Evolution of the eukaryotic membrane trafficking system as revealed by comparative genomic and phylogenetic analysis of adaptin, golgin, and SNARE proteins

by

Lael Dan Barlow

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physiology, Cell, and Developmental Biology

Department of Biological Sciences
University of Alberta

Abstract

All eukaryotic cells possess a complex system of endomembranes that functions in trafficking molecular cargo within the cell, which is not observed in prokaryotic cells. This membrane trafficking system is fundamental to the cellular physiology of extant eukaryotes, and includes organelles such as the endoplasmic reticulum, Golgi apparatus, and endosomes as well as the plasma membrane. The evolutionary history of this system offers an over-arching framework for research on membrane trafficking in the field of cell biology. However, the evolutionary origins of this system in the evolution from a prokaryotic ancestor to the most recent common ancestor of extant eukaryotes is a major evolutionary transition that remains poorly understood. A leading paradigm is described by the previously proposed Organelle Paralogy Hypothesis, which posits that coordinated duplication and divergence of genes encoding organelle-specific membrane trafficking proteins underlies a corresponding evolutionary history of organelle differentiation that produced the complex sets of membrane trafficking organelles found in extant eukaryotes. This thesis focuses on investigating the evolution of families of proteins that sustain membrane traffic by organizing vesicle transport between specific compartments.

Comparative genomics and phylogenetic analysis methods were applied to trace the evolution of subunits of the Adaptor Protein complexes, which function in vesicle formation, revealing that losses of genes encoding these subunits in Fungi and duplications in plants are relatively recent events, consistent with a highly conserved role of members of this protein family in organizing the membrane trafficking system. Similar methods were applied to the golgin proteins, which extend from membranes of the Golgi apparatus to recognize and tether specific vesicles. This revealed considerable conservation of golgins specific to different regions of the Golgi, consistent with ancestral complexity of the Golgi inferred from conservation of morphological complexity

among eukaryotes. The Soluble *N*-ethylmaleimide sensitive factor Attachment protein REceptor (SNARE) protein superfamily comprises numerous proteins forming complexes that mediate vesicle fusion at specific locations in the cell. An extensive analysis of the SNAREs revealed evidence for additional SNARE proteins present in the last common ancestor of eukaryotes. Finally, the evolution of the four constituent families within the SNARE superfamily was explored and relevance to various potential scenarios of early evolution of the membrane trafficking system are discussed.

Taken together, the results of this work shed light on the connections between protein evolution and organelle evolution, and provide novel evidence for distinguishing between alternative scenarios for the evolution of the membrane trafficking system.

Preface

Chapter 2 (exclusive of the preface and afterword) has been published as Barlow, L.D., Dacks, J.B., Wideman, J.G., 2014. From all to (nearly) none: Tracing adaptin evolution in Fungi. *Cellular Logistics* 4, e28114. https://doi.org/10.4161/cl.28114. This is an open-access article licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License (https://creativecommons.org/licenses/by-nc/3.0/). All text in Chapter 2 (exclusive of the preface and afterword) is unmodified from this publication, with the exception of some additional detail added to figure legends. JGW and JBD conceived of and designed the study. JGW performed most of the similarity searches and drafted the manuscript. I performed all phylogenetic analyses, generated figures, and contributed to writing the manuscript.

Chapter 3 (exclusive of the preface and afterword) has been published as Larson, R.T., Dacks, J.B., and Barlow, L.D., 2019. Recent gene duplications dominate evolutionary dynamics of adaptor protein complex subunits in embryophytes. *Traffic* (in press) https://onlinelibrary.wiley.com/doi/abs/10.1111/tra.12698. I obtained permission from John Wiley & Sons, Inc. to reproduce this work in this thesis, and the license number is 4695550381953. All text in Chapter 3 (exclusive of the preface and afterward) is unmodified from this publication (with the exception of a minor correction to the definition of the acronym HSP). With my mentorship, RTL performed similarity searches and phylogenetic analyses for this work as part of one of her undergraduate projects in the Dacks laboratory. JBD initially conceived of the study. RTL and I designed and performed the analyses, and wrote the manuscript. I conceived of and wrote all computer scripts used for performing the analyses. RTL and I prepared figures.

Chapter 4 (exclusive of preface and afterword) was published as Barlow, L.D., Nỳvltovà, E., Aguilar, M., Tachezy, J., Dacks, J.B., 2018. A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biology* 16. https://doi.org/10.1186/s12915-018-0492-9. This article was distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). All text in Chapter 4 (exclusive of preface and afterword) is unmodified from this publication, with the exception of changes applied as indicated in the published correction to the article (https://rdcu.be/bR35t). EN and JT designed, performed and interpreted the experimental work on *Mastigamoeba balamuthi*, and also

wrote the relevant manuscript sections. MA and I performed genomic analysis of *M. balamuthi*. I performed all molecular evolutionary analyses. JBD and I designed the molecular evolutionary analyses, interpreted the results, and wrote the manuscript.

Bioinformatic analyses in Chapter 5 are in preparation for eventual submission as a first-author publication with Joel B. Dacks. The idea for this project was initially conceived of by JBD. I performed all the work, including conceiving of and writing all computer scripts used for performing the analyses. Though intriguing, experimental work done with *Dictyostelium discoideum* in Chapter 5 is inconclusive and therefore is not currently intended for publication with the phylogenetic results. I performed this experimental work during a 6 week visit to the laboratory of Robert R. Kay at the Medical Research Council Laboratory of Molecular Biology, where I was instructed in experimental methods primarily by Gareth Bloomfield. Therefore, it is important to note that this experimental work was done in collaboration with RRK and GB.

Acknowledgements

During my graduate studies I was supported by a Post-Graduate Scholarship-Doctoral (PGS-D) from the National Sciences and Engineering Council of Canada (NSERC), a President's Doctoral Prize of Distinction and a 75th Anniversary Award from the University of Alberta, a Short Term Travelling Fellowship from the Company of Biologists (Cambridge, England), two Holtz-Connor travel awards from the International Society of Protistologists, and a travel award from the International Society for Evolutionary Protistology. Significant funds for covering my stipend also came from NSERC Discovery grants awarded to my supervisor Joel Dacks.

Chapter 2 specific acknowledgements: I am greatful for the opportunity to work with my co-authors on the associated publication: Joel B. Dacks and Jeremy G. Wideman. This work was supported by funding awarded to Joel B. Dacks from the following sources: An NSERC Discovery Grant, a New Faculty Award from Alberta Innovates Technology Futures, and the Canada Research Chairs program. We (myself and the other authors of the associated publication) would like to thank the members of the Dacks lab for discussion, as well as K. James, B. Kopetski, C. Lavergne, S. McCormack, N. Ellis, and C. Thulin for organizational and administrative support.

Chapter 3 specific acknowledgements: It was an honour to work with my co-authors Raegan T. Larson and Joel B. Dacks on the associated publication. This project could not have happened without Raegan's hard work and enthusiasm for plants. This work was supported by (NSERC) Discovery grants (RES0021028, RES0043758, RES0046091) and Joel B. Dacks is the Canada Research Chair (Tier II) in Evolutionary Cell Biology. Raegan T. Larson was supported by an Office of the Provost and Vice President (Academic) Summer Student Award from the University of Alberta Faculty of Medicine and Dentistry. For this work, I was supported by my NSERC PGS-D award.

Chapter 4 specific acknowledgements: I very much appreciate the important contributions of my co-authors on the associated publication: Eva Nỳvltovà, Maria Aguilar, Jan Tachezy, and Joel B. Dacks. In particular, Eva and Jan provided all the *Mastigamoeba balamuthi* data including genomic data and localization experiments. We (myself and co-authors on the associated publication) would like to thank Christen M. Klinger for collaboration on informatics workflows used for running homology searches and all members of the Dacks lab, past and present, for helpful discussion. We also want to thank Drs Aaron Turkewitz, Paul Melancon, Alan Warren and Frances Brodsky for helpful discussion. In performing this work, I was supported by my NSERC PGS-D award.

Maria Aguilar was supported by an Alberta Innovates Technology Futures Postdoctoral Fellowship. This work was also supported by NSERC Discovery grant (RES0021028) awarded to Joel B. Dacks who is the Canada Research Chair (Tier II) in Evolutionary Cell Biology. Jan Tachezy was supported by the Czech Science Foundation (16-06123S), BIOCEV (CZ.1.05/1.1.00/02.0109), and LQ1604 NPU II provided by MEYS CR. We (myself and the other authors of the associated publication) also acknowledge the Imaging Methods Core Facility at BIOCEV supported by the Czech-BioImaging RI project (LM2015062 funded by MEYS CR).

Chapter 5 specific acknowledgements: For this work, I was supported by my NSERC PGS-D award, as well as a Short Term Travelling Fellowship from the Company of Biologists. I would like to thank Prof. Robert R. Kay for hosting me in his laboratory and providing materials and guidance regarding work with *Dictyostelium discoideum*. I would also especially like to thank Dr. Gareth Bloomfield for instructing me in working with *D. discoideum*. Other very helpful members (and former members) of the Kay lab were Dr. Peggy Paschke, Dr. Thomas D. Williams, and Dr. David Traynor. In addition, Prof. Andrew Simmonds, Dr. Gary Eitzen, and Dr. Paul LaPointe shared several reagents and pieces of equipment from their laboratories with me for cloning *D. discoideum* genes. Also, Maria Aguilar and Avinash Sheshachalam were very kind and helpful to me while I was cloning. Moreover, related work published previously was done in collaboration with Nerissa N. Nankissoor (an excellent undergraduate student who I mentored in the Dacks laboratory) as well as others including Mark C. Field and his student Divya Venkatesh.

Working with Joel Dacks first as an undergraduate and then as a graduate student was without a doubt one of the most significant opportunities I have been given in my life. Joel has also been a consistent source of encouragement, which often makes all the difference.

I am also very fortunate in having an excellent supervisory committee. I am especially grateful to my co-supervisor Sally Leys for taking me under her wing during changes to my program of study. Emmanuelle Cordat has been a reliable source of wisdom over the years, for which I am very thankful. I also benefited significantly from the guidance of the members of the examining committee for my PhD candidacy exam: Thomas Simmen, Paul Melancon, and Joanne Lemieux. Also, I would especially like to thank all the members of my PhD thesis defense examining committee including Warren Gallin and Toni Gabaldón for reading my thesis and providing thoughtful comments and valuable constructive criticism. Thanks also to Toni Gabaldón for travelling from Barcelona for the defense.

A long list of colleagues in the Dacks lab over the years have been a great pleasure to work

with, and collaborated with me on many productive projects. This list includes: In no particular order: Alexander Schlacht, Emily Herman, Mary Klute, Christen Klinger, Serah Jacob, Ben Pi, Jeremy Wideman, Laura Lee, Maria Aguilar, Inma Ramirez-Macias, Elisabeth Richardson, Shweta Pipalya, Nerissa Nankissoor, Alexander Filbert, Raegan Larson, Alexa Thompson, John McFarlane. Very helpful technical support was provided by Igor Sinelnikov.

Many more instructors, mentors, colleagues, and friends from various departments on campus have enriched my experience over the years at the U of A, certainly more than I can name here. To name just a few: Denis Lamoureux has been a valuable source of advice and encouraged me in pursuing graduate studies. I'd also like to thank, Allen Fu, Fred Mast, Avinash Sheshachalam, Matthew Anderson-Baron, Conrad Pridie, and Azra Lari, and members of the Cell Biology Student's Association.

Angelina, my wife, has been amazingly encouraging and supportive over the many years of graduate school, and this was absolutely essential. My parents were also especially encouraging, and supported my interest in biology from a young age. My siblings and my extended family also played important roles in supporting my graduate studies.

Contents

List of Tables

List of Figures

List of abbreviations and definitions of terms

AAA ATPases Associated with diverse cellular Activities
ACBD3 Acyl-Coenzyme A Binding Domain containing 3 protein
AGFG Arf GTPase activating protein with FG repeats
ALPS Amphipathic Lipid Packing Sensor
AMOEBAE Analysis of Molecular Evolution with BAtch Entry
AP Adaptor Protein
ATP Adenosine TriPhosphate
AU Approximately Unbiased

BIC Bayesian Information Criterion
BLAST Basic Local Alignment Search Tool

CASP CCAAT-Displacement Protein/cut Alternatively Spliced Product
CDC Cell Division Control
CENDIK Cerebral dysgenesis, Neuropathy, Ichthyosis, and Keratoderma
CG-NAP Centrosome and Golgi localized PKN-Associated Protein
CHC Clathrin Heavy Chain
CIPRES CyberInfrastructure for Phylogenetic RESearch
CLC Clathrin Light Chain
CME Clathrin-Mediated Endocytosis
COG Conserved Oligomeric Golgi complex
COPI COatomer Protein I
COPII COatomer Protein II

DNA DeoxyriboNucleic Acid

EE Early Endosome
EEA1 Early-Endosomal Autoantigen 1
ELW Expected Likelihood Weights
ER Endoplasmic Reticulum
ERGIC ER-Golgi Intermediate Compartment

FECA First Eukaryotic Common Ancestor
FHL Familial Hemophagocytic Lymphohistiocytosis

GAP GTPase Activating Factor
GCC Golgi Coiled-coil protein
GCP60 Golgi Complex-associated Protein of 60 kiloDaltons
GEF GTPase Exchange Factor
GFF3 General Feature Format version 3
GFP Green Fluorescence Protein
GM130 Golgi Matrix protein of 130 kiloDaltons

GMAP210 Golgi Microtubule-Associated Protein 210

GRAB GRIP-related Arf-binding

GRASP Golgi Reassembly and Stacking Protein

GRIP Golgin-97, RanBP2alpha, Imh1p and P230/golgin-245

GTP Guanosine TriPhosphate

HA Hemagglutinin Antigen

HMM Hidden Markov Model

HOPS Homotypic Fusion and Vacuole Protein Sorting

HSP High-scoring Segment Pair

HSP Hereditary Spastic Paraplegia

Inparalogues Paralogues originating subsequent to a given speciation event.

LE Late Endosome

LECA Last Eukaryotic Common Ancestor

logL Log Likelihood

MCMC Markov Chain Monte Carlo

MDH Malate DeHydrogenase

MDV Mitochondria-Derived Vesicle

ML Maximum Likelihood

Monophyly When a group of taxa (or sequences) are more closely related to each other than to any other relevant taxa, their relationship is monophyletic.

MTS Membrane Trafficking System

MUSCLE MUltiple Sequence Comparison by Log-Expectation

NCBI National Center for Biotechnology Information

NE Nuclear Envelope

NECC NeuroEndocrine Coiled-Coil protein

NPSN Novel Plant SyNtaxin

OPH Organelle Paralogy Hypothesis

Orthologues Copies of a gene originating through speciation events, with counterparts existing in different species.

Outparalogues Paralogues originating prior to a given speciation event.

Paralogues Copies of a gene originating through gene duplication, and coexisting in the same species.

Paraphyly A group of taxa (or sequences) are paraphyletic with respect to another group (a monophyletic group) when it is not monophyletic but shares a common ancestor with the second group.

PCR Polymerase Chain Reaction

PDI Protein Disulfide-Isomerase

PM Plasma Membrane

Posterior probability In Bayesian analysis, the posterior probability is the probability of the hypothesis (which may include a phylogenetic tree) given data (which may be an alignment of homologous genetic sequences). This is calculated as the product of the prior probability of the hypothesis and the likelihood of the data given the hypothesis.

RE Recycling Endosome

RELL Relative Estimated Log-Likelihood
RFP Red Fluorescence Protein

SAR Stramenopiles Alveolates and Rhizaria
SCOCO Short COiled-COil protein
SCYL1BP1 SCY1-like 1 binding protein 1
SM Sec1/Munc18 family protein
SNAP SyNaptosomal Associated Protein
SNAPC SyNaptosomal Associated Protein Carboxyl-terminal SNARE domain
SNAPN SyNaptosomal Associated Protein amiNo-terminal SNARE domain
SNARE Soluble *N*-ethylmaleimide sensitive factor Attachment protein REceptor
SPG Spastic Paraplegia Gene
Syn Syntaxin
SYP SYntaxin of Plants

TACK Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota
TBC Tre-2/Bub2/Cdc16
TGN *trans*-Golgi Network
TM9SF3 TransMembrane 9 SuperFamily member 3
TMD TransMembrane Domain
TMF TATA element modulatory factor
TRAPP TRAnsport Protein Particle complex
TSET [This is not an abbreviation. See Hirst *et al.* (2014).]

