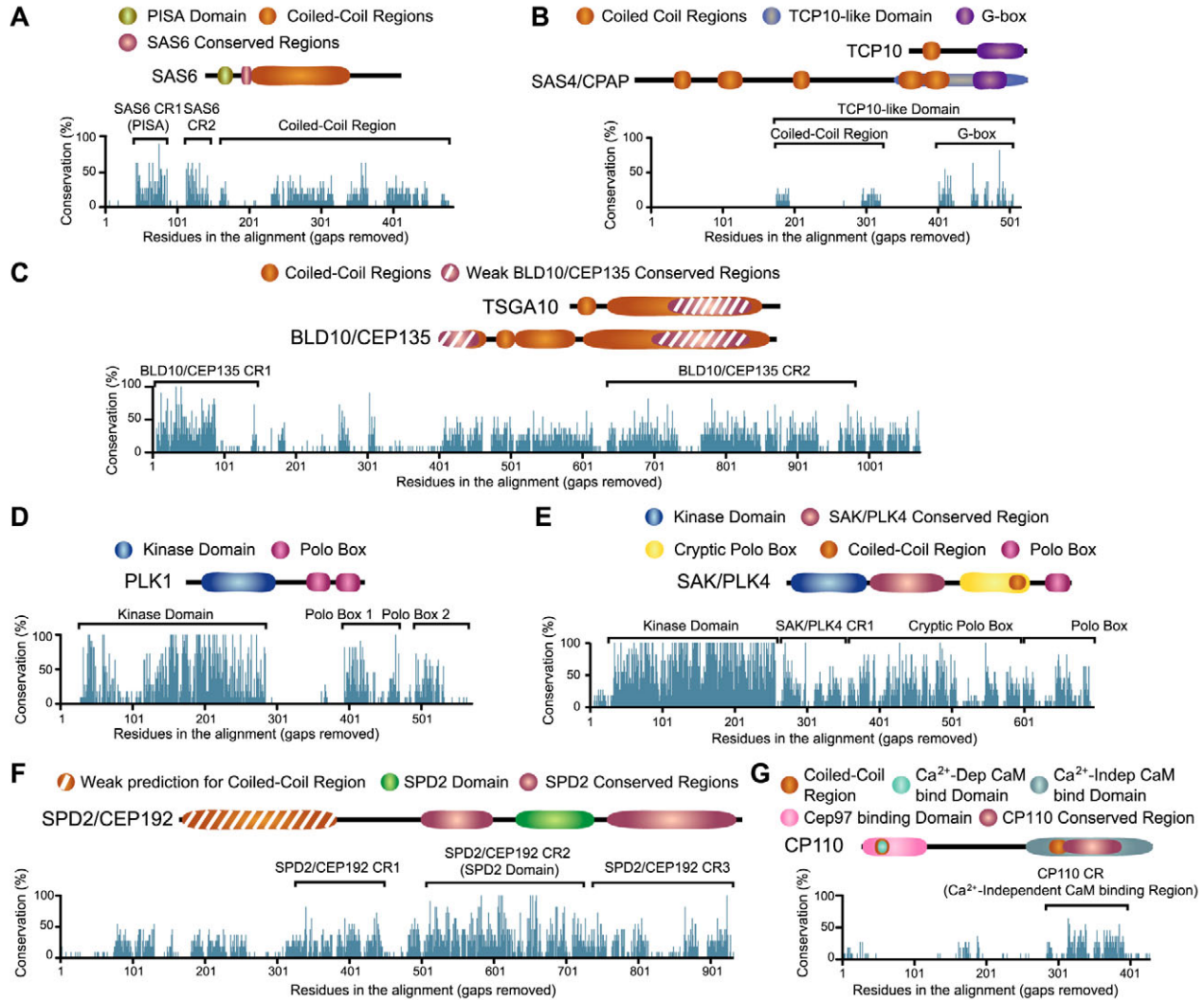
Fig. 3. Sequence features of the CBB-assembly proteins. For each protein family, we show the architecture of the human sequence, but it can vary slightly among different organisms (e.g. SAS4/CPAP in supplementary material Fig. S2). We further show a conservation plot of the multiple sequence alignment. Note that gaps of non-aligned residues spanning ≥25% of the total sequence were removed for clarity. All the regions identified in this study as being conserved among orthologs are referred to as conserved regions (CRs) and are represented in the schematic of each protein. HMMs of each region are available from www.evocell.org. (**A**) SAS6 contains a long coiled-coil region spanning residues ~160-480. A previously identified domain named PISA (Present In SAS6) (Leidel et al., 2005) is the first highlighted conserved region (SAS6 CR1: residues 44-87). This domain was further used as an additional criterion to classify SAS6 orthologs. We identified a second conserved region, SAS6 CR2 (residues 123-157), adjacent to the PISA domain. It contains a recently identified phosphorylation residue (Kitagawa et al., 2009). (**B**) SAS4/CPAP orthologs show conservation in a single region corresponding to the G-box (residues 1159-1264). This C-terminal conserved region is characterized by several nonamer motifs with a glycine or a glutamine in the first position, and is part of a larger region called the TCP10 domain (T complex protein 10), which was initially identified in the TCP10 protein (Hung et al., 2000). (**C**) BLD10/CEP135 contains long coiled-coil domains that mask their sequence similarity. Two conserved regions were identified (BLD10/CEP135 CR1: residues 6-147 and BLD10/CEP135 CR2: residues 678-1021). In vertebrates, we found a protein, TSGA10, that shares a high degree of similarity (65%) with BLD10/CEP135. Both domains, CR1 and CR2, were used to identify homologous sequences. (**D**,**E**) PLKs have an N-terminal kinase domain and one or two C-terminal 'polo boxes'. Human PLK1-3 and *Drosophila* Polo have two Polo boxes, whereas SAK/PLK4 has a single Polo box. SAK/PLK4 has distinct conserved regions: the previously named cryptic polo box (Swallow et al., 2005) (residues 696-836) and a second conserved region between the kinase domain and the cryptic polo box (CR1: residues 283-569) that includes a regulatory phosphodegron (Cunha-Ferreira et al., 2009b; Holland et al., 2010; Rogers et al., 2009). (**F**) SPD2/CEP192 has a conserved region previously named the SPD2 domain (SPD2/CEP192 CR2: residues 1169-1417). We identified two other conserved regions: CR1 (residues 841-1089) and CR3 (residues 1466-1918). (**G**) A conserved region of CP110 was identified as CP110 CR1 (residues 794:910), which localizes within a Ca^{2+}-independent calmodulin-binding region.
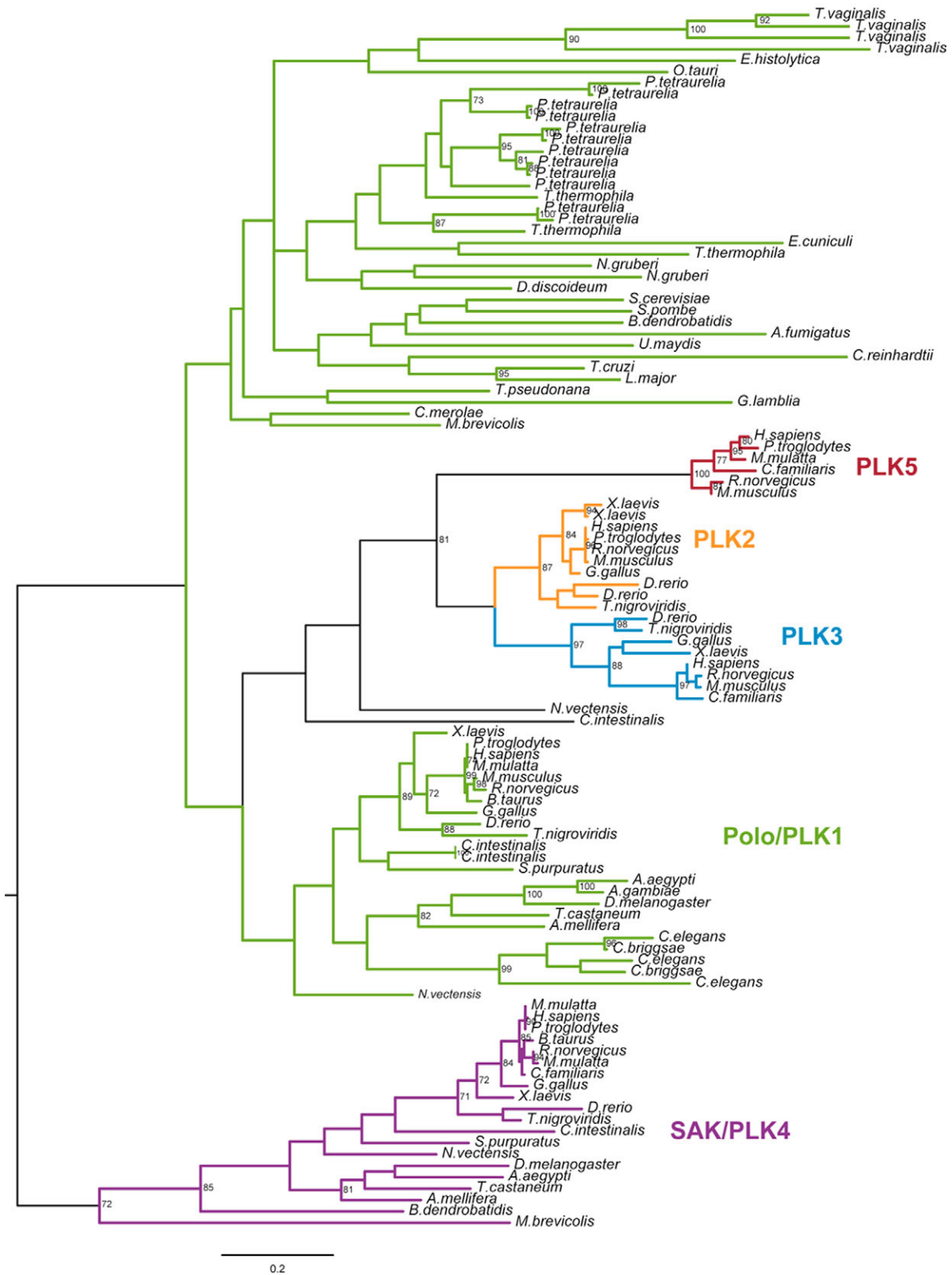
Fig. 4. The kinase domain is sufficient to distinguish all the PLK subfamilies. Maximum likelihood phylogenetic tree of the kinase domain of members of the different PLK subfamilies (a NiMA kinase was used as root). Branch labels represent bootstrap support in percentage. Using this strategy, we were able to distinguish the lineage of SAK/PLK4 from the other members of the PLK family, albeit with low bootstrap support. Please note that PLKs found outside opisthokonts cluster with budding yeast PLK (Cdc5), which is shown to be a functional homolog of human PLK1 (Lee and Erikson, 1997). The fact that *Drosophila* (as well as other insects) contains a single PLK that clusters with human PLK1 supports Polo/PLK1 being the ancestral kinase that underwent several duplication rounds in animals, to give rise to PLK1, 2, 3 and 5.

