Evolutionary history of clathrin-mediated endocytosis and the eisosome

Submitted by Luigi Cibrario

to the University of Exeter as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

August 2011

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

Abstract

Endocytosis is both an ancient and a diverse feature of the eukaryotic cell. Studying how it evolved can provide insight into the nature of the last common eukaryotic ancestor, and the diversification of eukaryotes into the known extant lineages. In this thesis, I present two studies on the evolution of endocytosis. In the first part of the thesis I report results from a large-scale, phylogenetic and comparative genomic study of clathrin-mediated endocytosis (CME). The CME pathway has been studied to a great level of detail in yeast to mammal model organisms. Several protein families have now been identified as part of the complex set of protein-protein and protein-lipid interactions which mediate endocytosis. To investigate how such complexity evolved, first, I defined the modular nature of the CME interactome (CME-I) by literature review, and then I carried out a systematic phylogenetic and protein domain architecture analysis of the proteins involved. These data were used to construct a model of the evolution of the CME-I network, and to map the expansion of the network's complexity to the eukaryotic tree of life. In the second part of the thesis, I present results from evolutionary and functional studies of the eisosome, a protein complex which has been proposed to regulate the spatial distribution of endocytosis in *S. cerevisiae*. The phylogeny of eisosomes components Pil1 and Lsp1 reported here, suggests that eisosomes are likely to have originated at the base of the fungi, and then diversified significantly via multiple gene duplications. I thus studied the localisation and function of Pil1 and Lsp1 homologues in *Magnaporthe oryzae* to investigate the role of eisosomes in filamentous fungi. Results suggests that eisosomes are linked with septal formation and integrity in *M. oryzae*, and that the septal specific Pil2 paralogue was lost in budding yeasts. Together, the data presented in this thesis describe the evolutionary history of a complex biological system, but also highlights the problem of asymmetry in the understanding of endocytic diversity in the eukaryotes.

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Acknowledgements

I am grateful to the BBSRC for funding this project and giving me the privilege to investigate the natural world first hand. I would like to thank Tom Richards for his guidance, his enthusiasm, and for the tireless drive for excellence he has always showed when supervising this project. I also thank Nick Talbot, to whom I am greatly indebted, for his help in developing the project and for the support he has offered at crucial stages of the PhD. Furthermore, I thank Bryony Williams, for her invaluable help in the later stages of the programme, and Martin Egan, for showing me the ropes in the lab and being generally a positive influence.

I was very lucky to be part of such a vibrant research community, so I thank everyone in GP who made the working place so enjoyable.

Among my friends and family, I thank mum and dad and everyone else who was there for me. I must also thank Riccardo, Emma, Theresa, Maria and Tony, for offering such warm hospitality, with little or no notice, and making the bad times not so bad at all, and the good ones even better. Above all I thank Karleigh, for her unconditional love, support and patience and for being just about the sweetest thing I know.

Abbreviations

ºC	degree Celsius
bp	base pair
CIA	24:1 Chloroform:isoamyl alcohol
DNA	deoxyribonucleic acid
cDNA	complementary DNA
DMSO	dimethyl sulfoxide
g	relative centrifugal force
g	gram
h	hour
kb	one thousand base pairs
LB	lysogeny broth
mRNA	messenger RNA
µM	micromolar
µg	microgram
µl	microlitre
µm	micrometre
mg	milligram
ml	millilitre
mm	millimetre
mM	millimolar
mins	minutes
M	molar
nt	nucleotide
PCR	polymerase chain reaction
pmol	picomole
RNA	ribonucleic acid
rpm	repetitions per minute
SOC	super optimal broth with catebolite repression
w/v	weight/volume
v/v	volume/volume

1 General Introduction

1.1 Endocytosis is a hallmark of the eukaryotic cell

Endocytosis is the process which allows cells to engulf and internalise the external particles and molecules which are either too large or cannot pass through the hydrophobic plasma membrane because of their polarity. This process was first reported by Elie Metchnikoff in 1883, when he demonstrated that specialised mobile cells in starfish larvae engaged in an immune response by swallowing infecting agents (Metchnikoff, 1883). The process was termed phagocytosis (from Greek word *phago* meaning 'to eat'). The internalisation of fluid medium via the invagination of the plasma membrane was first reported in 1931 (Lewis, 1931) with the introduction of the term 'pinocytosis' (from Greek word pino meaning 'drinking'). Following these early studies, endocytosis has emerged as an important cellular process, involved in a wide breadth of vital functions. These include the immune response (Stuart & Ezekowitz, 2008), nutrient uptake (Robibaro *et al*., 2001), cell signalling (Polo & Di Fiore, 2006), cell growth (Higuchi *et al*., 2009), cell differentiation (Romih & Jezernik, 1994), synaptic activity (Granseth *et al*., 2006), maintaining cell homeostasis (Covian-Nares *et al*., 2008) and recycling plasma membrane and its components (Schneider *et al*., 1979).

While much of the research to date has focused on metazoan and yeast model organisms, there are some studies of endocytosis which cover the wider diversity of the eukaryotes, for instance filamentous fungi (Fuchs & Steinberg, 2005), Viridiplantae (Irani & Russinova, 2009; Raven *et al*., 2009), and protists lineages such as alveolates (e.g. *Paramecium*) (Allen & Fok, 1980), kinetoplastids (e.g. *Trypanosoma*) (Gabernet-

Castello *et al*., 2009) and diplomonads (e.g. *Giardia*) (Hernandez *et al*., 2007). Because of the range of cellular functions mediated by endocytosis and the diversity of taxa it has been studied within, endocytosis is considered an important hallmark of eukaryotic cells (Field *et al*., 2006). In contrast, only recently have cellular uptake functions similar to endocytosis been identified in a prokaryotic organism, as experiments on the planctomycete *Gemmata obscuriglobus*, which has a compartmentalised cell plan and coat-like proteins (Santarella-Mellwig *et al*., 2010), demonstrate energy-dependent internalisation of green fluorescent protein (GFP) (Lonhienne *et al*., 2010). The evolutionary history of endocytosis may therefore provide insight into the transition between prokaryote and eukaryote life forms. Here I discuss how endocytosis relates to the origin of eukaryotes (eukaryogenesis) in light of the leading hypotheses regarding the origins of eukaryotes and the diversity and types of endocytosis identified in eukaryotic cells.

1.1.1 Endocytosis in eukaryogenesis

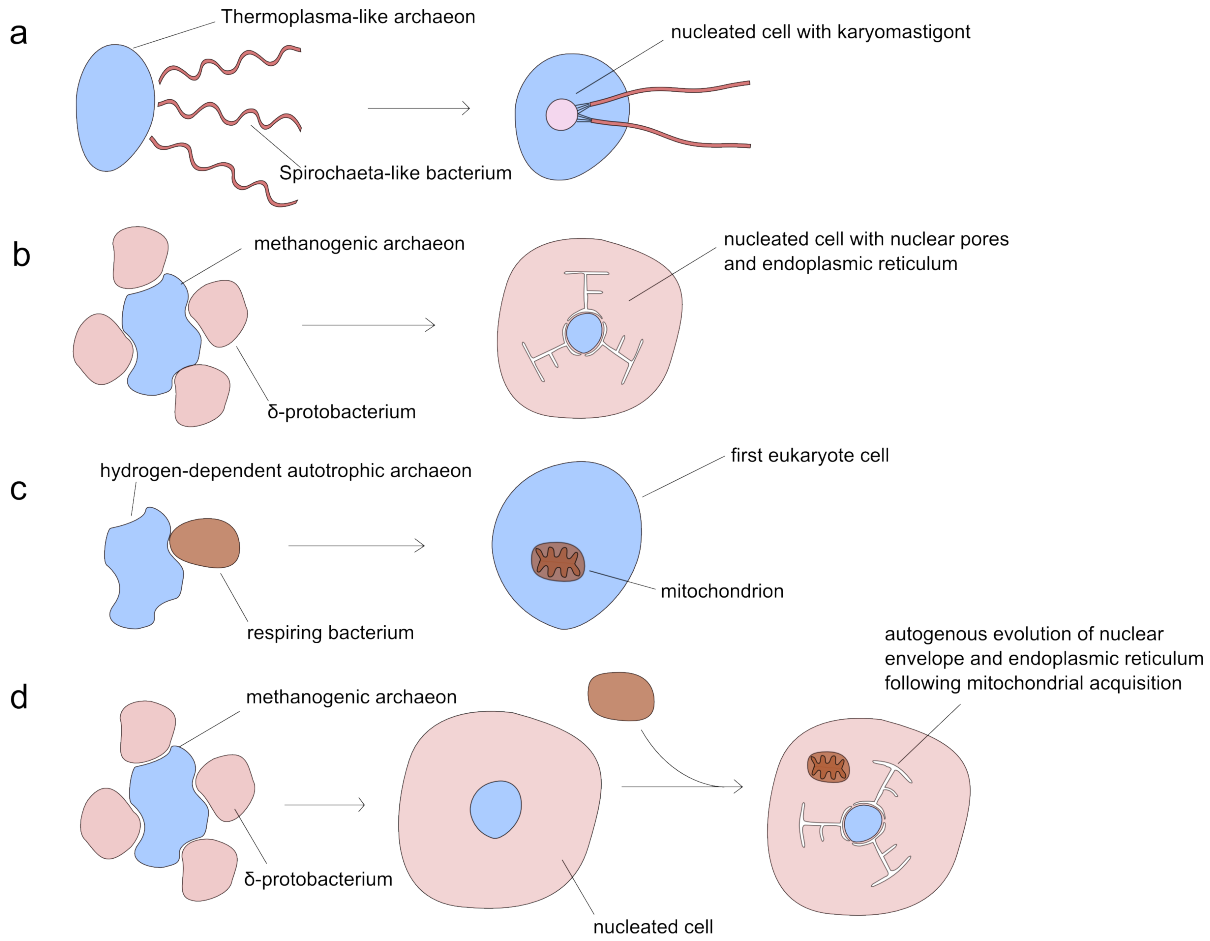
There are diverse hypotheses that account for the origin of eukaryotes, with different implications regarding the importance of endocytosis to this process (Cavalier-Smith, 2009; Gribaldo *et al*., 2010; Martin *et al*., 2001). One class of hypotheses states that eukaryotes originated from a chimeric cell which resulted from metabolic symbiosis (syntrophy) between two prokaryotes (Martin *et al*., 2001) (Figure 1.1 a-d) . For instance, Margulis *et al.* argue that the nucleus evolved from the merging of a Thermoplasma-like archaeon and Spirochaeta-like bacterium (Margulis *et al*., 2000) (Figure 1.1a). The syntrophic relationship conferred a selective advantage because the hypothetical archaeon generated hydrogen sulfide and the bacterium oxidised sulfide to

sulfur (Margulis *et al*., 2000). Another hypothesis argues that eukaryotes originated from syntrophy between a methanogenic archaeon and δ-protobacterium (Moreira & Lopez-Garcia, 1998) (Figure 1.1b). Key to this hypothesis is the transfer of hydrogen in anaerobic environments. The methane produced by the archaeon is metabolised into hydrogen by the bacterium, conferring selective advantage (Moreira & Lopez-Garcia, 1998). The hydrogen hypothesis also proposes that the first eukaryote cell originated from hydrogen-based syntrophy between two prokaryotes (Martin & Muller, 1998). It argues that the first eukaryote cell originated from a hydrogen-dependent and autotrophic archaeon, which engaged in metabolic symbiosis with a bacterium that respired but also produced hydrogen from anaerobic heterotrophic metabolism (Figure 1.1c). Importantly, however, this hypothesis argues that syntrophy led to the origin of the mitochondrion, and not the nucleus, as the defining eukaryotic acquisition (Martin & Muller, 1998).

The strength of syntrophic models of eukaryote origin is that they account for selective advantage of the novel chimeric cell. They also explain the different ancestry of eukaryotic genes with the Archaea and Bacteria, with informational genes (e.g. involved in transcription and translation) more closely related to Archaea and operations genes (e.g. involved in cellular metabolic processes) more closely related to Bacteria (Rivera & Lake, 2004). Yet the resolution among these genes is too weak to identify which prokaryotic lineages were involved in this endosymbiosis (Gribaldo *et al*., 2010).

However, because syntrophy putatively occurred between two prokaryotes, and phagocytosis is currently defined as a eukaryotic feature, it does not explain how one

SYNTROPHIC MODELS OF EUKARYOTE ORIGIN



AUTOGENOUS MODELS OF EUKARYOTE ORIGIN

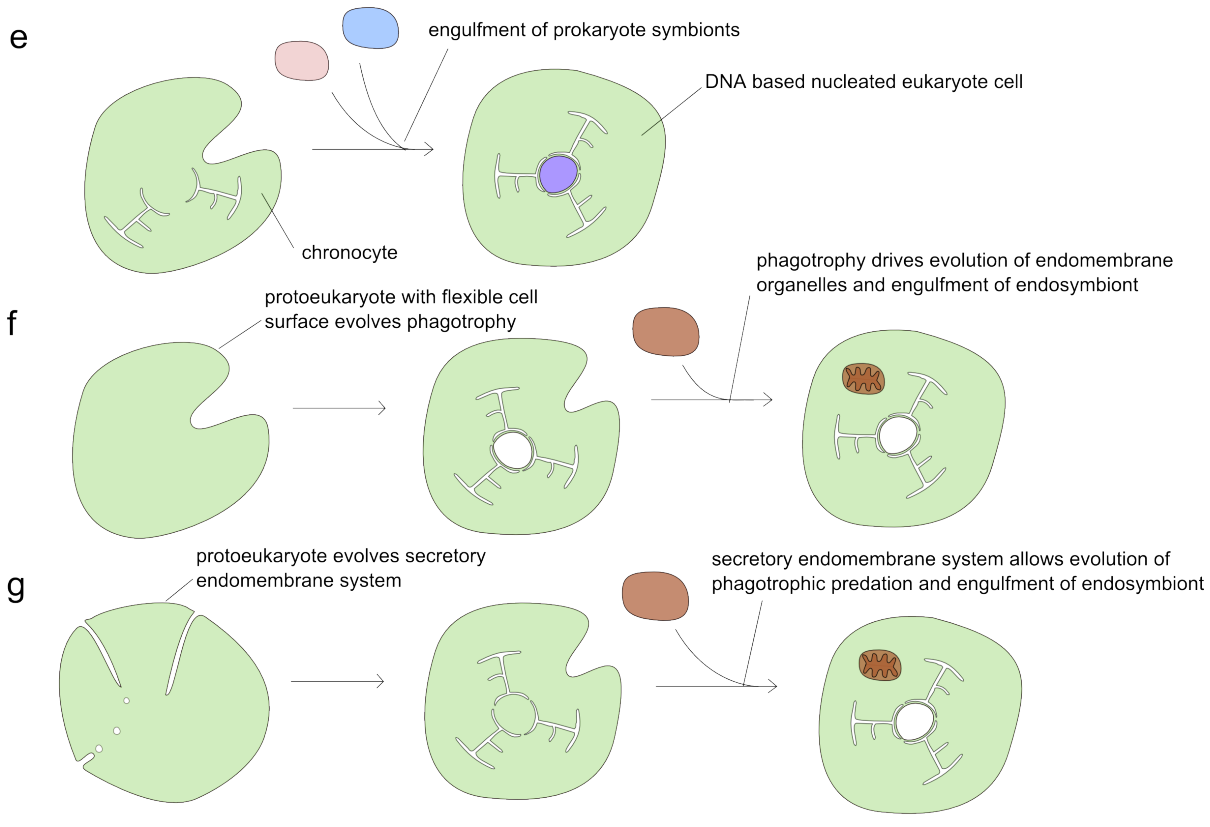


Figure 1.1 Syntrophic and autogenous scenarios of eukarote origin. A schematic summary of a selection of leading eukaryotgenesis hypotheses, devided into those proposing a syntrophic origin (a-d), and those proposing an autogenous origin of the eukaryotic cell (e-g). **a** Thermoplasma-like archaeon and Spirochaeta-like bacteria together formed a protoeukaryote "Thiodendron" stage based on syntrophic exchange of sulfur, from which a nucleated cell with a karyomastigont evolved (Margulis *et al.*, 2000). **b** Methanogenic archaeon and δ-protobacterium form a syntrophic consortium based on interspecies hydrogen transfer. Following bacterial cytoplasmic fusion the endoplasmic reticulum and nuclar pores evolve (Moreira & Lopez-Garcia, 1998). **c** Hydrogen dependent autotrophic archaeon enters syntrophic relationship with respiring bacterium, which is eventually engulfed and evolves into the mitochondrion (Martin & Muller, 1998). **d** Methanogenic archaeon and δ-protobacterium fuse to form nucleated cell. The mitochondrial ancestor then acquired, thus driving the autogenous evolution of the endomembrane system (Lopez-Garcia & Moreira, 2006). **e** A cell type distinct from Archaea and Eubacteria, i.e. the chronocyte, evolves cytoskeleton, endomembranes and phagotrophy, eventually engulfing prokaryotes and evolving a DNA-based nucleus (Hartman & Fedorov, 2002). **f** A protoeukaryote lineage, sister to the Archaea, having acquired a flexible surface coat, autogenously evolves phagotrophy, which in turn drives the evolution of the eukaryotic organelles and engulfment of the endosymbiont mitochondrial ancestor (Cavalier-Smith, 2002). **g** A protoeukaryote lineage evolves an endomembrane secretory system, which allows for the evolution of phagotrophic predation and eventually the engulfment of the mitochondrial endosymbiont (Jekely, 2007).

prokaryote could have engulfed the other. Endosymbiosis between two prokaryotes has, in fact, been observed (von Dohlen *et al.*, 2001) although it appears to be exceedingly rare. Having said that, the hydrogen hypothesis argues that phagocytosis, or any other form of endocytosis are not prerequisites for the host cell to acquire the symbiont. Rather, it argues that because the host would have benefited from sticking tightly to symbionts, it simply grew around the symbiont to prevent its escape (Martin & Muller, 1998).

In contrast to the syntrophic model of eukaryote origin, other hypotheses argue that a

protoeukaryote lineage autogenously evolved eukaryotic features such as the tubulin and actin cytoskeleton, the endomembrane system and, importantly, phagotrophy, which allowed it to engulf the free-living bacterial ancestor of mitochondria (Cavalier-Smith, 2002, 2010; Hartman & Fedorov, 2002; Jekely, 2007; O'Malley, 2010) (Figure 1.1 d-f). One such hypothesis, referred here as the neomuran-phagotrophy hypothesis, argues that phagocytosis was the prerequisite of all other eukaryote features (Cavalier-Smith, 2002). It suggests that the first key eukaryote acquisition was a flexible surface coat which was made possible by replacement of bacterial peptidoglycan cell wall components with N-linked glycoproteins (i.e. 'neomura', meaning new wall). This conferred cell surface flexibility. The first eukaryote could then lose the typically prokaryote exoskeleton and evolve phagocytosis, which would in turn lead to the origin of other eukaryote-specific traits (Figure 1.1 f)(Cavalier-Smith, 2002).

Another hypothesis, based on a study of ras GTPase diversification, argues that a secretory endomembrane system, ancestor of the endoplasmic reticulum, the Golgi apparatus and possibly even the nucleus, evolved prior to phagocytosis (Jekely, 2003, 2007). This model proposes that an endomembrane system, based on the formation of tubules and vesicles, evolved in concordance with a secretory system. This eventually allowed the membrane remodelling around the prey to be coordinated with the secretion of digestive enzymes and nutrient uptake, as necessitated by phagotrophic predation (Figure 1.1 g) (Jekely, 2003). A further hypothesis is based on a study of eukaryotic signature proteins (ESPs) (Hartman & Fedorov, 2002). ESPs are a minimal set of proteins present in diverse eukaryotic groups. The authors argue that this set of proteins means eukaryogenesis entail three cellular domains: Archaea, Bacteria and a distinct

cell named the Chronocyte (Hartman & Fedorov, 2002). The analysis of ESPs suggests that the chronocyte, a RNA-based cell which acted as the host cell in this eukaryogenesis model, had a complex cytoskeleton and an endomembrane system allowing it to engulf prokaryotic symbionts which in turn introduced DNA-based genetic systems and the nucleus (Figure 1.1e) (Hartman & Fedorov, 2002). The strength of eukaryogenesis hypotheses based on autogenous-first models is that they allow the proto-mitochondrial host to perform phagocytosis. It has in fact been argued that these hypotheses do not provide a satisfactory selective advantage for the protoeukaryote to evolve features such as a complex cytoskeleton and an endomembrane system (Martin *et al.*, 2001; Martin & Muller, 1998). However, one of these models proposes a strong selective advantage for cells that evolved phagocytic predation via the prior evolution of a functional secretory endomembrane system (Jekely, 2007). It has also been suggested that a complex cellular organisation and membrane-trafficking system conferred greater potential for increased cell volume, complexity of cellular function, and flexibility of how these functions are performed (Dacks & Field, 2007; Field *et al.*, 2011).

An interesting hypothesis combines elements of syntrophic symbiosis with autogenous evolution. As for the hypothesis by Moreira & Lopez-Garcia, it suggests that the nucleus evolved from syntrophic symbiosis between a methanogenic archaeon and a δ-protobacterium (Moreira & Lopez-Garcia, 1998). However, it also argues that following the acquisition of a respiring endosymbiont, the chimeric cell autogenously evolved a nuclear envelope and the endoplasmic reticulum from bacterial membranes (Figure 1.1d). The hypothesis specifies two selective forces, applied sequentially: first, the metabolic compartmentalisation to avoid deleterious interference of opposite anabolic

and catabolic pathways, and second the prevention of deleterious spreading of aberrant proteins in cytoplasm as introns become widespread in the nuclear genome (Lopez-Garcia & Moreira, 2006). This theoretical elaboration attempts to explain how some eukaryotic organelles have exogenous origins while others evolved autogenously, but it should be noted that it fundamentally belongs to the syntrophic model of eukaryogenesis, because it is still proposing that the primary event in the origin of eukaryotes is the syntrophic merge of two prokaryotes (Figure 1.1 d) (Lopez-Garcia & Moreira, 2006). Indeed, it is widely accepted that organelles such as mitochondria, plastids and respective derivations have exogenous origins but other eukaryotic organelles were autogenously acquired (Martin, 1999), so the real question is what is the primary event, the acquisition of exogenous organelles or of autogenous organelles?

According to autogenous-first models for eukaryogenesis, endocytosis and specifically phagocytosis were key acquisitions of the protoeukaryote cell, which predated the acquisition of mitochondria and even the nucleus and other organelles (Cavalier-Smith, 2002; Yutin *et al.*, 2009). Moreover, according to these models, the evolution of endocytosis was strongly linked with the evolution of the eukaryotic cytoskeleton and the endomembrane system (Cavalier-Smith, 2009; Jekely, 2007).

By contrast, syntrophic (or exogenous-first) models entail endocytosis-independent processes for the primary endosymbiotic event which led to origin of eukaryotes. As mentioned above, one such process has been described in the context of the hydrogen hypothesis (Martin & Muller, 1998). Also, the syntrophic hypothesis by Moreira posits that within colonial congregations of prokaryote cells engaged in metabolic symbiosis,

cells merged to form the chimeric eukaryote (Moreira & Lopez-Garcia, 1998).

It is difficult to test hypotheses of how the first eukaryote arose because no extant intermediate lineage has been sampled. The archezoan hypothesis first suggested that amitochondriate lineages such as *Giardia, Trichomonas, Entamoeba* and Microsporidia derived from primitive anaerobic eukaryotes which had not undergone mitochondrial endosymbiosis (Cavalier-Smith, 1983a; Cavalier-Smith, 1983b), thereby providing examples of a pre-mitochondrial phase of eukaryote evolution. Initially, phylogenies based on ribosomal RNA encoding gene were consistent with the hypothesis (Leipe *et al*., 1993; Sogin *et al*., 1989; Vossbrinck *et al*., 1987), but further studies revealed that 'archezoans' in fact harboured organelles which share common ancestry with mitochondria (Embley *et al*., 2003; Tovar *et al*., 1999; van der Giezen & Tovar, 2005; Williams *et al*., 2002). The early branching positions of archezoan candidates, were thus dismissed as an artifact due to long branch attraction (Embley & Hirt, 1998; Philippe *et al*., 2000). No other eukaryotic lineage derived from a pre-mitochondrial phase of cell evolution has been identified (Embley, 2006). Arguably, before the nature of the first eukaryote common ancestor can be defined we need to characterise the last common ancestor of known extant eukaryotic lineages.

1.1.2 Endocytosis in the last common eukaryotic ancestor

Evidence from recent phylogenomic studies suggests that the last common eukaryotic ancestor (LCEA) encoded several protein families responsible for a range of key eukaryotic cellular features. These include meiosis (Ramesh *et al.*, 2005), centrioles (Hodges *et al.*, 2010), a diversified DNA replisome (Liu *et al.*, 2009), intron-dominated

genomes (Koonin, 2009), and diversified actin and tubulin associated motor proteins (Richards & Cavalier-Smith, 2005; Wickstead & Gull, 2007; Wickstead *et al.*, 2010).

Importantly, data shows the LCEA also possessed a complex endomembrane system where specialisation and compartmentalisation evolved via gene family expansions of multiple paralogues (Dacks & Field, 2007; Dacks *et al.*, 2009; Dacks *et al.*, 2008). Phylogenomic studies have identified some of the gene families involved in the ancient endomembrane system, for which gene duplication coupled with specialisation was confirmed (Field *et al.*, 2006; Pereira-Leal & Seabra, 2001; Yoshizawa *et al.*, 2006). A link between the evolution of the endomembrane system and endocytosis is suggested by paralogues of endocytic adaptins and epsin-related proteins, localising at different sites of the membrane trafficking system such as endosomes, Golgi apparatus and endoplasmic reticulum (Dacks *et al.*, 2008; Gabernet-Castello *et al.*, 2009). This predicts the LCEA also possessed a complex and diversified endocytic system. Data showing that the endocytic proteins clathrin and the AP2 complex were present in the LCEA is consistent with the prediction (Field *et al.*, 2006). However, there is likely to be further complexity and diversity in the endocytic systems of the putative LCEA, especially in light of the mechanistic diversity of endocytic pathways studied thus far in eukaryotes (see Section 1.3) (Conner & Schmid, 2003; Kumari *et al.*, 2010). Clathrin and the AP2 complex putatively formed vesicle coats in the endomembrane system of the LCEA, but what about the rest of the clathrin-mediated endocytosis pathway? And what about other endocytic pathways in the LCEA? Did for example the LCEA evolve a range of specialised endocytic machineries, and what is the range of molecular mechanisms it employed for endocytic functions such as membrane deformation,

vesicle scission, vesicle transport and vesicle coat recycling? It is important to investigate endocytosis in the LCEA and provide further insight into nature of the ancestral cell.

1.1.3 Evolutionary history of endocytosis

Given the putative presence of endocytosis in the LCEA and its role in eukaryotic evolution it is important to investigate its evolutionary history. However, phagocytosis and pinocytosis are distinct processes, and within these broad categories there is further functional specification (Conner & Schmid, 2003). The evolutionary relationships among different endocytic pathways in not known. The approach I propose is thus to map pathway specific proteins and functions to eukaryotic diversity. This can be summarised in three steps:

1. Outline a consensus eukaryotic phylogeny from existing studies, highlighting major eukaryotic groups and alternative rooting hypotheses.

2. Identify and characterise endocytic genes involved in distinct endocytic pathways from literature review. This step involves a critical evaluation of the cell and molecular biology data which support the functional assignment prediction of candidate pathway components.

3. Predict the occurrence of distinct endocytic pathways across eukaryotic diversity by tracing the evolutionary distribution of endocytic genes specific to a given pathway.

1.2 Eukaryote diversification

Eukaryogenesis constitutes a major evolutionary leap which brought radical changes to the structure and function of the cell (Cavalier-Smith, 2006). In no specific order of importance or ancestry, these changes include a flexible cell surface and phagotrophy (Cavalier-Smith, 2009), a novel internal organisation regulated by a cytoskeleton and complement motor proteins (Cavalier-Smith, 1975; Richards & Cavalier-Smith, 2005; Wickstead *et al*., 2010) an endomembrane system of vesicle trafficking (Dacks & Field, 2007), compartmentalisation of informational and operational activity by a nuclear envelope (Koonin & Aravind, 2009; Neumann *et al*., 2010), a new system of duplication and sexual reproduction via mitosis and meiosis (Cavalier-Smith, 2010; Ramesh *et al*., 2005), a complex genomic structure featuring spliceosomal introns (Koonin, 2006, 2009), and a novel metabolic make-up characterised by compartmentalised biochemistry (i.e. mitochondria, peroxisomes, and in some eukaryotes, plastids) (Lane & Martin, 2010).

The effort to resolve the order of the cellular modifications and the early branching order of the eukaryotic tree has been hampered by the inability to identify an extant *bona fide* intermediate protoeukaryote lineage where only part of these changes have occurred (Embley, 2006). The eukaryote tree topology is therefore a matter of contention, as the early branching order has not been resolved and the root of tree has not been pinpointed with any degree of confidence (Roger & Simpson, 2009). The current consensus eukaryotic tree is represented by a polytomy of the major monophyletic eukaryote groups which have been recognised and supported by

molecular phylogenies (Baldauf, 2008). It has been suggested that this model reflects an evolutionary 'Big Bang' (i.e. a hard polytomy) whereby rapid eukaryote diversification occurred perhaps following mitochondrial endosymbiosis (Philippe & Adoutte, 1998; Philippe *et al*., 2000). However, the under-representation of protist taxa in broad eukaryote phylogenies (Dacks *et al*., 2002), the loss of evolutionary signal by the saturation of character change (Ho & Jermiin, 2004), large number of secondary endosymbiotic events and horizontal gene transfer events (Archibald, 2009; Archibald & Richards, 2010) and limitations in phylogenetic methods (Penny *et al*., 2001) may all contribute to the lack of resolution in deep parts of the eukaryotic tree.

Here, I briefly review the main eukaryotic groups which form the consensus tree used in this thesis as a phylogenetic framework. I also discuss evolutionary relationships between major groups and alternative rootings of the eukaryotic tree.

1.2.1 Major eukaryote super-groups

The traditional five kingdom structure divided the diversity of life into three levels of organisation: prokaryotic (kingdom Monera), unicellular eukaryotic (kingdom Protista), and multicellular eukaryotic (the Plantae, the Fungi and the Animalia kingdoms) (Whittaker, 1969). This was the first system to formally recognise Fungi as a kingdom. The organisation of the eukaryotes into protista and the three multicellular kingdoms was popularly used until the 21st century. However, with the increased availability of molecular data relative to diverse protists and the improvement of phylogenetic methods, a new organisation of eukaryotes was proposed (Simpson & Roger, 2004). The revised system improved biological realism by dividing eukaryotic lineages into

major monophyletic groups, thus ridding of the discrimination between unicellular and multicellular levels of organisation (Adl *et al*., 2005; Simpson & Roger, 2004). The major groups, or supergroups are briefly reviewed.

Opisthokonta are designated as the exclusive grouping of metazoans, fungi and multiple protist groups such as free-living choanoflagellates, parasitic Ichthyosporea, the amoeboids with filose pseudopods known as nucleariids, and the aerobic protists with long projecting tentacles known as filastereans (Cavalier-Smith & Chao, 1995, 2003; Ragan *et al*., 1996; Zettler *et al*., 2001; Shalchian-Tabrizi *et al*., 2008). In addition, the cellular slime mould *Fonticula* has recently been placed within the opisthokonts by phylogenetic means (Brown *et al*., 2009). This group has been confirmed as monophyletic by several single gene and multi gene phylogenetic analyses (Carr *et al*., 2008; Cavalier-Smith & Chao, 2003; Lang *et al*., 2002; Ruiz-Trillo *et al*., 2006; Shalchian-Tabrizi *et al*., 2008; Steenkamp *et al*., 2006). There are also morphological synapomorphies associated with opisthokonts, specifically unicellular motile stages bearing a single posterior flagellum and flattened mitochondrial cristae (Cavalier-Smith & Chao, 2003; Zettler *et al*., 2001). Finally, a molecular synapomorphy consisting in a unique ~12 amino acid insertion in the elongation factor protein EF1-α supports opisthokont monophyly (Baldauf & Palmer, 1993; Steenkamp *et al*., 2006) .

Amoebozoa include highly divergent lineages such as amoebae with broad pseudopodia (e.g. *Amoeba*), cellular and plasmodial slime moulds (e.g. *Dictyostelium* and *Physarum* respectively) and anaerobic, mitochondria-lacking commensal or parasitic lineages such as *Entamoeba* (Bapteste *et al*., 2002; Nikolaev *et al*., 2006; Smirnov *et al*., 2005). Early

rRNA-based phylogenies did not recover the monophyly of these groups, placing some taxa at the deepest branching level of the eukaryotes (Hinkle *et al*., 1994; Pawlowski *et al*., 1996). However, multi-gene analyses with improved phylogenetic methods produced phylogenies supporting the monophyly of the amoebozoan group and, importantly, helped determine that amitochondriate amoebozoans had secondarily lost their mitochondria (Baldauf *et al*., 2000; Bapteste *et al*., 2002).

According to the revised taxonomic system, the Archaeplastida are a supergroup consisting of three major lineages - Chloroplastida (land plants and green algae), Rhodophyta (red algae) and Glaucophyta (unicellular algae) - which descended from the eukaryotic host of the primary endosymbiotic plastid (Adl *et al.*, 2005). The origin of these lineages from a single primary endosymbiotic event is indicated by molecular evolution data which focuses on plastid-encoded genes and their comparison with nuclear-encoded genes (Bachvaroff *et al.*, 2005; Bhattacharya & Medlin, 1995; Douglas & Turner, 1991; Martin *et al.*, 1998b; Rodriguez-Ezpeleta *et al.*, 2005). The monophyly of the Archaeplastida is demonstrated by substantial phylogenomic data based on multi-gene analyses of both plastid and nuclear genes. (Moreira *et al.*, 2000; Reyes-Prieto & Bhattacharya, 2007; Rodriguez-Ezpeleta *et al.*, 2005; Rodriguez-Ezpeleta *et al.*, 2007). In addition, the presence of two ancestrally cyanobacterial membranes binding the plastids is an ultrastructural trait of the Archaeplastida, as it is in contrast with the presence of more than two plastid binding membranes in lineages that evolved from secondary or tertiary plastid endosymbiosis (Jarvis & Soll, 2002; Moreira & Philippe, 2001; Tomas & Cox, 1973).

The taxonomic group Rhizaria was initially proposed as an 'infrakingdom' including major amoeboid protist lineages characterised by root-like filose or reticulose pseudopodia, many of whom previously known as 'rhizopods', namely Cercozoa, Foraminifera, Retaria, Apusozoa and Heliozoa (Cavalier-Smith, 2002). Early molecular evolution data suggested phyla Cercozoa and Foraminifera were indeed closely related (Berney & Pawlowski, 2003; Keeling, 2001). However, a subsequent phylogenetic study based on RNA small subunits and actin, the first one to extensively sample putatively rhizarian taxa, redefined Rhizaria as monophyletic and one of the main eukaryotic supergroups, but with the collapse of Heliozoa, shown to be polyphyletic and therefore not a real group, and the exclusion of Apusozoa, the latter branching in an entirely different part of the eukaryote tree (Nikolaev *et al.*, 2004). More recent molecular evolution data, based on multi-gene phylogenetic analyses, confirmed monophyly of Rhizaria (Burki & Pawlowski, 2006; Burki *et al.*, 2007; Hampl *et al.*, 2009), and within Rhizaria, monophyly of Foraminifera and Radiozoa (Radiolaria minus Phaeodarea) to form the Retaria group (Moreira *et al.*, 2007). In addition, a single or double amino-acid insertion in universal eukaryotic protein polyubiquitin has been recognised as unique to Cercozoa and Foraminifera (Archibald *et al.*, 2003; Bass *et al.*, 2005).

Chromalveolata is a group proposed to contain the former kingdom Chromista (heterokonts, haptophytes and cryptophytes), and infrakingdom Alveolata (ciliates, apicomplexans and dinoflagellates), as they both putatively arose from a secondary endosymbiotic event whereby a heterotrophic bikont enslaved a red algae and subsequently became photosynthetic (Adl *et al.*, 2005; Cavalier-Smith, 1999; Harper *et*

al., 2005). This grouping can be validated by demonstrating that indeed there was a single secondary endosymbiotic event at the root of all chromalveolates, and also that the several diverse chromalveolate lineages are monophyletic. Both points are in fact contentious. On the issue of plastid inheritance, there are major lineages within the chromalveolates, such as the heterokont oomycetes, the alveolate ciliates and some dinoflagellates, that are not photosynthetic and are believed not to possess plastids (Archibald, 2009). According to the 'chromalveolate hypothesis', this is due to secondary loss of plastid and photosynthetic functions (Cavalier-Smith, 1999). So far, there have been significant findings of nuclear genes with plastid functions in non-photosynthetic organisms, such as the dinoflagellates *Oxyrrhis* and *Crypthecodinium* (Sanchez-Puerta *et al*., 2007; Slamovits & Keeling, 2008), which suggests these lineages only secondarily lost plastids and related functions. Similarly, secondary loss is suggested by the finding of 16 proteins of possible algal origin in non-photosynthetic ciliates *Tetrahymena* and *Paramecium* (Reyes-Prieto *et al*., 2008). However, critics of the chromalveolate hypothesis point out that horizontal gene transfer between cyanobacterial and ciliate lineages has been documented and therefore reduces the significance of this finding (Ricard *et al*., 2006), and a recent study comparing phylogenies obtained from plastid, mitochondrial and nuclear genes supports the idea of independent secondary plastid acquisitions within Chromalveolata (Baurain *et al*., 2010). With regards to monophyly of the chromalveolates, phylogenomic data based on multi-gene analyses suggest that alveolates and heterokonts are sister groups, and so are the haptophytes and cryptophytes, but that together the four lineage do not constitute a monophyletic chromalveolate supergroup (Burki *et al*., 2009; Burki *et al*., 2007; Hackett *et al*., 2007; Hampl *et al*., 2009; Harper *et al*., 2005; Minge *et al*., 2009; Patron

et al., 2007). This is in great part due to the unexpected branching position of Rhizaria, which in most multi-gene phylogenetic analyses is shown to be sister to the alveolates/heterokonts clade (Burki *et al*., 2009; Hackett *et al*., 2007; Minge *et al*., 2009), with one multi-gene analysis even suggesting Rhizaria and heterokonts are a monophyletic group, sister to the alveolates (Burki *et al*., 2007). This further complicates the 'chromalveolate hypothesis' as it would entail that all Rhizaria secondarily lost the red algal endosymbiont.

Excavata is a major eukaryote taxon proposed to include several of the unicellular flagellates previously thought to be deep-branching in the eukaryote tree (Cavalier-Smith, 2002). The defining criteria of this grouping are shared ultrastructural characters, the main one consisting of a distinctive suspension-feeding groove and an associated cytoskeletal system (Simpson, 2003). Some lineages have also been grouped in the Excavata according to molecular evidence (Cavalier-Smith, 2002; Dacks *et al*., 2001). The group is contentious because an initial comprehensive multi-gene phylogeny did not depict excavates as a monophyletic group (Simpson *et al*., 2006). However, a recent analysis focused on correcting long branch attraction artifacts indicates that Excavata are likely to be monophyletic and confirms three main sub-groups: Metamonada which are composed of anaerobes without classical mitochondria (i.e. *Giardia* and *Trichomonas*), Discoba which are composed of Discicristata (with discoid cristae on mitochondria i.e. *Trypanosoma* and *Naegleria*) and Jakobida, and *Malawinomas* (Hampl *et al*., 2009).

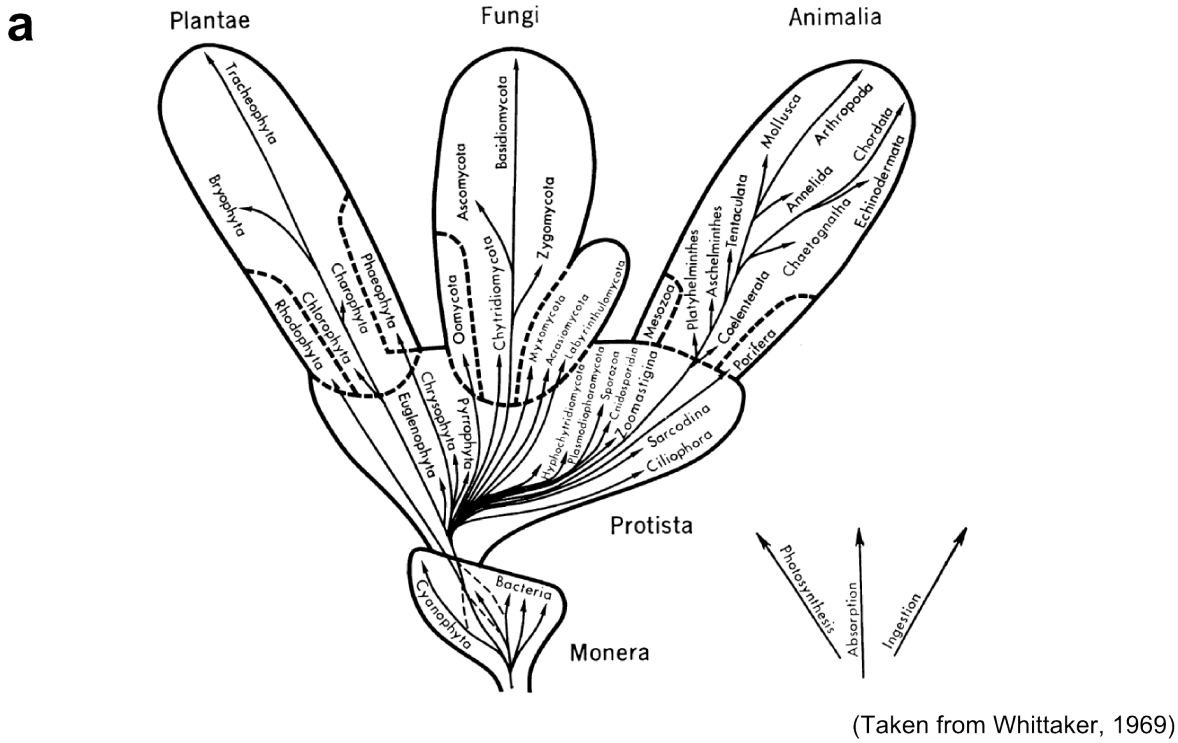


Figure 1.2 The diversity of eukaryotes: from 4 kingdoms to 6 'supergroups'. a The popular 5 kingdom system organised life forms according to three levels of organisation: Monera (prokaryotes), Protista (unicellular eukaryotes), and the three eukaryotic multicellular kingdoms (Plantae, Fungi and Animalia) (Whittaker, 1969). **b** A six supergroups system of eukaryote diversity is proposed. Rather than levels of cellular organisation, this sytem is based on the establishment of major monophyletic groups by molecular phylogenetics and improved knowledge of the diversity of protists (Simpson & Roger, 2004).

1.2.2 Outline of a consensus eukaryotic tree of life

Resolving the phylogenetic relationships between the major groups of eukaryotes is key to pinpointing the root of the eukaryotic tree. A leading hypothesis places the root between 'unikonts' and 'bikonts' (Stechmann & Cavalier-Smith, 2003) where unikonts evolved from a cell with a single cilium-bearing centriole and includes the Opisthokonta and Amoebozoa, whereas bikonts descend from a biciliate cell with a younger anterior cilium and include Archaeplastida and all other protists (Cavalier-Smith, 2002) (Figure 1.1a). This idea is based on the identification of a bikont specific derived gene fusion (Stechmann & Cavalier-Smith, 2002) and of unikont specific myosin domain structures (Richards & Cavalier-Smith, 2005). While the 'unikont' clade is robustly supported by phylogenomic evidence (Burki *et al*., 2007; Hampl *et al*., 2009), 'bikonts' are still contentious and not all analyses are consistent with this putative root (Roger & Simpson, 2009; Rogozin *et al*., 2009).

Other rooting scenarios for the eukaryotic tree have been proposed. For instance, two important phylogenetic analyses that sampled wide eukaryotic diversity, support a deep branching position of diplomonad *Giardia lamblia* (Bapteste *et al*., 2002; Morrison *et al*., 2007). One of those analyses also sampled trichomonad species *Trichomonas vaginalis,* and showed it to be monophyletic with the early-branching *Giardia* (Morrison *et al*., 2007). *Trichomonas vaginalis* and *Giardia lamblia* are anaerobes without canonical mitochondria, and have been proposed to be part of a monophyletic group called the Metamonada (Hampl *et al*., 2009). These data may thus be interpreted to mean that the root of the eukaryotes lies between the early-branching metamonads and the other eukaryotic lineages (Figure 1.1b). This hypothesis has been referred to as

'Metamonada first' (for instance in Wickstead *et al.*, 2010). However, a concern for this rooting is that the early-branching position of metamonads may be due to long branch attraction, as was proposed for other candidate 'archezoans' (Embley & Hirt, 1998; Philippe *et al*., 2000).

Yet another hypothesis, places the root of the eukaryotic tree between Euglenozoa and all other eukaryotes (Cavalier-Smith, 2010). This is mainly inspired by some primitive characteristics of Euglenozoa, namely the bacteria-like absence of mitochondrial outer-membrane channel Tom40 and DNA replication origin-recognition complexes (Cavalier-Smith, 2010) (Figure 1.1c).

The consensus eukaryotic tree, used as a phylogenetic framework for the studies presented in this thesis, makes no assumption on the root of the tree and is therefore depicted as a polytomy of four major eukaryote lineages: unikonts, Archaeplastida, excavates and a large group comprising alveolates, heterokonts, Rhizaria and haptophytes (Figure 1.1d). In addition, no assumption is made on contentious evolutionary relationships between the alveolates/heterokonts monophyletic group, Rhizaria and haptophytes. The excavates are depicted as a monophyletic group comprising metamonads and discicristates (Figure 1.1d).

Having outlined the taxonomic diversity and the consensus phylogeny of the eukaryotes, which will be used as framework for the evolutionary study presented in this thesis, the next step is to identify and describe the diversity of endocytic systems that have hitherto been investigated to significant molecular detail. In the next section,

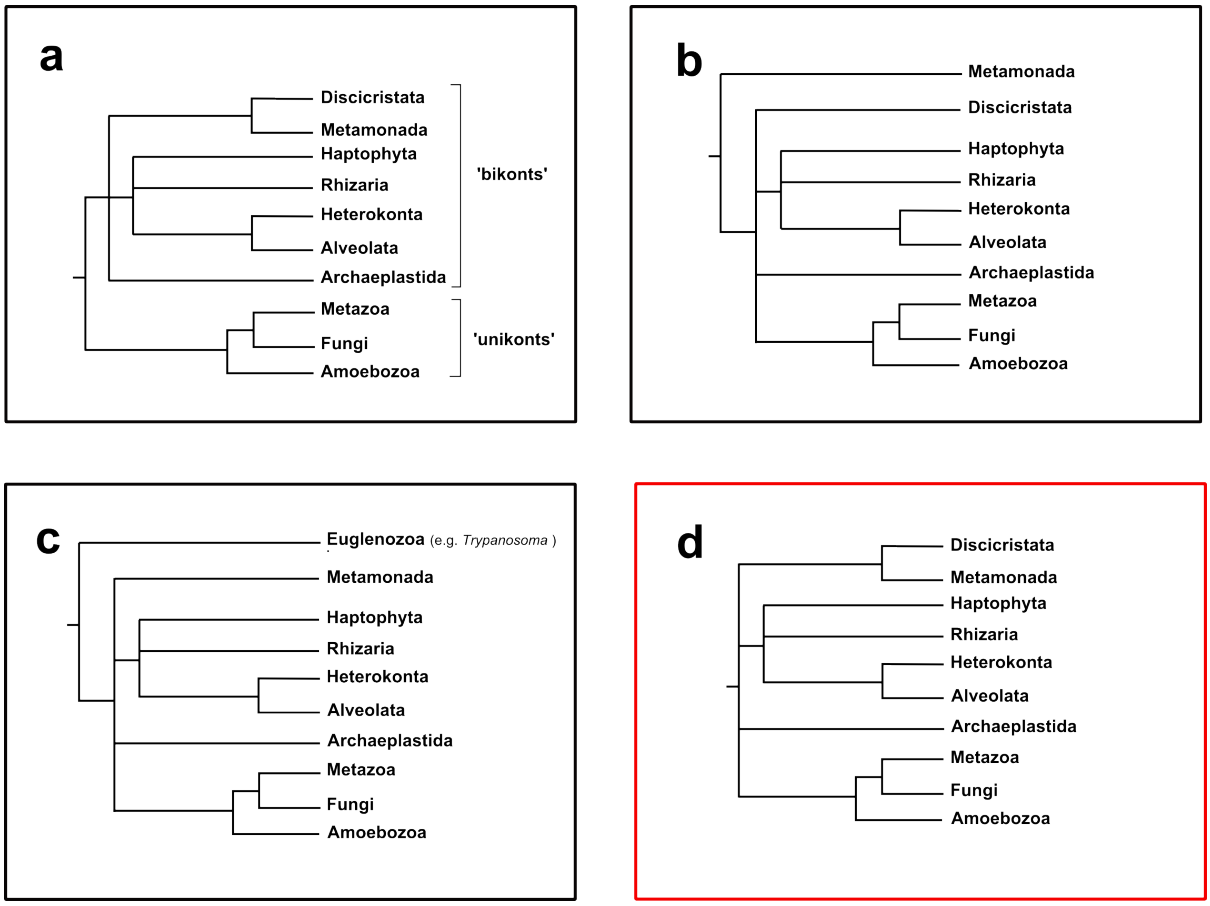


Figure 1.3 Alternative hypotheses for rooting the eukaryotic tree of life. The three schematic trees in black boxes represent alternative rooting hypotheses. **a.** Analysis of shared derived characters, such as gene fusions, suggests the root may be between unikonts and 'bikonts'. **b.** Some multi-gene phylogenetic analyses still place *Giardia* and *Trichomonas* at an early-branching position in the eukaryotic tree of life, suggesting the root is between metamonads and the rest of the eukaryotes. **c.** Euglenozoa have also been proposed as the early-branching eukaryote lineage, suggesting the root is between Euglenozoa and rest of the eukaryotes. **d.** The schematic tree in the red box is the consensus eukaryotic tree used in this thesis as phylogenetic framework, and does not make assumptions on the rooting of the tree. The tree is depicted as a polytomy of unikonts, Archaeplastida, excavates (metamonads + discicristates) and a clade comprising heterokonts, alveolates, haptophytes and Rhizaria. Monophyly of unikonts, of the alveolates + heterokonts group, and of excavates (metamonads + discicristates) is assumed.

different pathways and endocytic mechanisms will be briefly described, in light of relevant cell biological evidence.

1.3 Endocytosis comprises multiple distinct pathways

The cell has evolved several distinct processes to perform endocytosis (Figure 1.2). These are taxonomically defined by at least three factors: a) the nature of the internalised cargo, b) the level of cargo selectivity enforced by the endocytic apparatus, and c) the key molecular effectors responsible for carrying out the process. Most definitions given by cell biology text books follow at least one of these three criteria, though not always consistently. While a comprehensive and satisfactory classification system for endocytosis does not exist, we broadly outline here known types and their key differences.

1.3.1 Phagocytosis

Phagocytosis is the first endocytic process to be reported and studied extensively (Metchnikoff, 1883). It broadly refers to the engulfing of relatively large particles - larger than 0.5μm and up to ca. 8μm (Simon & Schmid-Schonbein, 1988) - via the local rearrangement of the actin microfilament cytoskeleton. It has been studied extensively in metazoan model organisms in relation to its role in the immune response (Ezekowitz *et al*., 1991; Greenberg & Grinstein, 2002), and in protists such as *Tetrahymena* (Jacobs *et al*., 2006), *Acanthamoeba* (Chambers & Thompson, 1976) and *Dictyostelium* (Waddell & Vogel, 1985), where it allows nutrient uptake via phagocytic predation.

In metazoan immune systems, phagocytosis is performed by specialised cells such as

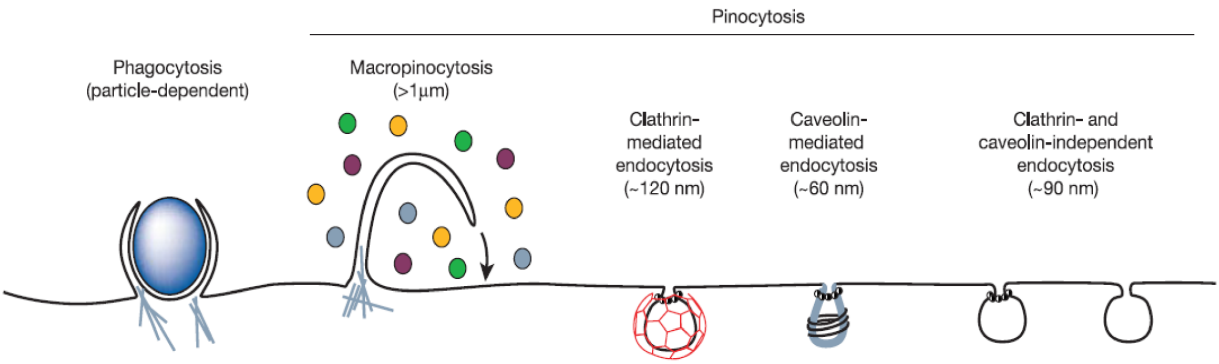


Figure 1.2 Endocytosis comprises mechanistically distinct pathways. Endocytosis is traditionally divided in phagocytosis (cell eating) and pinocytosis (cell drinking). However, within pinocytosis, diverse pathways with mechanistically distinct internalisation processes have been identified. These differ in vesicle size, vesicle coat composition, regulatory system, and function. (Taken from (Conner & Schmid, 2003)

neutrophils, monocytes and macrophages. These are known as 'professional' phagocytes to distinguish them from less specialised 'non-professional' cell types that perform phagocytosis (for instance, fibroblasts and epithelial cells) (Rabinovitch, 1995). The first step in the phagocytic pathway is opsonisation, whereby the particle is bound with either antibodies or specific recognition molecules called complement (Bennett *et al.*, 1963). Phagocytes then interact with the opsonised particle via either Fc receptors (FcR) which bind to the particle-bound antibodies; or complement receptors (CR), which interact with the particle via specific complement fragments (Chimini & Chavrier, 2000) The phagocytic mechanism differs depending on the type of receptor. In FcR-mediated phagocytosis, the particle is surrounded by thin plasma membrane extensions called pseudopodia while the receptors bind to the antibodies attached to the antigen. As the particle is engulfed, the pseudopodia join and fuse around it, forming the early phagosome (Uher *et al.*, 1981). In contrast, particles opsonised with complement fragment are internalised directly into the cytoplasm without the extension of

pseudopodia (May & Machesky, 2001; Newman *et al.*, 1991) .

Phagocytosis is not functionally restricted to the immune system. In metazoan cells, for example, it is employed in the clearing of apoptotic cells which is important in tissue development and tissue homeostasis (Finnemann & Rodriguez-Boulan, 1999; Reddien & Horvitz, 2000). Apoptotic cells can be engulfed by 'non-professional' neighbouring cells as well as 'professional' phagocytes. The latter are summoned via chemotactic factors released by the apoptotic cells (Lauber *et al*., 2003), and interact with signals such as phosphatidylserine, a phospholipid normally only present in the inside layer of the plasma membrane (Fadok *et al*., 2000). The inflammatory response, normally linked with phagocytic intervention in the immune system is then actively suppressed (Fadok *et al*., 1998; Meagher *et al*., 1992).

In protists, phagocytosis is known principally as a means of nutrient uptake by predation of other cells (Chambers & Thompson, 1976; Vogel *et al*., 1980; Waddell & Vogel, 1985). Studies on *Tetrahymena* have highlighted a phagotrophic digestive system: particles are internalised via a cell surface structure called cytostome, the content of the phagosome is digested in highly acidic environments, finally the phagosomes fuse with the cytoproct - a cell surface structure at the posterior end of the cell - and the residual contents are released (Allen & Wolf, 1979; Kitajima & Thompson, 1977). While the function and pathway of phagocytosis in *Tetrahymena* and in metazoan phagocyte are distinct, around 40% of the *Tetrahymena* phagosome proteome is shared (Jacobs *et al*., 2006) suggesting fundamental similarities and a common origin. Indeed, the adaptation of phagocytosis as a function of the immune

system likely occurred long before metazoan diversification, as specialised cell types within the social amoeba *Dictyostelium discoideum,* appear to provide immune-like functions by engulfing bacteria and sequestering toxins while navigating the organism (Chen *et al*., 2007).

1.3.2 Pinocytosis

Pinocytosis refers to uptake of fluid-phase medium and macromolecules. It includes all endocytic pathways that do not involve the engulfment of large particles and the formation of a phagosome. Within this broad category several distinct pathways have been studied. These differ in molecular mechanisms, recognition systems and size of the vesicle carrier.

1.3.2.1 Macropinocytosis

Macropinocytosis is dependent on cell surface ruffling. Ruffles are cytoplasm extensions formed by linear bands of outwardly polymerised actin filaments. Following appropriate stimuli the ruffles fold back on the plasma membrane forming a large vacuole called macropinosome (Swanson, 1989). This has been observed in mammalian cells such as macrophages and dendritic cells (Alpuche-Aranda *et al*., 1994; Racoosin & Swanson, 1992; Sallusto *et al*., 1995), but also in *Dictyostelium* (Hacker *et al*., 1997). The macropinosomes have neither vesicle coats, nor a concentration of receptors (Racoosin & Swanson, 1992) which suggests the process for internalisation of extra-cellular macromolecule is non-selective. However, there is evidence for regulation of macropinocytosis at different stages of its pathway. Cell surface ruffling can be stimulated by growth factor tyrosine kinase, phorbol esters and GTPase Rac1 (Bar-Sagi

& Feramisco, 1986; Ridley *et al*., 1992; Swanson, 1989). In addition, the maturation of macropinosomes is dependent on regulatory factors such as the GTPase Rab7 and lysosomal-membrane protein Lgp-1, which allow them to fuse with other endosomal and lysosomal vacuoles (Racoosin & Swanson, 1993). Interestingly, in phagocytes, macropinosome maturation is similar to phagosome maturation (Desjardins *et al*., 1994), whereas in other cell types macropinosomes are recycled back to the cell surface after their first stage of maturation (Hewlett *et al*., 1994).

1.3.2.2 Caveolin-mediated endocytosis

Caveolin-mediated endocytosis belongs to a class of endocytic pathways characterised by the formation of molecular coats around vesicles, which confer consistency in vesicle size. One such pathway is mediated by flask-shaped plasma membrane invaginations called caveolae. These are very consistent in size (60-80 nm in diameter) and feature a striated coat formed by parallel ~10 nm thick filaments (Somlyo *et al*., 1971; Stan, 2005). The protein caveolin is the main structural component of these plasma membrane pits. It forms dimers that bind cholesterol, inserted as a loop into the inner layer of the plasma membrane, and assembled with other caveolin dimers to create the striated coat (Murata *et al*., 1995). The internalisation of caveolae involves a complex signalling system which involves tyrosine-phosphorylation of its components (Shajahan *et al*., 2004). For instance, in endothelial cells serum albumin binds to the gp60 caveolae receptor which in turn activates the downstream G protein coupled Src kinase signalling pathway (Minshall *et al*., 2000). In addition, it has been shown that an increase in levels of cholesterol or treatment with synthetic glycosphingolipids stimulates caveolar endocytosis in a src kinase dependent way (Sharma *et al*., 2004).

Studies on the SV40 virus's exploitation of caveolae as way of entry into host cells has also shed light on the importance of actin cytoskeleton dynamics for caveolar endocytosis. Recruitment of actin patches is necessary as structural support of the preparatory stages of caveolae formation, andactin polymerisation is also required for internalisation of the caveolae into the cytosol (Pelkmans *et al*., 2002; Thomsen *et al*., 2002). Once internalised, the caveolae maintain structural stability until they fuse with an intermediate vacuole called the caveosome which in turn may fuse with the endoplasmic reticulum. The caveolar unit is then recycled back to the plasma membrane while maintaining its caveolin-cholesterol association (Pelkmans *et al*., 2004).

Caveolae have been studied extensively in mammalian cell types such as endothelial, epithelial and muscle cells but there is still debate regarding their function (Parton & Simons, 2007). Interestingly, the knock-out of the gene encoding caveolin-1 in mice caused the impairment of nitric oxide and calcium signalling in the cardiovascular system, but did not affect transport and transcytosis of cholesterol or serum albumin as expected (Drab *et al*., 2001). This suggests that caveolae and the associated endocytic system may be responsible for diverse cell signalling pathways. In addition, it has been suggested that caveolae play a role in storing and regulating the concentration of lipids in certain cell type plasma membranes. This is based on the extreme abundance of caveolae in adipocytes, and on data indicating that caveolin-1 overexpression facilitates uptakes of fatty acids and affects levels of free cholesterol, while caveolin-1 truncation mutant cells are affected by an imbalance of cholesterol and other lipid concentrations (Meshulam *et al*., 2006; Pol *et al*., 2001; Pol *et al*., 2004).

1.3.2.3 Clathrin-mediated endocytosis

Clathrin-mediated endocytosis (CME) is a type of receptor mediated pinocytosis characterised by the coating of plasma membrane pits with polymeric clathrin assemblies (Fotin *et al*., 2004). Clathrin monomers are recruited to plasma membrane regions rich in phosphoinositides by adaptor proteins which also bind cargo internalisation signals (Gaidarov & Keen, 1999). Several different adaptor proteins have been identified and are known to bind a diverse set of internalisation signals of cargo molecules and endocytic receptors (Traub, 2003). In addition, adaptor proteins like the AP2 complex interact with accessory proteins which mediate clathrin pit invagination, interaction with the actin cytoskeleton, vesicle scission and provide molecular scaffolding to coordinate the process (Maldonado-Baez & Wendland, 2006). Following the internalisation of the vesicle, the clathrin coat is disassembled by hydrolysing the phosphoinositide components, and the clathrin monomers are recycled back to the plasma membrane while the vesicle fuses with endosomes or lysosomes (Lemmon, 2001).

Clathrin-mediated endocytosis has been studied mainly in mammalian and budding yeast model organisms but it is of particular interest to neurobiologists because it is the prevalent pathway for the recycling of synaptic components during chemical signalling (Granseth *et al*., 2006). This has led to an effort to define the CME proteome using different approaches such as organelle-based proteomics (Blondeau *et al*., 2004; McPherson & Ritter, 2005), structural studies (Fotin *et al*., 2004; Shih *et al*., 1995) and biochemical protein interactions studies (Gaidarov & Keen, 1999; Hinshaw & Schmid, 1995). The resulting collection of data elucidates the molecular mechanisms behind the

functional stages of CME from clathrin recruitment and pit formation to vesicle scission and coat recycling. Different functional sub-modules are mediated and coordinated by a network of protein-protein and protein-phospholipid interactions which together can be described as the CME interactome (Lafer, 2002; Schmid & McMahon, 2007). For an in-depth review of the CME interactome, and definition of this process into functional modules, see Chapter 3.

1.3.2.4 Clathrin-independent endocytosis

In addition to the pinocytosis pathways listed so far, a number of potentially distinct endocytic pathways have been documented. In the literature these are referred as clathrin-independent endocytosis because the clathrin coated vesicle is the most extensively studied endocytic carrier and is universally present in all cell types and lineages sampled. However, within this broad category a more refined understanding of the pathways' characteristics is emerging.

A number of important ligand and receptor molecules are reported to be internalised in a clathrin-independent manner. For instance, major histocompatibility class I (MHCI) and the interleukin 2 receptor (IL-2R) are endocytosed following an alternative route (Neefjes *et al*., 1990; Subtil *et al*., 1994). They are both associated with a membrane trafficking pathway regulated by GTP-binding protein ADP-ribosylation factor 6 (ARF6) (Radhakrishna & Donaldson, 1997). Experiments on HeLa cells showed ARF6 to be distributed to the tubular membrane compartments where IL-2R alpha subunit (tac) and MHCI (Radhakrishna & Donaldson, 1997) co-localised. The formation of membrane tubules is induced by EHDI (homologue of CME regulator) and regulated by

nucleotide cycling on ARF6 and microtubules (Caplan *et al*., 2002). Internalisation of the tubular endosomes was confirmed to be independent from that of clathrin-coated vesicles, however, clathrin vesicles and ARF6 regulated tubular endosomes converge and fuse with early endosome (Naslavsky *et al*., 2003), suggesting distinct ways of entry into the cell may then follow the same trafficking and recycling route.

Research into the uptake of glycosylphosphatidylinositol anchored proteins (GPI-APs) has highlighted another distinct pinocytic pathway regulated by small GTPase cdc42 (Sabharanjak *et al*., 2002). This endocytic mechanism relies on the organisation of plasma membrane microdomains - known as lipid rafts - rich in cholesterol, sphingolipids and GPI-APs (Lakhan *et al*., 2009; Varma & Mayor, 1998). Experiments on CHO and Cos-7 cells have shown that GPI-APs are internalised in a clathrin and caveolin independent manner, forming distinct GPI-AP enriched early endosomal compartments (GEEC), and fusing with the recycling endosomal compartment (RE) (Sabharanjak *et al*., 2002). A study on the cholera toxin subunit (CTB) entry strategy into target cells has revealed a shared endocytic pathway with GPI-APs and identified GEEC as an uncoated tubular and ring-shaped structure (Kirkham *et al*., 2005).

A further distinct endocytic pathway may be the one mediated by flotillin proteins. Initially it was found that in mice with caveolin-1 knock out and loss of most of caveolin-2 expression, there were residual caveolae-like invaginations in endothelial cells that were slightly larger than typical caveolae (Drab *et al*., 2001; Zhao *et al*., 2002). Flotillin-1 and flotillin-2, proteins with a similar topology to caveolin-1, were found to be concentrated in areas of the plasma membrane which were budding into the

cell but were distinct from both clathrin pits and caveolin-1 positive caveolae (Frick *et al*., 2007; Glebov *et al*., 2006), and coassembly of the two proteins was found to induce membrane curvature and formation of invaginations morphologically similar to caveolae (Frick *et al*., 2007). However, expression of the two flotillins in caveolin-1 knock-out mice fibroblasts did not seem to result in an increase of caveolar structures, but rather caused an increase in tubular structures similar to the clathrin-independent endocytic carriers known to mediate the cdc42-regulated clathrin-independent pathway (Kirkham *et al*., 2008). A link between flotillins and this pathway is also supported by the fact that loss of flotillin-1 expression causes a reduction in the uptake of the GPI-AP CD59 (Frick *et al*., 2007; Glebov *et al*., 2006). With regards to the regulation of flotillin-mediated endocytosis, it has been reported that flotillins are endocytosed as a response to the expression of Fyn kinase, and that loss of function of the kinase impairs the internalisation of flotillin microdomains (Riento *et al*., 2009).

1.3.2.5 Evolution of pinocytic diversity and functional overlaps amongst distinct pathways

Having briefly described the remarkable diversity of pinocytic systems, it is important to consider how different pathways relate to each other, especially in light of the evolution of the eukaryote cell. Why did distinct endocytic systems evolve? What is the functional overlap among different systems, and how flexible and adaptable are the endocytic effectors that seem to be specific to a particular system? Considering its involvement in numerous vital cell functions, pinocytosis should be regarded as essential for every eukaryotic cell. However, there is a question whether any pinocytic pathway alone is necessary and sufficient for all endocytic functions required by the

cell. This is due to the presence of multiple pathways in the majority of model organisms studied and by the fact that impairment of a specific pinocytic pathway in most experiments is not lethal. For instance, CME is the most conserved pathway studied (Field *et al.,* 2007) and it is thought to be essential for constitutive nutrient uptake, membrane component recycling, cellular homoeostasis and signalling in eukaryotic cells (Brodsky *et al*., 2001; Conner & Schmid, 2003; Di Fiore & De Camilli, 2001). However, it has been demonstrated that in yeast and mammals cell types suppressing the expression of clathrin does not result in lethality (Neumann-Staubitz *et al*., 2010). For instance, absence of clathrin expression in a mammalian cell line, while substantially inhibiting endocytosis, did not have a noticeable effect on lysosome composition and cells were viable (Wettey *et al*., 2002). Likewise, massive depletion or full knock-down of clathrin by small interference RNA in HeLa cells, also resulted in severe disruptions of endocytosis of specific receptors, but did not result in cell death (Hinrichsen *et al*., 2003; Huang *et al*., 2004; Motley *et al*., 2003). Also, in budding yeast cells, clathrin deletion mutants are substantially affected in their growth rate but normally survive, bar the deletion of clathrin deficiency suppressor Scd1 (Lemmon & Jones, 1987; Payne & Schekman, 1985). The main notable exception to this trend is *Trypanosoma brucei,* which has been shown to necessitate expression of clathrin in both bloodstream and procyclic stages to be viable (Allen *et al*., 2003). This finding is significant because it suggests that while clathrin is ancient and important to distantly related protists, for instance *Trypanosoma, Giardia* (Hernandez *et al*., 2007), and *Paramecium* (Wiejak *et al*., 2004), clathrin-independent pathways only evolved in higher eukaryotes such Metazoa and Fungi. This could be the result of an increased necessity for regulatory flexibility of the endocytic systems, as multicellularity led to

cell specialisation and required a higher level of signalling coordination. It may also be associated with increase of functional demands of the cell. Clathrin-independent endocytic pathways for instance have been linked with functions like plasma membrane repair, cellular polarisation and cell spreading (Gauthier *et al*., 2009; Grande-Garcia *et al*., 2007; Idone *et al*., 2008).

There is also the question of how distinct the pathways described truly are. While initially clathrin-independent pathways were only theorised as a necessary alternative route for the internalisation of molecules in the absence of clathrin, there is now more substantial evidence with regards to unique features of these pathways such as the morphology of the carriers and regulatory systems. These were summarised in the previous section. Nonetheless, there are examples of functional overlaps among different pinocytic pathways, whereby a specific function is mediated by the same protein in different pathways. For instance, recruitment of dynamin is required in caveolin-mediated endocytosis (Pelkmans *et al*., 2002; Shajahan *et al*., 2004), suggesting a similar vesicle budding mechanism to CME. However, evidence indicates that dynamin is not required for macropinocytosis, nor for the clathrin-independent endocytic pathways regulated by Arf6 or Cdc42 (Kirkham *et al*., 2005; Naslavsky *et al*., 2003; Sabharanjak *et al*., 2002). The putative flotillin-mediated endocytic pathway does not seem to require dynamin either (Glebov *et al*., 2006). The SNX9 protein is also involved in more than one pathway, with evidence showing that it coordinates actin dynamics in CME, macropinocytosis, and Cdc42 regulated clathrin-independent pathways (Wang *et al*., 2010; Yarar *et al*., 2007). Remarkably, AP2 has been shown to play a role in the clathrin-independent endocytic pathway regulated by Arf6, though

only in a post-endocytic trafficking capacity (Lau & Chou, 2008).

1.3.3 Spatial regulation of endocytosis

So far we have described endocytic pathways according to its regulators, molecular mechanisms, carrier structure and interaction with the endomembrane system. A further aspect is the organisation and spatial distribution of the sites of endocytosis on the plasma membrane. What regulates this spatial distribution of endocytosis and how?

Some protists have specialised membrane micro-domains and structures dedicated to endocytosis. Kinetoplastida for instance feature a small invagination at the base of the flagellum known as the flagellar pocket (Morgan *et al*., 2002). This cell surface compartment - composed by several structural subdomains and with a distinct membrane proteome - is where all endocytosis and exocytosis activity takes place (Field & Carrington, 2009). In addition, protists such as *Trypanosoma cruzi*, *Tetrahymena* and *Paramecium*, feature an oral apparatus formed by a specialised funnel-like membrane invagination (Gonda *et al*., 2000; Plattner & Kissmehl, 2003; Porto-Carreiro *et al*., 2000). The apparatus consists of a cytostome where food is collected by ciliary motility, and a cytopharynx where vacuoles are formed and internalised into the cytoplasm (Cunha-e-Silva *et al*., 2010; Gonda *et al*., 2000).

1.3.3.1 Eisosomes

Specialised endocytic structures have not been reported in opisthokonts and the consensus is that besides the presence of microdomains with higher endocytic activity there is limited differentiation of the plasma membrane (Field *et al*., 2006). The only

known exceptions are polarised cells such as epithelial cells of mammals, or hyphae in filamentous fungi, where endocytosis occurs in distinct cell surface domains (Fuchs *et al*., 2006; Wang *et al*., 2001). This raises the question of what coordinates the spatial distribution of sites of endocytosis in undifferentiated plasma membranes.

A cell surface protein complex called the eisosome has been identified as a candidate endocytic spatial coordinator (Walther *et al*., 2006). The main components of the eisosome are two paralagous cytoplasmic proteins, Pil1 and Lsp1, which assemble in large quantities (approximately 2,000-5,000 molecules of each paralogue per eisosome) to create stable complexes at fixed sites across the plasma membrane (Walther *et al*., 2006). In addition, the transmembrane proteins Sur7 and Nce102 have been shown to accumulate at sites of eisosome formation (Frohlich *et al*., 2009; Walther *et al*., 2006), suggesting that they are integral components of the complex, and potentially act as anchors of the cytoplasmic side of the eisosome to the plasma membrane.

Cell imaging studies in yeast have shown that the internalisation sites of the membrane dye FM4-64 coincide with the sites where eisosomes assemble and that loss of Pil1 causes eisosome organisation to falter and aberrant membrane invaginations to form (Walther *et al*., 2006). It has also been reported that Pil1 and Lsp1 are part of a signalling pathway - regulated by sphingolipid components long chain bases (LCB) and protein kinases Pkh1/2, Pkc1 and Ypk1/2 - which is involved in cell wall integrity and cell wall remodelling during growth (Levin *et al*., 1994; Luo *et al*., 2008; Zhang *et al*., 2004). Pkh1/2 control eisosome assembly and disassembly by phosphorylating Pil1 and Lsp1 and in turn Pil1 and Lsp1 negatively regulate pkh1/2 and its downstream kinase

cascades (Walther *et al*., 2007; Zhang *et al*., 2004). Together, this evidence shows that eisosomes are involved in both endocytosis and cell surface integrity, and may therefore coordinate these two functions to provide a spatial organisation of endocytosis. However, the data on eisosomes comes from a taxonomically limited selection of model organisms, with representatives from only ascomycete fungi (Alvarez *et al*., 2008; Vangelatos *et al*., 2010; Walther *et al*., 2006). Given the lack of experimental evidence from other major eukaryote phyla the significance of eisosomes within the evolution of eukaryote endocytic systems needs further testing.

1.4 Thesis aims

In this introduction I have outlined a consensus eukaryote tree and defined endocytic diversity from the existing literature, which together forms the conceptual framework for the experimental work presented in this thesis. The overall aim of this research is to trace the diversity of endocytosis across the diversity of eukaryotes. This will help to predict the likely characteristics of the last common eukaryotic ancestor and investigate how the ancestral cell diversified. It is an important effort because the transition between prokaryote and eukaryotes is the most drastic evolutionarily leap in the history of life (Cavalier-Smith, 2006) and yet it is still a poorly understood process (Gribaldo *et al*., 2010; Roger & Simpson, 2009).

The remarkable diversity of endocytic systems poses a problem for such research because each distinct system is complex and mediated by several protein effectors (Conner & Schmid, 2003), and bearing this in mind our knowledge of how certain types of endocytosis work is almost certainly incomplete. The sheer size of the collective

endocytic proteome means that analysis of the entire system is beyond the remit of this PhD programme. On the other hand, selecting a watered-down selection of endocytic genes involved distinct pathways, with the idea to offer an average representation of endocytic systems, would defeat one of the primary objectives of this study which is to create a taxonomic map of endocytic diversity in eukaryotes. Such a map would help further characterise the endocytic abilities of the LCEA, as well as outline the diversification of ancestral pathways within the eukaryotic tree of life.

The solution adopted here is to select clathrin-mediated endocytosis - one of the most prominent and better characterised endocytic pathway in the literature (Conner & Schmid, 2003) - as a case study. Our prior knowledge that clathrin and the AP2 complex are at least as old as the last common ancestor of the sampled eukaryotes (Field *et al*., 2006) suggests that extending the study to include all of the CME interacting partners will tell us more about the state of the LCEA and how it diversified. More importantly, existing knowledge on the cellular mechanisms behind CME shows that the process occurs thanks to the integration of the specialised endocytic apparatus with the regulatory elements of the actin cytoskeleton and its associated motor proteins (Dawson *et al*., 2006; Schmid & McMahon, 2007). This implies that by studying the evolution of the complete identified CME proteome, one can study: 1) the evolution of endocytosis 2) the evolution of actin cytoskeleton regulation and 3) the evolution of modular interactions between the actin cytoskeleton and CME. This approach reflects a view of the evolution of early eukaryotes, whereby features that were acquired in early eukaryotes, such as the nucleus, the endomembrane system, phagotrophy, sex and mitochondria, were linked in the way they emerged. Specifically, it has been proposed

that the eukaryote cytoskeleton and the endomembrane system together enabled the evolution of phagotrophy (Cavalier-Smith, 2009). CME constitutes an ideal example of a link between the two cellular realms (endomembrane system and cytoskeleton), and studying its origin and evolution may provide insight to this evolutionary perspective.