VAMP Vesicle Associated Membrane Protein
VTI1 Vacuolar protein sorting Ten Interacting protein 1

ZFPL1 Zinc Finger Protein-Like 1

Chapter 1

Introduction

1.1 The membrane trafficking system and the protein machinery behind it

A fundamental aspect of the biology of eukaryotic cells including human cells, plant cells, and the cells of numerous parasites is a complex system of internal membrane-bound organelles that populate the cytosol. A subset of these organelles undergo dynamic division and fusion processes, allowing various molecules to be transported among the organelles and exchanged between organelles and the cells environment. This is termed the membrane trafficking system, and it is the focus of much research in the field of cell biology as it is necessary for sustaining essentially every cellular process in eukaryotic cells (Figure 1.1).

The overall flow of material within the membrane trafficking system may be generally divided into the secretory pathway and the endocytic pathway. The secretory pathway is responsible for transporting proteins from their site of synthesis to the specific locations in the cell where they are required, or releasing them from the cell (Novick *et al.*, 1981). Integral membrane proteins and secreted proteins are synthesized by ribosomes on the membrane of the rough endoplasmic reticulum (ER) (Redman *et al.*, 1966). These proteins may then be transported in membrane-bound vesicles to the Golgi apparatus, which comprises a series of compartments where proteins may be modified by addition of sugar polymers. Subsequently proteins may be transported to the cell surface, possibly passing through endosomes on the way. Coated transport vesicles are involved in trafficking at both the ER-Golgi stage and the post-Golgi stage of the secretory pathway (Rothman and Fine, 1980). In many cells this secretory pathway is considered the default trafficking itinerary for proteins after synthesis (Rothman, 1987). In addition, proteins that function at specific organelles may be retained in these organelles rather than proceeding through to the cell surface.

The endocytic pathway diverges from the secretory pathway at the endosomal stage–specifically early endosomes, and proceeds through late endosomes to lysosomes (Helenius *et al.*, 1983; Schmid, 1988). Early endosomes are the delivery point for membrane and cargo internalized from the cell

surface via membrane-bound vesicles (Prescianotto-Baschong and Riezman, 2002). The next step in the endocytic pathway is the late endosomes, which typically move membrane and membrane proteins from their surface into their lumen via the formation of intralumenal vesicles (for this reason late endosomes are also termed multivesicular bodies) (Felder *et al.*, 1990). Late endosomes then deliver their contents to lysosomes (or vacuoles in the case of yeast cells and plant cells), which contain hydrolases for protein degradation (Futter *et al.*, 1996). In mammalian cells, mannose-6-phosphate receptors play an important role in diverting lysosomal enzymes from early endosomes to lysosomes after delivery from the Golgi to early endosomes (Griffiths *et al.*, 1988; Ludwig, 1991). In some cases, lysosomes may eventually fuse with the plasma membrane, releasing any remaining contents outside the cell (Andrews, 2000). The endocytic pathway is also responsible for digestion of cytosolic components collected in autophagosomes and of larger volumes of material gathered from the cell exterior in phagosomes.

The flow of traffic in both the secretory and endocytic pathways, and the control of protein location, may be understood essentially in terms of organelle maturation with selective recycling of membrane and proteins in the opposite direction to maturation. In the secretory pathway, proteins are delivered from the ER to *cis*-Golgi cisternae which then mature into *trans*-Golgi cisternae before proteins are exported to endosomes or the cell surface. While Golgi maturation was once a matter of debate, with some researchers arguing for stable cisternae (Orci *et al.*, 2000), the field of cell biology has now reached a consensus favouring the maturation model (Glick *et al.*, 1997; Matsuura-Tokita *et al.*, 2006; Glick and Nakano, 2009). Similarly, early endosomes mature into late endosomes which subsequently mature into lysosomes (Helenius *et al.*, 1983; Murphy, 1991; Rink *et al.*, 2005). Thus to sustain organelle maturation, maintaining the flow of cargo along the secretory and endocytic pathways, proteins are transported in vesicles which form at one organelle and fuse with another at an earlier stage in the secretory or endocytic pathways.

The mechanisms by which transport vesicles are formed and delivered to specific target membranes in the membrane trafficking system are therefore of central importance to understanding the membrane trafficking system and the cellular processes that it sustains. Discoveries of the protein machinery mediating these processes began in the 1980's, and resulted in several cell biologists being awarded the 2013 Nobel Prize in Physiology and Medicine (Bonifacino, 2014). Work to date on the molecular machinery of the membrane trafficking system spans 100s of proteins orchestrating a multitude of vesicle transport pathways in a variety of model eukaryotic cells, which necessarily makes 100s of primary sources relevant. For this reason, comprehensive review articles on this topic are referenced (in the following paragraphs), and more detailed primary literature will

be referenced in subsequent chapters where relevant. Work to date reveals that the processes of vesicle formation and of vesicle fusion often involve comparable (analogous or homologous) proteins playing similar roles but in distinct vesicle transport pathways (Figure 1.2) (see Bonifacino and Glick (2004) for review).

Vesicle formation begins with nucleation, which involves lateral aggregation of membrane proteins to be included in the nascent vesicle (Bonifacino and Glick, 2004). These proteins may include transmembrane receptors that bind to vesicle cargo in the lumen of the organelle. Peripheral membrane proteins including vesicle adaptors are also recruited to vesicle nucleation sites, and these bind to transmembrane cargo proteins and to proteins that form vesicle coats. Key vesicle adaptors are the homologous family of Adaptor Protein complexes, which are related to the Coatomer Protein I (COPI) (Schledzewski *et al.*, 1999) and TSET (Hirst *et al.*, 2014) complexes (see Dacks and Robinson (2017) for review). These protein complexes are discussed in much more detail in Chapters 2 and 3 of this thesis. Important vesicle coat proteins include subunits of the COPII complex which coats vesicles budding from ER membranes (Barlowe *et al.*, 1994), and clathrin heavy and light chains which coat vesicles budding from the *trans*-Golgi network, early endosomes, and plasma membrane (Pearse, 1976). The Retromer complex mediates formation of vesicles at late endosomes for recycling of proteins back to the *trans*-Golgi network (Seaman *et al.*, 1998; Haft *et al.*, 2000). Vesicle budding occurs as the aggregation of coat proteins deforms the membrane. Budding vesicles are separated from the donor organelle membrane by proteins that mediate membrane scission such as dynamin or related guanosine triphosphatase (GTPase) proteins (Hill *et al.*, 2001; Chi *et al.*, 2014).

Following formation vesicles may be transported along microtubules by the motor proteins dynein or kinesin (Schroer, 1988; Schroeder, 1990). Inbound vesicles are recognized by tether proteins extending from the vesicle's target membrane. One class of vesicle tether are the golgins which functions at the Golgi apparatus, and these are discussed in greater detail in Chapter 4. Other types of tethers include multi-subunit tethering complexes, such as the homotypic fusion and vacuole protein sorting (HOPS) complex which functions at late endosomes (Seals *et al.*, 2000), and the Exocyst complex which functions at the plasma membrane (TerBush *et al.*, 1996). Once tethered, vesicles are docked adjacent to the target membrane and subsequently undergo membrane fusion, releasing their contents to the lumen of the target membrane, and combining their membrane components with that of the target organelle. Proteins of the Soluble *N*-ethylmaleimide sensitive factor Attachment protein REceptor (SNARE) superfamily are essential for mediating membrane fusion, and are introduced in detail in Chapter 5 of this thesis.

In addition to the proteins mentioned above, a class of proteins central to the organization of many aspects of vesicle transport are various small GTPases. These proteins are pheripheral proteins that are recruited from the cytosol to fulfill their roles at specific membranes by GTP exchange factors (GEFs), and released from membranes by GTPase activating proteins (GAPs) (see Stenmark and Olkkonen (2001) and Gillingham and Munro (2007) for review). For example, the small GTPase Sar1 functions in membrane curvature and recruitment of COPII vesicle coat components at ER (Lee *et al.*, 2005; Barlowe *et al.*, 1993, 1994). Arf1 plays a comparable role in membrane deformation and recruitment of COPI vesicle coat components at Golgi membranes (Donaldson and Cassel, 1992; Beck *et al.*, 2008), as well as in recruiting clathrin (Stamnes and Rothman, 1993). Arl1, which is related to Sar1 and Arf1, is involved in attaching vesicle tethers at the *trans*-Golgi network (Setty *et al.*, 2003). In addition, proteins in the Rab family of small GTPases play key regulatory roles particularly in organizing membrane tethering and fusion at specific locations in the cell via interaction with effector proteins which include tethers (see Grosshans *et al.* (2006) for review). These include Rab1 which regulates the golgin tether protein p115 at the Golgi (Allan *et al.*, 2000), Rab5 which regulates the tether protein early-endosomal autoantigen 1 (EEA1) at early endosomes (Simonsen *et al.*, 1998), Rab7 which regulates the HOPs complex at late endosomes (Lürick *et al.*, 2017), and Rab11 which regulates the Exocyst tether complex at exocytic vesicles (Zhang *et al.*, 2004).

Research on the cell biology of the membrane trafficking system is important for understanding the mechanisms underlying diseases for two key reasons: Many congenital diseases result from defects in membrane trafficking, and membrane trafficking interfaces between human cells and pathogens or parasites. One disease caused by mutations in genes encoding vesicle formation proteins is Hereditary Spastic Paraplegia (HSP), this involves defects in trafficking from late endosomes toward the Golgi that result from mutations affecting adaptor complexes leading to accumulation of abnormal lysosomes (Hirst *et al.*, 2013b, 2015, 2018). Also, Hermansky-Pudlak Syndrome can result from defects in vesicle trafficking involved in biogenesis of lysosome-related organelles, and can lead to blindness (Bonifacino, 2004). Mutations affecting vesicle tethers such as those at the Golgi membranes can result in diseases such as skeletal dysplasia (Smits *et al.*, 2010). Mutations affecting various SNARE vesicle fusion proteins or their regulatory proteins have been implicated in a wide variety of diseases including Cerebral dysgenesis, Neuropathy, Ichthyosis, and Keratoderma (CEDNIK syndrome) (Sprecher *et al.*, 2005), Familial Hemophagocytic Lymphohistiocytosis (FHL) (zur Stadt, 2005; Feldmann *et al.*, 2003), renal dysfunction (Gissen *et al.*, 2004), and neurological disorders (Rohena *et al.*, 2013; Shen *et al.*, 2014).

Whether pathogens (including parasites) attack human cells from the outside or invade their interior, the membrane trafficking system is central to virulence and host defence mechanisms. Ubiquitous bacteria of the *Clostridium* genus produce neurotoxins causing the life-threatening conditions of botulism or lockjaw. These botulinum and tetanus neurotoxins have evolved to enter neurons and cleave SNARE proteins involved in neural transmission, resulting in paralysis (Jahn and Niemann, 1994; Simpson, 2004b). The toxin produced by *Vibrio cholerae*, the causative agent of cholera, must traverse several organelles in the membrane trafficking system of epithelial cells before it can alter their function (Lencer *et al.*, 1999). Intracellular pathogens, and parasites, such as the bacteria *Salmonella*, *Legionella*, or *Chlamydia* typically gain entry to human cells in phagosomes (Alonso and Portillo, 2004). The invasion process then continues as the pathogen secretes factors which may inhibit the maturation of the phagosome, and hence its digestion by the host cell (Shotland *et al.*, 2003; Horwitz, 1984; Shohdy *et al.*, 2005; Delevoye *et al.*, 2004).

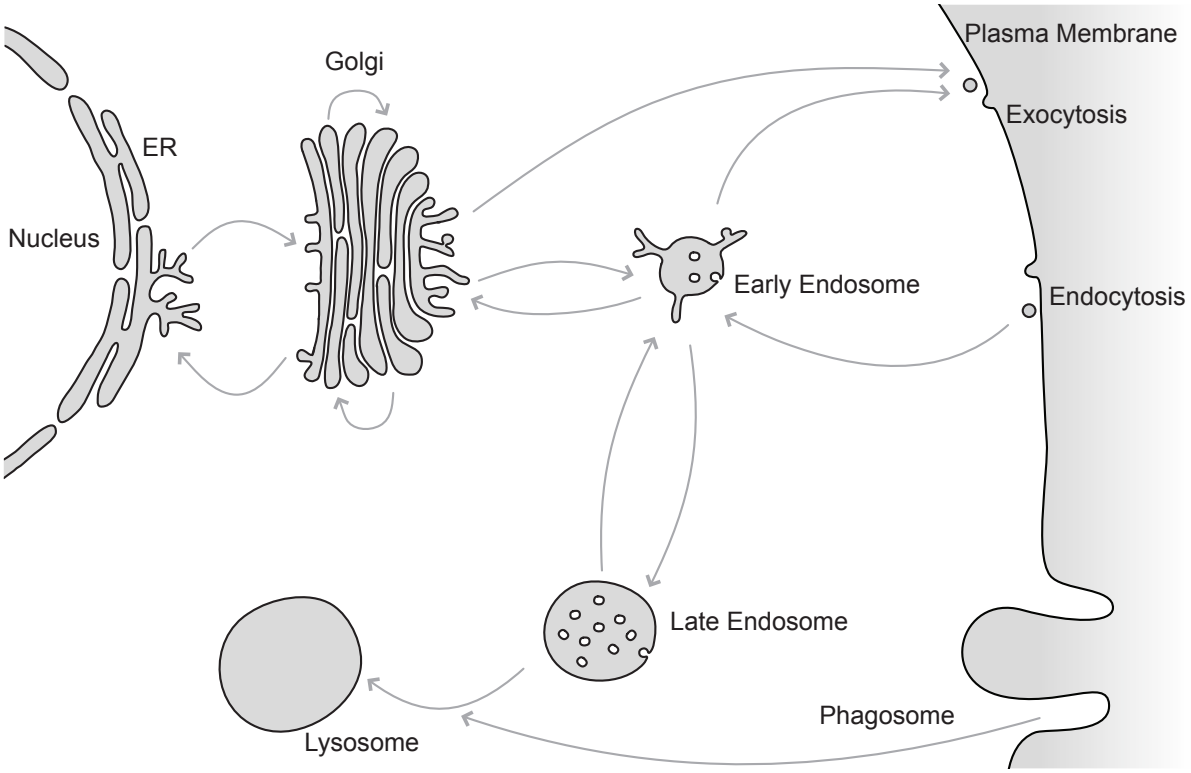


Figure 1.1: Schematic of the eukaryotic membrane trafficking system. This system comprises endomembrane organelles involved in membrane trafficking including the ER, Golgi, and endosomes, as well as the PM and the vesicle trafficking pathways that interconnect their membranes. Arrows indicate membrane trafficking pathways including those that occur via vesicle transport and organelle maturation. Here recycling endosomes, which exist in some species, are not distinguished from early endosomes or the *trans*-Golgi network (see section 5.2.2.2 for discussion of the issue of drawing relevant distinctions between endosomal organelles).