as shown by its disappearance in the yeasts, in which the CBB was lost concomitant with spindle pole body (SPB) emergence (Fig. 2; supplementary material Table S1). The evidence presented above strongly suggests that an ancestral Polo/PLK1 had both mitotic and CBB biogenesis functions. Upon duplication followed by subfunctionalization, this ancestral Polo/PLK1 generated Polo/PLK1 and SAK/PLK4, allowing the uncoupling of more general cell-cycle functions from CBB biogenesis.

SPD2 and CP110 emerged in a taxon-specific manner

A surprising observation is that SPD2/CEP192 and CP110, two proteins crucial for centriole biogenesis and function in humans, emerged in a taxon-specific manner (Fig. 2). SPD2/CEP192 is present in *Dictyostelium discoideum* (Fig. 2), having been lost in *Entamoeba hystolitica* and at the base of fungi. *D. discoideum* is a well-characterized amoeba that does not assemble CBBs. Instead, it has a microtubule-organizing center (MTOC) called the nucleus-associated body (NAB), where SPD2/CEP192 was recently shown to localize (Schulz et al., 2009). This suggests that the ancestral function of SPD2/CEP192 was pericentriolar material (PCM) recruitment to the MTOC, independent of the presence of CBBs. PCM proteins, such as SPD2, might have acquired a role in recruiting CBB-assembly proteins to the centrosome (Dammermann et al., 2004; Loncarek et al., 2008). In animals, SPD2/CEP192 is essential for CBB biogenesis in contexts in which less PCM is available. In agreement, *C. elegans* and *D. melanogaster* SPD2/CEP192 are essential for the recruitment of PCM to the PCM-naked sperm CBB and its duplication upon fertilization (Dix and Raff, 2007; Kemp et al., 2004; Pelletier et al., 2004). By contrast, *D. melanogaster* SPD2/CEP192 is dispensable for both PCM recruitment and CBB duplication in somatic cells (Dix and Raff, 2007; Giansanti et al., 2008).

CP110 only appeared in animals (Fig. 2). It localizes to a distal centriole compartment, and is needed for centriole reduplication in S-phase-arrested human cells and to define centriole length (Chen et al., 2002; Kleylein-Sohn et al., 2007; Kohlmaier et al., 2009; Schmidt et al., 2009). We hypothesize that CP110 was added to the centriole-assembly pathway in animals as an innovation. We found that a binding partner of CP110, CEP97, has a very similar phylogenetic distribution to CP110 (supplementary material Fig. S7). These results both suggest that the two proteins might work in a complex in all animals and validate the use of phylogenetic distributions as a screening strategy to find potential binding partners. *Drosophila* CP110 and CEP97 localize to centrioles and are necessary for centriole duplication in S2 cells (supplementary material Fig. S8A,B,D,E) (Dobbelaere et al., 2008). CP110 in humans participates in other processes, such as preventing centrioles from nucleating cilia (Kleylein-Sohn et al., 2007; Spektor et al., 2007) and cytokinesis (Tsang et al., 2006). It has been proposed that centrioles might play an important role in signaling the event of cellular abscission in cytokinesis (Piel et al., 2001). It is possible that CP110 emerged in animals to allow further coordination of centriole duplication with ciliogenesis and/or cytokinesis.

Extreme sequence divergence

Our expectation was that, considering the extreme structural conservation of CBBs, we were facing a highly conserved set of components. To our surprise, in the process of defining conserved regions in CBB-assembly components (Fig. 3; supplementary material Figs S1 and S5), we found their sequences to be highly divergent. We explored whether this divergence could underlie the

evolution of CBBs, using conservation scores, an estimate of the divergence of a pair of proteins or conserved protein regions (Lopez-Bigas and Ouzounis, 2004) (Fig. 5). A baseline for conserved molecules are the cell-cycle kinases, whose conservation is evident from the rescue of a *cdc2* fission yeast mutant and a *cdc5* budding yeast mutant by their human CDK1 and PLK1 counterparts, respectively (Lee and Erikson, 1997; Lee and Nurse, 1987). Their conservation scores (CS), calculated between the human sequence and either the *Drosophila* or zebrafish sequences, are $CS_{Drosophila}=0.75$; $CS_{Zebrafish}=0.86$ for CDK1, and $CS_{Drosophila}=0.51$; $CS_{Zebrafish}=0.76$ for PLK1. By contrast, SAK/PLK4 is much more divergent ($CS_{Drosophila}=0.18$; $CS_{Zebrafish}=0.25$; Fig. 5A,B). This divergence is more pronounced outside the kinase domain (Fig.
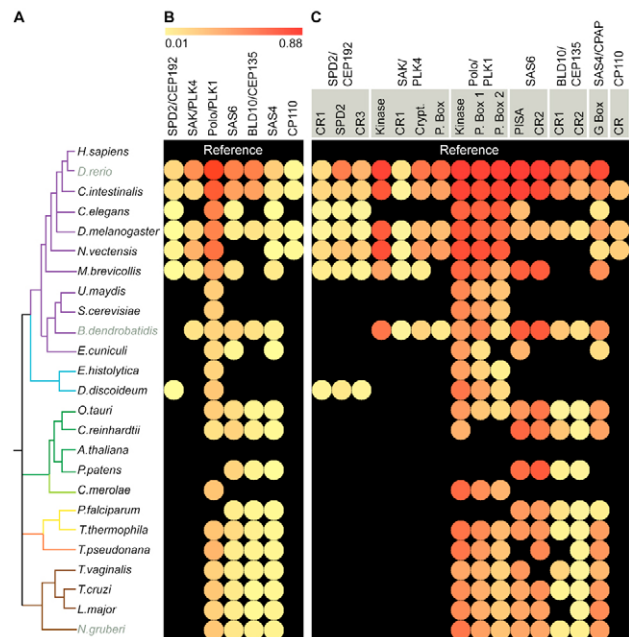

Fig. 5. Conservation of complete sequences and specific regions in the CBB-assembly machinery. (**A**) Simplified taxonomic tree of organisms representing the major eukaryotic branches, as shown in Fig. 2. (**B**) Heat map representing the global sequence conservation scores for each protein family. The conservation score is calculated as the BLAST score of the human protein against each ortholog normalized by the BLAST score of the human protein against itself (Lopez-Bigas et al., 2008; Lopez-Bigas and Ouzounis, 2004). This score can have a value between 0 and 1; 1 indicating full sequence conservation. We used the human ortholog as a reference protein. The color scale is shown at top. Colors closer to red indicate high conservation and, conversely, colors closer to yellow indicate low conservation. (**C**) Conservation scores for specific protein regions present in each protein family, using the same color scale as in B. The names of each conserved region are indicted in the grey boxes above the heat map (CR – conserved region; Crypt. – cryptic Polo box; P. Box – Polo box). Sequences are not scored if they are not present in a given organism. We also excluded sequences (*Danio rerio* CP110 CR, *C. elegans* SAS6 CR2, *Trichomonas vaginalis* SAS6 CR2, *Thalassiosira pseudonana* SAS6 PISA, *Physcomitrella patens* SAS4/CPAP G box and *E. cuniculi* Polo/PLK1 P. Box 2) that do not align with the human ortholog. *Nematostella vectensis* SAS6 and BLD10/CEP135 are truncated and hence not included. *Monosiga brevicollis* SAK/PLK4 P. Box, *Chlamydomonas reinhardtii* Polo/PLK1 P. Boxes 1 and 2, *T. pseudonana* BLD10/CEP135 CR1 and *Trypanosoma cruzi* BLD10/CEP135 CR2 are not present in the protein sequence (possibly as a result of gene annotation problems).