Studying the eisosomes also offers possibilities to investigate the origin of regulatory and mechanistic interactions in endocytosis. As mentioned in section 1.3.3, current evidence suggests the eisosome may mediate the recruitment of endocytic proteins to site of endocytosis, the spatial distribution of endocytosis across the plasma membrane, and the integrity of the cell wall during endocytosis (Walther *et al.*, 2007; Zhang *et al.*, 2004). This may be done by interacting with both endocytic proteins and the cytoskeleton or cytoskeletal regulators. The significance of eisosomes is therefore potentially two fold: on the one hand as markers of sites of endocytosis in eukaryotes, and on the other hand, as potential regulator of the interaction between the cytoskeleton and endocytosis.

The thesis aims are therefore to specifically address the evolution of endocytosis-cytoskeleton interaction, which can be broken down into the following objectives:

For the study on the evolution of CME the main objectives were to:

- Outline the comprehensive network of interactions which enables CME in the model organisms in which it has been studied.
- Determine the taxonomic distribution of the CME proteome across a broad selection of eukaryote diversity.

- Reconstruct the evolution of CME by performing phylogenetic analyses of its protein effectors.
- Reconstruct the evolution of CME and eisosomes by identifying synapomorphic, novel protein domain combination.
- Construct likely models of CME at different stages of their evolution by identifying synapomorphic novel interactions between network components.

For the study on the evolution of eisosomes the main objectives were to:

- Determine the taxonomic distribution of the eisosome across a broad selection of eukaryote diversity.
- Reconstruct the evolutionary history of eisosomes by performing phylogenetic analyses.
- Investigate eisosome function and localisation in additional fungi with multicellularity and cellular differentiation by experimental cell biology.

2 Materials and Methods

2.1 Bioinformatics

2.1.1 Genomic and proteomic data sampling

The criterion for sampling genomes was to select well studied model organisms with existing advanced cell biological data, completely or near-completely sequenced genomes and predicted proteins databases. A total of 28 genomes were selected. They include 11 opisthokonts (6 metazoan, 4 fungi and 1 choanoflagellate), 2 ameobozoans, 4 Archaeplastida (1 land plant, 2 green algae and 1 red algae), 7 'chromalveolates' (2 ciliates, 2 apicomplexa, 1 diatom, 1 oomycete and 1 haptophyte) and 4 excavates (2 Discicristata and 2 Metamonada). The full list of taxa and online access details is in Table 2.1.

In the comparative genomic study of eisosome associated gene families an additional 17 fungal complete genomes as listed in Table 2.2. The Taxonomically Broad expressed sequence tags Database (TBestDB) which includes a broad selection of protists (O'Brien *et al*., 2007) was searched via online access (Table 2.3). Finally, the EST database of chytridiomycete *Blastocladiella emersonii,* was searched via online access of the relevant genome project (Ribichich *et al*., 2005) (Table 2. 3).

2.1.2 Sequence similarity searches by BLAST and PSI-BLAST

To search for DNA and protein sequence similarity in the selected databases the Basic Local Alignment Search Tool was used (BLAST) (Altschul *et al*., 1990). Using a heuristic algorithm that approximates the Smith-Waterman algorithm (Smith & Waterman, 1981) BLAST calculates a measure of similarity between variable sub-

regions of the full length query and target sequences. This is done by searching for small query sequence sub-sections, known as words, in the target sequence, and evaluating the similarities between the neighbouring regions of the query and target sequences using a scoring matrix (Altschul *et al*., 1990). Alignments with a sufficiently high alignment scores, and sufficiently low Expect (E)-values (probability that the similarity is by chance) are listed in the output as candidate homologues.

The BLAST algorithm is implemented in a series of programs designed to compare either a DNA or protein sequence against either a DNA or protein database. In homology searches carried out for this thesis BLASTp was used to compare protein query sequences against predicted protein databases. Because some genes are absent in predicted proteomes, tBLASTn, which compares a protein query sequence against a translated DNA database, was also used in homology searches. When testing the identity of an experimentally sequenced DNA clone, BLASTx, which translates a DNA query sequence and compares it to a protein database, was used. Finally, when searching the NCBI Trace Archives for highly similar sequences, BLASTn, which compares a DNA query sequence against a DNA database, was used. In all the BLAST searches carried out with amino acid sequences as search seeds, the word size used was 3, the scoring matrix was BLOSUM62 (Henikoff & Henikoff, 1992), and the E-value threshold employed was 10. Accession numbers of candidate homologous were recorded and its sequences downloaded in fasta format for further analysis. For the BLAST searches with nucleotide sequences as search seeds, the word size used was 11, the scoring parameters were default (match/mismatch = 5/2; for gap costs, existence = 5, extension = 2).

Table 2.1 Predicted proteome and translated nucleotide databases of diverse eukaryotes used for phylogenomic studies in this thesis .

Species	**Taxon**	**Source**	**Online Access**
Homo sapiens	Metazoa	GenBank	
Danio rerio	Metazoa		http://www.ncbi.nlm.nih.gov/mapview/
Drosophila melanogaster	Metazoa		
Caenorhabditis elegans	Metazoa		
Saccharomyces cerevisiae	Metazoa		
Ustilago maydis	Metazoa		
Arabidopsis thaliana	Archaeplastida		
Trypanosoma brucei	Discicristata		
Monosiga brevicollis	Choanoflagellata		
Chlamydomonas reinhardtii	Archaeplastida		
Ostreococcus taurus	Archaeplastida		
Phythopthora ramorum	Heterokonta	Joint Genomic	http://genome.jgipsf.org/
Thalassiosira pseudonana	Heterokonta	Initiative	
Naeglaria gruberi	Discicristata		
Nematostella vectensis	Metazoa		
Emiliania huxleii	Haptophyta		
Entamoeba histolytica	Amoebozoa		http://www.tigr.org/parasiteProjects.shtml*
Tetrahymena thermophila	Alveolata	J. Craig Venter	
Trichomonas vaginalis	Metamonada	Institute	
Magnaporthe oryzae	Fungi		http://www.broadinstitue.org/science/
Batrachochytrium dendrobatidis	Fungi	BROAD	
Dictyostelium discoideum	Amoebozoa	DictyBase	http://dictybase.org/
Cyanidioschyzon merolae	Archaeplastida	CyanMerolae Genome project	http://merolae.biol.s.u-tokyo.ac.jp/
Paramecium tetraurelia	Alveolata	ParameciumDB	http://paramecium.cgmcnrs-gif.fr/
Plasmodium falciparum	Alveolata	PlasmoDB	http://plasmodb.org/plasmo/
Toxoplasma gondii	Alveolata	ToxoDB	http://toxodb.org/toxo/
Giardia lamblia	Metamonada	GiardiaDB	http://giardiadb.org/giardiadb/

Table 2.2 Additional fungal predicted proteomes and translated nucleotide databases selected for the phylogenetic study on the eisosome

Species	Taxon	Source	Online Access
Candida glabrata	Saccharomycotina	GenBank	http://www.ncbi.nlm.nih.gov/mapview/
Eremothecium gossypii	Saccharomycotina		
Kluyveromyces lactis	Saccharomycotina		
Debaryomyces hansenii	Saccharomycotina		
Pichia stipis	Saccharomycotina		
Yarrowia lipolytica	Saccharomycotina		
Trichoderma reseei	Pezizomycotina	Joint Genomic Initiative	http://genome.jgi-psf.org/
Mycospharella fijiensis	Pezizomycotina		
Phanerochaete chrysosporium	Basidiomycota		
Laccaria bicolor	Basidiomycota		
Postia placenta	Basidiomycota		
Phycomyces blakesleeanus	Zygomycota		
Candida albicans	Saccharomycotina	BROAD	http://www.broadinstitute.org/science/
Lodderomyces elongisporus	Saccharomycotina		
Botrytis cinerea	Pezizomycotina		
Puccinia graminis	Basidiomycota		
Rhizopus oryzae	Zygomycota		

Table 2.3 EST libraries used in Pil1 and Lsp1 homology searches

Species	Taxon	EST clusters	Source	Online access
Allomyces macrogynus	Chytridiomycota	5078	TBestDB	tbestdb.bcm.umontreal.ca/
Antonospora locustae	Microsporidia	2376		
Capsaspora owczarzaki	Filasterea	8870		
Monosiga ovata	Choanoflagellata	6433		
Mortierella verticillata	Zygomycota	5724		
Nuclearia simplex	Nucleariidae	3313		

Proterospongia sp. *Saitoella complicata* *Taphrina deformans*	Choanoflagellata Taphrinomycotina Taphrinomycotina	1303 3840 3919	TBestDB	
Blastocladiella emersonii	Chytridiomycota	4873	Suely L. Gomes lab.	blasto.iq.usp.br/

The BLAST algorithm is implemented in a series of programs designed to compare either a DNA or protein sequence against either a DNA or protein database. In homology searches carried out for this thesis BLASTp was used to compare protein query sequences against predicted protein databases. Because some genes are absent in predicted proteomes, tBLASTn, which compares a protein query sequence against a translated DNA database, was also used in homology searches. When testing the identity of an experimentally sequenced DNA clone, BLASTx, which translates a DNA query sequence and compares it to a protein database, was used. Finally, when searching the NCBI Trace Archives for highly similar sequences, BLASTn, which compares a DNA query sequence against a DNA database, was used. In all the BLAST searches carried out with amino acid sequences as search seeds, the word size used was 3, the scoring matrix was BLOSUM62 (Henikoff & Henikoff, 1992), and the E-value threshold employed was 10. Accession numbers of candidate homologous were recorded and its sequences downloaded in fasta format for further analysis. For the BLAST searches with nucleotide sequences as search seeds, the word size used was 11, the scoring parameters were default (match/mismatch = 5/2; for gap costs, existence = 5, extension = 2).

To detect putative and highly divergent homologues, the Position-Specific Iterative

(PSI)-BLAST program was used (Altschul *et al*., 1997). PSI-BLAST works by first performing a normal BLASTp search, then using an alignment of the resulting high scoring hits to create a position specific scoring matrix (PSSM), also known as profile, which assigns different scores across the alignment in relation to how conserved each position is (Altschul *et al*., 1997). The PSSM is then used for a second search, where new candidate homologues can be identified and included in the profile. The process is repeated until the user is satisfied that all putative gene homologues have been identified or until repeat iterations do not recover additional sequences. The parameters used for PSI-BLAST are the same as the ones used for BLASTp.

2.1.3 Conserved protein domain analysis

Sequence with similarity identified by BLAST were selected for further analysis. To annotated these candidate CME/eisosome proteins, I identified their protein domain structures using two methods. Firstly, protein sequences were searched against the Conserved Domain Database (CDD) (Marchler-Bauer *et al*., 2009), which comprises more than 38,000 multiple sequence alignment models in the form of PSSMs (version 2.26). Reverse position specific (RPS)-BLAST is used to search the protein query against the PSSMs database, returning as output any hit below a given E-value threshold. In addition, CDD features a low-complexity filter which recognises compositionally biased regions within a given protein and discards them from the search. The E-value employed in CDD analysis was 0.1 and the low-complexity filter was used in the protein domain analysis presented in this thesis.

The second method used to predict protein domains was a search of the PFAM database

(Bateman *et al*., 2000). PFAM is a collection of annotated protein families. Each protein family annotation consists of a seed multiple sequence alignment (MSA) composed of a selection of representative sequences, a full MSA featuring all protein family members detectable in the PFAMSEQ database (composed of SWISS-PROT and SP-TREMBL, (Bateman *et al*., 2000), and a profile hidden Markov model (profile HMM). The profile HMM is a position specific scoring system, built from a MSA, which can be used to calculate the probability that a given protein belongs to the protein domain family (Eddy, 1998). In PFAM they are built from the seed MSA with the HMMER package which implements an augmented version of the basic HMM method (Eddy, 1998). Putatively homologous protein sequences were searched against the PFAM database (current version 24.0 features 11'912 protein families, (Finn *et al*., 2010)) accessible online (http://pfam.sanger.ac.uk/search). Any hit with an E-value below the 1.0 threshold was recorded as having putative protein domain homology.

2.1.4 Preparing data-sets for phylogenetic analysis

2.1.4.1 Renaming sequences

FASTA files of protein sequences of interest were downloaded from the appropriate genome project online portal (Tables 2.1-2.2) for phylogenetic analysis. The title lines of sequences in FASTA format typically feature a label with accession number and database details, and a comment section with functional annotation and species names. For the analyses in this thesis the title lines of FASTA sequences were edited to included only the accession number and an abbreviation of the species it was sequenced from. This was done with the REFGEN software (Leonard *et al*., 2009) which was accessed on its online portal (http://projects.exeter.ac.uk/ceem/ refgen.html). REFGEN renames

any sequence from NCBI, JGI and Broad institute curated databases. For other databases formats, the title line of FASTA sequences was edited manually.

2.1.4.2 Multiple sequence alignment

Multiple sequence alignments (MSA) of groups of homologous sequences retrieved from database searches were created with the MUSCLE (multiple sequence comparison by log-expectation) software (Edgar, 2004b) (Edgar, 2004a). The MUSCLE algorithm is briefly summarised here in three main stages (for mathematical definitions see Edgar 2004a and Edgar 2004b). Firstly a preliminary MSA is constructed. This entails creating a guide binary tree by clustering a *k*mer distance matrix with the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) and constructing the progressive alignment by following the branching order of the guide tree. In prefix order, each sequence is assigned a profile calculated with the novel log-expectation score (Edgar, 2004b), and at each node a pair-wise alignment of the sequence profiles are constructed until a complete MSA is obtained. The second stage consists in re-estimating the guide tree using the Kimura distance (Kimura, 1983), which is more accurate than *k*mer distance but requires an alignment. As in the first stage the tree is constructed by clustering the distance matrix - in this case calculated with Kimura distance - with UPGMA and then producing a progressive alignment by following the tree's branching order. In the third stage the MSA goes through a process of refinement. The guide tree from the second stage is divided in two sub-trees by deleting an edge, a profile of the MSA in each subtree is calculated, and a new MSA is created by aligning the two profiles. The new MSA is kept if the sum of pairwise alignment score is improved and the process is repeated until convergence is reached.

2.1.4.3 Masking multiple sequence alignments

In most phylogenetic analyses evolutionary signal is improved by discarding gaps and highly variable regions from MSAs because they may constitute noise (Castresana, 2000; Talavera & Castresana, 2007). Gaps and regions of hyper-variability of the MSAs prepared for this thesis were removed manually. The MSAs analysed in the thesis varied greatly both in terms of length and degree of character conservation so no single set of suitable criteria for alignment masking could be found. Overall, as it is recommended by literature (Castresana, 2000; Talavera & Castresana, 2007), stringent criteria for conserved character selection were applied to long MSAs, whereas relaxed criteria for conserved character selection were applied to short MSAs. Alignment masking was carried out with the SEAVIEW software (version 4; Gouy *et al*., 2010) which allows to visualise the MSA with colour schemes highlighting different nucleotides or amino acid groups informative sub-sets of the alignment. Masked MSAs were exported in FASTA format for the phylogenetic inference. In those instances where MUSCLE partially misaligned some sequences the data sets were visualised with sequence editing software SE-AL (Rambaut, 2007) and the misaligned sections of the MSA manually corrected.

2.1.5 Phylogenetic analysis

2.1.5.1 Evolutionary model selection

Maximum likelihood and Bayesian methods were selected for phylogenetic analyses. To use these methods effectively, it is important to estimate the best-fitting evolutionary model. Informing the phylogenetic analysis with the optimal character substitution scoring system can significantly improve the resulting evolutionary model (Keane *et al*.,

2006). In addition, rates of nucleotide substitution vary significantly across a gene (Fitch & Margoliash, 1967) as different selective constraints apply across a gene in relation to the functional attributes of given sites. This variation can be partially represented by a gamma distribution, where the substitution rate is plotted against the proportion of the sites that demonstrate the substitution rate. The shape of the resulting curve is described as the α parameter (Yang, 1996). Another approach to correct for differential rate variation is by calculating and correcting for the proportion of invariable sites, because assuming that all sites can vary while some do not can result in model violation while assuming the remaining sites evolve at a constant rate (Steel *et al*., 2000). The invariable model can be used in conjunction with the gamma model (Gu *et al*., 1995). A further parameter used in evolutionary models is observed character frequencies, to correct for bias in amino-acid composition and the relative proportions of different character changes (Abascal *et al*., 2005).

The MODELGENERATOR software (http://bioinf.may.ie/software/ modelgenerator/) was used to estimate the optimal evolutionary model for masked MSAs. The program works by constructing a series of neighbour-joining trees using a combination of differing model parameters and substitution matrices. The trees are used to compare evolutionary models. For each model analysis, the branch lengths and parameters are optimised using the PAL library (Drummond & Strimmer, 2001). For protein data, the models MODELGENERATOR supports are 1 of 10 substitution matrices either on their own, or in combination with gamma shape for among-site rate variation (+G), proportion of invariable sites (+I), observed amino acid frequencies (+F), or any combination of the three parameters (Keane *et al*., 2006). The current version supports

96 different model combinations. To calculate the gamma shape value (α), MODELGENERATOR approximates the distribution in discrete categories to reduce computational intensity. For all model searching analyses, 8 discrete categories were used to approximate the gamma distribution. MODELGENERATOR returns for each model combination a likelihood value (i.e. the likelihood of observing the data given the evolutionary model). The model combination with the highest likelihood value was selected for phylogenetic analysis.

2.1.5.2 Fast maximum likelihood phylogenetic inference

The maximum likelihood (ML) method calculates the likelihood that a hypothetical phylogenetic tree generated the observed molecular data. For every hypothetical phylogenetic tree, the algorithm calculates the likelihood of each amino-acid or nucleotide, at each site in the MSA, according to a specified evolutionary model. The likelihoods are calculated separately for each site, and are then added together to constitute the overall likelihood of the tree topology. The hypothetical phylogenetic tree that has the highest likelihood of generating the data is chosen as the best tree.

The ML method can incorporate complicated evolutionary models. This gives it an advantage over simpler methods such as parsimony when analysing data with different rates of evolution among sites and among lineages (J. Felsenstein, 1981). However, it can be computationally very intensive, thus prohibitive for analyses with large numbers of sequences. To make ML analyses computationally viable, heuristic methods have been employed to create fast ML algorithms. These work via a stepwise optimisation of the tree's branch position and branch lengths until the likeliest tree can no longer be

improved. For the analyses presented in the thesis, two fast ML methods were used: PHYML (Guindon & Gascuel, 2003) and RAXML (Stamatakis, 2006).

PHYML has the following algorithm. Firstly an initial tree is constructed from pairwise evolutionary distance matrix of the sequences with fast distance-based neighbour-joining method BIONJ (Gascuel, 1997). In the second stage, the likelihood of the initial tree is calculated. The tree is then refined by iterative simultaneous modification of tree topology and branch length. This is done by calculating all possible branch modifications, applying a proportion of these changes to the tree, and verifying whether the likelihood of said tree has improved. If the likelihood has not improved a smaller proportion of branch changes is applied. The process is repeated until the tree can no longer be improved (Guindon & Gascuel, 2003). PHYML can estimate and adjust evolutionary model parameters as the analysis is carried out, but can also be instructed with a specified evolutionary model including a substitution matrix, gamma distribution, and proportion of invariable sites (Guindon & Gascuel 2003).

The RAXML algorithm works as follows. Firstly, the starting tree is calculated with the parsimony method DNAPARS (from PHYLIP package). According the author, this provides a starting tree with better likelihood values in comparison with neighbour-joining or random method, making the optimisation stage quicker (Stamatakis *et al*., 2005). Also, the parsimony tree is constructed with stepwise addition (J. Felsenstein, 1981) which allows to construct distinct starting trees according to random seeds provided by the user. This in turn allows the optimisation analyses to run from several distinct starting parsimony trees, thereby reducing the chances of being stuck on a local

optima in the tree search space (Stamatakis *et al*., 2005). The RAXML tree optimisation stage is based on FASTDNAML (Olsen *et al*., 1994) but features unique algorithm steps designed to ease the computational complexity of the search. Firstly, in a subtree rearrangement step, the RAXML algorithm only optimises the length of the three local branches adjacent to the insertion site, storing the 20 trees with best likelihood for global branch length optimisation at the end of the rearrangement step. Secondly, as the initial optimisation stage proceeds, if a subtree rearrangement produces a topology with higher likelihood, instead of completing the entire rearrangement step that topology is immediately selected as current tree and further improvements are applied to that tree (Stamatakis *et al*., 2005). RAXML also can be instructed with a specified substitution matrix, gamma distribution and proportion of invariable sites. The program is hard coded to correct for among-site rate variation with 4 discrete categories of evolution rate.

2.1.5.3 Bootstrap analysis

To estimate confidence levels in the phylogenies inferred with fast ML methods the bootstrap statistical approach was used. In short, bootstrap consists in creating pseudo-replicate data sets by randomly and independently sampling data points (represented in a MSA data set as a character column) with replacement, from a MSA, until a data set of the same size as the MSA is obtained. Each pseudo-replicate data set is then analysed with the same phylogenetic method as the original analysis. The resulting set of bootstraps derived trees provide a measure of the statistical variance of the original data set (Joseph Felsenstein, 1985). Confidence levels can be applied as a percentage to a given clade in the phylogenetic tree, where the percentage represents the frequency of

the clade in the bootstrap replicate analyses. 60% is considered here as the minimum value to present a clade as credibly monophyletic, and 90% is considered here the minimum value to present a clade as monophyletic with robust statistical support. Both PHYML and RAXML incorporate bootstrap analysis. In the analyses presented here, a minimum of 100 bootstrap ML repetitions were performed for data sets with 100 or less sequences and a minimum of 1000 bootstraps ML repetitions were performed for data sets with more than 100 sequences.

2.1.5.4 Bayesian phylogenetic analysis

In addition to ML methods a Bayesian approach to phylogenetic reconstruction was used. The Bayesian approach consists in calculating a posterior probability for a given tree by combining the prior probability of the tree with the likelihood of observing the molecular data given the tree (Huelsenbeck *et al*., 2001). This approach is implemented with the Markov chain Monte Carlo (MCMC) algorithm which allows an approximation of the posterior probability distribution. MCMC perturbs a current tree by changing tree topology parameters such as branch position and branch length, and/or evolutionary model parameters such as gamma distribution or proportion of invariable sites, if implemented. The probability of the new tree is then evaluated with Metropolis-Hastings probability (Larget & Simon, 1999). If the new tree is an improvement it is accepted as current tree state which will be subjected to further perturbations and evaluation cycles (Huelsenbeck *et al*., 2001). The process continues for a pre-determined number of times deemed sufficient to obtain an adequate sample of the posterior distribution of tree topologies, with the program sampling trees at pre-determined intervals. A concern for the Bayesian method is that the heuristic tree

searching process may miss areas of tree space with high likelihood values. Metropolis Coupling (MC), an MCMC convergence acceleration technique, is used in phylogenetic Bayesian implementations to reduce the chances of this happening (Huelsenbeck *et al*., 2001). MC consists in running X number of MCMC chains, where X minus 1 chains are 'heated' i.e. designed to perturb current trees to a greater or lesser extent (depending on the 'heat' parameter), so that distant peaks in the tree distribution space may be explored. The Bayesian method allows to sample a statistically significant number of trees from the probability optima, by discarding trees with sub-optimal likelihood values from the analysis. A consensus final tree is derived from the sampled trees, with posterior probability values represented by the frequencies of each monophyletic clade in the tree set. In the analyses presented here, Bayesian analysis was performed with MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003).

2.2 Molecular biology

2.2.1 Laboratory methods

All work in the molecular biology laboratory was conducted according to aseptic technique to minimise chances of contamination. Protocol procedures involving the handling of *Magnaporthe oryzae* and *Saccharomyces cerevisiae* cultures were all conducted in a class II laminar flow cabinet. All flasks, glass bottles, pipette tips, 1.5 ml microcentrifuge tubes (Eppendorf), Oakridge tubes (Nalgene), Falcon conical tubes (Becton Dickinson biosciences), Miracloth (Calbiochem), Pestle and mortars, commercial blenders (Waring), bacterial and fungal growth media was sterilised by autoclaving for 15 minutes at 121 ºC. All chemicals were ordered from Sigma unless stated otherwise. Powder free nitril gloves were used in the handling of all equipment

used for the experiments. Table 1 in Appendix 1 includes full list of laboratory products suppliers and respective addresses.

2.2.2 Growth and maintenance of fungal stocks

The Guy-11 *Magnaporthe oryzae* (*M. oryzae*) strain was obtained from Prof. Nick J. Talbot's laboratory (Biosciences, College of Life and Environmental Sciences, University of Exeter, UK). The strain was grown on complete media (CM) (Talbot *et al*., 1993). CM is 1% (w/v) dextrose, 0.2% peptone, 0.1% yeast extract, 0.1% trace elements, 0.1% vitamin supplement (0.001 g L^{-1} biotin, 0.001 g L^{-1} pyridoxine, 0.001 g L^{-1} thiamine, 0.001 g L^{-1} riboflavin, 0.001 g L^{-1} *p*-aminobenzoic acid, 0.001 g L^{-1} nicotinic acid), 0.6% $NaNO_3$, 0.05% KCl, 0.05% $MgSO_4$, 0.15% KH_2PO_4 (pH to 6.5), and 1.5 % agar. The strain was incubated on CM agar plates in steady conditions of 26 ºC room temperature and 12 hour light and dark cycles. To create long-term stocks, *M. oryzae* was grown through filter paper disks (3 mm, Whatman International), which were then desiccated and kept at -20 ºC.

Saccharomyces cerevisiae strains were obtained from Dr. Tobias Walther laboratory (Organelle Architecture and Dynamics, Max Planck Institute of Biochemistry, Germany). They are all in the W303 genetic background (ura3-52; trp1Δ 2; leu2-3,112; his3-11; ade2-1; can1-100) (Baudin-Baillieu *et al*., 1997). The strains used are wild type and *pil1*Δ (*pil1*Δ::KAN, mating type a and α). The *S. cerevisiae* strains were grown on yeast extract (1%), peptone (2%), and dextrose (2%) media (YPD) with the addition of 2% agar for plates. For growth, the cultures were incubated at 30 ºC for 2 days. For short term storage the yeast strains were kept in YPD agar plates at temperatures of 2-8

ºC for up to 30 days. For long-term storage, individual yeast colonies were resuspended in 500 µl of liquid YPD media, mixed with 500 µl of 50% glycerol and kept at -80 ºC.

2.2.3 Nucleic acid analysis

2.2.3.1 DNA extraction

Total *M. oryzae* genomic DNA was extracted with the following method. A *M. oryzae* liquid culture was prepared by inoculating 200 ml of liquid CM (normal CM recipe but without agar) with a ~2 cm^2 sized mycelium sample recovered from an agar plate, and blending it in a sterile commercial blender (Waring). The resulting inoculated culture was grown in a sterile flask at 24°C for 48 hours with shaking (150 rpm). The mycelium from the liquid culture was collected by filtration through sterile Miracloth (Calbiochem) in a class II laminar flow cabinet. It was then wrapped in an aluminium foil parcel and placed in liquid nitrogen for ~10 seconds to freeze dry it. Subsequently the mycelium was placed in a mortar and ground to a fine powder which was then added to a sterile Oakridge tube containing 4 ml of 2 X CTAB. 2 X CTAB buffer is 0.0055 M Hexadecyltrimenthylammonium bromide (CTAB), 0.1 M Tris(hydroxymethyl)amino-methane (Tris), 0.0078 M Ethylenediaminetetraacetic acid (EDTA) and 0.7 NaCl. The tube was shaken and incubated in 65 ºC water bath for 20 minutes. An equal volume of chloroform:pentanol (24:1) was added and the tube shaken for 20 minutes at room temperature. The sample was then centrifuged at 14,000 x *g* for 10 minutes in a Beckman J2-MC high speed centrifuge. The resulting supernatant was transferred to another Oakridge tube with an equal amount of chloroform:pentanol and the centrifugation step repeated. The supernatant was transferred to a new Oakridge tube, and an equal volume of isopropanol was added to precipitate the nucleic acids. After

five minutes of incubation on ice the solution was centrifuged at 15,700 x *g* for 10 minutes. The resulting pellet was resuspended in 500 µl of TE (10 mM Tris, 1 mM EDTA [pH 8]) and then reprecipitated by adding 0.1 volumes of 3 M sodium acetate (pH 5.2) and two volumes of 95% ethanol and incubating at -20 ºC for 10 minutes. The solution was transferred to a 1.5 ml centrifuge tube and the DNA collected by centrifuging for 20 minutes at 17,000 x g. The resulting pellet was washed with 500 µl of 70% ethanol, dried for 20 minutes and finally resuspended in 25-100 µl nuclease free H_2O with 10µg ml^{-1} of RNase A. The total genomic DNA sample was stored at -20 ºC.

2.2.3.2 RNA extraction

To minimise chances of RNase contamination, all equipment, surfaces and gloves were thoroughly cleaned with 70% ethanol and with RNase Decontamination Solution RNaseZap (Ambion) prior to carrying out RNA extraction and analysis. Only nuclease free 1.5 ml centrifuge tubes and filter pipette tips were utilised to avoid RNA degradation. Only Diethyl pyrocarbonate (DEPC) treated water (Invitrogen) or nuclease free water (Promega) was utilised in RNA related protocols.

2.2.3.2.1 Extraction of *M. oryzae* total RNA

The RNA extraction protocol used for *M. oryzae* entails the preparation of an extraction buffer with 0.1 M LiCl, 0.1 M Tris (pH 8), 10 mM EDTA and 1% SDS. A *M. oryzae* liquid culture was prepared as described in section 2.2.2.1. The mycelium was harvested in Miracloth, freeze dried in liquid nitrogen, and ground to a powder which was then added to a sterile Oakridge tube containing 5ml of the extraction buffer and 5ml of phenol. After inverting for 1 minute 5ml of chloroform was added. The tube was mixed

again by inverting for 1 minute and then centrifuged at 15,700 x g, 4 ºC, for 30 minutes. The aqueous phase (top, clear phase) was transferred to a new sterile Oakridge tube, 1 volume of 4M LiCL was added and the sample was then incubated at 4 ºC overnight. The sample was then centrifuged at 15,700 x g, 4 ºC for 20 minutes. The resulting pellet was washed with 70% ethanol before being resuspended in 500µl DEPC treated water. The sample was transferred to a 1.5 ml sterile, nuclease-free centrifuge tube to which 500 µl phenol:CIA was added. The sample was mixed by inverting for 30 seconds before being centrifuged at 17'000 x g, 4 ºC, for 10 minutes. The aqueous phase was transferred to a new centrifuge tube and the RNA was precipitated by adding 0.1 volumes of 3M sodium acetate (pH 5.2) and 2 volumes of ethanol before being incubated overnight at -20 ºC. The sample was then centrifuged at 17'000 x g, 4 ºC for 20 minutes. The resulting RNA pellet was washed with 70% ethanol and resuspended in 100µl DEPC treated water. The total RNA solution was stored at -80 ºC.

2.2.3.2.2 Extraction of *S. cerevisiae* total RNA

To extract total RNA from *S. cerevisiae,* liquid cultures were prepared by incubating at 30 ºC with shaking (200 rpm) 50ml of liquid YPD inoculated with a single yeast colony. The cultures were allowed to grow until stationary phase is reached (~16 hours for YPD, ~48 hours for drop-out media). The cultures were then diluted to an optical density of OD660 = 0.3 (~3.85 x 106 cells/ml) and allowed to grow at 30 ºC with shaking for a further 3 hours. Yeast cells were collected by centrifuging at 800 x *g* for 5 minutes in Falcon tubes. Having discarded the supernatant the cells were resuspended with 1 ml of Trizol reagent (Invitrogen) to lyse the cells and transferred to a nuclease-free 1.5 ml centrifuge tube. To aid homogenisation, glass beads were added to the lysed

cells which were first incubated for 10 minutes at room temperature and then incubated for 10 minutes in a thermoshaker at 60 ºC. After vortexing for 15 seconds 200 µl of chloroform were added followed by vigorous shaking by hand and incubation at room temperature for 3 minutes. The sample was centrifuged at 12'000 x g, 4 ºC, for 15 minutes. The top RNA containing aqueous phase was carefully transferred to a fresh nuclease-free 1.5ml centrifuge tube with extra care not to disturb the red organic phase or the white DNA containing interphase. To precipitate the RNA, 500 µl of isopropyl-alcohol were added before incubating at room temperature for 15 minutes and centrifuging the sample at 12'000 x *g* for 10 minutes at 4 ºC. The supernatant was discarded and the RNA pellet washed with 75% ethanol. After the washing step, the pellet was air dried for 5-10 minutes and then dissolved in 50 µl nuclease-free water.

2.2.4 DNA manipulations

2.2.4.1 DNA digestion with restriction enzymes

All restriction enzymes (RE) and RE buffers were ordered from Promega. The optimal RE buffer composition was selected according to the Compatible Buffers search tool available at http://www.promega.com/guides/re_guide/Default.htm. Plasmid DNA digestions were prepared by mixing in a sterile microcentrifuge tube the desired amount of plasmid DNA (between 50 ng and 1 µg), with 1 µl of each desired restriction enzyme at 12 units/µl concentration, 2 µl of RE buffer at 10 X concentration (the RE buffer composition varied according to the RE used), 0.2 µl of acetylated bovine serum albumin at 1 µg/µl, and sterile, nuclease-free water to 20 µl volume. The reactions were incubated at 37 ºC for at least 2 hours and up to 16 hours, depending on the amount of plasmid DNA to be digested and the predicted enzyme activity.

2.2.4.2 Ligation of DNA fragments

DNA fragments were ligated with the Promega T4 Ligase DNA kit. A molar ratio of 1:3 vector:insert was used for ligation reactions. The appropriate quantities of DNA vector and ligand, were determined with the following formula: (ng vector x kb size of insert)/kb size of vector. Ligation reactions were thus prepared by mixing appropriate quantity of vector with the appropriate quantity of insert, 1 µl T4 DNA Ligase Buffer (10X), 1 µl T4 DNA ligase (1 Weiss unit/µl) and nuclease-free H_2O to make a final volume of 10 µl. The ligase reactions were incubated for 3 hours at room temperature.

2.2.4.3 DNA gel electrophoresis

Agarose gels were prepared by creating 0.6% (w/v) to 1.5% (w/v) agarose solution with a TBE buffer (0.09 M Tris-borate and 0.002M EDTA). The solution was heated to boiling temperatures to melt the agarose and allowed to cool to ~60 ºC. Ethidium Bromide was added to stain DNA (final concentration 0.5 µg ml^{-1}). The solution was allowed to cool to ~50 ºC and poured in the appropriate gel tray. The DNA samples of interest were stained with gel loading dye and loaded into the gel wells. The gel was then subjected to electrical current of maximum 110 v. 1kb plus (Invitrogen) and HyperLadder I (Bioline) size markers were used to determine the length of DNA fragments. The DNA fragments were visualised with a gel documentation system (Image Master® VDS with a Fujifilm Thermal Imaging system FTI-500, Pharmacia Biotech).

2.2.4.4 Gel purification of DNA fragments

DNA fragments were purified from the agarose gel using the Wizard® SV Gel and PCR Clean-Up System (Promega). Fragments were excised from the gel with a sterile razor and placed in a 1.5 ml microcentrifuge tube. For every 10 mg of agarose gel 10 µl of membrane binding solution (4.5 M guanidine isothiocyanate and 0.5 M potassium acetate [pH 5.0]) were added. The membrane binding solution acts as a chaotropic agent which is necessary for DNA to bind the silica membrane in the SV Minicolumn (Promega). The microcentrifuge was vortexed and then placed in a 60 ºC water bath for 15 minutes or until the gel was completely dissolved. A SV Minicolumn was placed in a collection tube and the dissolved DNA fragment solution transferred into it. Following 1 minute of incubation at room temperature the SV Minicolumn/collection tube unit was centrifuged at 16'000 x *g* for 1 minute. The liquid in the collection tube was discarded and the SV Minicolumn placed back into it. The column was washed by adding 700µl of Membrane Wash Solution (10 mM potassium acetate [pH 5.0], 16.7 µM EDTA [pH 8.0] and 80% ethanol) and centrifuging at 16'000 x *g* for 1 minute. The collection tube was emptied and the SV Minicolumn placed back into it. The wash step was repeated by adding 500µl of Membrane Wash Solution and centrifuging at 16'000 x *g* for 5 minutes. The SV Minicolumn was transferred into a new microcentrifuge tube and the DNA was eluted by adding 50 µl of nuclease-free water and centrifuging at 16'000 x *g* for 1 minute.

2.2.5 Cloning of PCR products

Routine cloning of amplified DNA fragments was performed using the StrataClone™ PCR Cloning Kit (Agilent technologies) which is based on the combined activities of

topoisomerase I and Cre recombinase. (Abremski *et al*., 1983; Shuman, 1994). The vector mix included in the kit contains two DNA fragments. One end of the arms is charged with topoisomerase I and also feature a modified uridine overhang, the other end contains a *loxP* recognition sequence. Purified DNA fragments from PCR reactions are ligated via the hybridisation of poly adenine (A) tail with the uridine overhangs. Cre recombinase then catalyses ion of the *loxP* recognition sequences to form a circular DNA molecule (pSC-A) with resistance to ampicillin and a *lacZ′* α-complementation cassette for blue-white screening.

Ligation reaction mixtures were prepared by adding 3 µl StrataClone Cloning buffer, 2 µl of purified DNA fragment (5-50 ng) and 1 µl StrataClone vector mix to a sterile microcentrifuge tube. The reactions were incubated for 30 minutes and then placed in ice. 1 µl of the cloning reaction mixture was added to a tube of thawed StrataClone SoloPack® competent cells. These cells express the gene encoding Cre recombinase which is necessary for the transformation to work. The genotype of SoloPack® competent cells is Tet^{r}Δ (*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F′ *pro*AB *lac*I^q ZΔM15 Tn10 (Tet^r) Amy Cam^r]. The cells were incubated on ice for 20 minutes and then heat shocked at 42 ºC for 45 seconds. Following another two minutes on ice, 250 µl SOC medium (pre-warmed to 42 ºC) were added to the competent cells which were then allowed to recover for 1 hour at 37 ºC with shaking (250 rpm). LB plates (1% tryptone, 0.5% yeast extract, 86mM NaCl [pH to 7.5] and 1% agar) were treated with ampicillin (1µl per ml of media) and X-gal (40µl of a 2% solution spread on set plates). The transformation mixture was placed on the LB plates in different amounts (20 µl, 80 µl and 200 µl) and incubated overnight at 37 ºC.

Bacterial colonies that successfully transformed with pSC-A containing the DNA fragment were recognised by white coloration, bacterial colonies that transformed with pSC-A without the DNA fragment were recognised by blue coloration. White colonies were picked, placed in 4 ml LB with 4 µl ampicillin, and grown overnight at 37 ºC with shaking (200 rpm).

2.2.6 Transformation of bacterial hosts

To transform bacterial hosts with plasmid DNA, the XL-1 Blue, transformation-competent, *Escherichia coli* strain was used. The genotype of XL-1 Blue is: *sup*E44 *hsd*R17 *rec*A1 *end*A1 *gry*A46 *thi rel*A1 *lac*$^-$ [F' *pro* AB$^+$ *lac*I^q *lac*ZΔM15 *Tn*10 (tet^r)]. An aliquot of 50 ng of plasmid DNA was mixed with a 100 µl aliquot of XL-I Blue competent cells in a 10 ml transformation tube, and incubated for 30 mins on ice. The cells were then heat shocked at 42 ºC for 45 seconds, before adding 800 µl of SOC (2% tryptone , 0.5% yeast extract, 0.05% sodium chloride, 20mM glucose, 10 mM, magnesium sulfate and10 mM, magnesium chloride). The cells were allowed to recover by incubation at 37 ºC for 1 hour with gentle shaking (125 rpm). Three aliquots of the cells (routinely, 20 µl, 80 µl and 200 µl) were then plated on LB plates treated with ampicillin. Plates were incubated overnight, successful transformants were selected by resistance to ampicillin.

2.2.7 Plasmid DNA preparation

The Promega Wizard® Plus Minipreps DNA Purification System was used to carry out plasmid DNA preparations. Overnight bacterial cultures were harvested by transferring to a microcentrifuge tube and centrifuging at 10'000 x *g* for 5 minutes. The supernatant

was discarded and the tube inverted and blotted dry on paper towel to remove excess media. 250 µl of Cell Resuspension Solution (50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 100 µg/ml RNase A) was used to resuspend the pellet by repeated pipetting. 250 µl of Cell Lysis solution (0.2 M NaOH, 1% SDS) were added to lyse the cells. SDS solubilises phospholipid and protein components of the plasma membrane which causes lysis and release of cell content. After brief mixing by inverting the microcentrifuge tube the cells were allowed to lyse for 1 minute at room temperature. To inactivate any endonucleases released from cells lysing, 10 µl of Alkaline Protease Solution (components not specified in the accompanying documentation) were added, the tube was inverted four times, and the lysed cells incubated at room temperature for 5 minutes. 350 µl of Neutralisation Solution (4.09 M guanidine hydrochloride, 0.759 M potassium acetate, 2.12 M glacial acetic acid) were added to the lysed cells and the tube inverted four times to mix. The high salt concentration causes potassium dodecyl sulfate to coprecipitate with chromosomal DNA and cellular debris. The bacterial lysate was then centrifuged at 14'000 x *g* for 10 minutes at room temperature.

DNA purification units was prepared by inserting one Spin Column in a Collection Tube for each sample. The clear aqueous top phase (~700 µl) was transferred to the spin column with extra care not to disturb the white organic phase. The DNA purification units were centrifuged at 14'000 x *g* for 1 minute at room temperature. The flow-through in the collection tube was discarded and 750 µl of Column Wash Solution (60 mM potassium acetate, 8.3 mM Tris-HCl, 0.04 mM EDTA, 60% ethanol) was added to the spin column. The purification unit was centrifuged at 14'000 x *g* for 1 minute, the flow-through discarded, and the step repeated with 250 µl of Column Wash Solution.

The Spin Column was then placed in a new sterile 1.5 ml microcentrifuge tube and the DNA eluted by adding 100 µl of nuclease-free water, incubating for 1 minute at room temperature, and centrifuging at 14'000 for 1 minute. The plasmid DNA preparations were stored at -20 ºC.

2.2.8 RNA manipulations

2.2.8.1 RNA gel electrophoresis

The integrity of total RNA extracted from *Magnaporthe oryzae* and *Saccharomyces cerevisiae,* was verified by denaturing gel electrophoresis. To denature samples a solution consisting of a total RNA sample, 50% (v/v) formamide, 2.2 M formaldehyde, 1 x MOPS/EDTA buffer (20 mM 3-[N-morpholino]-propanesulfoni acid, 5mM sodium acetate, 1mM EDTA [pH 7.0]) was incubated at 65 ºC for 15 minutes. 1µl of Ethidium bromide (10 mg/ml) was added to aid visualisation of RNA fragments. The gel was prepared by melting 1.2% (w/v) agarose in deionised water, allowing it to cool to ~60 ºC, and adding 40% formaldehyde (which gives 2.2 M) and 1 X MOPS/EDTA. An RNA marker was used during electrophoresis to determine size and molecular mass estimation of the RNA sample (0.24-9.5 kb ladder, Invitrogen).

2.2.8.2 Reverse-transcription PCR

To reverse transcribe RNA samples and amplify a DNA sequence of interest from the synthesised cDNA strands the Titanium™ One-Step RT-PCR Kit (Clontech) was used. This system allows to perform RT-PCR in one step by having both reverse transcriptase and DNA polymerase in the reaction mixture. Reactions were prepared by adding to the RNA sample (20 ng - 1 µg) 1 X One-Step Buffer (400 mM Tricine, 200 mM KCl, 30

mM $MgCl_2$, 37.5 µg/ml BSA), 1 X dNTP (deoxyribonucleotide triphosphate) mix (final concentration 0.2 mM each), 0.4 units/ml of Recombinant RNase Inhibitor, 50% (v/v) Thermostabilising Reagent (trehalose solution (Carninci *et al*., 1998), precise concentrations not mentioned in kit documentation), 20% (v/v) GC-Melt (U.S. Patent No. 5,545,539), 0.4 µM Oligo(dT)Primer, 1 X Titanium Taq RT Enzyme Mix (includes MMLV-RT, Titanium Taq Polymerase and TaqStart Antibody) and 0.9 µM of each experimental primer. The samples were then incubated in a thermal cycler instructed with the following program: 1. 50 ºC for 1 hour (required for cDNA synthesis); 2. 94 ºC for 5 minutes; 3. 30-35 cycles (depending on expected copy number of mRNA molecule of interest) comprising 94 ºC for 30 seconds, 65 ºC for 30 seconds, and 72 ºC for 1 minute per 1 kb of amplicon; 5. A final extension step at 72 ºC for 2 minutes. Results were analysed by gel electrophoresis and visualised with gel documentation system.

2.2.9 Treatment and reverse transcription of total RNA for rapid amplification of 5' and 3' cDNA ends

To determine the 5' and 3' ends of *MOPIL2* coding sequence the GeneRacer® Kit with *SuperScript® III RT* (Invitrogen) was used. This protocol works by first removing the 5' phosphates of non-mRNA and truncated mRNA within total RNA samples, thereby eliminating them from subsequent ligation with GeneRacer RNA oligo. The 5' cap structure of full length mRNA is then removed and the RNA oligo attached to the exposed 5' phosphates. Reverse transcription of the RNA sample is then carried out, and the 5' and 3' ends of the coding sequence of interest can be amplified using primers specific to the ligated RNA oligo sequence, in conjunction with gene specific primers.

2.2.9.1 Dephosphorylation of non-mRNA and truncated mRNA from *M. oryzae* total RNA

The total RNA sample was treated with calf intestinal phosphatase (CIP) by mixing 4 µl of the RNA (~600 ng/µl) with 1 µl CIP buffer (10X), 1 µl RNaseOut™ (40 U/µl), 1 µl CIP (10 U/µl) and 3 µl nuclease free H_2O in a 1.5 ml microcentrifuge tube. The reaction was mixed and incubated at 50 ºC for 1 hour and then placed on ice after brief centrifugation.

2.2.9.2 RNA precipitation

To precipitate the RNA, 90 µl nuclease-free H_2O and 100 µl phenol:chloroform were added to the RNA sample and vortexed vigorously for 30 seconds. The sample was then centrifuged at 16'000 x *g* for 5 minutes at room temperature. The top aqueous phase was transferred to a new 1.5 ml microcentrifuge tube to which 2 µl mussel glycogen (10 mg/ml) and 10 µl NaOAc (3 M) was added. After mixing, 220 µl ethanol (95%) was added and the tube vortexed briefly. The mixture was frozen in dry ice for ten minutes and then centrifuged at 16'000 x *g* for 20 minutes at 4 ºC to pellet the RNA. The supernatant was removed by pipetting so as not to disturb the pellet. 500 µl ethanol (70%) was added to the pellet and the tube inverted several times and vortexed briefly. The tube was centrifuged at 16'000 x *g* for 2 minutes at 4 ºC. The ethanol was removed by pipetting and the tube centrifuged again for a 1 minute. Remaining ethanol was removed and pell*et al*lowed to air dry for 2 minutes at room temperature. The pellet was resuspended in 7 µl nuclease-free H_2O.

2.2.9.3 Removing Cap structures from full length mRNA

Dephosphorylated RNA was treated with tobacco acid pyrophosphatase (TAP) to remove cap structures from full length mRNA. To do this, 7 µl dephosphorylated RNA was mixed with 1 µl TAP buffer (10X), 1 µl RNaseOut™ (40 U/µl), 1 µl TAP (0.5 U/µl) in a 1.5 ml microcentrifuge tube. The mixture was vortexed briefly and centrifuged to collect liquids and then incubated for 1 hour at 37 ºC. The tube was pulse centrifuged and placed on ice. The RNA was then precipitated as described in section 2.2.7.2.

2.2.9.4 Ligation of RNA oligonucleotide to decapped full length mRNA

7 µl of the dephosphorylated, decapped full length mRNA were added to the lypholised GeneRacer™ RNA oligo (0.25 µg) which was resuspended by gentle pipetting and mixing. The RNA solution was incubated for 5 minutes at 65 ºC to relax secondary structure and then placed on ice for 2 minutes. To ligate the RNA oligo to decapped RNA, 1 µl ligase buffer (10X), 1 µl ATP (10 mM), 1 µl RNaseOut™ (40 U/µl) and 1 µl T4 RNA ligase (5 U/µl), were added to the RNA solution and mixed by pipetting. The tube was incubated for 1 hour at 37 ºC, pulse centrifuged and placed on ice. The RNA was precipitated as described in section 2.2.7.2 except the RNA pellet was resuspended in 10 µl nuclease-free H_2O.

2.2.9.5 Reverse-transcription of mRNA

1 µl GeneRacer™ Oligo dT Primer (50 µM), 1 µl dNTP mix (10 mM each nucleotide), and 1 µl nuclease-free H_2O were added to the 10 µl RNA solution obtained in section 2.2.7.4 in a 1.5 ml microcentrifuge tube. The tube was incubated for 5 minutes at 65 ºC to unravel any secondary RNA structure and then chilled on ice for 1 minute. The

following reagents were then added: 4 µl first strand buffer (5X), 1 µl DTT (0.1 M), 1 µl RNaseOut™ (40 U/µl) and 1 µl SuperScript™ III reverse transcriptase (200 U/µl). The reagents were mixed well, pulse centrifuged, and incubated at 50 ºC for 45 minutes. The reverse transcription (RT) reaction was inactivated by incubating the tube at 70 ºC for 15 minutes and then chilling on ice for 2 minutes. The tube was pulse centrifuged and 1 µl RNase H (2U) was added to the reaction mix. Finally, the RT reaction was incubated at 37 ºC for 20 minutes and then stored at -20 ºC or used immediately for PCR reaction.

2.2.10 DNA sequencing

Sequencing of DNA fragments of interest within plasmid DNA was carried out by Beckman Coulter Genomics (Hope End, Takeley, Essex UK). 10µl solutions containing 100 ng/ml plasmid DNA were sent for primer extension (PE) sequencing with universal primers M13F and M13R unless otherwise stated. The sequencing was performed with the Sanger method using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Fluorescently labelled DNA fragments were analysed with the 3730xl DNA analyser (Applied Biosystems). Sequencing data was returned in the form of .AB1 format files. Sequencing reads were edited and assembled into contigs with Sequencher (GenoCodes Corporation).

2.2.11 Fungal transformation

2.2.11.1 Transformation of *Magnaporthe oryzae*

A *M. oryzae* liquid culture was prepared as described in section 2.2.2.1. The mycelium was harvested by filtration through sterile Miracloth and washed with distilled water before being transferred to a Falcon tube with 40 ml OM buffer (1.2 M $MgSO_4$, 10 mM

$NaPO_4$ [pH 5.8]) and a 5% solution of the lytic enzyme beta-1,3-glucanase (glucanex). The mycelium solution was then incubated at 30 ºC for 3 hours with gentle shaking (75 rpm) to allow the removal of the cell wall. The resulting protoplasts were transferred to a sterile Oakridge tube and overlaid with an equal volume of cold (~4 ºC) ST buffer (0.6 M sucrose, 0.1 M Tris-HCl [pH 7]). The protoplast solution was then centrifuged at 1'500 x g, 4 ºC, in a swinging bucket rotor. The protoplast containing OM/ST interphase was transferred to a new Oakridge tube and filled with cold STC buffer (1.2 M sucrose, 10 mM Tris-HCl [pH 7.5], 10 mM $CaCl_2$). The resulting solution was centrifuged at 1'500 x g, 4 ºC, for 10 minutes. The protoplast pellet was washed twice with 10 ml STC, and then resuspended in 1 ml STC. The concentration of protoplasts was measured by counting using a heamocytometer.

A transformation solution was prepared by combining in a microcentrifuge tube the protoplast solution (~10^7 cells/ml) with the insert plasmid DNA (5 µg - 10 µg) in a total volume of 150µl. The mixture was incubated at room temperature for 20 minutes after which 1 ml of PTC (60% PEG 400, 10 mM Tris-HCl [pH 7.5], 10 mM $CaCl_2$) was added and gently mixed. The protoplast solution mixture was added to 150 ml of BDCM media (1.7 g L^{-1} yeast nitrogen base without amino acids and ammonium sulfate, 2 g L^{-1} ammonium nitrate, 1 g L^{-1} aspargine, 10 g L^{-1} glucose, 0.8 M sucrose and 15 g L^{-1} agar, melted and cooled to 45 ºC) . The media was then poured into 5 Petri dishes, which were incubated in the dark for 16 h at 24 ºC. For selection of sulfonylurea resistant transformants, the plates were overlaid with 15 ml of BDCM containing 50 µg/ml of chlorimuron ethyl.

2.2.11.2 Transformation of *Saccharomyces cerevisiae*

To transform *Saccharomyces cerevisiae* strains, the method based on the use of Lithium Acetate, single-stranded carrier DNA and polyethylene glycol (Gietz & Woods, 2002) was used. For each transformation procedure, 10ml of YPD liquid medium was inoculated with a colony of the desired mutant *S. cerevisiae* strain and incubated overnight at 30°C with shaking (200 rpm). The budding yeast cell concentration of the overnight culture was estimated by measuring the optical density of the culture with a spectrophotometer. The culture was diluted to an OD_{600} of 0.4 in 50ml of YPD medium and incubated at 30 ºC with shaking (200 rpm) for 3 h. *S. cerevisiae* cells were then pelleted by centrifugation at 2'500 rpm for 5 mins, and resuspended in 40 ml of 1 x TE (10 mM Tris, pH 7.5; 1mM EDTA). The cells were pelleted again by centrifugation at 2'500 rpm for 5 mins, but were then resuspended in 2 ml of a LiAc/TE solution (100 mM Lithium Acetate, pH 7.5; 0.5 x TE). The cells were then incubated at room temperature for 10 mins. For each transformation, 1 µg of transformation vector, and 100 µg denatured sheared salmon sperm DNA were mixed with 100 µl of the cells suspended in LiAc/TE. At this point, 700 µl of a LiAc/PEG-3350/TE (1 x Lithium Acetate, pH 7.5; 40% PEG-3350; 1 x TE) were added and mixed well into the cell suspension solution, which was then incubated at 30 ºC for 30 mins. After adding 88 µl of DMSO and mixing well, the cell solution was subjected to heat shock at 42 ºC for 7 mins. The cells were then centrifuged for 10 seconds at 12'000 x g, and the supernatant was discarded. After resuspending in 1 ml of 1 x TE, the cells were pelleted again, but then resuspended in 100 µl of 1 x TE. Finally, the cells were plated on synthetic, minimal defined medium lacking in uracil (6.7 g L^{-1} yeast nitrogen base, 20 g L^{-1} glucose, 0.1 g L^{-1} adenine, 0.1 g L^{-1} arginine, 0.1 g L^{-1} cysteine, 0.1 g L^{-1} leucine, 0.1 g L^{-}

[1] lysine, 0.1 g L^{-1} threonine, 0.1 g L^{-1} tryptophan, 0.05 g L^{-1} aspartic acid, 0.05 g L^{-1} histidine, 0.05 g L^{-1} isoleucine, 0.05 g L^{-1} methionine, 0.05 g L^{-1} phenylalanine, 0.05 g L^{-1} proline, 0.05 g L^{-1} serine, 0.05 g L^{-1} tyrosine, 0.05 g L^{-1} valine, and 20 g L^{-1} agar). Successful transformants were prototrophic for uracil, thus grew in the uracil deficient minimal medium.

3 Modular definition of the CME interactome network

3.1 Introduction

Clathrin-mediated endocytosis (CME) is a form of pinocytosis which has been studied extensively in the attempt to understand the molecular details of neural synaptic activity (Granseth *et al*., 2006; Mueller *et al*., 2004). In the introduction of this thesis, I identified the evolutionary history of clathrin-mediated endocytosis (CME) as an important case study for understanding the origin and diversification of endocytosis in eukaryotes. There are multiple factors that make it so.

First of all it is ancestral to all eukaryotes, as demonstrated by the presence of key CME proteins clathrin and the AP2 α and β subunits in all the main eukaryotic supergroups (Dacks *et al*., 2008; Field *et al*., 2006), with phylogenetic analyses of the adaptin protein family showing that AP complexes share a common origin but diversified prior to the branching of the major eukaryote supergroups (Boehm & Bonifacino, 2001; Dacks *et al*., 2008; Schledzewski *et al*., 1999). Considering that the four recognised AP complexes are involved at different compartments of the endomembrane system, and that only the AP2 complex is known to work in endocytosis (Boehm & Bonifacino, 2001; Robinson, 2004), this suggests that the process of endocytosis via the clathrin coated vesicle was already specialised in the last common eukaryotic ancestor (LCEA). Moreover, molecular evolution data indicates that other protein families involved in CME, such as the AP180 adaptor, epsin and dynamin, were too present in the LCEA (Elde *et al*., 2005; Field *et al*., 2006; Gabernet-Castello *et al*., 2009), adding further credibility and detail to the CME pathway in LCEA.

Secondly, as a result of a massive effort to understand endocytic activity in synapses, an impressive degree of molecular detail has been achieved in the description of the CME machinery. This includes the identification of a near complete inventory of the functional proteins that play a role in the pathway (Blondeau *et al*., 2004; McPherson & Ritter, 2005) and the understanding of what they do and how they interact (see sections 3.3.1-3.3.6 for a descriptions).

Importantly, a key aspect of the CME process is the interaction of core endocytic structural and adaptor proteins with the actin cytoskeleton. Current data has identified multiple proteins that mediate this interactions. As mentioned in the introduction, studying the evolution of the complex system behind cytoskeletal regulation of plasma membrane dynamics is crucial to understanding the early autogenous evolution of the eukaryotic endomembrane system, and it is an important question of this study.

In light of this question, it is proposed here to study CME holistically, *i.e.* as the known integrated system of protein-protein and protein-lipid interactions, known as the CME interactome (CME-I) (Schmid & McMahon, 2007), that carry out the functional requirements of this process. This provides the opportunity to investigate the evolution of CME by studying the evolution of interacting proteins in relation to each other and to the specific sub-processes they mediate. According to this approach, a synapomorphy can be identified not only by a novel protein family, but by an interaction, identified by experimental analysis, between any two given protein families. If, for example, origin of protein family A (Figure 3.1) is pinpointed at point X in the eukaryotic radiation, and

origin of protein family B is pinpointed to at point Y in the eukaryotic radiation which is more derived than branch X, the interaction between A and B is a synapomorphy which should be pinpointed to branch Y (Figure 3.1). By applying these criteria to the whole network of interacting proteins it is possible to construct a model depicting the most likely CME-I network on different branches of eukaryotic tree of life. The resulting model will allow to estimate whether major modifications have occurred to the cellular process and at which stages of eukaryotic evolution.

The main questions covered in this and the next chapter are therefore the following. Considering that CME is ancient and likely to have been significantly specialised in the LCEA, and that the known mammalian and yeast models of CME are very complex, involve a system of cytoskeletal regulation, and are best described by an interactome of more than 40 proteins, how did the complex CME system studied in yeast and mammals evolve? Also, how did the cytoskeletal regulatory system of CME evolve, and what can it tell us about autogenous co-evolution? Finally, I ask what is the degree of conservation of the CME interactome across the diversity of eukaryotes?

The aim of this chapter is to construct a model of the CME-I that can be used as the foundation for the phylogenomic study presented in Chapter 4. This has in part already been done. Previous analyses and perspectives have outlined interactomes related to CME. Notably, Lafer (Traffic, 2002), summarised clathrin binding partners in the format of an interaction web, and the model partially overlaps with the one presented here. However, it covers the entirety of vesicular trafficking mediated by clathrin coated vesicles, both in an endocytic and a secretory capacity, and the functional and modular

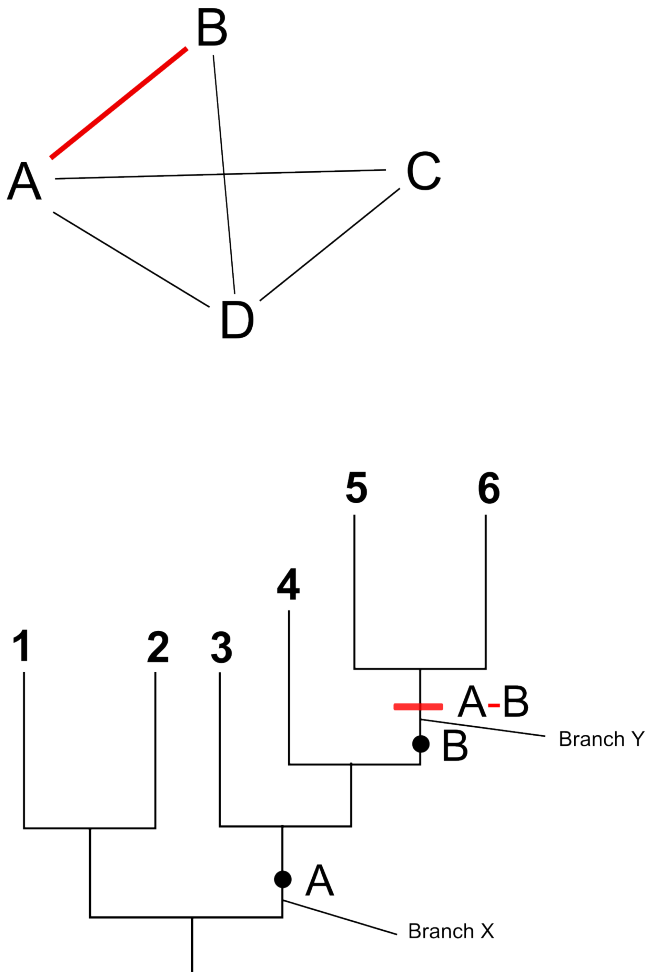


Figure 3.1 Novel interaction between two network components as a synapomorphy. Considering the interaction between hypothetical network components A and B from an hypothetical interactome composed by A, B, C and D, if A is acquired in branch X and B is acquired in branch Y which is more derived than X, the origin of the A-B interaction is pinpointed to branch Y. Therefore, while A is a synapomorphy (shared derived character) of taxa 3-6, and B is a synapomorphy of taxa 5-6, the A-B interaction is a synapomorphy specific to taxa 5-6. In the schematic tree, black cirlces represent acquisition of network component, and red line represents acquisition of interaction between two components.

characterisation of the network is minimal. Two further examples are by Drubin (Novartis Found Symp, 2005) and Traub (Nat. Rev. Mol. Cell. Biol., 2009), where the former looks at protein networks in endocytosis and actin cytoskeleton in yeast, and the latter at the network of clathrin adaptors and respective internalisation signals. The interactome model presented by Schmid & McMahon (Nature, 2007), is the closest to the one described in this chapter, both in terms of protein selection and modular correlation to functional stages of CME. However, in none of these interactome models there is specific information regarding the binding abilities of conserved functional domains. This is an important issue because even a brief review of the CME proteome can highlight several protein families with complex protein domain architectures, with some functional domains being shared between more that one protein family, but in different combinations. For an evolutionary study of the CME-I network, it is thus important to pinpoint when specific functional domains where acquired within a protein family, because this will allow us to also pinpoint when the specific interaction mediated by the functional group was acquired. Therefore the model of the CME-I needed here has to highlight the protein domain architectures of its components.

The CME pathway can be described as linear sequence of steps, i.e. formation of clathrin coated pits, membrane invagination, vesicle scission, internalisation of vesicle, vesicle uncoating and recycling of vesicle coat components (Brodin *et al*., 2000), recognition of cargo internalisation signals (Traub, 2003), recruitment of clathrin to sites of endocytosis (Ford *et al*., 2001), deformation of the plasma membrane (Futterer & Machesky, 2007), recruitment of dynamin to site of vesicle scission (Hinshaw & Schmid, 1995), polymerisation of actin filaments (Yamada *et al*., 2009),

dephosphorylation of vesicle coat components (Verstreken *et al*., 2003). These molecular functions are spatially and temporally coordinated within the CME-I. In this chapter, I present a model of the CME-I network by 1) reviewing the molecular components of this system, 2) creating a visual model depicting the intricate set of interactions, and 3) drawing a functional link between sections of the interactome and functional requirement in CME.

3.2 Materials and methods

3.2.1 Definition of the CME-I network

The protein components of the CME-I network were identified using literature review. The criteria for inclusion is proof of involvement in CME molecular machinery by functional molecular and cell biological studies and physical interaction to at least two CME proteins. Protein domain architectures were identified by searching conserved domain database CDD and Hidden Markov Model database PFAM, using each CME-I network protein as seed. To visualise the CME-I network, a diagram was drawn out depicting each protein-protein and protein-lipid interaction catalogued in the study. This visual model includes the name of the protein family and the protein domain architecture of each CME-I protein. Protein domain architectures were identified by searching conserved domain database CDD and Hidden Markov Model database PFAM, as described in section 2.1.3, with each CME-I network protein as seed.

Where available, information on the location of the binding sites was used to draw lines representing interactions. In addition, AP2 subunit binding motifs and clathrin heavy chain binding motifs were included in the model. Known interactions between

endocytic adaptors and cargo receptors, and internalisation signal motifs were also included (Figure 3.2). Information on protein function of CME-I network proteins was used to to divide the CME-I network into functional modules.

The formatting convention used to refer to CME-I proteins in the original literature varies depending on the model organism they were studied in. In this study, protein names derived from acronyms are written in capital letters, whereas all other protein names are written in smaller case except when starting a sentence. When referring to protein domains, the naming convention adopted by PFAM is used in this study.

3.3 Results

Data from 65 published experiments was used to create an inventory of the known protein-protein and protein-lipid interactions which make up the CME-I network (see Table 3.1, for references). The literature review focused on robust experimental evidence to confirm interactions between two components. A total of 39 proteins belonging to 21 gene families were included in the interactome. For each CME-I network protein, Table 3.1 lists the SWISS-PROT database accession numbers, known function, known CME interacting partners, experimental background information, taxonomic information, and source literature. A visual model of the CME-I network was created (Figure 3.2). In the model, CME-I components were described in terms of their protein family and known protein domain architecture. Knowledge on the properties of protein domains or binding motifs, was used to pinpoint binding sites for each interaction.

Table 1. CME-I network proteins. Protein names and accession numbers are retrieved from *Homo sapiens* UNIPROTKB/SWISS-PROT database (Schneider, 2009). Putative functions, interaction partners and the experimental background are specified from the referenced literature. The keys to references are fully listed below.

Protein name	Accession number	Function	Interaction partners	Experimental background	Literature
Clathrin heavy chain	Q00610	Structural component of vesicle coat.	AP2 β subunit, β-arrestins, ARH, disabled 2, numb, epsins, amphiphysins, SNX9, HSC70, auxilin.	Murine, bovine, budding yeast, COS-7 and HEK-293 cell lines.	1-11.
Clathrin light chain	P09496	Structural component of vesicle coat.	Clathrin heavy chain (CHC).	Bovine, budding yeast.	12,13.
AP2 α subunit	O95782	Recruits endocytic proteins to plasma membrane.	PI(4,5)P2, disabled 2, numb, epsins, AP180, SNX9, EPS15, auxilin.	Budding yeast, murine, bovine and human cell lines.	5, 14-16, 1, 6, 9.
AP2 β subunit	P63010	Binds and recruits clathrin to plasma membrane.	CHC, β arrestins, ARH.	Murine, human cell lines.	2-4.
AP2 μ subunit	Q96CW1	Interacts with plasma membrane and internalisation signals.	PI(4,5)P2, tyrosine internalisation signal.	Murine, human cell lines.	17, 18.
AP2 σ2 subunit	P53680	Structural integrity of AP2 complex.	AP2 α, β and μ subunits.	Murine, human cell lines.	17
β-arrestin 1	P49407	Adaptor protein. Involved in G-protein coupled receptor (GPCR) endocytosis.	PI(4,5)P2, AP2 β subunit, CHC., G-protein coupled receptor (GPCR).	Bovine, COS-7, HEK-293 and COS-1 cell lines	3, 19-21.
β-arrestin 2	P32121	Adaptor protein. Involved in G-protein coupled receptor (GPCR) endocytosis.	PI(4,5)P2, AP2 β subunit, CHC., GPCR.	Bovine, COS-7, HEK-293 and COS-1 cell lines	3, 19-21.

ARH	Q5SW96	Adaptor protein. Involved in low density lipid receptor (LDLR) endocytosis	PI(4,5)P2, AP2 β, CHC, low density lipid receptor (LDLR).	Murine.	4
Disabled-2	P98082	Adaptor protein. Involved in low density lipid receptor (LDLR) endocytosis.	PI(4,5)P2, AP2 β, CHC, LDLR.	Murine.	5
Numb	P49757	Adaptor protein. involved in endocytosis of epidermal growth factor (EGF), transferrin and notch receptors.	PI(4,5)P2, AP2 α, EPS15, epidermis growth factor (EGF) receptor, transferrin.	Murine, fruit fly.	14, 22-24.
AP180	O60641	Adaptor protein. Involved in endocytosis of EGF and transferrin.	PI(4,5)P2, CHC, AP2 α, EPS15, EGF, transferrin.	Murine, COS-7 cell line.	1
CALM	Q13492	Adaptor protein. Involved in endocytosis of EGF and transferrin.	PI(4,5)P2, CHC, AP2 α, EGF, transferrin.	Murine, COS-7, HeLa cell lines.	1, 25.
HIP1	O00291	Links endocytic machinery with actin cytoskeleton.	PI(4,5)P2, CHC, AP2 α, HIP1R.	Human, budding yeast.	6, 26, 27.
HIP1R	O75146	Links endocytic machinery with actin cytoskeleton.	CHC, filamentous actin (F-actin), HIP1.	Mammalian, COS-7, HeLa cell lines	28, 29.
Epsin 1	Q9Y6I3	Adaptor protein. Involved in endocytosis of ubiquitinated cargo. Induces membrane curvature.	PI(4,5)P2, CHC, AP2 α, EPS15, intersectin 1, ubiquinated EGF.	Murine, budding yeast.	30-33, 7, 15.
Epsin 2	O95208	Adaptor protein. Involved in endocytosis of ubiquitinated cargo.	PI(4,5)P2, CHC, AP2 α, EPS15.	Murine, human	30, 34.
Epsin 3	Q9H201	Putative adaptor protein.	PI(4,5)P2.	Murine.	30

Amphiphysin 1	P49418	Induces membrane deformation and curvature. Involved in actin polymerisation.	Plasma membrane, CHC, AP2 α, dynamin 1, synaptojanin 1, N-WASP.	Murine, budding yeast, COS-7, Ser-W3 cell lines.	35-40.
Amphiphysin 2	O00499	Induces membrane deformation and curvature.	Plasma membrane, CHC, dynamin 1.	Murine.	41-43.
Endophilin	Q99962	Induces membrane deformation and curvature.	Plasma membrane, dynamin, synaptojanin.	Murine, nematode.	44, 45.
TOCA-1	Q5T0N5	Induces membrane deformation and curvature. Involved in actin polymerisation.	PI(4,5)P2 ,Plasma membrane, N-WASP, dynamin.	Murine, human.	46, 47.
FBP-17	Q96RU3	Induces membrane deformation and curvature. Involved in actin polymerisation.	PI(4,5)P2 ,Plasma membrane, N-WASP, dynamin 1-3.	Human, amphibian.	48
CIP4	Q15642	Induces membrane deformation and curvature. Involved in actin polymerisation.	PI(4,5)P2 ,Plasma membrane, N-WASP, dynamin 2.	Murine, L6 GLUT4myc	49
Tuba	Q6XZF7	Links membrane bending to actin polymerisation and dynamin activiy.	Dynamin 1, N-WASP.	Murine, human.	50
SNX9	Q9Y5X1	Links membrane bending to actin polymerisation and dynamin activity.	CHC, AP2 α, dynamin 1, synaptojanin 1, N-WASP.	Murine, K562 cell line.	9, 51.
Dynamin 1	Q05193	Mediates vesicle scission.	Amphiphysin 1- 2, TOCA-1, FBP-17, tuba, SNX9, intersectin 1, ABP1, auxilin.	Murine, human, COS-7 cell line.	38, 44, 46-48, 50-52.
Dynamin 2	P50570	Mediates vesicle scission.	FBP-17, CIP4, ABP1.	Murine, human.	53, 54, 48.
Dynamin-3	Q9UQ16	Mediates vesicle scission.	FBP-17.	Human.	48

EPS15	P42566	Coordinates adaptor proteins.	AP2 α, numb, AP180, CALM, epsin 1.	Murine, yeast, COS-7 cell line.	15, 32, 1, 33.
EPS15R	Q9UBC2	Coordinates adaptor proteins.	AP2 α, numb.	Human.	22, 55.
Intersectin 1	Q15811	Scaffolding. Links vesicle scission and uncoating functions.	Dynamin 1, synaptojanin 1, epsin 1, N-WASP.	Murine, human, Xenopus.	52, 56, 57.
Intersectin 2	Q9NZM3	Scaffolding. Links vesicle scission and uncoating functions.	N-WASP.	Human.	58.
ABP1	Q9UJU6	Coordinates cortical actin patches.	F-actin, dynamin 1 & 2, amphiphysin, N-WASP.	Murine, budding yeast.	59, 53, 60, 61.
N-WASP	O00401	Induces actin polymerisation.	Amphiphysin 1, endophilin, TOCA1, FBP-17, CIP4, tuba, SNX9, intersectin 1-2, ABP1.	Human, murine, amphibian.	47-51, 56.
Auxilin	O75061	Recruits HSC70 to clathrin lattice	CHC, AP2 α, dynamin 1, HSC70.	Bovine, murine.	11, 63-65.
Synaptojanin 1	O43426	Hydrolises phosphoinositides inducing uncoating process.	Amphiphysin 1, endophilin, tuba, SNX9, intersectin 1.	Murine, COS-7.	47, 51, 39, 45.
Synaptojanin 2	O15056	Hydrolyses phosphoinositides inducing uncoating process.	PI(4,2)P2.	A549 cell line.	62
HSC70	P11142	Uncoats clathrin lattice in ATP dependant way.	CHC, auxilin.	Bovine.	63

Key to literature references:

1. (Ford *et al.*, 2001); 2. (Shih *et al.*, 1995); 3. (Laporte *et al.*, 1999); 4. (Mishra *et al.*, 2002b); 5.(Mishra *et al.*, 2002a); 6. (Mishra *et al.*, 2001); 7. (Drake *et al.*, 2000); 8. (Miele *et al.*, 2004); 9. (Lundmark & Carlsson, 2003); 10. (Blondeau *et al.*, 2004); 11. (Eisenberg & Greene, 2007); 12. (Newpher *et al.*, 2006); 13. (Chen *et al.*, 2002); 14. (Santolini *et al.*, 2000); 15. (Traub *et al.*,

1999); 16. (Scheele *et al.*, 2003); 17. (Collins *et al.*, 2002); 18. (Owen & Evans, 1998); 19. (Krupnick *et al.*, 1997); 20. (Goodman *et al.*, 1996); 21. (Gaidarov *et al.*, 1999); 22. (Salcini *et al.*, 1997); 23. (Dho *et al.*, 2006); 24. (Frise *et al.*, 1996); 25. (Tebar *et al.*, 1999); 26. (Engqvist-Goldstein *et al.*, 1999); 27. (Senetar *et al.*, 2004); 28. (Engqvist-Goldstein *et al.*, 2001); 29. (Engqvist-Goldstein *et al.*, 2004); 30. (Itoh *et al.*, 2001); 31. (Chen *et al.*, 1998); 32. (Kazazic *et al.*, 2009); 33. (Maldonado-Baez *et al.*, 2008); 34. (Rosenthal *et al.*, 1999); 35. (Peter *et al.*, 2004); 36. (Slepnev *et al.*, 2000); 37. (David *et al.*, 1996); 38. (Grabs *et al.*, 1997); 39. (McPherson *et al.*, 1996); 40. (Yamada *et al.*, 2009); 41. (Casal *et al.*, 2006); 42. (Ramjaun & McPherson, 1998); 43. (Wigge *et al.*, 1997); 44. (Ringstad *et al.*, 1997); 45. (Song & Zinsmaier, 2003); 46. (Ho *et al.*, 2004); 47. (Itoh *et al.*, 2005); 48. (Kamioka *et al.*, 2004); 49. (Hartig *et al.*, 2009); 50. (Salazar *et al.*, 2003); 51 (Shin *et al.*, 2007); 52. (Yamabhai *et al.*, 1998); 53. (Kessels *et al.*, 2001); 54. (Krendel *et al.*, 2007); 55. (Iannolo *et al.*, 1997); 56. (Hussain *et al.*, 2001); 57. (Sengar *et al.*, 1999); 58. (McGavin *et al.*, 2001); 59. (Goode *et al.*, 2001); 60. (Colwill *et al.*, 1999); 61. (Pinyol *et al.*, 2007); 62. (Rusk *et al.*, 2003); 63. (Heymann *et al.*, 2005); 64. (Newmyer *et al.*, 2003); 65. (Scheele *et al.*, 2001).