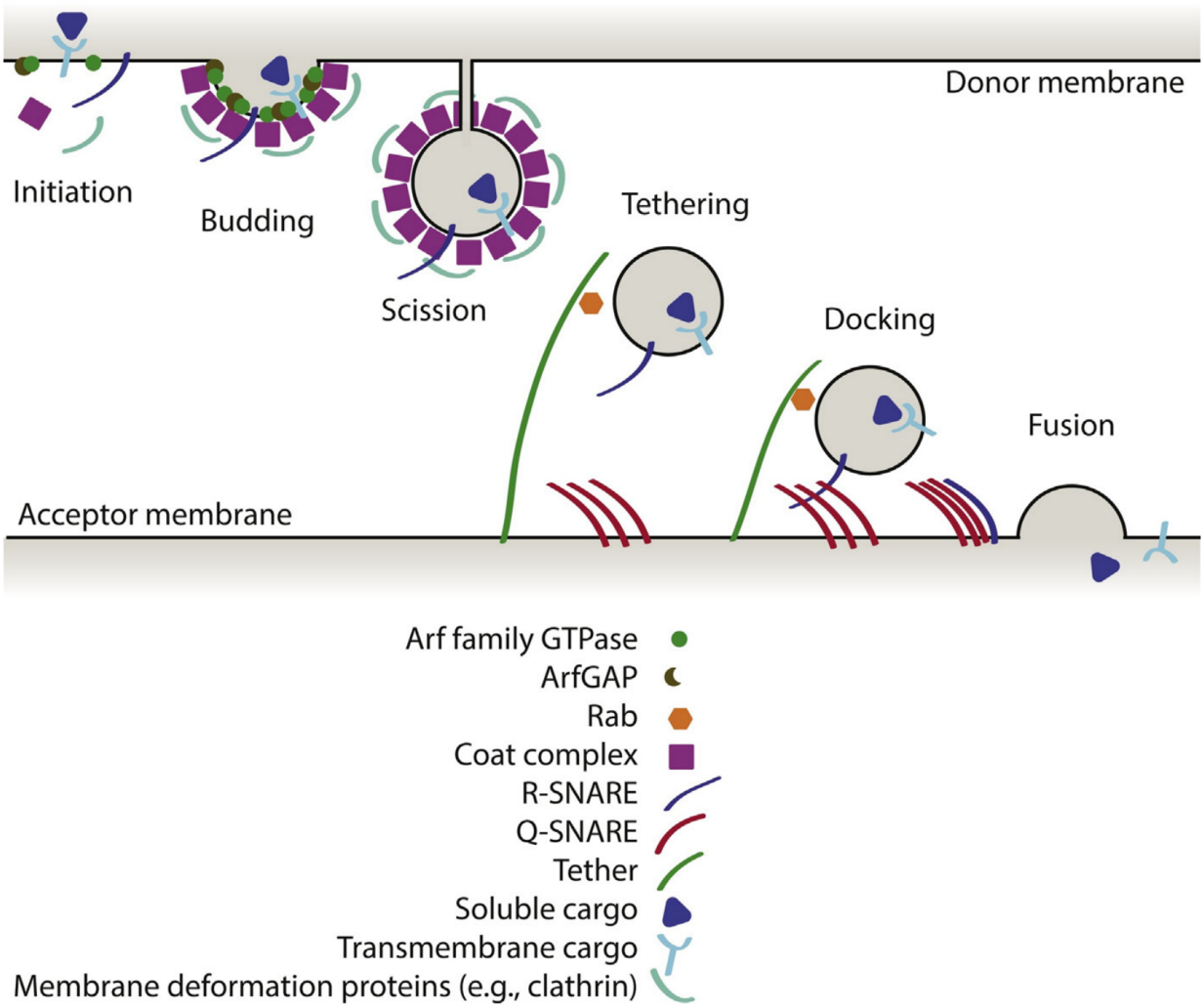


Figure 1.2: Vesicle trafficking in the eukaryotic membrane trafficking system. Various vesicle trafficking pathways proceed via similar mechanisms beginning with vesicle budding, continuing through vesicle scission, transport, and tethering, and concluding with membrane fusion. Note that Arf family GTPases and ArfGAP proteins are pheripheral membrane proteins and not integral membrane proteins, as might be suggested by their positioning in this figure. This figure reproduced from Barlow and Dacks (2018) Figure 1.

1.2 Overview of eukaryotic evolution

Extant eukaryotic species, their cells, and their membrane trafficking systems are results of evolution occurring over billions of years. The earliest fossil evidence of prokaryotic cells (cells without a nucleus, such as bacteria and archaea) is found in rocks dated to approximately 3.4 billion years ago (Wacey *et al.*, 2011). The earliest fossil evidence of eukaryotic life is dated later at approximately 1.6 billion years ago (Butterfield, 2015). The age of the last common ancestor of extant eukaryotes is estimated to be between 1.866 and 1.679 billion years (Betts *et al.*, 2018). Animals and land plants did not appear until much later at approximately 0.540 and 0.470 billion years ago, respectively (Morris, 1989; Kenrick *et al.*, 2012; Morris *et al.*, 2018).

The origin of eukaryotes was a major evolutionary transition, as the closest known prokaryotic relatives of eukaryotes was quite unlike extant eukaryotes, lacking features such as a complex endomembrane system. Analyses of metatranscriptomic sequence data collected from the deep sea suggest that the closest living relatives of eukaryotes are archaeal lineages collectively termed Asgard archaea (Zaremba-Niedzwiedzka *et al.*, 2017) (discussed further below). A recent preprint claims to report the first micrographs of an Asgard archaeon, and while it did suggest that this archaeon may have an unusual cellular morphology and ecology, it did not suggest existence of an endomembrane system in that species (Imachi *et al.*, 2019). The current paucity of information about the evolutionarily critical Asgard archaeal lineages stems from the tremendous difficulty of isolating and culturing these organisms from the deep sea. Indeed, Imachi *et al.* (2019) reported apparent doubling times of 14 to 25 days in liquid medium after enrichment in a custom-made bioreactor (Aoki *et al.*, 2014). Further methodological advances may allow culturing of such organisms under conditions that better simulate their natural environment. Nevertheless, other archaea such as *Igneococcus hospitalis* as well as planctomycete bacteria such as *Gemmata obscuriglobus* have been shown to possess endomembrane systems (Acehan *et al.*, 2014; Heimerl *et al.*, 2017). However, these do not rival the complexity of eukaryotic endomembrane systems. That is, they do not appear to contain more than one type of functionally distinct compartment, or organelle, within the system. Also, no evidence has been reported suggesting that they are homologous to eukaryotic endomembrane systems. Consistent with this, analyses of genomic sequences from archaeal lineages closely related to eukaryotes found very few direct counterparts of eukaryotic genes involved in endomembrane system organization (Klinger *et al.*, 2016b).

Since their origin, eukaryotes have diverged producing a large diversity of extant lineages, and the major lineages, or supergroups diverged early in eukaryotic evolution (Parfrey *et al.*, 2011). The fossil record indicates that distinct lineages of eukaryotes began to diversify over 1 billion

years ago, as fossils of multicellular red algae have been dated to approximately 1.15 billion years ago (Butterfield, 2015). Studies based on analysis of genetic sequences from a diversity of representative eukaryotic species consistently support classification of extant eukaryotes into four major lineages: Amorphea, Metamonada, Discoba, and Diaphoretickes (Adl *et al.*, 2012, 2018) (Figure 1.3). Amorphea includes two taxonomic supergroups: Amoebozoa, which consists primarily of single-celled amoebae such as *Dictyostelium discoideum* and *Entamoeba histolytica*, and Opisthokonta which includes both Holozoa (including Metazoa) and Holomycota (including Fungi). Metamonada includes only protists (single-celled eukaryotes) such as *Giardia intestinalis*, the causative agent of beaver fever (giardiasis), and *Trichomonas vaginalis* the causative agent of trichomoniasis, as well as *Monocercomonoides exilis*, the eukaryote with no mitochondria or related organelles (Karnkowska *et al.*, 2016). Discoba also includes only protists such as the soil-dwelling amoeba *Naegleria gruberi*, the photosynthetic protist *Euglena gracilis*, and *Trypanosoma brucei* the causative agent of African sleeping sickness (Hampl *et al.*, 2009). Finally, Diaphoretickes, like Amorphea, includes several taxonomic supergroups: Stramenopiles Alveolates and Rizaria (SAR) form a supergroup containing protists such as the ciliate *Tetrahymena thermophila* (an alveolate) and *Plasmodium falciparum* the causative agent of malaria (also an alveolate), but also multicellular brown algae such as *Macrocystis* spp. (stramenopiles) which constitute giant kelp forests (Steneck *et al.*, 2002; Burki *et al.*, 2008). Archaeplastida includes multicellular red algae and land plants (Embryophyta) as well as their respective single-celled relatives (Adl *et al.*, 2005). Diaphoretickes also contains the photosynthetic algal lineages Haptista, which includes the haptophyte *Emiliania huxleyi*, and Cryptista, which includes the cryptophyte *Cryptomonas* (Cavalier-Smith, 2003; Adl *et al.*, 2018). There is currently no strong consensus regarding which of these four main eukaryotic lineages (Amorphea, Metamonada, Discoba, or Diaphoretickes), or combination thereof, represents the earliest branch in the eukaryotic tree of life (see discussion).

One of the striking features of the eukaryotic tree of life, which follows from the early divergence of eukaryotes described above, is the independent origins of multiple multicellular eukaryotic lineages and of multiple parasitic eukaryotic lineages. Consistent with the independent origins of eukaryotes with these lifestyles, multicellular and parasitic lineages are in most cases closely related to one or more lineages of free-living protists that swim using flagella and ingest (phagocytose) other cells such as bacterial cells as a food source, representing the ancestral state of eukaryotic cells (Koumandou *et al.*, 2013). These protist lineages occupy deep branches in several taxonomic groups that include multicellular or parasitic lineages, having, for example, di-

verged after the divergence of the Amorphea and Diaphoretickes groups, but before diversification of animals and plants. Metazoans are related to protists such as *Syssomonas* sp. (Hehenberger *et al.*, 2017). Fungi including yeast are related to protists including *Fonticula alba* (Brown *et al.*, 2009). The parastic amoebozoan *Entamoeba histolytica* is related to the free-living *Mastigamoeba balamuthi* (Gill *et al.*, 2007; Kang *et al.*, 2017). The parasitic *Giardia intestinalis* is related to the free-living *Carpediemonas membranifera* (Takishita *et al.*, 2012). The parasitic kinetoplastid *Trypanosoma brucei* is related to the free-living kinetoplastid *Bodo saltans* (Simpson, 2004a). The giant kelp *Macrocystis* is closely related to the marine protist *Cafeteria roenbergensis* (both are stramenopiles) (Derelle *et al.*, 2016). Apicomplexan parasites including the deadly *Plasmodium falciparum* are related to ciliates including *Tetrahymena thermophila* (both are alveolates) (Fast *et al.*, 2002). The multicellular red algae are related to the flagellated phagocytic protists including *Rhodelphis marinus* (Gawryluk *et al.*, 2019). Land plants are related to the phagocytic protist *Mesostigma veridae* (Karol, 2001).

The diverse, independent origins of eukaryotes that are of particular interest to cell biologists, namely animals, fungi, plants, and parasites implies that understanding the similarities and differences between the cell biology of these different eukaryotes from an evolutionary perspective requires an evolutionary framework that extends back to the origin of eukaryotes over 1 billion years ago (Dacks and Doolittle, 2001). For example, assertions regarding homology of cellular features such as organelles or protein complexes in mammalian and parasite cells may be useful as they are suggestive of extensive similarity in structure and function. However, such assertions imply conservation of the respective features over more than 1 billion years of independent evolution. This parallel evolution often allows remarkable conservation, but also may result in dramatic differences between eukaryotes, even the loss of entire protein complexes or organelles, or convergent organelle evolution (see below). The value of an evolutionary framework in such cases is in providing justification for inferring whether homology, analogy, or lineage-specific novelty or loss might be more likely. This may allow cell biologists to select the most appropriate model systems, those that share a high degree of conservation of the cellular feature of interest. To distinguish between evolutionary possibilities with respect to any given cellular feature, the separate evolutionary paths must be traced. Thus the Last Eukaryotic Common Ancestor (LECA), the most recent common ancestor of extant eukaryotes, is of particular interest for understanding eukaryotic evolution, because this hypothetical organism provides a starting point from which to trace the parallel cellular evolution of all extant eukaryotic lineages.

Cellular features present in the LECA may be inferred based on broad conservation across the

eukaryotic supergroups, and such inferences have shown that the LECA was complex (Koumandou *et al.*, 2013). Comparative analyses suggest that in addition to a nucleus the LECA possessed flagella, peroxisomes, and mitochondria, although each of these has been subsequently lost in one or more lineages (Karnkowska *et al.*, 2016; Gabaldón *et al.*, 2016). All extant eukaryotic cells contain a system of non-endosymbiotic endomembrane organelles, which allows compartmentation of both internally synthesized molecules and ingested material (Figure 1.1). The membrane trafficking system has been most thoroughly investigated in model mammalian and yeast cells. However, features of the membrane trafficking system conserved among extant eukaryotes include the endoplasmic reticulum (ER), Golgi apparatus, endosomes, and lysosomes or related organelles, as well as the plasma membrane (PM). Of particular relevance to this thesis, the LECA also possessed a complex membrane trafficking system with organelles including an ER, Golgi apparatus, endosomes, lysosomes, and possibly others (Koumandou *et al.*, 2013).

The genetic basis for the origin of many complex autogenously-derived eukaryote-specific cellular features present in the LECA, including of its membrane trafficking system, appears to a large degree to be gene duplication. For example, both the nuclear pore, through which molecules pass between the nucleoplasm and cytoplasm, and the kinetochore, which attaches chromosomes to microtubules during cell division, are independent examples of very large protein complexes which appear to be the result of duplication and divergence of relatively few basic protein subunits (Obado *et al.*, 2016; Field and Rout, 2019; Tromer *et al.*, 2019). The protein constituents of the intraflagellar transport system that maintains the eukaryotic flagellum also appear to have originated through gene duplications (van Dam *et al.*, 2013). A considerable number of eukaryote-specific protein folds may have originated prior to the LECA as well (Kauko and Lehto, 2018).

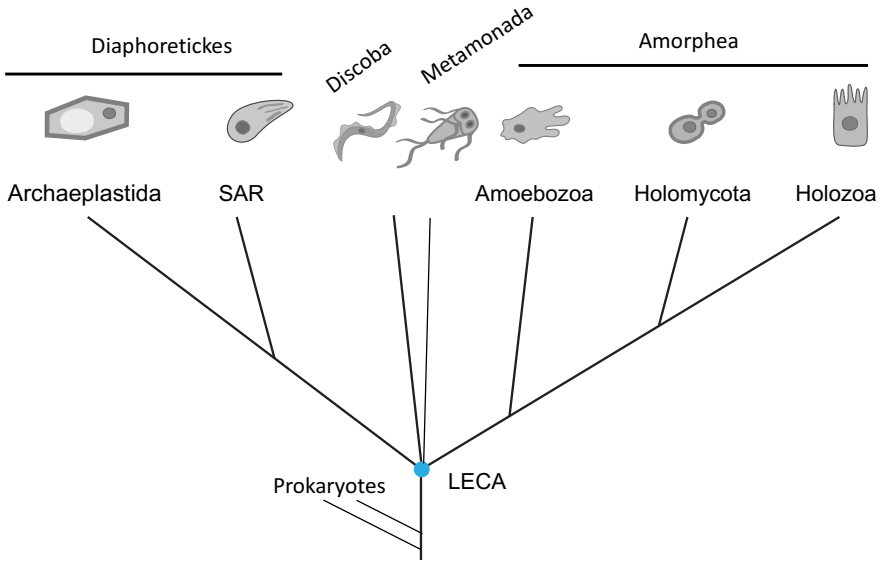


Figure 1.3: Phylogenetic relationships between major taxonomic groups in the eukaryotic tree of life. Based on Adl *et al.* (2018). Archaeplastida includes plants, green algae, and red algae. SAR includes Stramenopiles, Alveolates, and Rhizaria. Discoba includes Trypanosomes, *Euglena*, and *Naegleria*. Metamonada includes *Giardia* and *Monocercomonoides*. *Amoebozoa* includes *Dictyostelium*, *Mastigamoeba*, and *Entamoeba*. Holomycota includes Fungi and their single-celled relatives. Holozoa includes animals and single-celled relatives such as choanoflagellates. This figure was adapted from Barlow and Dacks (2018) Figure 2A.

1.3 The role of comparative evolutionary work in the field of cell biology

The diversity of eukaryotes and their membrane trafficking systems is vast and perhaps even overwhelming to cell biologists interested in elucidating the fundamentals of cell biology. Much remains unknown about the cell biology of popular model systems, let alone their many distant relatives. It is worth considering whether investigating aspects of the cell biology of these diverse eukaryotes and working towards synthesizing this into more detailed evolutionary theories, as I attempt to do in this thesis, is a valuable endeavor in the field of cell biology. The vast majority of published cell biological research is focused on opisthokonts including mammals and fungi (Lynch *et al.*, 2014), and important reasons for this include that mammalian and fungal model systems are highly tractable. Also, Sean Munro, a leading cell biologist, suggests another reason that this is a trend that is likely to continue into the future as well: "It is unlikely that the planet's tax payers will be willing to pay for enough cell biologists to investigate every last intriguing invertebrate or bizarre bikont, and thus future work is likely to focus on particular key cells [*sic*] types, especially those found in tax payers themselves." (Munro, 2013). For these reasons, I will briefly consider the role that comparative evolutionary work has played in the field of cell biology, and the potential for further phylogenetic analyses to advance this field.