5A,C), which leads us to hypothesize that there was a fast change in the regulation of this enzyme on the evolutionary timescale.

We tested this hypothesis experimentally, taking advantage of the fact that overexpression of both *D. melanogaster* and human SAK/PLK4 leads to overduplication of centrioles (Bettencourt-Dias et al., 2005; Habedanck et al., 2005; Kleylein-Sohn et al., 2007; Rodrigues-Martins et al., 2007b). Whereas human SAK/PLK4

induced centriole amplification in human osteosarcoma cells (U2OS), the *D. melanogaster* counterpart did not, despite being able to localize to centrioles (Fig. 6A,B) and being expressed at similar or higher levels (supplementary material Fig. S9A). The reverse was also true, human SAK/PLK4 did not induce centriole amplification in *Drosophila* S2 cells (Fig. 6C,D; supplementary material Fig. S9B). It is thus possible that the divergence of these sequences has
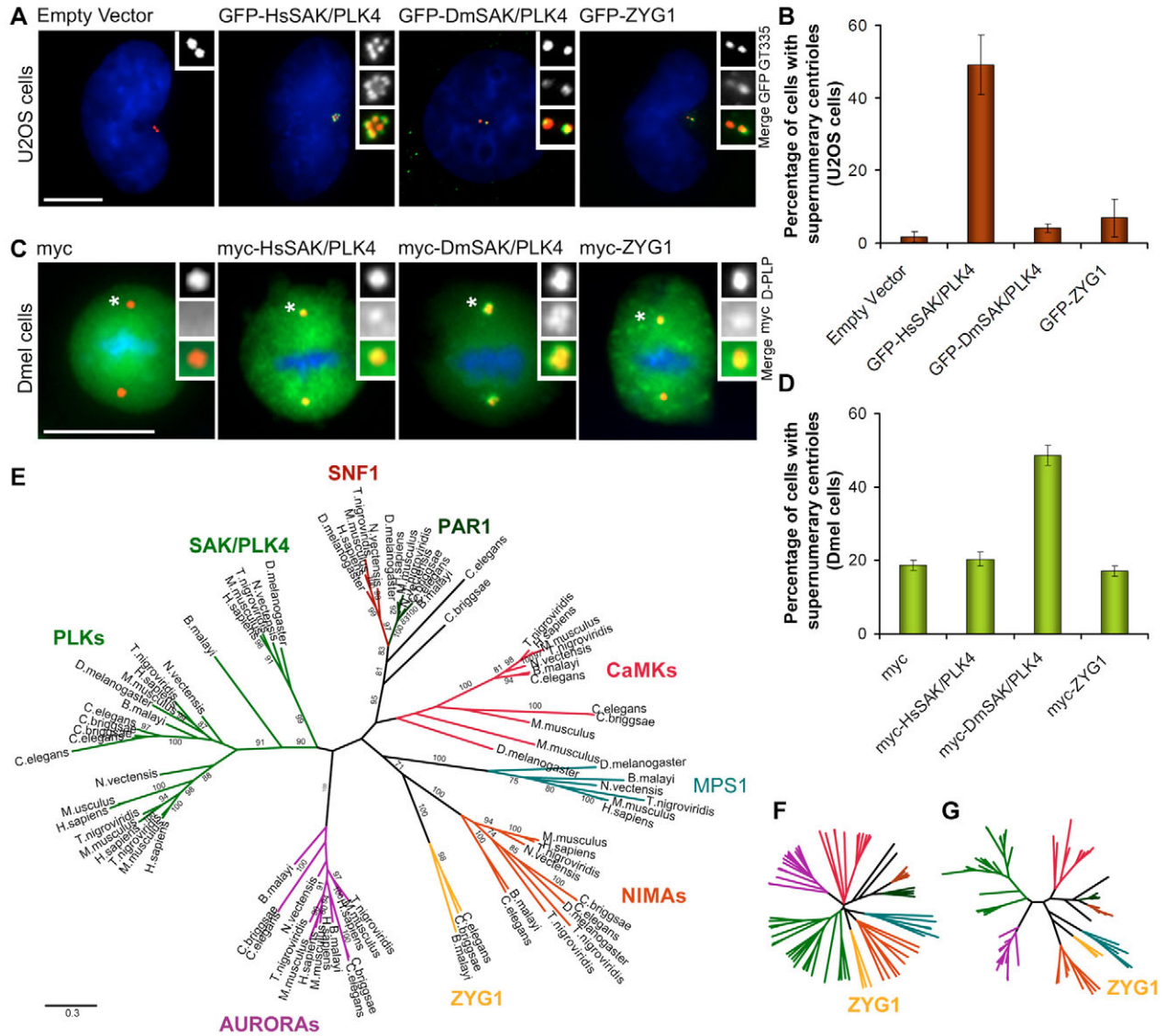

Fig. 6. Divergence and loss of cross-species complementation in CBB biogenesis. (**A**) Human osteosarcoma cells (U2OS) transfected with GFP-HsSAK/PLK4 show centriole amplification. GFP-DmSAK/PLK4 and GFP-ZYG1 localize to U2OS centrioles, but do not promote amplification (insets show 4.5× magnification). Centrioles are labeled in red (polyglutamylated tubulin-GT335); GFP-HsSAK/PLK4, GFP-DmSAK/PLK4 and GFP-ZYG-1 are in green; DNA is in blue. Scale bar: 10 μm. (**B**) Percentage of cells showing supernumerary centrioles (more than four). The mean ± s.e.m. of three experiments is shown, with a minimum of 100 cells counted in each experiment. (**C**) Expression of myc-DmSAK/PLK4, but not myc-HsSAK/PLK4 or myc-ZYG1, leads to centriole amplification in *Drosophila* tissue-culture cells. These proteins localize to centrioles during mitosis (insets show 2.25× magnification of the centrosome area, which is marked by an asterisk). Centrioles are labeled in red (D-PLP); myc-HsSAK/PLK4, myc-DmSAK/PLK4 and myc-ZYG-1 are in green; DNA is in blue. Scale bar: 10 μm. (**D**) Percentage of cells showing supernumerary centrioles (more than four). The mean ± s.e.m. of three experiments is shown, with a minimum of 200 cells counted in each experiment. Note that wild-type Dmel cells, as with other *Drosophila* tissue-culture cells, have a large population showing centriole amplification. (**E**) Maximum likelihood phylogenetic tree of the catalytic domains of different kinase families closely related in sequence to PLKs and ZYG1. The phylogenetic analysis was performed using a small subset of animal sequences. The number on the branches represents the number of times the branch was observed in 100 bootstraps. (**F,G**) Neighbor joining (F) and Bayesian (G) trees of the same kinase sequences. The kinase families are colored as in E, supporting the results obtained in E.

functional implications, leading to changes in protein regulation in a taxon-specific manner.

Taxon-specific divergence might be extreme in *C. elegans*, for which we did not find a SAK/PLK4 ortholog (Figs 2 and 4). The kinase ZYG1 in worms plays an important role upstream of SAS6 and SAS4, similar to human SAK/PLK4 (Bettencourt-Dias et al., 2005; Delattre et al., 2006; Habedanck et al., 2005; Kleylein-Sohn et al., 2007; Pelletier et al., 2006), and has been speculated to be its ortholog (Bettencourt-Dias et al., 2005; Song et al., 2008). When expressed in human and *Drosophila* cells, ZYG1 localized to centrosomes (Fig. 6A,C), although it did not induce centriole amplification (Fig. 6A-D). We further investigated the relationship of these kinases. We analyzed the phylogeny of their kinase domains and compared the structures of the C termini of ZYG1 and SAK/PLK4. We found a strongly supported monophyletic group of PLKs that included the known *C. elegans* PLKs 1-3, but not ZYG1, which is more similar to the centrosome kinases NIMA and MPS1 (Fig. 6E-G). Using fold recognition (3D-PSSM) (Kelley et al., 1999), we detected polo boxes in the C termini of both Polo/PLK1 and SAK/PLK4 kinases, but not in ZYG1 (data not shown). Moreover, we generated hidden Markov models (HMMs) of the so-called 'cryptic polo box' domain of animal SAK/PLK4, which targets it to the centrosome (Habedanck et al., 2005). This model was able to detect the distantly related SAK/PLK4 of the fungi *Batrachochytrium dendrobatidis*, but no *C. elegans* protein. The lack of both sequence similarity and supportive phylogenetic models (Fig. 6E-G) strongly supports the hypothesis that these molecules are not orthologs, that is, they do not share the same ancestry. Instead, the fact that ZYG1 can localize to centrosomes in *Drosophila* and human cells, and that it also plays a role upstream of SAS6 and SAS4 in *C. elegans* suggests a scenario of convergent evolution of ZYG1 and SAK/PLK4.