3.3.1 Description of the CME-I network

Here I briefly review the protein components of the CME-I network in relation to their function within the interactome. To manage the complexity of the CME system, I divided it into five functional modules: (i) core, (ii) membrane bending, (iii) vesicle scission, (iv) actin attachment, and (v) vesicle uncoating.

3.3.2 The core module of the CME-I network

The core module is composed of clathrin (heavy and light chains), the AP2 complex, epsins, AP180, CALM, EPS15 and the following monomeric adaptors: β arrestins, disabled 2, numb and ARH. These proteins mediate the formation of the clathrin pit and the binding of cargo (Ford *et al.*, 2001; Gaidarov & Keen, 1999; Traub, 2003; B. Wendland & Emr, 1998). AP2 is an heterotetrameric complex made of two large

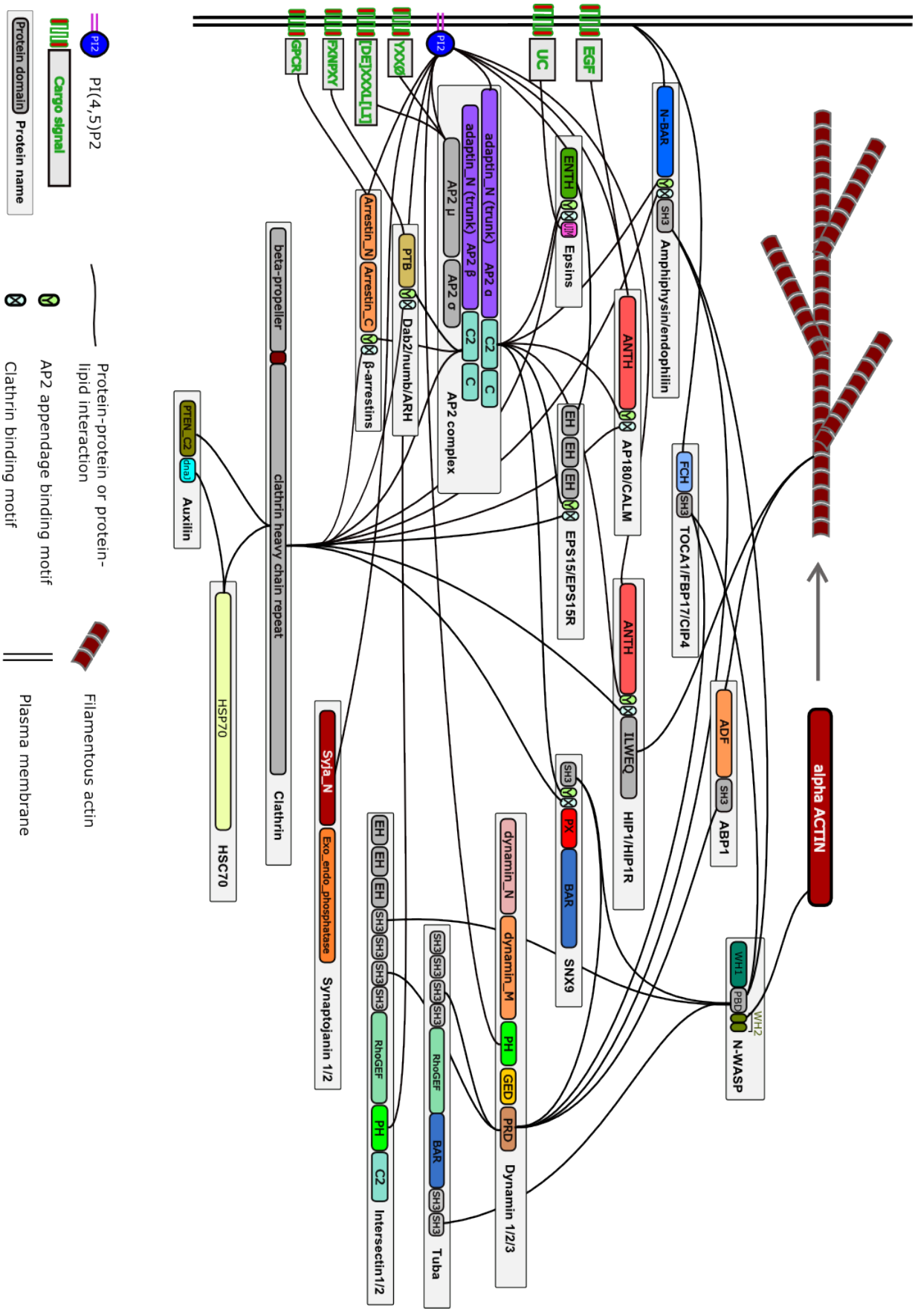

alpha ACTIN
N-BAR
SH3
Amphiphysin/endophilin
FCH
SH3
TOCA1/FBP17/CIP4
ADF
SH3
ABP1
WH1
PBD
WH2
N-WASP
EGF
UC
GPCR
PXNPXY
[DE]XXXL[LI]
YXXØ
PI2
ENTH
ANTH
Epsins
AP180/CALM
EH
EH
EH
EPS15/EPS15R
ANTH
ILWEQ
HIP1/HIP1R
SH3
PX
BAR
SNX9
adaptin_N (trunk)
AP2 α
adaptin_N (trunk)
AP2 β
AP2 μ
AP2 σ
C2
C
C2
C
AP2 complex
PTB
Dab2/numb/ARH
Arrestin_N
Arrestin_C
β-arrestins
beta-propeller
clathrin heavy chain repeat
Clathrin
PTEN_C2
fhaA
Auxilin
HSP70
HSC70
Syja_N
Exo_endo_phosphatase
Synaptojanin 1/2
dynamin_N
dynamin_M
PH
GED
PRD
Dynamin 1/2/3
EH
EH
EH
SH3
SH3
SH3
SH3
RhoGEF
PH
C2
Intersectin1/2
SH3
SH3
SH3
SH3
SH3
RhoGEF
BAR
SH3
SH3
Tuba
PI2
PI(4,5)P2
Cargo signal
Protein domain
Protein name
Protein-protein or protein-lipid interaction
AP2 appendage binding motif
Clathrin binding motif
Filamentous actin
Plasma membrane

Figure 3.2 Connectivity diagram depicting the CME-I network. Data from 65 published studies (Table 3.1) is used to outlined the CME-I network. Protein-protein and protein-lipid interaction have been demonstrated by co-immunoprecipitation, structural crystallography and two-hybrid screening. Protein domains are shown in coloured bars while protein names are in grey boxes. The X and the Y bars represent respectively clathrin binding motifs and AP2 appendage binding motifs. With the exception of UIM, protein domains are in scale. The rest of the diagram is not in scale. Internalisation signals are green in grey boxes (EGF = epidermis growth factor; UC = ubiquitinated cargo; GPCR = G-protein coupled receptor; YXXØ, [DE]XXXL[LI] and FXNPXY represent consensus internalisation signal motifs). Blue elipse labeled PI2 represents phosphatidylinositol 4,5-biphosphate. Protein-protein and protein-lipid binding interactions are shown with black lines.

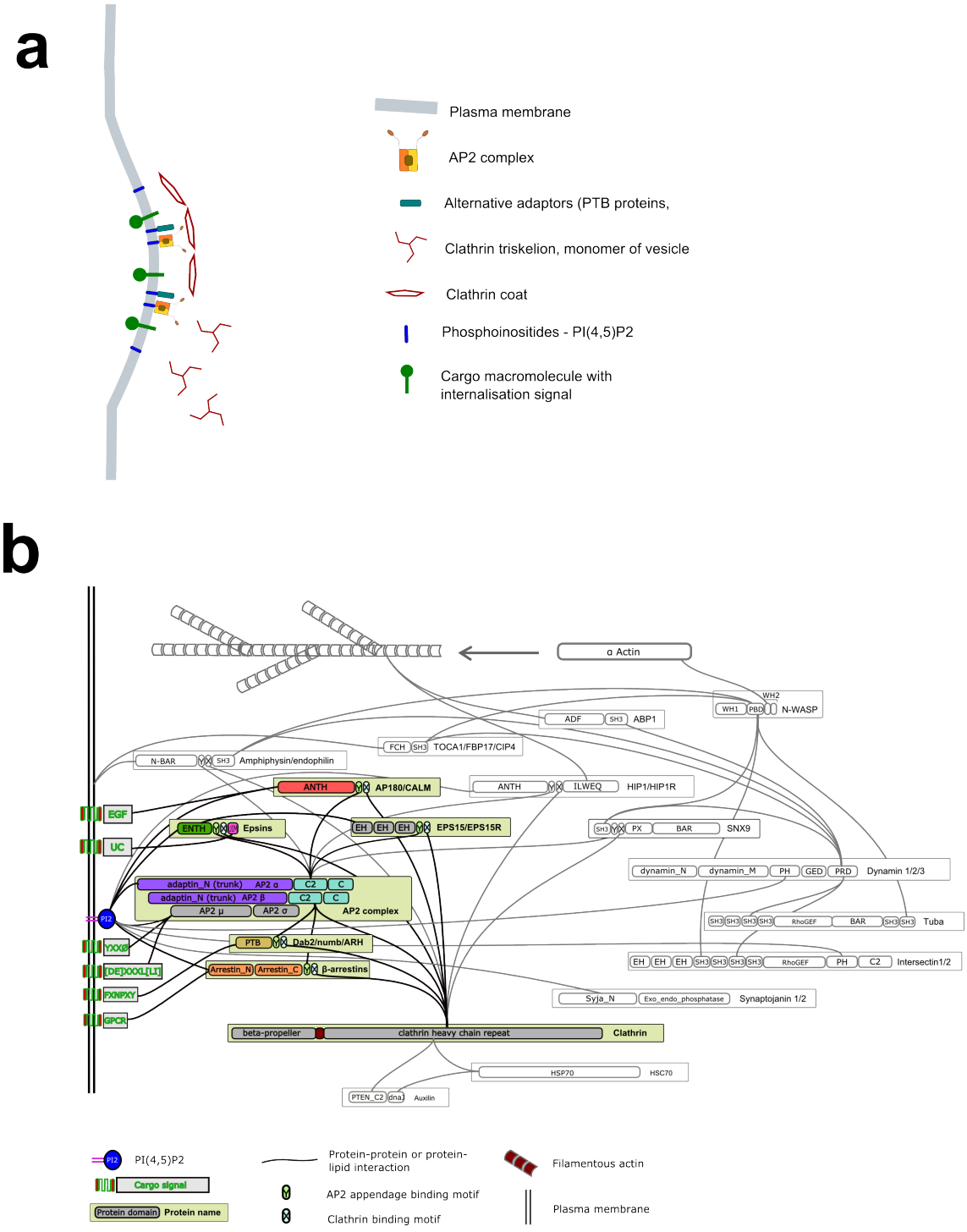
subunits (α and β2), one medium subunit (μ) and one small subunit (σ). The AP2 complex binds plasma membrane component phosphatidylinositol 4,5-biphosphate [PI(4,5)P2] via the large α subunit (Gaidarov & Keen, 1999), while recruiting clathrin at the site of pit formation via the β2 subunit (Traub *et al*., 1996). In addition, the μ subunit binds to cargo internalisation motifs such as YXXØ and [DE]XXXL[LI] (Figure 3.3). EPS15 binds to the AP2 complex, epsins and AP180/CALM (Maldonado-Baez *et al*., 2008; Tebar *et al*., 1996; Traub *et al*., 1999) and it is believed to provide stability to the initial clathrin-AP2 complex (Schmid *et al*., 2006).

Monomeric adaptors bind to clathrin heavy chain, PI(4,5)P2 and the AP2 complex (Traub, 2003). In addition, they engage with endocytic receptors unrecognised by the AP2 complex. For instance, β arrestins 1 and 2 engage with active, phosphorylated cargo G protein-coupled receptor (Laporte *et al*., 1999). Meanwhile, ARH, disabled 2 and numb engage with low density lipoprotein receptors such as VLDL, apoER2 and megalin via an N-terminal phosphotyrosine-binding (PTB) domain (Traub, 2003).

Epsins and the AP180 and CALM paralogues, belong to the a protein superfamily which is characterised by an N-terminal, PI(4,5)P2 binding, protein domain known as the ENTH/ANTH domain (Legendre-Guillemin *et al*., 2004). Similarly to monomeric adaptors, they also bind clathrin, the AP2 complex and specific cargo internalisation signals (e.g. transferrin and ubiquinated cargo) (Ford *et al*., 2001). However, in addition to the adaptor-like functions, epsins are also proposed to create membrane curvature via an amphipathic α helix - formed upon epsin's binding to PI(4,5)P2 - which inserts into one leaflet of the membrane's lipid bilayer thereby causing it to bend (Ford *et al*., 2002; Horvath *et al*., 2007) (Figure 3.3).

3.3.3 The membrane bending module

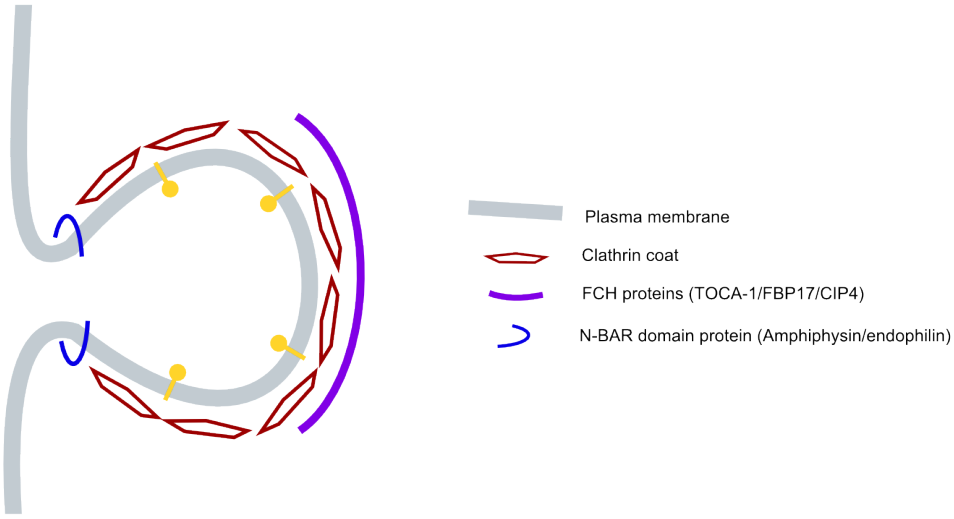
The membrane bending module consists of the BAR domain protein family which are divided into N-BAR and FCH sub-families. The two N-BAR proteins commonly studied are mammalian amphiphysin 1-2 and endophilin (Rvs167 and Rvs161 in *Saccharomyces cerevisiae*) which both have a N-terminal BAR domain and a C-terminal SH3 domain (Dawson *et al*., 2006). They form crescent shaped dimers that can sense, induce and maintain create curvature of the membrane. They do this via two known mechanisms: by binding to negatively charged phospholipids using electrostatic forces (Peter *et al*., 2004) and via an amphipathic α helix which is inserted in one leaflet of the lipid bilayer causing displacement of the phospholipids (Dawson *et al*., 2006; Masuda *et al*., 2006). FCH proteins are structurally related although the curvature they create is shallower by comparison (Shimada *et al*., 2007). It is proposed that FCH proteins cause the deepening of the early clathrin coated pit by binding to the



membrane, creating end-to-end oligomerisation of protein dimers which surround the

Figure 3.3 Core module of the CME-I. a. The AP2 complex and alternative monomeric adaptors simultaneously bind PI(4,5)2, cargo internalisation signal, and clathrin, inducing formation of clathrin coated plasma membrane pits. **b.** Connectivity diagram highlighting CME-I network components involved in core module. Network components involved in core module are in colour, non-core module components are in grey

a



b

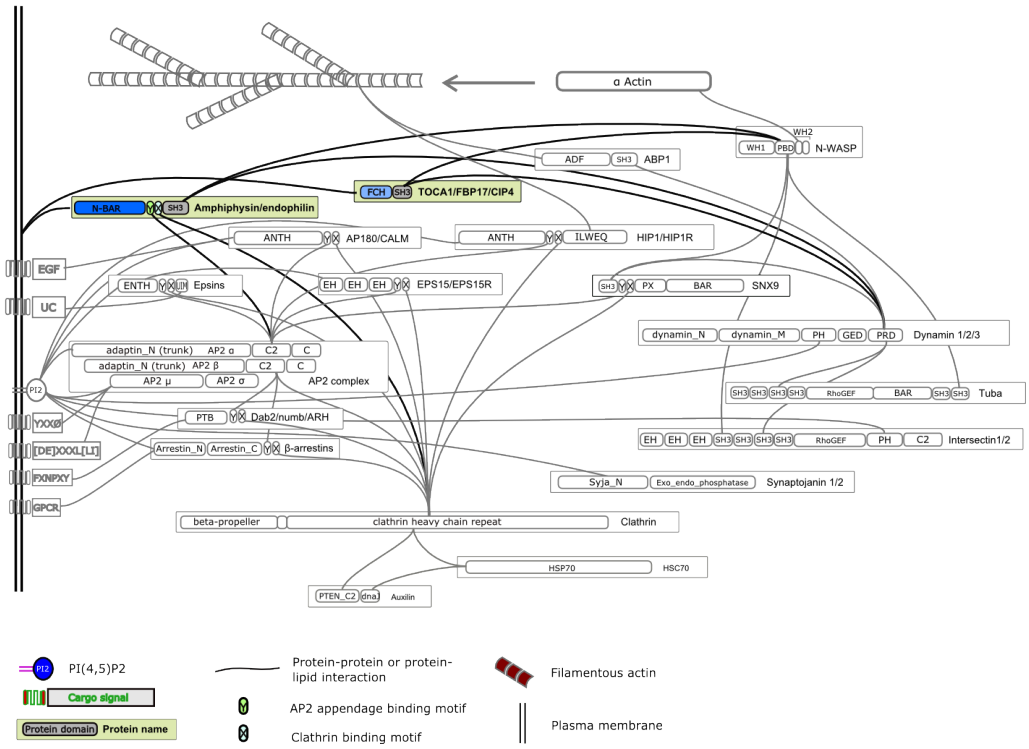


Figure 3.4 Membrane bending module of the CME-I. a. N-BAR and FCH proteins mediate membrane bending in CME **b.** Connectivity diagram highlighting CME-I network components involved in the membrane bending module. Network components involved in membrane bending module are in colour, other components are in grey.

budding vesicle (Futterer & Machesky, 2007) (Figure 3.4).

3.3.4 The vesicle scission module

The vesicle scission module is composed of a sub-group of the dynamin protein family, known in mammals as 'classical dynamins' (Praefcke & McMahon, 2004). Dynamin is a large and modular GTPase featuring four main functional domains (Urrutia *et al*., 1997). A large N-terminal domain of more than 300 amino acids features GTP binding motifs necessary for guanine nucleotide binding and hydrolysis. At the C-terminal we find, sequentially, a PI(4,5)P2 binding pleckstrin homology (PH) domain, a GTPase effector domain (GED) which activates dynamin self assembly and GTPase activity, and a proline rich domain (PRD), a site of interaction with other endocytic effectors (Urrutia *et al*., 1997). Dynamin is thought to self assemble into a ring or helical structure around the deeply invaginated neck of a nascent vesicle, and cause its fission by constricting and severing invaginated pits following GTP hydrolysis-driven conformational changes (Hinshaw & Schmid, 1995; Marks *et al*., 2001) (Figure 3.5).

3.3.5 The actin attachment module

Interaction of the actin cytoskeleton with the endocytic effectors is a key feature in CME (Qualmann *et al*., 2000). Actin filaments provide support for the early stages of vesicle formation, then polymerise at the site where the vesicle's neck narrows, and finally drive the inward movement of the vesicle following its dynamin mediated scission (Lamaze *et al*., 1997) (Figure 3.6). The components of the actin attachment module link the endocytic machinery to the actin cytoskeleton either directly, by binding to actin filaments, or indirectly, by inducing actin polymerisation via a

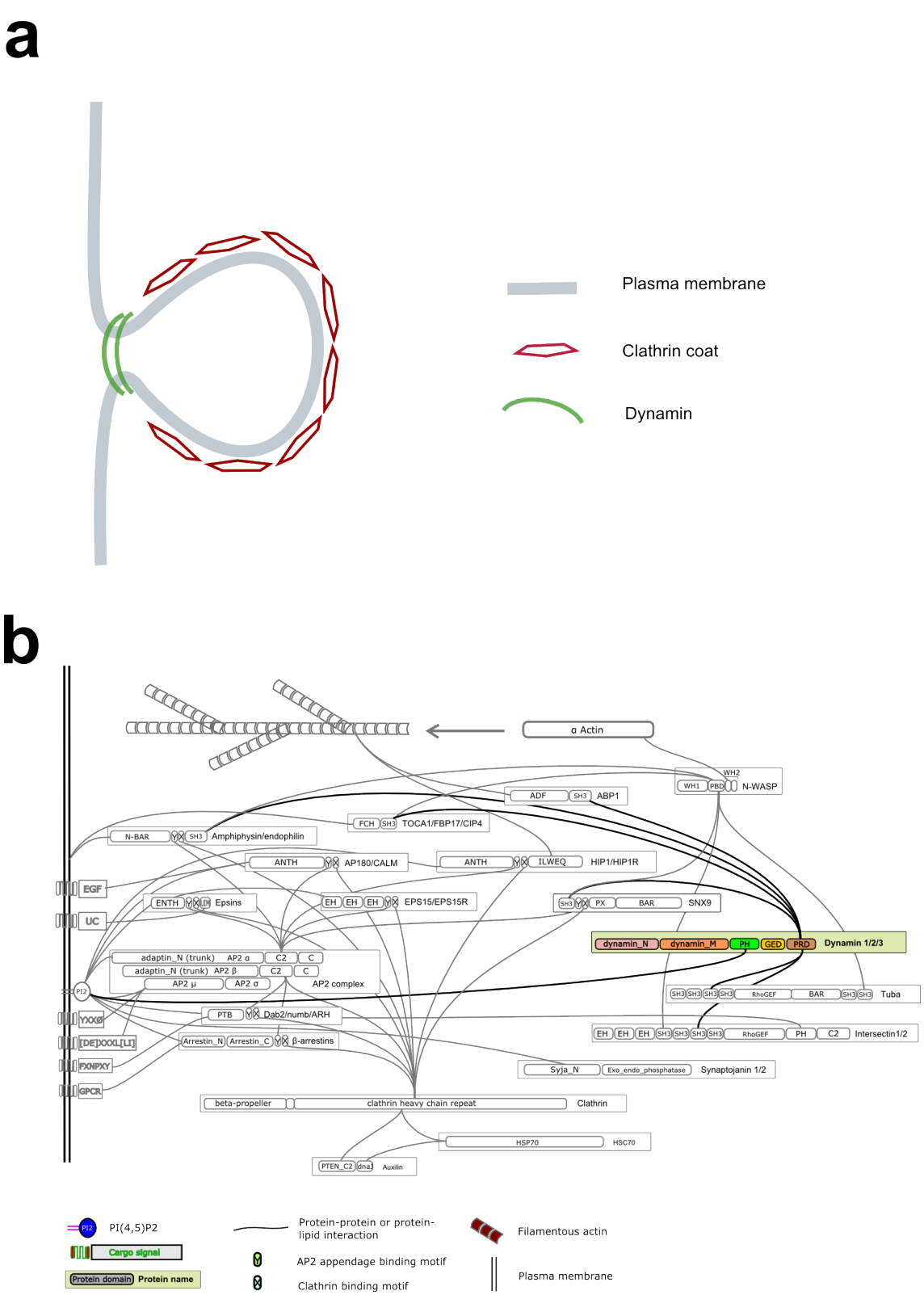


Figure 3.5 Vesicle scission module of the CME-I. a. Dynamin molecules assemble into a helical structure around the vesicle neck and causes scission by conformational changes. **b.** Connectivity diagram highlighting dynamins and its interactions with other CME-I network components.

signalling route. This module includes the HIP1 and HIP1R paralogues, ABP1, tuba, intersectins, SNX9 and N-WASP.

HIP1 has typical adaptor characteristics in that they bind to PI(4,5)P2, clathrin and the AP2 complex (Engqvist-Goldstein *et al*., 1999). Importantly, it can form heterodimers with HIP1R which in turn binds to filamentous actin via a C-terminal I/LWEQ talin domain (Engqvist-Goldstein *et al*., 2004). ABP1 also binds actin filaments directly. It does so via the N-terminal ADF domain which promotes rapid filament turnover (Goode *et al*., 2001). In addition, ABP1 features a C-terminal SH3 domain which has been shown to interact with proline rich domains (PRD) of dynamin (Kessels *et al*., 2001).

By contrast, tuba, intersectins and SNX9 regulate actin polymerisation indirectly, via interaction with N-WASP (Rohatgi *et al*., 1999). The proposed mode of function of N-WASP is to bind to the plasma membrane via a central basic domain and a GTPase binding domain, while two C-terminal Wasp Homology (WH2) domains bind to actin monomers and a Central/Acidic (CA) domain binds and activates the ARP2/3 complex (Svitkina, 2007). Actin filaments are recruited to N-WASP via the ARP2/3 complex, and are attached to the WH2-associated actin monomers (Rohatgi *et al*., 1999). N-WASP detaches to recruit more globular actin, and re-attaches to the barbed end of the actin filament (Co *et al*., 2007). As the cycle repeats, actin filaments are elongated causing propulsion off the plasma membrane. Crucially, N-WASP features a centrally located proline rich domain that engages in interactions with the SH3 domains of other proteins involved in CME. For instance, SNX9 stimulates actin assembly by binding the PRD of

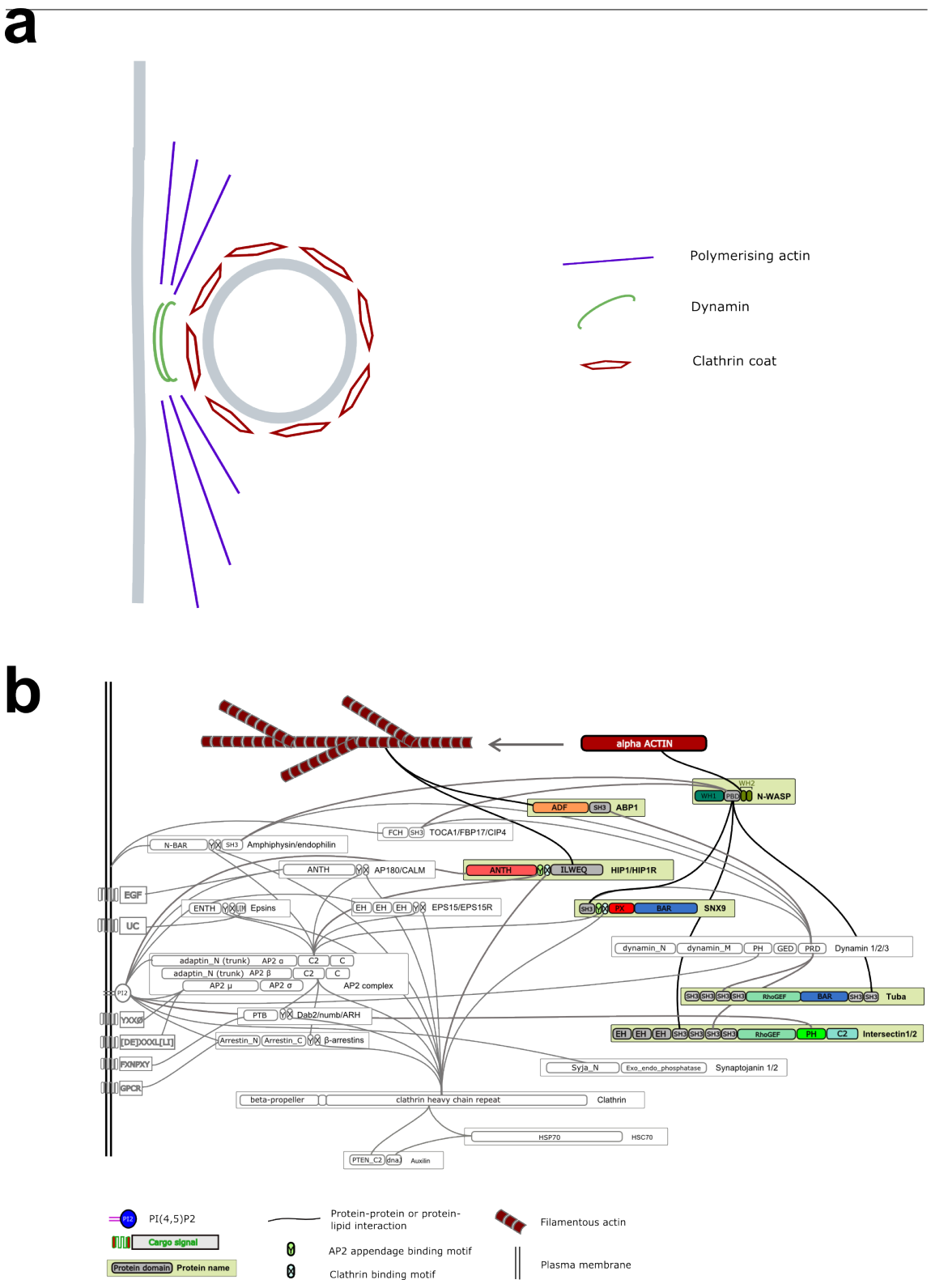


Figure 3.6 Actin attachment module of the CME-I network. a The Actin cytoskeleton drives inward transport of clathrin coated vesicles. **b** Connectivity diagram highlighting CME-I network components involved in actin attachment module. Network components involved in the actin attachment module are in colour, other components are in grey.

N-WASP via its N-terminal SH3 domain (Yarar *et al*., 2007). This links actin modulation to the CME core machinery as SNX9 also binds clathrin and the AP2 complex (Lundmark & Carlsson, 2003). Tuba features a centrally located BAR domain - homologous to the N-BAR domains in amphiphysin and endophilin – downstream of a RhoGEF domain. At the N- and C-terminals there are 4 and 2 SH3 domains, respectively. The SH3 domains at the N-terminal bind to N-WASP while the C-terminal SH3 domains are shown to selectively bind to dynamins (Salazar *et al*., 2003). As with SNX9, the SH3 mediated interaction with N-WASP induces actin assembly (Salazar *et al*., 2003). Intersectins also include multiple SH3 domains that mediate the binding to N-WASP. This upregulates the guanine nucleotide exchange factor (GEF) activity of intersectins, generating GTP-bound CDC42, which in turn activates N-WASP actin nucleation activity (Hussain *et al*., 2001) (Figure 3.6).

3.3.6 The vesicle uncoating module

Once the vesicle has been fully internalised the clathrin coat is removed and its polyhedral structure broken down so that clathrin triskelion monomers can be recycled (Ungewickell & Hinrichsen, 2007). The uncoating module is based on the combined functions of molecular chaperone HSC70, its co-factor auxilin, and inositol phosphatase synaptojanin (Massol *et al*., 2006; Perera *et al*., 2006). Synaptojanin's N-terminal sac1-like inositol phosphatase domain (Syja_N) can hydrolyse PI(4,5)P2 to PI which induces the uncoating process (Cremona *et al*., 1999). HSC70 has a low intrinsic ATPase activity which is stimulated by co-factor auxilin (Higgins & McMahon, 2002). The latter's C-terminal DNAJ (also known as Hsp40) domain recruits HSC70 to the clathrin coat and induces its uncoating activity (Massol *et al*., 2006) (Figure 3.7).

3.4 Discussion

In this chapter I defined the system that mediates CME as the set of protein-protein and protein-lipid interactions known as the CME-I network. As mentioned in the introduction of this chapter, the important novelty of this version of CME-I network is incorporation of information regarding protein domain architecture and binding ability of specific motifs. This is missing in previous versions of clathrin interactome networks, but it's important in order to determine how the CME-I network evolved, from the LCEA to the mammalian and budding yeast cells. The CME-I model presented here will thus be used as the framework for the phylogenomic study reported in Chapter 4, which will not only investigate the evolution of this endocytic process, but the origin and evolution of its distinct functions and sub-parts, and how they have become integrated to form CME 'as we know it'. Another aim of this analysis is to provide an insight into how a complex eukaryotic cellular processes evolve, focusing on the importance of gene and whole genome duplication, gene fusion and protein domain rearrangements in driving the diversification of the endomembrane system and the cytoskeleton, as has been previously reported (Dacks *et al*., 2008; Richards & Cavalier-Smith, 2005; Wickstead *et al*., 2010).

It should be noted that some CME-I network proteins are involved in more than one function. For instance, epsins share functional properties with other monomeric endocytic adaptors in CME as they bind clathrin, the AP2 complex, $PI(4,5)P_2$, and a specific cargo signal (in epsin's case ubiquitinated protein signal) (Drake *et al*., 2000; Itoh *et al*., 2001; Kazazic *et al*., 2009). They are thus classified as components of the

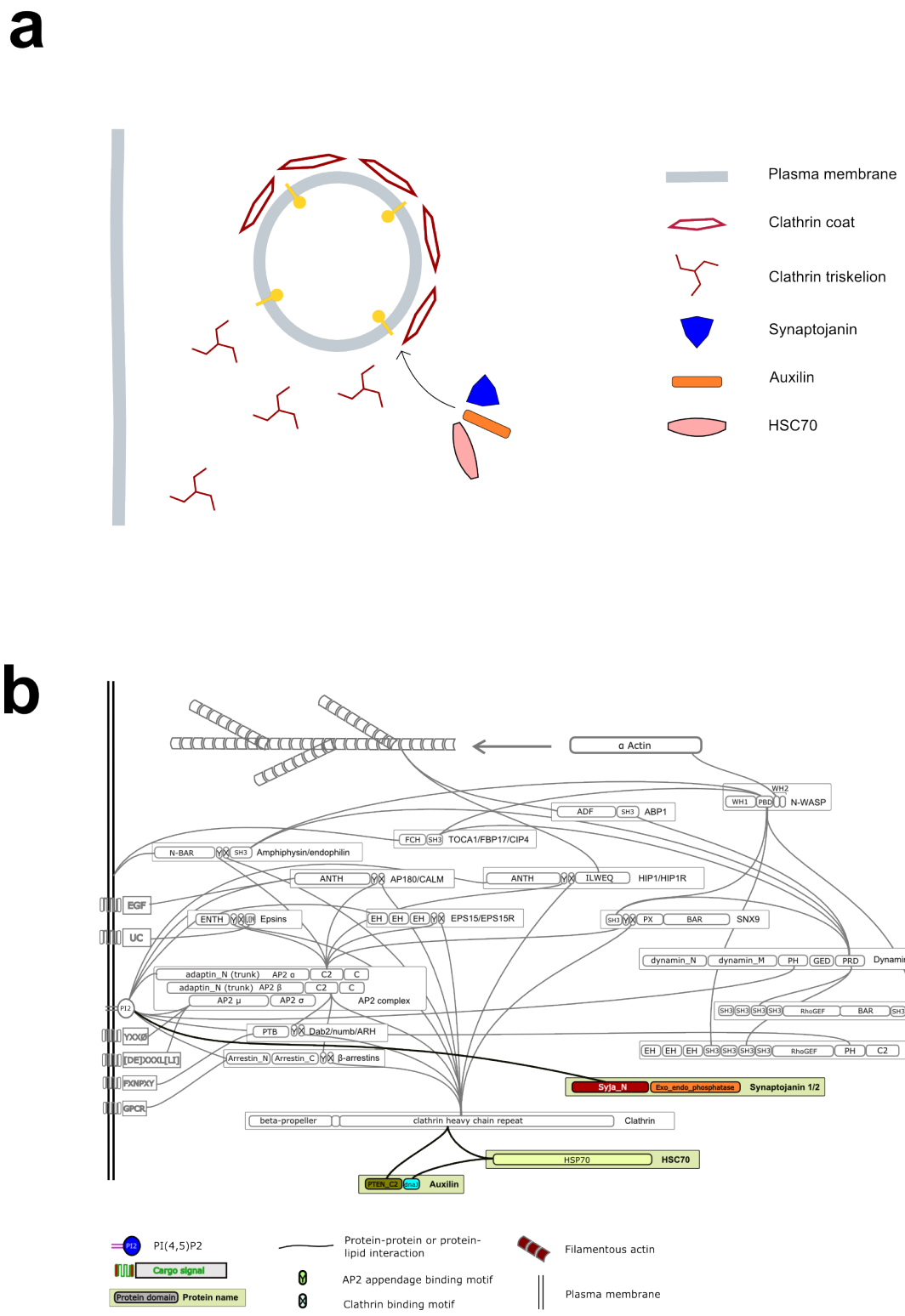


Figure 3.7 Vesicle uncoating module of the CME-I network. a Synaptojanin, auxilin and HSC70 combine to mediate the detachment of the clathrin coat components from the vesicle. Coat components are then recycled back to the plasma membrane. **b** Connectivity diagram highlighting CME-I network components involved in the vesicle uncoating module. Network components involved in the vesicle uncoating module are in colour, other components are in grey.

CME-I core module. However, evidence shows they can also bend membranes with a mechanism that is similar to amphiphysin (Ford *et al*., 2002; Horvath *et al*., 2007). Other proteins such as amphiphysins, endophilin, TOCA-1, FBP17 and CIP4 are characterised as membrane bending proteins because of their N-terminal, crescent shaped BAR domains. However, their C-terminal SH3 domains, bind proline rich domain (PRD) in N-WASP, an interaction shown to induce actin polymerisation (Yamada *et al*., 2009; Ho *et al*., 2004), and the PRD in 'classical dynamins' (Grabs *et al*., 1997) involved in vesicle scission. The functional classification outlined in this chapter should thus not be taken as a rigid and impermeable set of categories. Rather, the main purpose is to deconstruct the complexity of the CME-I network into manageable sub-parts.

Another issue is the specificity of CME-I proteins to the CME pathway. As mentioned in the general introduction, proteins such as dynamin, SNX9, N-WASP and amphiphysins are involved in other distinct endocytic pathways. In addition, some proteins, or at least paralogues of the proteins, play roles in completely separate cellular functions (Table 3.2 summarises involvement of these proteins in non-CME endocytic pathways and in other cellular functions). Thus when considering the results of comparative genomics and phylogenetic analyses, it should be noted that the presence of a specific protein does not necessarily mean it is involved in CME. Bioinformatics offers an important but nonetheless limited predictive tool, and involvement in CME can only be conclusively confirmed by cell biological data.

A further limitation of the CME-I defined here is that it may not be complete. The way proteins were searched was by association with clathrin, the AP2 complex, dynamin and

Table 3.2 List of CME proteins which also play roles in non-CME endocytic pathways and/or other cellular functions.

Protein name	**Function in non-CME endocytic pathways**	**Other cellular functions**
Clathrin	____	Vesicle trafficking between Golgi and endosomes (Duwel & Ungewickell, 2006). Mitosis - stabilises fibres of the mitotic spindle (Royle *et al*., 2005).
AP2 complex	Involved in post-endocytic trafficking in ARF6 regulated clathrin-independent endocytosis (CIE) (Lau & Chou, 2008).	____
ARH	____	Centrosome assembly and cytokinesis (Lehtonen *et al*., 2008)
Numb	____	Cadherin-based cell adhesion (Rasin *et al.*, 2007)
Epsin	Couples ubiquinated EGF receptor to caveolin-mediated endocytosis (Sigismund *et al*., 2005)	____
Amphiphysin	Important for actin polymerisation during phagocytosis (Yamada *et al*., 2007)	____
FBP17	Induces membrane tubulation and interacts with dynamin in caveolin-mediated endocytosis (Kamioka *et al*., 2004). In macrophages, recruits WASP proteins and dynamin to the plasma membrane which is necessary for phagocytic cup formation. (Tsuboi *et al*., 2009).	____
SNX9	Involved in dorsal ruffle formation in Macropinocytosis and coordination of actin dynamics in cdc42 mediated CIE pathway (Yarar *et al*., 2007)	____
Dynamin	Involved in vesicle scission in caveolin-mediated endocytosis (Pelkmans *et al*., 2002). Involved in RhoA regulated CIE (Sauvonnet *et al*., 2005). Dynamin-2 also plays a role in phagocytosis in macrophages (Gold *et al*., 1999)	Dynamin-related proteins are involved in division of organelles, cytokinesis and microbial pathogen resistance (Miyagishima *et al*., 2003; H. M. Thompson *et al*., 2002).
EPS15	Couples ubiquinated EGF receptor to caveolin-mediated endocytosis (Sigismund *et al*., 2005).	____

ABP1	Localises to phagocytic cups, and interacts with myosin IK to regulate phagocytosis efficiency (Dieckmann *et al*., 2010).	
N-WASP	Activated by cdc42 in Fc gamma receptor-mediated phagocytosis (Park & Cox, 2009).	____

some other known key endocytic proteins. The model of the CME-I network presented here may thus be biased towards proteins that interact with clathrin and the AP2 complex. In addition, proteins involved in CME may have been characterised after thisstudy was carried out, or considered as potentially involved in CME but for paucity of evidence. For instance, in a previous analysis of the CME-I, proteins such as tom1, NECAP-1 and stonin2 were proposed as potential endocytic adaptors, (see supplementary material in (Schmid & McMahon, 2007), but not enough evidence was available for inclusion in the CME-I network at the time when this project started. The CME-I network, and associated analyses, should therefore be regularly updated.

4 Evolutionary history of the CME-I network

4.1 Introduction

The molecular requirements of CME are mediated and coordinated by a complex set of protein-protein and lipid-protein interactions known as the CME interactome (CME-I) network (Schmid & McMahon, 2007). A total of 39 proteins, belonging to 21 gene families, were included in the known, functionally modulated CME-I network described in Chapter 3 (Table 3.1). In this chapter I present a phylogenomic study of the CME-I network.

As mentioned in Chapter 3, an important aim of this study is to investigate the evolutionary history of the complex molecular machinery required for viable CME in budding yeast and mammalian model organisms. This is particularly significant because this molecular machinery involves multiple modes of actin cytoskeleton regulation, and reconstructing the evolution of interactions between membrane deforming endocytic proteins with the actin cytoskeleton may provide clues concerning the evolution of the eukaryotic endomembrane system. The study will thus focus on patterns of modification and expansion in the evolution of the CME-I network. By identifying synapomorphies in the form of novel protein families, novel protein domain architectures and novel interactions between protein families, putative CME-I networks on different branches of the eukaryotic tree can be reconstructed and compared, thereby highlighting key events in the evolutionary history of CME.

This study also aims to add further detail to the model of CME in the LCEA. Previously

published, comparative genomic, phylogenetic and cell biological studies focusing on CME proteins such as clathrin, the AP2 complex, epsins, EPS15, AP180 and dynamins, suggests that a significant part of the CME-I network was present in the LCEA (Dacks *et al*., 2008; Elde *et al*., 2005; Field *et al*., 2007; Gabernet-Castello *et al*., 2009; Miyagishima *et al*., 2008; Schledzewski *et al*., 1999). However, the evolutionary histories of these protein families have been studied individually, without context of a protein interaction network. The data reported in this chapter will indicate if other proteins or protein domains which characterise the CME-I were likely to be present in the LCEA and will thus predict the organisation of the minimal set CME-I proteins in the ancestral CME-I network.

Finally, the study aims to provide an estimation of the diversity of the CME pathway in a sample of distantly related, eukaryotic lineages. Cell biological studies of CME proteins are biased to opisthokont taxa. The main exceptions to this bias are studies on kinetoplastid protist *Trypanosoma*, where an advanced model of the functions and mechanism of CME is emerging (Allen *et al.*, 2003; Chanez *et al*., 2006; Correa *et al*., 2007; Gabernet-Castello *et al*., 2009; Hung *et al*., 2004; Morgan *et al*., 2001; Morgan *et al*., 2004). There also notable studies of CME carried out on the ciliates *Paramecium* and *Tetrahymena* (Elde *et al*., 2005; Ramoino *et al*., 2006; Wiejak *et al*., 2004), on apicomplexan *Toxoplasma gondii* (Nichols *et al*., 1994) and the metamonad *Giardia* (Hernandez *et al*., 2007; Rivero *et al*., 2010). However, data from these studies only cover the role of clathrin, the AP2 complex, dynamins and epsins. Results presented in this chapter will be used to assess how conserved the rest of the CME-I network is in these diverse, non-opisthokont eukaryotic lineages as well as other protists such as

metamonad *Trichomonas,* Discicristata *Naegleria*, heterokonts *Thalassiosira* and *Phytophthora* and haptophyte *Emiliania*. This will provide a broad map of CME diversity across eukaryotes.

The experimental outline of this study consists in determining the taxonomic distribution of CME-I network components within a broad sample of diverse eukaryotes. Because paralogues often encode proteins with non-identical function (Koonin, 2005), the evolutionary relationships between network component and putative homologues, i.e. relation by vertical descent (orthology), or relation by gene duplication (paralogy), are qualified via phylogenetic analysis. To investigate the evolution of the CME functional repertoire, a protein domain analysis of CME-I network proteins and respective homologues is also carried out. The protein domain architectures of candidate homologues are then mapped to the phylogenies of the CME-I network protein families included in the study. Taken together, these data are used to reconstruct the evolution of the CME-I network.

4.2 Materials and methods

4.2.1 Identification of candidate CME-I protein homologues

The amino acid sequences of the CME-I network proteins were downloaded from the *Homo sapiens* predicted proteome on SWISS-PROT (M. Schneider *et al*., 2009). Using CME-I network proteins as seeds, BLASTp and tBLASTn searches were performed against a selection of diverse eukaryotic genome as described in sections 2.1.1-2 of the Materials and Methods chapter. A gathering threshold of < 1 x e-5 was applied to the sequence similarity results. Sequences with significant sequence similarity were tested

by reciprocal BLASTp searches. The protein domain architecture of putative homologues was predicted by searching the PFAM and CDD databases as described in section 2.1.3. Sequences that featured at least partially identical protein domain structures to the query CME-I network protein were retained for further analyses.

To ensure that the maximum number of candidate homologues were retrieved, for each database search, the sequence with the highest E-value among those found to be significantly similar in the first BLASTp or tBLASTn search, was selected as seed to perform a further BLASTp or tBLASTn search against the same database. To exclude false negative results in the database searches two methods were used. Firstly, BLASTp and tBLASTn searches were performed against databases producing negative results, using a candidate homologue from the most closely related organism sampled in the study, as seed (For instance if a search for query sequence A indicates absence in *Entamoeba histolytica,* but presence in *Dictyostelium discoideum,* the A candidate homologue in *Dictyostelium discoideum* is selected as query sequence for a new search against the *Entamoeba histolytica* database). Secondly, a selection of predicted proteomes - one for each of the ten main eukaryotic groups sampled in this study - were searched using the PSI-BLAST programme (Altschul *et al*., 1997) and a given CME-I network protein as seed, thus creating a position specific substitution matrix (PSSM). The PSSM was downloaded and used to perform PSI-BLAST searches against databases that produced negative results in the BLASTp and tBLASTn searches. Putative homologues found using the two methods described were analysed with PFAM and CDD and considered for further analysis in case of full or partial domain identity.

4.2.2 Phylogenetic analyses of CME-I network protein families

Data sets were prepared for phylogenetic analyses by carrying out multiple sequence alignments (MSA), alignment masking, and substitution model selection, using the methods described in section 2.1.4 of the Materials and Methods chapter. A total of 21 data sets for 20 protein families (the AP2α adaptin subunits and the AP2β adaptin subunits were analysed separately) were created. For each data-set, fast amino acid maximum likelihood and Bayesian phylogenetic analyses were performed using RAXML and MRBAYES respectively, as described in section 2.1.5. of Chapter 2. For all analyses, the best-fitting RtRev amino acid substitution matrix was used. For RAXML analyses, the hard-coded Γ distribution setting with 4 rate categories was used. For MRBAYES analyse, a Γ distribution setting of 8 rate categories was used. RAXML bootstrap analyses were used to assess topological support for the phylogenies. For data sets with more than 100 sequences, 1,000 bootstrap replicate analyses were carried out, whereas for data sets containing less than 100 sequences, 100 bootstrap replicate analyses were carried out. This is because when using RAXML to do a combined ML tree search and bootstrap analysis, every fifth tree produced by the bootstrap replicates is used as starting tree for the ML topology search (Stamatakis, 2006). A higher number of bootstraps will thus improve the final tree in very large data-sets. For all MRBAYES analyses, a minimum of 1'000'000 MCMC generations were performed, except for data-sets with more than 100 sequences where 2'500'000 MCMC generations were performed. In all runs, trees were sampled every 100 generations. To determine posterior probabilities for bipartitions, trees with sub-optimal likelihood values were removed using the 'burnin' function (the number of trees removed was graphically determined by plotting the likelihood values), and the remaining tree sample was

summarised using the 'sumt' function. In this study, orthology and inparalogy between the query protein and a putative homologue is suggested only when maximum likelihood and Bayesian phylogenetic analysis supports monophyly of the query protein and putative homologue with support values equal or higher than 80/0.90.

4.2.3 Mapping CME-I network evolution to the eukaryotic tree of life

For each sequence included in the 21 data sets prepared for phylogenetic analyses, the protein domain architecture was identified using PFAM (Finn *et al*., 2010) and mapped to the phylogenetic trees obtained in this study. The phylogenetic distribution of protein domain architectures, was used to pinpoint protein domains acquisition or paralogue-specific protein domain rearrangements. According to principles of evolution based on Dollo's law (Le Quesne, 1974), the acquisition of a complex character (in this case a protein domain architecture) is almost certainly a unique event, whereas a secondary reversal to a state of absence is by comparison far more probable. For protein domain architectures that distributed to a monophyletic group in the phylogenetic trees, a single origin was thus pinpointed to the last common ancestor of the taxa in the monophyletic group, even if it entailed secondary loss in a minority of the sequences in the monophyletic group. The taxonomic distribution of protein domain architectures in 15 strongly supported, monophyletic eukaryotic groups was thus determined. Using the Dollo based principle, this data was in turn used to map acquisition of protein domains and protein domain rearrangements on the model of the eukaryotic tree described in section 1.2.2.

Protein domain architectures present in few and distantly related taxa, and protein

domain architectures that were mapped to phylogenetic trees with very poor resolution, were excluded from the eukaryotic map as further analyses were required to identify a likely origin, but not performed in this study. The likely origin for interactions between two CME components was pinpointed according to the approach described in Section 3.1. This means that the acquisition of an interaction is pinpointed to the point of acquisition of the more derived of the two interacting proteins. The data on the origin of protein domains was combined with the data on origin of protein-protein and protein-lipid interactions to generate a model describing the probable CME-I network at different stages of eukaryotic evolution.

4.3 Results

4.3.1 Taxonomic distribution of CME-I proteins.

An initial sequence similarity search was performed on GenBank non-redundant bacterial and archaeal databases (Sayers *et al*., 2010). Only homologues of HSC70 and the DNAJ domain found at the C-terminal of auxilin proteins (Holstein *et al*., 1996) were found in prokaryotic genome databases. These belong to the heat shock protein family and operate as molecular chaperones in a range of functions outside of CME (Chappell *et al*., 1986; Walsh *et al*., 2004). Together, this confirms that CME is an eukaryotic cellular process. To study the taxonomic distribution of CME-I proteins I carried out similarity searches in 27 diverse eukaryotic predicted proteomes and translated nucleotide databases. For each protein family I collected sequences with significant similarity, aligned them, and conducted fast maximum likelihood and Bayesian phylogenetic analyses. I thus obtained phylogenetic trees of clathrin heavy chains, clathrin light chains, AP α subunits, AP2β subunits, AP2μ subunits, AP2σ

subunits, PTB proteins (ARH, disabled 2 and numb), arrestins, AP180/CALM, HIP1/HIP1R, epsins, amphiphysins/endophilin, FCH proteins (TOCA-1, FBP17 and CIP4), tuba, SNX9, dynamins, EH proteins (EPS15/EPS15R and intersectins), ABP1, N-WASP, auxilins and synaptojanins (Figures 4.3-23).

A characteristic found in all but the AP2μ, AP2σ and AP180/CALM phylogenies is the presence of gene duplications specific to *Homo sapiens* and *Danio rerio*. This finding is consistent with the hypothesis that whole genome duplication events occurred close to the origin of vertebrates, before the radiation of jawed vertebrates (Dehal & Boore, 2005; Kasahara, 2007). The implication of this pattern is that for most of the CME-I proteins, orthologues can only be confirmed in *Danio rerio* while all sequences outside of the *Homo sapiens/Danio rerio* clade are conservatively defined as paralogues. I therefore distinguished paralogues related by ancient duplication from paralogues that are orthologous to the parent protein of the *Homo sapiens/Danio rerio* specific duplications by using terminology previously described by Koonin (Annu. Rev. Genet., 2005). Distantly related paralogues are more likely to perform non-identical functions to the query protein and are referred to as outparalogues. Paralogues that are orthologous to the parent protein of *Homo sapiens/Danio rerio*-specific duplications are instead referred to here as inparalogues (Figure 4.1). Overall, the combined results from the homology searches and phylogenetic analyses suggest a mosaic pattern in the taxonomic distribution of CME-I network proteins (Figures 4.2-23). The results are thus divided according to distinct patterns of conservation in the eukaryotic genomes sampled.

4.3.1.1 Conserved core CME-I protein network

Database searches for clathrin heavy chain, AP2α and AP2β subunits revealed the presence of homologues with high sequence similarity in every taxon sampled (Figure 4.2). This suggests these key components of the core CME-I network are strongly conserved among eukaryotes. The phylogenetic analyses for the three protein families confirm the pattern of vertebrate specific gene duplications (Figures 4.3, 4.5 and 4.6). This entails true orthology to the respective query proteins can only be confirmed for homologues found in *Danio rerio*. For clathrin heavy chain and AP2β subunits, the phylogenetic analyses suggests that in all non-vertebrate taxa sampled, inparalogues are present, although for AP2β it is weakly supported (Figures 4.3 and 4.6). The AP2α

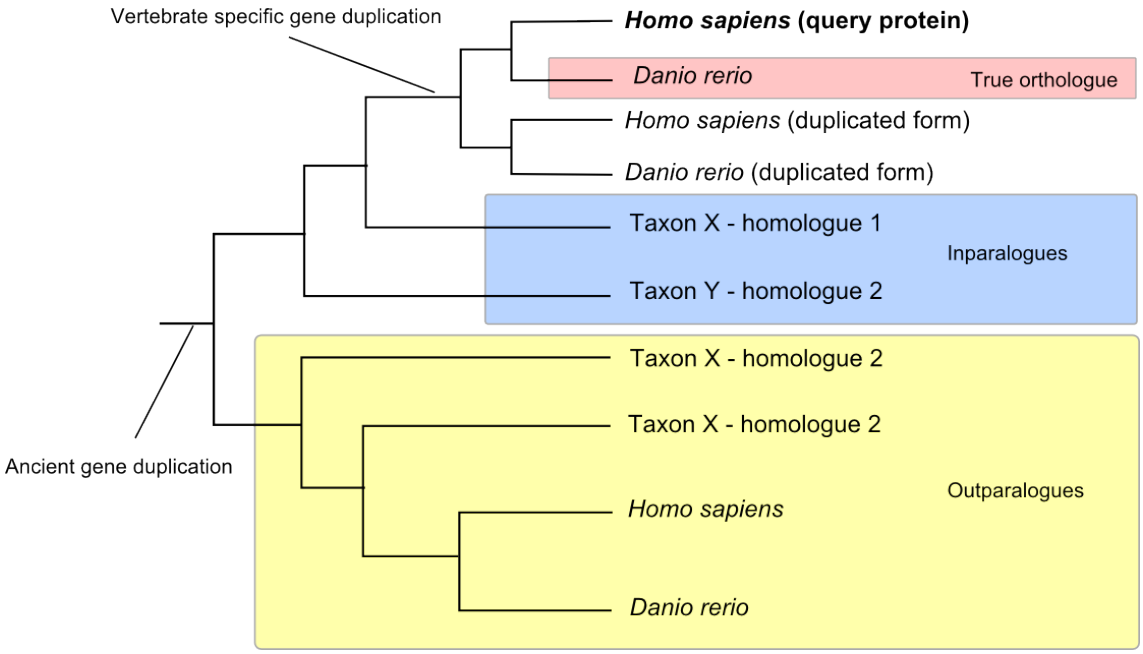


Figure 4.1 Distinction of inparalogues from outparalogues when assessing homology to query protein. The schematic tree depicts hypothetical CME-I protein phylogeny with typical pattern of vertebrate specific duplication. True orthologues are related by vertical descent (shown in red); type A paralogues are homologues that are orthologous with parent protein of vertebrate specific duplication (shown in blue); type B paralogues are homologues related by ancient duplication (shown in yellow).

phylogeny suggests that in all non-vertebrate taxa sampled, except *Trypanosoma brucei* and *Cyanidioschyzon merolae,* inparalogues are present. The AP2α homologues found in *Trypanosoma brucei* and *Cyanidioschyzon merolae* are outparalogues (i.e. related by ancient duplication) (Figure 4.5). In the phylogeny, the three AP2α *Trypanosoma* homologues are AP1γ, AP3δ subunits and AP4ε, the two AP2α *Cyanidioschyzon* homologues are AP1γ and AP3δ subunits (Figure 4.5).

Homologues to the AP2μ and AP2σ subunits were also found in all taxa sampled in the phylogenomic study (Figure 4.2). However, the phylogenies for the two protein families did not show vertebrate specific duplications found in the other two AP2 subunits (Figures 4.7 and 4.8). Orthology was confirmed for holozoan AP2μ homologues. For the rest of the sampled taxa weakly supported inparalogy was indicated except for *Trypanosoma brucei* and *Cyanidioschyzon merolae* where outparalogy was indicated. Out of the three AP2μ outparalogues found in *Trypanosoma brucei,* one is in the AP1μ clade, one is in the AP3μ clade and one is in the AP4μ clade. The *Cyanidioschyzon merolae* AP2μ outparalogue is in the AP3μ clade (Figure 4.7). Strongly supported orthology was confirmed for holozoan AP2σ homologues. For the rest of the sampled taxa weakly supported inparalogy was indicated except for *Ustilago maydis*, *Batrachochytrium dendrobatidis, Trypanosoma brucei* and *Cyanidioschyzon merolae* where outparalogy was indicated (Figure 4.2). The AP2σ outparalogues found in these four taxa group within the AP1σ and the AP3σ clades except for the *Trypanosoma brucei* which are present in the AP1σ, AP3σ and AP4σ clades (Figure 4.8).

Clathrin light chain homologues were found in all sampled taxa except *Entamoeba*

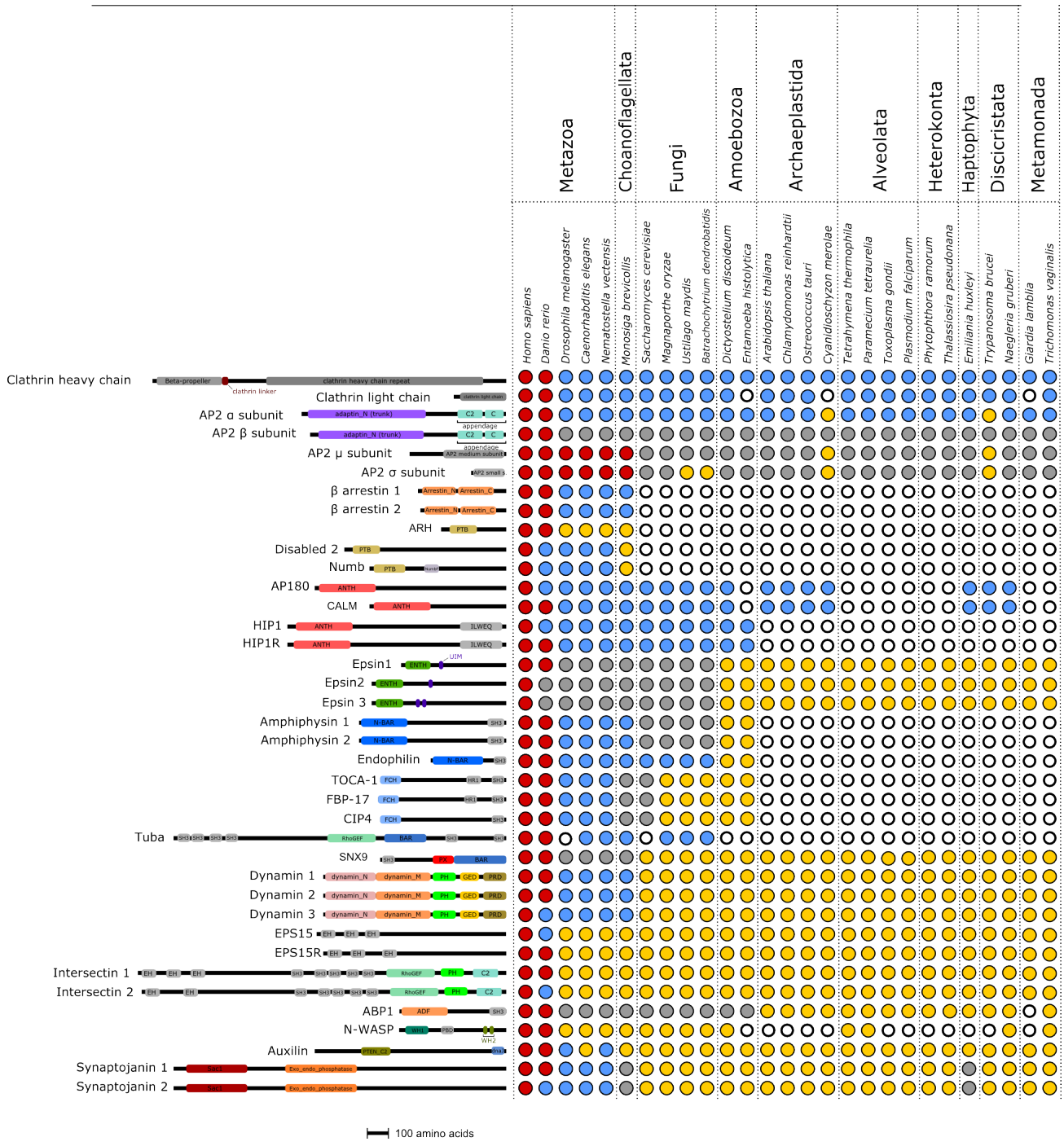


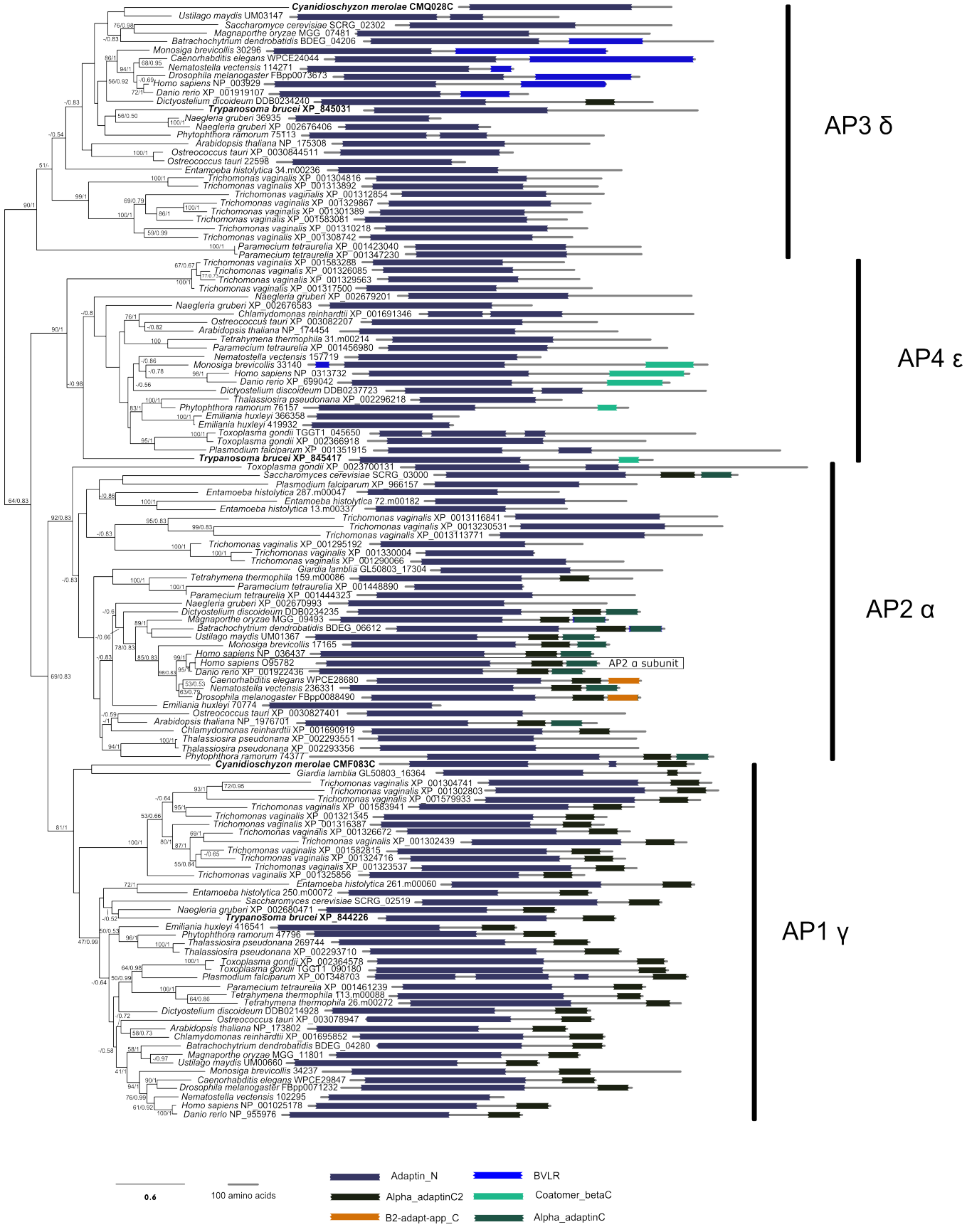
Figure 4.2 Taxonomic distribution of CME-I network proteins. Results from BLASTp, PSI-BLAST and tBLASTn searches against 27 predicted proteomes and translated nucleotide databases representing eukaryotic diversity. Query proteins with respective protein domain structures are shown on the Y-axis, sampled taxa are shown on the X-axis. Coloured circles indicate presence of homologous predicted protein(s). Red circle indicates orthology supported by phylogenetic analysis with minimum support of 80/0.90 (ML bootstrap/Bayesian posterior probability). Blue circle indicates inparalogy with strong support (>=80/0.90). Grey circle indicates that respective phylogeny suggests inparalogy but without strong support. Yellow circle indicates that putative homologue is outparalogue of query. Empty circles indicate absence of homology.

histolytica, Cyanidioschyzon merolae and *Giardia lamblia.* However, it should be noted that in *Chlamydomonas*, alveolates, heterokonts, *Emiliania* and *Trypanosoma brucei,* clathrin light homologues were found with the PSI-BLAST strategy explained in Section 4.2.1, and only very weak sequence similarity was detected. The clathrin light chain protein sequence is relatively small (~200-250 amino acids) and poorly conserved (Figure 4.4). It is possible that the clathrin light chain, in the three taxa where it was not found, is too divergent to be detected even with PSI-BLAST.

Together, this data indicates that clathrin, the main structural component of CME vesicle coats, is highly conserved across eukaryotic diversity and that the four AP complexes evolved prior to the diversification of eukaryotes, except for AP1β and AP2β which evolved during eukaryote diversification. These are not however novel findings. Conservation of clathrin across eukaryotic diversity has already been reported (Field *et al*., 2007). The evolutionary pattern of AP complexes was first suggested by phylogenies including a small sample of diverse eukaryotes (Elde *et al*., 2005; Schledzewski *et al*., 1999), and then conclusively demonstrated with a more broadly sampled, and phylogenetically more robust analysis (Dacks *et al*., 2008). Recently, a fifth AP complex was discovered, and the likely branching order of the five AP complexes determined (Hirst *et al*., 2011) The data also suggests the AP2 complex is highly conserved in the eukaryotes, but that it is absent in *Trypanosoma brucei* and *Cyanidioschyzon merolae.* This finding has also been previously reported (Dacks *et al*., 2008; Field *et al*., 2007; Morgan *et al*., 2001). A final result to consider is absence of the AP2σ subunit in *Batrachochytrium dendrobatidis* and *Ustilago maydis*. Given the unquestionable ancient origin of the four AP complexes, these absences must be due to

secondary loss.

Figure 4. 3 Phylogenetic tree of clathrin heavy chain. Box highlights query protein. For this and all other phylogenetic trees presented in this chapter, maximum likelihood and Bayesian MCMC analyses were run on multiple sequence alignment data-set of the protein sequences, and putative protein domain structures were calculated by searching HMMs database PFAM. In this and all other phylogenetic trees presented in this chapter, branch support values represent percentage of ML bootstrap support and Bayesian posterior probability. Protein domains are illustrated in scale next to respective proteins. Protein domain names are indicated at the bottom.


Cyanidioschyzon merolae CMQ028C
Ustilago maydis UM03147
Saccharomyce cerevisiae SCRG_02302
Magnaporthe oryzae MGG_07481
Batrachochytrium dendrobatidis BDEG_04206
Monosiga brevicollis 30296
Caenorhabditis elegans WPCE24044
Nematostella vectensis 114271
Drosophila melanogaster FBpp0073673
Homo sapiens NP_003929
Danio rerio XP_001919107
Dictyostelium discoideum DDB0234240
Trypanosoma brucei XP_845031
Naegleria gruberi 36935
Naegleria gruberi XP_002676406
Phytophthora ramorum 75113
Arabidopsis thaliana NP_175308
Ostreococcus tauri XP_0030844511
Ostreococcus tauri 22598
Entamoeba histolytica 34.m00236
Trichomonas vaginalis XP_001304816
Trichomonas vaginalis XP_001313892
Trichomonas vaginalis XP_001312854
Trichomonas vaginalis XP_001329867
Trichomonas vaginalis XP_001301389
Trichomonas vaginalis XP_001583081
Trichomonas vaginalis XP_001310218
Trichomonas vaginalis XP_001308742
Paramecium tetraurelia XP_001423040
Paramecium tetraurelia XP_001347230
Trichomonas vaginalis XP_001583288
Trichomonas vaginalis XP_001326085
Trichomonas vaginalis XP_001329563
Trichomonas vaginalis XP_001317500
Naegleria gruberi XP_002679201
Naegleria gruberi XP_002676583
Chlamydomonas reinhardtii XP_001691346
Ostreococcus tauri XP_003082207
Arabidopsis thaliana NP_174454
Tetrahymena thermophila 31.m00214
Paramecium tetraurelia XP_001456980
Nematostella vectensis 157719
Monosiga brevicollis 33140
Homo sapiens NP_0313732
Danio rerio XP_699042
Dictyostelium discoideum DDB0237723
Thalassiosira pseudonana XP_002296218
Phytophthora ramorum 76157
Emiliania huxleyi 366358
Emiliania huxleyi 419932
Toxoplasma gondii TGGT1_045650
Toxoplasma gondii XP_002366918
Plasmodium falciparum XP_001351915
Trypanosoma brucei XP_845417
Toxoplasma gondii XP_0023700131
Saccharomyces cerevisiae SCRG_03000
Plasmodium falciparum XP_966157
Entamoeba histolytica 287.m00047
Entamoeba histolytica 72.m00182
Entamoeba histolytica 13.m00337
Trichomonas vaginalis XP_0013116841
Trichomonas vaginalis XP_0013230531
Trichomonas vaginalis XP_0013113771
Trichomonas vaginalis XP_001295192
Trichomonas vaginalis XP_001330004
Trichomonas vaginalis XP_001290066
Giardia lamblia GL50803_17304
Tetrahymena thermophila 159.m00086
Paramecium tetraurelia XP_001448890
Paramecium tetraurelia XP_001444323
Naegleria gruberi XP_002670993
Dictyostelium discoideum DDB0234235
Magnaporthe oryzae MGG_09493
Batrachochytrium dendrobatidis BDEG_06612
Ustilago maydis UM01367
Monosiga brevicollis 17165
Homo sapiens NP_036437
Homo sapiens O95782
Danio rerio XP_001922436
AP2 α subunit
Caenorhabditis elegans WPCE28680
Nematostella vectensis 236331
Drosophila melanogaster FBpp0088490
Emiliania huxleyi 70774
Ostreococcus tauri XP_0030827401
Arabidopsis thaliana NP_1976701
Chlamydomonas reinhardtii XP_001690919
Thalassiosira pseudonana XP_002293551
Thalassiosira pseudonana XP_002293356
Phytophthora ramorum 74377
Cyanidioschyzon merolae CMF083C
Giardia lamblia GL50803_16364
Trichomonas vaginalis XP_001304741
Trichomonas vaginalis XP_001302803
Trichomonas vaginalis XP_001579933
Trichomonas vaginalis XP_001583941
Trichomonas vaginalis XP_001321345
Trichomonas vaginalis XP_001316387
Trichomonas vaginalis XP_001326672
Trichomonas vaginalis XP_001302439
Trichomonas vaginalis XP_001582815
Trichomonas vaginalis XP_001324716
Trichomonas vaginalis XP_001323537
Trichomonas vaginalis XP_001325856
Entamoeba histolytica 261.m00060
Entamoeba histolytica 250.m00072
Saccharomyces cerevisiae SCRG_02519
Naegleria gruberi XP_002680471
Trypanosoma brucei XP_844226
Emiliania huxleyi 416541
Phytophthora ramorum 47796
Thalassiosira pseudonana 269744
Thalassiosira pseudonana XP_002293710
Toxoplasma gondii XP_002364578
Toxoplasma gondii TGGT1_090180
Plasmodium falciparum XP_001348703
Paramecium tetraurelia XP_001461239
Tetrahymena thermophila T13.m00088
Tetrahymena thermophila 26.m00272
Dictyostelium discoideum DDB0214928
Ostreococcus tauri XP_003078947
Arabidopsis thaliana NP_173802
Chlamydomonas reinhardtii XP_001695852
Batrachochytrium dendrobatidis BDEG_04280
Magnaporthe oryzae MGG_11801
Ustilago maydis UM00660
Monosiga brevicollis 34237
Caenorhabditis elegans WPCE29847
Drosophila melanogaster FBpp0071232
Nematostella vectensis 102295
Homo sapiens NP_001025178
Danio rerio NP_955976
AP3 δ
AP4 ε
AP2 α
AP1 γ
0.6
100 amino acids
Adaptin_N
Alpha_adaptinC2
B2-adapt-app_C
BVLR
Coatomer_betaC
Alpha_adaptinC

(previous page) **Figure 4.5 Phylogenetic tree of AP2 α subunit.** The phylogeny includes the AP γ, δ and ε subunits which belong to the AP1, AP3 and AP4 complexes, respectively. The *Trypanosoma brucei* and *Cyanidioschizon merolae* outparalogues are highlighted in bold lettering.