Historically, comparative (evolutionary) study of diverse eukaryotes including protists has been considered relevant to the field of cell biology. the field of cell biology is concerned with description, scientific investigation, and explanation of eukaryotic biology at the cellular level of organization. Regarding the definition of cell biology, it is important to note that the term biology refers not only to study of extant living organisms but to study of their origins and evolution as well (Hine, 2019). So, cell biology includes study of the evolution of cells, as well as their development, physiology, and disease states. Pioneers in the field of cell biology recognized that the evolutionary origins of the organelles and proteins that they studied promised deeper levels of understanding. For example, Gunter Blobel sought to understand his Nobel prize-winning work on signal peptides in evolutionary terms, making several proposals regarding "the phylogeny of biological membranes, of cells, and of cellular compartments" (Blobel, 1980). Consideration of endomembrane origins provided an evolutionary context for his findings. Also, Christian de Duve discovered peroxisomes and was intensely interested in the evolution of peroxisomes, choosing to present on this topic at the first peroxisome conference (de Duve, 1982). Protists including the free-living ciliate *Tetrahymena pyroformis*, have been useful in revealing conserved cell biological and molecular mechanisms. For example, dynein a molecular motor involved in moving vesicles along microtubules was first described in a study that isolated it from the ciliate *Tetrahymena py-*

roformis (Gibbons and Rowe, 1965). Moreover, the concluding sentence of the article reporting Christian de Duve's Nobel prize-winning discovery of the lysosome justifies his use of *Tetrahymena pyroformis* for comparison with mammalian cells: "The finding that similar particles are present in a protozoan is of great interest and suggests that they have a long evolutionary history and presumably perform some vital function." (Baudhuin *et al.*, 1965).

While evolutionary studies of free-living protists are useful for developing theoretical context of many topics in cell biology, perhaps the strongest rationale for developing unifying evolutionary theories in this field is to address the problem of pathogenic and parasitic protists. As detailed above, protists have evolved parasitic lifecycles independently many times in diverse eukaryotic lineages (also see Walker *et al.* (2011) for review). Many of these parasites infect humans and cause numerous deaths each year. For example, in 2017 *Plasmodium* spp. caused an estimated 290 million cases of malaria which resulted in an estimated 435,000 deaths worldwide (WHO, 2018). The membrane trafficking system in particular is relevant for understanding virulence mechanisms. For example, *Trypanosoma brucei* the causative agent of African Sleeping Sickness rapidly cycles its plasma membrane through endosomes to evade the immune system (Field and Carrington, 2009). Several pathogenic protists including the trypanosomatids are considered neglected diseases, which disproportionately affect populations in developing countries (WHO, 2010). Given the need to address the cell biology of diverse pathogenic and parasitic protists and the resource-intensive nature of experimental cell biology, it is certainly worth considering what progress may be made by computational analysis of genetic sequence data.

In principle, phylogenetic analysis of genetic sequence data addresses hypotheses which are distinct from those regarding cellular physiology. Phylogenetic hypotheses in the field of cell biology describe the evolutionary relationships between cellular features (proteins, protein complexes, or organelles) such as 1) whether two features are homologous and 2) which homologous features share the most recent common ancestor. Genetic sequences contain evidence to distinguish between alternative hypotheses regarding the evolutionary relationships between genes based on assumptions about what evolutionary changes (mutations) may be more likely to occur (Zuckerkandl and Pauling, 1965). In fact, gene sequences provide a particularly information-rich form of evidence from which to evaluate alternative phylogenetic trees (Fitch and Margoliash, 1967). In contrast to phylogenetic hypotheses, hypotheses that describe aspects of the physiology, development, or disease states of cells can be directly tested via experimentation. Experimental methods elucidate the structure and causal relationships between cellular features in extant cells. These types of hypotheses are separate, because the structure and/or function of any given cellular fea-

ture cannot be fully equated with its evolutionary history. Such a direct equivalence is precluded by the possibility of change in protein function over time, and the occurrence of genes with multiple different functions (Todd *et al.*, 1999; Ponting, 2001).

Although phylogenetic hypotheses may be distinguished from other hypotheses in cell biology, they are highly intertwined. This can make the role of evolutionary analysis in cell biology unclear. The structure and functional roles of cellular features such as proteins and organelles are results of evolutionary processes, and so evolutionary ancestry investigated by phylogenetic analysis is closely connected to the function of these features. Accordingly, many bioinformatic approaches have been developed which attempt to predict protein function based on evolutionary analysis. For example, phylogenetic profiling is a method that attempts to identify functional relationships between proteins based on correlation of gene loss events (Ranea *et al.*, 2007; Carvalho-Santos *et al.*, 2010; Burns *et al.*, 2018). And, genes which are orthologous (derived from the same gene that existed in a common ancestor of two organisms), are often predicted to be functionally similar (Tatusov, 2000). This raises the question of what methodological approaches are needed to advance beyond prediction to test phylogenetic hypotheses. Indeed, in practice, it is common and often advantageous for cell biological studies to proceed through the following sequence of steps: phylogenetic/computational analysis, hypotheses formulation, and finally hypothesis testing via experimental methods. In other words, computational methods are used to generate hypotheses and predictions, while wet bench work is used to test those predictions. Thus it may seem that all computational methods, including phylogenetic methods are only applicable to hypothesis formulation, or that the intertwined phylogenetic and experimental methods are applicable to an overlapping set of hypotheses.

Phylogenetic methods are required to test phylogenetic hypotheses. Due to the occurrence of convergent evolution, gene/protein function alone is not conclusive evidence of evolutionary history. Moreover, the descriptive nature of phylogenetic analysis does not preclude hypothesis testing (Dunn and Munro, 2016). In fact, methods for explicitly testing phylogenetic hypotheses have been established for several decades (Huelsenbeck, 1997). However, the relevance of phylogenetic hypotheses, including those considered in subsequent chapters of this thesis, to the cell biology of extant eukaryotes is predicated on the idea that protein function is conserved over time, and that orthologous proteins are likely to retain similar functions. For evaluating these functional predictions, it is necessary to take a comparative approach to experimental results among a sampling of organisms representing the most distantly related lineages deriving from a common ancestor of interest. Therefore, experimental observation is required to test hypotheses regarding

the ancestral function of phylogenetically related genes/proteins. For these reasons, it may be most useful to view phylogenetic and experimental approaches to cell biology as interdependent, but individually sufficient to make important advances by testing hypotheses relevant to the evolutionary history of eukaryotic cells. Such distinctions become relevant when considering hypotheses, such as the Organelle Paralogy Hypothesis (Dacks and Field, 2007) (discussed below) which describe a type of evolutionary scenario (or mechanism) (Mast *et al.*, 2014) involving change of gene/protein function over time.

Important achievements of phylogenetic studies in cell biology include testing endosymbiotic hypotheses for the origins of organelles including mitochondria and chloroplasts. Lynn Margulis was a key proponent of endosymbiotic hypotheses for organelle origins, renewing interest in this topic in the late 1960s (Sagan, 1967). At the time, she did not have a conclusive test of her hypothesis that mitochondria are derived from alphaproteobacteria. It was later argued that molecular phylogenetic evidence could test the two alternative hypotheses (Gray and Doolittle, 1982; Gray, 1992). Phylogenetic analysis to test this hypothesis was first performed by Gray and Cedergren (1984), and subsequent work has confirmed this (see Roger *et al.* (2017) for review). Based on apparent absence of mitochondria in several protist eukaryotes, and phylogenetic results showing that these diverged at early timepoints in eukaryotic evolution, several eukaryotic lineages were thought to have diverged from other eukaryotes prior to the mitochondrial endsymbiosis (Cavalier-Smith, 1987). Phylogenetic advances in analysis were instrumental in eventually revealing that those tree topologies resulted from artefactual clustering of divergent branches in the phylogenies of eukaryotes (Philippe and Germot, 2000), implying that the LECA possessed mitochondria. This was important for establishing the relationship of mitochondria to divergent mitochondria-like organelles found in parasitic protist lineages including *Giardia* and *Entamoeba* (Van Der Giezen, 2009).

The possible endosymbotic origin of peroxisomes is also a question that has been addressed using phylogenetic methods. Peroxisomes contain numerous enzymes with functions including reduction of reactive oxygen species, which is important for aerobic organisms (de Duve and Baudhuin, 1966). Christian de Duve first proposed an autogenous (non-endosymbiotic) origin of peroxisomes as being derived from the endoplasmic reticulum (de Duve, 1969), and also later suggested that an endosymbiotic origin is possible (de Duve, 1982). Observations that may be consistent with an endosymbiotic origin of peroxisomes include the existence of dedicated protein import machinery, analogous to the mitochondrial import machinery, that translocates proteins from the cytosol to the peroxisomal lumen (Fujiki and Lazarow, 1985). Also, although peroxisomes can be

generated from endoplasmic reticulum membranes (Waterham *et al.*, 1993), peroxisomes undergo a growth and division cycle analogous to enosymbiotic organelles (Lazarow and Fujiki, 1985).

A challenge to the application of phylogenetic methods to investigating the origins of peroxisomes is that, unlike mitochondria, peroxisomes lack organellar genomes. However, several peroxisome-specific proteins are homologous to non-peroxisomal proteins, providing potential for informative genetic sequences for phylogenetic investigation of peroxisome origins. For example, Pex1 and Pex6 form peroxisome-specific adenosine triphosphatases (ATPases) Associated with diverse cellular Activities (AAA) ATPases. These are homologous to the endoplasmic reticulum-associated AAA ATPase Cell Division Control (CDC) 48 protein, and also homologous to prokaryotic AAA ATPases (Beyer, 1997; Rabinovich *et al.*, 2002). A comprehensive evolutionary analysis of peroxisomal proteins investigated whether peroxisomal proteins overall are more closely related to prokaryotic or eukaryotic homologues, and found that a majority of classifiable sequences are eukaryote-derived but a considerable proportion are prokaryote-derived (Gabaldón *et al.*, 2006). However, differences in functional roles of these different protein sets strongly suggested an autogenous origin of peroxisomes, as the eukaryote-derived proteins are those that are involved in organizing the biogenesis of peroxisomes (Gabaldón *et al.*, 2006), and similar analyses arrived at the same conclusion (Schlüter *et al.*, 2006). Thus, phylogenetic analysis played an essential role in testing alternative hypotheses regarding the origin of peroxisomes.

1.4 Debate regarding the relative timing of organelle origins in eukaryogenesis

While phylogenetic studies have been successful in establishing whether organelles including mitochondria and peroxisomes are endosymbiotic or not (as discussed above), currently debated questions include the relative timing of the origins of the mitochondria and the autogenous organelles in the membrane trafficking system. In general, hypotheses regarding this question may be categorized as mitochondria-early (in which the mitochondrial endosymbiosis pre-dates the origin of the membrane trafficking system) and mitochondria-late (the reverse order of events) (Roger *et al.*, 2017). These two alternatives have potentially quite different implications for early membrane trafficking evolution. However, the relative timing of the origins of the various organelles in the membrane trafficking system may be further considered as a somewhat separate question, and several contrasting hypotheses have been proposed regarding this as well. These scenarios are relevant to interpretation of results in Chapter 5 of this thesis.

The separate evolutionary origins of the initial mitochondrial endobiont and of the host organism are relevant for evaluating alternative hypotheses for understanding the potential nature

and timing of their initial interaction. Evidence to date consistently suggests a close relationship between mitochondria and alphaproteobacteria. Considerable phylogenetic evidence initially suggested that mitochondria are specifically related to obligate intracellular parasitic alphaproteobacteria of the order *Rickettsiales*, which represent early branches in the tree of alphaproteobacteria (Sicheritz-Ponten *et al.*, 1998; Fitzpatrick *et al.*, 2006; Wang and Wu, 2015; Munoz-Gomez *et al.*, 2019). This suggested a scenario whereby mitochondria originated from *Rickettsia*-like parasites which invaded a pre-eukaryotic cell (whether complex or prokaryote-like) and populated its cytosol in a manner similar to extant *Rickettsiales* and subsequently transitioned from a parasitic to endosymbiotic lifestyle, as argued by several authors (Davidov and Jurkevitch, 2009; Wang and Wu, 2014; White *et al.*, 2018). However, the relationship between mitochondria and *Rickettsiales* may be a methodological artefact resulting from long-branch attraction and convergence in the relative composition of nucleotides in the sequences (Rodríguez-Ezpeleta and Embley, 2012; Roger *et al.*, 2017). Moreover, a recent study that addressed these concerns and included sequences from novel alphaproteobacterial lineages strongly suggested that mitochondria are an outgroup to extant alphaproteobacterial lineages rather than being related specifically to *Rickettsiales* or other parasitic lineages (Martijn *et al.*, 2018). Thus, whether the ancestor of mitochondria was a parasitic bacterium is currently unclear. Moreover, an origin of mitochondria from outside the *Rickettsiales* is perhaps also more consistent with the absence in *Rickettsiales* of inner membrane invaginations homologous to mitochondrial cristae which were likely in the ancestral alphaproteobacteria (Munoz-Gomez *et al.*, 2017). However, the propensity for convergent evolution of this morphological trait is not completely understood.

The host of the original mitochondrial endobiont was almost certainly an archaeon, but its cellular structure and lifestyle are also unclear. Phylogenetic analysis has shown that key eukaryotic genes involved in house-keeping processes such as DNA replication are of archaeal origin (Cox *et al.*, 2008). In fact, it has been shown that eukaryotes are closely related to a clade of archaeal lineages including Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) (Guy and Ettema, 2011; Williams *et al.*, 2012). More recently, phylogenetic analysis has shown that a sister group to the TACK archaea are the closest living relatives of eukaryotes (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017; Spang *et al.*, 2017). This group is now termed the Asgard superphylum, and is represented mainly by metatranscriptomic sequences from the deep sea. Consistent with a close relationship to eukaryotes, Asgard archaea are predicted to possess several genes that are shared only with eukaryotes (Klinger *et al.*, 2016b; Zaremba-Niedzwiedzka *et al.*, 2017). This suggests some potential for limited membrane trafficking capability in the archaeal

ancestor of eukaryotes. Whether any Asgard archaea possess endomembranes or are capable of membrane trafficking is currently unknown, however a recent preprint claims to report the first micrographs of an Asgard archaeon and these reveal no internal membranes in this organism (Imachi *et al.*, 2019).

Current arguments for an early origin of mitochondria, predating complex eukaryotic features such as the membrane trafficking system are based primarily on the idea that metabolic advantages provided by mitochondria would have been essential for sustaining such complex eukaryotic features (Lane and Martin, 2010; Martin *et al.*, 2017). Indeed, in typical eukaryotes mitochondria play an important role in ATP production, and organization of the membrane trafficking system of extant eukaryotes is highly dependent on energy derived via ATP hydrolysis by various ATPases. For this reason mitochondria-early hypotheses fit best with the view that the relationship between the host and the mitochondrial endobiont was metabolically symbiotic from a very early stage (Martin and Müller, 1998). Also, it is argued that the initial host of mitochondria would not have needed to be phagocytic, considering several examples of prokaryotes with endosymbionts (Martin *et al.*, 2017).

A relatively early origin of eukaryotic complexity or at least the capacity for phagocytosis was the first possibility to be suggested, as it provides a clear means by which the mitochondria could be initially internalized (Sagan, 1967; de Duve, 1969). Arguments for a late origin of mitochondria are also based on apparent dispensability of aerobic ATP production by mitochondria, and on phylogenetic evidence for the relative timing of events in eukaryogenesis. As noted by Hampl *et al.* (2019), mitochondria are not absolutely required for sustaining eukaryotic complexity, as exemplified by the metamonad *Monocercomonoides exilis* which is entirely lacks mitochondrial organelles (Karnkowska *et al.*, 2016, 2019). Phylogenetic evidence for a late mitochondrial acquisition includes the finding by Pittis and Gabaldón (2016) that eukaryotic proteins of alphaproteobacterial origin that function at mitochondria appear to have been acquired by eukaryotes at a later timepoint than other bacteria-derived genes. Also, a recent study used fossil evidence and molecular clock analysis to estimate the timing of divergence of prokaryotic and eukaryotic lineages, and indicated that alphaproteobacteria (to which mitochondria are related) originated at a relatively late time point compared to the timing of eukaryotic diversification, most consistent with a relatively late origin of mitochondria (Betts *et al.*, 2018).