We were surprised to observe that the structural components of the UNIMOD were also very divergent, contrary to other structural proteins, such as tubulins, actins and myosins (Fig. 5B and data not shown). We wondered whether the presence of coiled coils could contribute to UNIMOD divergence. Coiled-coil conservation varies substantially, according to their function: protein-protein interaction motifs diverge very little, whereas protein domains that work as spacers and rods are more divergent [e.g. skeletal muscle myosin and nuclear mitotic apparatus protein (NuMA) diverge 2.1% and 18% between rat and human, respectively] (White and Erickson, 2006). We observed medium (8-12%) to high divergence (22%) of UNIMOD coiled coils, suggesting that these sequences function as spacers or rods (White and Erickson, 2006) and thus contribute to UNIMOD divergence. Supporting this hypothesis for coiled-coil function as rods and spacers is the fact that *Chlamydomonas reinhardtii* BLD10 coiled-coil truncations lead to the assembly of smaller cartwheel spokes (Hiraki et al., 2007; Matsuura et al., 2004) (supplementary material Fig. S6).

In principle, high protein divergence could potentially mask the ancient origin of the non-UNIMOD proteins. However, we think that this is not the case for two main reasons. First, we found proteins with regions showing some degree of similarity but different protein architecture in all eukaryotic branches (Fig. 2; supplementary material Fig. S4). Second, when comparing conserved domains that define the UNIMOD, such as PISA and G-box domains, flagellated fungi and *Chlamydomonas* are less divergent from human than *Drosophila* proteins; however, SPD2, SAK/PLK4 and CP110 were found in *Drosophila* but in none of these other branches.

Tissue specificity through subfunctionalization

We found two paralogs of SAS4/CPAP and BLD10/CEP135 in vertebrates, TCP10 and TSGA10, respectively (Figs 2 and 3). These vertebrate paralogs display the conserved G box and BLD10/CEP135 conserved region 2 (CR2), respectively. These duplicates are, in general, shorter than the ancestor family member present in organisms such as *Chlamydomonas* and *Drosophila*; in the case of TSGA10, it lacks BLD10/CEP135 CR1 (Figs 2, 3, Fig. 7A). What could be the role of these vertebrate paralogs in CBB assembly? *Chlamydomonas* and human BLD10/CEP135 have been shown to be important for early steps in CBB assembly (Hiraki et al., 2007; Kleylein-Sohn et al., 2007; Matsuura et al., 2004). TSGA10 is mainly expressed in testes and its absence is also associated with male sterility in humans (Modarressi et al., 2000). This protein localizes to the flagellum of mouse and bovine sperm (Behnam et al., 2006; Modarressi et al., 2004), suggesting a role in the assembly of sperm flagella. We propose two scenarios to explain this function of TSGA10 in the assembly of sperm flagella: subfunctionalization (partition of ancestral functions into the two duplicates) or neofunctionalization (acquisition of a new function by one duplicate).

We proceeded to test these scenarios in a model organism, *D. melanogaster*, which contains a single BLD10/CEP135 family member. These scenarios can be distinguished by the presence (subfunctionalization) or absence (neofunctionalization) of a *Drosophila* BLD10 (DmBLD10) function in flagella biogenesis, besides the expected role in centriole biogenesis. To test this, we used two approaches, RNAi in tissue culture cells and a mutant fruit-fly stock for BLD10/CEP135 (supplementary material Fig. S8A,C,D; Fig. S10A). We confirmed that DmBLD10 protein is absent from hemizygous mutant spermatocytes, whereas it localizes along centrioles in wild-type flies (supplementary material Fig. S10B). In line with its putative described ancestral function, we and others found that the protein localizes in the centrosomes of Dmel cells and RNAi leads to a decrease in centrosome number (supplementary material Fig. S8A,C-E) (Bettencourt-Dias et al., 2005; Dobbelaere et al., 2008; Rodrigues-Martins et al., 2007a). A role in centriole biogenesis is further supported by the observation that *DmBLD10* mutant spermatocytes show shorter centrioles and premature centriole disengagement associated with defects in meiosis I of spermatogenesis (Fig. 7B-D; supplementary material Fig. S10D-F), similar to other mutants in which centriole biogenesis is impaired (Rodrigues-Martins et al., 2007a). We thus conclude that DmBLD10 is involved in centriole biogenesis, although the consequences of its absence are not as severe compared with *SAS6* mutants (supplementary material Fig. S10G-I) (Bettencourt-Dias et al., 2005; Blachon et al., 2009; Peel et al., 2007; Rodrigues-Martins et al., 2007a).

We investigated a possible role for DmBLD10 in sperm formation. As in humans lacking TSGA10, *DmBLD10* mutant males were sterile, suggestive of sperm malfunction (supplementary material Fig. S10C). The male infertility phenotype was not due to the inability of short centrioles to build axonemes, because the number of axonemes in 64 spermatid cysts of *DmBLD10* mutants was similar to the one observed in the wild type (supplementary material Fig. S10D; Fig. S11A). However, we observed that the central microtubule pair, a structure essential for flagellum motility, was absent in mutant axonemes (Fig. 7E,F). The central pair is nucleated from a distal area of the basal body called the transition zone (McKean et al., 2003). Accordingly, we observed DmBLD10 to localize in a more distal region of the basal body (supplementary material Fig. S11B).
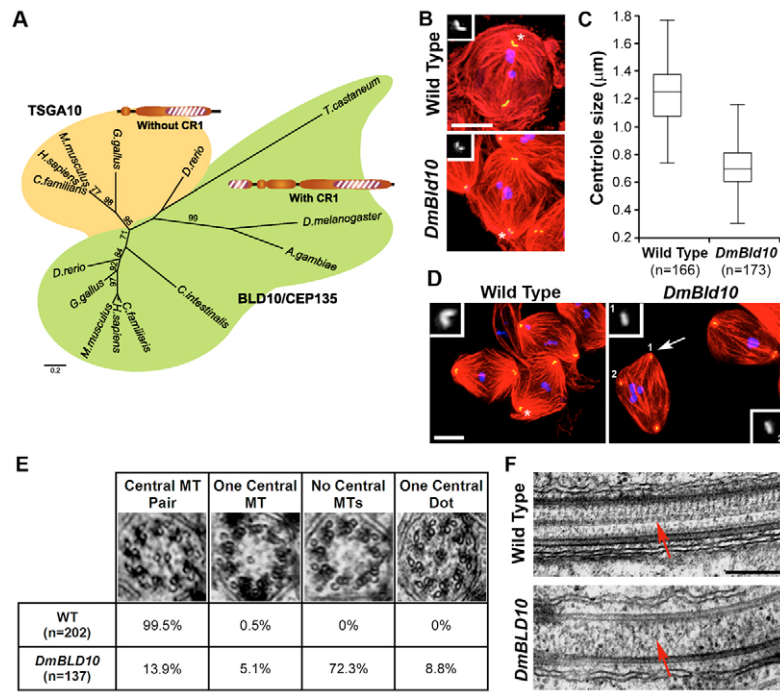
Fig. 7. *Drosophila* **BLD10/CEP135 represents an ancestor of the BLD10/CEP135 and TSGA10 family, and functions in both centriole and flagella biogenesis.** (**A**) Maximum likelihood phylogenetic tree of the CR2 region of BLD10/CEP135 and TSGA10 for a subset of vertebrates and invertebrates. The number on the branches represents the number of times the branch was observed in 100 bootstraps. The presence of CR1 is indicated. A green background indicates proteins with CR1 domain, whereas the orange background includes proteins with no CR1 domain. Note that most ancestral proteins possess a CR1, similar to HsBLD10/CEP135B, but not HsTSGA10. (**B,C**) Centrioles are shorter in *DmBLD10* mutants than in the wild type. The inset shows the GFP-PACT signal corresponding to the centrosome marked with an asterisk (1.5× magnification). Scale bar: 10 μm. (**C**) Centrioles were measured in cells undergoing meiosis I using the GFP-PACT centriolar marker. (**D**) Centrioles split precociously during meiosis I in *DmBLD10* mutants. Mutants occasionally show tripolar spindles (arrow) that correlate with the presence of single centrioles. The inset shows GFP-PACT (4× magnification). Scale bar: 10 μm. (**E**) Axonemes of mutant *DmBLD10* sperm are structurally abnormal. Note that the axonemes maintain the ninefold symmetry, but defects in the central microtubule pair are observed. These defects were quantified and are shown as percentages for each of the categories represented in the table. (**F**) Longitudinal section showing the lack of the central pair in the mutant axoneme (arrows). Scale bar: 300 nm.