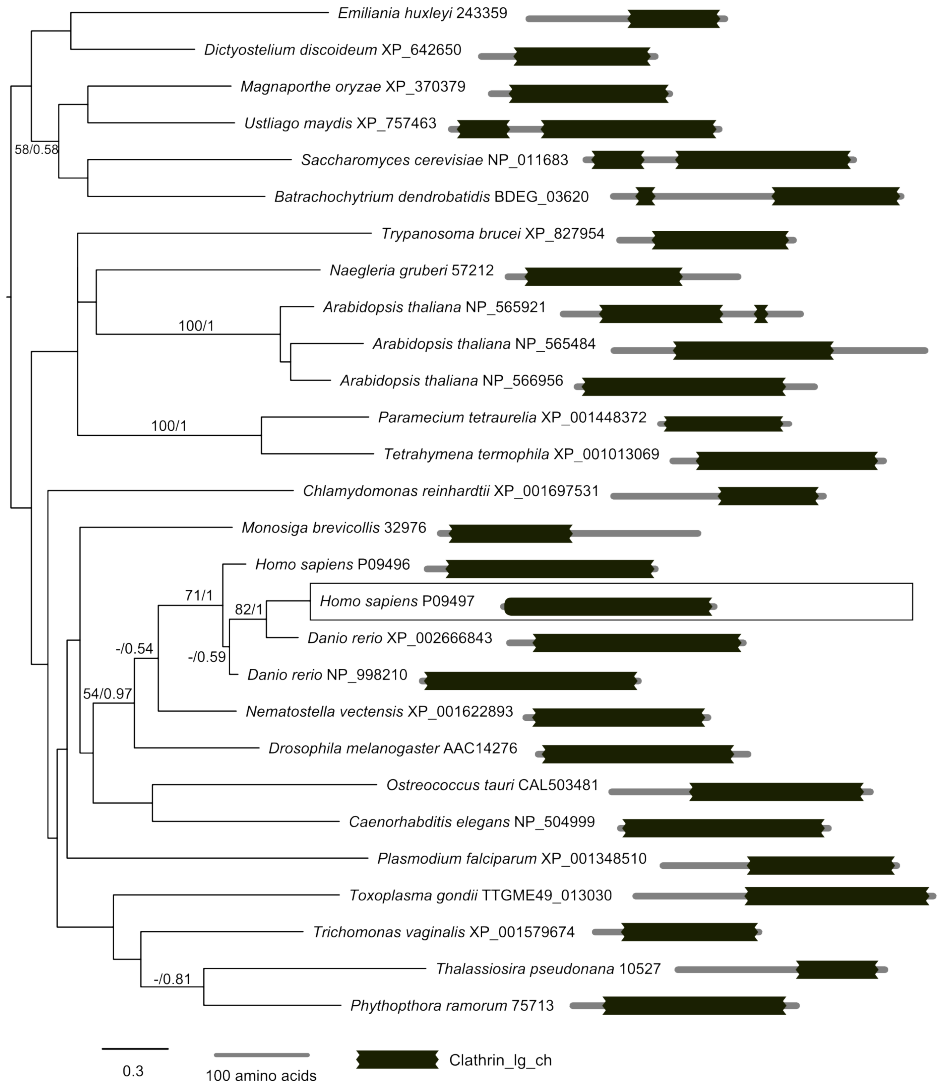


Figure 4.4 Phylogenetic tree of clathrin light chain. Box highlights query protein. Vertebrate specific gene duplication is suggested but the paralogues are not resolved.

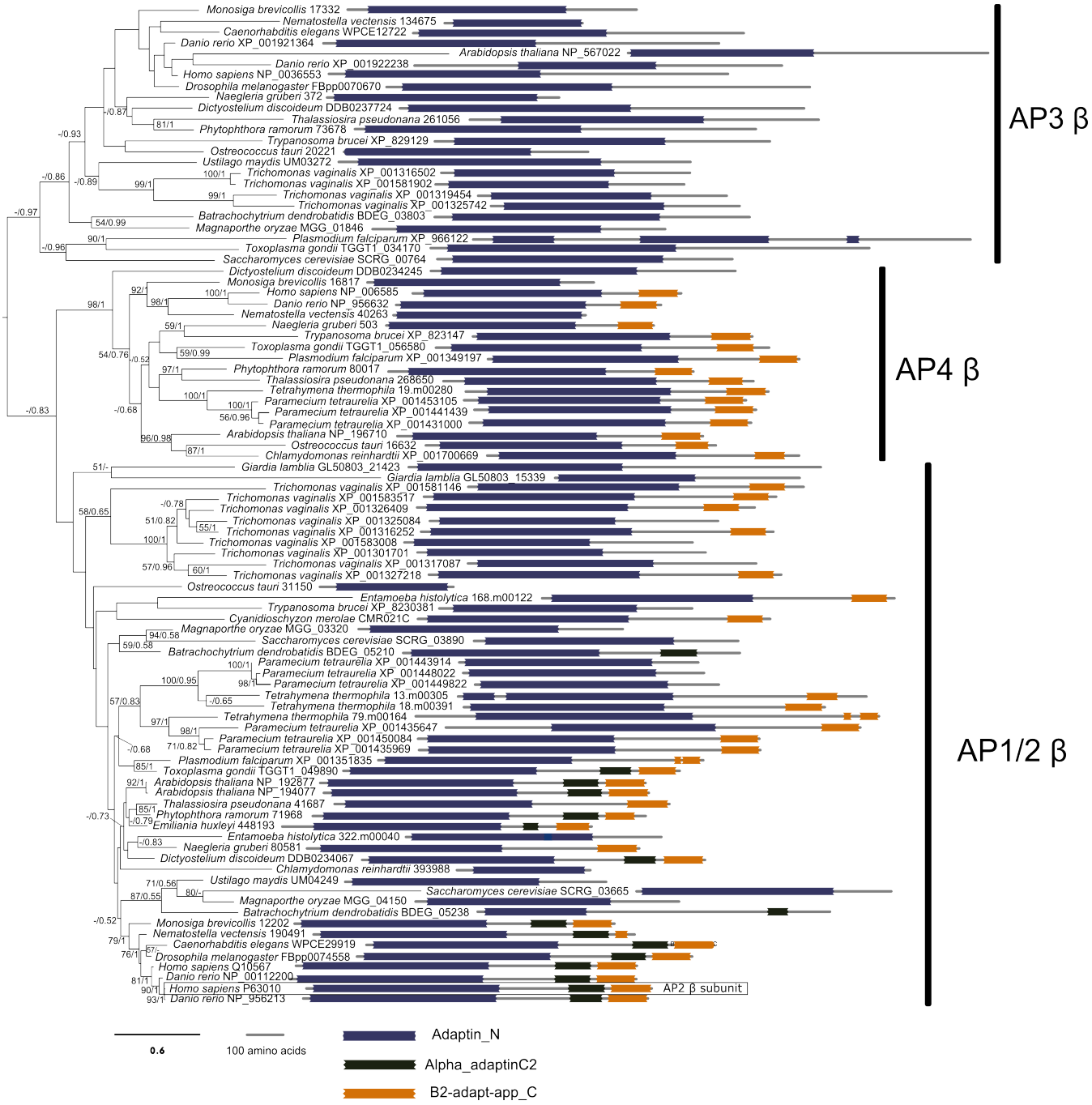


Figure 4.6 Phylogenetic tree of AP2 β subunits. The phylogeny includes the β subunits of the AP1, AP3 and AP4 complexes. The AP3β clade and AP4β clade are resolved, suggesting pre-LCEA origin, while AP1β and AP2β form one poorly resolved clade, suggesting they evolved during eukaryote diversification as previously reported (Dacks *et al.* 2008). Box highlights query protein. Ptotein domain names are indicated at the bottom.

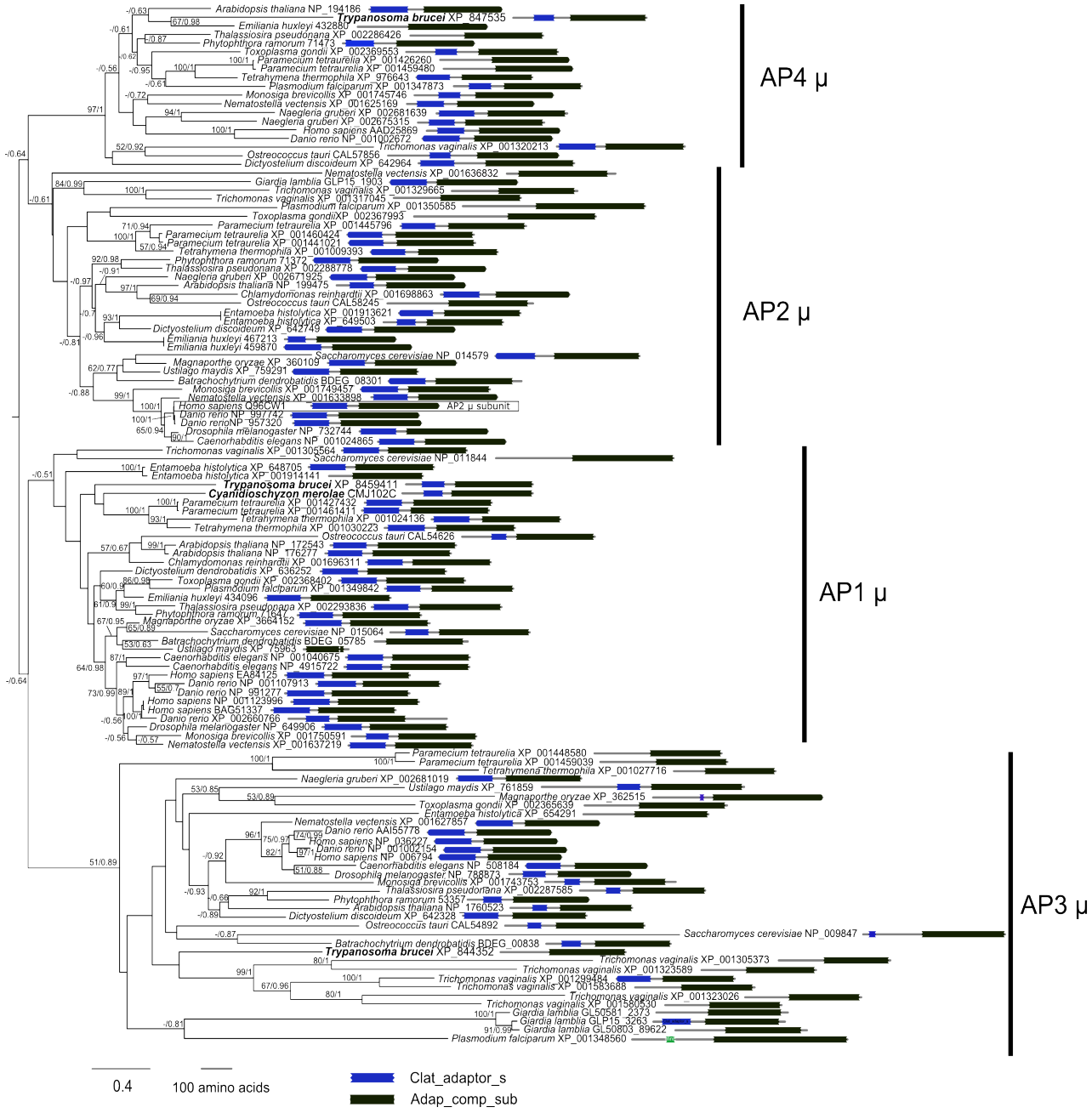


Figure 4.7 Phylogenetic tree of AP2 μ subunits. The phylogeny includes the μ subunits of the AP1, AP3 and AP4 complexes. The clades corresponding to the four subunits are labelled accordingly. The four APμ clades are resolved, but with the exception of AP4μ they are supported by relatively weak support values. Box highlights query protein. The *Trypanosoma brucei* and *Cyanidioschyzon merolae* AP2μ outparalogues are highlighted by bold lettering.

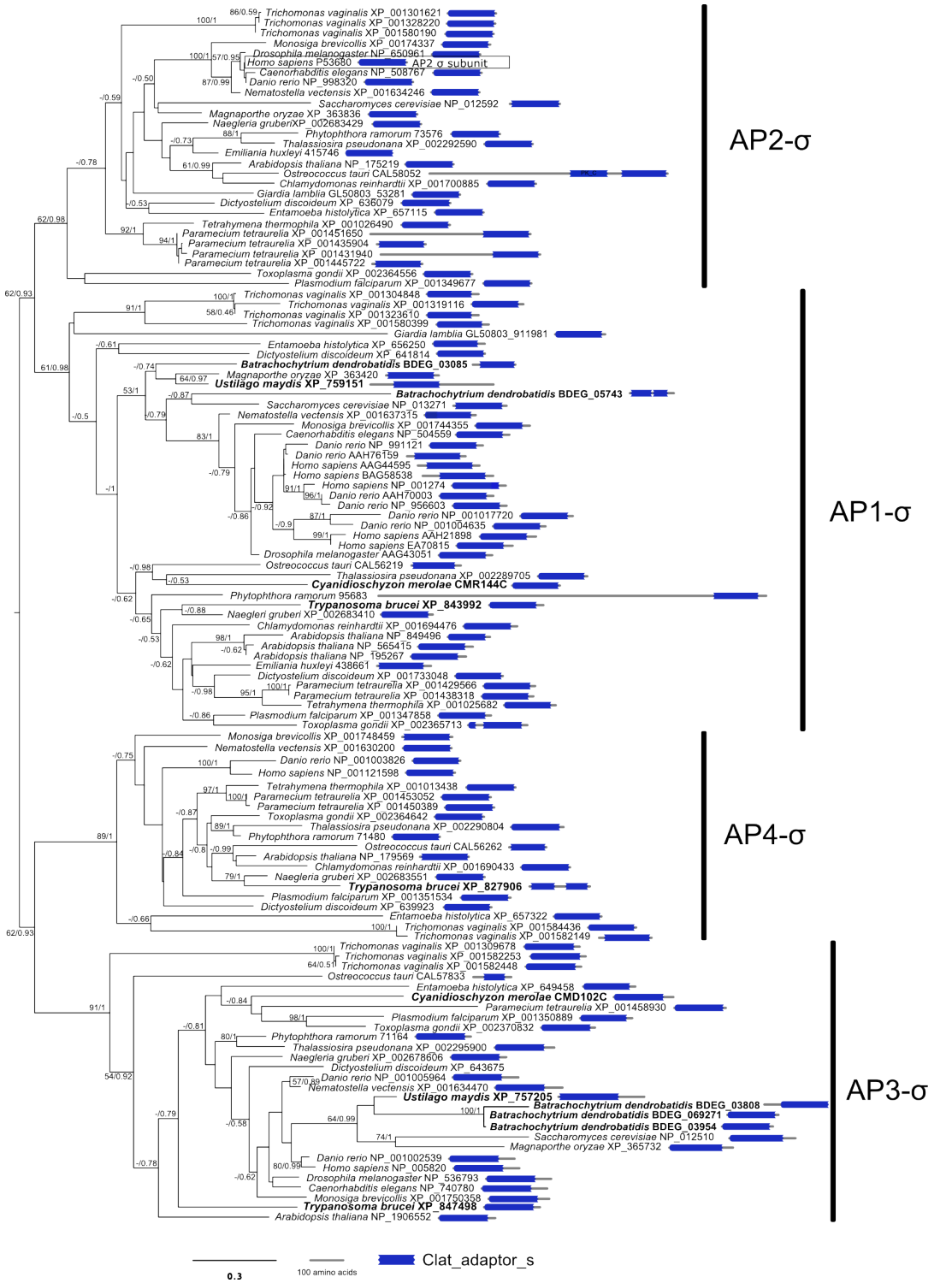


Figure 4.8 Phylogenetic tree of AP2 µ subunits. The phylogeny includes the σ subunits of the AP1, AP3 and AP4 complexes. The clades corresponding to the four subunits are labelled accordingly. The four APσ clades are resolved with moderate to strong statistical support. Box highlights query protein. The *Trypanosoma brucei, Cyanidioschyzon merolae, Ustilago maydis* and *Batrachochytrium dendrobatidis* AP2µ outparalogues are highlighted by bold lettering.

4.3.1.2 Epsins, dynamins, SNX9, synaptojanins, auxilin and EPS15/EPS15R/ intersectins are conserved but have complex evolutionary histories

The epsins, dynamins, SNX9 and synaptojanins have complex phylogenetic patterns. For the four protein families, homologues were found in all the taxa sampled. However, respective phylogenetic analyses suggest gene duplications and/or protein domain rearrangements have resulted in multiple and divergent paralogous protein forms which have diversified across the eukaryotic phylogeny (Figures 4.9, 4.10, 4.11 and 4.12). All four protein families are characterised by at least one ancient gene duplication, and it is likely that distantly related paralogues evolved non-identical functions to the query proteins. They all share the same pattern of vertebrate specific duplications which means orthology could only be confirmed in *Danio rerio*, so here I concentrate on the taxonomic distribution of inparalogues to assess the possibility that CME variants of these protein families are present in the taxa sampled. According to the epsins phylogeny, inparalogues are confirmed for homologues found in non-vertebrate opisthokont taxa, but not with strong statistical support (Figure 4.9). All other homologues were conservatively defined as outparalogues. This result is consistent with the only other published eukaryotic wide phylogeny of the epsins (Gabernet-Castello *et al.*, 2009). In synaptojanins (Figure 4.10), inparalogy is strongly supported for homologues present in non-vertebrate metazoan taxa, and weakly supported for *Monosiga brevicollis* and, interestingly, *Emiliania huxleyi*. All other synaptojanin homologues were conservatively defined as outparalogues. The dynamins are known to be diverse and carry several distinct functions within the cell (Praefcke & McMahon, 2004), so it is important for this study to pinpoint the putative origin of the endocytic variant. This study strongly supports presence of inparalogues in metazoan and

choanoflagellate taxa (together known as Holozoa, Lang *et al*., 2002) (Figure 4.11) which suggests the dynamin paralogue has its origin at the root of the Holozoa. This is a novel finding given that the two prior phylogenetic eukaryotic-wide studies of dynamins did not sample *Monosiga brevicollis,* or other closely related choanozoans in the analyses (Elde *et al*., 2005; Miyagishima *et al*., 2008). Similarly to dynamins, SNX9 belongs to a wider group of proteins known as sorting nexins, which are employed in a variety of putative functions within membrane trafficking and protein sorting (Seet & Hong, 2006; Worby & Dixon, 2002). The SNX9 phylogeny also suggests inparalogues are present in the holozoan taxa sampled, but only with weak statistical support (Figure 4.12). The origin of the SNX9 endocytic paralogue could thus be pinpointed at the root of the Holozoa, but with low confidence.

Auxilin features an N-terminal domain which is homologous to the PTEN phosphatase and can act on phosphoinositides such as PI(4,5)P2 (Lee *et al*., 1999). It also feature a C-terminal DNAJ domain. Homologues to at least part of the protein were found in all the taxa sampled, but the phylogenetic tree was based on the alignment of the PTEN_C2 domain because it is the most conserved section of the protein. The auxilin phylogeny confirms the presence of orthologues in *Danio rerio*, and inparalogues in *Nematostella vectensis* and *Drosophila melanogaster* but not in *Caenorhabditis elegans* (Figure 4.13). This data suggests the auxilin lineage originated at the root of the Metazoa.

The analyses on the EPS15/EPS15R and intersectin protein families point towards a similar evolutionary history to epsins, synaptojanins, SNX9 and dynamins but for the following reasons: for taxa outside of Holozoa, the homology hits are very weak and

based on the eukaryotic universal Ef Hand (EH) functional domain which is very short (~60 amino acids) and repeated across the N-terminals of EPS15/EPS15R and intersectins. In addition, in the intersectins the EH domains represent roughly a fifth of the total functional repertoire of the protein, which is fully conserved only in *Homo sapiens* and *Danio rerio,* and partially conserved in the rest of the metazoans sampled. While multiple duplication and protein domain recombinations have occurred, the paucity of conserved characters for a global alignment means the phylogeny is poorly resolved and the origin of the endocytic variants can not be pinpointed with confidence (Figure 4.14).

4.3.1.3 Evolutionary distribution of ABP1, N-WASP and AP180/CALM

ABP1 and N-WASP belong to large and diverse protein families which are involved in a range of actin-related cellular functions. ABP1 belongs to the ADF/cofilin family which mediates actin depolymerisation (Bernstein & Bamburg, 2010). In the genomic searches more than 180 ADF/cofilin proteins were identified, and an initial phylogenetic tree was produced to identify the section of the tree with ABP1 (data not shown). The ABP1 data sub-set was then used for a further phylogeny. Although it is overall poorly resolved, the resulting tree highlights the typical vertebrate-specific gene duplication and suggests that ABP1 may be specific to unikonts (Figure 4.15). However, the statistical support for the unikont clade is weak (-/0.67). Notably, outparalogues of ABP1 were found in all sampled taxa except for *Giardia lamblia.* N-WASP belongs to the WASP/WAVE family which mediates actin nucleation in different cell processes (Pollitt & Insall, 2009). Similarly to the ABP1 analysis, an initial global phylogeny was used to identify the section of the tree where N-WASP branched. A full analysis of the N-WASP sub-set

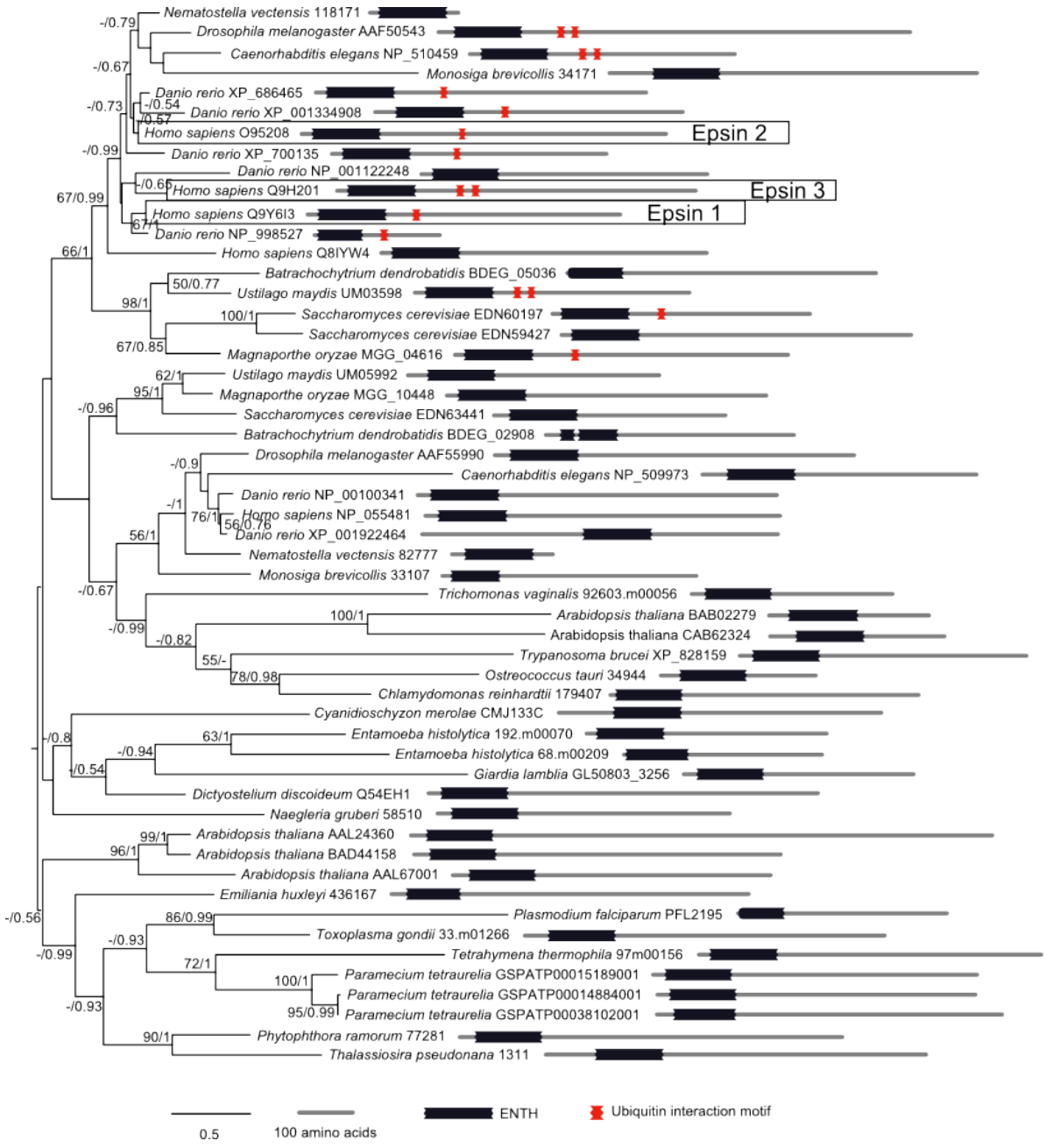


Figure 4.9 Phylogenetic tree of epsins. The phylogeny suggests the three human epsins are the result of typical vertebrate-specific gene duplications. It provides strong support for inparalogy in Holozoa and weak support for inparalogy in Fungi. It also suggests an opisthokont origin of the ENTH-UIM domain structure, but with moderate statistical support (66/1). The phylogeny is poorly resolved at the deep branching level, indicating possible an ancient gene duplication during eukaryote diversification.

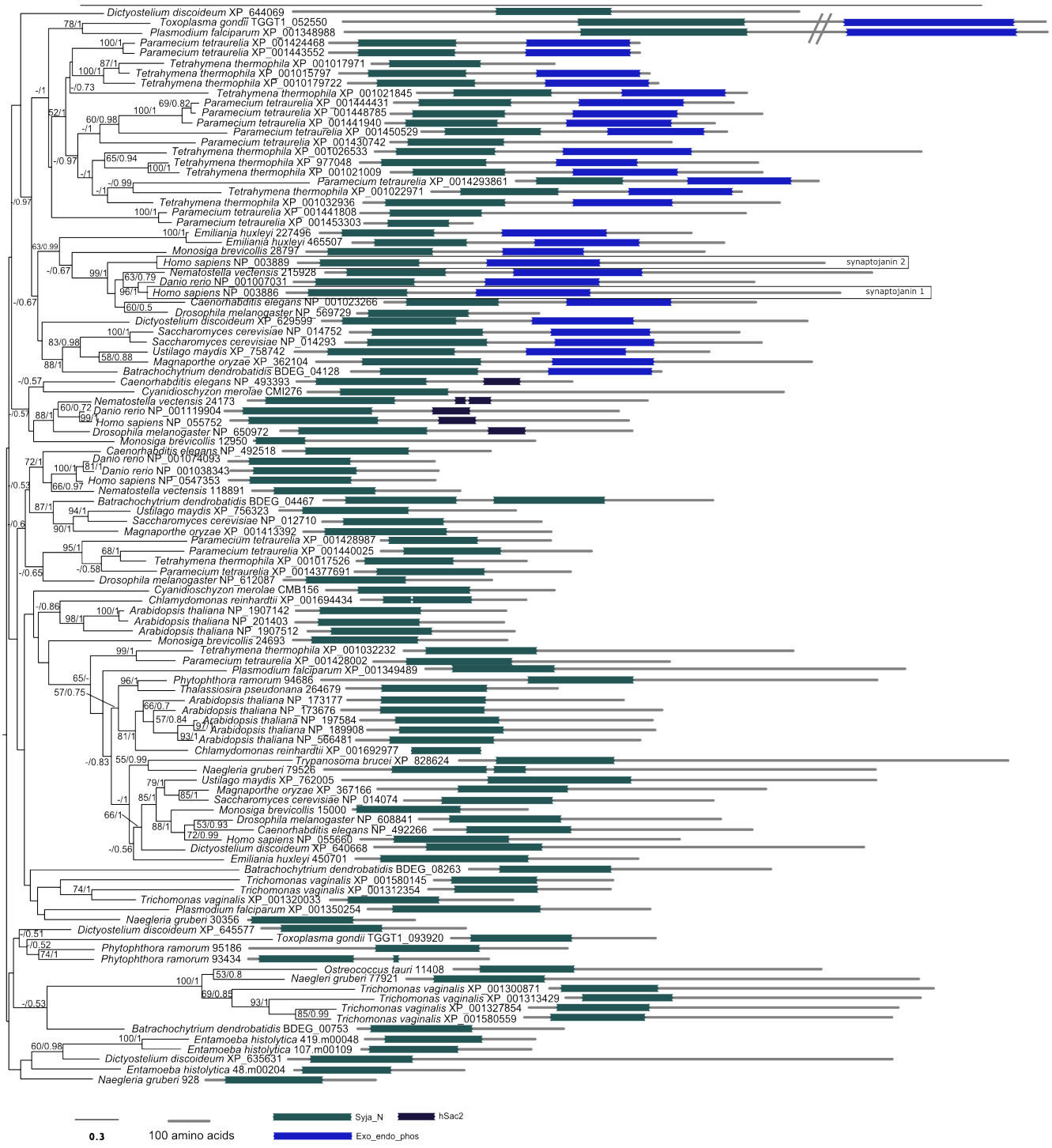


Figure 4.10 Phylogenetic tree of synaptojanins. Analyses were based on the alignment of the N-terminal Syja_N domain. The Giardia lamblia GL50803_8589 homologue was removed due to long branch. Overall the phylogeny is poorly resolved, especially at deep branching level. Metazoan and fungal synaptojanin groups are resolved but opisthokont monophyly is not supported statistically (-/0.67). This may be due to the unusual branching position of the two *Emiliania huxleyi* homologues (227496 and 465507) as sister to the Holozoa. The phylogenetic distribution of the Syja_N-Exo_endo_phos domain structure is limited to opisthokonts, *Dictyostelium,* alvoelates and the haptophyte. The phylogeny suggests that deep level gene duplications are likely to have occurred.

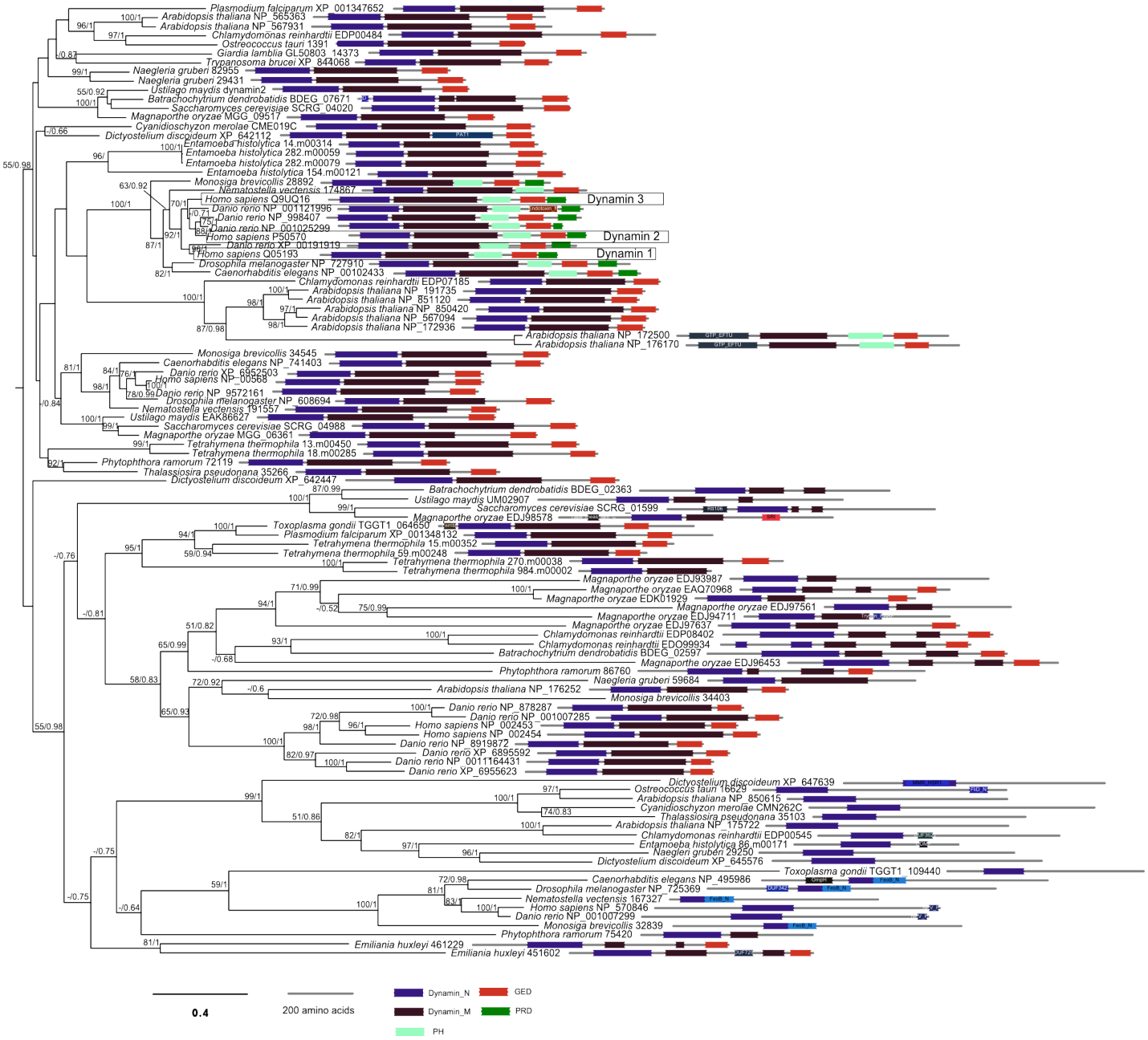


Figure 4.11 Phylogenetic tree of dynamins. To reduce computational demand the 11 *Paramecium tetraurelia* and 8 *Trichomonas vaginalis* dynamin homologues were excluded from the analysis. The holozoan clade comprising the 'classical' dynamins is well supported (100/1), although its revolutionary relationship with other dynamins lineages is unresolved. The phylogeny suggests deep-level gene duplications have occurred but it is inconclusive as to the branching order of distantly related paralogues. The protein domain architecture of endocytic dynamins is specific to holozoa, although interestingly two *Arabidopsis thaliana* homologues (NP_172500 and NP_176170) are predicted to have a PH domain that is otherwise specific to 'classical' dynamins.

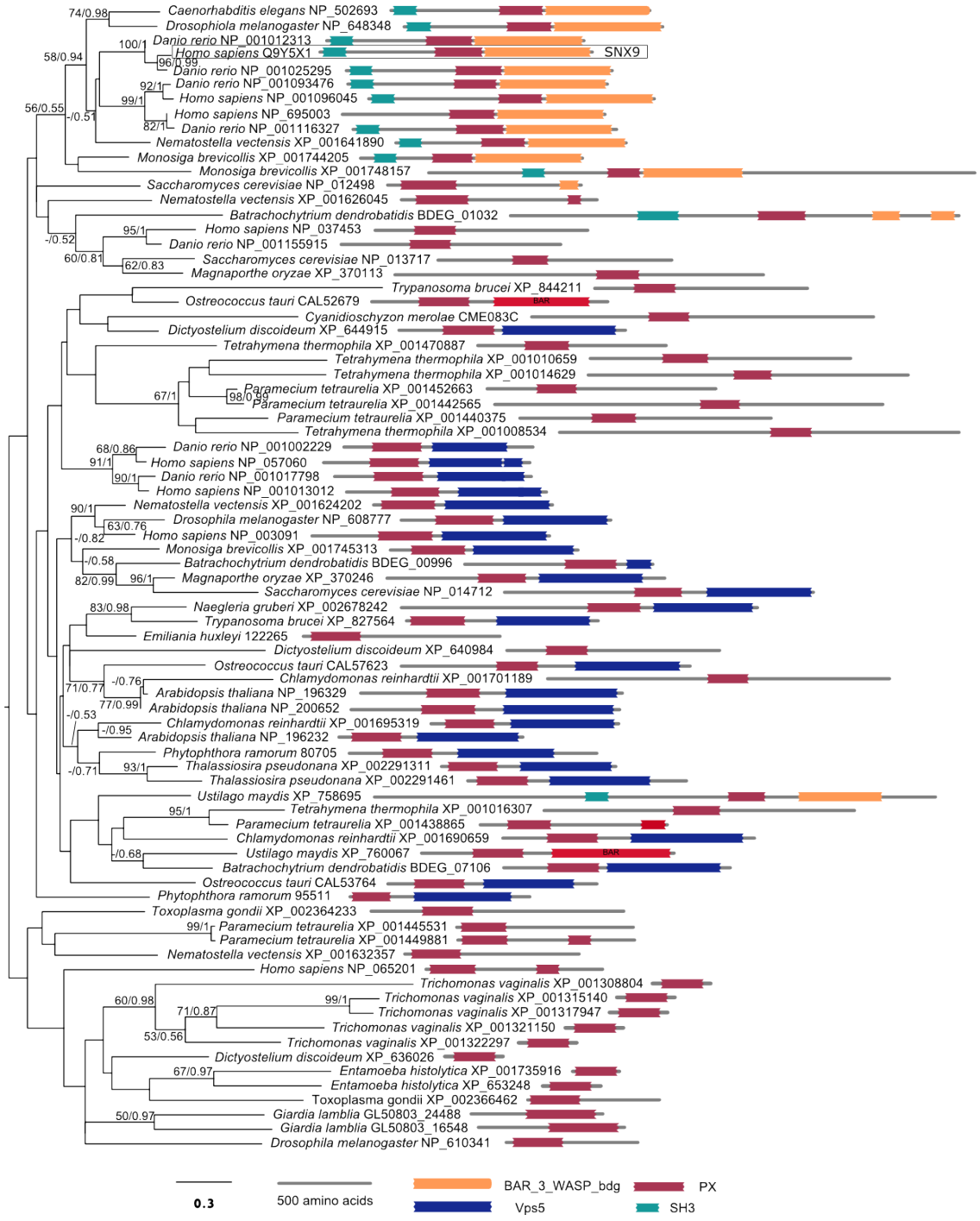


Figure 4.12 Phylogenetic tree of SNX9. The Plasmodium falciparum XP_001349109 homologue was removed from analysis due to long branch. The phylogeny indicates that multple vertebrate-specific gene duplications produced SNX9, with inparalogy confirmed for other metazoan and *Monosiga brevicollis* with weak to moderate statistical support. As for epsins, synaptojanins and dynamins, phylogenetic analysis suggests a complex evolutionary history with deep-level gene duplications, but the tree overall lacks resolution. The SH3-PX-BAR_3_WASP domain structure appears to be specific to Holozoa, although a *Batrachochytrium dendrobatidis* homologue (BDEG_01032) and a *Ustilago maydis* homologue (XP_758695) feature the same protein domains albeit in much larger protein sizes.

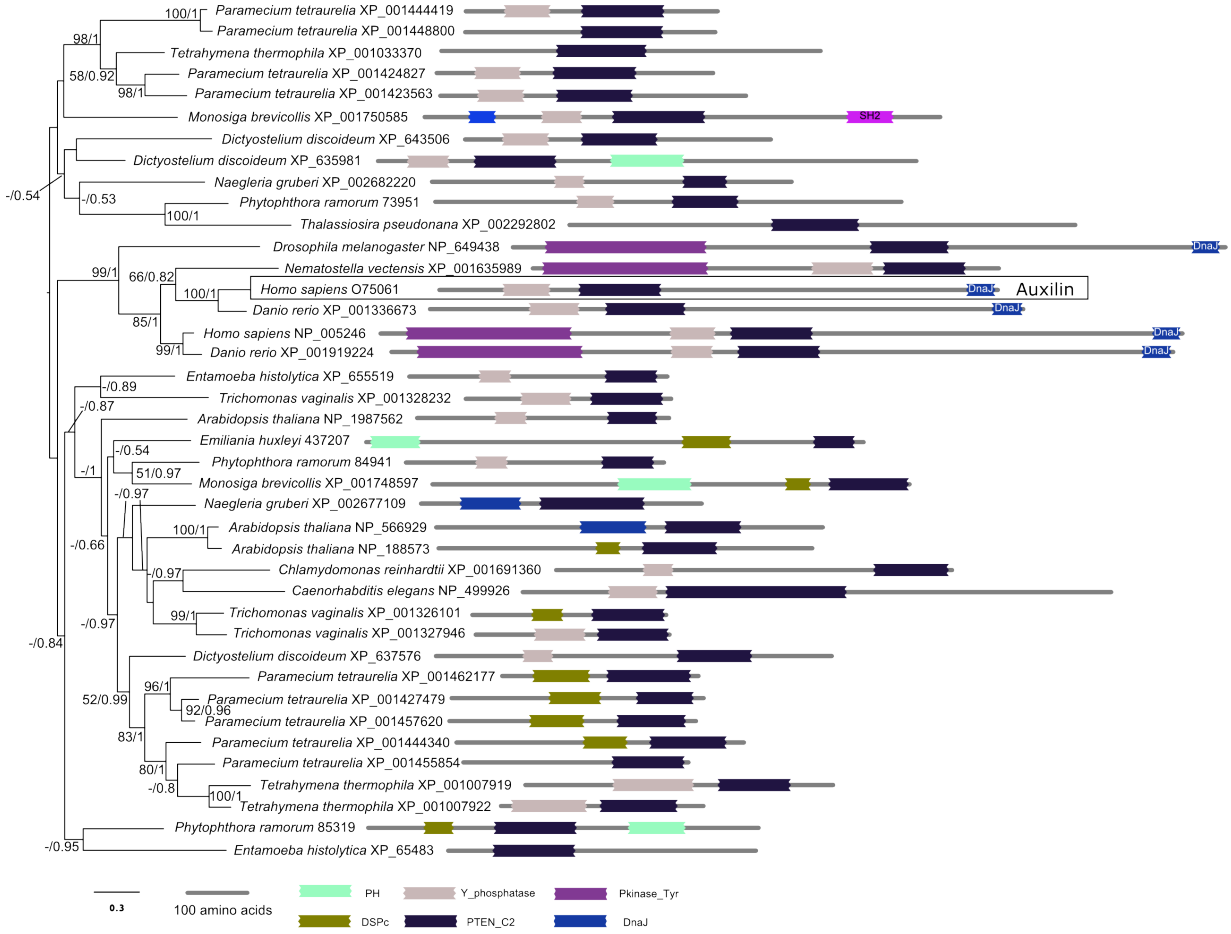
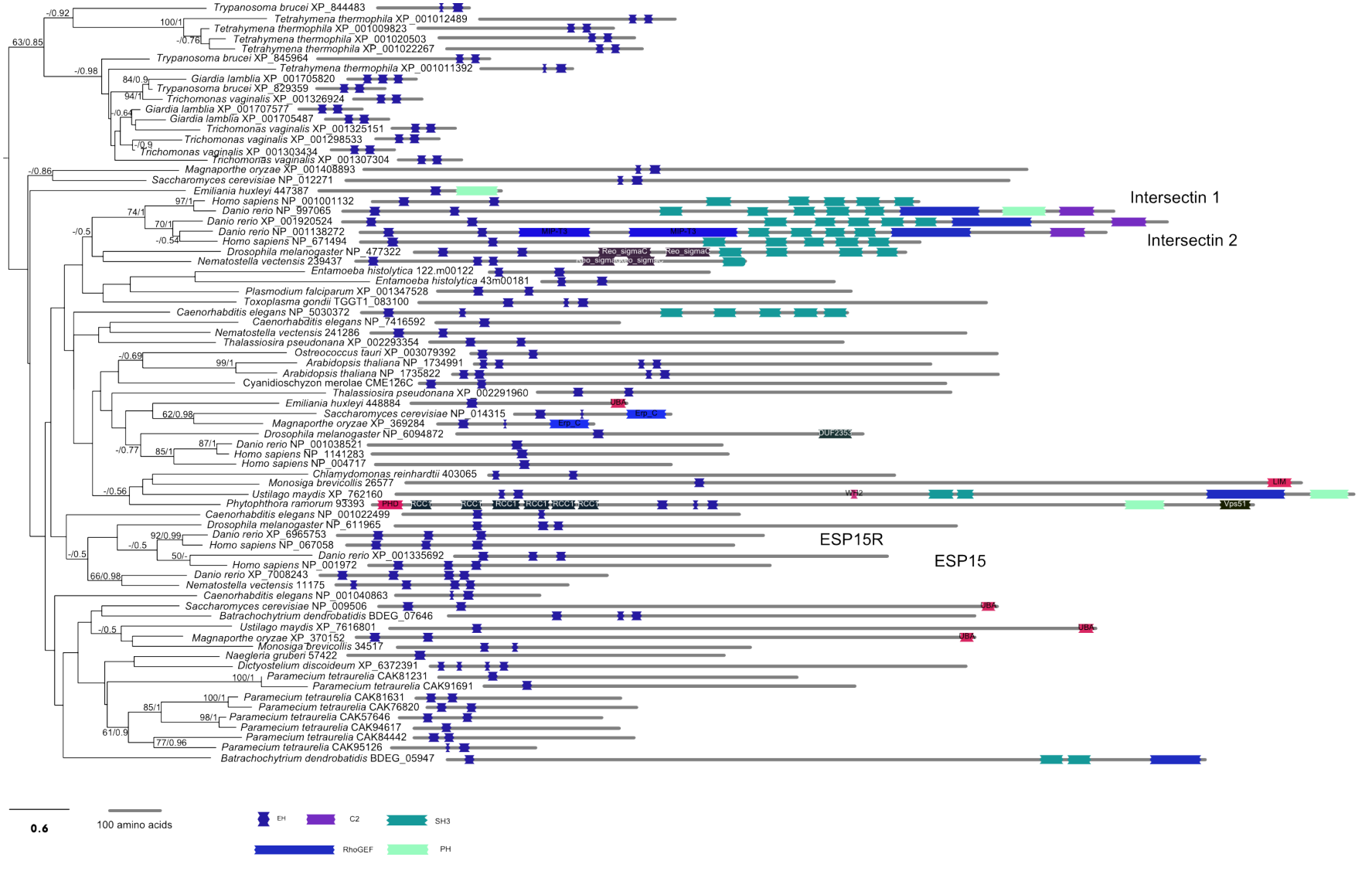


Figure 4.13 Phylogenetic tree of auxilin. The tree is based on the alignment of the central PTEN_C2 homology domain. The domain is missing in the Fungi, Apicomplexa, *Cyanidioschyzon merolae, Trypanosoma brucei* and *Giardia lamblia* which are thus not featured in this analysis. *Ostreococcus tauri* XP_003080200 long branch homologoue was removed. The auxilin clade is clearly resolved but it only includes Metazoa minus *Caenorhabditis elegans*, for which the PTEN_C2-DNAJ protein could not be found.


Trypanosoma brucei XP_844483
Tetrahymena thermophila XP_001012489
Tetrahymena thermophila XP_001009823
Tetrahymena thermophila XP_001020503
Tetrahymena thermophila XP_001022267
Trypanosoma brucei XP_845964
Tetrahymena thermophila XP_001011392
Giardia lamblia XP_001705820
Trypanosoma brucei XP_829359
Trichomonas vaginalis XP_001326924
Giardia lamblia XP_001707577
Giardia lamblia XP_001705487
Trichomonas vaginalis XP_001325151
Trichomonas vaginalis XP_001298533
Trichomonas vaginalis XP_001303434
Trichomonas vaginalis XP_001307304
Magnaporthe oryzae XP_001408893
Saccharomyces cerevisiae NP_012271
Emiliania huxleyi 447387
Homo sapiens NP_001001132
Danio rerio NP_997065
Danio rerio XP_001920524
Danio rerio NP_001138272
Homo sapiens NP_671494
Drosophila melanogaster NP_477322
Nematostella vectensis 239437
Entamoeba histolytica 122.m00122
Entamoeba histolytica 43m00181
Plasmodium falciparum XP_001347528
Toxoplasma gondii TGGT1_083100
Caenorhabditis elegans NP_5030372
Caenorhabditis elegans NP_7416592
Nematostella vectensis 241286
Thalassiosira pseudonana XP_002293354
Ostreococcus tauri XP_003079392
Arabidopsis thaliana NP_1734991
Arabidopsis thaliana NP_1735822
Cyanidioschyzon merolae CME126C
Thalassiosira pseudonana XP_002291960
Emiliania huxleyi 448884
Saccharomyces cerevisiae NP_014315
Magnaporthe oryzae XP_369284
Drosophila melanogaster NP_6094872
Danio rerio NP_001038521
Homo sapiens NP_1141283
Homo sapiens NP_004717
Chlamydomonas reinhardtii 403065
Monosiga brevicollis 26577
Ustilago maydis XP_762160
Phytophthora ramorum 93393
Caenorhabditis elegans NP_001022499
Drosophila melanogaster NP_611965
Danio rerio XP_6965753
Homo sapiens NP_067058
Danio rerio XP_001335692
Homo sapiens NP_001972
Danio rerio XP_7008243
Nematostella vectensis 11175
Caenorhabditis elegans NP_001040863
Saccharomyces cerevisiae NP_009506
Batrachochytrium dendrobatidis BDEG_07646
Ustilago maydis XP_7616801
Magnaporthe oryzae XP_370152
Monosiga brevicollis 34517
Naegleria gruberi 57422
Dictyostelium discoideum XP_6372391
Paramecium tetraurelia CAK81231
Paramecium tetraurelia CAK91691
Paramecium tetraurelia CAK81631
Paramecium tetraurelia CAK76820
Paramecium tetraurelia CAK57646
Paramecium tetraurelia CAK94617
Paramecium tetraurelia CAK84442
Paramecium tetraurelia CAK95126
Batrachochytrium dendrobatidis BDEG_05947
Intersectin 1
Intersectin 2
ESP15R
ESP15
0.6
100 amino acids
EH
C2
SH3
RhoGEF
PH

(previous page) **Figure 4.14 Phylogenetic tree of EPS15, EPS15R and intersectins.** The phylogeny is mostly unresolved and support values are low even for established clades. This is due to the paucity of conserved characters available (125 for an alignment of 101 sequences) when aligning the EH domains. However it still highlights the vertebrate-specific duplications which produced EPS15 and EPS15R, and the two intersectins.

again resulted in a poorly resolved tree where robust statistical support was only found for vertebrate clades (Figure 4.16). Notably, the N-WASP section of the WASP/WAVE tree, did not include Archaeplastida, heterokonts, apicomplexa, *Entamoeba histolytica, Emiliania huxleyi, Trypanosoma brucei* and *Giardia lamblia*, suggesting secondary loss of the ancient paralogue from which N-WASP evolved. Secondary loss was also suggested by results of the AP180/CALM analysis. As previously reported, homologues were not found in *Entamoeba,* alveolates, heterokonts and metamonads (Field *et al.*, 2007). However, a previously unreported *Emiliania huxleyi* (JGI acc. n. 454329) predicted protein with a short N-terminal ANTH domain was found. The phylogeny indicates that A180 and CALM are the result of a vertebrate specific gene duplication. Interestingly, as many as 8 *Arabidopsis* specific gene duplications were also indicated (Figure 4.17).

4.3.1.4 HIP1/HIP1R, amphiphysins, endophilin and FCH are unikont-specific

The analyses of the HIP1/HIP1R, amphiphysins, endophilin and TOCA-1/FBP-17/CIP4 proteins produced similar results. For all these proteins, homologues were only found in unikont taxa, and vertebrate-specific gene duplications were indicated by respective phylogenies (Figure 4.18-20). HIP1/HIP1R proteins feature an N-terminal conserved domain which is similar to the ANTH domain in AP180/CALM protein family. However, the sequence similarity between the two ANTH domains is very weak and

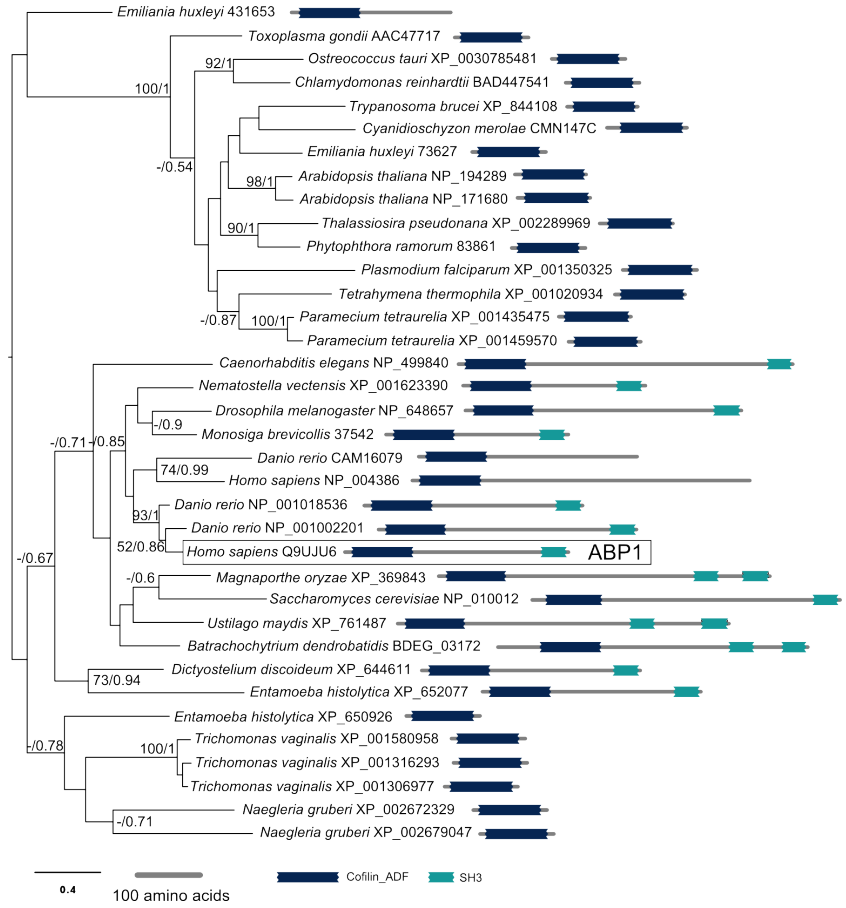


Figure 4.15. Phylogenetic tree of ABP1. The analysis is based on the alignment of the ADF/cofilin protein domain. Tree represents the sub-branch of the ADF/cofilin protein family comprising ABP1 and representative outgroup species. Analysis suggests ABP1 are specific to unikonts, although the finding is weakly supported. The ADF-SH3 domain distributes to unikonts.

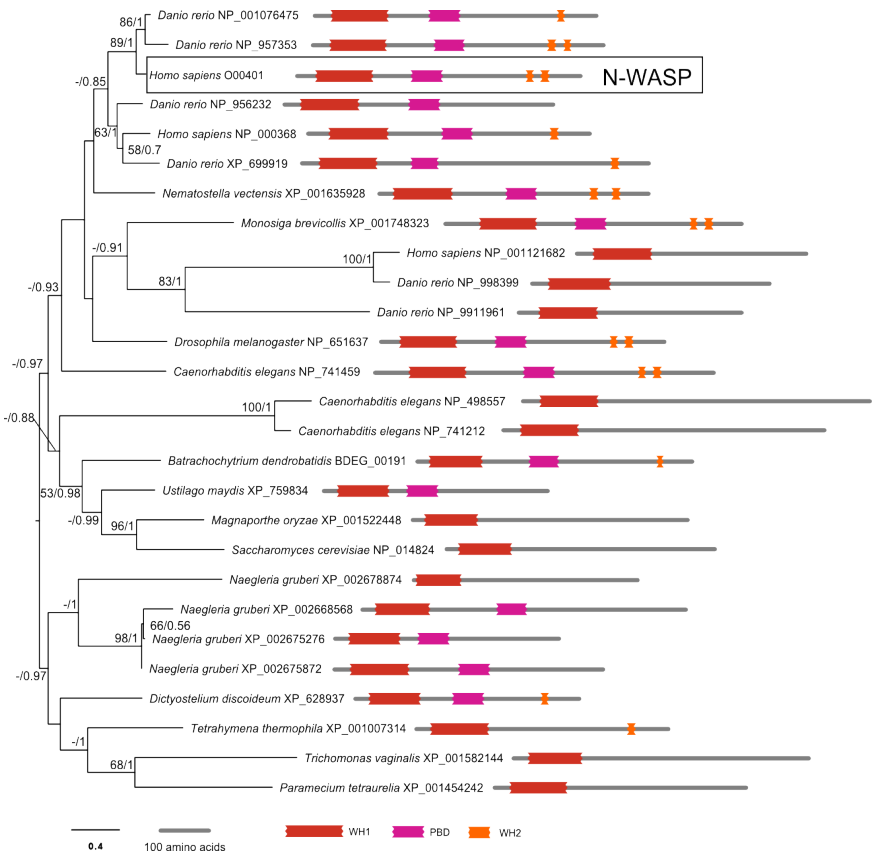


Figure 4.16 Phylogenetic tree of N-WASP. The analysis is based on the alignment of the WH1 protein domain. Tree represents the sub-branch comprising N-WASP and representative outgroup species. Overall poor resolution of the tree means analysis is inconclusive as to origin of the N-WASP paralogue.

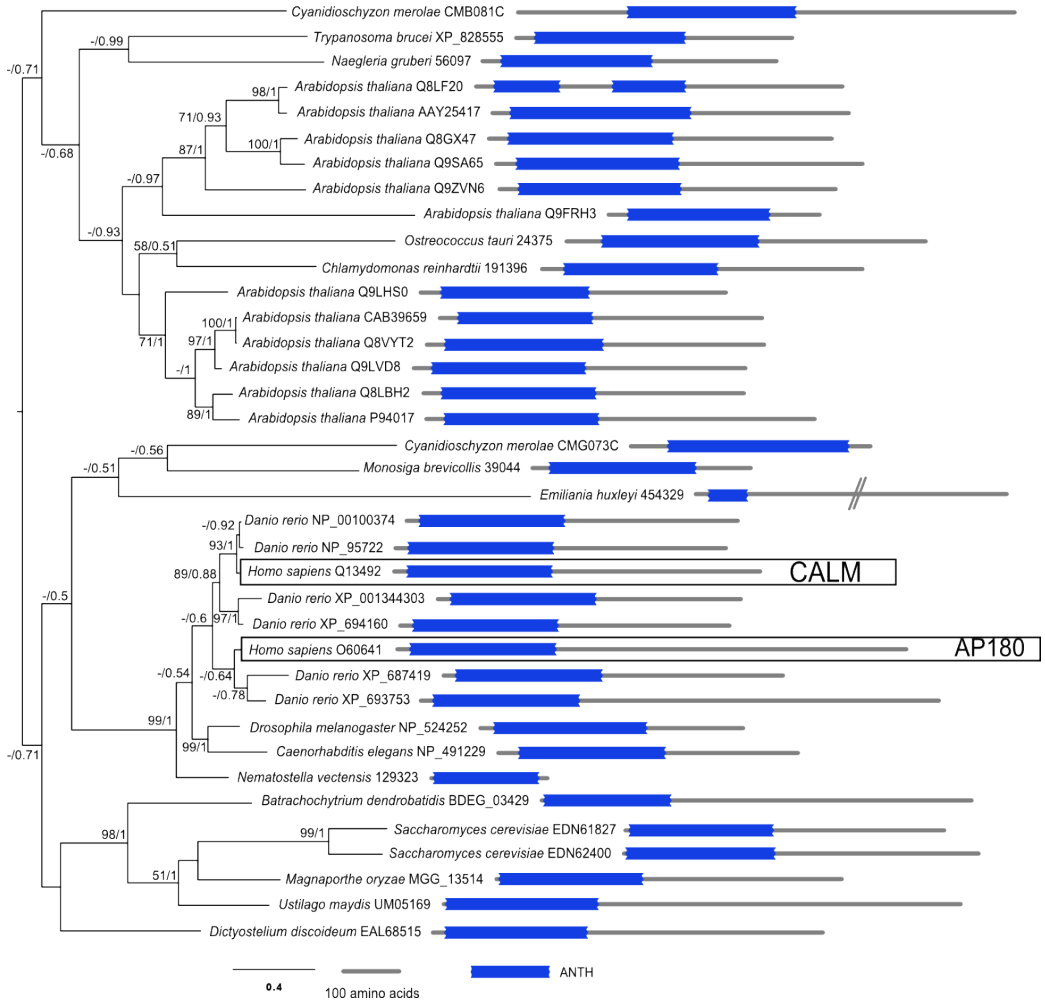


Figure 4.17 Phylogenetic tree of AP180 and CALM. Phylogeny highlights vertebrate-specific gene duplications which produced AP180 and CALM, although the AP180 clade is weakly supported. Several *Arabidopsis thaliana*-specific duplications are also indicated

insufficient to carry out a phylogenetic analysis sampling both protein families. HIP1/HIP1R proteins were thus analysed separately. Overall the HIP1/HIP1R phylogeny constitutes a strongly supported and well resolved unikont phylogeny (Figure 4.18). Amphiphysins and endophilin were analysed together as they share the same protein domain structure and a sufficient degree of sequence similarity for an inclusive multiple sequence alignment. The phylogeny suggests, albeit with weak statistical support, that an ancestral amphiphysin/endophilin gene duplicated prior to the

diversification of metazoans from Fungi, producing the two sub-families (Figure 4.19). Within the endophilins, further duplications occurred at the base of the metazoans and of the vertebrates, while the endophilin gene was independently lost in fungal taxa *Ustilago maydis* and *Saccharomyces cerevisiae.* Within the amphiphysins there was a fungal specific duplication in addition to the vertebrate-specific duplication. These duplications receive moderate support values in the phylogeny (Figure 4.19). For both the endophilin and amphiphysin clades, inparalogues were confirmed for homologues in non-vertebrate opisthokonts, whereas the clade comprising amoebozoans *Dictyostelium discoideum* and *Entamoeba histolytica* was defined as an outparalogue group. The phylogenetic tree of FCH proteins is not as robustly resolved as the amphiphysin/endophilin tree (Figure 4.20). Therefore, while highlighting the vertebrate specific gene duplications which produced TOCA-1, FBP17 and CIP4, the analysis did not resolve deep-level evolutionary relationships. Inparalogues could only be confirmed for holozoan taxa whereas other non-holozoan homologues were conservatively defined as outparalogues.

4.3.1.5 Taxonomic distribution of tuba, β-arrestins and PTB proteins

Tuba is a protein characterised by a central RhoGEF-BAR protein domain core, surrounded by SH3 domains at the N- and C-terminal. Homologues were only found in the opisthokonts, although the SH3 domains are absent in the Fungi sampled. The gene encoding tuba was not found in *Saccharomyces cerevisiae* and *Drosophila melanogaster,* which is likely due to secondary loss (Figure 4.21). Homologues of β arrestins and the ARH/disabled 2/numb group, were only found in the holozoan taxa sampled in this study (Figure 4.2). The β arrestin phylogeny reveals three *Homo*

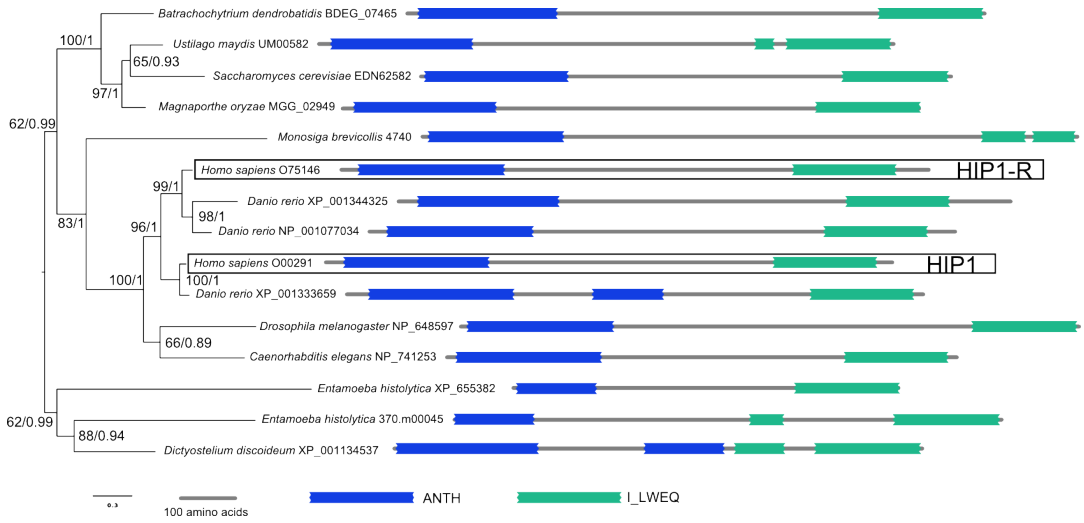
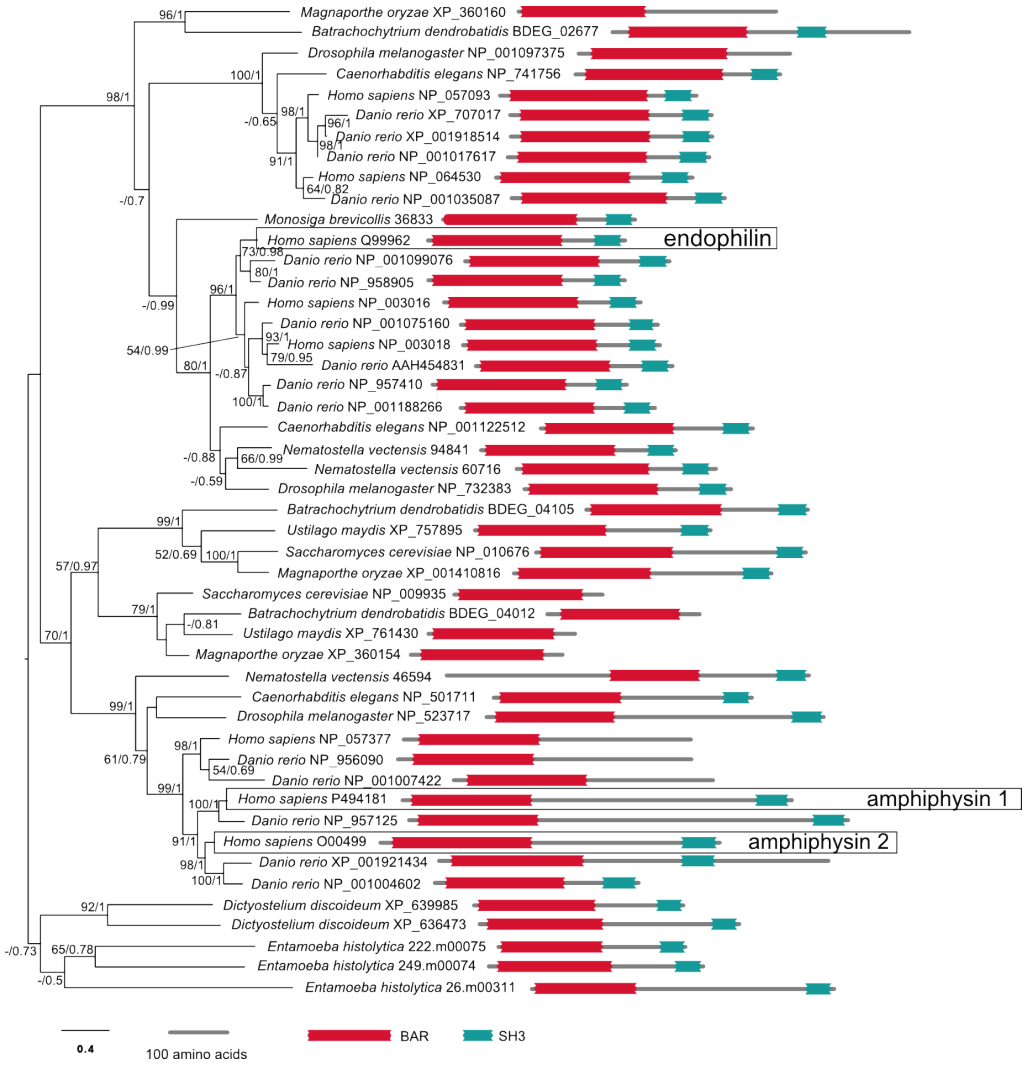


Figure 4.18 Phylogenetic tree of HIP1 and HIP1R. Phylogeny constitutes a strongly supported unikont phylogeny.



(previous page) **Figure 4.19 Phylogenetic tree of amphiphysins and endophilin.** A *Monosiga brevicollis* putative homologue (34505 in JGI) was discarded because of long branch. The tree was rooted with amoebozoan sequences, producing a topology which suggests amphiphysin and endophilin arose as a result of a gene duplication at the root of the opisthokonts. It also indicates, with moderate statistical support, that there was a holozoa-specific endophilin duplication and a fungal-specific amphiphysin duplication.

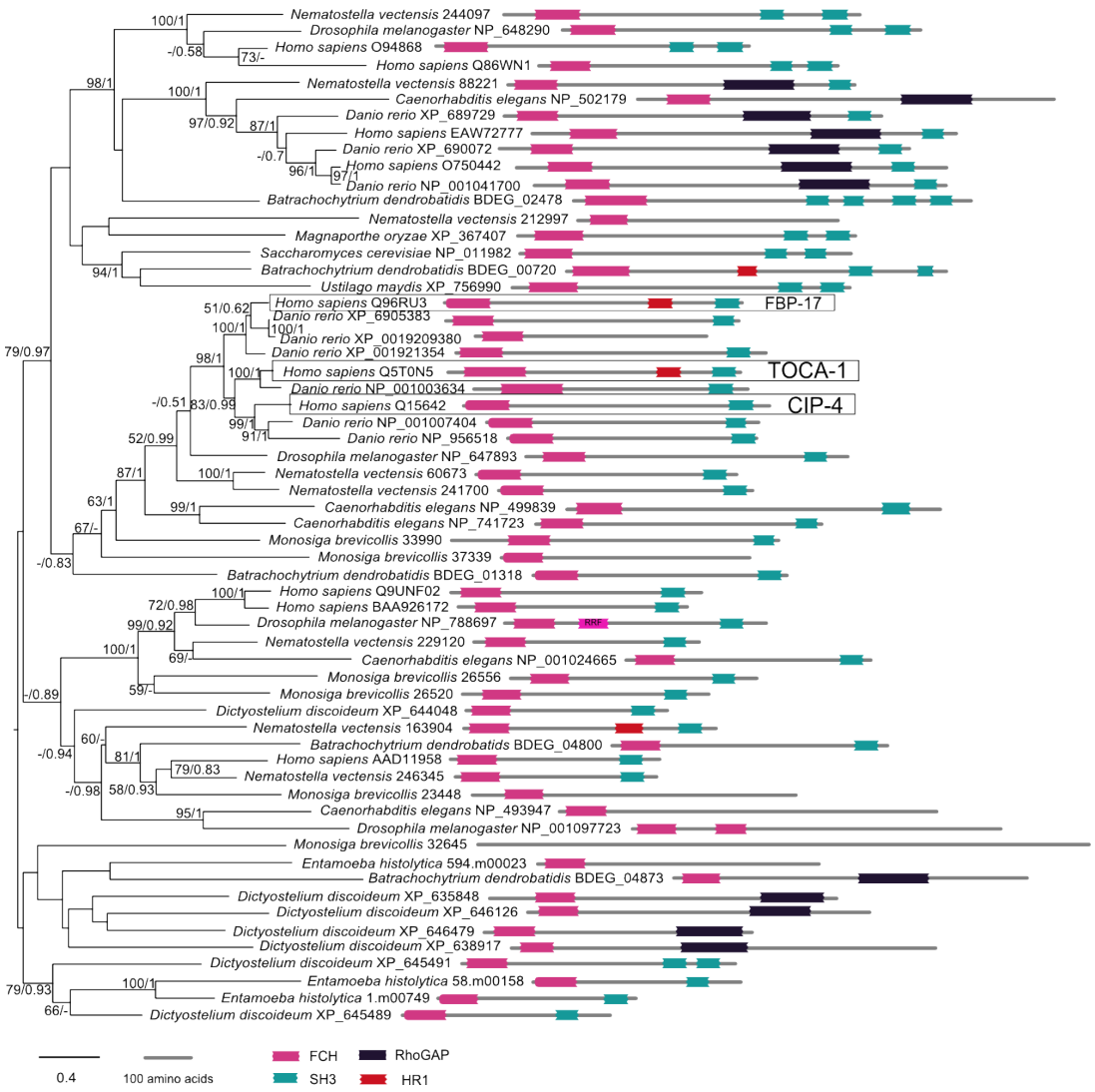


Figure 4.20 Phylogenetic tree of CIP4, FBP17 and TOCA-1. Analyses were based on alignment of the FCH domain. Tree topology indicates CIP1, FBP17 and TOCA-1 resulted from verterbate-specific duplications. Deep-level branching order is not resolved. Inparalogy confirmed only for holozoan taxa.

sapiens/Danio rerio specific duplications and three *Drosophila melanogaster* specific duplications. In the phylogeny, the β arrestins homologue from *Mus musculus* were added to help define the four vertebrate paralogues (Figure 4.22). ARH, disabled 2 and numb belong to the protein family characterised by the phospholipid binding PTB domain. Phylogenetic analyses of a diverse sample of PTB proteins suggests that the disabled 2 and the numb paralogues arose at the base of the Metazoa, while the ARH lineage arose in vertebrates. The branching position of the six *Monosiga brevicollis* PTB homologues suggests they are outparalogues of all three query proteins (Figure 4.23).

4.3.2 Evolution of the CME-I functional repertoire

The phylogenomic study presented in section 4.3.1 focused on the presence or absence of homologous genes in a selection of diverse eukaryotic predicted proteomes. The study was complicated by multiple gene duplications and the presence of diverse paralogues which may mediate different functions (Koonin, 2005). The apparently straightforward question "is protein A encoded in genome X?" should thus be amended to "is an orthologue of protein A is encoded in genome X?" and if a paralogue is identified, a further question is added: "what is the evolutionary relationship between the two paralogues?" Figure 4.2 summarises database searching and the phylogenetic analyses carried out to answer these questions.

Another approach is changing the question from "is protein A encoded by genome X?" to "is the function associated with protein A encoded by genome X?". This logic is effective because it circumvents the issues of multiple paralogues. By studying the phylogenetic distribution of protein domain architectures, it is possible to pinpoint the

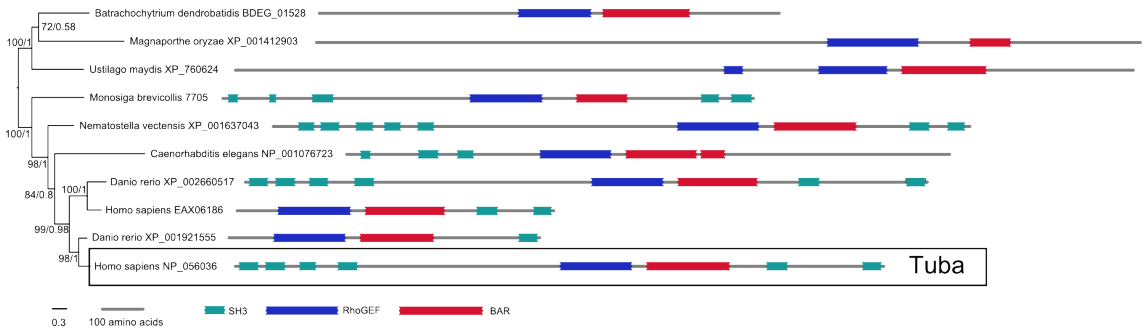


Figure 4.21 Phylogenetic tree of tuba. The phylogeny is strongly supported and suggests the characteristic SH3 domains at either terminal of tuba were a holozoan innovation. Note that tuba was not found in *Saccharomyces cerevisiae* and *Drosophila melanogaster*.

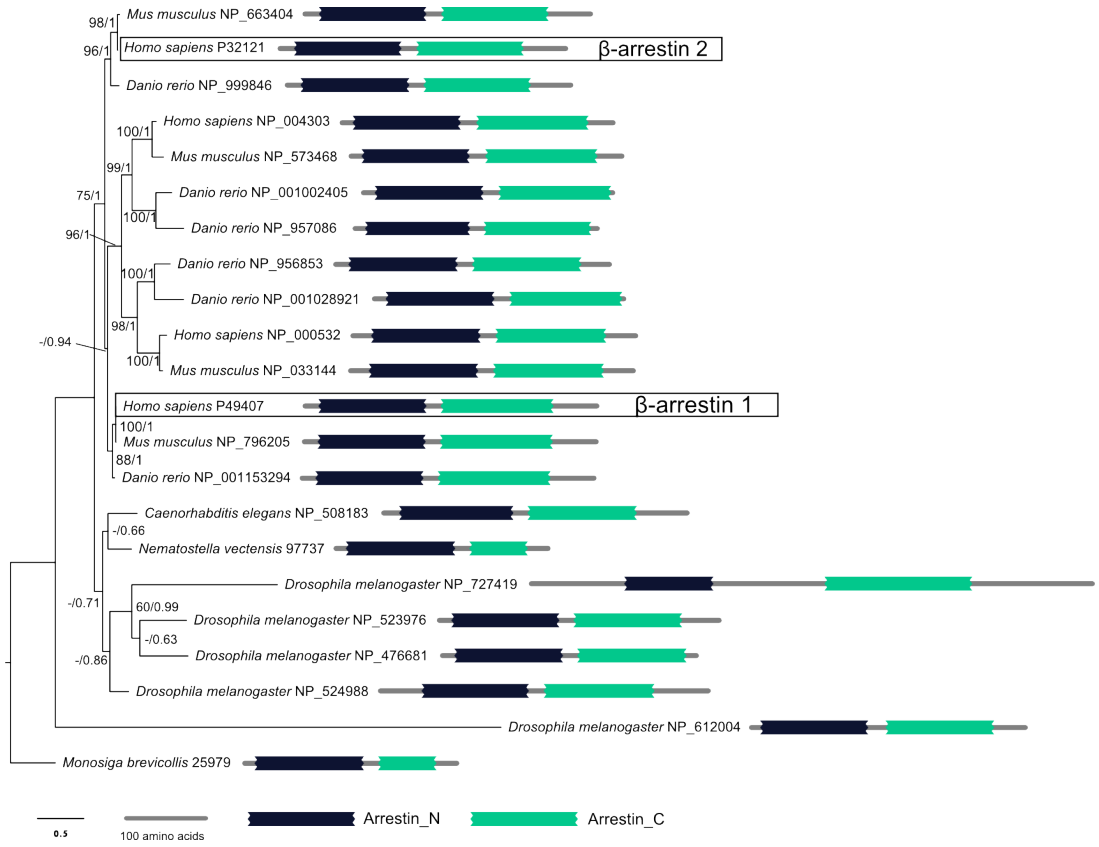


Figure 4.22 Phylogenetic tree of β-arrestins. The *Mus musculus* homologues of β-arrestins were included to help resolve clades which resulted from three verterbate-specific gene duplications. Note also three possible *Drosophila melanogaster*-specific gene duplications.