Whether the mitochondrial acquisition pre-dated or post-dated the origin of eukaryote-specific autogenous organelles, various possibilities exist for the relative timing of the origin of the different autogenous organelles which were present in the LECA. These include an ER, Golgi, and

endosomes in addition to phagosomes (Koumandou *et al.*, 2013). Previously proposed hypotheses with specific predictions regarding the relative timing of the origin of these membrane trafficking organelles may be compared whether they are intended to fit mitochondria-early or mitochondria-late scenarios, because none involve derivation of organelles from both host and endosymbiont independently.

First, de Duve and Wattiaux (1966) proposed a scenario of autogenous endomembrane evolution beginning with an initial undifferentiated organelle or invagination with both secretory and absorptive functions. However, the wording used by de Duve and Wattiaux (1966) is somewhat ambiguous, not seeming to imply a particular order in which the ER, Golgi, and endosomes/lysosomes differentiated from this original organelle, and possibly implying organellar differentiation preceding the capacity for membrane trafficking. Blobel (1980) also proposes an origin of endomembranes via invagination of an undifferentiated "pluripotent" plasma membrane. But, instead of an initial combined secretory and absorptive function, Blobel (1980) proposed that the ER (and nuclear envelope) was formed first, and further speculated that the lysosomes/endosomes and Golgi were derived secondarily from either the ER or the plasma membrane. Several similar proposals are made, which posit an undifferentiated ER/Golgi-like organelle as the first membrane trafficking organelle, but are ambiguous regarding the source and timing of endosome origin (Jékely, 2003, 2007; Mironov *et al.*, 2007; Jékely, 2008). Christian de Duve later updated his hypothesis, proposing less ambiguously that the Golgi differentiated from the ER after differentiation of the ER (ER/Golgi) from an initial undifferentiated organelle (de Duve, 2007). Cavalier-Smith (2009) described a hypothesis similar to that of de Duve and Wattiaux (1966) and de Duve (2007) involving an initial organelle with both secretory and absorptive function, although in this case this primordial organelle is explicitly a phagosome. However, Cavalier-Smith (2009) clearly specifies that a primordial ER/Golgi organelle first differentiated from endosomes, and then the ER and Golgi differentiated. Finally, Gould *et al.* (2016) suggest that the first endomembrane organelle was not derived from the plasma membrane, but instead from mitochondria-derived vesicles. Nevertheless, Gould *et al.* (2016) explicitly propose that all endomembrane organelles are derived from a single primordial organelle composed of fused mitochondria-derived vesicles. While Gould *et al.* (2016) suggest that an ER-like organelle originated first, they are ambiguous regarding the timing of endosome origin in relation to the ER and Golgi. Thus, to date the most detailed hypothesis for the evolutionary relationships between the ER, Golgi, and endosomes is that proposed by Cavalier-Smith (2009). Which of these hypotheses may be best supported by molecular evolutonary data remains to be determined.

1.5 Organelle paralogy hypotheses for evolution of the membrane trafficking system

Explanation of the origin of the eukaryotic membrane trafficking system and its constituent organelles and transport pathways in terms of a series of differentiation events would ideally be based on some form of systematic evidence. Phylogenetic evidence was effective in elucidating the evolution of peroxisomes based on analysis of peroxisome-specific proteins (as discussed above) (Gabaldón *et al.*, 2006). Also, it is clear that individual origins of proteins or protein complexes whose functions are integral and essential to the maintenance of specific organelles or transport pathways, and which have orthologues with conserved function across eukaryotes, played some important role in the origin of the respective organelles and transport pathways. For example, paralogy between vesicle coat/adaptor complexes specific to different organelles was recognized and suggested to be involved in the evolution of organelles (Schledzewski *et al.*, 1999; Cavalier-Smith, 2002). This raises the question of whether a phylogenetic approach might be applied to distinguish between scenarios of membrane trafficking organelle evolution.

A breakthrough in this area of research was the proposal by Dacks and Field (2007) that the evolutionary history of organelles in the membrane trafficking system may be evident from the evolution of the families of membrane trafficking proteins with members which are specific to those organelles. They reasoned that, when resolved phylogenetic trees describing the relationships between members of such protein families are constructed, internal branches of the trees should correspond not only to the relationships between genes/proteins, but also between the organelles at which the particular proteins act. If this is the case, Dacks and Field (2007) proposed, nodes connecting the branches should correspond to intermediate ancestral organelles that existed prior to the LECA (Figure 1.4) (Dacks and Field, 2007; Dacks *et al.*, 2008; Mast *et al.*, 2014). This mechanism of organelle evolution is now referred to as the Organelle Paralogy Hypothesis (OPH) (Schlacht *et al.*, 2014; Mast *et al.*, 2014).

The coordinated evolution of proteins and organelles proposed by the OPH implies that duplications of proteins and corresponding organelle differentiations would generally both occur before any subsequent duplication or organelle differentiation events. However, more precise constraints on the timing of these events are not specified (Dacks and Field, 2007). The OPH would be consistent with scenarios involving rapid duplication of numerous membrane trafficking proteins via whole genome duplication and rapid distinction between sub-populations of organelles. The OPH would also be consistent with scenarios of gradual, progressive gene duplication events and correspondingly slow differentiation of organelles. However, Dacks and Field (2007) do imply that gene duplications would initiate the process of organelle differentiation, meaning that gene duplications

would precede corresponding organelle differentiation events to some extent.

The OPH suggests a potentially feasible means by which evidence may be found to distinguish between alternative hypotheses for the origin of the organelles in the membrane trafficking system. What is required for this is a set of phylogenetic trees of protein families with members that are specific to different organelles (*e.g.*, the ER, Golgi, endosomes, and PM). Increasingly, relevant sequence data is available from genomes of eukaryotes across the eukaryotic tree of life, and methods for systematic phylogenetic analysis of these sequences are established. A general prediction of the OPH is that the phylogenies of an essential subset of membrane trafficking protein families will be in agreement (have consistent topologies) regarding the respective organelle-specificity of family members (which would be the case if these proteins do not frequently acquire functions at new organelles). The main prediction regarding cellular physiology is that the same key subset of membrane trafficking protein family members share the same organelle specificity/function across eukaryotic diversity, as is generally observed (Klinger *et al.*, 2016a). A corollary of this functional prediction is that these same proteins are defining features of organelles because their conserved functions are integral to encoding organelle-specificity of protein machinery that collectively mediates different (vesicle) transport pathways. Specific hypotheses nested within an OPH framework may describe particular scenarios of paralogous organelle evolution on the basis of both the phylogenies of membrane trafficking proteins and the (conserved) functions of the family members in diverse eukaryotes.

While the OPH is perhaps the only such framework currently put forward to explain the origin of endomembrane organelles it represents only a subset of the possible scenarios, and no organelle paralogy hypotheses have yet been proposed to describe the pre-LECA origin and evolution of organelles in detail. A key assumption of the OPH is that endomembrane organelle evolution can be described by a bifurcating tree. This excludes evolutionary events such as organelle mergence, and origins of multiple different organelles from a single progenitor (an organelle that does not change in its function after differentiation of a new organelle). Assumptions underlying the OPH are considered in more detail in Chapter 5. Despite advances in resolving the phylogenetic trees of the Adaptor protein complex subunits and of Rab GTPases (Elias *et al.*, 2012; Hirst *et al.*, 2014), the details of pre-LECA autogenous organelle evolution have evaded inference from protein phylogenies. Indeed, it has been suggested that the evolution of protein families may be more an indicator of the evolution of membrane trafficking pathways than of organelles in the membrane trafficking system (Elias *et al.*, 2012), implying only a loose connection between protein evolution and organelle evolution.

The alternative hypothesis to the OPH is that organelles originate through neofunctionalization of membrane trafficking proteins. In neofunctionalization scenarios, some gene/protein paralogues retain ancestral function while others acquire very different functions (*i.e.*, are neofunctionalized) either on the same organelle or at different organelles. Neofunctionalization is incompatible with attempts to infer early evolution of the endomembrane system from the topology of membrane trafficking protein phylogenies. This is because the phylogenetic affinities of proteins would be unrelated to their organelle specificities. Using species as an analogy for organelles, attempting to infer organelle evolution from the phylogeny of proteins that had been neofunctionalized would be analogous to attempting to infer the relationship between species based only on phylogenies of genes that had been horizontally transferred between their genomes. For this reason, neofunctionalization hypotheses for organelle origin cannot be tested based on protein phylogenies as the OPH can. The possibility of co-option, or retargeting, of proteins to different organelles through the course of early eukaryotic evolution has previously been identified as particularly problematic for reconstructing organelle evolution based on protein targeting/function observed in extant eukaryotes (Gabaldón and Pittis, 2015). Examples of neofunctionalization of Rab proteins have been predicted in evolutionary analyses (Petrželková and Eliáš, 2014), and a dominant role for neofunctionalization in the evolution of the Rab protein family has been argued previously, at least in the case of lineage-specific paralogous expansion (Diekmann *et al.*, 2011). Cases of neofunctionalization of membrane trafficking protein paralogues strongest form of evidence against applicability of the OPH to organelle evolution, because if this turns out to be the dominant mode of protein evolution, then this would favour neofunctionalization over the OPH.

While this caveat regarding neofunctionalization applies generally to any interpretation based on the OPH, the OPH posits not that every membrane trafficking protein family will be highly informative, but that some combination of key membrane trafficking proteins will be particularly informative for understanding evolution of organelles in the membrane trafficking system (Dacks and Field, 2007). Also, theoretical work using minimal biophysical models supports the hypothesis that duplication and specialization of vesicle formation and fusion machinery alone is sufficient for distinct organelles to originate (Ramadas and Thattai, 2013). Thus, whether the OPH allows strong and detailed inferences regarding organelle evolution to be made remains an open question.

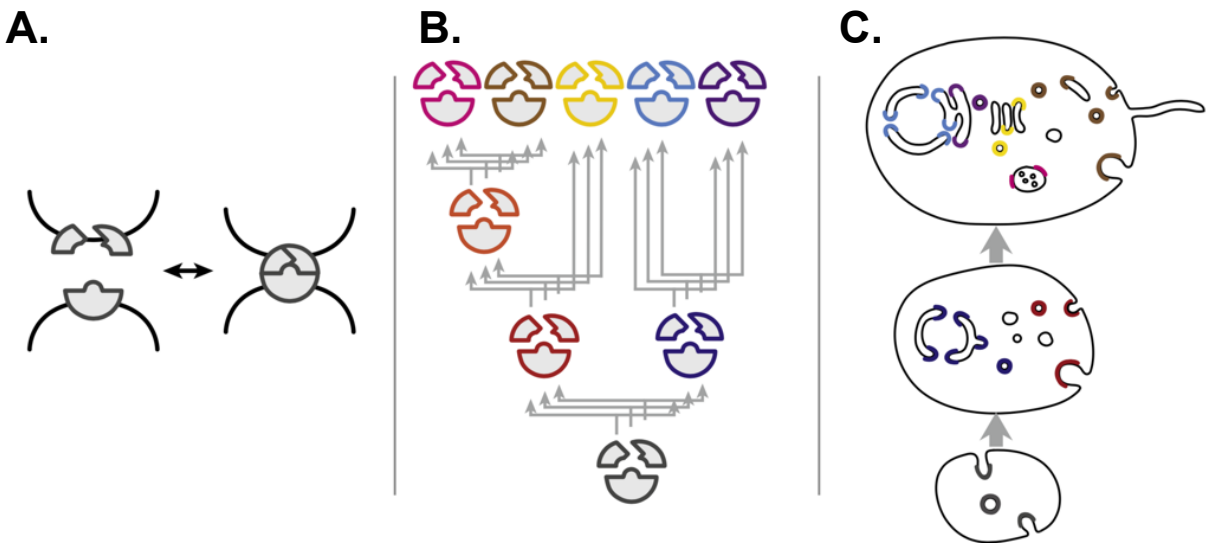


Figure 1.4: The organelle-paralogy hypothesis mechanism for endomembrane evolution. (**A**) Proteins function coordinately to mediate membrane trafficking steps such as vesicle budding or fusion. (**B**) Genes encoding these proteins undergo duplication and sequence divergence resulting in sets of machinery capable of mediating distinct membrane trafficking pathways. (**C**) This paralogous protein evolution coincides with, or allows, a similar pattern of paralogous organelle origin and evolution as distinct sets of protein machinery acquire organelle specificity. Over time, this results in the diversity of organelles in the membrane trafficking system. This figure was reproduced from Mast *et al.* (2014) Figure 2C.

1.6 Knowledge gaps regarding the early evolution of membrane trafficking

Knowledge of the evolution of the membrane trafficking system (and of other aspects of cell biology) may be divided into knowledge of evolutionary states or events occurring at one of three different stages: Post-LECA, LECA, and pre-LECA. Knowledge of later time points may be used to make inferences about evolution at earlier time points. That is, observation and comparison of extant eukaryotes of diverse lineages is required for inferring which traits were present in the LECA. This distinguishes traits that arose in a specific lineage from those that are inherited from a common ancestor and subsequently lost in some cases (Dacks and Doolittle, 2001). In turn, the evolutionary events in pre-LECA eukaryote evolution rely heavily on knowledge of the LECA. For these reasons, reconstructing features of the LECA has become a focal point of evolutionary cell biology (Koumandou *et al.*, 2013). However, important knowledge gaps exist (or existed prior to the work in this thesis being performed) in each of these three areas, post-LECA, LECA, and pre-LECA, with relevance to the evolution of the membrane trafficking system and specifically the OPH.

The OPH describes a continuous process of organelle evolution, so if it is expected to explain autogenous organelle evolution at the earliest stages of eukaryotic evolution (pre-LECA), then it should be able to explain those at later stages in specific lineages (post-LECA) (Dacks *et al.*, 2008). However, post-LECA evolution of the relevant membrane trafficking protein families in specific lineages has not been investigated thoroughly. Previous studies performing the type of work necessary to fill this knowledge gap include analyses of SNARE proteins in Embryophyta (Sanderfoot, 2007), Rab protein evolution in Archaeplastida (Petrželková and Eliáš, 2014), and adaptor protein complex evolution in kinetoplastid parasites (Manna *et al.*, 2013). Much work is needed to elucidate the evolution of these proteins (and other proteins potentially relevant to the OPH) across all extant lineages. This would include analysis of adaptins in Fungi and Embryophyta. Fungi and Embryophyta are of particular interest because they contain relatively well-characterized cell biological model systems, and many species within these lineages have publicly available sequenced genomes.

Many features of the complex LECA and its membrane trafficking system are well established (Koumandou *et al.*, 2013), with numerous organelles being inferred as present due to broad conservation of the organelles and their organelle-specific proteins among extant eukaryotes. As mentioned above, these include an ER, Golgi, endosomes, and lysosomes. However, molecular evidence for the presence of several other organelles or organelle subdomains is lacking. For example, many eukaryotes are phagocytic leading to the consensus that the LECA possessed phagosomes,

yet, unlike many other organelles, no protein machinery specific to phagosomes has been found to be conserved among diverse phagocytic eukaryotes (Yutin *et al.*, 2009). The Golgi, a key membrane trafficking organelle, was also present in the LECA, and many proteins, including membrane trafficking proteins such as the subunits of the COPI vesicle coat complex (Hirst *et al.*, 2014), are specific to the Golgi. The Golgi is not only inferred to be present in the LECA, but to be present as a complex stack of functionally differentiated cisternae (Mowbrey and Dacks, 2009). The OPH is thought to be relevant to this type of differentiation between organelle subdomains (Dacks and Field, 2007). However, conservation of the protein machinery that would underlie this differentiation of organelle subdomains remains unknown.