Our results and those from a recent report (Mottier-Pavie and Megraw, 2009) suggest that *DmBLD10* mutant males are infertile because this molecule is needed for the assembly of the central microtubule pair of the axoneme. These data clearly support the subfunctionalization scenario, whereby two distinct ancestral functions of BLD10/CEP135 were present in a single protein in animals and were split between duplicates in vertebrates (Fig. 2). In this respect, it is interesting that TCP10, the duplicate of SAS4/CPAP, is mainly expressed in testes and was originally identified as a member of the t-complex locus linked to male sterility (Cebra-Thomas et al., 1991; Schimenti et al., 1988). It will be important to investigate whether this molecule is also involved in flagella biogenesis.

The origin of the CBB-assembly machinery
Our detailed bioinformatics analysis of each protein family revealed the conserved regions (Fig. 3; supplementary material Figs S1-S6) that characterize each protein. These regions, considered together with the UNIMOD, represent a genomic identifier of the CBB. A long-standing debate revolves around the origin of these structures, with suggestions that the flagellum and its basal body have a bacterial origin, resulting from endosymbiosis (Dolan et al., 2002). We can now use these conserved regions to investigate whether the CBB ancestral core has bacterial counterparts. We generated profile HMMs of the conserved regions identified in this study and used them to search a database of 586 bacterial and 50 archaeal genomes. With the exception of the kinase domain of Polo, which is related to many protein kinase domains in bacteria and archaea (Kannan et al., 2007), we could not detect any positive hits suggestive of putative homologous sequences. This result indicates a eukaryotic origin of the CBB.

Discussion
The conservation of the morphology of the CBB structure contrasts with the diversity of contexts in which it assembles and operates

in eukaryotic life. Focusing on the phylogenetic distribution of six proteins essential for centriole assembly in humans, we found that, in contrast to the previously observed conservation of ciliary and flagella components (Avidor-Reiss et al., 2004; Li et al., 2004), CBB- assembly mechanisms evolved in a stepwise fashion (Figs 2 and 8). We propose that a subset of these proteins, which belong to what we call the universal module (UNIMOD), are necessary to define the CBB structure: the ninefold symmetry and the recruitment and tethering of centriolar microtubules. These proteins have a similar phylogenetic distribution to that previously observed for ciliary and flagella components, and it is likely that new centriole components, such as POC1 (Keller et al., 2009; Pearson et al., 2009), will also fall into this subset. Furthermore, the set of proteins needed to form a centriole is likely to be larger than the UNIMOD, including proteins that also have non-centriolar functions and are present in organisms that do not have CBBs, such as α- and γ-tubulins and centrin. Mechanisms such as duplication with subfunctionalization of ancestral components (e.g. PLK and the BLD10/CEP135 families, Figs 6 and 7), divergence (e.g. SAK/PLK4, Figs 4, 5 and 6) and the emergence of new genes (e.g. SPD2/CEP192 and CP110; Fig. 2) play important roles in the evolution of CBB biogenesis. We have shown experimentally that subfunctionalization might have played a role in CBB evolution at least twice. In the case of BLD10/CEP135, duplication and subfunctionalization with the generation of TSGA10 is likely to be important in the development of tissue-specific mechanisms of CBB assembly and flagella formation (Fig. 7). In the case of the PLK family, the appearance of SAK/PLK4 with subfunctionalization (Fig. 4) is likely to play a role in uncoupling the regulation of CBB biogenesis from other cell-cycle events performed by PLKs. We have also shown experimentally that divergence in the PLK4 family leads to loss of cross-species complementation (Figs 5 and 6), which might create conditions for further development of species-specific regulation of CBB-assembly mechanisms. Finally, the emergence of novel molecules might have allowed adaptation
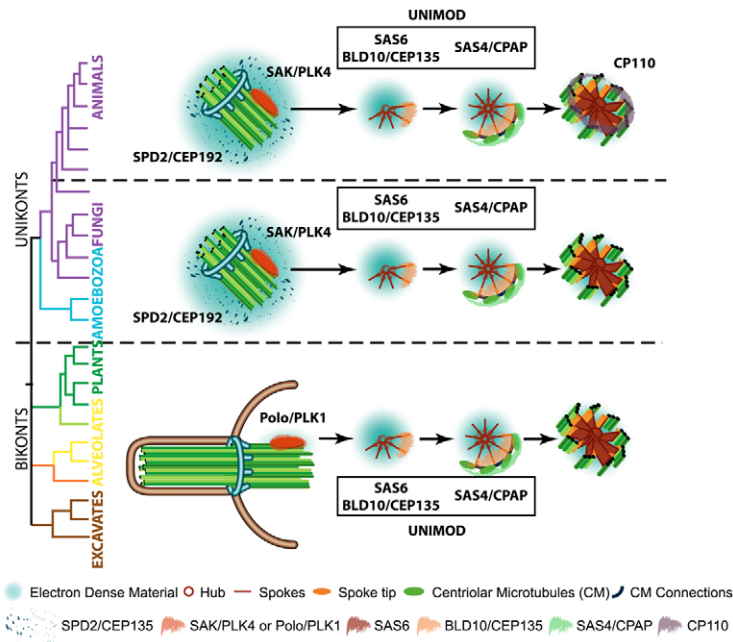
Fig. 8. Stepwise evolution of CBB assembly in eukaryotes. Our work suggests that an ancestral and universal module composed of SAS6, BLD10/CEP135 and SAS4/CPAP was present in the LECA and dictates or enforces CBB ninefold symmetry. We propose that CBB biogenesis is controlled by a trigger: an ancestral Polo/PLK1 kinase in bikonts that has both cell-cycle and CBB biogenesis functions, and its SAK/PLK4 duplicate in unikonts, which is involved only in CBB biogenesis. Note that, because of the lack of completely sequenced genomes of amebozoans containing CBBs, such as *Physarum polycephalum*, we cannot pinpoint whether SAK/PLK4 emergence occurred at the base of opisthokonts (purple clade) or before their divergence from amebozoa. Other proteins, such as SPD2/CEP192 and CP110, emerged in a taxon-specific manner to add new functions or regulatory steps to the conserved CBB structure, as discussed in the text. SPD2/CEP192 emerged in unikonts and its ancestral function is likely to be in PCM recruitment. SPD2/CEP192 might play an important role in CBB formation in contexts in which there is little PCM, such as duplication of the basal body upon fertilization. CP110 might play an important role in coordinating CBB assembly with cilia biogenesis and cytokinesis (see references in main text).

to new contexts of assembly and new functions of the structure. The appearance in unikonts of SPD2/CEP192 (Fig. 2), a molecule whose ancestral function is thought to be in PCM recruitment, might have permitted, in animals, CBB biogenesis in contexts in which there is less PCM, such as duplication of the basal body upon fertilization (Dix and Raff, 2007; Kemp et al., 2004; Pelletier et al., 2004). In animals, CP110 might have coupled the assembly of CBBs to the acquisition of new functions, such as cilia assembly and cytokinesis (Kleylein-Sohn et al., 2007; Spektor et al., 2007; Tsang et al., 2006). Overall, our results strongly support the notion that the molecular machinery that defines the CBB structure is an innovation that emerged in the LECA. This structure evolved through the emergence and divergence of new components that adapted CBB biogenesis and function to the diversification of subcellular contexts and tissue types in which they assemble and function (Fig. 8).

In its evolutionary mechanisms, the CBB machinery is similar to multiprotein complexes and protein-trafficking pathways (Dacks and Field, 2007). In the former, a conserved core that presumably defines the basic function of the complex (Gavin et al., 2006; Snel and Huynen, 2004) can acquire tissue- and organism-specific functions by duplication and specialization of specific components (Pereira-Leal and Teichmann, 2005), as well as recruitment of novel interactions. Our observation of heterogeneous phylogenetic distributions (Fig. 2) revealed extensive species-specific adaptations, which suggests that we have uncovered an approach to identify novel CBB biogenesis players and functions using phylogenetic profiling. We show, for example, that both CP110 and CEP97, which are biochemical partners, appeared in animals (Fig. 2; supplementary material Fig. S7). Our study reveals that it is possible to extend the predictive power of evolutionary-based approaches by considering phylogenetic distributions of genes together with biological structures, and that this will be helpful in predicting both protein functions and interactions. In the future, it will be important to increase the repertoire of species whose genome is sequenced and to thoroughly describe the morphology and function of their CBBs.