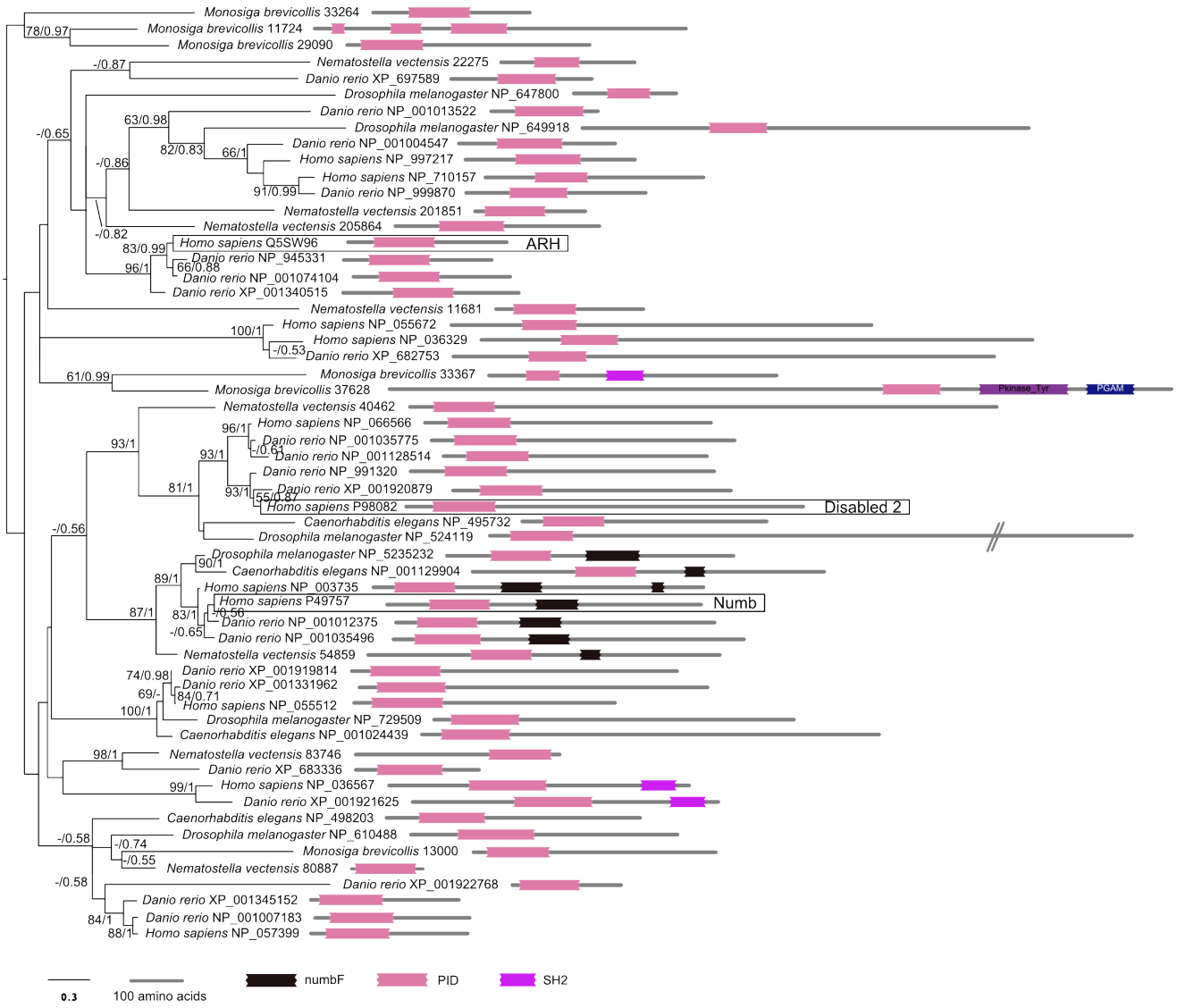


Figure 4.23 Phylogenetic tree of ARH, disabled-2 and numb. The results from the phylogenetic analyses robustly support a metazoan origin of disabled-2 and numb, and a verterbrate origin of ARH. Deep-level branching order is poorly resolved. The six *Monosiga brevicollis* sequences are conservatively classified as outparalogues of the query proteins.

putative origin of a protein domain or of a protein domain rearrangement, based on fundamental assumptions regarding the evolution of complex characters (Le Quesne, 1974). For instance, ABP1 has an N-terminal actin de-polymerisation factor (ADF) domain and a C-terminal SH3 domain. Since all predicted proteomes we sampled feature at least one protein with the ADF domain (except for *Giardia*) but the ADF-SH3

protein domain architecture phylogenetically distributes only to a monophyletic group consisting of unikonts, ADF is predicted to be eukaryotic cenancestral (present in the LCEA) domain but that ADF-SH3 is a shared derived character of the unikonts. A model for the origin of protein domains architectures involved in CME was thus created.

4.3.2.1 Mapping protein domain architectures of the CME-I network to the eukaryotic tree of life

For a section of the CME-I network, it was uncomplicated to predict likely origin of protein domain architectures. This is because, for most of these protein families, a single, typical protein domain architecture was conserved in all the sequences sampled for the phylogenetic analyses, or, if variation in protein domain architecture was detected, it was minor and distributed to clades with high statistical support. This was found for clathrin heavy and light chain, the four AP2 subunits, β arrestins, PTB proteins (ARH, disabled 2 and numb), the AP180/CALM family, the HIP1/HIP1R family, epsins, N-BAR proteins (amphiphysins and endophilin) and FCH proteins (TOCA-1, FBP17 and CIP4) (see Figures 4.3-9, 4.17-20, 4.22 and 4.23).

For the rest of the CME-I network, the phylogenetic distribution of protein domain architectures suggests major protein domain rearrangements have occurred. These may include gene fusions, protein domain losses, and re-shuffling of protein domain combinations. In some cases, the unique protein domain architectures do not distribute to monophyletic groups with support. Here, the evidence on the evolution of these protein families is analysed to assess whether these protein domain rearrangement

events can be mapped on the eukaryotic phylogeny.

4.3.2.2 Evolution of protein domain architectures in epsins, SNX9, dynamins and synaptojanins

The epsins, SNX9, dynamin and synaptojanin phylogenomic studies all described complex evolutionary histories characterised by multiple gene duplications and protein domain recombinations (Figure 4.9-12). The ENTH domain of the epsins was found in all the genomes sampled, which suggests it was acquired in the LCEA. However, in Metazoa and Fungi, an N-terminal ENTH domain was accompanied by a C-terminal ubiquitin interaction motif (UIM) (Figure 4.9). This suggests the ENTH-UIM combination was acquired in the last common opisthokont ancestor (LCOA) (Figure 4.9). Together, these results confirm previously reported data on the evolution of epsins (Field *et al*., 2007; Gabernet-Castello *et al*., 2009).

The SNX9 protein is characterised by the following protein domain structure: an N-terminal SH3 domain which interacts with the PRDs of endocytic dynamins and N-WASP (Shin *et al*., 2007), and a C-terminal combination of a phosphoinositide binding PX and a BAR-related domain (known as BAR_3_WASP in PFAM and as BAR_SNX9 in CDD), which together mediate membrane remodelling in the late stages of CME (Pylypenko *et al*., 2007). This protein domain combination is present in all holozoan taxa sampled as well as fungal taxa *Batrachochytrium dendrobatidis* and *Ustilago maydis* (Figure 4.12). The SNX9 phylogeny is poorly resolved and does not support monophyly of proteins containing this domain structure (Figure 4.12) but it is based on the global alignment of the small sized PX domain (~90 amino acids) which means the

evolutionary signal in the data-set is limited. However, under Dollo parsimony, the best interpretation of this gene family distribution data is that the SH3-PX-BAR domain was acquired in the LCOA and then lost in the ascomycete Fungi sampled (Figure 4.24).

The dynamin paralogues which are experimentally shown to be involved in the vesicle scission module of CME are frequently referred to in the literature as 'classical dynamins' (Praefcke & McMahon, 2004). These dynamins are distinguished from other dynamin sub-families such as dynamin-like or dynamin-related proteins, Mx proteins, profusions and guanylate-binding proteins which cover a variety of cellular functions including mitochondrial division, cytokinesis and pathogen resistance (Praefcke & McMahon, 2004). The domain structure that characterises 'classical dynamins' consists of an N-terminal GTPase domain (known in PFAM as 'Dynamin_N'), a middle Dynamin_M domain followed by the phosphoinositide binding PH domain, the GTPase effector domain (GED) and finally a C-terminal proline rich domain (PRD) (Praefcke & McMahon, 2004). PRDs are not included in the CDD or PFAM databases, but they are recognisable by high proline content. In this study, PRDs are assigned to protein regions with proline contents equal or higher than 25%. The protein domain analysis of the dynamin homologues indicates that the Dynamin_N-Dynamin_M-GED domain combination is present in all the eukaryotic taxa. In contrast, the presence of a PH domain between the Dynamin_M and GED domains is specific to metazoans, choanoflagellates, and two *Arabidopsis thaliana* sequences (NP_172500 and NP_176170, Figure 4.11). Finally, the PRDs, which are instrumental in mediating interactions with endocytic proteins such as amphiphysins, tuba, SNX9 and ABP1 are only present in choanoflagellates and metazoan organisms. The complete domain

structures of classical dynamins are therefore present only in choanoflagellates and metazoans, which suggests it was acquired in the last common holozoan ancestor (LCHA) (Figure 4.24). It is interesting to note that land plants feature the PH domain between the Dynamin_M and GED domains, a feature otherwise unique to holozoans dynamins and absent in every other taxon sampled including green and red algae. The dynamin phylogeny does not support monophyly of the two *Arabidopsis thaliana* sequences with the holozoan homologues (Figure 4.11) although the relationship of classical dynamins with its sister clade is poorly resolved. The evidence concerning the origin of the *Arabidopsis* PH domain is inconclusive.

Synaptojanins have two inositol phosphatase domains: the N-terminal 'Syja_N' domain which can hydrolyse phosphoinositides PI(3)P, PI(4)P and PI(3,5)P_2 to PI in vitro (Guo *et al*., 1999), and the central 'Exo_endo_phos' domain which hydrolyses PI(4,5)P_2 and PI(3,4,5)P_3 (Cremona *et al*., 1999). The analysis of these domains show that every organism considered in this study includes in its genome at least one gene that encodes a protein with the Syna_N domain and at least one gene that encodes a protein with the Exo_endo_phos domain. However, the combination of the two domains is not universal. Indeed, the analyses show that it is present only in opisthokonts, *Dictyostelium discoideum* and alveolates (Figure 4.10). This phylogenetic distribution can be explained by a horizontal gene transfer between opisthokonts and alveolates, by homoplasy, or by multiple gene loss in Archaeplastida, metamonads, discicristatae, heterokontophytes and haptophytes. The phylogenetic analysis of synaptojanins is poorly resolved at deep nodes and does not support any of the three evolutionary scenarios (Figure 4.10).

4.3.2.3 Evolution of the EPS15/EPS15R and intersectin protein domain architectures

EPS15 and EPS15R feature three N-terminal repetitions of the EH protein domain. Results show that this trait is found in predicted proteomes in a wide diversity of eukaryotes, and every predicted proteome searched includes at least one protein with at least one EH domain suggesting it is eukaryotic cenancestral feature. Intersectins share the N-terminal EH domain repetitions of EPS15 and EPS15R but have additional structural features representative of a composite functional repertoire. These consist of a central repetition of 5 SH3 domains and a C-terminal domain complex, made up of a RhoGEF (also known as Dbl Homology), a PH domain and a C2 domain, which suggests a role as a guanine exchange factor (GEF) for Rho-like GTPases (Pucharcos *et al*., 2000). The full intersectin domain architecture is present only in *Homo sapiens* and *Danio rerio* (Figure 4.14). In the rest of the metazoan taxa intersectins do not feature the C-terminal GEF complex. Interestingly, in the fungal species *Batrachochytrium dendrobatidis* and *Ustilago maydis* proteins were found with domain structures featuring two N-terminal EH domains, two central SH3 domains, and a C-terminal RhoGEF-PH domain complex (in *Batrachochytrium* the PH domain is not present). In ascomycete fungi, choanoflagellates, and the rest of the eukaryotic taxa sampled in this study, intersectin specific structural features are absent (Figure 4.14). This suggests that the intersectin protein domain architecture was acquired in the LCOA, with secondary loss of the SH3 domains and GEF unit in choanoflagellates and ascomycetes fungi , and secondary loss of the GEF unit in invertebrate metazoans. However, the phylogenetic tree of intersectins is poorly resolved, so these putative protein domain rearrangements

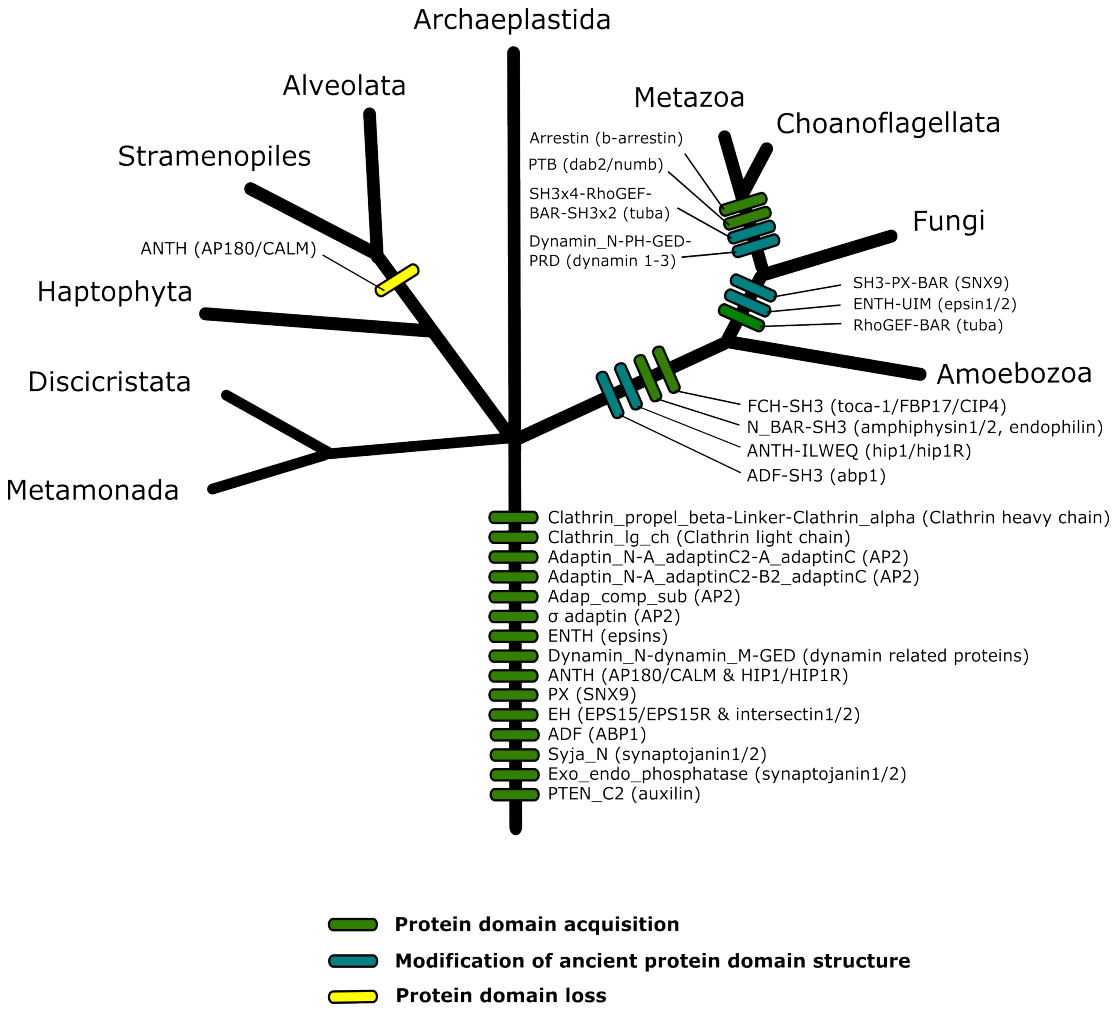


Figure 4.24 Putative origin of the CME protein domain repertoire. The acquisition of protein domain is mapped to a schematic model of the eukaryotic tree of life based on current consensus and includes 10 major phyla. Green bars represent putative acquisition of novel protein domains, teal bars represent putative acquitision of a novel protein domain combination, and red bars represent putative loss of ancient protein domains.

cannot be mapped on the eukaryotic tree with confidence.

4.3.2.4 Evolution of N-WASP and auxilin protein domain architectures

The protein domain architecture of N-WASP consists in three protein domains recognised by PFAM and CDD: an N-terminal WASP homology 1 (WH1) domain (also

known as EVH1) which is bound by WASP interacting proteins (WIPs) to stabilise conformation during inactivity (Svitkina, 2007), a central P21-Rho-binding (PBD) domain (also known as CRIB) which binds Cdc42 or Rho-like small GTPases (G. Thompson *et al*., 1998), and two C-terminal WASP homology 2 (WH2) domains, which are in fact very short conserved motifs, made of approximately 18 amino acid, which bind to actin monomers and to the end of polymerising actin filaments (Co *et al*., 2007). The combination of the three functional units was found in the metazoan and the choanoflagellate taxa, in chytrid *Batrachochytrium dendrobatidis*, in amoebozoan *Dictyostelium discoideum* and in ciliate *Tetrahymena thermophila* even though for the latter the domain recognition of PBD and WH2 were statistically very weak in PFAM (E-values respectively of 0.28 and 0.027) and not detected in CDD. In addition, the combination of WH1 and PBD domains was found in basidiomycete *Ustilago maydis* and heterolobosean *Naegleria gruberi.* Finally the WH1 domain was found on its own in ascomycetes taxa, in *Paramecium tetraurelia* and *Trichomonas vaginalis*. These data show that, outside the holozoan clade, the taxonomic distribution of the three N-WASP functional domains is sparse and mostly absent. However, because of the size of the WH2 and the PBD domains (~18 amino acids and ~40 amino acids respectively), it is difficult to recognise them with HMMs or PSSMs, even in closely related homologous sequences, so in those genomes where a solitary WH1 domain was detected, these regions of the gene may have simply diverged enough not to be recognised by domain recognition tools. Nonetheless, the origin of the N-WASP protein domain architecture could not be pinpointed because of the its sparse phylogenetic distribution of, and the weak topological support in the N-WASP phylogenetic tree.

Auxilin has an N-terminal regions with high similarity to the phosphatase and C2 domain of the PTEN gene. This protein domain is identified in PFAM and CDD as PTEN_C2, and is necessary for binding to phosphoinositides (Massol *et al*., 2006). The DNAJ protein domain at the C-terminal of auxilin, is of prokaryotic origin and interacts with HSC70 to mediate the uncoating of clathrin coated vesicles (Lemmon, 2001). The PTEN-like domain is present in Metazoa, choanoflagellates, *Ustilago maydis, Magnaporthe oryzae*, amoebozoans, Viridiplantae, ciliates, *Toxoplasma gondii*, heterokonts, haptophytes, *Naegleria gruberi* and metamonads (Figure 4.13) which suggests it was encoded by a gene in the LCEA (Figure 4.24). However, the combination of the PTEN-like domains and DNAJ is only present in *Homo sapiens, Danio rerio*, *Drosophila melanogaster* and *Caenorhabditis elegans*.

4.3.2.5 Mapping expansion of the CME functional repertoire to the eukaryotic tree of life

The putative origins of most of the protein domain and protein domain combinations that mediate the functions necessary for CME were mapped to a consensus eukaryotic tree of life (Figure 4.24) demonstrating five steps in the evolution of CME complexity. The five steps are summarised:

1) Clathrin_propel_beta-Linker-Clathrin_alpha (present in clathrin heavy chain), Clathrin_lg_ch (clathrin light chain), Adaptin_N-α_adaptinC2-α_adaptinC (AP2 α subunit), Adaptin_N-α_adaptinC2-β2_adaptinC (AP2 β subunit), Adapt_comp_sub (AP2 μ subunit), Clat_adaptor_σ (AP2 σ subunit), ANTH (AP180/CALM), ENTH (epsins), PX (SNX9), Dynamin_N-Dynamin_M-GED (dynamins), EH (EPS15/EPS15

& intersectins), ADF (ABP1), Syna_N (synaptojanins), Exo_endo_phos (synaptojanins) and PTEN_C2 (auxilin) were acquired in the LCEA.

2) BAR-SH3 (amphiphysins and endophilins), FCH-SH3 (TOCA-1/FBP17/CIP4), ADF-SH3 (ABP1) and ANTH-ILWEQ (HIP1/HIP1R) were acquired in the LCUA.

3) ENTH-UIM (epsins), RhoGEF-BAR (tuba) and SH3-PX-BAR (SNX9), were acquired in the LCOA.

4) Arrestin_N & Arrestin_C (β arrestins), PTB (ARH, numb and disabled-2), Dynamin_N-Dynamin_M-PH-GEF-PRD (dynamin 1/2/3), SH3-RhoGEF-BAR-SH3 (tuba) were acquired in the LCHA.

5) Finally, the PTEN_C2-DNAJ domain combination was acquired in the last common ancestor of *Homo sapiens, Danio rerio, Drosophila melanogaster* and *Caenorhabditis elegans.*

The putative origins of the EH-SH3-RhoGEF-PH-C2 structure of intersectins, the WH1-PBD-WH2 structure of N-WASP and the Syja_N-Exo_endo_phos structure of synaptojanins are contentious because of their sparse taxonomic distributions and poorly resolved phylogenies. With current data, their origin cannot be pinpointed with confidence.

4.3.3 Reconstructing expansions in complexity of the CME-I network

4.3.3.1 Model for studying changes in complexity of the CME-I network

Taken together, the data presented in sections 4.3.1 and 4.3.2 depict a complex picture regarding the origin and evolutionary diversification of CME. The phylogenomic data summarised in Figure 4.2 reveals that true orthology of the majority of proteins which make up the CME-interactome (CME-I) in *H. sapiens* can only be confirmed in *Danio rerio*. Even the genes encoding for the most conserved of CME-I proteins have undergone multiple duplications, losses, and recombination of functional domains, as can be observed for instance in the epsins, synaptojanins and auxilins . The evolution of multiple paralogues may reflect the evolution of diverse and specialised functions, a process which has been proposed to be a key aspect in the evolution of specificity in the eukaryotic endomembrane system (Dacks *et al*., 2009). The results from the protein domain study presented in section 4.3.2 indicate that 15 of the 30 functional protein domain architectures present in the CME-I proteome were acquired in the LCEA, while 4 were acquired in the LCUA, 3 were acquired in the LCOA, 4 were acquired in the LCHA, and only 1 protein domain structure is specific to metazoans (the remaining three could not be mapped to the eukaryotic tree with confidence) (Figure 4.24). Also, out of the 11 protein domain structures acquired in the LCUA, the LCOA and the LCHA, 7 were a result of recombination of existing protein domains and only 4 were acquired *de novo*. This suggests that the molecular functions associated with all but 4 of the protein domains necessary for CME was already present in the LCEA. Therefore, the evolutionary history of CME is multifaceted, because it is both conserved and ancient while in contrast has many derived forms. This can be explained by an evolutionary model whereby CME was already complex in the LCEA, but then

diversified and gained further complexity via gene duplications and rearrangements of protein domains architectures.

To investigate this evolutionary process, I referred to the data on known interactions within the CME-I network reviewed in chapter 3, and related it to the evolutionary ancestry data to create a model describing putative CME-I network at different stages of eukaryote evolution. The identification of novel interactions as shared derived characters was key to the construction of this model. The rationale explained in sections 3.1 and 3.2.3 was applied to identify putative origins for novel interaction (Appendix 3). This data was combined with the data on protein domain architecture evolution to construct a model of CME-I network evolution.

A difficulty encountered in formulating these networks is that binding motifs necessary for certain types of protein-protein physical interactions are only a few amino acids in length and are rarely recognised by protein domain recognition. For instance, the AP2 complex can bind to clathrin heavy chain via a LLNLD motif on the linker region of the β subunit, whereas epsin 1 binds to clathrin via either a LMDLAD or a LVDLD motif in the central or C-terminal region of the protein, respectively (Dell'Angelica, 2001). This was found for mammalian systems, but no data exists regarding clathrin binding sites in any other eukaryotic system. It is therefore difficult to investigate whether a certain motif existed at a given stages of eukaryotic evolution. The model therefore relies on the assumption that a hypothetical ancestral protein which features the same protein domain architecture as its extant homologue, will also feature the same clathrin and AP2 binding motifs as its extant homologue.

4.3.3.2 The CME-I network in the LCEA

The model predicts the LCEA featured clathrin and the AP2 complex which interacted with the following hypothetical proteins: an epsin ancestor featuring an ENTH domain and clathrin and AP2 large subunit appendage binding sites; an AP180 ancestor featuring an ANTH domain, a clathrin and a AP2 large subunit appendage binding site; an EPS15 ancestor with EH domains, a clathrin and a AP2 large subunit appendage binding sites. Interaction with $PI(4,5)P_2$ phosphoinositide was mediated by the AP2 large α subunit, the ENTH domain and the ANTH domain. Interaction with YXXØ and [DE]XXXL[LI] internalisation signals was mediated by the AP2 μ subunit. The uncoating process was putatively mediated by three hypothetical proteins, each with one of the three phosphatase domains Syja, Exo_endo_phos and PTEN_C2, and by domains HSC70 and DNAJ. In addition, the model suggests the actin binding ADF domain and the Dynamin_N-Dynamin_M-GED domain structure was present in the LCEA, but the data do not suggest these domains were integrated in the CME-I network. The putative LCEA CME-I network is depicted in Figure 4.25.

4.3.3.3 Expansion of network complexity in the LCUA CME-I

The first step in complexity is traceable to the LCUA. The network analysis predicts the acquisition of the following novel protein families: a hypothetical HIP1/HIP1R ancestor featuring the ANTH-ILWEQ protein domain combination which interacted with $PI(4,5)P_2$ via the ANTH domain, interacted with filamentous actin via the ILWEQ domain, and featured clathrin and AP2 large subunit appendage binding sites; a hypothetical amphiphysin and endophilin ancestor featuring the BAR-SH3 protein domain combination which bound the plasma membrane with the BAR domain and the

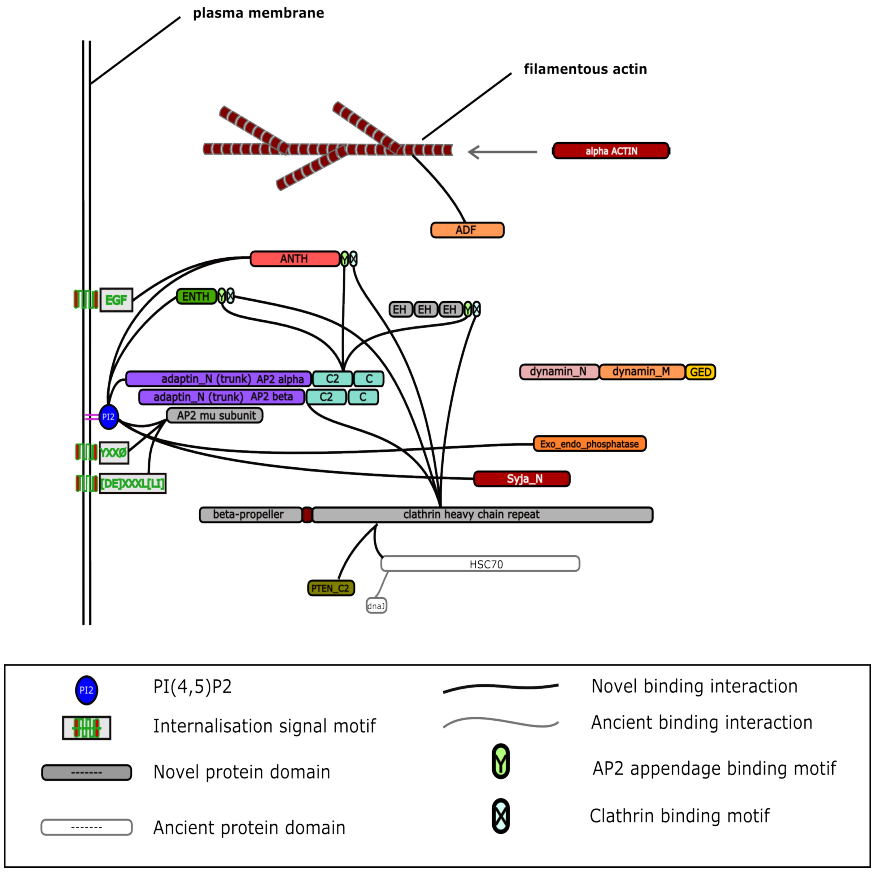


Figure 4.25 Connectivity diagram depicting the putative CME-I network topology in the last common eukaryotic ancestor. In this Figure and in Figures 4.26-28, novel protein domains are shown in coloured bars while ancient protein domains are white bars with grey margins. The X and the Y bars represent respectively putative clathrin binding motifs and putative AP2 appendage binding motifs. Novel binding interactions are shown with black lines. Ancient interactions are depicted by grey lines. Green writing in grey box represent novel internalisation signals. PI2 represents phosphoinositide PI(4,5)P2.

PRD of N-WASP via the SH3 domain and featured clathrin and AP large subunit appendage binding sites; a hypothetical TOCA-1/FBP17/CIP4 ancestor featuring the FCH-SH3 protein domain combination, which bound plasma membrane with the BAR domain and the PRD region of N-WASP via the SH3 domain. Finally, the ADF acquired a C-terminal SH3 domain but no novel interactions were acquired as a result. The putative LCUA CME-I network is depicted in Figure 4.26.

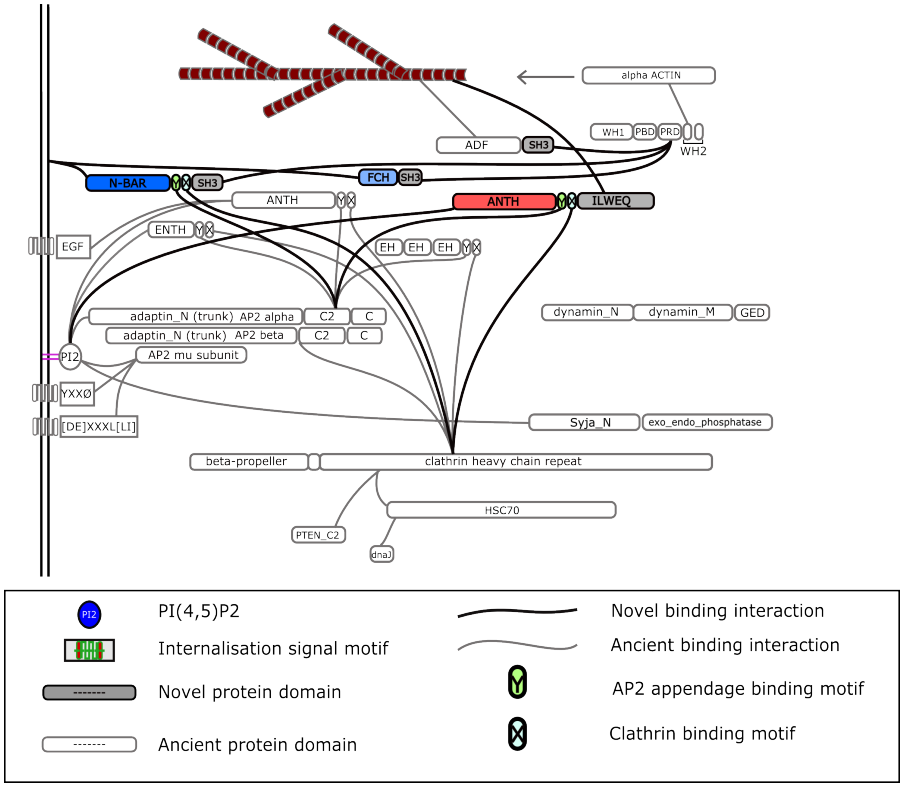


Figure 4.26 Connectivity diagram depicting the putative CME-I network in the last common unikont ancestor. This model predicts that membrane bending and actin regulation functions studied in mammalian CME were acquired in the LCUA.

4.3.3.4 Expansion of network complexity in the LCOA CME-I

The second step in complexity is in the LCOA where SH3-PX-BAR, RhoGEF-BAR and ENTH-UIM protein domain combinations were acquired. However, these additions only translate into four novel putative interactions within the CME-I: a hypothetical SNX9 ancestor featuring the SH3-PX-BAR interacted with N-WASP via the SH3 domain and featured clathrin and AP2 large subunit appendage binding sites; the acquisition of the UIM domain at the C-terminal of the hypothetical epsin ancestor allowed for interactions with ubiquinated cargo; the acquisition of the RhoGEF-BAR protein domain combination was not associated with novel interactions according to the data used for the model. The putative LCOA CME-I network is depicted in Figure 4.27.

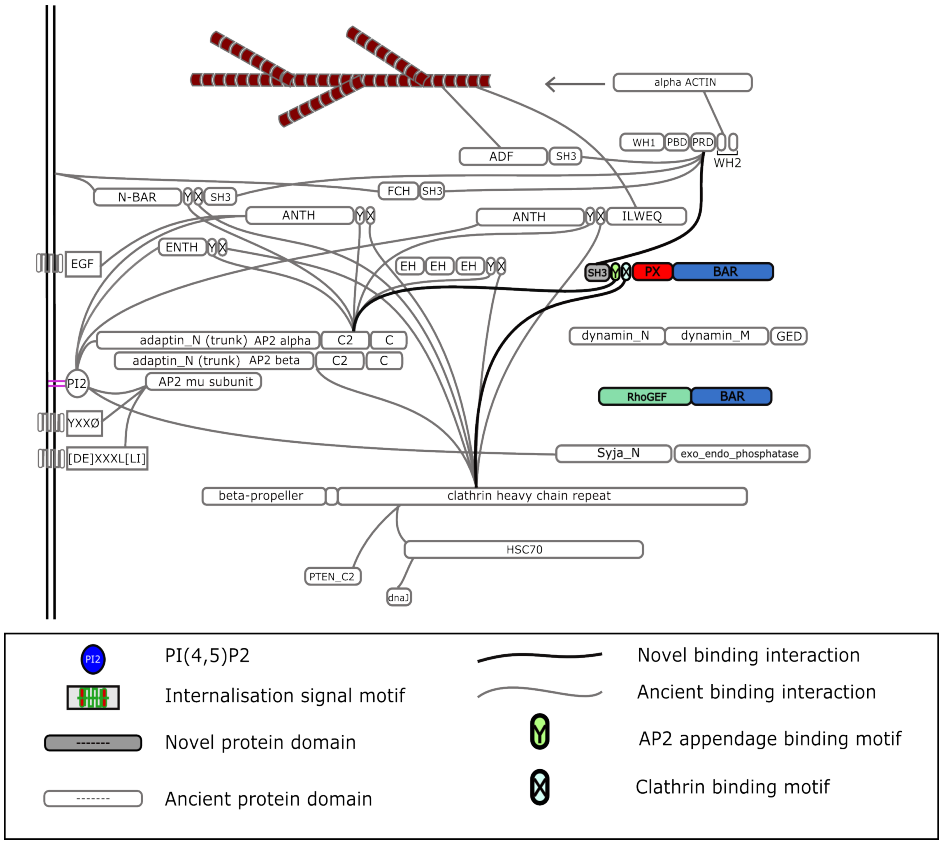


Figure 4.27 Connectivity diagram depicting the putative CME-I network in the last common opisthokont ancestor. The model predicts the SNX9 protein domain architecture was acquired, signifying futher advance in CME actin regulation function. The RhoGEF-BAR domain found in tuba was also acquired.

4.3.3.5 Expansion of network complexity in the LCHA CME-I

The third leap in complexity is in the LCHA. The model predicts the acquisition of a novel form of dynamin featuring the Dynamin_N-Dynamin_M-PH-GED-PRD domain structure, the acquisition of multiple SH3 domains at the N- and C-terminal of the RhoGEF-BAR domain proteins, and the acquisition of the PTB and the arrestin domain which are found in alternative CME monomeric adaptors. These novel protein forms introduced the following novel interactions: the hypothetical dynamin1/2/3 ancestor featuring the Dynamin_N-Dynamin_M-PH-GED-PRD protein domain combination interacted with the SH3 domains of the amphiphysin/endophilin ancestor, of the TOCA-

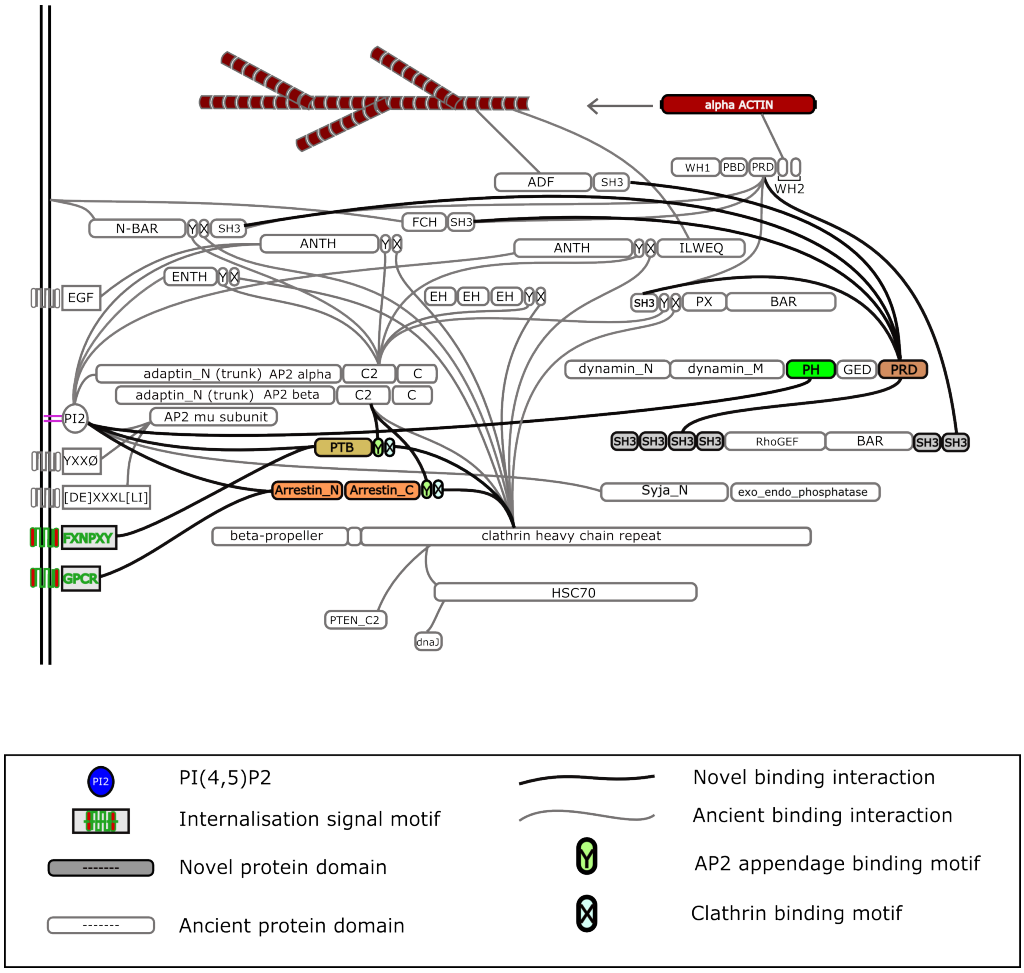


Figure 4.28 Connectivity diagram depicting the putative CME-I network topology in the last common holozoan ancestor. Significant expansion in complexity of the CME-I network was mediated by acquisition of classical dynamin protein domain architecture, the tuba specific SH3 domains and monomeric CME adaptors β-arrestins and PTB proteins (ARH, disabled-2 and numb).

1/FBP17/CIP4 ancestor, of the ABP1 ancestor and of the tuba ancestor via its PRD domain; the hypothetical ancestor of tuba featuring the SH3-RhoGEF-BAR-SH3 domain structure interacted with N-WASP via one of its two C-terminal SH3 domains; the hypothetical ancestor of ARH, disabled-2 and numb featuring the PTB domain interacted with PI(4,5)P_2 and the FXNPXY internalisation signal via the PTB domain and bound to clathrin and AP2 large subunit via putative binding sites; the hypothetical

ancestor of β-arrestin1/2 featuring the arrestin_N-arrestin_C domain structure interacted with PI(4,5)P_2 and G-protein coupled receptors via the arrestin_N domain, and bound to clathrin and AP2 large subunit via putative binding sites. The putative LCOA CME-I network is depicted in Figure 4.28.

4.4 Discussion

In this chapter I presented a study of the evolutionary history of CME. The study is based on three parts: a comparative genomic and phylogenetic analysis of the CME-I, the mapping of putative protein domain acquisition and rearrangements to the eukaryotic tree of life, and the construction of a model depicting CME's expansion of complexity in relation to eukaryotic evolution.

Here, I first discuss the significance of the results in relation to previously published data. Starting with the taxonomic distribution of CME-I components, this study relates in particular to two previously published comparative genomic studies (Field *et al*., 2007; Schmid & McMahon, 2007). Some of the results presented here are thus redundant. For instance the taxonomic distribution of Clathrin heavy chain and the AP2 complex among diverse eukaryotes had already been reported, including the fact that AP2 is missing in *Trypanosoma* and *Cyanidioschyzon* (Field *et al*., 2007). The same study also looked at epsins, dynamins, EPS15 and dab-2, reporting similar results (Field *et al*., 2007). However, the taxonomic distribution of β-arrestins, HIP1, Amphiphysin, FCH proteins, SNX9, intersectins, ABP1, N-WASP, auxilin and synaptojanin, have not been reported before, at least not across eukaryotic diversity. Indeed, the comparative genomic data by Schmid & McMahon (Nature, 2007) concerns most of those proteins,

but its taxon sampling is limited to Metazoa and one apicomplexan protist (*Plasmodium falciparum*), which is not sufficient for a robust eukaryotic study. The study reported in this thesis also uses advanced phylogenetic methods to determine orthology, which is, arguably, a more exhaustive and reliable method than sole analysis of protein domain structure or protein size alone. Thanks to this method, the trend of vertebrate-specific gene duplications was highlighted, which in turn entailed further distinction between closely related paralogues (inparalogues) to distantly related paralogues (outparalogues). Comparing results on the same proteins between previous study and this one, this method proved useful in proteins such as dynamin, which was reported to have orthologues in Fungi, Amoebozoa and *Arabidopsis thaliana* (Field *et al.,* 2007), while our phylogeny robustly indicated that classical dynamins are specific to Holozoa (Figure 4.11). Two previous analyses suggested the dynamins were highly derived, although with a reduced taxon sampling which did not feature basal metazoans, basal fungi or any choanozoans (Elde *et al*., 2005; Miyagishima *et al*., 2008), thus not pinpointing origin at the base of the Holozoa. Therefore, while certainly not without problems - notably the difficulties in determining evolutionary relationships in poorly resolved phylogenies - this approach arguably provides an improved predictive potential for comparative genomic studies such as this one.

Furthermore, by studying the phylogenetic distribution of protein domain architecture predictions, some important protein domain rearrangements were pinpointed on the eukaryotic tree of life, albeit with varying degrees of statistical confidence. These include dynamins, ABP1, tuba, SNX9, HIP1 and epsins. While this method has been used effectively in evolutionary studies of motor proteins such as myosins and kinesins

(Richards & Cavalier-Smith, 2005; Wickstead *et al.*, 2010), it has not been systematically used on CME studies, with the exception of an epsin study where the ENTH-UIM structure of epsins was reported as being opisthokont-specific (Gabernet-Castello *et al.*, 2009). However, tracing the occurrence of protein domain rearrangements is important towards understanding the evolution of CME. This was illustrated in the reconstruction of CME-I network complexity expansion in unikonts and holozoans (Figure 4.26 and 4.28), where protein domain rearrangement are suggested to play an important role.

Finally, by sampling a broad diversity of eukaryotes, and including in the analyses all the paralogues that were similar enough to align with confidence and not produce long branches, some of the phylogenies reported here represent a valuable evolutionary characterisation of the respective protein families. For several CME proteins, phylogenies have been previously reported, often within the format of protein family reviews. However, in those analyses, the taxonomic sampling used is narrow and the phylogenies are inferred with distance and parsimony methods, which are not as successful at resolving phylogenies as parametric, likelihood methods such as the ones used in this thesis. These include phylogenies for β-arrestin (Alvarez, 2008b), BAR domain proteins (amphiphysins, endophilin, tuba and SNX9) (Habermann, 2004), FCH proteins (Ahmed *et al.*, 2010), sorting nexins (SNX9) (Cullen, 2008; Seet & Hong, 2006), ABP1 (Lappalainen *et al.*, 1998) and auxilin (Bai *et al.*, 2010). Phylogenies of other CME proteins, namely AP180/CALM, HIP1/HIP1R, N-WASP and synaptojanins, are reported here for the first time.

With regards to the primary aim set out in the chapter's introduction, it is proposed here that the holistic approach used in this study produces a robust model of the evolutionary history of CME. The starting point is the LCEA cell which, as indicated by previous studies (Field *et al*., 2007; Gabernet-Castello *et al*., 2009), formed clathrin-coated endocytic vesicles with adaptor components such as an early AP2 complex, and ancient relatives of epsins, EPS15 and AP180. The detail that the two dephosphorylating protein domains of synaptojanins (Syja_N and exo_endo_phos) were encoded in the LCEA and were potentially involved in vesicle uncoating in that early system, is added here. But how did CME further evolve into the complex mechanism that is studied in mammalian cell types?

Here I discuss the two stages where CME underwent substantial modifications (Figure 4.26-28). As mentioned in section 4.3.2, two novel protein domains and two novel protein domain architectures were acquired in the LCUA (Figure 4.3). Interestingly, all four acquisitions are either directly or indirectly involved in actin cytoskeleton regulation. The ADF and ILWEQ domains bind filamentous actin directly (Lappalainen *et al*., 1998; McCann & Craig, 1997), while the SH3 domains at the C-terminal of N-BAR and FCH proteins bind the PRD region in N-WASP which in turn activates actin polymerisation (Hartig *et al*., 2009; Yamada *et al*., 2009). The other acquired function is membrane deformation, as both the N-BAR and FCH domains proteins mediate membrane bending (Dawson *et al*., 2006). When considering the putative CME-I in the LCEA (Figure 4.25), no explicit molecular link exists between the core CME components and the actin cytoskeleton, much as there is no explicit plasma membrane bending effector. In addition, no SH3-PRD interactions can be traced to the LCEA.

Taken together, these data suggest that fundamental modifications occurred in the LCUA, whereby the integration of actin cytoskeleton dynamics with core CME machinery was acquired and linked to a novel membrane deformation system. Crucially, the integration of the novel functions was mediated by the SH3-PRD protein-protein interaction which were also acquired in the LCUA. The overall expansion in complexity of the CME-I network in the LCUA also provides strong evidence in favour of unikonts being monophyletic.

The second major set of modifications in CME can be traced to the LCHA. The 'classical' dynamin protein domain structure was acquired, as was the presence of multiple SH3 domains at the N- and C-terminal of a RhoGEF-BAR central region found in tuba (Figure 4.3). These two novel protein domain combinations introduced the system of vesicle scission, and its integration with actin polymerisation, as it is studied in mammalian CME. As with the changes which occurred in the LCUA, the SH3-PRD protein-protein interactions play an important part in these modifications. The PRDs of the novel endocytic dynamins bind to at least five different SH3 domains, namely the ones at the C-terminal of N-BAR, FCH, and ABP1 proteins, and the ones at the N-terminal of tuba and SNX9 (David *et al*., 1996; Salazar *et al*., 2003; Hartig *et al*., 2009; Kessels *et al*., 2001; Shin *et al*., 2007). All five SH3-featuring proteins are also linked to actin polymerisation by binding the PRD of N-WASP (Bu *et al*., 2009; Cestra *et al*., 2005; Dawson *et al*., 2006; Pinyol *et al*., 2007; Shin *et al*., 2007). In the LCHA, the PTB domain found in the ARH/disabled 2/numb family and the arrestin domain found in β-arrestin1/2 were also acquired (Figure 4.24). These two acquisitions are related as both families are known to be alternative monomeric CME adaptors (Traub, 2003). The

acquisition of these two protein families may be linked to an increased requirement of cellular signalling specificity due to cellular diversification in basal holozoan organisms.

The way these results relate to the modular nature of the CME-I network described in Chapter 3 is interesting. The core module, for instance, likely evolved in the LCEA, except for endocytic adaptors β arrestins and PTB proteins which evolved in the LCHA. The membrane bending module evolved in the LCUA, while the actin attachment module evolved mainly in the LCUA, but also in the LCOA and the LCHA. The vesicle scission module evolved in the LCHA with the rise of the 'classical dynamins'. Part of the uncoating module is prokaryotic in origin (HSC70 and the DNAJ protein domain), while the dephosphorylating protein domains Syja_N and Endo_exo_phosphatase evolved in the LCEA. The auxilin paralogue however evolved with the vertebrates. This indicates a correlation between the modular nature of the CME-I network and its evolutionary history.

With regards to the diversity of CME across eukaryotic diversity, this study was not able to provide a significant amount of information as most of the proteins analysed are poorly conserved and may only be present as a divergent paralogue. For instance, the data shows that all non-opisthokont taxa encode dynamin-related proteins, and also proteins that feature the EH domain, proteins with the dephosphorylating domain Syja_N, and, as previously reported, proteins with the ENTH domain (Gabernet-Castello *et al*., 2009). Yet how those proteins work in diverse protists or Plantae needs to be investigated at a cell biological level. Overall, a problem with understanding

endocytic diversity in eukaryotes is the taxonomic bias with which the CME-I network was defined. The literature used refers mostly to studies performed on mammalian cell types, and in a smaller part to studies performed on yeast cells (see Table 3.1 for list of cell types and model organism). As a result, the list of query proteins used as search sequences in the homology searches reflects the diversification and specificity of CME in mammals and yeast but not of other eukaryotic lineages. Considering the split between opisthokonts and amoebozoans for instance, I listed the CME-I components and protein functional domains that were likely acquired in the LCOA, but the analysis does not account for amoebozoan-specific CME-related innovations. The issue of asymmetry will be discussed further in Chapter 7.

A further issue concerns data security. The bioinformatics-based research carried out in this study was informed by a rich literature of cell biological data regarding CME. Therefore, it is important to evaluate how reliable and robust the data really is. As explained in the method section of this chapter, certain criteria were applied in identifying the protein components of the CME-I. These were in short that a protein had to be shown to physically interact with at least two established CME proteins (e.g. clathrin, AP2, epsin, EPS15, AP180, dynamin), and that the role played within CME had to be elucidated by functional studies. For instance, data indicates that amphiphysin binds the appendage of AP2α subunit and clathrin heavy chain via specific motifs, and also that during vesicle formation, amphiphysin recruits dynamin and interacts with N-WASP to modulate the actin cytoskeleton (David *et al.*, 1996; McMahon *et al.*, 1997; McPherson *et al.*, 1996; Slepnev *et al.*, 2000; Yamada *et al.*, 2009). However, the quantity of data available on a specific protein was not taken into account and it may

provide better data security to only consider proteins that have been confirmed by a minimum number of independent studies. The other data security problem is establishing the specificity of selected proteins to CME function. This is especially important given that multiple distinct endocytic pathways have been documented (see section 1.3). The ideal scenario is selecting only proteins for which there is evidence of involvement in CME and evidence for its non-involvement in other endocytic pathway, or even a separate cellular processes. However that is seldom the case. Table 3.2 lists proteins that are involved in clathrin-independent endocytosis and other cellular functions. A trend that was noticed is that those proteins involved in actin cytoskeleton modulation tend to be involved in endocytic pathways other than CME. Evidence shows that this is the case for amphiphysin (Yamada *et al*., 2007), FBP17 (Tsuboi *et al*., 2009), SNX9 (Yarar *et al*., 2007) ABP1 (Dieckmann *et al*., 2010) and N-WASP (Park & Cox, 2009). Furthermore, dynamin is documented to mediate vesicle scission in caveolin-mediated endocytosis and also to be important in phagocytosis (Gold *et al*., 1999; Pelkmans & Helenius, 2002; Sauvonnet *et al*., 2005). Dynamin-related proteins also mediate non-endocytic functions, i.e. organelle division and cytokinesis (Miyagishima *et al*., 2003; Thompson *et al*., 2002). However, even proteins that are typically considered CME markers such as epsin and EPS15, can play a role in caveolin-mediated endocytosis (Sigismund *et al*., 2005), and the AP2 complex is involved in post-endocytic trafficking in ARF6-regulated clathrin-independent endocytosis (Lau & Chou, 2008). Finally it must also be noted that clathrin itself can play a role in a non-endocytic function, i.e. mitosis (Royle *et al*., 2005). In light of this evidence, some of the results presented in this chapter should be considered with caution, and the importance of testing predictions with cell biological studies is iterated.

5 Comparative genomic and phylogenetic study of eisosome components Pil1 and Lsp1

5.1 Introduction

Understanding endocytosis entails the consideration of several molecular mechanisms, such as the deformation of plasma membranes, the formation of vesicle coats, the identification of cargo internalisation signals, the inward transport of the vesicle and the recycling of vesicle coat components (Conner & Schmid, 2003; Kumari *et al*., 2010; Ungewickell & Hinrichsen, 2007; Wu *et al*., 2009) . In Chapter 3 and Chapter 4 of this thesis, results were presented regarding the organisation of the system of protein-protein and protein-lipid interactions that fulfil the molecular requirements of CME, and its likely evolution. Results presented thus far, therefore provide an insight into the complex nature of endocytosis, but do not address the spatial distribution of endocytosis on the plasma membrane.

Little is known about how cells regulate the spatial organisation of endocytosis. In section 1.3.3, endocytic membrane micro-domains and cell surface structures which have been observed in protists such as *Trypanosoma* and *Tetrahymena* were mentioned (Gonda *et al.*, 2000; Porto-Carreiro *et al.*, 2000). In these protists, sites of cargo entry and exit appear to be limited to these specialised regions. Similarly, in certain cell types such as fungal hyphae and mammalian epithelial cells, endocytosis and exocytosis occurs in specialised cell surface structures that support morphogenesis and polar growth (Fuchs *et al.*, 2006; Wang *et al*., 2001). However, in cells where specialised plasma membrane domains cannot be recognised, where endocytosis and exocytosis

appears to be distributed evenly across the cell surface, the mechanism for determining sites of endocytosis is still poorly understood.

Endocytic protein complexes known as eisosomes have been suggested to mark spatial sites of endocytosis in *Saccharomyces cerevisiae* cells (Walther *et al*., 2006). Eisosomes are composed of large and stable assemblies of the cytoplasmic proteins Pil1 and Lsp1, which form at the plasma membrane (Walther *et al*., 2006) (Figure 5.1). Experimental data has shown that these protein complexes affect both endocytosis and plasma membrane integrity, because deletion of Pil1, results in loss of endocytic efficiency and severe disruption of endocytic site distribution, coupled with the occurrence of aberrant plasma membrane invaginations (Walther *et al*., 2006).

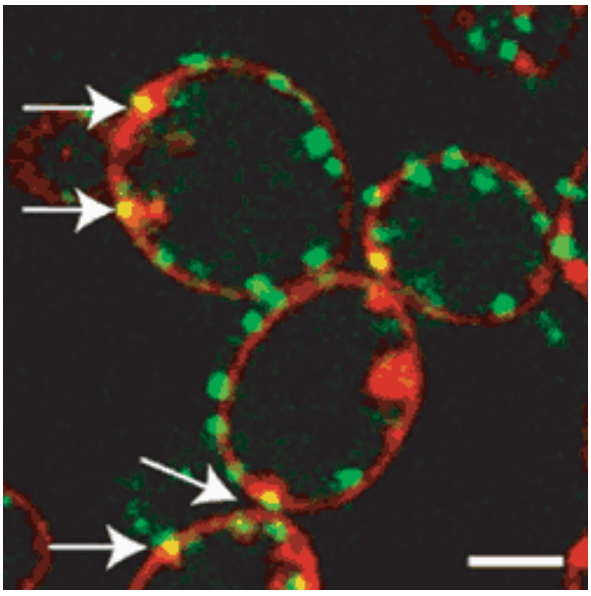


Figure 5.1 Eisosomes assemble into punctate complexes at the plasma membrane of *S. cerevisiae* cells. Arrows point to the yellow patches which is where Pil1-GFP co-localises with lipophilic marker FM4-64 (red), at proposed early endocytic vesicles. Green patches represent the Pil1-GFP signal (Taken from Swaminathan, 2006).

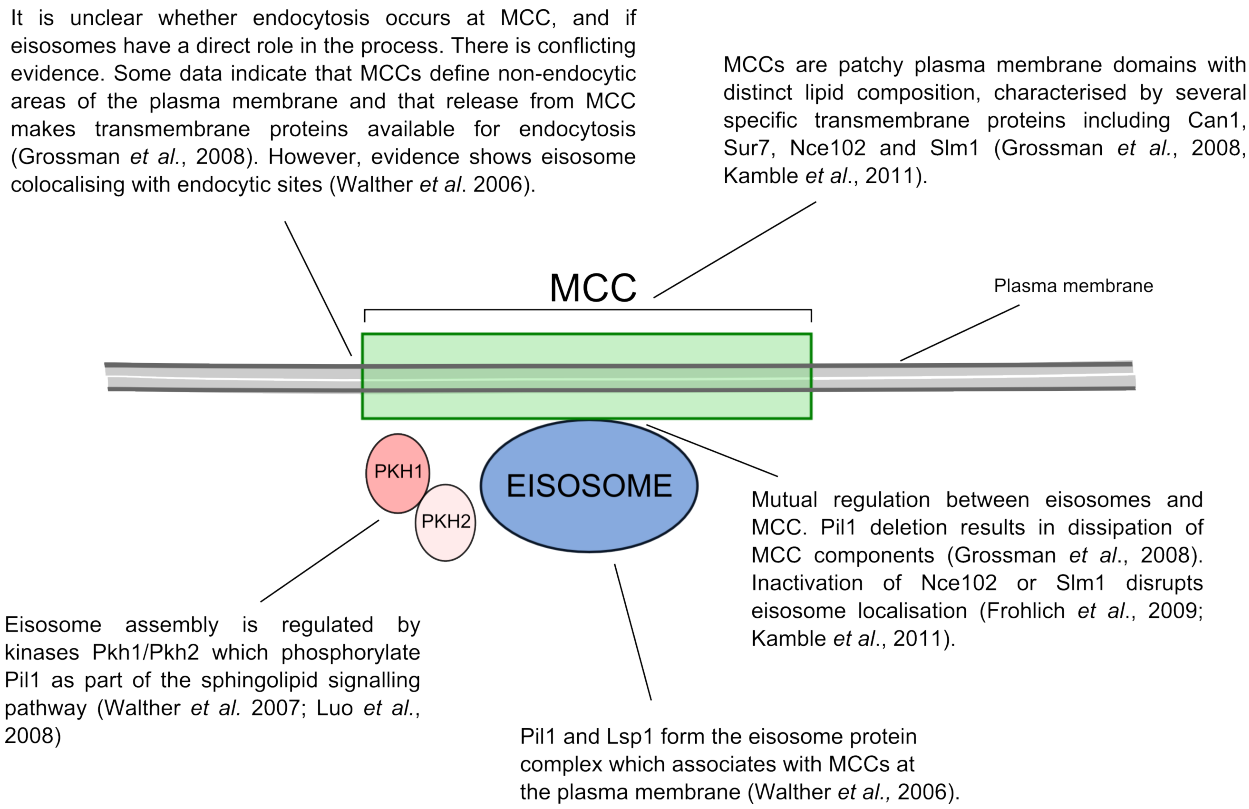


Figure 5.2 Schematic model of the eisosome/MCC system. Diagram summarises existing data on eisosome regulation and function. All graph components are schematic and not in scale. References for the data is included in diagram annotation.

Eisosomes were also found to colocalise with plasma membrane micro-domains, known as the membrane compartment of Can1 (MCC), which feature several proteins with transmembrane domains involved in transport and sphingolipid signalling (Grossmann *et al*., 2008). Two MCC proteins, Nce102 and Slm1, have been shown to play a role in eisosome organisation (Frohlich *et al*., 2009; Kamble *et al*., 2011). Conversely, the deletion of Pil1 causes all MCC markers to dissipate (Grossmann *et al*., 2008). In addition, eisosomes appear to operate as part of the signalling pathway, consisting of

long chain bases (LCB) and kinases Pkh1/2, Pkc1 and Ypk1/2, which is involved in cell wall integrity and cell wall remodelling during growth (Luo *et al*., 2008; Zhang *et al*., 2004). The data demonstrate that Pil1 and Lsp1 negatively regulate Pkh and its activation of the Pkc and Ypk pathways (Zhang *et al*., 2004). In addition, Pkh1 and Pkh2 have been shown to phosphorylate Pil1 and Lsp1 *in vitro* and *in vivo,* which activates eisosome assembly (Luo *et al*., 2008; Walther *et al*., 2007), whereas inactivation of Ypk kinases and the disruption of LCB synthesis causes disassembly of eisosomes (Luo *et al*., 2008). A schematic diagram of the emerging eisosome/MCC model can be seen in Figure 5.2.

The composition of eisosomes is still an open question. It is debatable whether MCC proteins, such as Sur7, Nce102 and Slm1, should be considered integral parts of eisosomes, or rather that eisosomes and MCCs are distinct and only interact transiently in a regulatory capacity. Here, I define eisosomes conservatively as Pil1 and Lsp1 assemblies at the plasma membrane, and I define MCC proteins that regulate eisosomes as eisosome related proteins (ERP). With regard to function, current published experimental data does not exhaustively conclude whether eisosomes play an active role in the molecular mechanisms involved in endocytosis, or simply help to define plasma membrane micro-domains with the required protein and lipid composition. However, the putative involvement of eisosomes in endocytosis (Walther *et al*., 2006), their regulatory interactions with MCCs (Grossmann *et al*., 2008) and their role in the LCB/kinase sphingolipid signalling pathway (Luo *et al*., 2008) makes a strong case for eisosomes to be involved at least in the spatial organisation of endocytosis.

Studies on eisosomes have so far been limited to budding yeast *Saccharomyces*

cerevisiae (Kamble *et al*., 2011; Luo *et al*., 2008; Walther *et al*., 2006), the yeast species *Ashbya gossypii* (Seger *et al*., 2011), and to the filamentous ascomycete *Aspergillus nidulans* (Vangelatos *et al*., 2010). In the study of *Aspergillus nidulans,* Pil1 and Lsp1 homologues were shown to be present in all ascomycetes, and a phylogeny suggested that multiple gene duplications occurred in the evolution of eisosomes in ascomycete fungi, but no further evolutionary data has yet been presented.

In this chapter, I investigate the importance of eisosomes to eukaryotic endocytosis with bioinformatics tools. The experimental outline is to carry out a comparative genomic study of Pil1 and Lsp1 in a diverse range of eukaryotic organism to establish the taxonomic distribution of eisosomes. I then carried out a phylogenetic analysis to reconstruct the evolutionary history of Pil1 and Lsp1. Finally, a secondary structure prediction analysis of a sample of Pil1 and Lsp1 homologues is presented. The data from these analyses is used to predict the likely origin and evolution of eisosomes.

5.2 Materials and methods

5.2.1 Identifying Pil1 and Lsp1 homologues

A total of 27 genome projects, representing a broad diversity of eukaryotes diversity were selected for database searches (see Table 2.1) for details of genome projects and online sources). Pil1 and Lsp1 sequences were searched against predicted proteomes and translated nucleotide databases using BLASTp, PSI-BLAST and tBLASTn, as described in section 2.1.2. Further database searches were then performed against predicted proteomes and translated nucleotide databases from an additional set of 17 fungal genome projects (see Table 2.2 for details of genome project and online sources).

The additional searches were performed with BLASTp, PSI-BLAST and tBLASTn, as described in section 2.1.2. Pil1 and Lsp1 sequences were also searched against the *Blastocladiella emersonii* expressed sequence tags database (EST) (Ribichich *et al*., 2005) http://blasto.iq.usp.br/ and TBestDB, a database comprising of EST libraries from 49 organisms (O'Brien *et al*., 2007) (http://tbestdb.bcm.umontreal.ca/searches/). Both searches were performed using tBLASTn (Altschul *et al*., 1990). Finally, The trace files archive available on NCBI (Sayers *et al*., 2010) was searched using BLASTn. In all cases, threshold levels for significance of sequence similarity was $e < 1 \times 10^{-5}$.

5.2.2 Cloning and sequencing putative Pil1 homologue from a *Blastocladiella emersonii* cDNA library

The EST BeE60C21E01 from cDNA library of the chytridiomycete *Blastocladiella emersonii* (Ribichich *et al*., 2005) http://blasto.iq.usp.br/, was found to share significant sequence similarity with both Pil1 and Lsp1 (E-value = 7e-15 for Lsp1, 7e-19 for Pil1). The corresponding cDNA clone (in pSPORT1) was obtained from Prof. Suely L. Gomes (Departamento de Bioquimica, Universidade de São Paulo, Brazil). The cDNA clone was sequenced with vector-specific primers T7 and SP6 and with the following clone-specific primers. Sense: BlemF1 5'-AGCAGTTTGCGGGGCACCA-3' and BlemF2 5'-CTGGCGTAATAGCGAAGA-3'. Anti-sense: BlemR1 5'-ACCTTGTGCTCGGCG-TCG-3' and BlemR2 5'-TCTTCGCTATTACGCCAG-3'. The resulting sequencing reads covered both strands of the full length cDNA and were edited and assembled into a single contig with Sequencher (see section 2.2.10 for details).

5.2.3 Cloning and sequencing putative Pil1 homologue from a *Capsaspora*

***owczarzaki* cDNA library**

An EST from the filasterean *Capsaspora owczarzaki* (NUL00001676), was identified from the TBestDB EST database (O'Brien *et al.*, 2007) (http://tbestdb.bcm.umontreal.ca/searches/) and found to share significant sequence similarity with Pil1 (E-value = 2e-07). The corresponding cDNA clone was obtained from Prof. Andrew Roger's laboratory (Department of Biochemistry and Molecular Biology, University of Dalhousie). The cDNA clone was sequenced with vector specific primers (M13 forward and M13 reverse) and with the following clone specific primers. Sense: lc2Rev 5'-ATCTTGAGCGATACAACCGAAA-3' and lc3Rev 5'-ATTCAAACGCCTATGCCAAC-3'; anti-sense: lc2Fwd 5'-TGAAGGACAGCAAG-ACCATT-3' and lc3Fwd 5'-TGAAGGACAGCAAGACCATT-3'. The resulting sequencing reads covered both strands of the full length cDNA and were edited and assembled into a single contig with Sequencher (see section 2.2.10 for details).

5.2.4 Phylogenetic analyses of Pil1 and Lsp1

Pil1 and Lsp1 homologues were recovered from a selection of predicted proteomes, representing the diversity of Fungi (see Tables 2.1-2.2 for details on taxa and online access of genome projects). The sequences were aligned with MUSCLE (Edgar, 2004a). Hyper-variable regions and gaps were removed with SEAVIEW (Gouy *et al.*, 2010). The masked data set was analysed to infer optimal model parameters with MODELGENERATOR (Keane *et al.*, 2006). The LG model with proportion of invariable site (0.04) and Γ distribution ($\alpha = 3.5$) was selected. A Bayesian phylogenetic analysis was carried out with MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003), performing 1,000,000 MCMC generations, sampling trees every 100 generations. The

RtREV amino acid substitution score matrix was used as the LG model is not implemented by the current version. Trees with low likelihood values were discarded and a consensus tree with posterior probability node support was derived from the remaining sampled trees. Fast maximum likelihood (ML) analyses were run with RAXML and PHYML, performing 1,000 bootstrap replicates each. The LG model was implemented in both analyses. For the PHYML analysis among-site rate variation was corrected by α value (3.5) and 8 discreet categories, describing the Γ distribution. For the RAXML analysis, the hard coded setting of 4 categories was routinely used.

5.2.5 Secondary structure prediction of Pil1 and Lsp1 homologues

A total of 11 sequences representing the diversity of Pil1 and Lsp1 homologues were selected for secondary structure prediction. The sequences were analysed using the JPRED 3 secondary structure prediction server, which implements the JNET algorithm (Cole *et al*., 2008) (http://www.compbio.dundee.ac.uk/www-jpred/). JNET predicts secondary structures of a given amino acid sequence (the three possible states are α-helix, β-sheet and coiled-coil) by training a neural network secondary structure prediction algorithm with multiple sequence alignment profiles in the form of position specific scoring matrices (PSSM) and Hidden Markov Model (HMM) profiles (Cuff & Barton, 2000). The use of PSSMs and HMMs in secondary structure prediction algorithm has been shown to improve significantly the accuracy of predictions (Cole *et al*., 2008; Cuff & Barton, 2000; Jones, 1999). JNET calculates confidence values for secondary structure prediction at each amino acid residue by subtracting the score for the second highest scoring state from the score for highest scoring state (Cuff & Barton, 2000).

5.3 Results

5.3.1 Taxonomic distribution of Pil1 and Lsp1

The amino acid sequences for *S. cerevisiae* Pil1 and Lsp1 were used as query sequences for homology searches against 28 eukaryotic predicted proteome and translated nucleotide genomic databases. Sequences with significant similarity to Pil1 and Lsp1 were found in *Magnaporthe oryzae, Ustilago maydis* and *Batrachochytrium dendrobatidis*. No sequence similarity to Pil1 or Lsp1 homologues was found with either BLASTp, PSI-BLAST or tBLASTn searches, in non-fungal organisms. This preliminary result suggests the Pil1 and Lsp1 paralogues might be specific to Fungi. To investigate this possibility further, I performed Pil1 and Lsp1 homology searches against predicted proteomes and translated nucleotide databases of a diverse group of fungi and opisthokont protists. These include 6 members of the Saccharomycotina (*Ashbya gossypii*, *Candida glabrata*, *Candida albicans, Debaryomyces hansenii, Lodderomyces elongisporus* and *Yarrowia lipolytica*) 3 representative Pezizomycotina species (*Aspergillus fumigatus, Botrytis cinerea* and *Neurospora crassa*), 1 member of the Taphrinomycotina (*Schizosaccharomyces pombe*), 5 members of the Basidiomycota (*Cryptococcus neoformans, Laccaria bicolor*, *Phanerochaete chrysosporium, Postia bicolor, Puccinia graminis*) 2 representative Zygomycota species (*Phycomyces blakesleeanus* and *Rhizopus oryzae*), 1 Chytridiomycota species (*Allomyces macrogynus*), 1 Microsporidia species (Encephalitozoon *cuniculi*), 1 member of the Choanoflagellata (*Salpingoeca rosetta)*, 1 member of the Ichthyosporea (*Sphaeroformans arctica*) and 1 Filasterea species (*Ministeria brivans*). In addition, I searched the Pil1 and Lsp1 sequences against expressed sequence tags (EST) databases of a selection of additional fungal and opisthokont protist taxa. These include 2

Taphrinomycotina species (*Saitoella complicata* and *Taphrina deformans*), 1 Zygomycota species (*Morteriella verticillata*), 2 members of the Chytridiomycota (*Allomyces macrogynus* and *Blastocladiella emersonii*), 1 Microsporidia species (*Antonospora locustae)*, 2 Choanoflagellata species (*Proterospongia sp.* and *Monosiga ovata*), 1 Filasterea species (*Capsaspora owczarzaki*) and 1 Nucleariidae species (*Nuclearia simplex*).

Predicted proteins and/or translated nucleotide sequences with significant similarity to Pil1 and Lsp1 were found in all the saccharomycete, pezizomycete, taphrinomycete, basidiomycete and zygomycete taxa sampled. For the chytridiomycete species *Batrachochytrium dendrobatidis*, a putative homologue was found (sequence hit: BDEG_06158; E-value: 1.2e-06 with Pil1 as query, and 2.0e-06 with Lsp1 as query) but the other chytridiomycete taxon sampled, *Allomyces macrogynus,* lacked any sequences that are significantly similar to either Pil1 or Lsp1. No sequences with significant similarity to Pil1 or Lsp1 were found in the microsporidian *Encephalitozoon cuniculi*, the choanoflagellate *Salpingoeca rosetta*, the ichthyosporean *Sphaeroformans arctica,* or the filasterean *Ministeria brivans* (Figure 5.3a).

Translated EST sequences with significant similarity to Pil1 and Lsp1 were found in taphrinomycete *Saitoella complicata*, the zygomycete *Morteriella verticillata* and the chytridiomycete *Blastocladiella emersonii.* In the filasterean species *Capsaspora owczarzaki,* a tBLASTn search with Pil1 as query returned EST NUL00001676 with an E-value of 2e-07, 29% amino acid identity and 48% similarity. However, a tBLASTn search with Lsp1 as query did not return a significantly similar sequence. No sequences

with significant similarity to Pil1 or Lsp1 were found in the EST databases of the taphrinomycete species *Taphrina deformans,* the chytridiomycete species *Allomyces macrogynus*, the microsporidian species *Antonospora locustae*, the nucleariid species *Nuclearia simplex* and choanoflagellate species *Proterospongia sp.* and *Monosiga ovata* (Figure 5.3b).

5.3.2 Alignment and phylogenetic analysis of Pil1 and Lsp1

To investigate the evolutionary history of Pil1 and Lsp1, I retrieved all of the putative Pil1 and Lsp1 homologues found in the comparative genomic study. In addition, I sequenced the full length cDNA fragment of BeE60C21E01 from a *Blastocladiella emersonii* EST library and NUL00001676 from a *Capsaspora owczarzaki* EST library (see sections 5.2.2 and 5.2.3). I also assembled a contig from *Spizellomyces punctatus* whole-genome shotgun data held in the NCBI trace archives (see section 5.2.1). All sequences retrieved were then aligned to create a multiple sequence alignment (MSA) in amino-acid format. The MSA was edited to exclude gaps and hypervariable regions and analysed to infer the optimal parameters for its evolutionary model. I then carried out a phylogenetic analysis of the masked data-set, using two fast-maximum likelihood (ML) methods (PHYML and RAXML) and one Bayesian method (MRBAYES), and using bootstrap replicates and Bayesian posterior probability values (PP), as statistical support for tree topologies (see section 5.2.4 for details on methods).