Finally, one of the main objectives of the OPH is to understand the evolutionary mechanism underlying the increase in organellar complexity in the membrane trafficking system prior to the LECA (Dacks and Field, 2007). This depends on first reconstructing the complement of proteins within key protein families that was present in the LECA, and then determining the evolutionary relationships between these proteins. The three most relevant protein families are the Adaptor Protein (AP) complexes (along with the related COPI and TSET complexes), the Rab GTPase protein family, and the SNARE protein superfamily (Dacks and Field, 2007). The LECA complement of APs and their interrelationships have been studied extensively (Hirst *et al.*, 2011, 2014). The LECA complement of Rab proteins and their interrelationships have been investigated intensely as well (Elias *et al.*, 2012). From these studies, a detailed hypothesis describing a specific sequence of events in the early evolution of membrane trafficking organelles has not emerged. This is due in part to incomplete understanding of the ancestral functions of some of the AP complexes (as well as TSET), and the still limited resolution of evolutionary relationships between Rab proteins as well as very different functions of apparently closely related Rabs. However, the LECA complement of SNARE proteins has not been thoroughly investigated using phylogenetic methods. It is currently unknown precisely which SNAREs represent those present in the LECA, and the relative timing of duplication events giving rise to these paralogues in pre-LECA evolution is entirely unexplored.

1.7 Organization of this thesis

The over-arching goal of the work in this thesis is to reconstruct the sequence of key steps in the evolution of the eukaryotic endomembrane system through inference from genomic sequence data and previous functional characterization of relevant proteins in various eukaryotic lineages. Pursuit of this goal is divided into two main sub-goals which address important knowledge gaps

identified above:

1. Test key predictions of the OPH with respect to the evolutionary history of relevant membrane trafficking protein families in eukaryotes (Chapters 2, 3, and 4).

2. With applicability of the OPH as a premise, test alternative scenarios for endomembrane organelle evolution as revealed by phylogenetic analysis of the SNARE protein families (Chapter 5).

Chapters 2, 3, and 4 include analyses of the evolution of particular membrane trafficking protein families across taxonomic groups in which they have not been analyzed previously. Each of these chapters is relevant to testing the applicability of the OPH with respect to the particular protein families, by examining how closely organelle evolution may be connected to the evolution of these proteins. Chapter 2 tests whether loss of membrane trafficking proteins correspond to losses of organelles. This is a corollary of the prediction that these proteins maintain organelle specificity over time. This is done by examining the evolution of adaptor protein complex subunits in Fungi and closely related eukaryotes. Chapter 3 tests the prediction that duplications of relevant membrane trafficking proteins occur coordinately with differentiation of novel organelles. This is done by examining the evolution of subunits of adaptor protein, COPI, and TSET complexes among embryophytes. Chapter 4 tests the prediction that the Golgi-specific golgin tether proteins evolved coordinately with sub-compartments of the Golgi apparatus, which may have differentiated prior to the LECA. The content of chapters 2, 3, and 4 have been published as peer-reviewed articles by Barlow *et al.* (2014), Larson *et al.* (2019), and Barlow *et al.* (2018), respectively (see preface for author contributions).

Chapter 5 builds upon the results in chapter 2 to 4, which are largely consistent with explanation of organelle evolution in terms of the OPH. Chapter 5 examines the evolution of the four SNARE protein families across a broad taxonomic sampling of eukaryotes. SNARE proteins are key proteins for understanding early organelle evolution in terms of the OPH, because of the conservation of the function of specific SNAREs in each family to the ER, Golgi, endosomes, and the plasma membrane. The work included in chapter 5 differentiates between alternative scenarios of organelle evolution based on phylogenetic analysis of SNARE families.

Chapter 2

From all to (nearly) none: Tracing adaptin evolution in Fungi[1]

2.1 Preface

As discussed in Chapter 1 section 1.5, a phylogenetic prediction of the OPH is that the phylogenies of an essential subset of membrane trafficking protein families will be in agreement (have consistent topologies) regarding the respective organelle-specificity of family members at all stages of eukaryotic evolution (see Figure 1.4 which is from Mast *et al.* (2014) Figure 2C). This would be the case if these proteins do not frequently acquire functions at new organelles. Also, a prediction regarding protein function is that orthologues of members within the same key subset of membrane trafficking protein family members share the same organelle specificity/function across eukaryotic diversity. Assuming strong organelle-specificity, this also implies that when such organelle-specific proteins are duplicated, then organelles should duplicate/differentiate as well, and that when such proteins are lost through gene mutation, then the organelles will be lost as well.

As originally conceived by Dacks and Field (2007) and Dacks *et al.* (2008) the OPH included the adaptor protein complexes (see Figure 1.4) as playing a key role in evolution of the membrane trafficking system, due to the highly paralogous evolution of the complex subunits. Therefore, testing the predictions of the OPH with respect to this family of proteins may be informative for understanding whether the OPH may be generally applicable to the evolution of the membrane trafficking system. The fungal model system *Saccharomyces cerevisiae* and a small sampling of other fungi were previously shown to lack genes encoding subunits of Adaptor Protein complexes 4 and 5 (Markus Boehm and Juan S. Bonifacino, 2001; Hirst *et al.*, 2011). However, the precise pattern of evolution of the AP complexes in this lineage remained unclear. Accordingly, work in this chapter examines the evolution of AP complexes in Fungi and close relatives, and this

[1]The contents of this chapter, exclusive of preface and afterword, have been previously published as: Barlow, L.D., Dacks, J.B., and Wideman, J.G. (2014). From all to (nearly) none: Tracing adaptin evolution in Fungi. *Cellular Logistics*, 4(1):e28114. doi:10.4161/cl.28114.

is relevant for answering the question of what happens when key membrane trafficking protein complexes are lost in evolution.

2.2 Introduction

The five heterotetrameric Adaptor Protein (AP or adaptin) complexes function as cargo adaptors that recruit coat proteins during vesicle formation at various stages and locations in the membrane trafficking system (Markus Boehm and Juan S. Bonifacino, 2001; Hirst *et al.*, 2013b). AP-1 mediates formation of clathrin coated vesicles (CCVs) that traffic between the *trans*-Golgi network (TGN) and early endosomes[2]. AP-2 recruits clathrin to the plasma membrane and is involved in vesicle formation during clathrin-mediated endocytosis (CME) (Markus Boehm and Juan S. Bonifacino, 2001). AP-3 is involved in vesicle transport between tubular endosomes and lysosomes (Robinson *et al.*, 2010)[3][4]. AP-4 is involved in vesicle transport from the TGN to endosomes and the cell surface (Burgos *et al.*, 2010; Hirst *et al.*, 1999; Dell'Angelica *et al.*, 1999) and has been linked to human disorders such as Alzheimers disease (Burgos *et al.*, 2010), cerebral palsy (Moreno-De-Luca *et al.*, 2011), and spastic paraplegia (Abou Jamra *et al.*, 2011). The recently discovered AP-5 complex has also been linked to Alzheimers disease and spastic paraplegia suggesting a functional relation to AP-4 (Robinson *et al.*, 2010; Słabicki *et al.*, 2010; Boukhris *et al.*, 2008). And, like AP-4, AP-5 is suggested to play a role in the late endosomal pathway (Hirst *et al.*, 2013a, 2011).

Not only do different AP complexes function in different endosomal pathways, different AP complexes also interact with different coat proteins. As mentioned above, AP-1 and AP-2 interact with clathrin to form CCVs. AP-3 has been suggested to interact with clathrin (Dell'Angelica *et al.*, 1998); however, AP-3 does not require clathrin for function in *Saccharomyces cerevisiae* (Simpson *et al.*, 1996) and clathrin function is dispensable for AP-3 function in mammalian cells (Peden *et al.*, 2002; Zlatic *et al.*, 2013). Alternatively, it has been shown that Vps41, a protein with a clathrin heavy chain domain, likely plays a role similar to clathrin in AP-3 vesicles (Rehling *et al.*, 1999; Asensio *et al.*, 2013). While the coat protein for AP-4 vesicles is currently unknown, two human proteins (SPG11 and SPG15) interact with AP-5, are structurally similar to clathrin, and thus likely perform a homologous function in AP-5-mediated processes (Hirst *et al.*, 2013a,

[2]Loss of the AP-1 gamma subunit is embryonic lethal in mice (Zizioli *et al.*, 1999)

[3]While Barlow *et al.* (2014) only cite Robinson *et al.* (2010) here, AP-3 was not the focus of that article, and more direct evidence for this function of AP-3 in transport from tubular endosomes to lysosomes was shown by Cowles *et al.* (1997), Le Borgne *et al.* (1998), and Peden *et al.* (2004).

[4]Mutations in genes encoding AP-3 subunits have been implicated in Hermansky-Pudlak syndrome (Badolato and Parolini, 2007).

2011).

AP complexes are homologous to the coat protein complex COPI (Schledzewski *et al.*, 1999), with both structural and functional similarities retained. COPI comprises two large subunits (γ and β-COP), a medium subunit (δ-COP) and a small subunit (ζ-COP). Likewise, each AP complex comprises two large subunits (γ1, α2, δ3, ε4, ζ5, and β1-5) a medium subunit (μ1-5) and a small subunit (σ1-5). Since all five AP complexes as well as COPI have been shown to have eukaryote-wide distribution, it is thought that the paralogous expansions that gave rise to these complexes all occurred before the divergence of the major eukaryote lineages (Hirst *et al.*, 2011; Field *et al.*, 2007). Thus, it can be inferred that all five AP complexes were present in the Last Eukaryote Common Ancestor (LECA).

Although all five AP complexes were present in the LECA and are widely distributed across extant eukaryote diversity, each AP complex (except AP-1) has been lost in various lineages. The clade of salivarian trypanosomes including *Trypanosoma brucei* has been shown to have lost AP-2 but retain clathrin and maintain a modified version of CME (Manna *et al.*, 2013). AP-3 has been lost in two independent lineages of Apicomplexa (Nevin and Dacks, 2009). AP-4 has been lost several times: in Fungi, *e.g.*, *S. cerevisiae*; several kinetoplastids *e.g.*, the *Leishmania* genus, *Phytomonas serpens*, as well as *Trypanosoma congolense* (Manna *et al.*, 2013); in diplomonads, *e.g.*, *Giardia intestinalis*; and in the Archaeplastida in both red algae, *e.g.*, *Cyanidioschyzon merolae*, and green algae, *e.g.*, *Chlamydomonas reinhardtii* (Field *et al.*, 2007). AP-5 is the most often lost adaptin complex and has been lost at least once in every major group of eukaryotes (see Hirst *et al.* (2011) for further details). All eukaryotes investigated thus far, retain clathrin except the intracellular parasites, the Microsporidia (Field *et al.*, 2007; Fedorov and Hartman, 2004). Remarkably, the microsporidian *Encephalitozoon cuniculi* was reported to lack clathrin but maintain AP subunits (Fedorov and Hartman, 2004).

Microsporidia are extremely divergent, extremely reduced, intracellular pathogens (Vávra and Lukeš, 2013). Due to their apparent lack of mitochondria and Golgi, they were once considered to be basal, early-diverging eukaryotes (Sogin, 1997). With the discovery of extremely reduced mitochondria called mitosomes (Williams *et al.*, 2002), and Golgi-homologs in Microsporidia (Takvorian *et al.*, 2013; Takvorian and Cali, 1994), this idea has been abandoned. Phylogenetic advances have now firmly placed Microsporidia in the fungal kingdom; thus, instead of being basal eukaryotes, microsporidia now represent perhaps the most divergent and reduced of all known eukaryotes (Corradi *et al.*, 2010). Reduction of other systems and processes in microsporidia has been investigated (*e.g.* metabolic pathways, transporter proteins, mitochondrial protein import pathways,

genome content) (Corradi *et al.*, 2010; Keeling and Fast, 2002; Katinka *et al.*, 2001; Heinz *et al.*, 2012); however, the extent to which AP and clathrin loss has occurred in diverse microsporidia is currently unknown. Thus, in this study we investigate the evolution of AP complexes in the fungal lineage with a particular focus on the enigmatic intracellular parasites, the Microsporidia.

2.3 Results and Discussion

2.3.1 The last common ancestor of Holomycota contained a complete set of adaptins

Since some animals (*e.g.*, humans) retain all five currently identified AP complexes, it can be inferred that the last common ancestor of animals also had at least five AP complexes. However, since the diversity of fungal genomes now available has not been systematically analyzed for presence of AP complexes the complement of AP complexes present in the last common fungal ancestor remains unknown. Knowing the complement of AP complexes in the fungal ancestor would allow us to trace the pattern of AP loss from the fungal ancestor across the fungal tree of life to the more familiar taxa in the Dikarya (*e.g.*, *Cryptococcus neoformans* and *S. cerevisiae*). Thus, in order to identify the complement of AP complexes likely present in the fungal ancestor we searched the recently sequenced genome of *Fonticula alba*, a member of the fonticulids, the sister-group to Fungi (Brown *et al.*, 2009). Using BLAST and HMMer homology[5] searching algorithms we identified a complete set of adaptins in *F. alba* (Figure 2.1). The only subunit that we could not confidently identify was β5 for which we found a putative homolog whose identity could not be confirmed by reciprocal homology searches. However, since only β1, β3, and the putative β5 sequences were retrieved by our original HMMer analysis we think that this sequence could represent an extremely divergent β5 subunit. All other *F. alba* adaptin subunits were verified by reciprocal homology searching into either the *Homo sapiens* or *Thecamonas trahens* genomes for reference. Since animals are known to contain all five AP complexes, the presence of all five AP complexes in *F. alba* indicates that the ancestor of Fungi and *Fonticula* retained the full adaptin complement present in the LECA and likely had a more complex endosomal trafficking system than that seen in *S. cerevisiae*.

[5]The word "similarity" would be more accurate here, as homology is inferred from similarity.

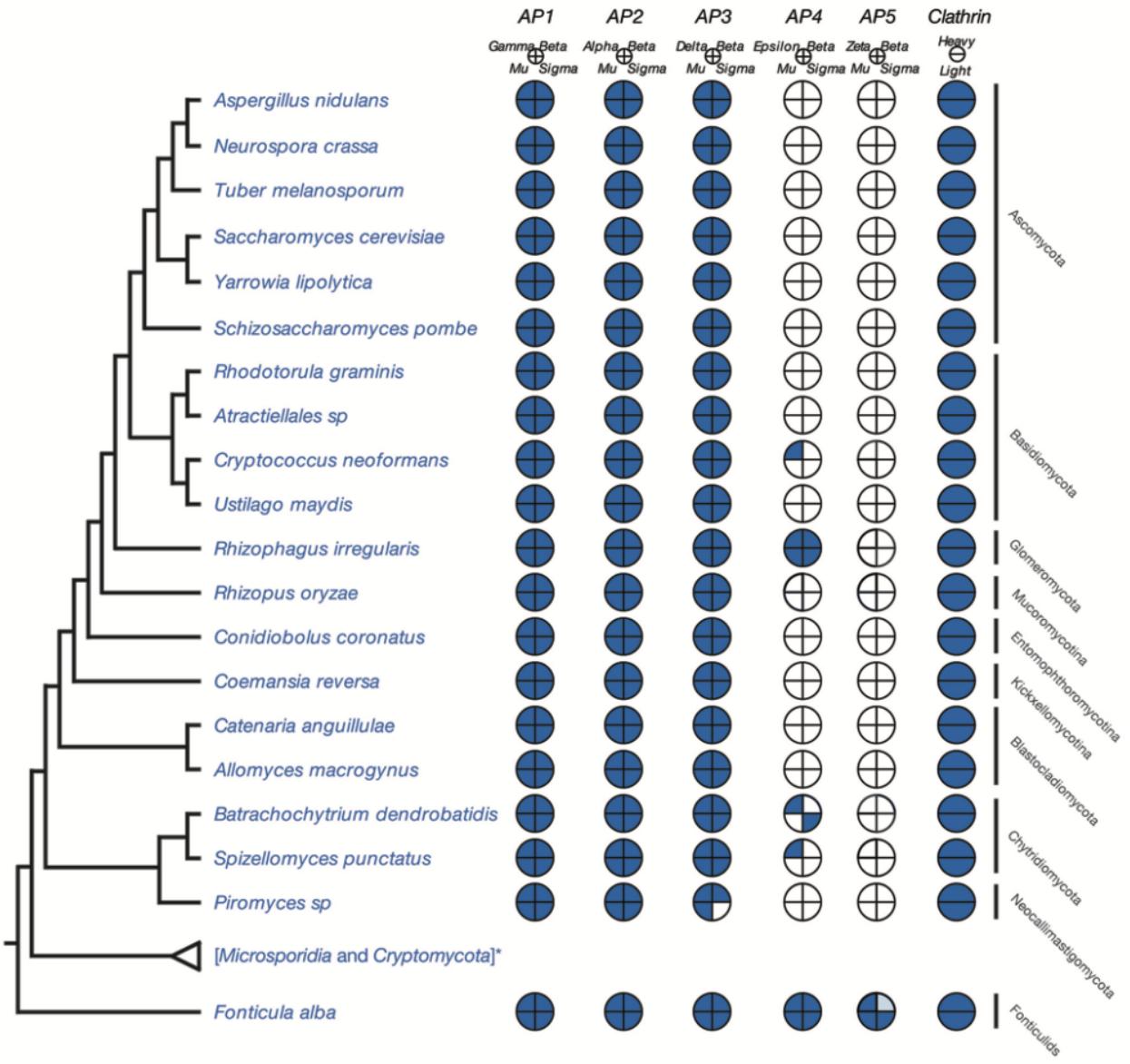


Figure 2.1: Coulson plot showing clathrin and adaptin subunits in Fungi and *F. alba*. Homologs of adaptin and clathrin components were found using a combination of BLAST and HMMer algorithms. Presence of identified protein sequences is represented by a filled-in pie piece; failure to identify a candidate ortholog represented by an open pie piece. The tree is based on the topology obtained by James *et al.* (2006). All subunits of the Adaptor Protein (AP) complexes 1 to 5 as well as clathrin heavy and light chains were identified in the genome of the holomycotan *Fonticula alba* (bottom row of subplots), and AP-1, 2, 3, and clathrin chains are highly conserved among the other genomes representing Holomycota (including Fungi). However, no orthologs of the AP-5 complex were identified, in other genomes queried, and orthologs of AP4 subunits were lost in almost all genomes. The conservation of at least the AP-4 ε subunit in representatives of Chytridomycota, Glomeromycota, and Basidiomycota suggests that losses of AP-4 subunits occurred multiple times independently. This figure is reproduced from Barlow *et al.* (2014) Figure 1.