We were surprised to observe species in which CBBs have not been described, but whose genomes contain SAS6 and SAS4: the algae *Ostreococcus* and the microsporidiae *Encephalitozoon cuniculi* and *Enterocytozoon bienusi* (Fig. 2). The *Ostreococcus* genome also encodes orthologs of axonemal dyneins (Wickstead and Gull, 2007) and other centriolar proteins, such as POC1 (Keller et al., 2009). However, many flagella components are missing from the *Ostreococcus* genome (Merchant et al., 2007). We propose that this organism might have an elusive CBB remnant, with no associated flagella, such as that described in the non-flagellated, non-sequenced green algae *Kirchneriella* (Pickett-Heaps, 1971). The significance of the presence of these proteins, although severely truncated (supplementary material Fig. S2), in the highly reduced genomes of microsporidial intracellular parasites remains to be determined. Further cell biology research in these enigmatic organisms should reveal mechanisms coupling the loss of cellular structures to the evolution of their molecular assembly machinery or alternatively unveil other functions exhibited by these proteins.

Materials and Methods
Sequence analysis
We used the following approaches for the identification and classification of homologous proteins. (1) We searched for putative orthologs using BLASTP and iterative BLASTP in non-redundant protein databases (Altschul et al., 1990; Altschul et al., 1997; Schaffer et al., 2001) using the full human sequence of each family in eukaryotic species with complete, draft assembly or ongoing genome sequencing (supplementary material Table S2). We considered proteins to be orthologs as reciprocal best hits in BLASTP to the full human sequence (Overbeek et al., 1999). Top-scoring hits were further characterized and specific conserved regions were mapped for each family in multiple sequence alignments (Fig. 3). (2) To further query genome databases, we used regions of high conservation, either previously defined by others or identified in this study, in multiple sequence alignments of the bona fide members of each family. (3) We further investigated the negative results by querying the databases using family members of closely related species or using profile HMMs created with bona fide members of the family or specific conserved regions (using HMMER 2.3.2) (Eddy, 1998). (4) We used TBLASTN (Altschul et al., 1997) whenever sequences were too divergent or much shorter than other members of the family to search for the full protein sequence. (5) We further considered as orthologs those sequences that, although not obeying the first criterion for orthology (see above), were bidirectional best hits to members of the family in closely related species or to the most conserved regions in the human sequence (shown in Fig. 2 as grey boxes). (6) When possible, our orthology assignments were aided by phylogenetic analysis.

Correlation Molecule:CBB was calculated using the formula: 100×[number of species showing correlation (*p*)/total number of species], where *p* is the total of species containing both CBB and the molecule, and species missing both CBB and the molecule. Only sequenced species and species for which ultrastructure information exists were considered in this correlation (supplementary material Tables S1 and S2). Putative homologs that do not strictly satisfy our orthology criteria (grey squares in Fig. 2) were considered as negative hits. Multiple sequence alignments were performed using Muscle 3.6 with the default settings (Edgar, 2004a; Edgar, 2004b). The alignments were represented using Jalview v2.3 (Waterhouse et al., 2009) with the BLOSUM62 color settings. The species used in the alignments are underlined in supplementary material Table S2. Organism-specific sequences larger than five residues were removed from the alignment and are highlighted in supplementary material Fig. S5. Protein conservation values (Fig. 3) were obtained from these alignments using Jalview v2.3 (Waterhouse et al., 2009) – each residue of the alignment is classified from 0 to 11 according to the percentage of aligned residues (these values are shown as a percentage). This information was shown graphically for each subset of protein orthologs. Regions in the alignment with more than 25% gaps are not scored and hence not included. HMMs were built using HMMer (http://hmmer.wustl.edu/) (Eddy, 1998) and these models were used to query specific genomes. A hit was considered significant if the e-value was lower than 0.1 and the bit-score was positive. We used this strategy for BLD10/CEP135, SPD2/CEP192, CP110 and the cryptic polo-box domain of known SAK/PLK4 orthologs, but still we were unable to find further orthologs. Phylogenies were inferred using: (i) neighbor joining (Saitou and Nei, 1987) as implemented in ClustalW 2.0 (Thompson et al., 1994) (1000 bootstraps); (ii) maximum likelihood (Felsenstein, 1981) in the Phylip 3.5 package (ProML and Bootstrap) (Jones-Taylor-Thornton matrix and 100 bootstraps) (J. Felsenstein, PHYLIP: phylogenetic inference package. PhD Thesis, University of Washington, 1993; Larkin et al., 2007); and (iii) the Bayesian method implemented in MrBayes v.3.1.2 (with Blosum62, fixed amino acid rate mode and the program running until the error standard deviation was lower than 0.01). Trees were drawn using FigTree v.1.0 (http://tree.bio.ed.ac.uk/software/).

The coiled-coil prediction was performed using Marcoil1.0 (Delorenzi and Speed, 2002) with 50% threshold in supplementary material Figs S2 and S4. For the representation of the architecture of each human protein, we used the probability per residue and represented it graphically.

The accession numbers of the proteins used in this study are available from our web site at http://www.evocell.org.

Fly stocks

Two *DmBLD10* mutant alleles, *PBac{PB}CG17081c04199* (Thibault et al., 2004) and *Df(3L)Brd15 [71A1-72C2]* (Galewsky and Schulz, 1992) (Bloomington Stock Center), were employed in this study. We confirmed the mapping of *c04199* by inverse PCR (data not shown). All analysis was done on hemizygous flies and thus we refer to those flies as *DmBLD10* mutants throughout the text. Transgenic flies were originated by injection of the plasmid construct (http://www.thebestgene.com). GFP-PACT (Martinez-Campos et al., 2004) flies were kindly provided by Jordan Raff (Gurdon Institute, Cambridge, UK). W1118 stocks were used as wild type. All flies were reared according to standard procedures.

Constructs

All the vectors used in this study were constructed using the Gateway system (Invitrogen). *Drosophila* SAK/PLK4 entry vector has been described elsewhere (Bettencourt-Dias et al., 2005). Human SAK/PLK4 was amplified from IMAGE clone 5273226 and cloned into pDONR221 vector. ZYG1 entry vector was kindly provided by Kevin O'Connell (NIDDK, National Institutes of Health, Bethesda, USA). *Drosophila* and human SAK/PLK4 and ZYG1 coding sequences were then recombined into the destination vectors pcDNA-pDEST53 (Invitrogen) and pAMW. *DmBLD10* cDNA (LD35990) was purchased from the DGC gold BDGP collection (Berkeley, USA) (Stapleton et al., 2002) and cloned into pDONR221 vector. The integrity of the sequence was confirmed by sequencing prior to recombination into destination vectors pMT N-terminal GFP (kindly provided by João Rocha, University of Cambridge, UK) for expression in Dmel cells and pUbq-GFP for expression in flies (kindly provided by Renata Basto, Institut Curie, France). *Drosophila* CP110 was cloned from genomic DNA into pDONR221 vector. The integrity of the sequence was confirmed by sequencing prior to recombination into destination vector pMT N-terminal GFP.

Transfection of constructs, RNAi and treatment of cells

RNAi and transfections of *Drosophila* cell lines were performed as previously described (Bettencourt-Dias et al., 2005).

U2OS cells, kindly provided by Pierre Gonczy (ISREC, Switzerland), were maintained in DMEM (Advanced-DMEM; Gibco) supplemented with 10% FCS (Gibco), 1×L-glutamine-penicillin-streptomycin (Gibco), according to standard tissue-culture techniques. 1×10^5 cells were seeded per well. 700 ng of vector DNA was combined with 100 μl Opti-MEM (Gibco) and 0.5 μl Plus Reagent (Invitrogen), and incubated at room temperature 5 minutes before addition of 1.25 μl Lipofectamine (Invitrogen). Cells were transfected for 6-8 hours, after which the medium was replaced by 1 ml of antibiotic-free complete media. These cells were further incubated for 36 hours to allow protein expression prior to fixation.

Western blotting and reverse transcriptase (RT)-PCR

Standard procedures were used for western blotting. Extracts of U2OS cells were prepared, resuspending the cells in 150 μl lysis buffer (50 mM HEPES pH 8, 200 mM NaCl, 5 mM EDTA, 1% NP-40 and protease inhibitors); all procedures were carried out on ice. Protein concentration was quantified using the Bradford reagent (BioRad) and the same amount of protein applied in the gel.

Total mRNAs were extracted from cells using the RNeasy mini kit (QUIAGEN) and RNase-free DNase set kit (QUIAGEN), according to the manufacturer's instructions. cDNA synthesis was carried out using the Transcriptor First Strand cDNA Synthesis kit (ROCHE). PCR of the gene of interest was carried out using the same primers used for dsRNA synthesis. Amplification products of *eIF4a* cDNA were used as loading control.

Immunostaining and imaging

U2OS cells were fixed for 3 minutes in dry ice-cold methanol, permeabilized and washed in PBSTB (PBS containing 0.1% Triton X-100 and 1% BSA), and stained for polyglutamylated tubulin. Dmel cells were plated on glass coverslips and fixed 1 hour later in 4% formaldehyde in PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 4 mM MgCl$_2$). Cells were permeabilized and washed in PBSTB, and stained for *Drosophila* pericentrin-like protein (D-PLP). DNA was stained with DAPI Vectashield mounting medium (H-1200, Vector Laboratories). Cell imaging and counting were performed on a Leica DMRA2 microscope equipped with a Photometrics Cool SNAP HQ camera. All figure panels were prepared for publication using Adobe Photoshop (Adobe Systems).