The trees obtained with the fast-ML and Bayesian analyses were rooted with *C. owczarzaki* NUL00001676. However, this rooting depicts *C. owczarzaki* NUL00001676 as a long branch, and produced an unlikely clade comprising chytridiomycete and

a

Taxa	Taxonomic group		Pil1	Lsp1
Saccharomyces cerevisiae	Saccharomycotina	Ascomycota	●	●
Candida glabrata			●	●
Ashbya gossypii			●	●
Candida albicans			●	●
Debaryomyces hansenii			●	●
Lodderomyces elongisporus			●	●
Yarrowia lipolytica			●	●
Magnaporthe oryzae	Pezizomycotina		●	●
Aspergillus fumigatus			●	●
Botrytis cinerea			●	●
Neurospora crassa			●	●
Schizosaccharomyces pombe	Taphrin.		●	●
Ustilago maydis	Basidiomycota		●	●
Cryptococcus neoformans			●	●
Laccaria bicolor			●	●
Phanerochaete chrysosporium			●	●
Postia bicolor			●	●
Puccinia graminis			●	●
Phycomyces blakesleeanus	Zygomycota		●	●
Rhizopus oryzae			●	●
Batrachochytrium dendrobatidis	Chytridiomycota		●	●
Allomyces macrogynus			○	○
Enchephalitozoon cuniculi	Microsporidia		○	○
Salpingoeca rosetta	Choanoflagellata		○	○
Sphaeroformans arctica	Ichthyosporea		○	○
Ministeria brivans	Filasterea		○	○

b

Taxa	Taxonomic group	Pil1	Lsp1
Saitoella complicata	Taphrinomycotina	●	●
Taphrina deformans		○	○
Morteriella verticillata	Zygomycota	●	●
Allomyces macrogynus	Chytridiomycota	○	○
Blastocladiella emersonii		●	●
Antonospora locustae	Microsporidia	○	○
Proterospongia sp.	Choanoflagellata	○	○
Monosiga ovata		○	○
Capsaspora owczarzaki	Filasterea	●	○
Nuclearia simplex	Nucleariidae	○	○

Figure 5.3 Taxonomic distribution of Pil1 and Lsp1 in Fungi and opisthokont protists. a Results from BLASTp, PSI-BLAST and tBLASTn searches against predicted proteomes and translated nucleotide databases. Pil1 and Lsp1 are the query proteins. Black circle represents presence of significant sequence similarity between query protein and a predicted protein and/or translated nucleotide sequence. Empty circle represents absence of significant sequence similarity. **b** Results from tBLASTn searches against expressed sequence tags (EST) libraries. Black circle represents presence of significant sequence similarity between query protein and a translated EST. Taxa and taxonomic classification are indicated. Empty circle represents absence of significant sequence similarity. For details of database sources see Table 5.2.1-3.

basidiomycete sequences (See Appendix 4). The branch position of *C. owczarzaki* NUL00001676 may therefore be an artifact due to long branch attraction. A second set of fast-ML and Bayesian phylogenetic analyses was performed with the exclusion of *C. owczarzaki* NUL00001676. The trees were rooted with a clade comprising the three chytridiomycete species (*Batrachochytrium dendrobatidis*, *Spizellomyces punctatus,* and *Blastocladiella emersonii*). To summarise the statistical node support from different phylogenetic analyses, the tree topology derived from the Bayesian analysis was selected as template and the bootstrap support values from PHYML and RAXML analyses were assigned, together with the Bayesian PP values, to the relevant nodes (Figure 5.4).

Overall, the phylogenetic analyses indicate that Pil1 and Lsp1 have a complex evolutionary history, characterised by at least five gene duplications and one gene loss events (Figure 5.4). The gene duplication that produced the Pil1 and Lsp1 paralogues occurred in the last common ancestor of all the Saccharomycotina species sampled, minus *Yarrowia lipolytica* which branched off prior to the proposed duplication event. This result is supported by robust statistical support for both the Pil1 clade and the Lsp1 clade (Figure 5.4). The results are also consistent with a gene duplication event in the last common ancestor of the Ascomycota. This event, putatively produced a lineage ancestral to Pil1 and Lsp1 that was retained in all the ascomycete taxonomic groups, and a paralogous lineage that was retained in the Pezizomycotina and the Taphrinomycotina but not in the Saccharomycotina (Figure 5.4). The statistical support for the Pil1/Lsp1 paralogue clade is 0.85 PP, 50% RAXML bootstraps and 65% PHYML bootstraps (0.85/50/65). The statistical support for its paralogous clade

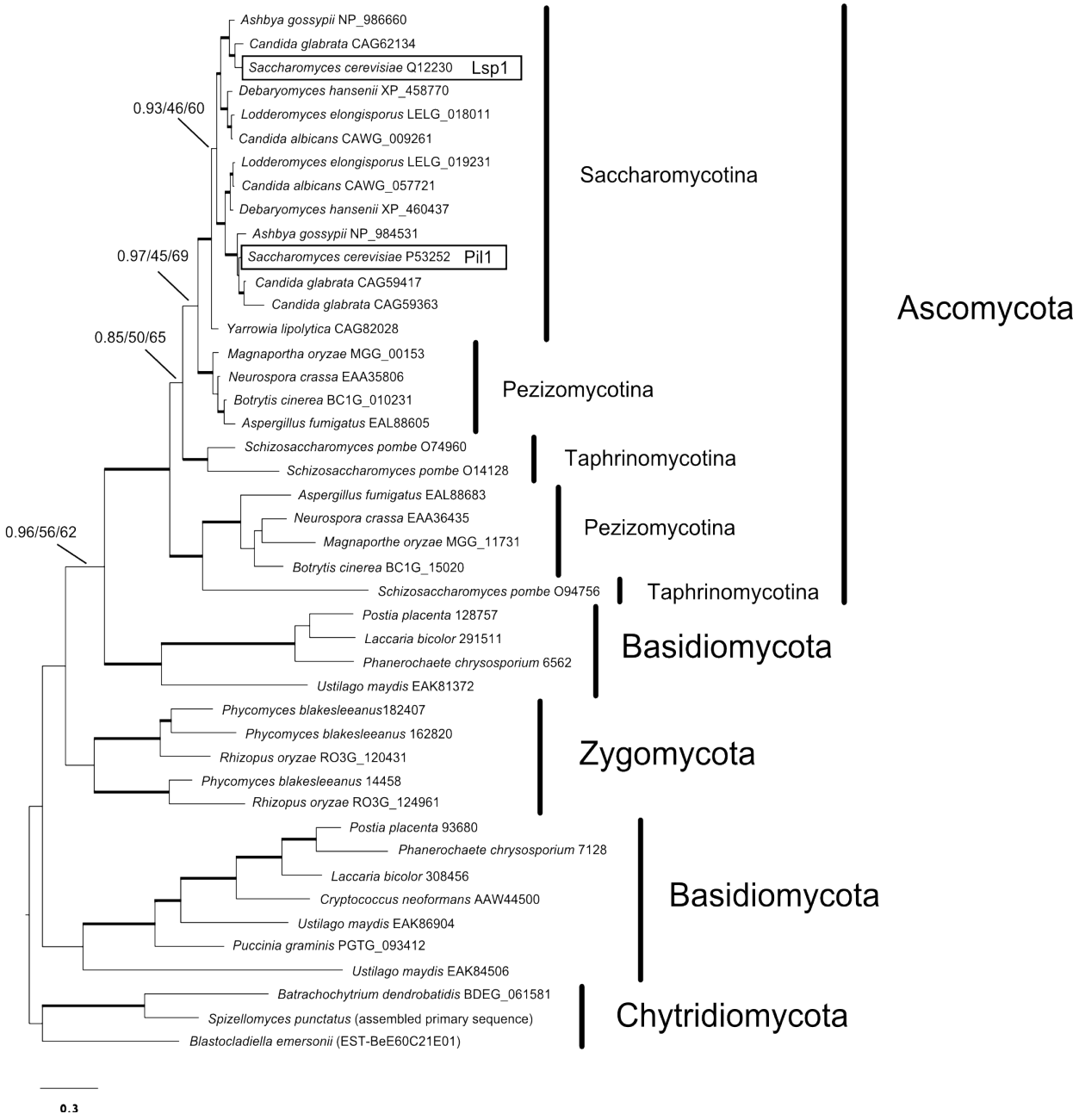


Figure 5.4 Phylogenetic tree of Pil1 and Lsp1. The topology was derived as the consensus tree from Bayesian phylogenetic analysis. Thick branches represent statistical support values of at least 0.80 Bayesian posterior probability (PP), and 60% from both RAXML and PHYML using 1000 bootstrap replicates. In key nodes support values are indicated in order as Bayesian PP, RAXML bootstrap support and PHYML bootstrap support. Species names, database accession numbers and taxonomic groups are indicated.

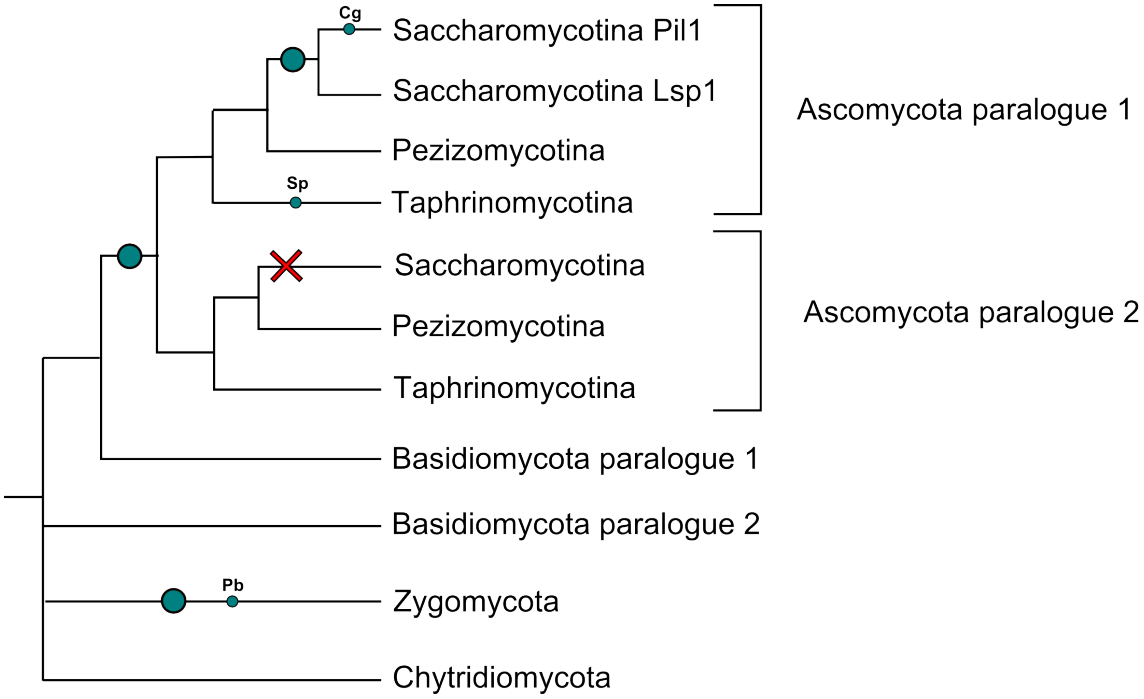


Figure 5.5 Schematic tree of Pil1 and Lsp1 evolutionary history. The schematic tree summarises phylogenetic analyses of Pil1 and Lsp1 and highlights putative gene duplication and gene loss events. The blue-filled circle represents a proposed gene duplication event, the red cross represents a proposed gene loss event. The smaller blue-filled circles represent gene duplications specific to a species sampled in the study. Cg is *Candida glabrata,* Sp is *Schizosaccharomyces pombe,* Pb is *Phycomyces blakesleeanus.* Branches are annotated according to taxonomy and paralogue group. Nodes with two out of three support values less than 50 % are collapsed. The blue-filled circle represents a proposed gene duplication event, the red cross represents a proposed gene loss event.

meanwhile is 0.99/60/69. The phylogenetic tree also revealed two monophyletic basidiomycete clades. One clade has a support value of 0.96/56/62 for monophyly with the sister ascomycete clade, while the position of the other clade is poorly resolved (see Figure 5.4). The gene duplication event which produced the two clades therefore can not be pinpointed with certainty. Finally, the tree revealed a *Candida glabrata*-specific gene duplication, a *Schizosaccharomyces pombe*-specific gene duplication, a *Phycomyces blakesleeanus*-specific gene duplication and a gene duplication in the last

common ancestor of the zygomycete species *Phycomyces blakesleeanus* and *Rhizopus oryzae.* A schematic tree summarising key gene duplication and loss events suggested by the phylogenetic analyses is shown in Figure 5.5.

5.3.3 Secondary structure prediction of Pil1 and Lsp1 homologues

A total of 11 amino acid sequences were selected for secondary structure prediction analysis of Pil1 homologues. These are *Saccharomcyes cerevisiae* Pil1 and Lsp1, *Magnaporthe oryzae* MGG_00153 and MGG_11731, *Ustilago maydis* EAK81372, *Cryptococcus neoformans* AW44499, *Phycomyces blakesleeanus* 79864, *Rhizopus oryzae* RO3G_12496, *Blastocladiella emersonii* BeE60C21E01, *Batrachochytrium dendrobatidis* BDEG_06185 and *Capsaspora owczarzaki* NUL00001676. The α-helix, β-sheet and coil secondary structure states were predicted for the 11 amino acid sequences, using the JNET algorithm (see section 5.2.5 for details on JNET). The secondary structures were compared by aligning the 11 sequences with MUSCLE and annotating each sequence with secondary structures supported by confidence value of 7 or higher (Figure 5.6).

The secondary structure of Pil1 and Lsp1 is characterised by α-helices and is strongly conserved in the putative homologues selected. In particular, a structure of five successive α-helices is shared in all of the sequences with only two deviations. *Batrachochytrium dendrobatidis* BDEG_06185 lacks the first of the five α-helices, while in *Capsaspora owczarzaki* NUL00001676 the third and fourth α-helices are predicted to be a single large α-helix. The NUL00001676 structure also features a 7 amino acid insert in the $3^{rd}/4^{th}$ α-helix and a single amino acid insert in the 5^{th} α-helix

(Figure 5.6).

5.4 Discussion

The comparative genomic data I have reported in this chapter suggests that eisosomes are specific to the kingdom Fungi. To pinpoint the origin of the two known cytoplasmic components of eisosomes, Pil1 and Lsp1, homology searches in basal fungal and opisthokont protist taxa were carried out. A homologue of Pil1 and Lsp1 was found in two of out three chytridiomycete taxa, but was absent in both microsporidian species examined. Pil1 and Lsp1 were absent in all opisthokont protists except for a putative homologue in *Capsaspora owczarzaki* (NUL00001676). This finding was not expected because the the opisthokont protist taxon is sister to metazoans whereas Pil1 and Lsp1 are absent in the Nucleariidae taxon that was sampled. However, the phylogenetic tree of Pil1 and Lsp1 suggests that the *Capsaspora* candidate homologue is a long branch (Appendix 4). In addition, by rooting the tree with the candidate *Capsaspora* homologue, an unlikely branch including basidiomycetes and chytridiomycete is produced. The secondary structure prediction study also shows that the *Capsaspora* candidate homologue is partly divergent from the other sequences. Taken together, these data make homology of the NUL00001676 sequence questionable. To further investigate the origin of eisosomes, more genomic data should be added to the current selection of basal fungal and opisthokont protist genome projects (O'Brien *et al*., 2007; Ruiz-Trillo *et al*., 2007). Phylogenetic analyses showed that multiple gene duplication and at least one gene loss event have occurred in the evolutionary history of Pil1 and Lsp1. Two gene duplications and one gene loss in was identified in ascomycetes as well as lineage specific duplications in *Candida glabrata, Schizosaccharomyces pombe* and

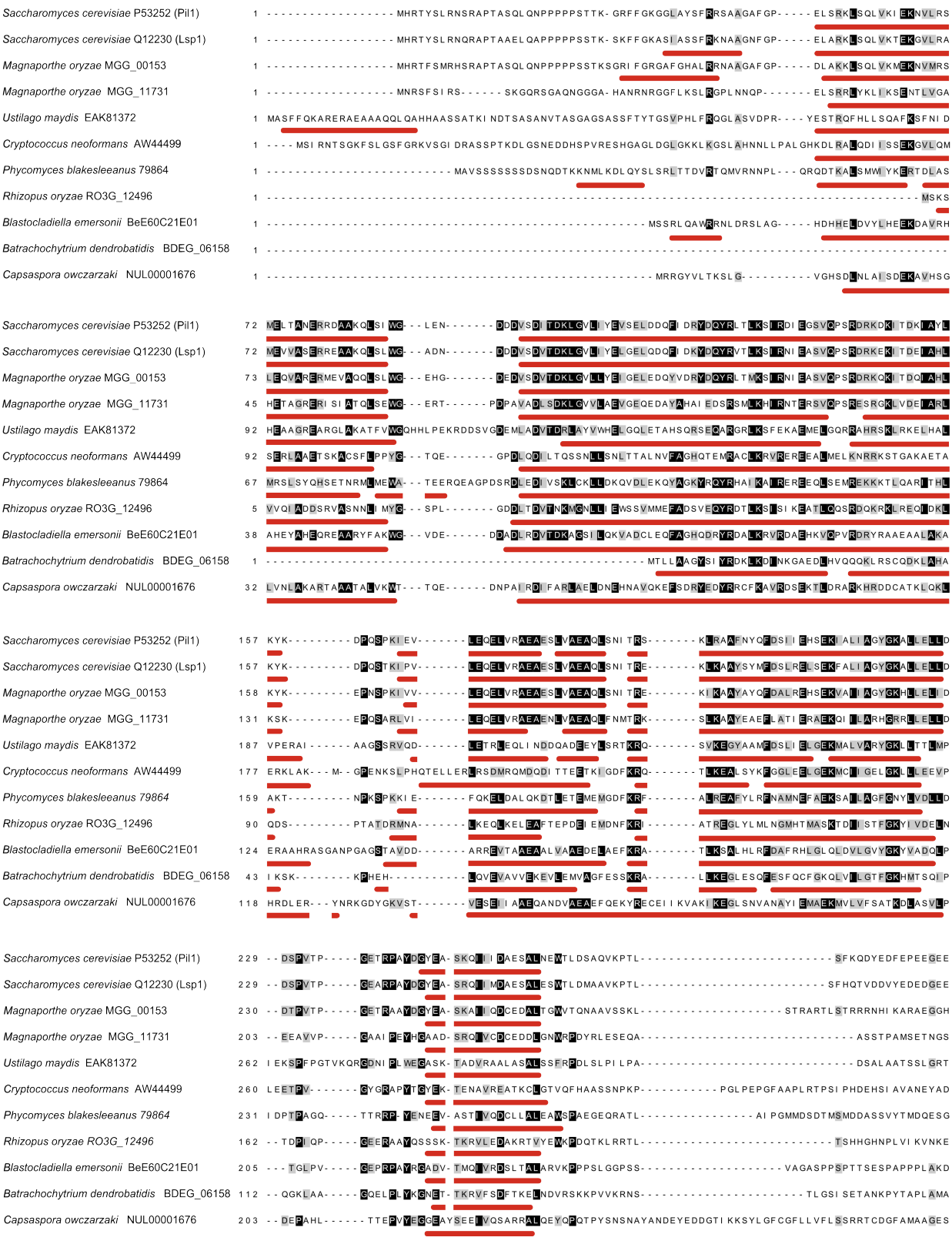

Saccharomyces cerevisiae P53252 (Pil1)
Saccharomyces cerevisiae Q12230 (Lsp1)
Magnaporthe oryzae MGG_00153
Magnaporthe oryzae MGG_11731
Ustilago maydis EAK81372
Cryptococcus neoformans AW44499
Phycomyces blakesleeanus 79864
Rhizopus oryzae RO3G_12496
Blastocladiella emersonii BeE60C21E01
Batrachochytrium dendrobatidis BDEG_06158
Capsaspora owczarzaki NUL00001676

Figure 5.6 Secondary structure prediction of Pil1 and Lsp1 homologues. Multiple sequence alignment was performed with MUSCLE and visualised with BOXSHADE. Black boxes indicate conserved invariant amino acids wheras grey boxes indicate conserved substitutions. Each sequence taxon name and database accession number are indicated. Red bars represent putative α helices. Only secondary structures with JNET confidence levels equal to or higher than 7 are represented.

the two zygomycete taxa that were sampled. There are two basidiomycete paralogues too, but the tree does not explain when they duplicated. Taken together, the data suggest that the function of Pil1 and Lsp1 diversified significantly during the evolution of fungi.

Here, I propose that eisosomes evolved as an adaptation to the evolution of the fungal cell wall, and that they characterised a fungal specific diversification of endocytosis. This hypothesis entails that eisosomes, and the functions they mediate, are conserved across diverse fungi. In the next Chapter, this notion is investigated by presenting data on the function of the eisosome in the filamentous fungus *Magnaporthe oryzae*.

6 Functional characterisation of *Magnaporthe oryzae* MoPil1 and MoPil2

6.1 Introduction

Eisosomes are large and stable protein complexes which localise at the cell surface of *Saccharomyces cerevisiae*. Their putative involvement in, endocytosis (Walther *et al.*, 2006), the sphingolipid signalling pathway (Luo *et al.*, 2008; Zhang *et al.*, 2004), and the regulation of the membrane compartment of Can 1 (MCC) (Grossmann *et al.*, 2008) (Frohlich *et al.*, 2009; Kamble *et al.*, 2011) suggests a role in membrane dynamics and spatial regulation of endocytosis. The results from the phylogenomic analysis presented in chapter 5, demonstrated that Pil1 and Lsp1 homologues are present in all the ascomycete, basidiomycete, and zygomycete fungi sampled, as well as in two out of the three chytridiomycete fungi included in the study, but that they are absent from all non-fungal organisms. This suggests that eisosomes were likely acquired at the base of the fungal tree and could thus represent a case of fungal specific diversification of endocytosis. This is important because, in chapter 4, the results suggested that CME significantly diversified in the Holozoa, but no data exists on endocytic diversification that is specific to the holozoan sister groups (i.e. the Fungi and nucleariids).

However, thus far, the evidence of the involvement of eisosomes in endocytosis comes from cell biological studies of only budding yeast (Grossmann *et al.*, 2008; Luo *et al.*, 2008; Walther *et al.*, 2007; Walther *et al.*, 2006). The Fungi in fact comprise groups with very diverse morphology and ecology, for instance fungi with flagellated cells (e.g. Chytridiomycota) and multicellular fungi with filamentous growth and aerial spore

dispersal (e.g. pezizomycotina; James *et al.*, 2006). Homologues of Pil1 and Lsp1 were found in all major fungal divisions, but the phylogenetic analyses presented in chapter 5, suggested that multiple, division specific, gene duplications and gene losses have occurred (see Figure 5.3). This may reflect a significant level of functional diversification of the eisosome across fungal diversity, and may mean the role of eisosomes in endocytosis is a derived function, specific to budding yeasts. To test whether eisosomes are important to the diversification of endocytosis in fungi, Pil1 and Lsp1 need to be characterised in fungal model organisms other than budding yeast.

For instance, the phylogenetic analysis presented in section 5.3.2, suggested that a gene duplication in the last common ancestor of ascomycete fungi resulted in a paralogue which is ancestral to both Pil1 and Lsp1, and a paralogue which was retained in Pezizomycotina and Taphrinomycotina but was subsequently lost in the Saccharomycotina (Figure 5.4). There may thus be significant diversity of eisosome function within ascomycetes. The aim of this chapter, is to investigate this hypothesis by characterising Pil1 and Lsp1 homologues in the pezizomycete fungus *Magnaporthe oryzae*.

M. oryzae is an ideal candidate for a comparative study of eisosome function in Ascomycota for the following reasons: (i) It is an economically important and devastating rice crop pathogen, which has been studied extensively in molecular biology laboratories, to engineer effective treatment against its spread (Talbot, 2003). (ii) *M. oryzae* can be transformed using a range of different markers, such as hygromycin B and sulfonylurea, and targeted gene replacement can be carried out to study gene

function (Talbot & Foster, 2001). (iii) The life cycle of *M. oryzae* includes different developmental stages, such as the formation of aerial conidiophores, development of of appresoria, and vegetative hyphal growth, all with distinct cell organisation and morphology (Wilson & Talbot, 2009) (Figure 6.1). (iv) *M. oryzae* is a pezizomycete, the sister group of saccharomycetes (James *et al.*, 2006), and thus represents the most closely related group of filamentous fungi to S. cerevisiae.

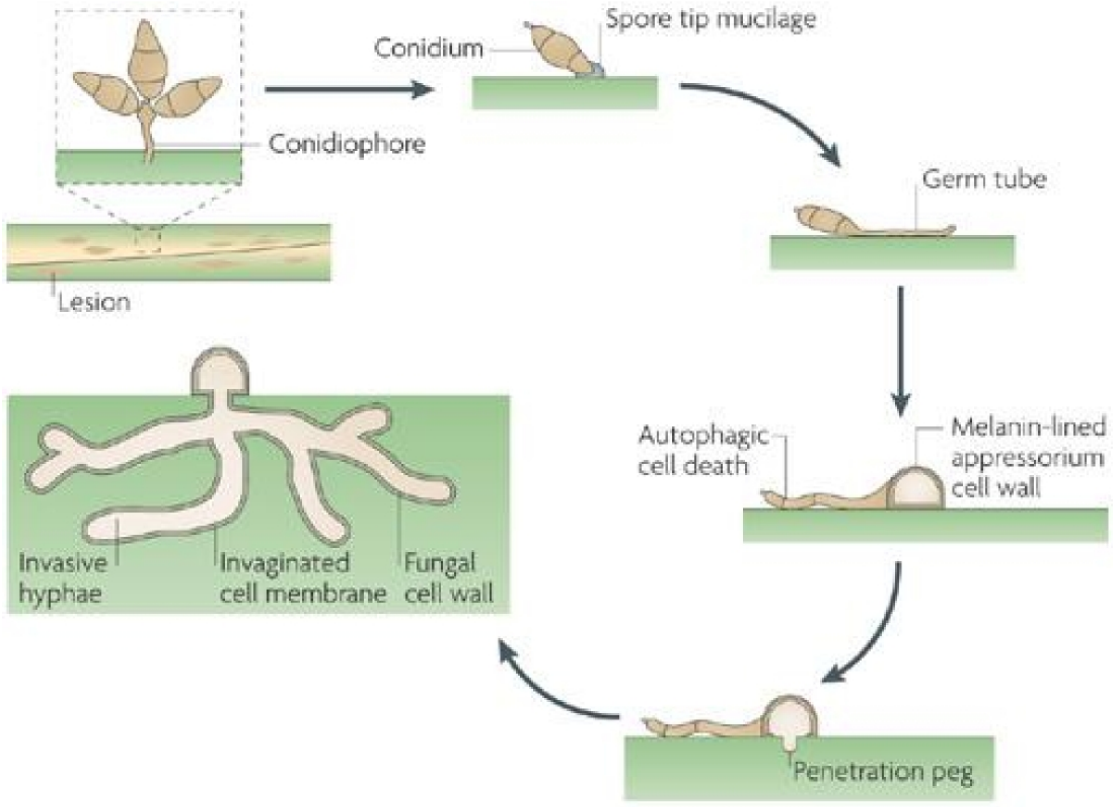


Figure 6.1 The infective life cycle of *Magnaporthe oryzae*. Life cycle begins with an asexual spore (conidium) which germinates on a host leaf cunicle. A germ tube expands from the condium and matures into an infection-ready appressorium. The turgour pressure of the dome-shaped cell eventually allows entry into the plant host via the extension of a penetration peg. Invasive hyphae ramify in the host tissue. Sporulation occurs at sites of host lesion. Figure taken from (Wilson & Talbot, 2009).

To distinguish between ascomycete specific paralogues, I name the two paralogues which resulted from a putative gene duplication event in the last common ancestor, Pil1 and Pil2 (Figure 6.2). According to the phylogenetic analysis presented in chapter 5, the MGG_00153 *M. oryzae* homologue is in the Pil1 clade which has 0.93 posterior probability (PP), 60% PHYML bootstraps support and 50% RAXML bootstraps support. By contrast, the MGG_11731 *M. oryzae* homologue is in the Pil2 clade which has 0.99 PP, 68% PHYML bootstraps support and 68% RAXML bootstraps support (Figure 6.2). MGG_00153 is thus referred to as *MoPIL1,* and MGG_11731 is thus referred to as *M. oryzae MoPIL2*.

The first aim for this chapter, is to study the sub-cellular localisation of MoPil1 and MoPil2 at different developmental stages of *M. oryzae.* Although *M. oryzae* is heterothallic and can reproduce sexually (Valent *et al*., 1991), its infective life cycle is asexual (Wilson & Talbot, 2009) (Figure 6.1). The infective process starts with the attachment of a three-celled conidium (asexual, non-motile spore) to the host's leaf cuticle, via an adhesive released by an apical spore compartment during hydration (Hamer *et al*., 1988; Wilson & Talbot, 2009). A narrow, polarised germ tube emerges from the spore before differentiating into a single celled, dome shaped, infective structure called the appressorium. Via development of turgor in the appressorium, massive pressure is exerted onto the leaf's cuticle, which is eventually pierced by the extension of a penetration peg from the base, allowing invasion of the host (Howard & Valent, 1996). From the penetration peg, bulbous, invasive hyphae develop, invading epidermal cells and ramifying through the host's tissue (Wilson & Talbot, 2009) (Figure 6.1). Sporulation occurs from disease lesions, with formation of aerial conidiophores

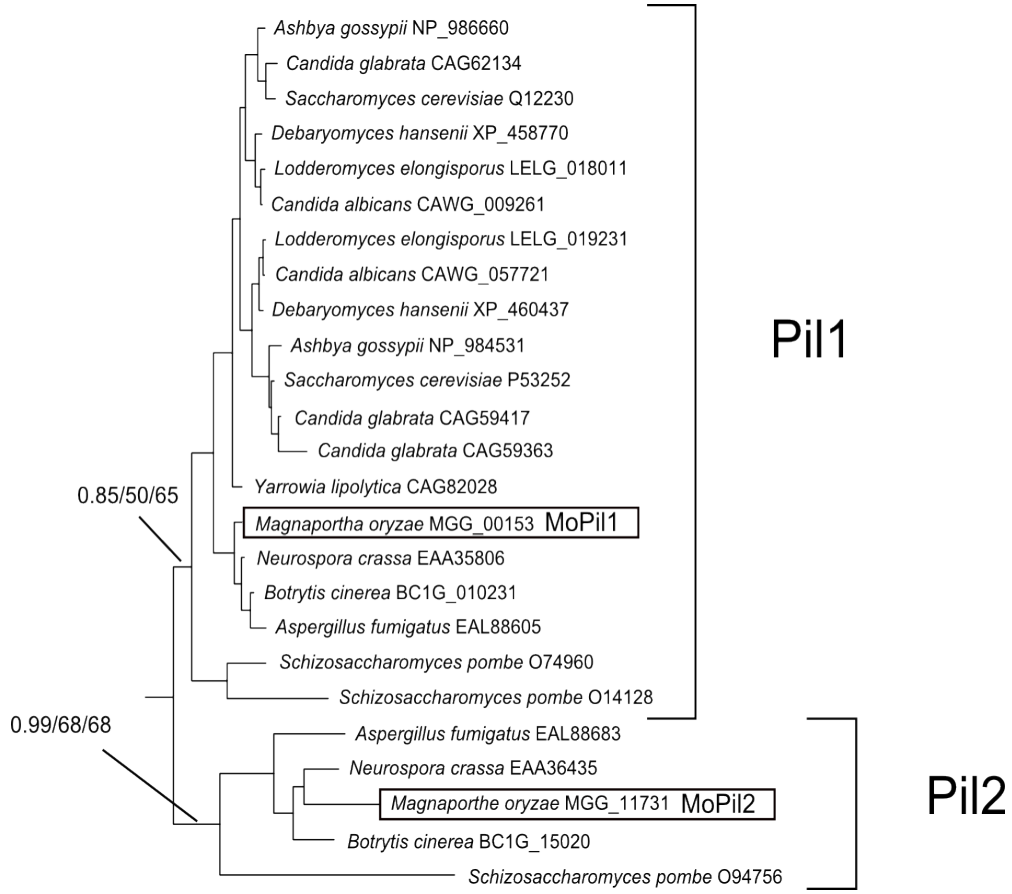


Figure 6.2 Defining Pil1 and Pil2 paralogues in fungi. The ascomycete sub-tree is taken from the Pil1 and Lsp1 phylogeny presented in chapter 5. The Pil1 and Pil2 paralogues are annotated. MGG_00153 is in the Pil1 clade, wheras MGG_11731 is in the Pil2 clade. The numeric values represent, in order, Bayesian posterior probability, the percentage RAXML bootstrap support and the percentage PHYML bootstrap support.

carried to new host plants via wind or dew splash (Talbot, 2003). In laboratory conditions, *M. oryzae*'s infective developmental stages can be analysed by inducing the maturation of the infective structures on hydrophobic plastic surfaces (Dean, 1997). Vegetative hyphae can be grown on media and sporulation is induced by increased humidity.

The second aim of this part of the project, is to compare the function of MoPil1 and MoPil2, with *S. cerevisiae* eisosome components. A useful method to compare gene function is to test for functional complementation. This consists in transforming a null mutation strain for a native gene, with a homologous gene from a different organism, and testing whether the inserted gene functionally complements the host's deleted gene. Data shows that, although the function of Pil1 and Lsp1 is non-redundant, a Pil1 null mutation will cause severe disruption to plasma membrane integrity and distribution of endocytic sites, whereas deletion of Lsp1 will not (Walther *et al.*, 2006). Because of this greater deleterious impact on the wild type phenotype, functional complementation of MoPil1 and MoPil2 is tested against Pil1.

The experimental outline for this chapter thus consists in:

Creating *M. oryzae* strains expressing MoPil1 and MoPil2 fused with different reporter genes, and study sub-cellular localisation pattern of the two proteins, at different stages of *M. oryzae*'s life cycle.

Creating a *S. cerevisiae pil1*Δ strains expressing MoPil1, and create a *S. cerevisiae pil1*Δ strain expressing MoPil2. Then, compare the phenotype of the transformed strains with phenotype of the wild type strain to test test for functional complementation.

6.2 Materials and methods

6.2.1 Generating *M. oryzae* strains expressing MoPil1-GFP and MoPil2-GFP.

A 3385 bp DNA fragment comprising the predicted MGG_00153 promoter region and

coding sequence was amplified using primers 5Pil1: 5'- ATGGATCCCACATTGTTC-TTCCTAACCGGC-3', and 3Pil1: 5'-ATGAATTCGGCAAGAACGGCTTCCTGGG-3' which feature a 5' end flanking *Bam*HI and *Eco*RI recognition sequence, respectively. The PCR reaction was prepared by mixing 100 ng of *Magnaporthe oryzae* genomic DNA (100 ng/µl), 1 µl 5Pil1 (20 pmol/µl), 1 µl 3Pil1 (20 pmol/µl), 5 µl Green GoTaq ® Flexi Buffer (5X), 2.5 µl $MgCl_2$ solution (25 pmol/µl), 0.5 GoTaq® DNA polymerase (5u/µl) and 14 µl nuclease free H_2O. Polymerase Chain Reactions (PCRs) were incubated in a thermal cycler with the following program: step 1. 95 ºC for 5 minutes; step 2. a cycle of 95 ºC for 30 seconds, 60 ºC for 30 seconds and 72 ºC for 3 minutes and 30 seconds, repeated 30 times; step 3. 72 ºC for 10 minutes. The amplicons were analysed by DNA gel electrophoresis to identify the predicted amplified DNA fragment. The amplicon was excised from the gel, purified and cloned into *Bam*HI and *Eco*RI sites of the linearised pCB1532 vector (Sweigard, 1997). The modified vector was then used to transform competent *E. coli* cells (XL-1 Blue strain, see section 2.26 for genotype). Successful transformants were screened by resistance to ampicillin. Plasmid preparations of a sample of transformed bacterial colonies were carried out. A 1015 bp DNA fragment consisting of the sGFP gene and NOS terminator were amplified from the pCAMBgfp vector (Sesma & Osbourn, 2004) with primers 5GFP: 5'-ATGAATTCATGGTGAGCAAGGGCGAGGAG-3' and 3NosT: 5'-ATATCG-ATACTCATGTTTGACAGCTTATCATCGG-3' which feature 5' end flanking EcoRI and ClaI recognition sequences, respectively. The PCR, was prepared by mixing 1 µl pCAMBgfp plasmid DNA (50 ng/µl), 1 µl 5GFP1 (20 pmol/µl), 1 µl 3NosT (20 pmol/µl), 5 µl Green GoTaq ® Flexi Buffer (5X), 2.5 µl $MgCl_2$ solution (25 pmol/µl), 0.5 GoTaq® DNA polymerase (5u/µl) and 14 µl nuclease free H_2O. Each reaction was

incubated in a thermal cycler with the following program: step 1. 95 ºC for 5 minutes; step 2. a cycle of 95 ºC for 30 seconds, 60 ºC for 30 seconds and 72 ºC for 1 minute and 30 seconds, repeated 30 times; step 3. 72 ºC for 5 minutes. The PCR products were analysed by DNA gel electrophoresis to identify the amplified DNA fragment. The amplicon was excised from the gel, purified and cloned into *Eco*RI and *Cla*I sites of the linearised pCB1532+MGG_00153 vector via DNA ligation. The modified vector was then used to transform a batch of competent *E. coli* cells. Successful transformants were screened by resistance to ampicillin and plasmid preparations of a sample of transformed bacterial colonies were carried out. To confirm MoPil1 (promoter and ORF) and GFP were in-frame, the construct was sequenced using sense primer constrF 5'-GAAGCGAGCAAGGCTATCATTC-3' and anti-sense primer, constrR 5'-CGTCGTCCTTGAAGAAGATGGT-3'. The resulting sequencing reads were edited and assembled into a contig using Sequencher (GenoCodes Corporation). The *M. oryzae* Guy-11 strain (Leung *et al*., 1988) was then transformed with the pcb1532+MGG_00153+GFP vector. For details on methodologies related to DNA manipulation, cloning, sequencing and fungal transformation, please see Sections 2.2.3-2.2.7, 2.2.10 and 2.2.11.1. The *Magnaporthe oryzae* strain expressing Pil1-GFP and Pil2-RFP was made in collaboration with Dr. M. J. Egan (School of Biosciences, University of Exeter).

6.2.2 Microscopy

Microscopy of MoPil1-GFP strains was performed in collaboration with Dr. M. J. Egan (School of Biosciences, University of Exeter) using a Zeiss LSM 510 Meta confocal microscope. Excitation of fluorescently labeled proteins was carried out using Blue

diode (405nm), Argon (458, 477, 488, 504nm), Helium-Neon (543nm), and all images were recorded following examination under a 63X Plan-Apo/1.4 NA Oil with DIC capability. Offline image analysis was carried out using the LSM image browser (Zeiss). Figures were subsequently prepared from digitised images in Adobe Photoshop 7.0 (Adobe Systems inc.). Microscopy images of MoPIL2-RFP localisation were collected by Dr. M. J. Egan (School of Biosciences, University of Exeter).

6.2.3 Cloning the MGG_00153 coding sequence from *M. oryzae* cDNA

Primers were designed to amplify the coding sequence (CDS) of MGG_00153 by consulting version 6 of the BROAD-MIT *Magnaporthe oryzae* genome database (Dean *et al*., 2005) on http://www.broad.mit.edu/annotation/genome/magnaporthe_grisea/Home.html. The sense primer includes the start codon and a 5' BamHI recognition sequence: Pil1_magnaF 5'-GGATCCATGCATCGCACGTTCT-3'. The anti-sense primer includes the stop codon and a 5' EcoRI recognition sequence: Pil1_magnaR 5'-GAATTCTTAGGCAAGAACGGCTTC-3'. A PCR reaction was prepared by mixing 1.5 µg *M. oryzae* cDNA, 1µl Pil1_magnaF (20 pmol/µl), 1 µl Pil1_magnaR (20 pmol/µl), 5 µl Green GoTaq ® Flexi Buffer (5X), 2.5 µl $MgCl_2$ solution (25 pmol/µl), 0.5 GoTaq® DNA polymerase (5u/µl) and 12.5 µl nuclease free H_2O. A 1,098 base pair DNA fragment was amplified by incubating the PCR reaction in a thermal cycler with the following program: step 1. 95 ºC for 5 minutes; step 2. a cycle of 95 ºC for 30 seconds, 56 ºC for 30 seconds and 72 ºC for 1 minute, repeated 30 times; step 3. 72 ºC for 10 minutes. The products of the PCR, was analysed by DNA gel electrophoresis to identify the amplified DNA fragment. The amplicon was excised from the gel, purified and cloned into pSC-A (Agilent Technologies) which was used to transform a batch of *E.*

coli competent cells. Plasmid preparations of bacterial colonies that successfully transformed were carried out. The cloned DNA fragment were sequenced with the M13 primers (forward and reverse) to confirm presence of correct MGG_00153 CDS. Sequencing reads were assembled with Sequencher (GenoCodes Corporation). For further details on the methodologies used in the cloning procedure, please see Sections 2.2.3-22.7 and 2.2.10.

6.2.4 Rapid amplification of 5' and 3' MGG_11731 cDNA ends

To empirically determine the coding sequence of MGG_11731 I carried out rapid amplification of cDNA 5' and 3' ends (RACE). First I treated total *M. oryzae* RNA to add the GeneRacer™ RNA Oligo (Invitrogen) to the 5' end of full length mRNA only. I then reverse transcribed (RT) the mRNA with the GeneRacer™ Oligo dT Primer (Invitrogen) for cDNA first strand synthesis (the procedures used are described in section 2.2.9). To amplify the 5' of MGG_11731, a PCR was prepared with 3 µl GeneRacer™ 5' end primer (10 µM), 1 µl anti-sense gene specific primer: GPS_n3 5'-TTGCTCTTGAGGCGTGCAATCTCAT-3' (10 µM), 1 µl RT template, 5 µl High Fidelity PCR Buffer (10X), 1 µl dNTP solution (10 mM each), 0.5 µl Platinum® *Taq* DNA Polymerase High Fidelity, 2 µl $MgSO_4$ (50 mM) and 36.5 µl nuclease-free H_2O. To amplify the 3' end of MGG_11731, a PCR was prepared with 3 µl GeneRacer™ 3' end primer (10 µM), 1 µl sense gene specific primer: GPS_n5 5'-AAGCACATTCGCAACACGGAAAGGA-3' (10 µM), 1 µl RT template, 5 µl High Fidelity PCR Buffer (10X), 1 µl dNTP solution (10 mM each), 0.5 µl Platinum® *Taq* DNA Polymerase High Fidelity, 2 µl $MgSO_4$ (50 mM) and 36.5 µl nuclease-free H_2O. The two PCRs were incubated in a thermal cycler with the following reaction

conditions: Step 1. 94 ºC for 2 minutes; Step 2. a cycle of 94 ºC for 30 seconds and 72 ºC for 1 minute, repeated 5 times; Step 3. a cycle of 94 ºC for 30 seconds and 70 ºC for 1 minute, repeated 5 times; Step 4. a cycle of 94 ºC for 30 seconds, 65 ºC for 30 seconds, and 72 ºC for 50 seconds, repeated 25 times; Step 5. 72 ºC for 10 minutes. All PCR products were analysed by gel electrophoresis. Amplicons were excised, purified, and cloned into pSC-A (Agilent Technologies) which was sequenced with vector specific primers M13f and M13r. (For details on these procedures, please see Section 2.2.4).

6.2.5 Cloning MGG_11731 coding sequence from *M. oryzae* cDNA

To clone the coding sequence of MGG_11731, primers were designed from the MGG_11731 5' and 3' end sequences, which were obtained using RACE, as described in section 6.2.4. The sense primer includes the start codon and a 5' end *Bam*HI recognition sequence: F11731 5'-GGATCCATGAATCGAAGCTTCTCCA-3'. The anti-sense primer includes the stop codon and a 5' end *Eco*RI recognition sequence: R11731 5-GAATTCCTACGCCACTTCACTCGCTT-3. A PCR reaction was prepared by mixing 1.5 µg *Magnaporthe oryzae* cDNA, 1µl F11731 (20 pmol/µl), 1 µl R11731 (20 pmol/µl), 5 µl Green GoTaq® Flexi Buffer (5X), 2.5 µl $MgCl_2$ solution (25 pmol/µl), 0.5 GoTaq® DNA polymerase (5u/µl) and 12.5 µl nuclease free H_2O. A XX base pair DNA fragment was amplified by incubating the PCR reaction in a thermal cycler set with the following program: 1. 95 ºC for 5 minutes; 2. a cycle of 95 ºC for 30 seconds, 56 ºC for 30 seconds and 72 ºC for 1 minute, repeated 30 times; 3. 72 ºC for 10 minutes. The amplicon was cloned into a plasmid vector and sequenced as described in sections 2.2.5 and 2.2.10.

6.2.6 Building *S. cerevisiae* transformation vectors

The pYES2:MoPil1 and pYES2:MoPil2 vectors were created by cloning the cDNA coding sequences of MGG_00153 and MGG_11731 in the *S. cerevisiae* transformation vector, pYES2 (Invitrogen). The *S. cerevisiae pil1*Δ mutant strain was transformed with pYES2+MGG_001153, pYES2+MGG_11731 and pYES2 (empty vector for control) using the acetate/single-stranded carrier DNA/polyethylene glycol transformation method (Gietz & Woods, 2002), as described in section 2.2.11.2. Successful transformants were selected on uracil-deficient synthetic complete (SC-U) media agar plates (see Section 2.2.11.2 for full media composition). Yeast colony PCR was performed with the zymolyase method (H. R. Chen *et al.*, 1995) (see section 2.2.9.2 for details). Expression of MGG_ 00153 and MGG_11731, was induced by growing transformed strains in activation media (SC-U with galactose, see section of 2.2.9.3 for full media composition). To test expression of MGG_00153 and MGG_11731, reverse-transcription PCR was performed on RNA extracted from *pil1*Δ:MGG_00153 and *pil1*Δ:MGG_11731 strains grown overnight on activation media (See sections 2.2.2.2.2 and 2.2.6.2 for details). The primers used for the RT-PCR reactions were Pil1_nstdF 5'-ACCTGGCTAAGAAGCTGTCG-3' and Pil1_nstdR 5'-GGTGTGACTGGTGTGTC-GTC-3' for MGG_00153 and Pil2_nstdF 5'-CTCAACAATCAGCCCGAG-3' and Pil2_nstdR 5'-ATTATCTGTTTTTCGGCACG-3' for MGG_11731.

6.2.7 Phenotypic study of transformed *S. cerevisiae* strains

To analyse the phenotypes of mutant and transformed strains, S. cerevisiae cultures were subjected to the following treatments: a sensitivity assays with Calcofluor White, a

sensitivity assay with Congo Red; a sensitivity assay with SDS; and cell viability assay following heat shock treatment. For the Calcofluor White, Congo Red and SDS sensitivity assays, agar plates consisting of YPD and three different concentrations of the desired chemical (20 and 40 µg/ml for Calcofluor White assays, 200 and 400 µg/ml for Congo Red assays and 10 and 40 µg/ml for SDS assays) were prepared. *S. cerevisiae* cell cultures were grown to OD_{600} = ~0.7 in liquid activation media. The cultures were first diluted to approx. 1 x 106 cells/ml, and serial dilutions were carried out. A 5 µl aliquot of each serial dilution was used to inoculate in all assays. For the cell viability assays, *S. cerevisiae* cell cultures were grown overnight in activation media (uracil was added for non pYES2 transformed strains) at 30 ºC with shaking. The cultures were diluted to OD_{600} = 0.3 and grown for a further 3 h into log-phase (OD_{600} = 0.5 - 1). The cultures were then serially diluted to a final concentration of 1.2 x 10^4 cells/ml. For each strain, one culture was incubated at 30 ºC and another heat shocked at 47 ºC for 80 min. An aliquot of 30 µl of the heat shocked cell suspensions, and an aliquot of 30 µl of the non-heat shocked cell suspension, were plated in triplicate on YPD agar plates which were then incubated at 30 ºC for 48 hours. *S. cerevisiae* colonies were counted and cell viability expressed as the number of viable yeast cells after heat shock over number of viable yeast cells kept at 30 ºC. The experiments were repeated three times.

6.3 Results

6.3.1 Generating a *M. oryzae* strain expressing MoPil1-GFP

To investigate the expression and localisation of *MoPIL1*, a reporter gene construct comprising the promoter region and open reading frame (ORF) of MoPil1 and green

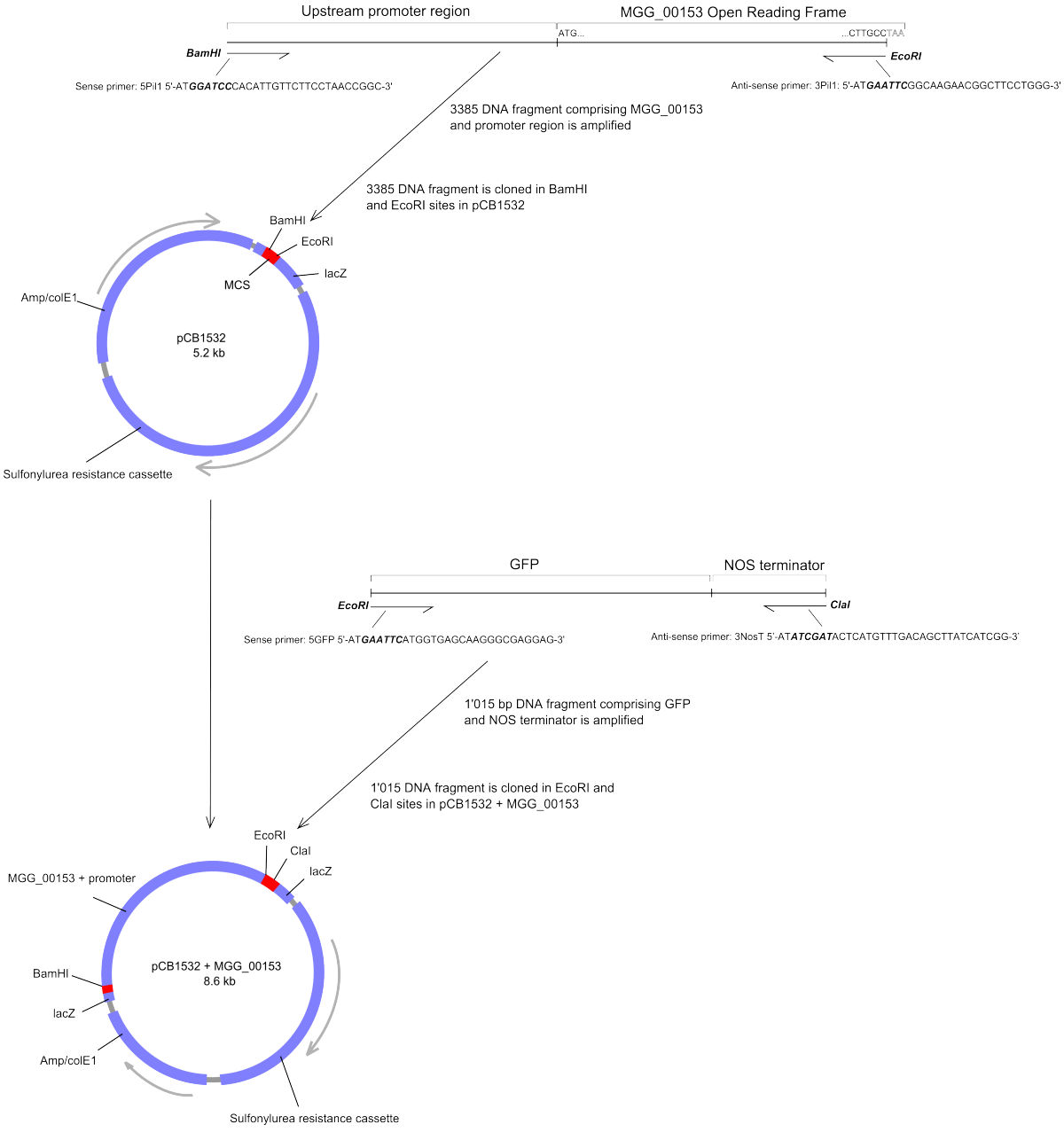


Figure 6.3 Cloning strategy to generate MoPil1-GFP gene fusion. Schematic representation depicting the cloning strategy used to generate a MoPil1-GFP construct in the *M. oryzae* transformation vector pCB1532.

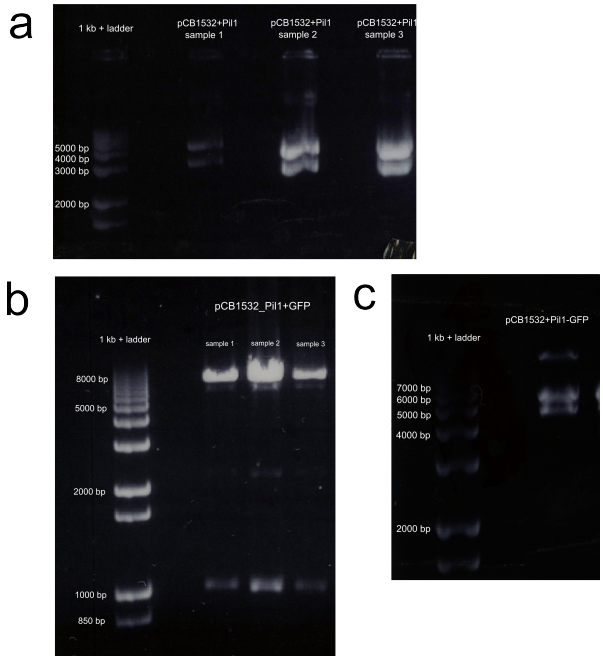


Figure 6.4 Generating a MoPil1-GFP gene fusion construct. a Restriction enzyme digestion of pCB1532+MGG_00153 (promoter and ORF) samples with *Bam*HI and *Eco*RI producing a 5.2 kb DNA fragment (pCB1532) and a 3,385 DNA fragment (MGG_00153 promoter and ORF). **b** Restriction enzyme digestion of pCB1532-MGG_00153+GFP *Eco*RI and *Cla*I producing a 8.6 kb DNA fragment (pCB1532-MGG_00153) and a 1.0 kb DNA fragment (GFP + NOS terminator). **c** Restriction enzyme digestion of a pCB1532+MGG_00153+GFP with *Bam*HI and *Cla*I producing a 5.2 kb DNA fragment (pCB1532) and a 4.4 kb DNA fragment (MGG_00153 + GFP). 1 kb + ladder was used (Invitrogen)

fluorescent protein (GFP) was constructed (See Figure 6.3 for cloning strategy). To do this the promoter region and ORF of *MoPIL1* was cloned from *M. oryzae* genomic DNA. The clone was then ligated to the linearised transformation vector pCB1532 (Figure 6.4a). A 1015 bp DNA fragment consisting of GFP and the NOS terminator from the pCAMBgfp plasmid (Sesma & Osbourn, 2004) was then isolated. The MoPil1-GFP construct was created by ligating a GFP-NOS_terminator DNA fragment to *MoPil1* (Figure 6.3 & Figure 6.4b). Presence of the MoPil1-GFP construct in pCB1532 was confirmed by restriction enzyme digestion (Figure 6.4c). To confirm the Pil1

promoter and ORF was cloned in frame with GFP, the construct was sequenced in both directions with MoPil1 and GFP specific primers. The *M. oryzae* Guy-11 strain was transformed with the pCB1532-MoPil1-GFP vector and transformants selected by resistance to sulfonylurea (Sweigard, 1997).

6.3.2 Sub-cellular localisation of MoPil1-GFP and MoPil2-RFP in *M. oryzae* conidia and hyphae

The sub-cellular localisation of MoPil1-GFP was observed in different cell types and at the different developmental stages of the *M. oryzae* life cycle (Wilson & Talbot, 2009). In the three celled conidium, MoPil1-GFP localised to discrete puncta that were evenly distributed at the cell cortex (Figure 6.5a). This is consistent with localisation of eisosomes in *S. cerevisiae* cells (Walther *et al*., 2006). As the germ tube extended from the apical cell and appressorium began to form, the MoPil1 signal became diffuse in the cytosol of the germ tube and nascent appressorium, while maintaining punctate patches in the three conidial cells (Figure 6.5b,c). However, in mature appresoria, MoPil1-GFP localised to discreet puncta, in a ring formation around the central appressorium pore (Figure 6.5d). In *M. oryzae* mycelium, the localisation of Pil1-GFP was found to be diffuse in actively growing hyphal tips, but in older hyphae, it localised to a discrete patch in the centre of an established septum, or as a cluster of patches during the process of septation (Figure 6.6a,b). To compare the sub-cellular localisation of MoPil1 with MoPil2, a *M. oryzae* strain co-expressing MoPil1-GFP and MoPil2-RFP was generated. Unlike MoPil1-GFP, the localisation of MoPil2-RFP in conidia was either diffuse in the cytoplasm, or was concentrated into bright patches which coincided with septal pores (Figure 6.7 a,b). In mature appresoria, Pil2-RFP did not co-localise with MoPil1-GFP

punctate patches, but was diffusely distributed around the central pore (Figure 6.8). In hyphae, Pil1-GFP and Pil2-RFP colocalised to septa and to sparse punctate patches along the periphery of hyphae (Figure 6.9).

6.3.3 Functional complementation of *S. cerevisiae Δpil1* mutant by *M. oryzae* MoPil1 and MoPil2

The functional relatedness of MoPil1 and MoPil2 was analysed by complementation analysis with a *pil1*Δ mutant of *S. cerevisiae*. To do this, differences in response to stress factors in *S. cerevisiae* wild type and *pil1*Δ strains were identified. Transformation vectors with the coding sequences of MoPil1 and MoPil2 were then constructed and used to transform *S. cerevisiae pil1*Δ strains. Finally, stress resistance assays were carried out to test whether MoPil1 and MoPil2 recover wild type phenotype.

6.3.3.1 Differential stress resistance in *S. cerevisiae* wild type and *pil1*Δ strains

Deletion of Pil1 is non-lethal in *S. cerevisiae* and it is therefore important to identify detectable phenotypic variations between *S. cerevisiae* wild type and *pil1*Δ strains. According to Walther *et al.* (Walther *et al*., 2006), deletion of Pil1 causes aberrant plasma membrane invaginations. However, this phenotype can be confirmed only by transmission electron microscopy. A simpler way of testing phenotypic variation is to look at stress resistance to either heat shock and to chemicals. In Zhang *et al.* (Zhang *et al*., 2004), the authors report that Pil1 and Lsp1 negatively regulate resistance to heat stress. By performing cell viability studies, they showed that *S. cerevisiae pil1*Δ cells, are twice as resistant to heat stress as wild type cells. I confirmed this observation by treating *S. cerevisiae* wild type and *pil1*Δ mutant strains (see section 2.2.1 for genetic

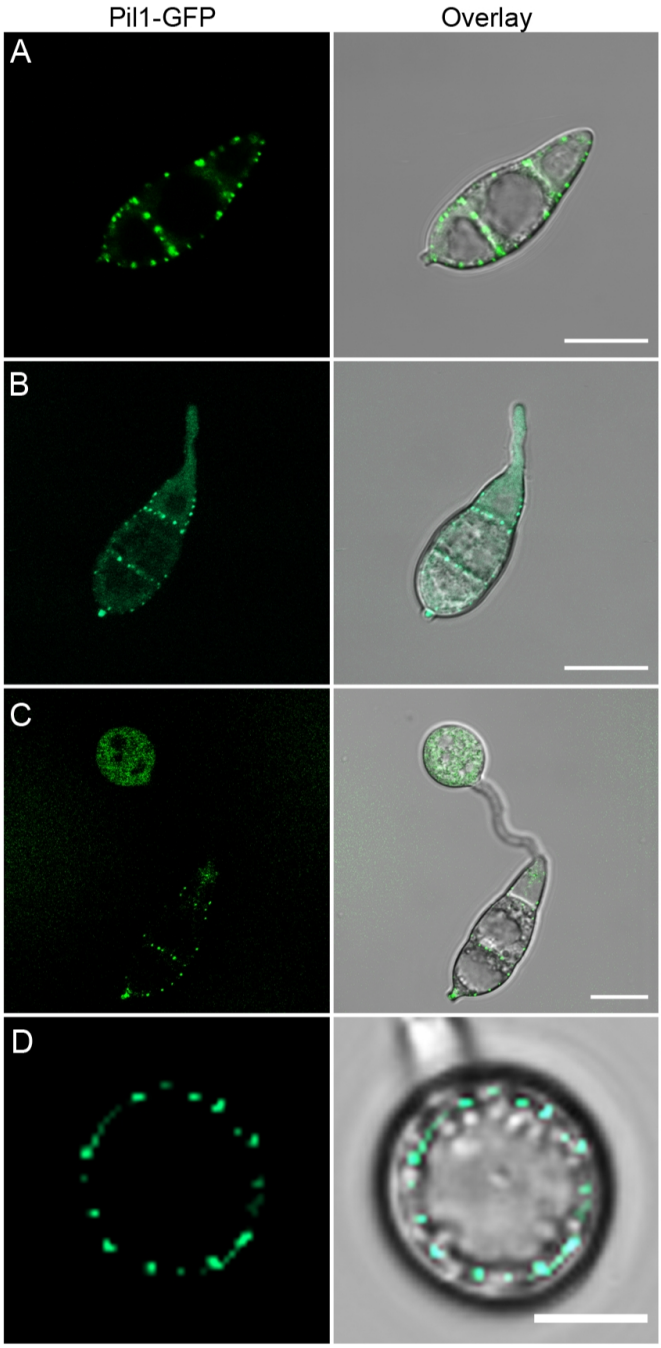


Figure 6.5 Sub-cellular localisation of MoPil1-GFP at different stages of *M. oryzae* conidia germination and appressorium formation. Conidia were inoculated onto glass coverslips and incubated in a moist chamber at 26ºC. **a** MoPil1-GFP localised to bright punctate patches at the cell cortex of conidia. **b,c** MoPil1-GFP signal was diffuse within the cytoplasm in the elongating germ tube and nascent appressorium, but maintained punctate patches in conidia. **d** MoPil1-GFP signal was distributed in bright punctate patches around the central pore in infection-ready appresoria. Scale = 10 μm

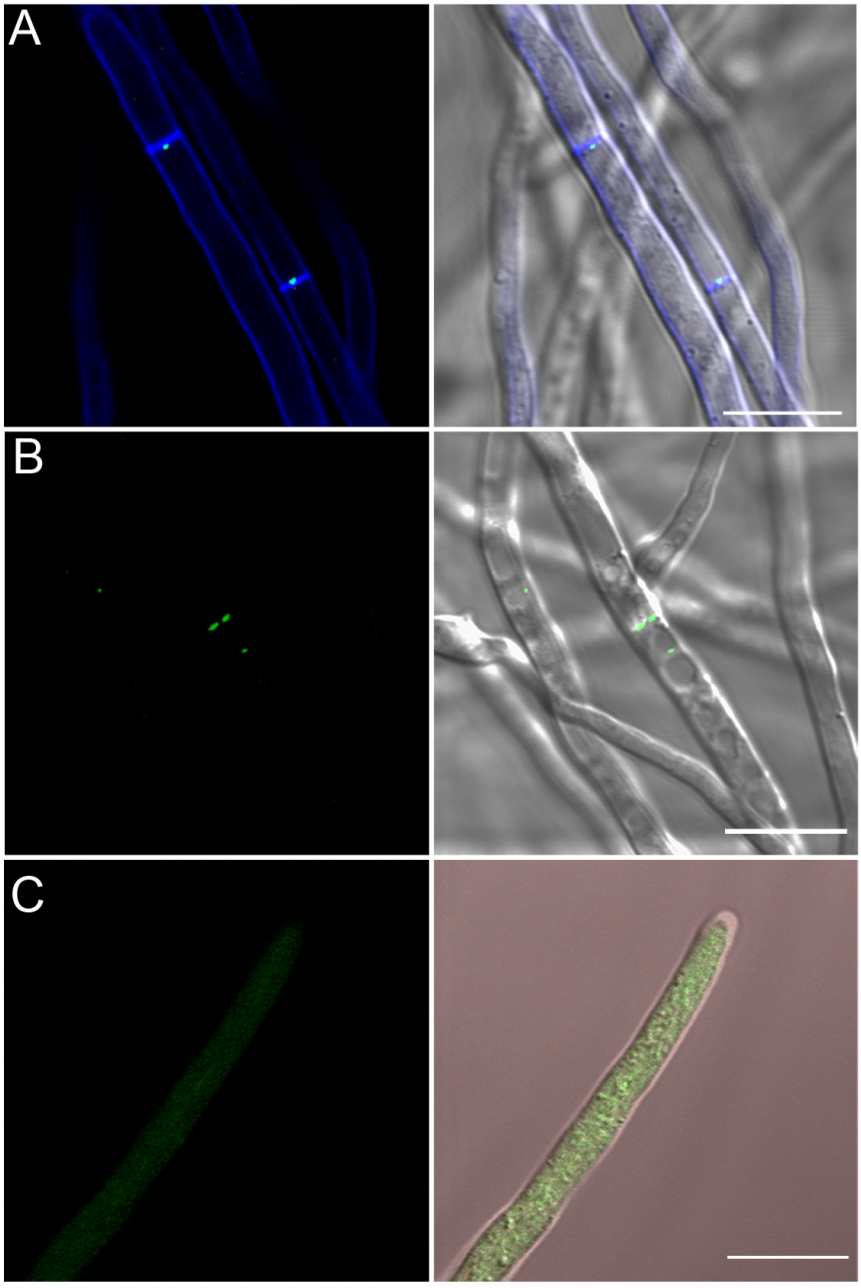


Figure 6.6 Sub-cellular localisation of MoPil1-GFP in *M. oryzae* hyphae. Vegetative hyphae were grown in liquid CM for 48 h, and incubated in Calcofluor White for 5 min prior to confocal imaging. **a** MoPil1-GFP localises at discrete patches consistent with the sites of septal spores (blue signal is Calcofluor White). **b** MoPil1-GFP localises with septa. **c** MoPil1-GFP signal was diffuse in the cytoplasm of actively growing hyphal tips. Scale = 10 μm.

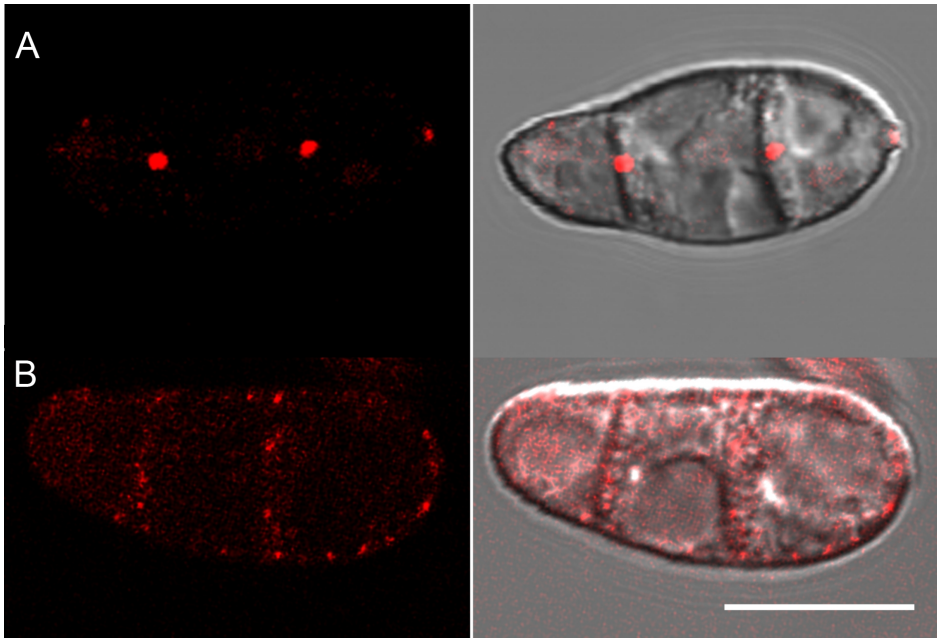


Figure 6.7 Sub-cellular localisation of MoPil2-RFP in *M. oryzae* conidia. Conidia were inoculated onto glass coverslips and incubated in a moist chamber at 26ºC. **a** MoPil2-RFP localised at discrete patches consistent with the sites of septal spores. **b** MoPil2-GFP signal was dispersed in the cytoplasm of conidia. Some patches on the cell periphery can be observed. Scale = 10 µm. (Images obtained by Egan M.J., see section 6.2.2)

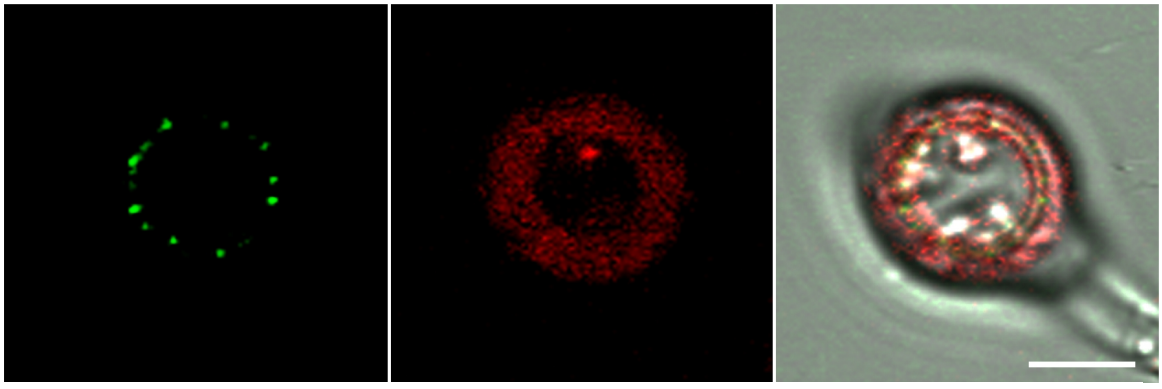


Figure 6.8 MoPil1-GFP and MoPil2-RFP do not colocalise in mature appresoria of *M. oryzae*. MoPil1-GFP localised at punctate patches around the central pore of infection-ready appresoria. MoPil2-RFP signal is diffuse in the cytoplasm around the central pore. Scale = 10 µm (Images obtained by Egan M.J., see section 6.2.2)

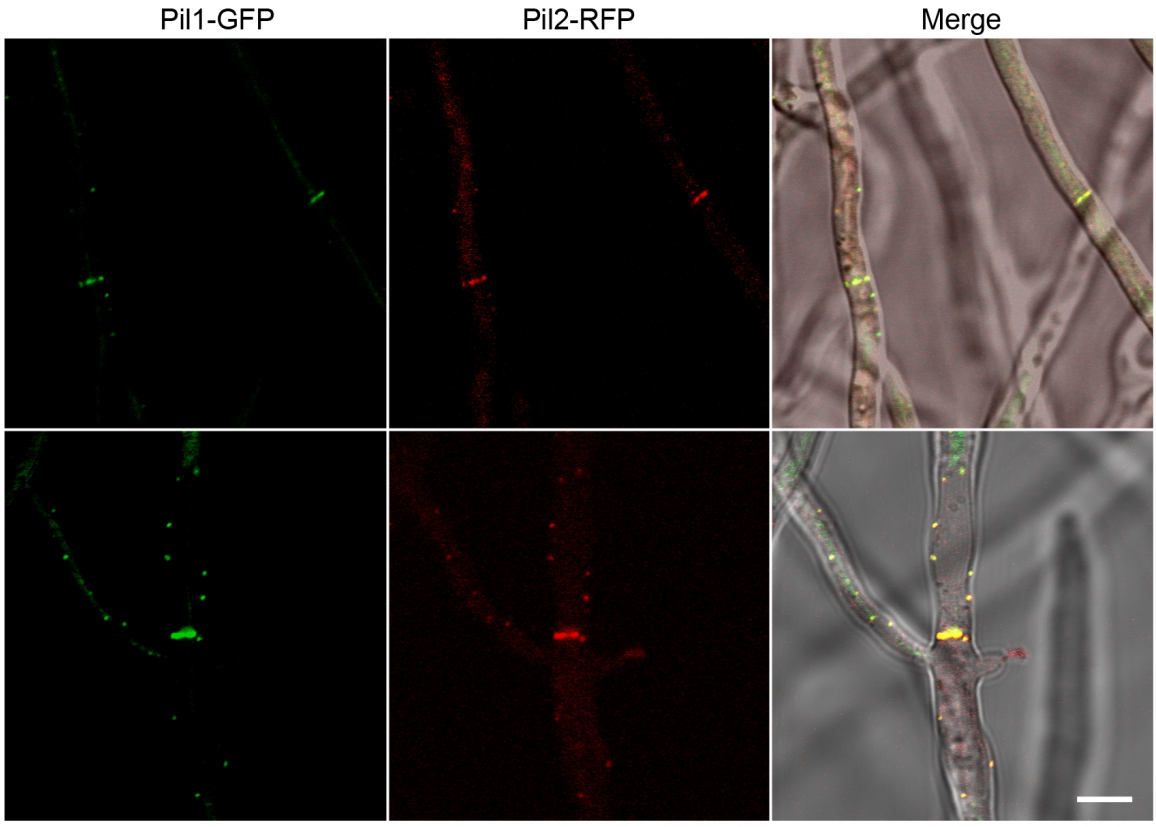


Figure 6.9 MoPil1-GFP and MoPil2-RFP co-localise in vegetative hyphae of *M. oryzae*. MoPil1 and MoPil2 localisation was examined in the hyphae of a *M. oryzae* strain co-expressing Pil1-GFP and Pil2-RFP fusion proteins. Both MoPil1-GFP and MoPil2-RFP localise in septa and in sparse punctate patches at the cell periphery of the hyphae. Scale = 10 μm (Images obtained by Egan M.J, see section 6.2.2)

specification) with heat stress at 48 ºC for 80 mins (see section 6.2.7). The average cell viability of triplicate experiments was 0.409 (standard deviation = 0.053) for wild type strain and 0.757 (standard deviation = 0.065) for *pil1*Δ strains (Figure 6.14a). P-value for Student's t-test on the two data samples is 0.002 (Figure 6.14b). Cell viability after heat stress at 48 ºC for 80 mins was therefore 1.85 higher for *pil1*Δ than for wild type. To test for differential stress resistance in *pil1*Δ and wild type cells, log phase *S. cerevisiae* cells were also treated with cell wall integrity disruptors Calcofluor White

and Congo Red (Nikolaou *et al*., 2009), and with Sodium Dodecyl Sulfate (SDS). Mutant *pil1*Δ cells treated with Calcofluor White, showed higher rates of growth than wild type cells. This was found for the 1 x 10^3 cells/ml suspension in the 20 μg/ml assay, and the 1 x 10^4 cells/ml suspension in the 40 μg/ml assay (Figure 6.10). Wild type cells treated with Congo Red, showed higher rates of growth than mutant *pil1*Δ cells. This was found for the 1 x 10^3 cells/ml suspension in the 200 μg/ml assay, and the 1 x 10^4 cells/ml suspension in the 400 μg/ml assay (Figure 6.10). Wild type cells at 1 x 10^2 cells/ml concentration, treated with 40 μg/ml SDS, showed higher growth rates than *pil1*Δ cells (Figure 6.10).

6.3.3.2 Construction of pYES2:MoPil1 and pYES2:MoPil2 transformation vectors

The 1098 bp DNA fragment comprising the coding sequence of MoPil1, was cloned in a sequencing vector, and then ligated into the *S. cerevisiae* transformation vector pYES2, as described in sections 6.2.3 and 6.2.6. Presence of MoPil1 coding sequence was confirmed by restriction enzyme digestion (Figure 6.11). A sample of pYES2:MoPil1 vectors, was sequenced to ensure the ORF does not contain errors introduced by the PCR or by cloning.