2.3.2 Multiple loss of AP-4 in the fungal tree of life

Having established that the ancestral holomycotan had all five AP complexes we wanted to determine the distribution of AP complexes across the known diversity of fungi. It has been previously shown that *S. cerevisiae* has maintained AP-1 to 3 (Markus Boehm and Juan S. Bonifacino, 2001). However, the extent to which adaptin complexes are retained in other fungi has not been systematically assessed. As above, we used BLAST and HMMer homology searching algorithms, to search the genomes of representatives from relevant fungal lineages (see James *et al.* (2006)) including Ascomycota, Basidiomycota, Glomeromycota, Mucoromycotina, Entomophthoromycotina, Kickxellomycotina, Blastocladiomycota, Chytridiomycota and Neocallimastigomycota. We found homologs of all AP-1 to 3 subunits in representatives from each of these major clades (Figure 2.1). The only subunit we could not identify was σ3 from *Piromyces* sp., a member of the Neocallimastigomycota. Since σ subunits are very short, failure to detect could be explained by extreme sequence divergence or an incompleteness of the database.

Since it had been previously reported that the basidiomycete *C. neoformans* has retained ε4 but the ascomycete *S. cerevisiae* has not (Field *et al.*, 2007), we were interested in uncovering the extent to which AP-4 subunits have been retained in other fungi. Our homology searches confirmed that no AP-4 subunits are detectable in *S. cerevisiae* or any other ascomycete (Figure 2.1); and furthermore, while we confirmed the previously reported presence of ε4 in *C. neoformans* (Field *et al.*, 2007), we could not find evidence for the presence of any other AP-4 subunits in *C. neoformans* or any other basidiomycete analyzed in our study. No AP-5 subunit was detected in any ascomycete or basidiomycete genome. The presence of ε4 in *C. neoformans* indicates that the ε4 subunit has been lost independently in at least two basidiomycete lineages. It is unknown what role a lone ε4 might play in endosomal trafficking in *C. neoformans* and is thus an attractive candidate model for studying degenerate AP function.

Next, we searched for AP-4 and AP-5 in basally diverging fungi (fungi other than ascomycetes and basidiomycetes) to determine the pattern of AP loss in less familiar taxa. While we could not find any evidence for AP-5 in any genome analyzed, in searching the genome of the glomeromycete, *R. irregularis*, we found a complete AP-4 complex (Figure 2.1). AP-4 subunits were also identified in the chytridiomycetes *Spizellomyces punctatus* (ε) and *Batrachochytrium dendrobatidis* (ε and σ). No AP-4 subunits could be detected in genomes sampled from other fungal lineages including Mucoromycotina, Entomophthoromycotina, Kickxellomycotina, Blastocladiomycota, and Neocallimastigomycota, suggesting that AP-4 has been lost numerous times over the course of fungal evolution. In fact, if the currently accepted fungal tree is correct (James *et al.*,

2006, 2013), then AP-4 has been independently lost seven times.

2.3.3 AP-1 β and AP-2 β were present in the last common ancestor of the Holomycota

The only AP subunits not present in the LECA were β1 and β2 (Dacks *et al.*, 2008). Instead, the LECA contained a single β1/2 protein that likely interacted with both AP-1 and AP-2 complexes. This is evidenced by phylogenetic analyses and the presence of only a single β1/2 protein in several extant organisms spanning the tree of eukaryotes including *Drosophila melanogaster*, *Caenorhabditis elegans*, *Monosiga brevicollis*, *Phytophthora ramorum*, *Dictyostelium discoideum* and *Oryza sativa* (Markus Boehm and Juan S. Bonifacino, 2001; Dacks *et al.*, 2008). In our investigation of the adaptin complement of Fungi we noticed that every genome analyzed contained two β1/2 proteins (Figure 2.1). Since the *S. cerevisiae* β1/2 proteins Apl1p and Apl2p have been determined to act independently in AP-1 (β1 = Apl2p) and AP-2 (β2 = Apl1p) (Rad *et al.*, 1995; Kirchhausen, 1990), we wanted to determine if these *S. cerevisiae* proteins are the result of an ancient duplication in Holomycota, or if multiple lineage-specific duplications have occurred. In order to answer this question, we reconstructed the phylogeny of holomycotan β1-4 subunits (Figure 2.2). Our analysis robustly reconstructed the β2, β3, and β4 clades. Although the clade designated as β1 was not well supported in our analysis, it was strongly excluded from all other clades. In addition, since no β1 protein was ever robustly placed within the β2 clade it can be inferred that the β1 proteins did not arise from independent lineage-specific duplications. Taken together, these results indicate that the β1 proteins most likely form a monophyletic clade resulting from a single duplication event involving the ancient β1/2 protein. Thus, it can be concluded that the common ancestor of Holomycota had both β1 and β2 subunits that arose from a single ancient duplication of the ancestral β1/2 protein.

That the β1/2 duplication occurred early in the evolution of the holomycotan lineage is rather surprising. In Holozoa (multicellular animals and their protistan relatives), the sister lineage to Holomycota, many basal taxa including the choanoflagellate *M. brevicollis* as well as several invertebrates like *C. elegans* and *D. melanogaster*, contain only a single β1/2 subunit. This suggests that the duplication seen in animals occurred rather late in the course of animal evolution (*i.e.*, in the vertebrate lineage). Similarly, in the excavate lineage, while the heterolobosean *Naegleria gruberi* has a single β1/2 subunit, the kinetoplastids have undergone a duplication and thus have both β1 and β2 (Manna *et al.*, 2013). As more genomes become available, it will become possible to determine precisely when β1/2 duplications occurred in lineages like excavates, plants, and animals (see Dacks *et al.* (2008)).

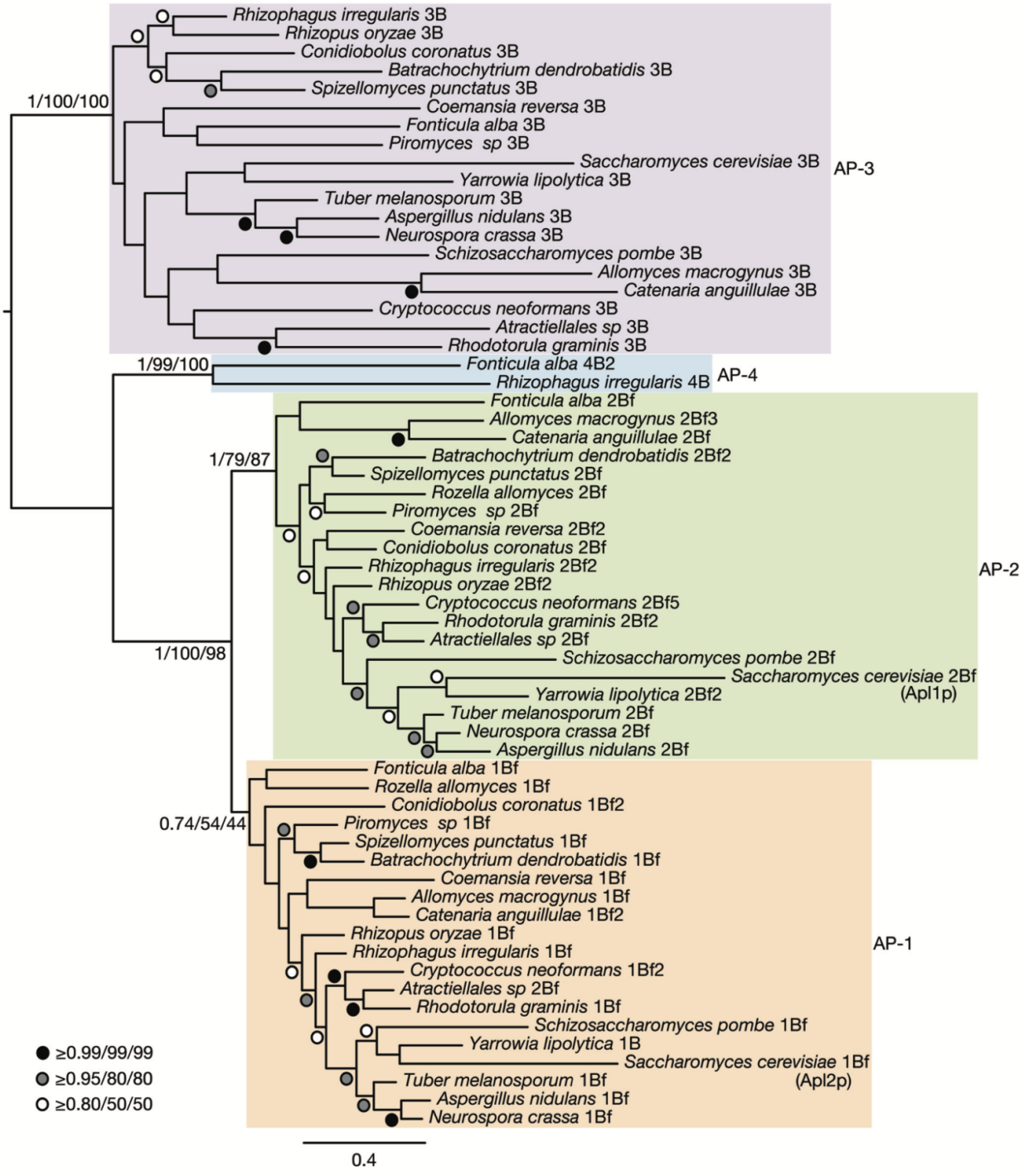

Rhizophagus irregularis 3B
Rhizopus oryzae 3B
Conidiobolus coronatus 3B
Batrachochytrium dendrobatidis 3B
Spizellomyces punctatus 3B
1/100/100
Coemansia reversa 3B
Fonticula alba 3B
Piromyces sp 3B
Saccharomyces cerevisiae 3B
AP-3
Yarrowia lipolytica 3B
Tuber melanosporum 3B
Aspergillus nidulans 3B
Neurospora crassa 3B
Schizosaccharomyces pombe 3B
Allomyces macrogynus 3B
Catenaria anguillulae 3B
Cryptococcus neoformans 3B
Atractiellales sp 3B
Rhodotorula graminis 3B
1/99/100
Fonticula alba 4B2
Rhizophagus irregularis 4B
AP-4
Fonticula alba 2Bf
Allomyces macrogynus 2Bf3
Catenaria anguillulae 2Bf
1/79/87
Batrachochytrium dendrobatidis 2Bf2
Spizellomyces punctatus 2Bf
Rozella allomyces 2Bf
Piromyces sp 2Bf
Coemansia reversa 2Bf2
Conidiobolus coronatus 2Bf
Rhizophagus irregularis 2Bf2
Rhizopus oryzae 2Bf2
AP-2
Cryptococcus neoformans 2Bf5
Rhodotorula graminis 2Bf2
Atractiellales sp 2Bf
1/100/98
Schizosaccharomyces pombe 2Bf
Saccharomyces cerevisiae 2Bf
(Apl1p)
Yarrowia lipolytica 2Bf2
Tuber melanosporum 2Bf
Neurospora crassa 2Bf
Aspergillus nidulans 2Bf
Fonticula alba 1Bf
Rozella allomyces 1Bf
Conidiobolus coronatus 1Bf2
0.74/54/44
Piromyces sp 1Bf
Spizellomyces punctatus 1Bf
Batrachochytrium dendrobatidis 1Bf
Coemansia reversa 1Bf
Allomyces macrogynus 1Bf
Catenaria anguillulae 1Bf2
Rhizopus oryzae 1Bf
AP-1
Rhizophagus irregularis 1Bf
Cryptococcus neoformans 1Bf2
Atractiellales sp 2Bf
Rhodotorula graminis 1Bf
Schizosaccharomyces pombe 1Bf
Yarrowia lipolytica 1B
Saccharomyces cerevisiae 1Bf
(Apl2p)
Tuber melanosporum 1Bf
Aspergillus nidulans 1Bf
Neurospora crassa 1Bf
≥0.99/99/99
≥0.95/80/80
≥0.80/50/50
0.4

Figure 2.2: Phylogenetic tree of holomycotan β-adaptin sequences. Bayesian phylogenetic analysis was performed using MrBayes, and maximum likelihood analysis was performed using PhyML and RAxML. The MrBayes tree topology is shown, and support for the nodes defining the adaptin clades (boxed) are shown in the following order: MrBayes/PhyML/RAxML. For all other nodes, the values are symbolized as inset. Significance thresholds for branch support were considered to be MrBayes prior probability of ≥ 0.80 (out of a maximum of 1.0), and bootstrap percentages of ≥ 50 (out of a maximum of 100) from bootstrap analyses using PhyML or RAxML. The topology and significant support for the clades containing AP-1 β (Apl2) and AP-2 β (Apl1) indicates early duplication of Adaptor Protein (AP) 1/2 β subunits prior to the last common ancestor of holomycota rather than in a more recent ancestor of the yeast *Saccharomyces cerevisiae*. This figure is reproduced from Barlow *et al.* (2014) Figure 2.

2.3.4 Microsporidia lack clathrin but retain (cryptic) adaptin subunits

After analyzing AP complexes in other fungi, we wanted to determine the adaptin complement present in the most basal clade of fungi that comprises Microsporidia and Cryptomycota (James *et al.*, 2013). Using alignments generated from our set of identified fungal AP components we constructed HMMs and searched for adaptins in the genomes of the cryptomycete *Rozella allomyces* and several microsporidians. In similar findings to most other fungi, we identified AP-1 to 3 subunits in the *R. allomyces* genome but could not identify any AP-4 subunits. Since only the δ3 and μ3 subunits could be found in the *R. allomyces* genome it is possible that AP-3 may be undergoing a process of degeneration. However, our inability to identify the other AP-3 subunits could again reflect extreme sequence divergence or the incompleteness of the database rather than genuine loss. As more cryptomycete genomes are sequenced further investigation will determine the extent of AP loss in this lineage.