Testes from pharate adults were dissected in 183 mM KCl, 47 mM NaCl, 1 mM EDTA and 10 mM Tris-HCl (pH 6.8), transferred to poly-L-lysine glass slides (Sigma) and frozen in liquid nitrogen as previously described (Cenci et al., 1994). Fixation was done for 8 minutes in dry ice-cold methanol followed by 10 minutes in acetone. DNA was stained with TOTO-3-iodide. Testes were mounted using Vectashield mounting media for fluorescence (Vector Laboratories). Testes were imaged as a Z-series (0.5 μm apart) on a Leica SP5 high-speed and high-resolution spectra confocal microscope. Images are presented as maximum-intensity projections. For phase contrast microscopy analysis, testes were dissected in 0.7% NaCl solution and analyzed on an Olympus IMT-2 inverted microscope equipped with a Leica DC 200.

Transmission electron microscopy analysis of testes

Testes from 3- to 5-day-old adults were dissected and fixed in 2.5% glutaraldehyde in PBS (pH 7.2) for 2 hours at 4°C. Testes were post-fixed in OsO$_4$ 1% for 1 hour and treated with 1% uranyl acetate for 30 minutes. Samples were then dehydrated in a graded series of alcohols (70%, 90% for 10 minutes each and three times in 100% for 10 minutes). Testes were incubated in propylene oxide three times for 10 minutes, followed by 1:1 propylene-oxide and resin twice for 15 minutes (Glauert, 1984). Samples were embedded and solidified for 16-48 hours at 60°C. Thin sections (60-80 nm) were cut in a Leica Reichert Ultracut S ultramicrotome, collected on copper grids, and stained with uranyl acetate and lead citrate (Hayat, 1989). Samples were examined and photographed at 80 kV using either a Philips CM10 or a Morgagni 268 transmission electron microscope.

Antibodies

Mouse GT335 anti-polyglutamylated tubulin antibody was kindly provided by Carsten Janke (CNRS, France). The origin of the other antibodies was as follows: chicken anti-D-PLP (Rodrigues-Martins et al., 2007b); rat anti-tubulin-YL1/2 (Oxford Biosciences, USA; 1:50); mouse anti-myc (Santa Cruz Biotechnologies; 1:500); mouse anti-GFP (Roche; 1:50); rabbit anti-actin (Sigma; 1:5000). Secondary antibodies were purchased from Jackson Immunoresearch Laboratories, USA, and used at 1:100 for immunostaining and 1:10,000 for western blot. The DmBLD10 antibody was generated in chicken against the peptide C-LADDRYNQARTREVS (residues 1037-1051) by Pacific Immunology Corp (Ramona, California).

Primers used for dsRNA synthesis, RT-PCR and cloning

Primers used to synthesize GFP dsRNA and for RT-PCR: forward, TAATACG-ACTCACTATAGGGGAGACTTCAGCCGCTACCCC, reverse, TAATACGACTCA-CTATAGGGAGATGTCGGGCAGCACG; to synthesize *Drosophila* SAK/PLK4 (CG7186) dsRNA: forward, TAATACGACTCACTATAGGGGAGAATACGGG-AGGAATTTAAGCAAGTC, reverse, TAATACGACTCACTATAGGGAGATTA-TAACGCGTCGGAAGCAGTCT; to synthesize DmBLD10 (CG17081) dsRNA and for RT-PCR: forward, TAATACGACTCACTATAGGGGAGAACCACCCACACG-ACCAAA, reverse, TAATACGACTCACTATAGGGGAGAGATCCTTTCCCTT-CTTCTT; to synthesize *Drosophila* CP110 (CG14617) dsRNA and for RT-PCR: forward, TAATACGACTCACTATAGGGGAGAAAGAAGCGCGAGGTGCAGCT, reverse, TAATACGACTCACTATAGGGGAGAATGCGATTATGCCGCCTTGG; as control for RT-PCR, eIF4a: forward, TAATACGACTCACTATAGGGAGAGAA-ATGAGATACCTCAGGATGGCCC, reverse, TAATACGACTCACTATAGGG-AGAACGTTAGTGCCGCCAATGCA; for DmBLD10 cloning: forward, GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAATATCAACGATGGTG-ACTTT, reverse, GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAAGAG-TCTTCGATGGCACCCG; for *Drosophila* CP110 cloning: forward, GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATGCGACGTGGTGA-

Supplementary material available online at
http://jcs.biologists.org/cgi/content/full/123/9/1414/DC1

References

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

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Archambault, V. and Glover, D. M. (2009). Polo-like kinases: conservation and divergence in their functions and regulation. *Nat. Rev. Mol. Cell Biol.* **10**, 265-275.

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.

Azimzadeh, J. and Bornens, M. (2004). The centrosome in evolution. In *Centrosomes in Development and Disease* (ed. E. A. Nigg), pp. 93-122. Weinheim: Wiley-VCH.

Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science* **300**, 1703-1706.

Baron, D. M., Ralston, K. S., Kabututu, Z. P. and Hill, K. L. (2007). Functional genomics in Trypanosoma brucei identifies evolutionarily conserved components of motile flagella. *J. Cell Sci.* **120**, 478-491.

Behnam, B., Modarressi, M. H., Conti, V., Taylor, K. E., Puliti, A. and Wolfe, J. (2006). Expression of Tsga10 sperm tail protein in embryogenesis and neural development: from cilium to cell division. *Biochem. Biophys. Res. Commun.* **344**, 1102-1110.

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

Bettencourt-Dias, M. and Glover, D. M. (2007). Centrosome biogenesis and function: centrosomics brings new understanding. *Nat. Rev. Mol. Cell Biol.* **8**, 451-463.

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. *Curr. Biol.* **15**, 2199-2207.

Blachon, S., Cai, X., Roberts, K. A., Yang, K., Polyanovsky, A., Church, A. and Avidor-Reiss, T. (2009). A proximal centriole-like structure is present in Drosophila spermatids and can serve as a model to study centriole duplication. *Genetics* **182**, 133-144.

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

Cebra-Thomas, J. A., Decker, C. L., Snyder, L. C., Pilder, S. H. and Silver, L. M. (1991). Allele- and haploid-specific product generated by alternative splicing from a mouse t complex responder locus candidate. *Nature* **349**, 239-241.

Cenci, G., Bonaccorsi, S., Pisano, C., Verni, F. and Gatti, M. (1994). Chromatin and microtubule organization during premeiotic, meiotic and early postmeiotic stages of Drosophila melanogaster spermatogenesis. *J. Cell Sci.* **107**, 3521-3534.

Chen, Z., Indjeian, V. B., McManus, M., Wang, L. and Dynlacht, B. D. (2002). CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Dev. Cell* **3**, 339-350.

Cunha-Ferreira, I., Bento, I. and Bettencourt-Dias, M. (2009a). From zero to many: control of centriole number in development and disease. *Traffic* **10**, 482-498.

Cunha-Ferreira, I., Rodrigues-Martins, A., Bento, I., Riparbelli, M., Zhang, W., Laue, E., Callaini, G., Glover, D. M. and Bettencourt-Dias, M. (2009b). The SCF/Slimb ubiquitin ligase limits centrosome amplification through degradation of SAK/PLK4. *Curr. Biol.* **19**, 43-49.

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

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

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. *J. Cell Biol.* **180**, 771-785.

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

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

Delorenzi, M. and Speed, T. (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617-625.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A. and Rout, M. P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380.

Dix, C. I. and Raff, J. W. (2007). Drosophila Spd-2 recruits PCM to the sperm centriole, but is dispensable for centriole duplication. *Curr. Biol.* **17**, 1759-1764.

Dobbelaere, J., Josue, F., Suijkerbuijk, S., Baum, B., Tapon, N. and Raff, J. (2008). A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in Drosophila. *PLoS Biol.* **6**, e224.

Dolan, M. F., Melnitsky, H., Margulis, L. and Kolnicki, R. (2002). Motility proteins and the origin of the nucleus. *Anat. Rec.* **268**, 290-301.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755-763.

Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113.

Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368-376.

Gabaldon, T., Snel, B., van Zimmeren, F., Hemrika, W., Tabak, H. and Huynen, M. A. (2006). Origin and evolution of the peroxisomal proteome. *Biol. Direct* **1**, 8.

Galewsky, S. and Schulz, R. A. (1992). Drop out: a third chromosome maternal-effect locus required for formation of the Drosophila cellular blastoderm. *Mol. Reprod. Dev.* **32**, 331-338.

Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B. et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-636.

Giansanti, M. G., Bucciarelli, E., Bonaccorsi, S. and Gatti, M. (2008). Drosophila SPD-2 is an essential centriole component required for PCM recruitment and astral-microtubule nucleation. *Curr. Biol.* **18**, 303-309.