MGG_11731 sequence information on version 6 of the BROAD-MIT *Magnaporthe oryzae* genome database (Dean *et al*., 2005), was used to design primers to amplify the MoPil2 coding sequence. However, PCR failed with different concentrations of template cDNA and at different annealing temperatures, whereas in all positive control PCRs, a DNA fragment was amplified using *M. oryzae* genomic DNA as template (Figure 6.12a). The PCRs with nested MoPil2 specific primers generated an amplicon,

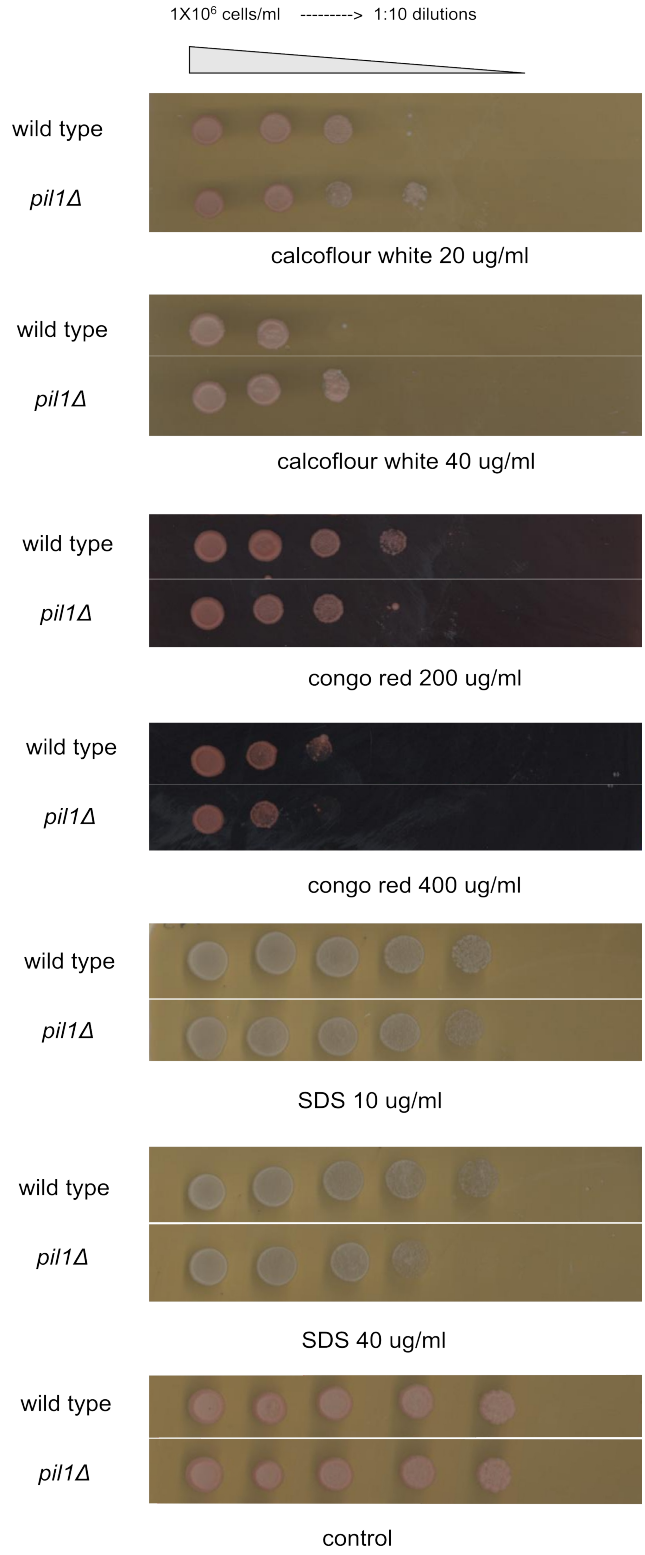


Figure 6.10 Stress resistance to Calcoflour White, Congo Red and SDS in *S. cerevisiae* wild type and *pil1*Δ strains. Cells were grown to log phase, diluted to starting concentration of ~1 x 10^6 cells/ml with four subsequent 1:10 serial dilutions. 5 µl samples from each of the dilutions were plated on YPD agar treated with 20 and 40 µg/ml Calcofluor White, 200 and 400 µg/ml Congo Red, and 10 and 40 µg/ml and on untreated YPD agar for control.

which demonstrated that MoPil2 was expressed at the time of RNA extraction, and the RNA or cDNA was not subsequently degraded. To empirically determine the 5' and 3' ends of the MoPil2 CDS, rapid amplification of cDNA ends (RACE) was carried out (Figure 6.12b). The 5' and 3' ends of MoPil2 CDS were sequenced and ligated to obtain the full MoPil2 CDS. The data also demonstrated transcript information for MGG_11731 on version 6 of the BROAD-MIT *M. oryzae* genome database was incorrect at the 5' end.

The sequence data thus obtained was used to design primers to amplify MoPil2 CDS. The coding sequence of MoPil2 was cloned in a sequencing vector, and then ligated into *S. cerevisiae* transformation vector pYES2, as described in section 6.2.5 and 6.2.6. Presence of MoPil2 coding sequence was confirmed by restriction enzyme digestion (Figure 6.12c). A sample of pYES2:MoPil1 vectors, was sequenced to ensure the ORF does not contain errors introduced by the PCR or by cloning.

6.3.3.3 Expression of MoPil1 and MoPil2 in transformed *S. cerevisiae* cells

The *S. cerevisiae pil1*Δ strain used for transformation is from the W303 strain background. The genetic specifications are *ura*3-52; *trp1*Δ 2; *leu*2-3,112; *his*3-11; *ade*2-1; *can*1-100; *pil1*Δ::KAN. To confirm absence of Pil1, I carried out a colony PCR reaction with Pil1 specific primers with a control reaction using wild type colonies as template (see Figure 6.13a). Three transformed *S. cerevisiae* strains were generated using methods described in section 6.2.6. I transformed a *pil1*Δ mutant with pYES2:MoPil1 and also transformed a *pil1*Δ mutant with pYES2:MoPil2. The *pil1*Δ strain was also transformed with the empty pYES2 vector as a control experiment.

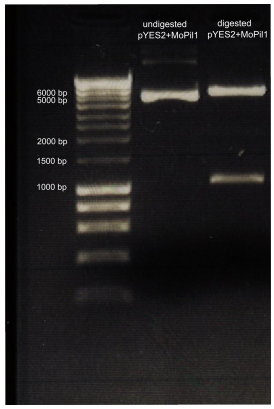


Figure 6.11 Restriction Enzyme digestion of pYES2:MoPil1. An aliquot of 100 ng of pYES2:MoPil1 was digested with *Bam*HI and *Eco*RI to confirm presence of the MoPil1 coding sequence. The digestion produced a 1.1 kb DNA fragment (MoPil1 CDS) and a 5.9 kb DNA fragment (linearised pYES2). An aliquot of 100 ng of undigested pYES2:MoPil1 was run for control. Hyperladder I was used to determine size of DNA fragments (Bioline)

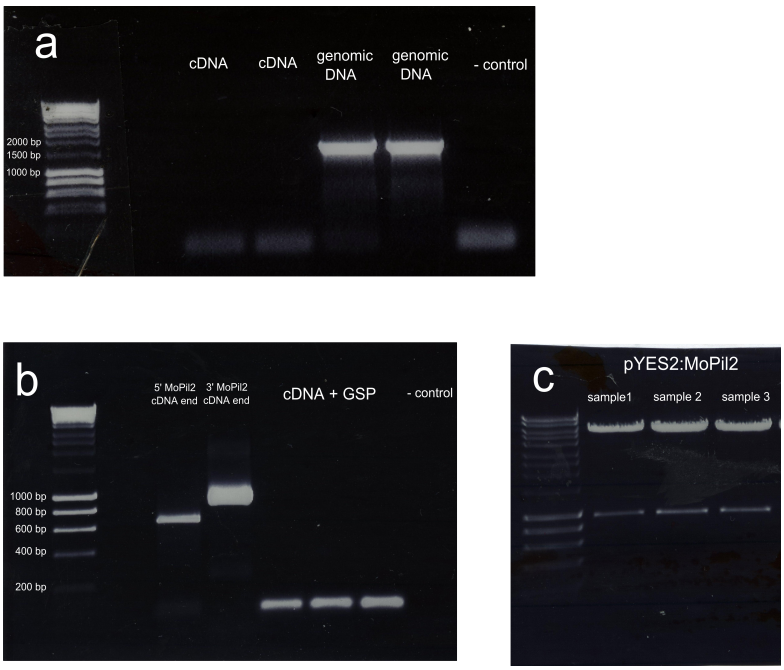


Figure 6.12 Cloning the MoPil2 coding sequence. a Primers were designed from the *M. oryzae* database but failed to amplify MoPil2 CDS from cDNA template. **b** Amplification of 5' and 3' MoPil2 cDNA ends. Reverse trascribed RNA treated to exclude non mRNA and truncated mRNA from total RNA, and with GenePrimer RNA Oligo ligated to 5' end, was used as template. **c** Restriction enzyme digestion of ~100 ng pYES2:MoPil2 with BamHI and EcoRI. Digestion produced 949 base pair DNA fragment identified as MoPil2 CDS and 5.9 kb DNA fragment identified as linearised pYES2). Hyperladder I was used to determine DNA fragment size.

Transformed colonies were selected by growth in uracil deficient chemically growth medium and maintained in the same media with glucose. I confirmed the presence of MoPil1 and MoPil2 in each respective transformed *pil1*Δ strains, by colony PCR (Figure 6.13b). To induce expression of MoPil1 and MoPil2, the transformed strains were transferred to induction media, which is chemically defined, lacks uracil, and uses galactose as its carbon source. To test that both genes were expressed, *pil1*Δ:MoPil1, *pil1*Δ:MoPil2 and *pil1*Δ:pYES2 cells cultures were grown to log phase in both glucose based and galactose based chemically defined medium. Total RNA was then extracted from each cell culture and used as template for one-step RT-PCR reactions, with each reaction including the appropriate gene-specific primers. For the *pil1*Δ:MoPil1 and *pil1*Δ:MoPil2 cultures grown in galactose, gene-specific DNA fragments were amplified, for both *pil1*Δ:pYES2 cultures and for the other cultures grown in glucose, no DNA fragments were amplified (Figure 6.13c).

6.3.3.4 Analysis of *S. cerevisiae pil1*Δ:MoPil1 and *pil1*Δ:MoPil2 transformants

Although there was some difference in stress resistance to Calcofluor White and Congo Red in between wild type and *pil1*Δ *S. cerevisiae* strains, these were insufficient to test recovery of wild type phenotype by MoPil1 and MoPil2. Instead, resistance to heat stress was used as a method to determine functional complementation because clear differences could be observed (see section 6.2.7 for details on methods). Heat stress resistance assays were thus carried out on *pil1*Δ:MoPil1, *pil1*Δ:MoPil2 and *pil1*Δ:pYES2 (control) strains (the first heat shock assay run is shown in Appendix 5). The average cell viability for *pil1*Δ:MoPil1 is 0.452 with standard deviation 0.076, the average cell viability for *pil1*Δ:MoPil2 is 0.572 with standard deviation 0.081 and the

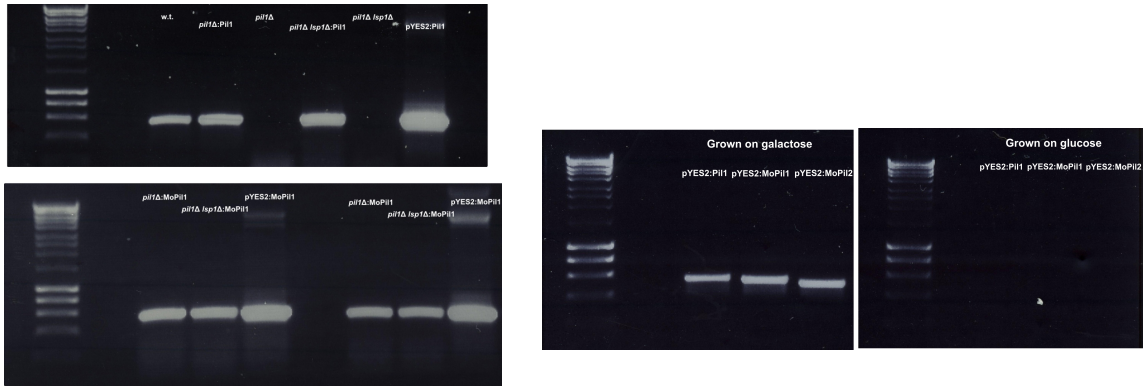


Figure 6.13 Expression of MoPil1 and MoPil2 in *S. cerevisiae pil1*Δ mutant strains. a. Colony PCR confirmed presence of inserted MoPil1 and MoPil2 in transformed *pil1*Δ mutant strains. **b.** Expression of MoPil1 and MoPil2 was tested by RT-PCR. Cultures were grown in galactose based minimal induction media and in glucose based minimal media. MoPil1 and MoPil2 specific primers were used for the RT-PCRs. Hyperladder I was used to determine size of DNA fragments (Bioline)

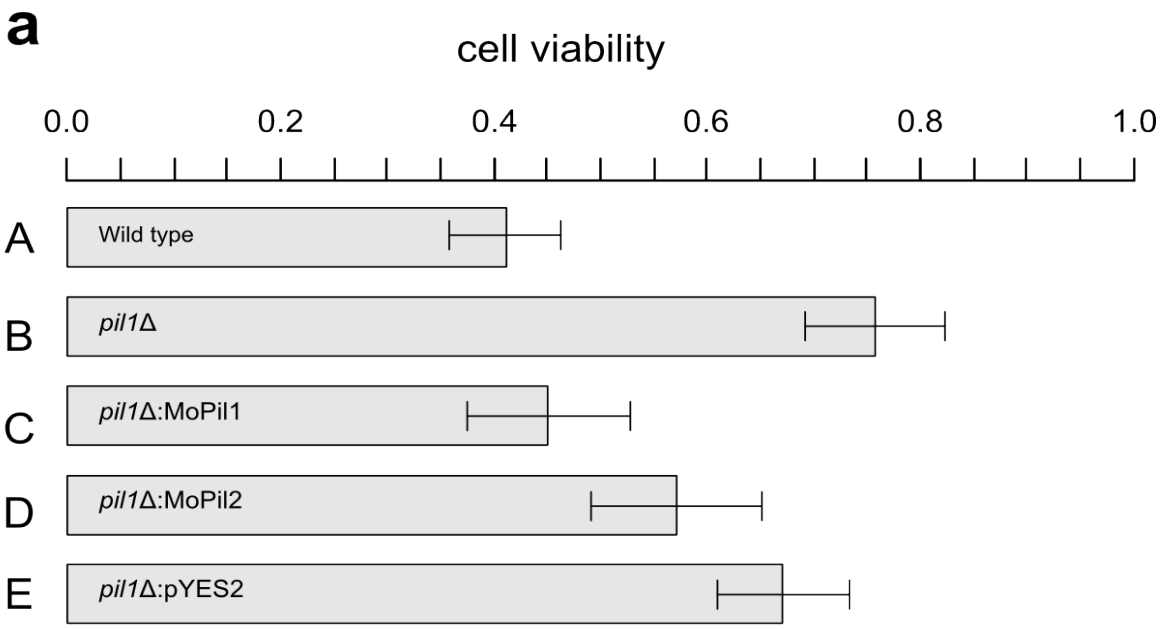
average cell viability for *pil1*Δ:pYES2 is 0.670 with standard deviation 0.064 (Figure 6.14a). Because these data sets are normally distributed, Student's t-tests were used to estimate statistical significance of the results. P-value for t-test between *pil1*Δ:MoPil1 and *pil1*Δ:pYES2 (control) data sets is 0.017 which rejects the null hypothesis that MoPil1 does not reduce resistance to heat stress (Figure 6.14b). However, P-value for the t-test between pil1Δ:MoPil2 and pil1Δ:pYES2 data sets was 0.158, which supports the null hypothesis that MoPil2 does not reduce resistance to heat stress (Figure 6.14b). I also calculated t-tests between *pil1*Δ:MoPil1 and wild type data sets and between *pil1*Δ:MoPil1 and *pil1*Δ data sets, obtaining p-values of 0.492 and 0.006, respectively. These results support the null hypothesis that wild type and *pil1*Δ:MoPil1 means do not differ, and reject the null hypothesis that *pil1*Δ:MoPil1and *pil1*Δ means do not differ. T-tests between *pil1*Δ:MoPil2 and wild type data sets and *pil1*Δ:MoPil2 and *pil1*Δ data sets give p-values of 0.043 and 0.035 which rejects null hypotheses that the

*pil1*Δ:MoPil2 mean does not differ from the wild type or the *pil1*Δ means (Figure 6.14b). Taken together, this data suggests that MoPil1 restores the wild type phenotype to a *pil1*Δ mutant, and therefore there is functional complementation between MoPil1 and *S. cerevisiae* Pil1. However, results are inconclusive for MoPil2 because the difference is not significant compared to *pil1*Δ:pYES2 (Figure 6.14b).

6.4 Discussion

The phylogenetic evidence presented in Chapter 5, suggests multiple gene duplications have occurred in the evolutionary history of eisosomes. The presence of diverse paralogues may therefore reflect divergence of eisosomal functions. This is supported by a recent publication (Vangelatos *et al.*, 2010), where the authors studied two Pil1 and Lsp1 homologues in *Aspergillus nidulans* and suggested that the organisation of eisosomes is different from that found in *S. cerevisiae* (Walther *et al*., 2006). They also found that deletion of the two proteins did not affect endocytosis (Vangelatos *et al*., 2010).

The hypothesis that eisosomes mediate a conserved endocytic function linked with the evolution of the fungal cell wall and fungal endocytosis, was investigated by characterising the Pil1 and Lsp1 homologues of filamentous fungus *M. oryzae*. This is important also because of the two gene duplications found in ascomycetes: one at the base of the Saccharomycotina which produced the Pil1 and Lsp1 paralogues and one in the last common ancestor of ascomycetes which produced the Pil1 and Pil2 paralogues (Figure 5.3). Pil2 was retained in Pezizomycotina and Taphrinomycotina but putatively lost in Saccharomycotina. Is Pil2 therefore related to a function that became obsolete for



b

	A	B	C	D	E
A	-	0.002	0.492	0.043	0.005
B	0.002	-	0.006	0.035	0.167
C	0.492	0.006	-	0.131	0.017
D	0.043	0.035	0.131	-	0.158
E	0.005	0.167	0.017	0.158	-

Figure 6.14 Functional complementation between *Saccharomyces cerevisiae* Pil1 and *Magnaporthe grisea* MoPil1 and MoPil2 a. Cell viability was measured by treating cells with heat shock at 47ºC for 80 minutes and plating on YPD medium. Values represented by barchart is the number of viable yeast cells after heat shock over number of viable yeast cells kept at 30ºC. Experiments were performed in triplicates, with values representing the means of the three runs and standard deviation expressed by error bars. **b.** P-value scores for t-tests calculated between the five data samples. Red boxes highlight probabilities of observed data assuming null hypothesis of 1% or lower, blue boxes highlight probabilities of observed data assuming the null hypothesis between 5 and 1%.

ascomycete yeasts?

The data in this chapter showed that MoPil1 is distributed at the plasma membrane of in punctate patches typical of the eisosomes observed in *S. cerevisiae,* whereas MoPil2 localised instead at septal pores, and did not colocalise to any extent with MoPil1. However, in vegetative hyphae, the two proteins colocalised in septa and in sparse puncta at the cell periphery. Functional complementation study suggests that MoPil1 restores wild type phenotype to a *pil1*Δ mutant, but the same could not be confirmed for MoPil2. In terms of involvement in endocytosis, the diffuse distribution in actively growing hyphae, and lack of localisation at the the hyphal tip where endocytosis is known to occur, suggests MoPil1 may not be needed for endocytosis in hyphae. However, recent unpublished data has shown that MoPil1 localises to sites of endocytosis and is necessary for correct endocytosis in *M. oryzae* conidia (M. J. Egan and N. J. Talbot , personal communication). MoPil2 localisation to septa in both conidia and hyphae also suggested that its function may be related to septation and or regulation of septal pores rather than endocytosis. This is supported by data showing that MoPil2 deletion does not affect endocytosis in *M. oryzae* conidia (M. J. Egan and N. J. Talbot , personal communication). Taken together, the data lead to the following conclusions: MoPil1 is functionally homologous to *S. cerevisiae* Pil1 and Lsp1, defines eisosomes in *M. oryzae* conidia where it is involved in endocytosis, but also localises to septa in hyphae. MoPil2 is functionally divergent from Pil1 proteins, is likely to be involved in septation and septal pore regulation in both conidia and hyphae, and is not likely to be involved in endocytosis. These conclusions would become even more robust by demonstrating that MoPil1-GFP localises in the typical Pil1 punctate distribution, while

showing that MoPil2 distributes differently, in *pil1*Δ *Saccharomyces cerevisiae* cells.

The putative involvement of MoPil2 in conidia and hyphal septation may explain why it was lost in Saccharomycotina where the predominant cell form is the budding yeast cell. The fact that the Taphrinomycotina species *Schizosaccharomyces pombe*, where a putative Pil2 homologue is present, can form elaborately branched hyphae which are divided by septation, is consistent with this hypothesis (Amoah-Buahin *et al*., 2005). However, two points can be made against this hypothesis: first, it should be noted that Saccharomycotina species such as *Candida albicans*, *Ashbya gossypii* and *Yarrowia lipolytica* are all pleiomorphic and can grow septate hyphae as well as pseudohyphae and yeast cells (Barth & Gaillardin, 1997; Sanchez-Martinez & Perez-Martin, 2001; J. Wendland & Walther, 2005). *S. cerevisiae* itself, the typical representative for yeast cells, when starved with nitrogen can undergo a dimorphic transition and form pseudohyphae (Gimeno *et al*., 1992). In addition, the loss of Pil2 could be explained by whole genome duplication followed by substantial gene loss and diversification, a pattern which has been postulated to have played an important role in the evolution of budding yeasts (Wolfe & Shields, 1997), although it should be noted that the putative WGD was pinpointed at a branch more derived than the Pil1/Lsp1 duplication (Kellis *et al*., 2004). The precise role of eisosomes therefore remains elusive. Some data, including results presented in this chapter, point towards a role in endocytosis (Walther *et al*., 2006) M. J. Egan and N. J. Talbot , personal communication), but this is refuted in more than one study (Grossmann *et al*., 2008; Vangelatos *et al*., 2010). The role of the eisosome in the LCB-pkh-ypk signalling pathway involved in cell wall remodelling during growth (Luo *et al*., 2008; Zhang *et al*., 2004) and its strong connection with the

transmembrane proteins of the membrane compartment of Can1 (MCC) (Grossmann *et al*., 2008) points towards a fundamental role in cell wall and plasma membrane integration and regulation. The putative function of MoPil1 and MoPil2's in septation at different developmental stages of *M. oryzae* life cycle is consistent with such a role. Other distinct functions may also have evolved as a result of the multiple gene duplications discussed here, because it is also shown by the fact that Pil1 and Lsp1, probably a result of whole-genome duplication in yeast (Kellis *et al*., 2004), show non-redundant functions (Walther *et al*., 2006). It may also be that eisosomes indirectly regulate endocytosis, by recruiting the proteins which define micro-domains across the plasma membrane. This is supported by the punctate and stable distribution at the cell surface which was observed in conidia expressing MoPil1-GFP. Large, fixed, protein complexes could provide an anchoring for endocytic proteins being transported to sites of endocytosis. Further insights into the nature of eisosomes could be obtained by studying Pil1/Lsp1 and Pil2 homologues in taxonomically and morphologically diverse fungal model organisms. Dimorphic yeasts such as *Candida albicans* and *Schizosaccharomyces pombe* would also be of particular interest, as well as the basidiomycete *Ustilago maydis* and zygomycete *Rhizopus oryzae,* for which a system for genetic transformation exists (Michielse *et al*., 2004).

7. General Discussion

In the general introduction of this thesis I addressed the role of endocytosis in eukaryogenesis as an important reason for studying how this cellular process evolved. Out of the numerous models that have been proposed two main types were identified: those that suggest that the first eukaryotic nucleated cell arose from syntrophy between two prokaryotes (Lopez-Garcia & Moreira, 2006; Margulis *et al*., 2000; Martin & Muller, 1998; Moreira & Lopez-Garcia, 1998) and those that instead argue for the emergence of a protoeukaryote lineage, either from within or sister to the Archaea, that autogenously evolved eukaryotic features (although there is disagreement over exactly what those are and in what order they were acquired), and only subsequently, via phagotrophy, they engulfed the α-protobacterium which later evolved into the mitochondrion (Cavalier-Smith, 2002; Hartman & Fedorov, 2002; Jekely, 2007). According to the autogenous models, endocytosis plays key roles, both in terms of the acquisition of the primary mitochondrial endosymbiont (phagocytosis) and the evolution of the endomembrane system (pinocytosis), whereas models based on syntrophy entail endocytosis-independent processes for the origin of the first eukaryotic cell. With regards to the study presented in this thesis, the question is whether it offers a real contribution to the debate over which type of model is more plausible. The quick answer is no, and that simply is because, using the tools of comparative genomics and phylogenetics, we cannot go further back than the last common ancestor of known extant eukaryotic taxa (LCEA) with sequenced DNA data. That cell, the LCEA, already possessed the mitochondrion, as well as all the major eukaryotic benchmarks, thus making it a remarkably complex cell, far surpassing initial predictions (DeGrasse *et al.,* 2009; Eme *et al*., 2009; Field & Dacks, 2009; Koonin, 2009; Liu *et al.,* 2009; Ramesh

et al., 2005; Richards & Cavalier-Smith, 2005). Whether that cell evolved from two prokaryotic organisms, exchanging sulfur or hydrogen, fusing membranes, merging genomes, and only then, evolving the autogenous membrane bound organelles to organise the metabolic compartments of the cell; or whether it evolved from a protoeukaryote cell that had become flexible, phagotrophic and predatory, able to engulf and enslave metabolically advantageous organisms, is not known, or at least it cannot be known via phylogenomic methods, because no *bona fide* intermediate eukaryotic lineages have been found (Embley & Martin, 2006).

So what is the real purpose of this thesis? Let's consider the idea is that although the actual origin of the eukaryotic cell remains an unanswered question, we do know that the way the cell evolved, from its very first step (the first common eukaryotic ancestor - FCEA) to the LCEA, combines element of exogenous and of autogenous evolution. Indeed, mitochondria certainly evolved from a bacterium, but other organelles such as the Golgi, the endoplasmic reticulum and peroxisomes most probably evolved from within the cell (Dacks & Field, 2007; Gabaldon *et al*., 2006; Martin, 1999; Schluter *et al*., 2006). Models of autogenous evolution, whether in the context of eukaryogenesis, or origin of organelles, are thus important to the field of eukaryotic cell evolution and are the focus of this thesis.

For instance, the model of autogenous eukaryotic origin based on the evolution of the Ras superfamily of GTPases (Jekely, 2003, 2007), suggests that eukaryotic endomembranes evolved from tubulation of the plasma membrane to form secretory membranes, from which vesicle scission first occurred (referred to in his writing as

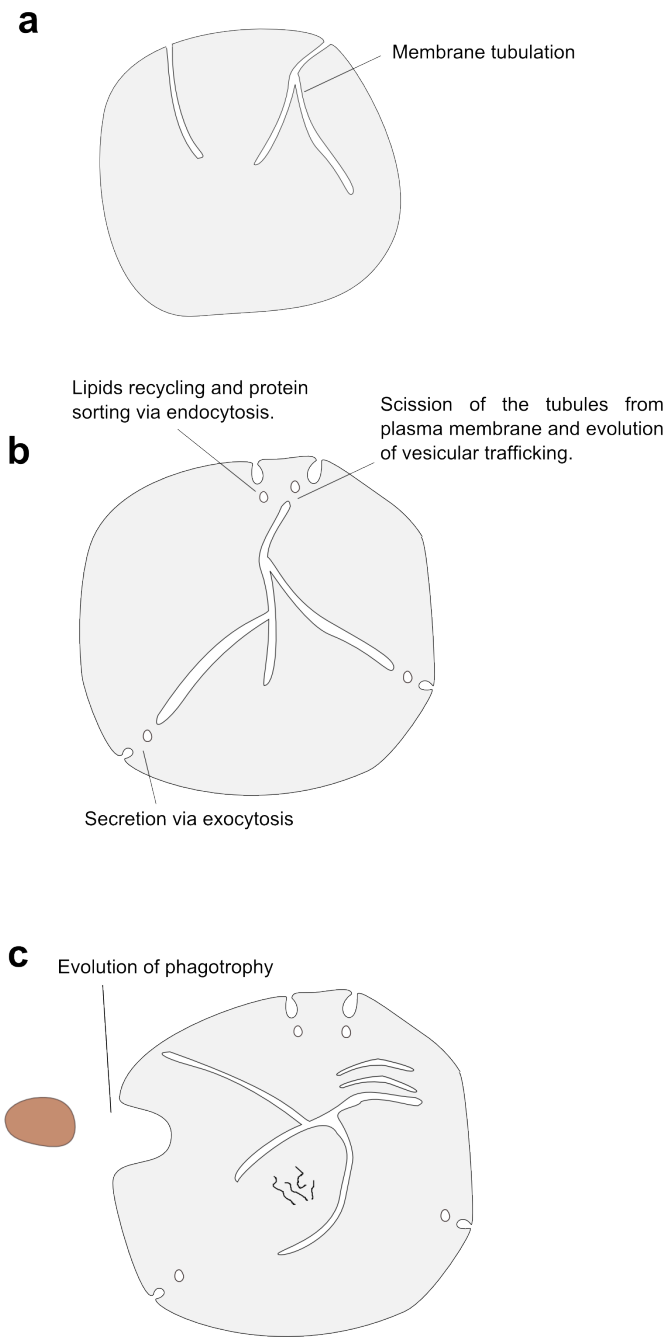


Figure 7.1 Autogenous scenario of endomembrane origin based on Ras GTPase evolution. a Secretory membranes evolved following initial membrane tubulations off the plasma membrane. **b** Scission of tubules allowed separation of tubules from the plasma membrane and led to origin of vesicle trafficking to recycle lipids and sort proteins. **c** Development of a viable membranous secretory system led to evolution of phagotrophic predation, and eventually to acquisition of the mitchondrial ancestor. The model thus proposes pinocytosis evolved prior to phagocytosis, and that the actin cytoskeleton coevolved with the vesicle trafficking system to support its functional demands (The figure was adapted from Jèkely, 2007).

"topological separation from the plasma membrane"), with the subsequent evolution of vesicular trafficking "to sort proteins and recycle lipids". This allowed the elaboration of a membranous secretory system that could release digestive enzymes and regulate nutrient uptake and therefore created the conditions for viable predatory phagocytosis which in turn led to endosymbiosis of the mitochondrion (Figure 7.1) (Jekely, 2003,

2007). Crucially, according to this model, the coevolution of eukaryotic endomembranes and cytoskeleton is fundamental, because membrane dynamics necessitated cytoskeletal support and associated action of motor proteins (Jekely, 2007). Prokaryotic precursors of cytoskeletal proteins, from which early eukaryotic filament systems could have evolved, notably include the tubulin homologue the actin-like proteins MreB and ParM (Lowe & Amos, 1999; van den Ent *et al*., 2001).

While carrying out research for this thesis, in the attempt to understand the way the actin cytoskeleton regulates and is regulated by the core endocytic proteins involved in CME, two main types of interaction were identified. As summarised in the actin attachment module described in Chapter 3, one type of actin interaction is based on direct binding by the endocytic protein to actin filament. This is true for the HIP1 and HIP1R proteins, that bind actin filaments via its C-terminal I/LWEQ domain (Engqvist-Goldstein *et al*., 2004; Senetar *et al*., 2004), and also for ABP1, which belongs to a protein family that can de-polymerise actin filaments via ADF/cofilin protein domains (Kessels *et al*., 2001; Quintero-Monzon *et al*., 2005). The other type of interaction with the actin cytoskeleton is indirect, i.e. via the mediation of the N-WASP protein. There are as many as five components of the CME-I that interact with the actin cytoskeleton this way - these are amphiphysins/endophilin, FCH proteins, intersectins, SNX9 and tuba - and they all do so via their SH3 domains that bind the proline rich domain (PRD) of N-WASP which in turn activates actin polymerisation (Bu *et al*., 2010; Cestra *et al*., 2005; Hussain *et al*., 2001; Shin *et al*., 2007; Tsujita *et al*., 2006; Yamada *et al*., 2009). Considering that CME is at least as old as the LCEA, the question is whether these modes of interaction and regulation of the actin cytoskeleton, as identified in

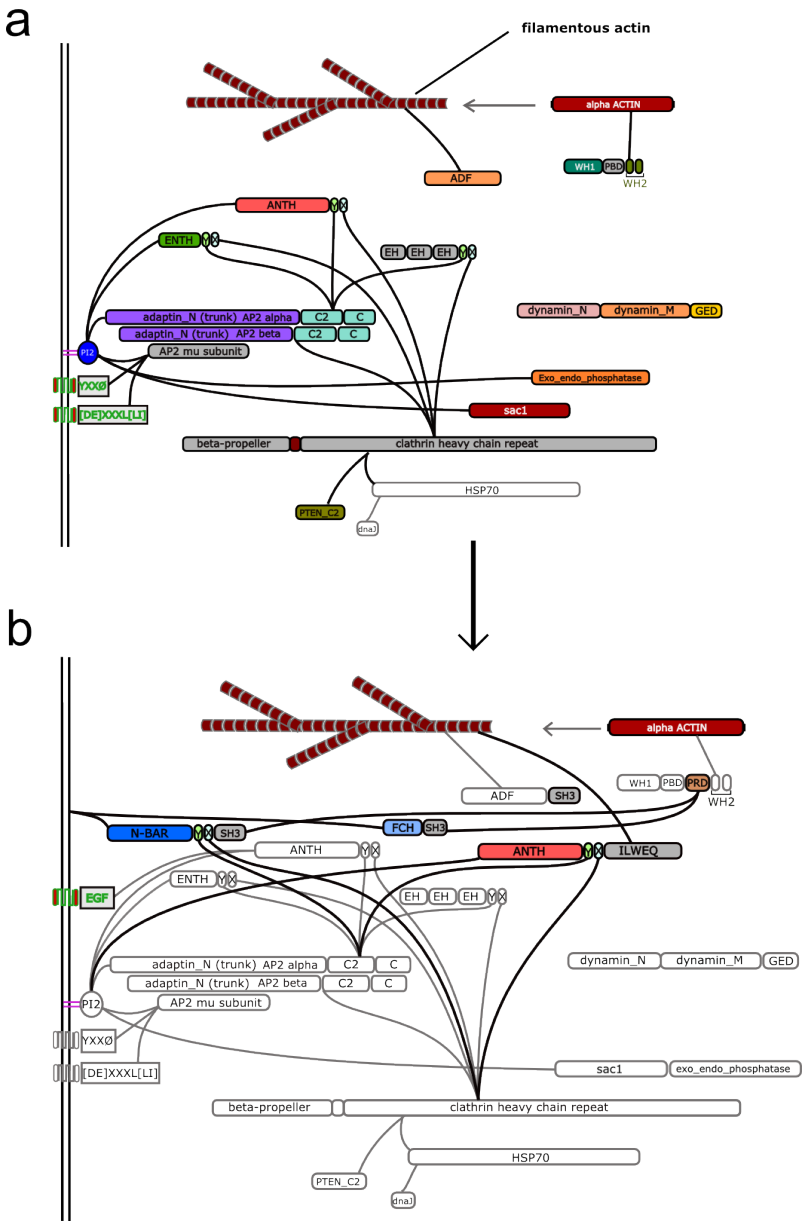


Figure 7.2 Evolution of actin modulation system in the last common unikont ancestor. a The likely CME-I system in the last common eukaryotic ancestor was viable but the model offers no information on how actin cytoskeleton function integrated with the endocytic machinery. **b** The known system of endocytic machinery-actin cytoskeleton modulation evolved in the last common unikont ancestor. Note direct binding to actin filaments by I/LWEQ domain and induction of actin nucleation via SH3-PRD mediated nteraction with N-WASP. Coloured bars represent novel protein domains, black lines represent novel interactions.

opisthokont models, are ancient too. In short, do they represent an ancient eukaryotic system of cytoskeletal support of endomembrane activity such as the ones theorised by Jekely (2007)? The data unfortunately indicates that this is not the case. The direct binding of actin by HIP1/HIP1R protein family can be traced back to the last common unikont ancestor (LCUA), because although the N-terminal ANTH domain of HIP1 and HIP1R was acquired in the LCEA, the C-terminal actin binding I/LWEQ domain was only found in opisthokonts and amoebozoans (i.e. unikonts). Likewise, the data from Chapter 4 shows that none of the SH3-PRD interactions which indirectly induce actin polymerisation are older than the LCUA, and in tuba's and SNX9's case they are more recent still, likely tracing to the last common holozoan ancestor. The protein domain architecture of ABP1 (ADF-SH3), probably originated in the LCUA (in any case certainly not in the LCEA), but the ADF/cofilin protein domain is cenancestral to eukaryotes. However the model presented in this thesis does not propose a mode of how it may have interacted with the CME machinery in the LCEA (Figure 7.2).

These results lead to further considerations. Since the LCEA did perform CME and possessed an actin cytoskeleton, but the former was not regulated by the latter in the ways described, there must have been different modes of regulation. It should be noted that a study of the origin and evolution of phagocytosis also suggested high variability of actin regulated phagocytic mechanisms in diverse taxa, but found that the mechanisms of actin polymerisation, based on regulation by Arp2/3 and WASp/WAVE proteins (the superfamily N-WASP belongs to) are conserved (Yutin *et al*., 2009). So despite differences in CME regulation in the LCEA, actin probably polymerised in the same way.

It is also significant that a large part of the actin attachment module of the CME-I originated in the LCUA (Figure 7.2). Firstly, it increases the number of recognised unikont-specific synapomorphies, further strengthening the support for unikont monophyly, which in large scale phylogenies is moderate to strong (Burki *et al*., 2007; Hampl *et al*., 2009; Minge *et al*., 2009). Secondly, the introduction of the SH3-PRD interaction may represent a significant evolution of how actin cytoskeleton and endocytosis (and by extension vesicular trafficking), mutually regulated, as is also corroborated by the unique presence of myosins with C-terminal SH3 domains in amoebozoans, metazoans and fungi (Richards & Cavalier-Smith, 2005). The novel system must have proved successful because in more derived branches such as the opisthokonts and holozoans, several more SH3 domains were acquired and combined with ancient protein domains. For instance intersectins acquired four SH3 domains and tuba even six (Salazar *et al*., 2003; Yamabhai *et al*., 1998). It would be interesting to further investigate the role of SH3 domains in actin cytoskeleton regulation within the endomembrane system. This could be done by identifying and characterising more proteins with SH3 domains and study their function and evolution.

In any case, this thesis has little bearing on the issue of the first or even the last common eukaryotic ancestor. Indeed, several studies already strongly indicated that LCEA performed CME (Allen *et al*., 2003; Dacks *et al*., 2008; Field *et al*., 2006; Gabernet-Castello *et al*., 2009), so there is no significant novelty other than an extension of the CME model in LCEA. Rather, it constitutes a detailed and comprehensive evolutionary history of CME, thanks in great part to its holistic approach, i.e. the decision to

investigate the evolution of the CME interactome as a whole. The only previous data of note is the comparative genomic study of the CME-I components by Schmid & McMahon (Nature, 2007), where they indicated presence or absence of CME proteins, but only in four metazoan taxa and the apicomplexan protist *Plasmodium falciparum.* Clearly, that taxon sampling cannot be deemed to account for the diversity of the eukaryotes, and the results are not considered within a model of eukaryote evolution. However, their study did not focus on the evolution of the interactome, but on the modular nature of this type of network, and its significance in terms of biological systems (Schmid & McMahon, 2007). By carrying out comparative genomics studies with proper eukaryotic wide taxon sampling, by investigating the phylogenetic distribution of the protein domain architectures that characterise CME proteins, and by constructing a model of CME-I evolution based on the resulting patterns, this thesis adds the evolutionary dimension to the mammalian-biased understanding of the interactome. Interestingly, it confirms that the 'hubs' of the network, i.e. clathrin and the AP2 complex, which are engaged in a disproportionally high number of interactions within the CME-I, are the most highly conserved proteins studied (and in AP2's case, the only one to have specialised before eukaryote diversification), suggesting that other CME proteins coevolved 'around' the hubs. This may seem an intuitive conclusion, but the data presented in this thesis provides actual empirical evidence to support it.

.

In addition, by investigating both the phylogeny of CME genes, and the taxonomic distribution of protein domain architecture, an interesting pattern emerged. The LCEA possessed most of the protein domains involved in CME, which is consistent with the validated view of the LCEA as remarkably complex. However, the phylogenetic trees

suggest that the mammalian CME-I is highly derived, as true orthologues of the query proteins could only be found in *Danio rerio*, and rarely in the rest of the animals. Presence of inparalogues in all major eukaryotic lineages was only found for clathrin and AP2, whereas for all other protein families that had widespread conservation, only distantly related paralogues were found for a significant section of eukaryotic diversity. The model depicting expansion of CME-I complexity, also suggested that several protein families arose by rearrangements of ancient protein domains with newer ones. Considering these apparent contrasting data, it seems rational to suggest that while the LCEA was complex, and performed a fully developed CME pathway, further complexity, specificity and diversity of the pathway evolved chiefly by multiple gene duplications and by reshuffling and rearrangements of protein domain architectures. In essence this may provide an example of how a system that is already relatively complex, becomes more complex still.

This in part relates to the model of endomembrane evolution, supported by a growing number of molecular evolution studies, which proposes that organellar complexity and specialisation was driven by "iterations of gene duplication and coevolution of organelle identity-/specificity-encoding machinery" (Dacks & Field, 2007). These protein families include for instance coatomers/adaptins proteins, syntaxins and other SNARE proteins sub-families and small GTPases (Cavalier-Smith, 2002; Dacks & Doolittle, 2002; Koumandou *et al*., 2007; Yoshizawa *et al*., 2006). Vesicle coat proteins in particular demonstrate this model effectively. All coatomers and adaptin proteins are believed to be homologous and share an ancient ancestor. Shared mechanisms of vesicle formation, the assembly into heterotetrameric complexes, primary sequence similarity and a similar

pattern of pre-LCEA gene duplications all support a common origin (Bonifacino & Glick, 2004; Duden *et al*., 1991; Schledzewski *et al*., 1999). Additional evidence comes from the elucidation of the tertiary structure of vesicle coat proteins, which indicates that clathrin/AP and the two coatomer complexes are composed of α-solenoid and/or β-propeller domains (Devos *et al*., 2004). Strikingly, the same tertiary structure based on α-solenoid and β-propeller domains was found in the nuclear pore complex core scaffold, which coats the curved surface of the nuclear envelope membrane, suggesting a common origin for nuclear pore complexes and coatomer/adaptin proteins (Alber *et al*., 2007; Devos *et al*., 2004). The model thus proposes that the full known diversity of coatomers, AP complexes and nuclear pore complex, which localise to different parts of the endomembrane system, ranging from the plasma membrane to endosomes, trans-Golgi network, the ER and even the nucleus, was produced by ancient, pre-LCEA gene duplications of a hypothetical protocoatomer protein (Field & Dacks, 2009). In fact, in this thesis, no protein family with the same pattern of pre-LCEA duplications was identified, other of course than AP2. But the respective gene families of dynamins, synaptojanins, N-WASPS, ABP1, SNX9 and EPS15 all underwent gene duplications which must have occurred during early eukaryotic diversification. This would explain the diversity of these gene families, and suggests that they contributed to the evolution of diversification and specificity of the eukaryotic endomembrane system, after the early endomembrane system based on coatomers/adaptin proteins and ESCRTs was already in place.

Regarding the evolution of protein domain architectures, while it has been discussed in the context of eukaryote evolution (Basu *et al.*, 2009), few studies have concerned

endomembrane evolution (Bar et al., 2008; Elias, 2010), and a systematic analysis is wanting. In this thesis, at least seven such cases of protein domain rearrangements have been identified, and they range from the mere acquisition of an SH3 domain at the C-terminal of ADF/cofilin proteins, to the intersectins which remarkably contain N-terminal ESP15 homology domains, as many as five SH3 centrally located SH3 domains, and C-terminal guanine nucleotide exchange factor for GTPases followed by PH and C2 domains. Importantly, the model of CME-I evolution presented in Chapter 4, highlights how the acquisition and rearrangements of these protein domains can significantly increase the flexibility and complexity of the interactome. Therefore, if we posit that a biological system is the result of coevolution of its proteome and that it functions via a network of interactions, it is consequently less evolutionarily probable for a novel protein domain to be acquired and 'fit in' within the existing network, than for an existing molecular signature to duplicate and evolve diverging functions while still maintaining the potential to operate within the network, and for multiple protein domains to arrange in novel combination, thus conferring increased functional potential and flexibility. This model is clearly speculative and based on only one well studied cellular process. To validate it, several more studies of different (thought possibly related) cellular processes would have to be carried out and show consistent results. The holistic interactome-based approach used in this thesis would prove useful for this type of further investigation, because it is systematic and allows for the consideration of the fundamental patterns discussed here.

Another important aspect highlighted by the evolutionary study of CME is the problem of the strong mammalian (and to a lesser degree budding yeast) bias in existing cell

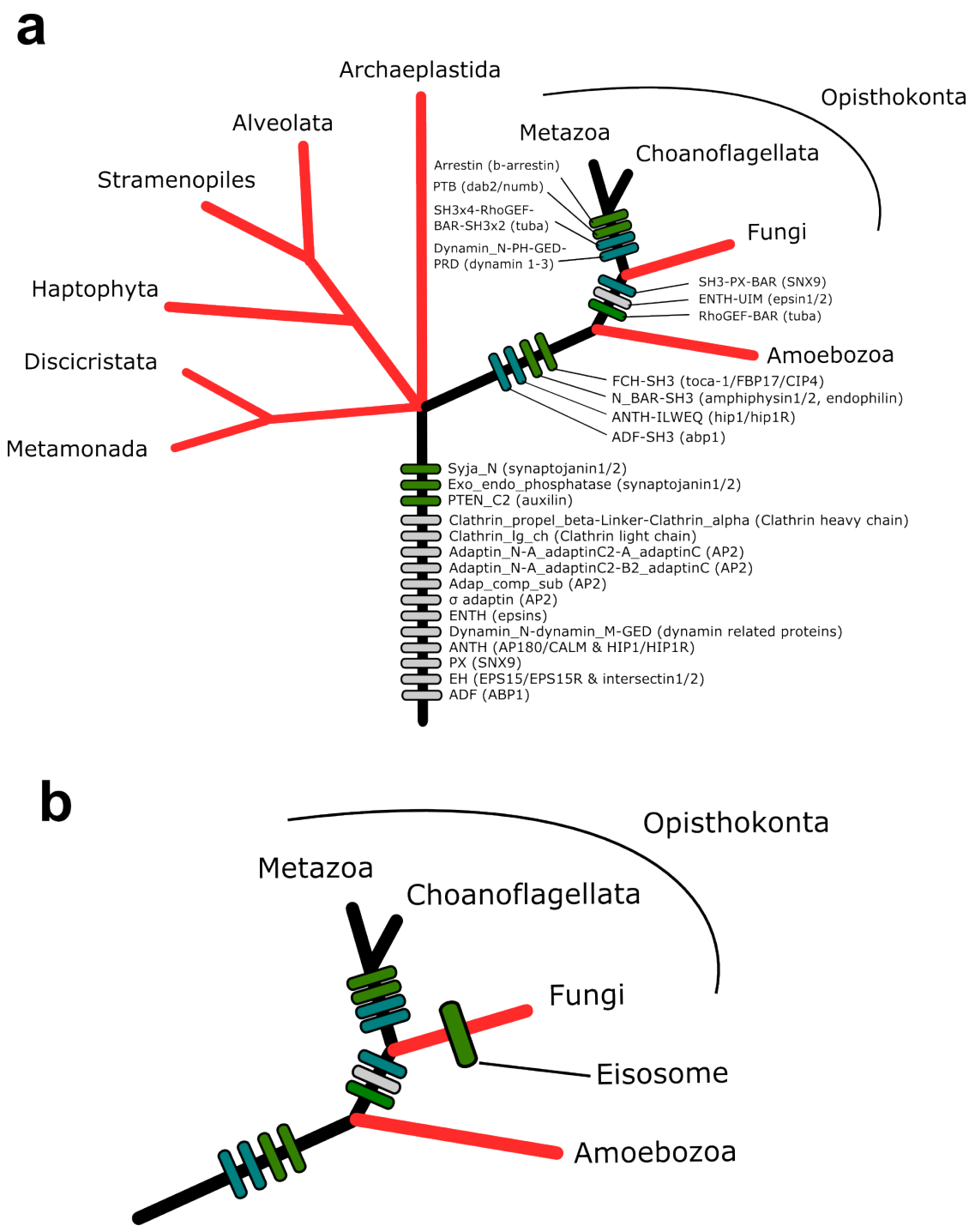


Figure 7.3 Asymmetry in the knowledge of CME diversification across eukarotes. a Black branches represent lineages for which CME diversification data is presented in Chapter 4 of this thesis. Red branches represent eukaryotic groups for which lineage-specific CME diversification is likely to have occurred but cannot be identified with comparative genomics. **b** Results from Chapters 5 and 6 have identified the eisosome as a fungal specific diversification of plasma membrane modulation and endocytic function. Green bars are synapomorphies reported in this thesis, Grey bars are previously reported synapomorphies. Blue bars are novel protein domain rearrangements.

biological data. These are by far the most commonly used model organisms to investigate the nature of the eukaryotic cell. The result of this bias is that the data used to inform evolutionary studies usually comes from a small section of the diversity of eukaryotes (one in six eukaryote supergroups). The phylogenomic study of Chapter 4 shows that CME is significantly derived, and important changes occurred in the unikont, the opisthokont and the holozoan branches. Two problems arise. Firstly, we cannot know anything about how CME diversified in those lineages that do not comprise the taxa where the cell biological data comes from. For instance we can only know about what characteristics the amoebozoans share with opisthokonts, but we cannot know about amoebozoan-specific diversification of CME. Another example is the Fungi. We know that in Holozoa, a series of synapomorphies were acquired that signified a change in the way CME is performed, but what about the fungal/nuclearid sister branch? This issue is even more dramatically illustrated when one considers that from this study we do not know anything about how CME diversified in Archaeplastida, heterokonts, alveolates, haptophytes, Rhizaria and excavates. That could be as many as 4 out of the 6 eukaryote supergroups (Figure 7.3a). The second problem is that we may overlook ancient protein families that were present in the LCEA but were lost in the opisthokont branch. Considering how common gene loss is in eukaryotic evolution this is expected to be the case. Therefore, this asymmetry in taxonomic source of cell biological information, not only limits knowledge of the diversity of eukaryotic cells, but also our understanding of the eukaryote cenancestral cell.

Ironically, an example of lineage-specific endocytic diversity is provided from within this thesis. Initially there was excitement over the discovery of eisosomes, as they may

have proved to be a universal endocytic anchor and key to understand fundamental mechanisms behind spatial distribution of endocytosis (Swaminathan, 2006; Walther *et al*., 2006). However, the data presented in this thesis indicates that the eisosome is a fungal-specific innovation (Figure 7.3b). Phylogenetic and cell biological research presented in this thesis showed that even within different sub-phyla of the ascomycete, the eisosome seem to have diversified in function following multiple gene duplications and a gene loss. Hitherto there has been no attempt to study eisosomes in basidiomycetes, zygomycetes or chytridiomycetes, but given the presence of lineage specific duplications and the high sequence diversity of the homologues found in those taxa there is every reason to suspect eisosomes are functionally diverse. An interesting question is also how the eisosomes interacts with the CME machinery if at all. Indeed, the CME-I is relatively well conserved in Fungi, and there is no question that CME is an important endocytic pathway, so it is rational to predict that eisosomes somehow are integrated within the CME-I system. Strikingly, recent publications suggest that eisosomes contain a domain that is distantly related to the BAR domains of amphiphysins and endophilin, and that they can tubulate liposomes in mammalian cells (Olivera-Couto *et al*., 2011; Ziolkowska *et al*., 2011). It was thus proposed that the membrane bending quality of eisosomes could mediate the formation of lipid clustering in plasma membrane domains (Olivera-Couto *et al*., 2011). These latest studies also prove the importance of structural studies, and the use of more sophisticated homology recognition system to identify distantly related homologues. In this thesis, protein domains were predicted by the established method of running a sequence through libraries of hidden Markov model profiles and position-specific scoring matrices, but that method did not suffice to detect the alleged BAR domains of Pil1 and Lsp1. Despite

this shortcoming, the functional study of MoPil1 and MoPil2 provides a credible link between eisosome function and septation in filamentous fungi, althought this link would have to be supported by studies in non-ascomycete filamentous fungi to become robust. It also demonstrates the value of combining bioinformatics with cell biology to improve our understanding of the cell. Indeed, the effective way to overcome the asymmetry problem discussed in this chapter, would be to use bionformatics analyses, such as the ones reported in Chapter 4, as a platform and a guiding map for functional, cell biological studies in diverse cell types and diverse eukaryotic organisms.

Appendix 1: Laboratory products and software suppliers used in this thesis

Company name	Address
Agilent Technologies UK Ltd.	Agilent Technologies UK Ltd 5 Lochside Avenue Edinburgh Park Edinburgh EH12 9DJ United Kingdom
Applied biosytems/Ambion	2130 Woodward St. Austin, TX 78744-1832 USA
Beckman Coulter Genomics	Beckman Coulter Genomics Hope End, Takeley Essex CM22 6TA United Kingdom
Becton Dickinson	1 Becton Drive Franklin Lakes, NJ USA 07417
Calbiochem-Merck KGaA	Frankfurter Str. 250 64293 Darmstadt Germany
Clontech-Takara	2 Avenue du President Kennedy 78100 Saint-Germain-en-Laye France
FUJIFILM UK Limited	Unit 10A St Martins Business Centre St Martins Way Bedfordshire MK42 0LF, UK
GenoCodes Corporation	Gene Codes Corporation 775 Technology Drive Suite 100A Ann Arbor, MI 48108 USA

Invitrogen Ltd	3 Fountain Drive Inchinnan Business Park Paisley UK PA4 9RF, UK
Nalgene	Outdoor Products Division 75 Panorama Creek Drive Rochester, NY 14625 U.S.A
Promega	Delta House Southampton Science Park Southampton SO16 7NS • United Kingdom
Sigma-aldrich	Sigma-Aldrich Company Ltd. The Old Brickyard New Road Gillingham Dorset SP8 4XT
Waring, Christison Scientific	Albany Road Gateshead NE8 3AT, UK
Whatman International	Whatman International Ltd. Springfield Mill, James Whatman Way, Maidstone, Kent, ME14 2LE

Appendix 2 Taxonomic distribution of protein domains architectures in CME-I network proteins across 15 monophyletic eukaryote groups. Black filled circles denote presence of protein domain architecture, whereas white circle denotes absence. Naming convention adopted by PFAM is used to name protein domain architectures

	Opisthokonts			Amoebozoa	Archaeplastida		Alveolates		Heterokontophyta		Haptophyta	Discicristata		Metamonada	
	Metazoa	Choanoflagellate	Fungi		Viridiplantae	Rhodophyta	Ciliates	Apicomplexa	Oomycetes	Diatoms		Kinetoplastida	Heterolobosea	Diplomonadida	Trichomonadida
Clathrin_propel_beta-Linker-Clathrin_alpha	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Clathrin_lg_ch	●	●	●	●	●	O	●	●	●	●	●	●	●	O	●
Adaptin_N-A_adaptinC2-A_adaptinC	●	●	●	●	●	O	●	●	●	●	●	O	●	●	●
Adaptin_N-Alpha_adaptinC2-B2_adaptinC	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Adapt_comp_sub	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Clat_adaptor_s	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Arrestin_N-Arrestin_C	●	●	O	O	O	O	O	O	O	O	O	O	O	O	O
PTB	●	●	O	O	O	O	O	O	O	O	O	O	O	O	O
ANTH	●	●	●	●	●	●	O	O	O	O	●	●	●	O	O
ANTH-ILWEQ	●	●	●	●	O	O	O	O	O	O	O	O	O	O	O
ENTH	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
ENTH-UIM	●	●	●	O	O	O	O	O	O	O	O	O	O	O	O
BAR-SH3	●	●	●	●	O	O	O	O	O	O	O	O	O	O	O
RhoGEF-BAR	●	●	●	O	O	O	O	O	O	O	O	O	O	O	O
SH3(X4)-RhoGEF-BAR-SH3(X2)	●	●	O	O	O	O	O	O	O	O	O	O	O	O	O
FCH-SH3	●	●	●	●	O	O	O	O	O	O	O	O	O	O	O
PX	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
SH3-PX-BAR	●	●	●	O	O	O	O	O	O	O	O	O	O	O	O
Dynamin_N-Dynamin_M-GED	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Dynamin_N-Dynamin_M-PH-GED-PRD	●	●	O	O	O	O	O	O	O	O	O	O	O	O	O
EH	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
ADF	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
ADF-SH3	●	●	●	●	O	O	O	O	O	O	O	O	O	O	O
WH1-PBD	●	●	●	●	O	O	●	O	O	O	O	O	●	O	●
Syja_N	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Exo_endo_phos	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Syja_N-Exo_endo_phos	●	●	●	●	O	O	●	●	O	O	●	O	O	O	O
PTEN_C2	●	●	●	●	●	O	●	O	●	●	●	O	●	O	●
PTEN_C2-DNAJ	●	O	O	O	O	O	O	O	O	O	O	O	O	O	O

Appendix 3 Putative origins of protein-protein and protein-lipid interaction in the CME-I network. Colour codes in matrix denotes origins on different branches of the eukaryote radiation, where green represents the last common eukaryotic ancestor, orange represents the last common unikont ancestor, pink represents the last common opisthokont ancestor and blue represents last common holozoan ancestor

Appendix 4: Pil1 and Lsp1 phylogeny with *C.owczarzaki* NUL00001676

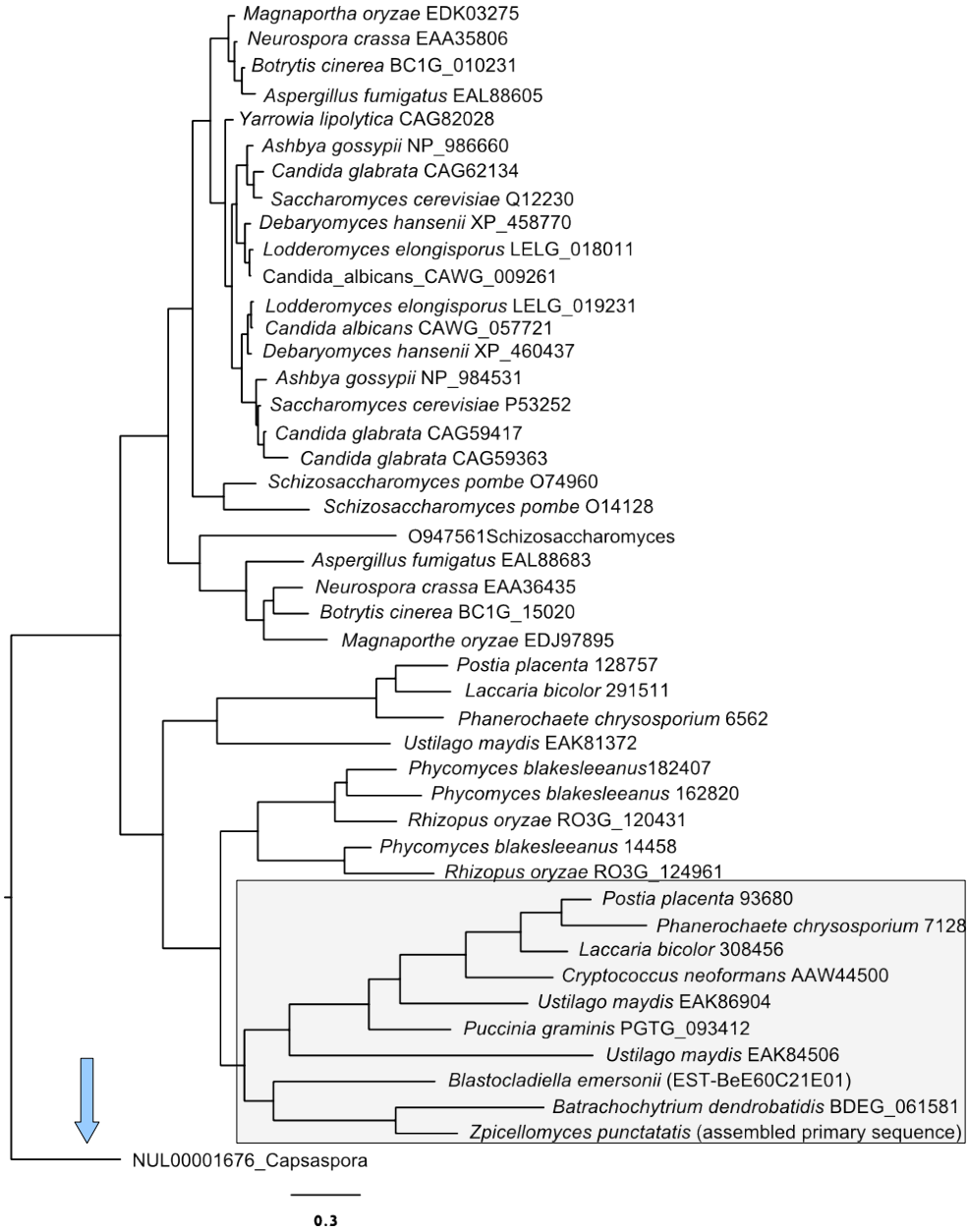


Figure 1 Phylogeny of Pil1 and Lsp1 with long branch sequence *C. owczarzaki*. NUL00001676. Tree was calculated with MRBAYES, arrow indicates the long branch sequence. The box highlights the clade comprising 'chtrids' and basidiomycetes.

Appendix 5: Heat shock assay of transformed *S. cerevisiae* pil1Δ mutants

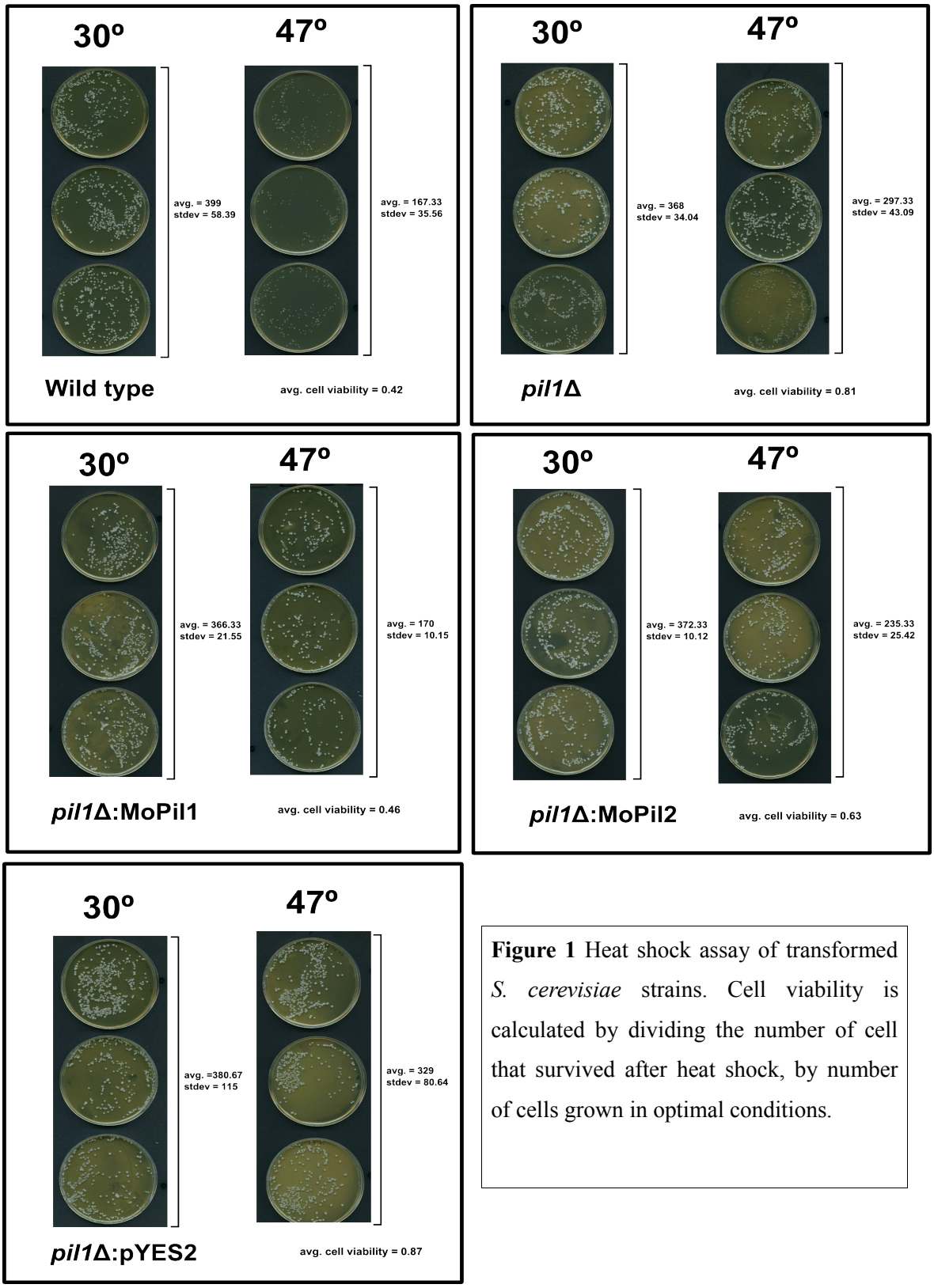


Figure 1 Heat shock assay of transformed *S. cerevisiae* strains. Cell viability is calculated by dividing the number of cell that survived after heat shock, by number of cells grown in optimal conditions.

Bibliography

Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics, 21*(9), 2104-2105.

Abremski, K., Hoess, R., & Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell, 32*(4), 1301-1311.

Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., *et al.* (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol, 52*(5), 399-451.

Ahmed, S., Bu, W., Lee, R. T., Maurer-Stroh, S., & Goh, W. I. (2010). F-BAR domain proteins: Families and function. *Commun Integr Biol, 3*(2), 116-121.

Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., *et al*. (2007). The molecular architecture of the nuclear pore complex. *Nature, 450*(7170), 695-701.

Alberts B, J. A., Lewis J, et. al. (2002). *Molecular Biology of the Cell. 4th edition. Chapter 13*. New York: Garland Science.

Allen, L. A., & Aderem, A. (1996). Mechanisms of phagocytosis. *Curr Opin Immunol, 8*(1), 36-40.

Allen, R. D., & Fok, A. K. (1980). Membrane recycling and endocytosis in *Paramecium* confirmed by horseradish peroxidase pulse-chase studies. *J Cell Sci, 45*, 131-145.

Allen, C. L., Goulding, D., & Field, M. C. (2003). Clathrin-mediated endocytosis is

essential in Trypanosoma brucei. *Embo J, 22*(19), 4991-5002.

Allen, R. D., & Wolf, R. W. (1979). Membrane recycling at the cytoproct of *Tetrahymena*. *J Cell Sci, 35*, 217-227.

Alpuche-Aranda, C. M., Racoosin, E. L., Swanson, J. A., & Miller, S. I. (1994). Salmonella stimulate macrophage macropinocytosis and persist within spacious phagosomes. *J Exp Med, 179*(2), 601-608.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.

Alvarez, C. E. (2008b). On the origins of arrestin and rhodopsin. *BMC Evol Biol, 8*, 222.

Alvarez, F. J., Douglas, L. M., Rosebrock, A., & Konopka, J. B. (2008). The Sur7 protein regulates plasma membrane organization and prevents intracellular cell wall growth in *Candida albicans*. *Mol Biol Cell, 19*(12), 5214-5225.

Amoah-Buahin, E., Bone, N., & Armstrong, J. (2005). Hyphal Growth in the Fission Yeast *Schizosaccharomyces pombe*. *Eukaryot Cell, 4*(7), 1287-1297.

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol, 55*(4), 539-552.

Archibald, J. M. (2009). The puzzle of plastid evolution. *Curr Biol, 19*(2), R81-88.

Archibald, J. M., Longet, D., Pawlowski, J., & Keeling, P. J. (2003). A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol Biol Evol, 20*(1), 62-66.

Archibald, J. M., & Richards, T. A. (2010). Gene transfer: anything goes in plant mitochondria. *BMC Biol, 8*, 147.

Ayala, J. (1994). Transport and internal organization of membranes: vesicles, membrane networks and GTP-binding proteins. *J Cell Sci, 107 (Pt 4)*, 753-763.

Bachvaroff, T. R., Sanchez Puerta, M. V., & Delwiche, C. F. (2005). Chlorophyll c–Containing Plastid Relationships Based on Analyses of a Multigene Data Set with All Four Chromalveolate Lineages. *Molecular Biology and Evolution, 22*(9), 1772-1782.

Bai, T., Seebald, J. L., Kim, K. E., Ding, H. M., Szeto, D. P., & Chang, H. C. (2010). Disruption of zebrafish cyclin G-associated kinase (GAK) function impairs the expression of Notch-dependent genes during neurogenesis and causes defects in neuronal development. *BMC Dev Biol, 10*, 7.

Baldauf, S. L., & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A, 90*(24), 11558-11562.

Baldauf, S. L. (2008). An overview of the phylogeny and diversity of eukaryotes. *Journal Of Systematics And Evolution, 46*(3), 263-273.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science, 290*(5493), 972-977.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., *et al.* (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc Natl Acad Sci U S A, 99*(3), 1414-1419.

Bar, M., Aharon, M., Benjamin, S., Rotblat, B., Horowitz, M., & Avni, A. (2008). AtEHDs, novel Arabidopsis EH-domain-containing proteins involved in endocytosis. *Plant J, 55*(6), 1025-1038.

Bar-Sagi, D., & Feramisco, J. R. (1986). Induction of membrane ruffling and fluid-phase pinocytosis in quiescent fibroblasts by ras proteins. *Science, 233*(4768), 1061-1068.

Barth, G., & Gaillardin, C. (1997). Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol Rev, 19*(4), 219-237.

Bartnicki-Garcia, S. (1987). The cell wall: a crucial structure in fungal evolution. In A. D. M. Rayner, C. M. Brasier & D. Moore (Eds.), *Evolutionary Biology of the Fungi* (pp. 389-403). Cambridge: Cambridge University Press.

Bass, D., Moreira, D., Lopez-Garcia, P., Polet, S., Chao, E. E., von der Heyden, S., *et al.* (2005). Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist, 156*(2), 149-161.

Basu, M. K., Poliakov, E., & Rogozin, I. B. (2009). Domain mobility in proteins: functional and evolutionary implications. *Brief Bioinform, 10*(3), 205-216.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., & Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res, 28*(1), 263-266.

Baudin-Baillieu, A., Guillemet, E., Cullin, C., & Lacroute, F. (1997). Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast, 13*(4), 353-356.

Baurain, D., Brinkmann, H., Petersen, J., Rodriguez-Ezpeleta, N., Stechmann, A., Demoulin, V., *et al.* (2010). Phylogenomic evidence for separate acquisition of

plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol, 27*(7), 1698-1709.

Bhattacharya, D., & Medlin, L. (1995). The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *Journal of Phycology, 31*(4), 489-498.

Bennett, B., Old, L. J., & Boyse, E. A. (1963). Opsonization of cells by isoantibody in vitro. *Nature, 198*, 10-12.

Berney, C., Fahrni, J., & Pawlowski, J. (2004). How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol, 2*, 13.

Bernstein, B. W., & Bamburg, J. R. (2010). ADF/cofilin: a functional node in cell biology. *Trends Cell Biol, 20*(4), 187-195.

Blondeau, F., Ritter, B., Allaire, P. D., Wasiak, S., Girard, M., Hussain, N. K., *et al.* (2004). Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc Natl Acad Sci U S A, 101*(11), 3833-3838.

Bonifacino, J. S., & Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. *Cell, 116*(2), 153-166.

Brodin, L., Low, P., & Shupliakov, O. (2000). Sequential steps in clathrin-mediated synaptic vesicle endocytosis. *Curr Opin Neurobiol, 10*(3), 312-320.

Brodsky, F. M., Chen, C. Y., Knuehl, C., Towler, M. C., & Wakeham, D. E. (2001). Biological basket weaving: formation and function of clathrin-coated vesicles. *Annu Rev Cell Dev Biol, 17*, 517-568.

Brown, M. W., Spiegel, F. W., & Silberman, J. D. (2009). Phylogeny of the "forgotten" cellular slime mold, Fonticula alba, reveals a key evolutionary branch within

Opisthokonta. *Mol Biol Evol, 26*(12), 2699-2709.

Bu, W., Chou, A. M., Lim, K. B., Sudhaharan, T., & Ahmed, S. (2009). The Toca-1-N-WASP complex links filopodial formation to endocytosis. *J Biol Chem, 284*(17), 11622-11636.

Burki, F., & Pawlowski, J. (2006). Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. *Mol Biol Evol, 23*(10), 1922-1930.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S. I., Jakobsen, K. S., *et al.* (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One, 2*(8), e790.

Burki, F., Inagaki, Y., Brate, J., Archibald, J. M., Keeling, P. J., Cavalier-Smith, T., *et al*. (2009). Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol, 1*, 231-238.

Caplan, S., Naslavsky, N., Hartnell, L. M., Lodge, R., Polishchuk, R. S., Donaldson, J. G., *et al.* (2002). A tubular EHD1-containing compartment involved in the recycling of major histocompatibility complex class I molecules to the plasma membrane. *Embo J, 21*(11), 2557-2567.

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., *et al.* (1998). Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc Natl Acad Sci U S A, 95*(2), 520-524.

Carr, M., Leadbeater, B. S., Hassan, R., Nelson, M., & Baldauf, S. L. (2008). Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci U S A, 105*(43), 16641-16646.

Casal, E., Federici, L., Zhang, W., Fernandez-Recio, J., Priego, E. M., Miguel, R. N., *et al.* (2006). The crystal structure of the BAR domain from human Bin1/amphiphysin II and its implications for molecular recognition. *Biochemistry, 45*(43), 12917-12928.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol, 17*(4), 540-552.

Cavalier-Smith, T. (1975). The origin of nuclei and of eukaryotic cells. *Nature, 256*(5517), 463-468.

Cavalier-Smith, T. (1983a). A 6 kingdom classification and a unified phylogeny. In W. Schwemmler & H. E. A. Schenk (Eds.), *Endocytobiology II* (pp. 265–279). Berlin: De Gruyter.

Cavalier-Smith, T. (1983b). Endosymbiotic origin of the mitochondrial envelope. *Endocytobiology II*, 265–279.

Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol, 46*(4), 347-366.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol, 52*(Pt 2), 297-354.

Cavalier-Smith, T. (2006). Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc Lond B Biol Sci, 361*(1470), 969-1006.

Cavalier-Smith, T. (2009). Predation and eukaryote cell origins: a coevolutionary perspective. *Int J Biochem Cell Biol, 41*(2), 307-322.

Cavalier-Smith, T. (2010a). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett, 6*(3), 342-345.

Cavalier-Smith, T. (2010b). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biol Direct, 5*, 7.

Cavalier-Smith, T., & Chao, E. E. (1995). The Opalozoan Apusomonas is Related to the Common Ancestor of Animals, Fungi, and Choanoflagellates. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 261*(1360), 1-6.

Cavalier-Smith, T., & Chao, E. E. (2003). Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol, 56*(5), 540-563.

Cestra, G., Kwiatkowski, A., Salazar, M., Gertler, F., & De Camilli, P. (2005). Tuba, a GEF for CDC42, links dynamin to actin regulatory proteins. *Methods Enzymol, 404*, 537-545.

Chambers, J. A., & Thompson, J. E. (1976). Phagocytosis and pinocytosis in *Acanthamoeba castellanii*. *J Gen Microbiol, 92*(2), 246-250.

Chapman-Andresen, C. (1984). The early days of pinocytosis. *Carlsberg Research Communications, 49*(2), 179-186.

Chappell, T. G., Welch, W. J., Schlossman, D. M., Palter, K. B., Schlesinger, M. J., & Rothman, J. E. (1986). Uncoating ATPase is a member of the 70 kilodalton family of stress proteins. *Cell, 45*(1), 3-13.

Chanez, A. L., Hehl, A. B., Engstler, M., & Schneider, A. (2006). Ablation of the single dynamin of T. brucei blocks mitochondrial fission and endocytosis and leads to a precise cytokinesis arrest. *J Cell Sci, 119*(Pt 14), 2968-2974.

Chen, C. Y., Reese, M. L., Hwang, P. K., Ota, N., Agard, D., & Brodsky, F. M. (2002). Clathrin light and heavy chain interface: alpha-helix binding superhelix loops via critical tryptophans. *Embo J, 21*(22), 6072-6082.

Chen, H., Fre, S., Slepnev, V. I., Capua, M. R., Takei, K., Butler, M. H., *et al.* (1998).

Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature, 394*(6695), 793-797.

Chen, H. R., Hsu, M. T., & Cheng, S. C. (1995). Spheroplast preparation facilitates PCR screening of yeast sequence. *Biotechniques, 19*(5), 744-746, 748.

Chen, G., Zhuchenko, O., & Kuspa, A. (2007). Immune-like phagocyte activity in the social amoeba. *Science, 317*(5838), 678-681.

Chimini, G., & Chavrier, P. (2000). Function of Rho family proteins in actin dynamics during phagocytosis and engulfment. *Nat Cell Biol, 2*(10), E191-196.

Co, C., Wong, D. T., Gierke, S., Chang, V., & Taunton, J. (2007). Mechanism of Actin Network Attachment to Moving Membranes: Barbed End Capture by N-WASP WH2 Domains. *Cell, 128*(5), 901.

Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res, 36*(Web Server issue), W197-201.

Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R., & Owen, D. J. (2002). Molecular architecture and functional model of the endocytic AP2 complex. *Cell, 109*(4), 523-535.

Colwill, K., Field, D., Moore, L., Friesen, J., & Andrews, B. (1999). In vivo analysis of the domains of yeast Rvs167p suggests Rvs167p function is mediated through multiple protein interactions. *Genetics, 152*(3), 881-893.

Conner, S. D., & Schmid, S. L. (2003). Regulated portals of entry into the cell. *Nature, 422*(6927), 37.

Correa, J. R., Atella, G. C., Menna-Barreto, R. S., & Soares, M. J. (2007). Clathrin in Trypanosoma cruzi: in silico gene identification, isolation, and localization of protein expression sites. *J Eukaryot Microbiol, 54*(3), 297-302.

Covian-Nares, J. F., Smith, R. M., & Vogel, S. S. (2008). Two independent forms of endocytosis maintain embryonic cell surface homeostasis during early development. *Dev Biol, 316*(1), 135-148.

Cremona, O., Di Paolo, G., Wenk, M. R., L¸thi, A., Kim, W. T., Takei, K., *et al.* (1999). Essential Role of Phosphoinositide Metabolism in Synaptic Vesicle Recycling. *Cell, 99*(2), 179.

Cuff, J. A., & Barton, G. J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins, 40*(3), 502-511.

Cullen, P. J. (2008). Endosomal sorting and signalling: an emerging role for sorting nexins. *Nat Rev Mol Cell Biol, 9*(7), 574-582.

Cunha-e-Silva, N., Sant'Anna, C., Pereira, M. G., & de Souza, W. (2010). Endocytosis in *Trypansoma cruzi*. *The Open Parasitology Journal, 4*, 98-101.

Dacks, J. B., & Doolittle, W. F. (2002). Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: implications for the ancient evolution of the eukaryotic endomembrane system. *J Cell Sci, 115*(Pt 8), 1635-1642.

Dacks, J. B., & Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci, 120*(Pt 17), 2977-2985.

Dacks, J. B., Marinets, A., Ford Doolittle, W., Cavalier-Smith, T., & Logsdon, J. M., Jr. (2002). Analyses of RNA Polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol Biol Evol, 19*(6), 830-840.

Dacks, J. B., Peden, A. A., & Field, M. C. (2009). Evolution of specificity in the eukaryotic endomembrane system. *Int J Biochem Cell Biol, 41*(2), 330-340.

Dacks, J. B., Poon, P. P., & Field, M. C. (2008). Phylogeny of endocytic components

yields insight into the process of nonendosymbiotic organelle evolution. *Proc Natl Acad Sci U S A, 105*(2), 588-593.

Dacks, J. B., Silberman, J. D., Simpson, A. G., Moriya, S., Kudo, T., Ohkuma, M., *et al.* (2001). Oxymonads are closely related to the excavate taxon Trimastix. *Mol Biol Evol, 18*(6), 1034-1044.

Daeron, M. (1997). Fc receptor biology. *Annu Rev Immunol, 15*, 203-234.

David, C., McPherson, P. S., Mundigl, O., & de Camilli, P. (1996). A role of amphiphysin in synaptic vesicle endocytosis suggested by its binding to dynamin in nerve terminals. *Proc Natl Acad Sci U S A, 93*(1), 331-335.

Dawson, J. C., Legg, J. A., & Machesky, L. M. (2006). Bar domain proteins: a role in tubulation, scission and actin assembly in clathrin-mediated endocytosis. *Trends In Cell Biology, 16*(10), 493-498.

Dean, R. A. (1997). Signal pathways and appressorium morphogenesis. *Annu Rev Phytopathol, 35*, 211-234.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., *et al.* (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature, 434*(7036), 980-986.

DeGrasse, J. A., DuBois, K. N., Devos, D., Siegel, T. N., Sali, A., Field, M. C., *et al*. (2009). Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics, 8*(9), 2119-2130.

Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol, 3*(10), e314.

Dell'Angelica, E. C. (2001). Clathrin-binding proteins: got a motif? Join the network!

Trends Cell Biol, 11(8), 315-318.

Desjardins, M., Huber, L. A., Parton, R. G., & Griffiths, G. (1994). Biogenesis of phagolysosomes proceeds through a sequential series of interactions with the endocytic apparatus. *J Cell Biol, 124*(5), 677-688.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., *et al*. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol, 2*(12), e380.

Dho, S. E., Trejo, J., Siderovski, D. P., & McGlade, C. J. (2006). Dynamic regulation of mammalian numb by G protein-coupled receptors and protein kinase C activation: Structural determinants of numb association with the cortical membrane. *Mol Biol Cell, 17*(9), 4142-4155.

Dieckmann, R., von Heyden, Y., Kistler, C., Gopaldass, N., Hausherr, S., Crawley, S. W., *et al*. (2010). A myosin IK-Abp1-PakB circuit acts as a switch to regulate phagocytosis efficiency. *Mol Biol Cell, 21*(9), 1505-1518.

Di Fiore, P. P., & De Camilli, P. (2001). Endocytosis and signaling. an inseparable partnership. *Cell, 106*(1), 1-4.

Douglas, S. E., & Turner, S. (1991). Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *J Mol Evol, 33*(3), 267-273.

Drab, M., Verkade, P., Elger, M., Kasper, M., Lohn, M., Lauterbach, B., *et al.* (2001). Loss of caveolae, vascular dysfunction, and pulmonary defects in caveolin-1 gene-disrupted mice. *Science, 293*(5539), 2449-2452.

Drake, M. T., Downs, M. A., & Traub, L. M. (2000). Epsin binds to clathrin by associating directly with the clathrin-terminal domain. Evidence for cooperative binding through two discrete sites. *J Biol Chem, 275*(9), 6479-6489.

Drubin, D. G., Kaksonen, M., Toret, C., & Sun, Y. (2005). Cytoskeletal networks and pathways involved in endocytosis. *Novartis Found Symp, 269*, 35-42; discussion 43-36, 223-230.

Drummond, A., & Strimmer, K. (2001). PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics, 17*(7), 662-663.

Duden, R., Griffiths, G., Frank, R., Argos, P., & Kreis, T. E. (1991). Beta-COP, a 110 kd protein associated with non-clathrin-coated vesicles and the Golgi complex, shows homology to beta-adaptin. *Cell, 64*(3), 649-665.

Duwel, M., & Ungewickell, E. J. (2006). Clathrin-dependent association of CVAK104 with endosomes and the trans-Golgi network. *Mol Biol Cell, 17*(10), 4513-4525.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics, 14*(9), 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*(5), 1792-1797.

Elias, M. (2010). Patterns and processes in the evolution of the eukaryotic endomembrane system. *Mol Membr Biol, 27*(8), 469-489.

Eisenberg, E., & Greene, L. E. (2007). Multiple roles of auxilin and hsc70 in clathrin-mediated endocytosis. *Traffic, 8*(6), 640-646.

Elde, N. C., Morgan, G., Winey, M., Sperling, L., & Turkewitz, A. P. (2005). Elucidation of clathrin-mediated endocytosis in tetrahymena reveals an evolutionarily convergent recruitment of dynamin. *PLoS Genet, 1*(5), e52.

Embley, T. M. (2006). Multiple secondary origins of the anaerobic lifestyle in eukaryotes. *Philos Trans R Soc Lond B Biol Sci, 361*(1470), 1055-1067.

Embley, T. M., & Hirt, R. P. (1998). Early branching eukaryotes? *Curr Opin Genet Dev, 8*(6), 624-629.

Embley, T. M., van der Giezen, M., Horner, D. S., Dyal, P. L., & Foster, P. (2003). Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Philos Trans R Soc Lond B Biol Sci, 358*(1429), 191-201; discussion 201-192.

Eme, L., Moreira, D., Talla, E., & Brochier-Armanet, C. (2009). A complex cell division machinery was present in the last common ancestor of eukaryotes. *PLoS One, 4*(4), e5021.

Engqvist-Goldstein, A. E., Warren, R. A., Kessels, M. M., Keen, J. H., Heuser, J., & Drubin, D. G. (2001). The actin-binding protein Hip1R associates with clathrin during early stages of endocytosis and promotes clathrin assembly in vitro. *J Cell Biol, 154*(6), 1209-1223.

Engqvist-Goldstein, A. E., Zhang, C. X., Carreno, S., Barroso, C., Heuser, J. E., & Drubin, D. G. (2004). RNAi-mediated Hip1R silencing results in stable association between the endocytic machinery and the actin assembly machinery. *Mol Biol Cell, 15*(4), 1666-1679.

Engqvist-Goldstein, A. E. Y., Kessels, M. M., Chopra, V. S., Hayden, M. R., & Drubin, D. G. (1999). An actin-binding protein of the Sla2/Huntingtin interacting protein 1 family is a novel component of clathrin-coated pits and vesicles. *Journal Of Cell Biology, 147*(7), 1503-1518.

Ezekowitz, R. A., Williams, D. J., Koziel, H., Armstrong, M. Y., Warner, A., Richards, F. F., *et al*. (1991). Uptake of Pneumocystis carinii mediated by the macrophage mannose receptor. *Nature, 351*(6322), 155-158.

Fadok, V. A., Bratton, D. L., Konowal, A., Freed, P. W., Westcott, J. Y., & Henson, P. M. (1998). Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-beta, PGE2, and PAF. *J Clin Invest, 101*(4), 890-898.

Fadok, V. A., Bratton, D. L., Rose, D. M., Pearson, A., Ezekewitz, R. A., & Henson, P. M. (2000). A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature, 405*(6782), 85-90.

Fadok, V. A., & Henson, P. M. (1998). Apoptosis: getting rid of the bodies. *Curr Biol, 8*(19), R693-695.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol, 17*(6), 368-376.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution, 39*(4), 783-791.

Field, M. C., & Carrington, M. (2009). The trypanosome flagellar pocket. *Nat Rev Microbiol, 7*(11), 775-786.

Field, M. C., & Dacks, J. B. (2009). First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr Opin Cell Biol, 21*(1), 4-13.

Field, M. C., Gabernet-Castello, C., & Dacks, J. B. (2006). Reconstructing the Evolution of the Endocytic System: Insights from Genomics and Molecular Cell Biology. *Origins and Evolution of Eukaryotic Endomembranes and Cytoskeleton*.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., *et al.* (2010). The Pfam protein families database. *Nucleic Acids Res, 38*(Database issue), D211-

222.

Field, M. C., Sali, A., & Rout, M. P. (2011). Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat. *J Cell Biol, 193*(6), 963-972.

Fitch, W. M., & Margoliash, E. (1967). A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochem Genet, 1*(1), 65-71.

Ford, M. G. J., Mills, I. G., Peter, B. J., Vallis, Y., Praefcke, G. J. K., Evans, P. R., *et al.* (2002). Curvature of clathrin-coated pits driven by epsin. *Nature, 419*(6905), 361-366.

Ford, M. G. J., Pearse, B. M. F., Higgins, M. K., Vallis, Y., Owen, D. J., Gibson, A., *et al.* (2001). Simultaneous binding of PtdIns(4,5)P-2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science, 291*(5506), 1051-1055.

Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T., *et al.* (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature, 432*(7017), 573-579.

Frick, M., Bright, N. A., Riento, K., Bray, A., Merrified, C., & Nichols, B. J. (2007). Coassembly of flotillins induces formation of membrane microdomains, membrane curvature, and vesicle budding. *Curr Biol, 17*(13), 1151-1156.

Frise, E., Knoblich, J. A., Younger-Shepherd, S., Jan, L. Y., & Jan, Y. N. (1996). The *Drosophila* Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage. *Proc Natl Acad Sci U S A, 93*(21), 11925-11932.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., *et al.* (2010). The genome of *Naegleria gruberi* illuminates early

eukaryotic versatility. *Cell, 140*(5), 631-642.

Frohlich, F., Moreira, K., Aguilar, P. S., Hubner, N. C., Mann, M., Walter, P., *et al.* (2009). A genome-wide screen for genes affecting eisosomes reveals Nce102 function in sphingolipid signaling. *J Cell Biol, 185*(7), 1227-1242.

Fuchs, U., Hause, G., Schuchardt, I., & Steinberg, G. (2006). Endocytosis is essential for pathogenic development in the corn smut fungus *Ustilago maydis*. *Plant Cell, 18*(8), 2066-2081.

Futterer, K., & Machesky, L. M. (2007). "Wunder" F-BAR domains: going from pits to vesicles. *Cell, 129*(4), 655-657.

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

Gabernet-Castello, C., Dacks, J. B., & Field, M. C. (2009). The single ENTH-domain protein of trypanosomes; endocytic functions and evolutionary relationship with epsin. *Traffic, 10*(7), 894-911.

Gauthier, N. C., Rossier, O. M., Mathur, A., Hone, J. C., & Sheetz, M. P. (2009). Plasma membrane area increases with spread area by exocytosis of a GPI-anchored protein compartment. *Mol Biol Cell, 20*(14), 3261-3272.

Gaidarov, I., & Keen, J. H. (1999). Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits. *Journal Of Cell Biology, 146*(4), 755-764.

Gaidarov, I., Krupnick, J. G., Falck, J. R., Benovic, J. L., & Keen, J. H. (1999). Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *Embo J, 18*(4), 871-881.

Gallop, J. L., Jao, C. C., Kent, H. M., Butler, P. J. G., Evans, P. R., Langen, R., *et al.*

(2006). Mechanism of endophilin N-BAR domain-mediated membrane curvature. *Embo Journal, 25*(12), 2898-2910.

Galtier, N., Gouy, M., & Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci, 12*(6), 543-548.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol, 14*(7), 685-695.

Gietz, R. D., & Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol, 350*, 87-96.

Gimeno, C. J., Ljungdahl, P. O., Styles, C. A., & Fink, G. R. (1992). Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. *Cell, 68*(6), 1077-1090.

Glebov, O. O., Bright, N. A., & Nichols, B. J. (2006). Flotillin-1 defines a clathrin-independent endocytic pathway in mammalian cells. *Nat Cell Biol, 8*(1), 46-54.

Gold, E. S., Underhill, D. M., Morrissette, N. S., Guo, J., McNiven, M. A., & Aderem, A. (1999). Dynamin 2 is required for phagocytosis in macrophages. *J Exp Med, 190*(12), 1849-1856.

Gonda, K., Komatsu, M., & Numata, O. (2000). Calmodulin and Ca2+/calmodulin-binding proteins are involved in *Tetrahymena thermophila* phagocytosis. *Cell Struct Funct, 25*(4), 243-251.

Goode, B. L., Rodal, A. A., Barnes, G., & Drubin, D. G. (2001). Activation of the Arp2/3 complex by the actin filament binding protein Abp1p. *J Cell Biol, 153*(3), 627-634.

Goodman, O. B., Jr., Krupnick, J. G., Santini, F., Gurevich, V. V., Penn, R. B., Gagnon, A. W., *et al.* (1996). Beta-arrestin acts as a clathrin adaptor in endocytosis of the beta2-adrenergic receptor. *Nature, 383*(6599), 447-450.

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annu Rev Plant Biol, 59*, 491-517.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol, 27*(2), 221-224.

Grabs, D., Slepnev, V. I., Songyang, Z., David, C., Lynch, M., Cantley, L. C., *et al.* (1997). The SH3 domain of amphiphysin binds the proline-rich domain of dynamin at a single site that defines a new SH3 binding consensus sequence. *J Biol Chem, 272*(20), 13419-13425.

Grande-Garcia, A., Echarri, A., de Rooij, J., Alderson, N. B., Waterman-Storer, C. M., Valdivielso, J. M., *et al*. (2007). Caveolin-1 regulates cell polarization and directional migration through Src kinase and Rho GTPases. *J Cell Biol, 177*(4), 683-694.

Granseth, B., Odermatt, B., Royle, S. J., & Lagnado, L. (2006). Clathrin-mediated endocytosis is the dominant mechanism of vesicle retrieval at hippocampal synapses. *Neuron, 51*(6), 773-786.

Grant, B. D., & Caplan, S. (2008). Mechanisms of EHD/RME-1 protein function in endocytic transport. *Traffic, 9*(12), 2043-2052.

Greenberg, S., el Khoury, J., di Virgilio, F., Kaplan, E. M., & Silverstein, S. C. (1991). Ca(2+)-independent F-actin assembly and disassembly during Fc receptor-mediated phagocytosis in mouse macrophages. *J Cell Biol, 113*(4), 757-767.

Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P., & Brochier-Armanet, C. (2010). The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol, 8*(10), 743-752.

Grossmann, G., Malinsky, J., Stahlschmidt, W., Loibl, M., Weig-Meckl, I., Frommer, W. B., *et al.* (2008). Plasma membrane microdomains regulate turnover of transport proteins in yeast. *J Cell Biol*.

Gu, X., Fu, Y. X., & Li, W. H. (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol Biol Evol, 12*(4), 546-557.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol, 52*(5), 696-704.

Guo, S., Stolz, L. E., Lemrow, S. M., & York, J. D. (1999). SAC1-like domains of yeast SAC1, INP52, and INP53 and of human synaptojanin encode polyphosphoinositide phosphatases. *J Biol Chem, 274*(19), 12990-12995.

Habermann, B. (2004). The BAR-domain family of proteins: a case of bending and binding? *EMBO Rep, 5*(3), 250-255.

Hacker, U., Albrecht, R., & Maniak, M. (1997). Fluid-phase uptake by macropinocytosis in *Dictyostelium*. *J Cell Sci, 110 (Pt 2)*, 105-112.

Hackett, J. D., Yoon, H. S., Li, S., Reyes-Prieto, A., Rummele, S. E., & Bhattacharya, D. (2007). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol Biol Evol, 24*(8), 1702-1713.

Harper, J. T., Waanders, E., & Keeling, P. J. (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol*

Microbiol, 55(Pt 1), 487-496.

Hamer, J. E., Howard, R. J., Chumley, F. G., & Valent, B. (1988). A mechanism for surface attachment in spores of a plant pathogenic fungus. *Science, 239*(4837), 288-290.

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G., *et al.* (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci U S A, 106*(10), 3859-3864.

Hansen, C. G., & Nichols, B. J. (2009). Molecular mechanisms of clathrin-independent endocytosis. *J Cell Sci, 122*(Pt 11), 1713-1721.

Hartig, S. M., Ishikura, S., Hicklen, R. S., Feng, Y., Blanchard, E. G., Voelker, K. A., *et al.* (2009). The F-BAR protein CIP4 promotes GLUT4 endocytosis through bidirectional interactions with N-WASp and Dynamin-2. *J Cell Sci, 122*(Pt 13), 2283-2291.

Hartman, H., & Fedorov, A. (2002). The origin of the eukaryotic cell: a genomic investigation. *Proc Natl Acad Sci U S A, 99*(3), 1420-1425.

Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A, 89*(22), 10915-10919.

Hernandez, Y., Castillo, C., Roychowdhury, S., Hehl, A., Aley, S. B., & Das, S. (2007). Clathrin-dependent pathways and the cytoskeleton network are involved in ceramide endocytosis by a parasitic protozoan, *Giardia lamblia*. *Int J Parasitol, 37*(1), 21-32.

Hewlett, L. J., Prescott, A. R., & Watts, C. (1994). The coated pit and macropinocytic pathways serve distinct endosome populations. *J Cell Biol, 124*(5), 689-703.

Heymann, J. B., Iwasaki, K., Yim, Y. I., Cheng, N., Belnap, D. M., Greene, L. E., *et al.* (2005). Visualization of the binding of Hsc70 ATPase to clathrin baskets: implications for an uncoating mechanism. *J Biol Chem, 280*(8), 7156-7161.

Higgins, M. K., & McMahon, H. T. (2002). Snap-shots of clathrin-mediated endocytosis. *Trends Biochem Sci, 27*(5), 257-263.

Higuchi, Y., Shoji, J. Y., Arioka, M., & Kitamoto, K. (2009). Endocytosis is crucial for cell polarity and apical membrane recycling in the filamentous fungus *Aspergillus oryzae*. *Eukaryot Cell, 8*(1), 37-46.

Hinkle, G., Leipe, D. D., Nerad, T. A., & Sogin, M. L. (1994). The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. *Nucleic Acids Res, 22*(3), 465-469.

Hinrichsen, L., Harborth, J., Andrees, L., Weber, K., & Ungewickell, E. J. (2003). Effect of clathrin heavy chain- and alpha-adaptin-specific small inhibitory RNAs on endocytic accessory proteins and receptor trafficking in HeLa cells. *J Biol Chem, 278*(46), 45160-45170.

Hinshaw, J. E. (2000). Dynamin and its role in membrane fission. *Annu Rev Cell Dev Biol, 16*, 483-519.

Hinshaw, J. E., & Schmid, S. L. (1995). Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature, 374*(6518), 190-192.

Hirst, J., Barlow, L. D., Francisco, G. C., Sahlender, D. A., Seaman, M. N., Dacks, J. B., *et al*. (2011). The fifth adaptor protein complex. *PLoS Biol, 9*(10), e1001170.

Ho, H. Y., Rohatgi, R., Lebensohn, A. M., Le, M., Li, J., Gygi, S. P., *et al.* (2004). Toca-1 mediates Cdc42-dependent actin nucleation by activating the N-WASP-WIP complex. *Cell, 118*(2), 203-216.

Ho, S. Y., & Jermiin, L. (2004). Tracing the decay of the historical signal in biological sequence data. *Syst Biol, 53*(4), 623-637.

Hodges, M. E., Scheumann, N., Wickstead, B., Langdale, J. A., & Gull, K. (2010). Reconstructing the evolutionary history of the centriole from protein components. *J Cell Sci, 123*(Pt 9), 1407-1413.

Holstein, S. E., Ungewickell, H., & Ungewickell, E. (1996). Mechanism of clathrin basket dissociation: separate functions of protein domains of the DnaJ homologue auxilin. *J Cell Biol, 135*(4), 925-937.

Horvath, C. A., Vanden Broeck, D., Boulet, G. A., Bogers, J., & De Wolf, M. J. (2007). Epsin: inducing membrane curvature. *Int J Biochem Cell Biol, 39*(10), 1765-1770.

Howard, R. J., & Valent, B. (1996). Breaking and entering: host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. *Annu Rev Microbiol, 50*, 491-512.

Howes, M. T., Mayor, S., & Parton, R. G. (2010). Molecules, mechanisms, and cellular roles of clathrin-independent endocytosis. *Curr Opin Cell Biol, 22*(4), 519-527.

Huang, F., Khvorova, A., Marshall, W., & Sorkin, A. (2004). Analysis of clathrin-mediated endocytosis of epidermal growth factor receptor by RNA interference. *J Biol Chem, 279*(16), 16657-16661.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science, 294*(5550), 2310-2314.

Hung, C. H., Qiao, X., Lee, P. T., & Lee, M. G. (2004). Clathrin-dependent targeting of receptors to the flagellar pocket of procyclic-form Trypanosoma brucei.

Eukaryot Cell, 3(4), 1004-1014.

Hussain, N. K., Jenna, S., Glogauer, M., Quinn, C. C., Wasiak, S., Guipponi, M., *et al.* (2001). Endocytic protein intersectin-l regulates actin assembly via Cdc42 and N-WASP. *Nat Cell Biol, 3*(10), 927-932.

Iannolo, G., Salcini, A. E., Gaidarov, I., Goodman, O. B., Jr., Baulida, J., Carpenter, G., *et al.* (1997). Mapping of the molecular determinants involved in the interaction between eps15 and AP-2. *Cancer Res, 57*(2), 240-245.

Idone, V., Tam, C., Goss, J. W., Toomre, D., Pypaert, M., & Andrews, N. W. (2008). Repair of injured plasma membrane by rapid Ca2+-dependent endocytosis. *J Cell Biol, 180*(5), 905-914.

Irani, N. G., & Russinova, E. (2009). Receptor endocytosis and signaling in plants. *Curr Opin Plant Biol, 12*(6), 653-659.

Itoh, T., Erdmann, K. S., Roux, A., Habermann, B., Werner, H., & De Camilli, P. (2005). Dynamin and the actin cytoskeleton cooperatively regulate plasma membrane invagination by BAR and F-BAR proteins. *Developmental Cell, 9*(6), 791-804.

Itoh, T., Koshiba, S., Kigawa, T., Kikuchi, A., Yokoyama, S., & Takenawa, T. (2001). Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science, 291*(5506), 1047-1051.

Jacobs, M. E., DeSouza, L. V., Samaranayake, H., Pearlman, R. E., Siu, K. W., & Klobutcher, L. A. (2006). The *Tetrahymena thermophila* phagosome proteome. *Eukaryot Cell, 5*(12), 1990-2000.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., *et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature, 443*(7113), 818-822.

Jarvis, P., & Soll, J. (2002). Toc, tic, and chloroplast protein import. *Biochim Biophys Acta, 1590*(1-3), 177-189.

Jekely, G. (2003). Small GTPases and the evolution of the eukaryotic cell. *Bioessays, 25*(11), 1129-1138.

Jekely, G. (2007). Origin of Eukaryotic Endomembranes: A Critical Evaluatoin of Different Model Scenarios. In G. Jekely (Ed.), *Eukaryotic Membranes and Cytoskeleton: Origin and Evolution*: Landes Bioscience & Springer Science + Business Media.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol, 292*(2), 195-202.

Kamble, C., Jain, S., Murphy, E., & Kim, K. (2011). Requirements of Slm proteins for proper eisosome organization, endocytic trafficking and recycling in the yeast *Saccharomyces cerevisiae*. *J Biosci, 36*(1), 79-96.

Kamioka, Y., Fukuhara, S., Sawa, H., Nagashima, K., Masuda, M., Matsuda, M., *et al.* (2004). A novel dynamin-associating molecule, formin-binding protein 17, induces tubular membrane invaginations and participates in endocytosis. *J Biol Chem, 279*(38), 40091-40099.

Kasahara, M. (2007). The 2R hypothesis: an update. *Curr Opin Immunol, 19*(5), 547-552.

Kazazic, M., Bertelsen, V., Pedersen, K. W., Vuong, T. T., Grandal, M. V., Rodland, M. S., *et al.* (2009). Epsin 1 is involved in recruitment of ubiquitinated EGF receptors into clathrin-coated pits. *Traffic, 10*(2), 235-245.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., & McLnerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on

empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol, 6*, 29.

Keeling, P. J. (2001). Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol Biol Evol, 18*(8), 1551-1557.

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature, 428*(6983), 617-624.

Kessels, M. M., Engqvist-Goldstein, A. E., Drubin, D. G., & Qualmann, B. (2001). Mammalian Abp1, a signal-responsive F-actin-binding protein, links the actin cytoskeleton to endocytosis via the GTPase dynamin. *J Cell Biol, 153*(2), 351-366.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution. *Cambridge University Press*.

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., *et al.* (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature, 451*(7180), 783-788.

Kirkham, M., Fujita, A., Chadda, R., Nixon, S. J., Kurzchalia, T. V., Sharma, D. K., *et al.* (2005). Ultrastructural identification of uncoated caveolin-independent early endocytic vehicles. *J Cell Biol, 168*(3), 465-476.

Kirkham, M., Nixon, S. J., Howes, M. T., Abi-Rached, L., Wakeham, D. E., Hanzal-Bayer, M., *et al*. (2008). Evolutionary analysis and molecular dissection of caveola biogenesis. *J Cell Sci, 121*(Pt 12), 2075-2086.

Kitajima, Y., & Thompson, G. A., Jr. (1977). Differentiation of food vacuolar membranes during endocytosis in Tetrahymena. *J Cell Biol, 75*(2 Pt 1), 436-445.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet, 39*, 309-338.

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct, 1*, 22.

Koonin, E. V. (2009). Intron-dominated genomes of early ancestors of eukaryotes. *J Hered, 100*(5), 618-623.

Koonin, E. V., & Aravind, L. (2009). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle, 8*(13), 1984-1985.

Koumandou, V. L., Dacks, J. B., Coulson, R. M., & Field, M. C. (2007). Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol Biol, 7*, 29.

Krendel, M., Osterweil, E. K., & Mooseker, M. S. (2007). Myosin 1E interacts with synaptojanin-1 and dynamin and is involved in endocytosis. *FEBS Lett, 581*(4), 644-650.

Krupnick, J. G., Goodman, O. B., Jr., Keen, J. H., & Benovic, J. L. (1997). Arrestin/clathrin interaction. Localization of the clathrin binding domain of nonvisual arrestins to the carboxy terminus. *J Biol Chem, 272*(23), 15011-15016.

Kumari, S., Mg, S., & Mayor, S. (2010). Endocytosis unplugged: multiple ways to enter the cell. *Cell Res, 20*(3), 256-275.

Lafer, E. M. (2002). Clathrin-protein interactions. *Traffic, 3*(8), 513-520.

Lakhan, S. E., Sabharanjak, S., & De, A. (2009). Endocytosis of glycosylphosphatidylinositol-anchored proteins. *J Biomed Sci, 16*, 93.

Lamaze, C., Fujimoto, L. M., Yin, H. L., & Schmid, S. L. (1997). The actin

cytoskeleton is required for receptor-mediated endocytosis in mammalian cells. *J Biol Chem, 272*(33), 20332-20335.

Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature, 467*(7318), 929-934.

Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W., & Burger, G. (2002). The closest unicellular relatives of animals. *Curr Biol, 12*(20), 1773-1778.

Laporte, S. A., Oakley, R. H., Zhang, J., Holt, J. A., Ferguson, S. S. G., Caron, M. G., *et al.* (1999). The beta(2)-adrenergic receptor/beta arrestin complex recruits the clathrin adaptor AP-2 during endocytosis. *Proceedings Of The National Academy Of Sciences Of The United States Of America, 96*(7), 3712-3717.

Lappalainen, P., Kessels, M. M., Cope, M. J., & Drubin, D. G. (1998). The ADF homology (ADF-H) domain: a highly exploited actin-binding module. *Mol Biol Cell, 9*(8), 1951-1959.

Larget, B., & Simon, D. L. (1999). Markov Chasin Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. *Molecular Biology and Evolution, 16*(6), 750.

Lau, A. W., & Chou, M. M. (2008). The adaptor complex AP-2 regulates post-endocytic trafficking through the non-clathrin Arf6-dependent endocytic pathway. *J Cell Sci, 121*(Pt 24), 4008-4017.

Lauber, K., Bohn, E., Krober, S. M., Xiao, Y. J., Blumenthal, S. G., Lindemann, R. K., *et al.* (2003). Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal. *Cell, 113*(6), 717-730.

Le Quesne, W. J. (1974). The Uniquely Evolved Character Concept and its Cladistic Application. *Systematic Biology, 23*(4), 513-517.

Lee, J. O., Yang, H., Georgescu, M. M., Di Cristofano, A., Maehama, T., Shi, Y., *et al.* (1999). Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell, 99*(3), 323-334.

Legendre-Guillemin, V., Wasiak, S., Hussain, N. K., Angers, A., & McPherson, P. S. (2004). ENTH/ANTH proteins and clathrin-mediated membrane budding. *J Cell Sci, 117*(Pt 1), 9-18.

Lehtonen, S., Shah, M., Nielsen, R., Iino, N., Ryan, J. J., Zhou, H., *et al*. (2008). The endocytic adaptor protein ARH associates with motor and centrosomal proteins and is involved in centrosome assembly and cytokinesis. *Mol Biol Cell, 19*(7), 2949-2961.

Leipe, D. D., Gunderson, J. H., Nerad, T. A., & Sogin, M. L. (1993). Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol Biochem Parasitol, 59*(1), 41-48.

Lemmon, S. K., & Jones, E. W. (1987). Clathrin requirement for normal growth of yeast. *Science, 238*(4826), 504-509.

Lemmon, S. K. (2001). Clathrin uncoating: Auxilin comes to life. *Curr Biol, 11*(2), R49-52.

Leonard, G., Stevens, J. R., & Richards, T. A. (2009). REFGEN and TREENAMER: automated sequence data handling for phylogenetic analysis in the genomic era. *Evol Bioinform Online, 5*, 1-4.

Leung, H., Borromeo, E. S., Bernardo, M. A., & Notteghem, J. L. (1988). Genetic analysis of virulence in the rice blast fungus *Magnaporthe grisea*. *Phytopathology, 78*, 1227-1233.

Levin, D. E., Bowers, B., Chen, C. Y., Kamada, Y., & Watanabe, M. (1994). Dissecting the protein kinase C/MAP kinase signalling pathway of *Saccharomyces cerevisiae*. *Cell Mol Biol Res, 40*(3), 229-239.

Lewis, W. H. (1931). Pinocytosis. *John Hopkins Hosp. Bull., 49*, 17-27.

Lonhienne, T. G., Sagulenko, E., Webb, R. I., Lee, K. C., Franke, J., Devos, D. P., *et al.* (2010). Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A, 107*(29), 12883-12888.

Lopez-Garcia, P., & Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. *Bioessays, 28*(5), 525-533.

Liu, Y., Richards, T. A., & Aves, S. J. (2009). Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol, 9*, 60.

Lowe, J., & Amos, L. A. (1999). Tubulin-like protofilaments in Ca2+-induced FtsZ sheets. *Embo J, 18*(9), 2364-2371.

Lundmark, R., & Carlsson, S. R. (2003). Sorting nexin 9 participates in clathrin-mediated endocytosis through interactions with the core components. *J Biol Chem, 278*(47), 46772-46781.

Luo, G., Gruhler, A., Liu, Y., Jensen, O. N., & Dickson, R. C. (2008). The sphingolipid long-chain base-Pkh1/2-Ypk1/2 signaling pathway regulates eisosome assembly and turnover. *J Biol Chem, 283*(16), 10433-10444.

Maldonado-Baez, L., Dores, M. R., Perkins, E. M., Drivas, T. G., Hicke, L., & Wendland, B. (2008). Interaction between Epsin/Yap180 Adaptors and the Scaffolds Ede1/Pan1 Is Required for Endocytosis. *Mol Biol Cell, 19*(7), 2936-2948.

Maldonado-Baez, L., & Wendland, B. (2006). Endocytic adaptors: recruiters,

coordinators and regulators. *Trends In Cell Biology, 16*(10), 505-513.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., *et al.* (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res, 37*(Database issue), D205-210.

Margulis, L., Dolan, M. F., & Guerrero, R. (2000). The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc Natl Acad Sci U S A, 97*(13), 6954-6959.

Marks, B., Stowell, M. H., Vallis, Y., Mills, I. G., Gibson, A., Hopkins, C. R., *et al.* (2001). GTPase activity of dynamin and resulting conformation change are essential for endocytosis. *Nature, 410*(6825), 231-235.

Martin, W. (1999). A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 266*(1426), 1387-1395.

Martin, W., Hoffmeister, M., Rotte, C., & Henze, K. (2001). An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem, 382*(11), 1521-1539.

Martin, W., & Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature, 392*(6671), 37-41.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., & Kowallik, K. V. (1998b). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature, 393*(6681), 162-165.

Massol, R. H., Boll, W., Griffin, A. M., & Kirchhausen, T. (2006). A burst of auxilin recruitment determines the onset of clathrin-coated vesicle uncoating. *Proc Natl*

Acad Sci U S A, 103(27), 10265-10270.

Masuda, M., Takeda, S., Sone, M., Ohki, T., Mori, H., Kamioka, Y., *et al.* (2006). Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms. *Embo Journal, 25*(12), 2889-2897.

May, R. C., & Machesky, L. M. (2001). Phagocytosis and the actin cytoskeleton. *J Cell Sci, 114*(Pt 6), 1061-1077.

Mayer, B. J. (2001). SH3 domains: complexity in moderation. *J Cell Sci, 114*(Pt 7), 1253-1263.

McCann, R. O., & Craig, S. W. (1997). The I/LWEQ module: a conserved sequence that signifies F-actin binding in functionally diverse proteins from yeast to mammals. *Proc Natl Acad Sci U S A, 94*(11), 5679-5684.

McGavin, M. K., Badour, K., Hardy, L. A., Kubiseski, T. J., Zhang, J., & Siminovitch, K. A. (2001). The intersectin 2 adaptor links Wiskott Aldrich Syndrome protein (WASp)-mediated actin polymerization to T cell antigen receptor endocytosis. *J Exp Med, 194*(12), 1777-1787.

McMahon, H. T., Wigge, P., & Smith, C. (1997). Clathrin interacts specifically with amphiphysin and is displaced by dynamin. *FEBS Lett, 413*(2), 319-322.

McPherson, P. S., Garcia, E. P., Slepnev, V. I., David, C., Zhang, X., Grabs, D., *et al.* (1996). A presynaptic inositol-5-phosphatase. *Nature, 379*(6563), 353-357.

McPherson, P. S., & Ritter, B. (2005). Peptide motifs: building the clathrin machinery. *Mol Neurobiol, 32*(1), 73-87.

Meagher, L. C., Savill, J. S., Baker, A., Fuller, R. W., & Haslett, C. (1992). Phagocytosis of apoptotic neutrophils does not induce macrophage release of thromboxane B2. *J Leukoc Biol, 52*(3), 269-273.

Metchnikoff, E. (1883). Untersuchungen iaber de mesodermalen Phagocyten einigen Wirbeltiere. *Biol. Zentr., 3*, 560-565.

Meshulam, T., Simard, J. R., Wharton, J., Hamilton, J. A., & Pilch, P. F. (2006). Role of caveolin-1 and cholesterol in transmembrane fatty acid movement. *Biochemistry, 45*(9), 2882-2893.

Michielse, C. B., Salim, K., Ragas, P., Ram, A. F., Kudla, B., Jarry, B., *et al.* (2004). Development of a system for integrative and stable transformation of the zygomycete *Rhizopus oryzae* by Agrobacterium-mediated DNA transfer. *Mol Genet Genomics, 271*(4), 499-510.

Miele, A. E., Watson, P. J., Evans, P. R., Traub, L. M., & Owen, D. J. (2004). Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain [beta]-propeller. *Nat Struct Mol Biol, 11*(3), 242.

Minge, M. A., Silberman, J. D., Orr, R. J., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., *et al.* (2009). Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc Biol Sci, 276*(1657), 597-604.

Minshall, R. D., Tiruppathi, C., Vogel, S. M., Niles, W. D., Gilchrist, A., Hamm, H. E., *et al.* (2000). Endothelial cell-surface gp60 activates vesicle formation and trafficking via G(i)-coupled Src kinase signaling pathway. *J Cell Biol, 150*(5), 1057-1070.

Mishra, S. K., Agostinelli, N. R., Brett, T. J., Mizukami, I., Ross, T. S., & Traub, L. M. (2001). Clathrin- and AP-2-binding sites in HIP1 uncover a general assembly role for endocytic accessory proteins. *J Biol Chem, 276*(49), 46230-46236.

Mishra, S. K., Keyel, P. A., Hawryluk, M. J., Agostinelli, N. R., Watkins, S. C., & Traub, L. M. (2002a). Disabled-2 exhibits the properties of a cargo-selective

endocytic clathrin adaptor. *Embo Journal, 21*(18), 4915-4926.

Mishra, S. K., Watkins, S. C., & Traub, L. M. (2002b). The autosomal recessive hypercholesterolemia (ARH) protein interfaces directly with the clathrin-coat machinery. *Proc Natl Acad Sci U S A, 99*(25), 16099-16104.

Moreira, D., & Lopez-Garcia, P. (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol, 47*(5), 517-530.

Moreira, D., Le Guyader, H., & Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature, 405*(6782), 69-72.

Moreira, D., & Philippe, H. (2001). Sure facts and open questions about the origin and evolution of photosynthetic plastids. *Res Microbiol, 152*(9), 771-780.

Moreira, D., von der Heyden, S., Bass, D., Lopez-Garcia, P., Chao, E., & Cavalier-Smith, T. (2007). Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol, 44*(1), 255-266.

Morgan, G. W., Allen, C. L., Jeffries, T. R., Hollinshead, M., & Field, M. C. (2001). Developmental and morphological regulation of clathrin-mediated endocytosis in *Trypanosoma brucei*. *J Cell Sci, 114*(Pt 14), 2605-2615.

Morgan, G. W., Hall, B. S., Denny, P. W., Carrington, M., & Field, M. C. (2002). The kinetoplastida endocytic apparatus. Part I: a dynamic system for nutrition and evasion of host defences. *Trends Parasitol, 18*(11), 491-496.

Morgan, G. W., Goulding, D., & Field, M. C. (2004). The single dynamin-like protein of Trypanosoma brucei regulates mitochondrial division and is not required for endocytosis. *J Biol Chem, 279*(11), 10692-10701.

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., *et al.* (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science, 317*(5846), 1921-1926.

Motley, A., Bright, N. A., Seaman, M. N., & Robinson, M. S. (2003). Clathrin-mediated endocytosis in AP-2-depleted cells. *J Cell Biol, 162*(5), 909-918.

Mueller, V. J., Wienisch, M., Nehring, R. B., & Klingauf, J. (2004). Monitoring clathrin-mediated endocytosis during synaptic activity. *J Neurosci, 24*(8), 2004-2012.

Murata, M., Peranen, J., Schreiner, R., Wieland, F., Kurzchalia, T. V., & Simons, K. (1995). VIP21/caveolin is a cholesterol-binding protein. *Proc Natl Acad Sci U S A, 92*(22), 10339-10343.

Naslavsky, N., Weigert, R., & Donaldson, J. G. (2003). Convergence of non-clathrin- and clathrin-derived endosomes involves Arf6 inactivation and changes in phosphoinositides. *Mol Biol Cell, 14*(2), 417-431.

Neefjes, J. J., Stollorz, V., Peters, P. J., Geuze, H. J., & Ploegh, H. L. (1990). The biosynthetic pathway of MHC class II but not class I molecules intersects the endocytic route. *Cell, 61*(1), 171-183.

Neumann, N., Lundin, D., & Poole, A. M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS One, 5*(10), e13241.

Neumann-Staubitz, P., Hall, S. L., Kuo, J., & Jackson, A. P. (2010). Characterization of a temperature-sensitive vertebrate clathrin heavy chain mutant as a tool to study clathrin-dependent events in vivo. *PLoS One, 5*(8), e12017.

Newman, S. L., Mikus, L. K., & Tucci, M. A. (1991). Differential requirements for cellular cytoskeleton in human macrophage complement receptor- and Fc

receptor-mediated phagocytosis. *J Immunol, 146*(3), 967-974.

Newmyer, S. L., Christensen, A., & Sever, S. (2003). Auxilin-dynamin interactions link the uncoating ATPase chaperone machinery with vesicle formation. *Dev Cell, 4*(6), 929-940.

Newpher, T. M., Idrissi, F. Z., Geli, M. I., & Lemmon, S. K. (2006). Novel function of clathrin light chain in promoting endocytic vesicle formation. *Mol Biol Cell, 17*(10), 4343-4352.

Nichols, B. A., Chiappino, M. L., & Pavesio, C. E. (1994). Endocytosis at the micropore of Toxoplasma gondii. *Parasitol Res, 80*(2), 91-98.

Nikolaev, S. I., Berney, C., Fahrni, J. F., Bolivar, I., Polet, S., Mylnikov, A. P., *et al.* (2004). The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc Natl Acad Sci U S A, 101*(21), 8066-8071.

Nikolaev, S. I., Berney, C., Petrov, N. B., Mylnikov, A. P., Fahrni, J. F., & Pawlowski, J. (2006). Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. *Int J Syst Evol Microbiol, 56*(Pt 6), 1449-1458.

Nikolaou, E., Agrafioti, I., Stumpf, M., Quinn, J., Stansfield, I., & Brown, A. J. (2009). Phylogenetic diversity of stress signalling pathways in fungi. *BMC Evol Biol, 9*, 44.

O'Brien, E. A., Koski, L. B., Zhang, Y., Yang, L., Wang, E., Gray, M. W., *et al.* (2007). TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res, 35*(Database issue), D445-451.

Olivera-Couto, A., Grana, M., Harispe, L., & Aguilar, P. S. (2011). The eisosome core is composed of BAR domain proteins. *Mol Biol Cell, 22*(13), 2360-2372.

Olsen, G. J., Matsuda, H., Hagstrom, R., & Overbeek, R. (1994). fastDNAmL: a tool for

construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci, 10*(1), 41-48.

Owen, D. J., & Evans, P. R. (1998). A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science, 282*(5392), 1327-1332.

Parfrey, L. W., Barbero, E., Lasser, E., Dunthorn, M., Bhattacharya, D., Patterson, D. J., *et al.* (2006). Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet, 2*(12), e220.

Park, H., & Cox, D. (2009). Cdc42 regulates Fc gamma receptor-mediated phagocytosis through the activation and phosphorylation of Wiskott-Aldrich syndrome protein (WASP) and neural-WASP. *Mol Biol Cell, 20*(21), 4500-4508.

Parton, R. G., & Simons, K. (2007). The multiple faces of caveolae. *Nat Rev Mol Cell Biol, 8*(3), 185-194.

Patron, N. J., Rogers, M. B., & Keeling, P. J. (2004). Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell, 3*(5), 1169-1175.

Patron, N. J., Inagaki, Y., & Keeling, P. J. (2007). Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol, 17*(10), 887-891.

Pawlowski, J., Bolivar, I., Fahrni, J. F., Cavalier-Smith, T., & Gouy, M. (1996). Early origin of foraminifera suggested by SSU rRNA gene sequences. *Mol Biol Evol, 13*(3), 445-450.

Payne, G. S., & Schekman, R. (1985). A test of clathrin function in protein secretion and cell growth. *Science, 230*(4729), 1009-1014.

Pelkmans, L., Burli, T., Zerial, M., & Helenius, A. (2004). Caveolin-stabilized

membrane domains as multifunctional transport and sorting devices in endocytic membrane traffic. *Cell, 118*(6), 767-780.

Pelkmans, L., Puntener, D., & Helenius, A. (2002). Local actin polymerization and dynamin recruitment in SV40-induced internalization of caveolae. *Science, 296*(5567), 535-539.

Penny, D., McComish, B. J., Charleston, M. A., & Hendy, M. D. (2001). Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J Mol Evol, 53*(6), 711-723.

Pereira-Leal, J. B., & Seabra, M. C. (2001). Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol, 313*(4), 889-901.

Perera, R. M., Zoncu, R., Lucast, L., De Camilli, P., & Toomre, D. (2006). Two synaptojanin 1 isoforms are recruited to clathrin-coated pits at different stages. *Proc Natl Acad Sci U S A, 103*(51), 19332-19337.

Peter, B. J., Kent, H. M., Mills, I. G., Vallis, Y., Butler, P. J. G., Evans, P. R., *et al.* (2004). BAR domains as sensors of membrane curvature: The amphiphysin BAR structure. *Science, 303*(5657), 495-499.

Philippe, H., & Adoutte, A. (1998). The molecular phylogeny of Eukaryota: solid facts and uncertainties. In G. Coombs, K. Vickerman, M. Sleigh & A. Warren (Eds.), *Evolutionary relationships among Protozoa* (pp. 25-56). London: Chapman & Hall.

Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., *et al.* (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc Biol Sci, 267*(1449), 1213-1221.

Pinyol, R., Haeckel, A., Ritter, A., Qualmann, B., & Kessels, M. M. (2007). Regulation

of N-WASP and the Arp2/3 complex by Abp1 controls neuronal morphology. *PLoS One, 2*(5), e400.

Plattner, H., & Kissmehl, R. (2003). Molecular aspects of membrane trafficking in paramecium. *Int Rev Cytol, 232*, 185-216.

Pol, A., Luetterforst, R., Lindsay, M., Heino, S., Ikonen, E., & Parton, R. G. (2001). A caveolin dominant negative mutant associates with lipid bodies and induces intracellular cholesterol imbalance. *J Cell Biol, 152*(5), 1057-1070.

Pol, A., Martin, S., Fernandez, M. A., Ferguson, C., Carozzi, A., Luetterforst, R., *et al*. (2004). Dynamic and regulated association of caveolin with lipid bodies: modulation of lipid body motility and function by a dominant negative mutant. *Mol Biol Cell, 15*(1), 99-110.

Pollitt, A. Y., & Insall, R. H. (2009). WASP and SCAR/WAVE proteins: the drivers of actin assembly. *J Cell Sci, 122*(Pt 15), 2575-2578.

Polo, S., & Di Fiore, P. P. (2006). Endocytosis conducts the cell signaling orchestra. *Cell, 124*(5), 897-900.

Porto-Carreiro, I., Attias, M., Miranda, K., De Souza, W., & Cunha-e-Silva, N. (2000). *Trypanosoma cruzi* epimastigote endocytic pathway: cargo enters the cytostome and passes through an early endosomal network before storage in reservosomes. *Eur J Cell Biol, 79*(11), 858-869.

Praefcke, G. J., & McMahon, H. T. (2004). The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat Rev Mol Cell Biol, 5*(2), 133-147.

Pucharcos, C., Estivill, X., & de la Luna, S. (2000). Intersectin 2, a new multimodular protein involved in clathrin-mediated endocytosis. *FEBS Lett, 478*(1-2), 43-51.

Pylypenko, O., Lundmark, R., Rasmuson, E., Carlsson, S. R., & Rak, A. (2007). The PX-BAR membrane-remodeling unit of sorting nexin 9. *Embo J, 26*(22), 4788-4800.

Qualmann, B., Kessels, M. M., & Kelly, R. B. (2000). Molecular links between endocytosis and the actin cytoskeleton. *J Cell Biol, 150*(5), F111-116.

Rabinovitch, M. (1995). Professional and non-professional phagocytes: an introduction. *Trends Cell Biol, 5*(3), 85-87.

Racoosin, E. L., & Swanson, J. A. (1992). M-CSF-induced macropinocytosis increases solute endocytosis but not receptor-mediated endocytosis in mouse macrophages. *J Cell Sci, 102 (Pt 4)*, 867-880.

Racoosin, E. L., & Swanson, J. A. (1993). Macropinosome maturation and fusion with tubular lysosomes in macrophages. *J Cell Biol, 121*(5), 1011-1020.

Radhakrishna, H., & Donaldson, J. G. (1997). ADP-ribosylation factor 6 regulates a novel plasma membrane recycling pathway. *J Cell Biol, 139*(1), 49-61.

Ragan, M. A., Goggin, C. L., Cawthorn, R. J., Cerenius, L., Jamieson, A. V., Plourde, S. M., *et al*. (1996). A novel clade of protistan parasites near the animal-fungal divergence. *Proc Natl Acad Sci U S A, 93*(21), 11907-11912.

Rasin, M. R., Gazula, V. R., Breunig, J. J., Kwan, K. Y., Johnson, M. B., Liu-Chen, S., *et al*. (2007). Numb and Numbl are required for maintenance of cadherin-based adhesion and polarity of neural progenitors. *Nat Neurosci, 10*(7), 819-827.

Rambaut, A. (2007). Se-Al. http://tree.bio.ed.ac.uk/software/seal/.

Ramesh, M. A., Malik, S. B., & Logsdon, J. M., Jr. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol, 15*(2), 185-191.

Ramjaun, A. R., & McPherson, P. S. (1998). Multiple amphiphysin II splice variants display differential clathrin binding: identification of two distinct clathrin-binding sites. *J Neurochem, 70*(6), 2369-2376.

Ramoino, P., Gallus, L., Beltrame, F., Diaspro, A., Fato, M., Rubini, P., *et al*. (2006). Endocytosis of GABAB receptors modulates membrane excitability in the single-celled organism Paramecium. *J Cell Sci, 119*(Pt 10), 2056-2064.

Raven, J. A., Beardall, J., Flynn, K. J., & Maberly, S. C. (2009). Phagotrophy in the origins of photosynthesis in eukaryotes and as a complementary mode of nutrition in phototrophs: relation to Darwin's insectivorous plants. *J Exp Bot, 60*(14), 3975-3987.

Reyes-Prieto, A., & Bhattacharya, D. (2007). Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol, 45*(1), 384-391.

Reyes-Prieto, A., Moustafa, A., & Bhattacharya, D. (2008). Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. *Curr Biol, 18*(13), 956-962.

Ribichich, K. F., Salem-Izacc, S. M., Georg, R. C., Vencio, R. Z., Navarro, L. D., & Gomes, S. L. (2005). Gene discovery and expression profile analysis through sequencing of expressed sequence tags from different developmental stages of the chytridiomycete *Blastocladiella emersonii*. *Eukaryot Cell, 4*(2), 455-464.

Ricard, G., McEwan, N. R., Dutilh, B. E., Jouany, J. P., Macheboeuf, D., Mitsumori, M., *et al*. (2006). Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics, 7*, 22.

Richards, T. A., & Cavalier-Smith, T. (2005). Myosin domain evolution and the primary

divergence of eukaryotes. *Nature, 436*(7054), 1113-1118.

Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D., & Hall, A. (1992). The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell, 70*(3), 401-410.

Riento, K., Frick, M., Schafer, I., & Nichols, B. J. (2009). Endocytosis of flotillin-1 and flotillin-2 is regulated by Fyn kinase. *J Cell Sci, 122*(Pt 7), 912-918.

Ringstad, N., Nemoto, Y., & De Camilli, P. (1997). The SH3p4/Sh3p8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain. *Proc Natl Acad Sci U S A, 94*(16), 8569-8574.

Rivera, M. C., & Lake, J. A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature, 431*(7005), 152-155.

Rivero, M. R., Vranych, C. V., Bisbal, M., Maletto, B. A., Ropolo, A. S., & Touz, M. C. (2010). Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in Giardia lamblia. *Biochem J, 428*(1), 33-45.

Robibaro, B., Hoppe, H. C., Yang, M., Coppens, I., Ngo, H. M., Stedman, T. T., *et al.* (2001). Endocytosis in different lifestyles of protozoan parasitism: role in nutrient uptake with special reference to *Toxoplasma gondii*. *Int J Parasitol, 31*(12), 1343-1353.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Loffelhardt, W., *et al.* (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol, 15*(14), 1325-1330.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A. J., Gray, M. W., Philippe, H., *et al*. (2007). Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol, 17*(16), 1420-1425.

Roger, A. J., & Simpson, A. G. (2009). Evolution: revisiting the root of the eukaryote tree. *Curr Biol, 19*(4), R165-167.

Rogozin, I. B., Basu, M. K., Csuros, M., & Koonin, E. V. (2009). Analysis of Rare Genomic Changes Does Not Support the Unikont-Bikont Phylogeny and Suggests Cyanobacterial Symbiosis as the Point of Primary Radiation of Eukaryotes. *Genome Biol Evol 2009*(0), 99-113.

Rohatgi, R., Ma, L., Miki, H., Lopez, M., Kirchhausen, T., Takenawa, T., *et al.* (1999). The Interaction between N-WASP and the Arp2/3 Complex Links Cdc42-Dependent Signals to Actin Assembly. *Cell, 97*(2), 221.

Romih, R., & Jezernik, K. (1994). Endocytosis during postnatal differentiation in superficial cells of the mouse urinary bladder epithelium. *Cell Biol Int, 18*(6), 663-668.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics, 19*(12), 1572-1574.

Rosenthal, J. A., Chen, H., Slepnev, V. I., Pellegrini, L., Salcini, A. E., Di Fiore, P. P., *et al.* (1999). The epsins define a family of proteins that interact with components of the clathrin coat and contain a new protein module. *J Biol Chem, 274*(48), 33959-33965.

Royle, S. J., Bright, N. A., & Lagnado, L. (2005). Clathrin is required for the function of the mitotic spindle. *Nature, 434*(7037), 1152-1157.

Ruiz-Trillo, I., Lane, C. E., Archibald, J. M., & Roger, A. J. (2006). Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts Capsaspora owczarzaki and Sphaeroforma arctica. *J Eukaryot Microbiol, 53*(5), 379-384.

Ruiz-Trillo, I., Burger, G., Holland, P. W., King, N., Lang, B. F., Roger, A. J., *et al.* (2007). The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet, 23*(3), 113-118.

Rusk, N., Le, P. U., Mariggio, S., Guay, G., Lurisci, C., Nabi, I. R., *et al.* (2003). Synaptojanin 2 functions at an early step of clathrin-mediated endocytosis. *Curr Biol, 13*(8), 659-663.

Sabharanjak, S., Sharma, P., Parton, R. G., & Mayor, S. (2002). GPI-anchored proteins are delivered to recycling endosomes via a distinct cdc42-regulated, clathrin-independent pinocytic pathway. *Dev Cell, 2*(4), 411-423.

Salazar, M. A., Kwiatkowski, A. V., Pellegrini, L., Cestra, G., Butler, M. H., Rossman, K. L., *et al.* (2003). Tuba, a novel protein containing bin/amphiphysin/Rvs and Dbl homology domains, links dynamin to regulation of the actin cytoskeleton. *J Biol Chem, 278*(49), 49031-49043.

Salcini, A. E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O., *et al.* (1997). Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. *Genes Dev, 11*(17), 2239-2249.

Sallusto, F., Cella, M., Danieli, C., & Lanzavecchia, A. (1995). Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J Exp Med, 182*(2), 389-400.

Sanchez-Martinez, C., & Perez-Martin, J. (2001). Dimorphism in fungal pathogens: *Candida albicans* and *Ustilago maydis*--similar inputs, different outputs. *Curr Opin Microbiol, 4*(2), 214-221.

Santarella-Mellwig, R., Franke, J., Jaedicke, A., Gorjanacz, M., Bauer, U., Budd, A., *et*

al. (2010). The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol, 8*(1), e1000281.

Santolini, E., Puri, C., Salcini, A. E., Gagliani, M. C., Pelicci, P. G., Tacchetti, C., *et al.* (2000). Numb is an endocytic protein. *J Cell Biol, 151*(6), 1345-1352.

Savill, J., Gregory, C., & Haslett, C. (2003). Cell biology. Eat me or die. *Science, 302*(5650), 1516-1517.

Sanchez-Puerta, M. V., Lippmeier, J. C., Apt, K. E., & Delwiche, C. F. (2007). Plastid genes in a non-photosynthetic dinoflagellate. *Protist, 158*(1), 105-117.

Sauvonnet, N., Dujeancourt, A., & Dautry-Varsat, A. (2005). Cortactin and dynamin are required for the clathrin-independent endocytosis of gammac cytokine receptor. *J Cell Biol, 168*(1), 155-163.

Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., *et al.* (2010). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res, 38*(Database issue), D5-16.

Scheele, U., Alves, J., Frank, R., Duwel, M., Kalthoff, C., & Ungewickell, E. (2003). Molecular and functional characterization of clathrin- and AP-2-binding determinants within a disordered domain of auxilin. *J Biol Chem, 278*(28), 25357-25368.

Scheele, U., Kalthoff, C., & Ungewickell, E. (2001). Multiple interactions of auxilin 1 with clathrin and the AP-2 adaptor complex. *J Biol Chem, 276*(39), 36131-36138.

Schledzewski, K., Brinkmann, H., & Mendel, R. R. (1999). Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin

of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. *J Mol Evol, 48*(6), 770-778.

Schluter, A., Fourcade, S., Ripp, R., Mandel, J. L., Poch, O., & Pujol, A. (2006). The evolutionary origin of peroxisomes: an ER-peroxisome connection. *Mol Biol Evol, 23*(4), 838-845.

Schmid, E. M., Ford, M. G., Burtey, A., Praefcke, G. J., Peak-Chew, S. Y., Mills, I. G., *et al.* (2006). Role of the AP2 beta-appendage hub in recruiting partners for clathrin-coated vesicle assembly. *PLoS Biol, 4*(9), e262.

Schmid, E. M., & McMahon, H. T. (2007). Integrating molecular and network biology to decode endocytosis. *Nature, 448*(7156), 883-888.

Schneider, M., Lane, L., Boutet, E., Lieberherr, D., Tognolli, M., Bougueleret, L., *et al.* (2009). The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J Proteomics, 72*(3), 567-573.

Schneider, Y. J., Tulkens, P., de Duve, C., & Trouet, A. (1979). Fate of plasma membrane during endocytosis. II. Evidence for recycling (shuttle) of plasma membrane constituents. *J Cell Biol, 82*(2), 466-474.

Seet, L. F., & Hong, W. (2006). The Phox (PX) domain proteins and membrane traffic. *Biochim Biophys Acta, 1761*(8), 878-896.

Seger, S., Rischatsch, R., & Philippsen, P. (2011). Formation and stability of eisosomes in the filamentous fungus *Ashbya gossypii*. *J Cell Sci, 124*(Pt 10), 1629-1634.

Senetar, M. A., Foster, S. J., & McCann, R. O. (2004). Intrasteric inhibition mediates the interaction of the I/LWEQ module proteins Talin1, Talin2, Hip1, and Hip12 with actin. *Biochemistry, 43*(49), 15418-15428.

Sengar, A. S., Wang, W., Bishay, J., Cohen, S., & Egan, S. E. (1999). The EH and SH3

domain Ese proteins regulate endocytosis by linking to dynamin and Eps15. *Embo J, 18*(5), 1159-1171.

Sesma, A., & Osbourn, A. E. (2004). The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature, 431*(7008), 582-586.

Shajahan, A. N., Timblin, B. K., Sandoval, R., Tiruppathi, C., Malik, A. B., & Minshall, R. D. (2004). Role of Src-induced dynamin-2 phosphorylation in caveolae-mediated endocytosis in endothelial cells. *J Biol Chem, 279*(19), 20392-20400.

Shalchian-Tabrizi, K., Minge, M. A., Espelund, M., Orr, R., Ruden, T., Jakobsen, K. S., *et al.* (2008). Multigene phylogeny of choanozoa and the origin of animals. *PLoS One, 3*(5), e2098.

Sharma, D. K., Brown, J. C., Choudhury, A., Peterson, T. E., Holicky, E., Marks, D. L., *et al.* (2004). Selective stimulation of caveolar endocytosis by glycosphingolipids and cholesterol. *Mol Biol Cell, 15*(7), 3114-3122.

Shih, W., Gallusser, A., & Kirchhausen, T. (1995). A clathrin-binding site in the hinge of the beta 2 chain of mammalian AP-2 complexes. *J Biol Chem, 270*(52), 31083-31090.

Shimada, A., Niwa, H., Tsujita, K., Suetsugu, S., Nitta, K., Hanawa-Suetsugu, K., *et al.* (2007). Curved EFC/F-BAR-domain dimers are joined end to end into a filament for membrane invagination in endocytosis. *Cell, 129*(4), 761-772.

Shin, N., Lee, S., Ahn, N., Kim, S. A., Ahn, S. G., YongPark, Z., *et al.* (2007). Sorting nexin 9 interacts with dynamin 1 and N-WASP and coordinates synaptic vesicle endocytosis. *J Biol Chem, 282*(39), 28939-28950.

Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis

using vaccinia DNA topoisomerase. *J Biol Chem, 269*(51), 32678-32684.

Sigismund, S., Woelk, T., Puri, C., Maspero, E., Tacchetti, C., Transidico, P., *et al*. (2005). Clathrin-independent endocytosis of ubiquitinated cargos. *Proc Natl Acad Sci U S A, 102*(8), 2760-2765.

Simon, S. I., & Schmid-Schonbein, G. W. (1988). Biophysical aspects of microsphere engulfment by human neutrophils. *Biophys J, 53*(2), 163-173.

Simpson, A. G. (2003). Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int J Syst Evol Microbiol, 53*(Pt 6), 1759-1777.

Simpson, A. G., Inagaki, Y., & Roger, A. J. (2006). Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol, 23*(3), 615-625.

Simpson, A. G., & Roger, A. J. (2004). The real 'kingdoms' of eukaryotes. *Curr Biol, 14*(17), R693-696.

Slamovits, C. H., & Keeling, P. J. (2008). Plastid-derived genes in the nonphotosynthetic alveolate Oxyrrhis marina. *Mol Biol Evol, 25*(7), 1297-1306.

Slepnev, V. I., Ochoa, G. C., Butler, M. H., & De Camilli, P. (2000). Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. *J Biol Chem, 275*(23), 17583-17589.

Smirnov, A., Nassonova, E., Berney, C., Fahrni, J., Bolivar, I., & Pawlowski, J. (2005). Molecular phylogeny and classification of the lobose amoebae. *Protist, 156*(2), 129-142.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular

subsequences. *J Mol Biol, 147*(1), 195-197.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., & Peattie, D. A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science, 243*(4887), 75-77.

Somlyo, A. P., Devine, C. E., Somlyo, A. V., & North, S. R. (1971). Sarcoplasmic reticulum and the temperature-dependent contraction of smooth muscle in calcium-free solutions. *J Cell Biol, 51*(3), 722-741.

Song, W., & Zinsmaier, K. E. (2003). Endophilin and synaptojanin hook up to promote synaptic vesicle endocytosis. *Neuron, 40*(4), 665-667.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688-2690.

Stamatakis, A., Ludwig, T., & Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics, 21*(4), 456-463.

Stan, R. V. (2005). Structure of caveolae. *Biochim Biophys Acta, 1746*(3), 334-348.

Stechmann, A., & Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. *Science, 297*(5578), 89-91.

Stechmann, A., & Cavalier-Smith, T. (2003). The root of the eukaryote tree pinpointed. *Curr Biol, 13*(17), R665-666.

Steel, M., Huson, D., & Lockhart, P. J. (2000). Invariable sites models and their use in phylogeny reconstruction. *Syst Biol, 49*(2), 225-232.

Steenkamp, E. T., Wright, J., & Baldauf, S. L. (2006). The protistan origins of animals and fungi. *Mol Biol Evol, 23*(1), 93-106.

Stuart, L. M., & Ezekowitz, R. A. (2008). Phagocytosis and comparative innate immunity: learning on the fly. *Nat Rev Immunol, 8*(2), 131-141.

Subtil, A., Hemar, A., & Dautry-Varsat, A. (1994). Rapid endocytosis of interleukin 2 receptors when clathrin-coated pit endocytosis is inhibited. *J Cell Sci, 107 (Pt 12)*, 3461-3468.

Svitkina, T. (2007). N-WASP Generates a Buzz at Membranes on the Move. *Cell, 128*(5), 828.

Swaminathan, S. (2006). Eisosomes: endocytic portals. [10.1038/ncb0406-310]. *Nat Cell Biol, 8*(4), 310-310.

Swanson, J. A. (1989). Phorbol esters stimulate macropinocytosis and solute flow through macrophages. *J Cell Sci, 94 (Pt 1)*, 135-142.

Sweigard, J. A., Carroll, A. M., Farrall, L. & Valent, B. (1997). A series of vectors for fungal transformation. *Fungal Genet. Newslett, 44*, 52-53.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol, 56*(4), 564-577.

Talbot, N. J. (2003). On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu Rev Microbiol, 57*, 177-202.

Talbot, N. J., & Foster, A. J. (2001). Genetics and genomics of the rice blast fungus *Magnaporthe grisea*: Developing an experimental model for understanding fungal diseases of cereals *Advances in Botanical Research* (Vol. Volume 34, pp. 263-287): Academic Press.

Tebar, F., Bohlander, S. K., & Sorkin, A. (1999). Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions

with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol Biol Cell, 10*(8), 2687-2702.

Tebar, F., Sorkina, T., Sorkin, A., Ericsson, M., & Kirchhausen, T. (1996). Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits. *J Biol Chem, 271*(46), 28727-28730.

Thompson, G., Owen, D., Chalk, P. A., & Lowe, P. N. (1998). Delineation of the Cdc42/Rac-binding domain of p21-activated kinase. *Biochemistry, 37*(21), 7885-7891.

Thomsen, P., Roepstorff, K., Stahlhut, M., & van Deurs, B. (2002). Caveolae are highly immobile plasma membrane microdomains, which are not involved in constitutive endocytic trafficking. *Mol Biol Cell, 13*(1), 238-250.

Tomas, R. N., & Cox, E. R. (1973). Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. I. Ultrastructure1,2. *Journal of Phycology, 9*(3), 304-323.

Tovar, J., Fischer, A., & Clark, C. G. (1999). The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol Microbiol, 32*(5), 1013-1021.

Traub, L. M. (2009). Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nat Rev Mol Cell Biol, 10*(9), 583-596.

Traub, L. M. (2003). Sorting it out: AP-2 and alternate clathrin adaptors in endocytic cargo selection. *Journal Of Cell Biology, 163*(2), 203-208.

Traub, L. M., Bannykh, S. I., Rodel, J. E., Aridor, M., Balch, W. E., & Kornfeld, S. (1996). AP-2-containing clathrin coats assemble on mature lysosomes. *J Cell Biol, 135*(6 Pt 2), 1801-1814.

Traub, L. M., Downs, M. A., Westrich, J. L., & Fremont, D. H. (1999). Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. *Proc Natl Acad Sci U S A, 96*(16), 8907-8912.

Tsuboi, S., Takada, H., Hara, T., Mochizuki, N., Funyu, T., Saitoh, H., *et al*. (2009). FBP17 Mediates a Common Molecular Step in the Formation of Podosomes and Phagocytic Cups in Macrophages. *J Biol Chem, 284*(13), 8548-8556.

Turk, J. L. (1991). Metchnikoff revisited. *J R Soc Med, 84*(10), 579-580.

Uher, F., Dobronyi, I., & Gergel, J. (1981). IgM-Fc receptor-mediated phagocytosis of rat macrophages. *Immunology, 42*(3), 419-425.

Ungewickell, E. J., & Hinrichsen, L. (2007). Endocytosis: clathrin-mediated membrane budding. *Curr Opin Cell Biol, 19*(4), 417-425.

Urrutia, R., Henley, J. R., Cook, T., & McNiven, M. A. (1997). The dynamins: redundant or distinct functions for an expanding family of related GTPases? *Proc Natl Acad Sci U S A, 94*(2), 377-384.

Valent, B., Farrall, L., & Chumley, F. G. (1991). *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics, 127*(1), 87-101.

van den Ent, F., Amos, L. A., & Lowe, J. (2001). Prokaryotic origin of the actin cytoskeleton. *Nature, 413*(6851), 39-44.

van der Giezen, M., & Tovar, J. (2005). Degenerate mitochondria. *EMBO Rep, 6*(6), 525-530.

Vangelatos, I., Roumelioti, K., Gournas, C., Suarez, T., Scazzocchio, C., & Sophianopoulou, V. (2010). Eisosome organisation in the filamentous ascomycete *Aspergillus nidulans*. *Eukaryot Cell*.

Varma, R., & Mayor, S. (1998). GPI-anchored proteins are organized in submicron domains at the cell surface. *Nature, 394*(6695), 798-801.

Verstreken, P., Koh, T. W., Schulze, K. L., Zhai, R. G., Hiesinger, P. R., Zhou, Y., *et al.* (2003). Synaptojanin is recruited by endophilin to promote synaptic vesicle uncoating. *Neuron, 40*(4), 733-748.

Viguera, A. R., Arrondo, J. L., Musacchio, A., Saraste, M., & Serrano, L. (1994). Characterization of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemistry, 33*(36), 10925-10933.

Vogel, G., Thilo, L., Schwarz, H., & Steinhart, R. (1980). Mechanism of phagocytosis in *Dictyostelium discoideum*: phagocytosis is mediated by different recognition sites as disclosed by mutants with altered phagocytotic properties. *J Cell Biol, 86*(2), 456-465.

von Dohlen, C. D., Kohler, S., Alsop, S. T., & McManus, W. R. (2001). Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature, 412*(6845), 433-436.

Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A., & Woese, C. R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature, 326*(6111), 411-414.

Waddell, D. R., & Vogel, G. (1985). Phagocytic behavior of the predatory slime mold, *Dictyostelium caveatum*. Cell nibbling. *Exp Cell Res, 159*(2), 323-334.

Walsh, P., Bursac, D., Law, Y. C., Cyr, D., & Lithgow, T. (2004). The J-protein family: modulating protein assembly, disassembly and translocation. *EMBO Rep, 5*(6), 567-571.

Walther, T. C., Aguilar, P. S., Frohlich, F., Chu, F., Moreira, K., Burlingame, A. L., *et al.*

(2007). Pkh-kinases control eisosome assembly and organization. *Embo J, 26*(24), 4946-4955.

Walther, T. C., Brickner, J. H., Aguilar, P. S., Bernales, S., Pantoja, C., & Walter, P. (2006). Eisosomes mark static sites of endocytosis. *Nature, 439*(7079), 998-1003.

Wang, E., Pennington, J. G., Goldenring, J. R., Hunziker, W., & Dunn, K. W. (2001). Brefeldin A rapidly disrupts plasma membrane polarity by blocking polar sorting in common endosomes of MDCK cells. *J Cell Sci, 114*(Pt 18), 3309-3321.

Wang, J. T., Kerr, M. C., Karunaratne, S., Jeanes, A., Yap, A. S., & Teasdale, R. D. (2010). The SNX-PX-BAR family in macropinocytosis: the regulation of macropinosome formation by SNX-PX-BAR proteins. *PLoS One, 5*(10), e13763.

Wendland, B., & Emr, S. D. (1998). Pan1p, yeast eps15, functions as a multivalent adaptor that coordinates protein-protein interactions essential for endocytosis. *Journal Of Cell Biology, 141*(1), 71-84.

Wendland, J., & Walther, A. (2005). *Ashbya gossypii*: a model for fungal developmental biology. *Nat Rev Microbiol, 3*(5), 421-429.

Wettey, F. R., Hawkins, S. F., Stewart, A., Luzio, J. P., Howard, J. C., & Jackson, A. P. (2002). Controlled elimination of clathrin heavy-chain expression in DT40 lymphocytes. *Science, 297*(5586), 1521-1525.

Whittaker, R. H. (1969). New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science, 163*(863), 150-160.

Wickstead, B., & Gull, K. (2007). Dyneins across eukaryotes: a comparative genomic

analysis. *Traffic, 8*(12), 1708-1721.

Wickstead, B., Gull, K., & Richards, T. A. (2010). Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol Biol, 10*, 110.

Wiejak, J., Surmacz, L., & Wyroba, E. (2004). Dynamin- and clathrin-dependent endocytic pathway in unicellular eukaryote Paramecium. *Biochem Cell Biol, 82*(5), 547-558.

Wigge, P., Kohler, K., Vallis, Y., Doyle, C. A., Owen, D., Hunt, S. P., *et al.* (1997). Amphiphysin heterodimers: potential role in clathrin-mediated endocytosis. *Mol Biol Cell, 8*(10), 2003-2015.

Williams, B. A., Hirt, R. P., Lucocq, J. M., & Embley, T. M. (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature, 418*(6900), 865-869.

Wilson, R. A., & Talbot, N. J. (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat Rev Microbiol, 7*(3), 185-195.

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature, 387*(6634), 708-713.

Worby, C. A., & Dixon, J. E. (2002). Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol, 3*(12), 919-931.

Wu, Y., Matsui, H., & Tomizawa, K. (2009). Amphiphysin I and regulation of synaptic vesicle endocytosis. *Acta Med Okayama, 63*(6), 305-323.

Yamabhai, M., Hoffman, N. G., Hardison, N. L., McPherson, P. S., Castagnoli, L., Cesareni, G., *et al.* (1998). Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains. *J Biol Chem, 273*(47), 31401-

31407.

Yamada, H., Ohashi, E., Abe, T., Kusumi, N., Li, S. A., Yoshida, Y., *et al*. (2007). Amphiphysin 1 is important for actin polymerization during phagocytosis. *Mol Biol Cell, 18*(11), 4669-4680.

Yamada, H., Padilla-Parra, S., Park, S. J., Itoh, T., Chaineau, M., Monaldi, I., *et al.* (2009). Dynamic interaction of amphiphysin with N-WASP regulates actin assembly. *J Biol Chem, 284*(49), 34244-34256.

Yarar, D., Waterman-Storer, C. M., & Schmid, S. L. (2007). SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Dev Cell, 13*(1), 43-56.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol, 11*(9), 367-372.

Yarar, D., Waterman-Storer, C. M., & Schmid, S. L. (2007). SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Dev Cell, 13*(1), 43-56.

Yoshizawa, A. C., Kawashima, S., Okuda, S., Fujita, M., Itoh, M., Moriya, Y., *et al.* (2006). Extracting sequence motifs and the phylogenetic features of SNARE-dependent membrane traffic. *Traffic, 7*(8), 1104-1118.

Yutin, N., Wolf, M. Y., Wolf, Y. I., & Koonin, E. V. (2009). The origins of phagocytosis and eukaryogenesis. *Biol Direct, 4*, 9.

Zettler, L. A. A., Nerad, T. A., O'Kelly, C. J., & Sogin, M. L. (2001). The nucleariid amoebae: more protists at the animal-fungal boundary. *J Eukaryot Microbiol, 48*(3), 293-297.

Zhao, Y. Y., Liu, Y., Stan, R. V., Fan, L., Gu, Y., Dalton, N., *et al*. (2002). Defects in

caveolin-1 cause dilated cardiomyopathy and pulmonary hypertension in knockout mice. *Proc Natl Acad Sci U S A, 99*(17), 11375-11380.

Zhang, X., Lester, R. L., & Dickson, R. C. (2004). Pil1p and Lsp1p negatively regulate the 3-phosphoinositide-dependent protein kinase-like kinase Pkh1p and downstream signaling pathways Pkc1p and Ypk1p. J Biol Chem, 279(21), 22030-22038.

Ziolkowska, N. E., Karotki, L., Rehman, M., Huiskonen, J. T., & Walther, T. C. (2011). Eisosome-driven plasma membrane organization is mediated by BAR domains. *Nat Struct Mol Biol, 18*(7), 854-856.