Glauert, A. (1984). *Fixation, Dehydration and Embedding of Biological Specimens.* North-Holland, Amsterdam: Elsevier Science.

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

Hammarton, T. C., Kramer, S., Tetley, L., Boshart, M. and Mottram, J. C. (2007). Trypanosoma brucei Polo-like kinase is essential for basal body duplication, kDNA segregation and cytokinesis. *Mol. Microbiol.* **65**, 1229-1248.

Hayat, M. (1989). *Principles and Techniques of Electron Microscopy: Biological Applications.* Basingtoke: Macmillan Press Ltd.

Hedges, S. B. (2002). The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838-849.

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

Holland, A. J., Lan, W., Niessen, S., Hoover, H. and Cleveland, D. W. (2010). Polo-like kinase 4 kinase activity limits centrosome overduplication by autoregulating its own stability. *J. Cell Biol.* **188**, 191-198.

Hung, L. Y., Tang, C. J. and Tang, T. K. (2000). Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the gamma-tubulin complex. *Mol. Cell. Biol.* **20**, 7813-7825.

Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. and Manning, G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biol.* **5**, e17.

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

Kelley, L., Maccallum, R. and Sternberg, M. (1999). Recognition of remote protein homologies using three-dimensional information to generate a position specific scoring matrix in the program 3D-PSSM. In *Third Annual Conference on Computational Molecular Biology*, pp. 218-225. The Association for Computing Machinery, New York.

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. *Dev. Cell* **6**, 511-523.

Kitagawa, D., Busso, C., Fluckiger, I. and Gonczy, P. (2009). Phosphorylation of SAS-6 by ZYG-1 is critical for centriole formation in C. elegans embryos. *Dev. Cell* **17**, 900-907.

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

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. *Curr. Biol.* **19**, 1012-1018.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Lee, K. S. and Erikson, R. L. (1997). Plk is a functional homolog of Saccharomyces cerevisiae Cdc5, and elevated Plk activity induces multiple septation structures. *Mol. Cell. Biol.* **17**, 3408-3417.

Lee, M. G. and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature* **327**, 31-35.

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

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

Liu, X. and Erikson, R. L. (2002). Activation of Cdc2/cyclin B and inhibition of centrosome amplification in cells depleted of Plk1 by siRNA. *Proc. Natl. Acad. Sci. USA* **99**, 8672-8676.

Loncarek, J., Hergert, P., Magidson, V. and Khodjakov, A. (2008). Control of daughter centriole formation by the pericentriolar material. *Nat. Cell. Biol.* **10**, 322-328.

Lopez-Bigas, N. and Ouzounis, C. A. (2004). Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res.* **32**, 3108-3114.

Lopez-Bigas, N., De, S. and Teichmann, S. A. (2008). Functional protein divergence in the evolution of Homo sapiens. *Genome Biol.* **9**, R33.

Mans, B. J., Anantharaman, V., Aravind, L. and Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3**, 1612-1637.

Martinez-Campos, M., Basto, R., Baker, J., Kernan, M. and Raff, J. W. (2004). The Drosophila pericentrin-like protein is essential for cilia/flagella function, but appears to be dispensable for mitosis. *J. Cell Biol.* **165**, 673-683.

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

McKean, P. G., Baines, A., Vaughan, S. and Gull, K. (2003). Gamma-tubulin functions in the nucleation of a discrete subset of microtubules in the eukaryotic flagellum. *Curr. Biol.* **13**, 598-602.

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

Modarressi, M. H., Taylor, K. E. and Wolfe, J. (2000). Cloning, characterization, and mapping of the gene encoding the human G protein gamma 2 subunit. *Biochem. Biophys. Res. Commun.* **272**, 610-615.

Modarressi, M. H., Behnam, B., Cheng, M., Taylor, K. E., Wolfe, J. and van der Hoorn, F. A. (2004). Tsga10 encodes a 65-kilodalton protein that is processed to the 27-kilodalton fibrous sheath protein. *Biol. Reprod.* **70**, 608-615.

Mottier-Pavie, V. and Megraw, T. L. (2009). Drosophila bld10 is a centriolar protein that regulates centriole, basal body, and motile cilium assembly. *Mol. Biol. Cell* **20**, 2605-2614.

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

Nigg, E. A. (2007). Centrosome duplication: of rules and licenses. *Trends Cell Biol.* **17**, 215-221.

Ohno, S. (1970). *Evolution by Gene Duplication*. New York: Springer-Verlag.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. and Maltsev, N. (1999). The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. USA* **96**, 2896-2901.

Pearson, C. G., Osborn, D. P., Giddings, T. H., Jr, Beales, P. L. and Winey, M. (2009). Basal body stability and ciliogenesis requires the conserved component Poc1. *J. Cell Biol.* **187**, 905-920.

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

Pelletier, L., 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. *Curr. Biol.* **14**, 863-873.

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

Pereira-Leal, J. B. and Teichmann, S. A. (2005). Novel specificities emerge by stepwise duplication of functional modules. *Genome Res.* **15**, 552-559.

Pickett-Heaps, J. D. (1971). The autonomy of the centriole: fact or fallacy? *Cytobios* **3**, 205-214.

Piel, M., Nordberg, J., Euteneuer, U. and Bornens, M. (2001). Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291**, 1550-1553.

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

Rodrigues-Martins, A., Riparbelli, M., Callaini, G., Glover, D. M. and Bettencourt-Dias, M. (2007b). Revisiting the role of the mother centriole in centriole biogenesis. *Science* **316**, 1046-1050.

Rogers, G. C., Rusan, N. M., Roberts, D. M., Peifer, M. and Rogers, S. L. (2009). The SCF Slimb ubiquitin ligase regulates Plk4/Sak levels to block centriole reduplication. *J. Cell Biol.* **184**, 225-239.

Rose, A., Schraegle, S. J., Stahlberg, E. A. and Meier, I. (2005). Coiled-coil protein composition of 22 proteomes-differences and common themes in subcellular infrastructure and traffic control. *BMC Evol. Biol.* **5**, 66.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425.

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. and Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**, 2994-3005.

Schimenti, J., Cebra-Thomas, J. A., Decker, C. L., Islam, S. D., Pilder, S. H. and Silver, L. M. (1988). A candidate gene family for the mouse t complex responder (Tcr) locus responsible for haploid effects on sperm function. *Cell* **55**, 71-78.

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. *Curr. Biol.* **19**, 1005-1011.

Schulz, I., Erle, A., Graf, R., Kruger, A., Lohmeier, H., Putzler, S., Samereier, M. and Weidenthaler, S. (2009). Identification and cell cycle-dependent localization of nine novel, genuine centrosomal components in Dictyostelium discoideum. *Cell Motil. Cytoskeleton* **66**, 915-928.

Snel, B. and Huynen, M. A. (2004). Quantifying modularity in the evolution of biomolecular systems. *Genome Res.* **14**, 391-397.

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. *Dev. Cell* **15**, 901-912.

Spektor, A., Tsang, W. Y., Khoo, D. and Dynlacht, B. D. (2007). Cep97 and CP110 suppress a cilia assembly program. *Cell* **130**, 678-690.

Stapleton, M., Carlson, J., Brokstein, P., Yu, C., Champe, M., George, R., Guarin, H., Kronmiller, B., Pacleb, J., Park, S. et al. (2002). A Drosophila full-length cDNA resource. *Genome Biol.* **3**, RESEARCH0080.

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

Swallow, C. J., Ko, M. A., Siddiqui, N. U., Hudson, J. W. and Dennis, J. W. (2005). Sak/Plk4 and mitotic fidelity. *Oncogene* **24**, 306-312.

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. *Nat. Cell Biol.* **11**, 825-831.

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L. et al. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat. Genet.* **36**, 283-287.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tsang, W. Y., Spektor, A., Luciano, D. J., Indjeian, V. B., Chen, Z., Salisbury, J. L., Sanchez, I. and Dynlacht, B. D. (2006). CP110 cooperates with two calcium-binding proteins to regulate cytokinesis and genome stability. *Mol. Biol. Cell* **17**, 3423-3434.

Tsou, M. F., Wang, W. J., George, K. A., Uryu, K., Stearns, T. and Jallepalli, P. V. (2009). Polo kinase and separase regulate the mitotic licensing of centriole duplication in human cells. *Dev. Cell* **17**, 344-354.

Warnke, S., Kemmler, S., Hames, R. S., Tsai, H. L., Hoffmann-Rohrer, U., Fry, A. M. and Hoffmann, I. (2004). Polo-like kinase-2 is required for centriole duplication in mammalian cells. *Curr. Biol.* **14**, 1200-1207.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. and Barton, G. J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

White, G. E. and Erickson, H. P. (2006). Sequence divergence of coiled coils-structural rods, myosin filament packing, and the extraordinary conservation of cohesins. *J. Struct. Biol.* **154**, 111-121.

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