Microsporidia have been suggested to lack clathrin but retain AP subunits (Fedorov and Hartman, 2004). In our analysis of microsporidian genomes we detected, at most, a single representative of each of the four AP subunits (Figure 2.3). The only exception was the presence of two closely related μ subunits in *Nosema apis* which, based on reciprocal BLAST analysis into the *N. apis* genome, appear to be the result of a relatively recent gene duplication or gene database error (data not shown). While clathrin is in all other fungi (Figure 2.1), neither the clathrin heavy nor light chain was detected in any microsporidian genome analyzed (Figure 2.3). Of note, a single complete AP complex was detected in several species across the diversity of microsporidia including *Encephalitozoon cuniculi*, *E. intestinalis*, *N. bombycis*, *Vittaforma corneae*, *Enterocytozoon bieneusi*, *Trachipleistophora hominis*, *Edhazardia aedis*, *Antonospora locustae* and *Vavraia culicis floridensis*. Three of the four AP subunits were detected in *N. apis* and *N. ceranae* (each was missing a different AP subunit). Only two subunits were detected in *Spraguea lophii* (μ and σ); and remarkably, only a β subunit could be detected in the basally diverging nematode parasite *Nematocida parisii*. Microsporida are notorious for having extremely divergent protein sequences (Slamovits *et al.*, 2006). Therefore, failure to detect could be caused by extreme sequence divergence; thus, our hypothesis of loss must be followed-up with experimental studies. As new genomes are sequenced, and potentially more basal microsporidians are discovered, more light will be shed on the evolutionary history of reduction in these enigmatic parasites.

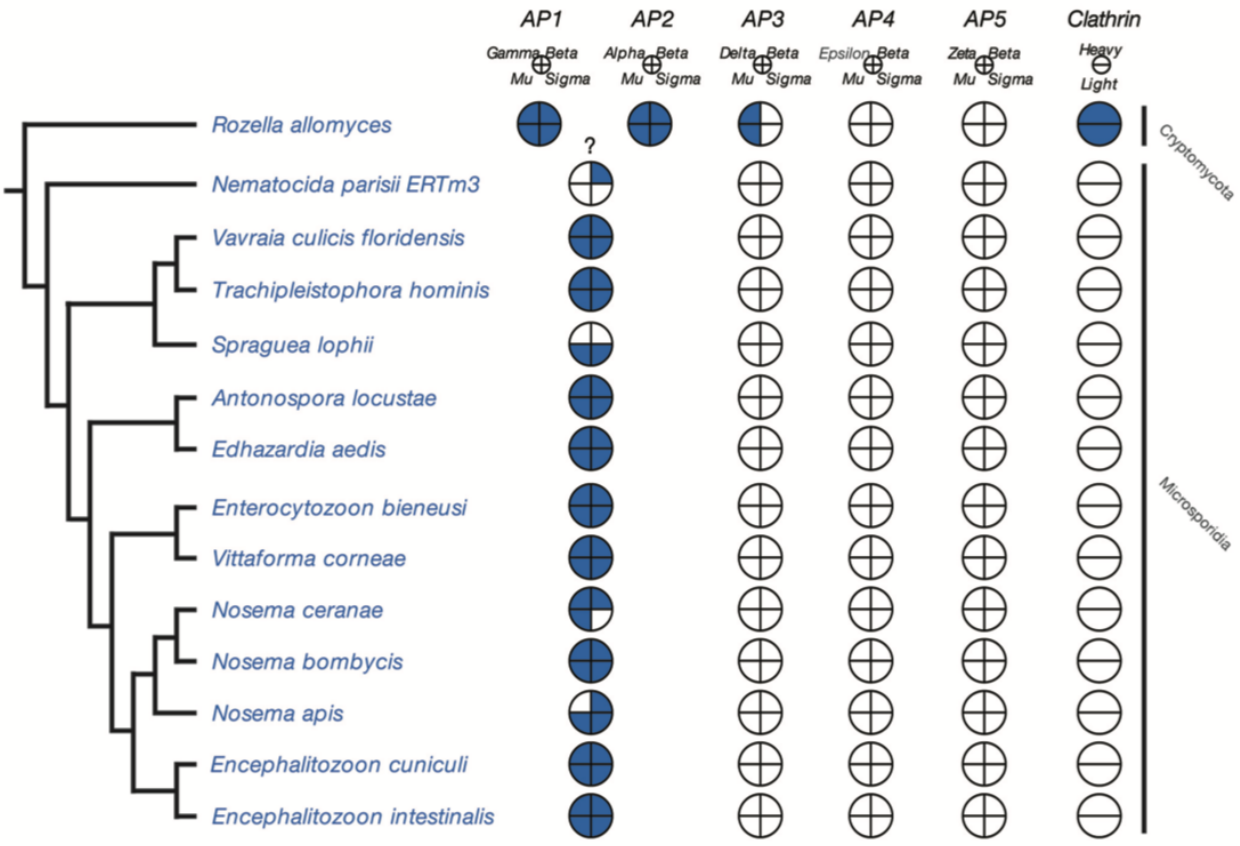


Figure 2.3: Coulson plot of adaptin subunits identified in cryptomycetes and microsporidians. Homologs of adaptin and clathrin components were found using a combination of BLAST and HMMer algorithms (as for results shown in Figure 2.1). Presence of identified protein sequences is represented by a filled-in pie piece; failure to identify a candidate ortholog represented by an open pie piece. The question mark indicates ambiguity in classification of identified homologues as orthologues of either AP-1 or AP-2 subunits. The tree is based on topologies obtained in previous studies (Vossbrinck and Debrunner-Vossbrinck, 2005; Capella-Gutiérrez *et al.*, 2012). While subunits of AP-1, 2 and 3 complexes as well as clathrin heavy and light chains were identified in the genome of the cryptomycotan *Rozella allomyces*, no clathrin chains and a much reduced set of AP complex subunits with similarity to AP1 and AP2 subunits were identified among the microsporidia. This figure is reproduced from Barlow *et al.* (2014) Figure 3.

2.3.5 The cryptic AP complex of Microsporidia is likely either AP-1 or AP-2

Thus far, we have demonstrated that a single AP complex is present in several microsporidian species; however, the identity of this complex, at this point, remains unclear. Since clathrin is not present in microsporidia we reasoned that if another coat could be found, its presence could provide further clues to the identity of microsporidian adaptins. However, while coat proteins like Vps41 and Spg11 were detected in other genomes (Spg11 was only found in *F. alba*), these proteins could not be detected in any microsporidian genome analyzed (data not shown) and could not provide information on the identity of the cryptic microsporidian adaptins.

Thus, we searched for other ways to determine the identity of the cryptic AP complex. Homology searching experiments using microsporidian AP protein sequences as queries retrieved adaptin subunits from AP-1 to 4 with approximately equal E-values, providing little guidance as to evolutionary provenance of the microsporidian AP complex (data not shown). We therefore reconstructed the phylogenies of each AP subunit from a set of basal fungi in an attempt to determine the identity of the microsporidian adaptins (Appendix A chapter 2 Figs. S1-S4). Unfortunately, these trees were largely uninformative as microsporidian sequences either did not resolve into a particular AP clade (Appendix A chapter 2 Figs. S1 and S2), or were robustly excluded from all AP clades (Figure 2.3). Microsporidian μ subunits were excluded from AP-1, 2, and 4 μ clades but were not strongly excluded from the AP-3 μ clade (Appendix A chapter 2 Fig. S4) raising the possibility that the microsporidian μ proteins are μ3. However, our analysis of μ alignments revealed the conservation of a tryptophan residue that is present in all fungal μ1 and μ2 proteins but absent in μ3 and μ4 proteins. This residue is also conserved in nearly all microsporidian μ proteins (not shown). Therefore, although our single-protein μ phylogeny places the microsporidian μ proteins in a clade with fungal μ3 proteins, we suggest that this is likely an artifact due to long branch attraction. Instead we support the more likely hypothesis that the microsporidian AP subunits are all derived from a single AP complex[6].

Since the single-protein phylogenies did not show robust discordance, and assuming that the four microsporidian AP subunits are derived from a single AP complex having a shared evolutionary history, we chose to use a concatenated phylogeny of AP subunits in order to increase the phylogenetic signal and resolve the identity of the microsporidian AP complex. However, to err on the side of caution, we excluded μ from our concatenated analysis in case it is indeed μ3, although we believe this to be unlikely.

[6]Of course, more complex scenarios involving complementation of subunits from several different AP complexes allowing combination into a new chimeric AP complex are plausible.

Our concatenated phylogeny provided a tree with excellent support (Figure 2.4), uniting the microsporidian AP with the clades of AP-1 and AP-2. Since the topology of the concatenation is similar to the topology seen in the γαδε phylogeny (Appendix A chapter 2 Fig. S3), we hypothesized that the γαδε phylogenetic signal was over-powering the phylogenetic signal in the β and σ proteins. Thus, we performed a concatenated phylogenetic analysis of only the β and σ subunits in order to get a more resolved phylogeny for these two subunits (Figure 2.5). This analysis provided robust exclusion of the microsporidian proteins from all AP clades except AP-1. Due to the identification of the conserved tryptophan in the μ1/2 and microsporidian μ proteins, as well as the β and σ single protein and concatenated phylogenies, we conclude that it is likely that the microsporidian complex is most likely either AP-1 or AP-2. Given that AP-1 has never been reported as lost in any eukaryote, and given the result of the β and σ concatenated tree, it is tempting to suggest that the microsporidian complex is derived from AP-1. However, this is admittedly highly speculative and requires functional confirmation.

If this conclusion is correct, the question that arises is: How does AP-1 or AP-2 function in the absence of clathrin? Fortunately, there are some examples in the literature that can help answer this question. First, in the apicomplexan *Toxoplasma gondii* clathrin acts at the TGN in conjunction with AP-1; however, clathrin is not present at the plasma membrane. Instead, AP-2 functions at the plasma membrane without clathrin (Pieperhoff *et al.*, 2013). Therefore, if *T. gondii* AP-2 can function without clathrin, then it is conceivable that the microsporidian AP complex may function similarly. Second, although microsporidia contain COPI and COPII it has been suggested that they do not contain true COPI or COPII vesicles; instead, membranes are proposed to be trafficked progressively from the ER through the Golgi compartment to the plasma membrane without vesicles (Beznoussenko *et al.*, 2007)[7]. If this is true, then the microsporidian AP complex could potentially function similarly to microsporidian COPI, but at the TGN-like compartments.

[7]The non-vesicular mechanism proposed by Beznoussenko *et al.* (2007) involves maturation of Golgi compartments directly from subdomains of the ER.

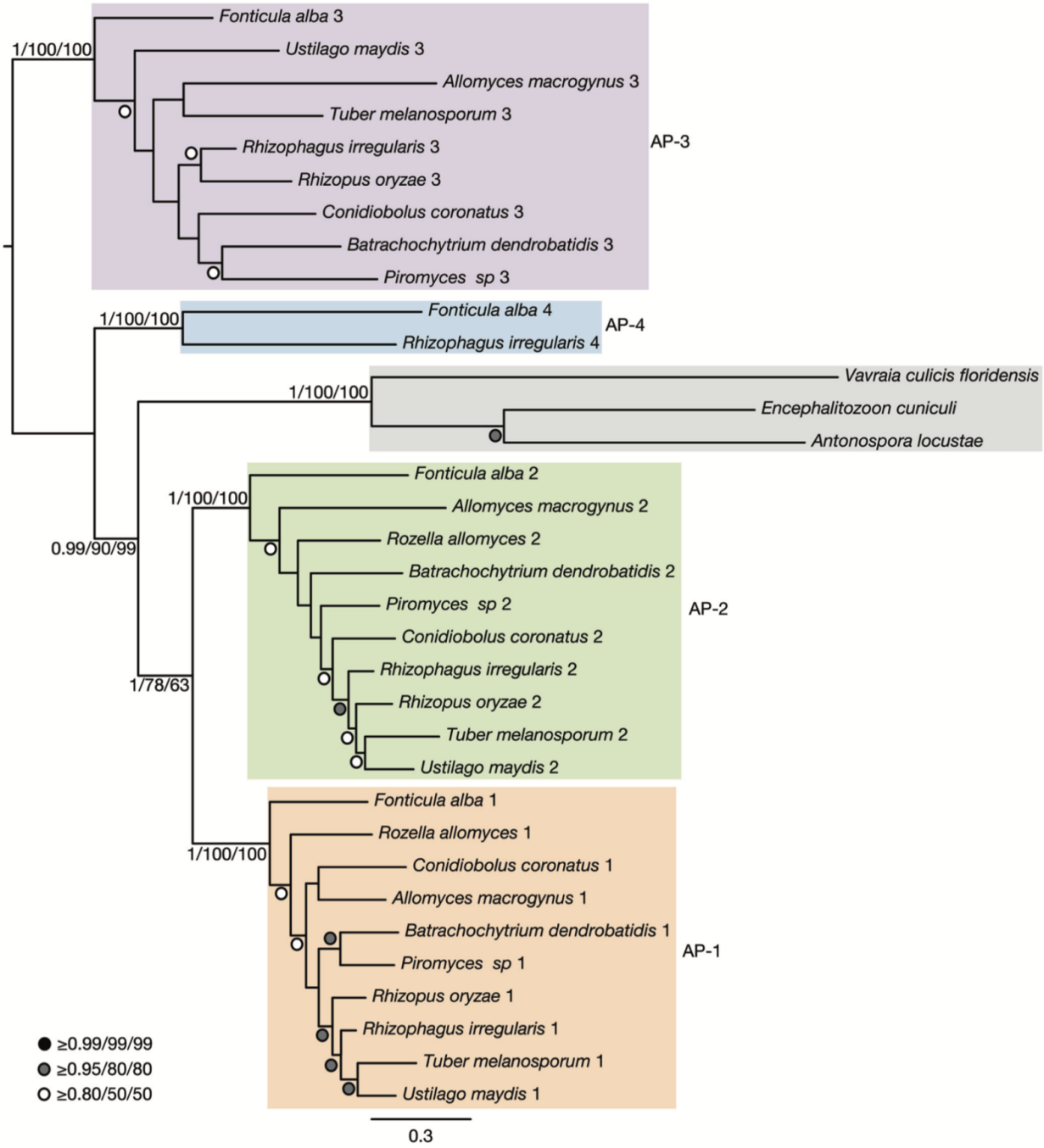

Fonticula alba 3
Ustilago maydis 3
Allomyces macrogynus 3
Tuber melanosporum 3
Rhizophagus irregularis 3
Rhizopus oryzae 3
Conidiobolus coronatus 3
Batrachochytrium dendrobatidis 3
Piromyces sp 3
AP-3
1/100/100
Fonticula alba 4
Rhizophagus irregularis 4
AP-4
1/100/100
Vavraia culicis floridensis
Encephalitozoon cuniculi
Antonospora locustae
1/100/100
Fonticula alba 2
Allomyces macrogynus 2
Rozella allomyces 2
Batrachochytrium dendrobatidis 2
Piromyces sp 2
Conidiobolus coronatus 2
Rhizophagus irregularis 2
Rhizopus oryzae 2
Tuber melanosporum 2
Ustilago maydis 2
AP-2
1/100/100
0.99/90/99
1/78/63
Fonticula alba 1
Rozella allomyces 1
Conidiobolus coronatus 1
Allomyces macrogynus 1
Batrachochytrium dendrobatidis 1
Piromyces sp 1
Rhizopus oryzae 1
Rhizophagus irregularis 1
Tuber melanosporum 1
Ustilago maydis 1
AP-1
1/100/100
≥0.99/99/99
≥0.95/80/80
≥0.80/50/50
0.3

Figure 2.4: Concatenated phylogenetic analysis of holomycotan AP complex subunits. An alignment of AP-1, 2, 3, and 4 subunit amino acid sequences was prepared with each hypothetical AP complex in each species represented by sequences of the β, γ/α/δ/ε, and σ subunits concatenated in that order. The MrBayes tree topology is shown, and support for the nodes defining the clades for different AP complexes (boxed) are shown in the following order: MrBayes/PhyML/RAxML. For all other nodes, the values are symbolized as inset. Significance thresholds for branch support were considered to be MrBayes prior probability of ≥ 0.80 (out of a maximum of 1.0), and bootstrap percentages of ≥ 50 (out of a maximum of 100) from bootstrap analyses using PhyML or RAxML. The topology and branch supports indicate that the microsporidian AP complexes are most closely related to the AP-1 and AP-2 complexes, robustly supporting a clade containing the microsporidian sequences and AP-1 and AP-2 sequences. This figure is reproduced from Barlow *et al.* (2014) Figure 4A.

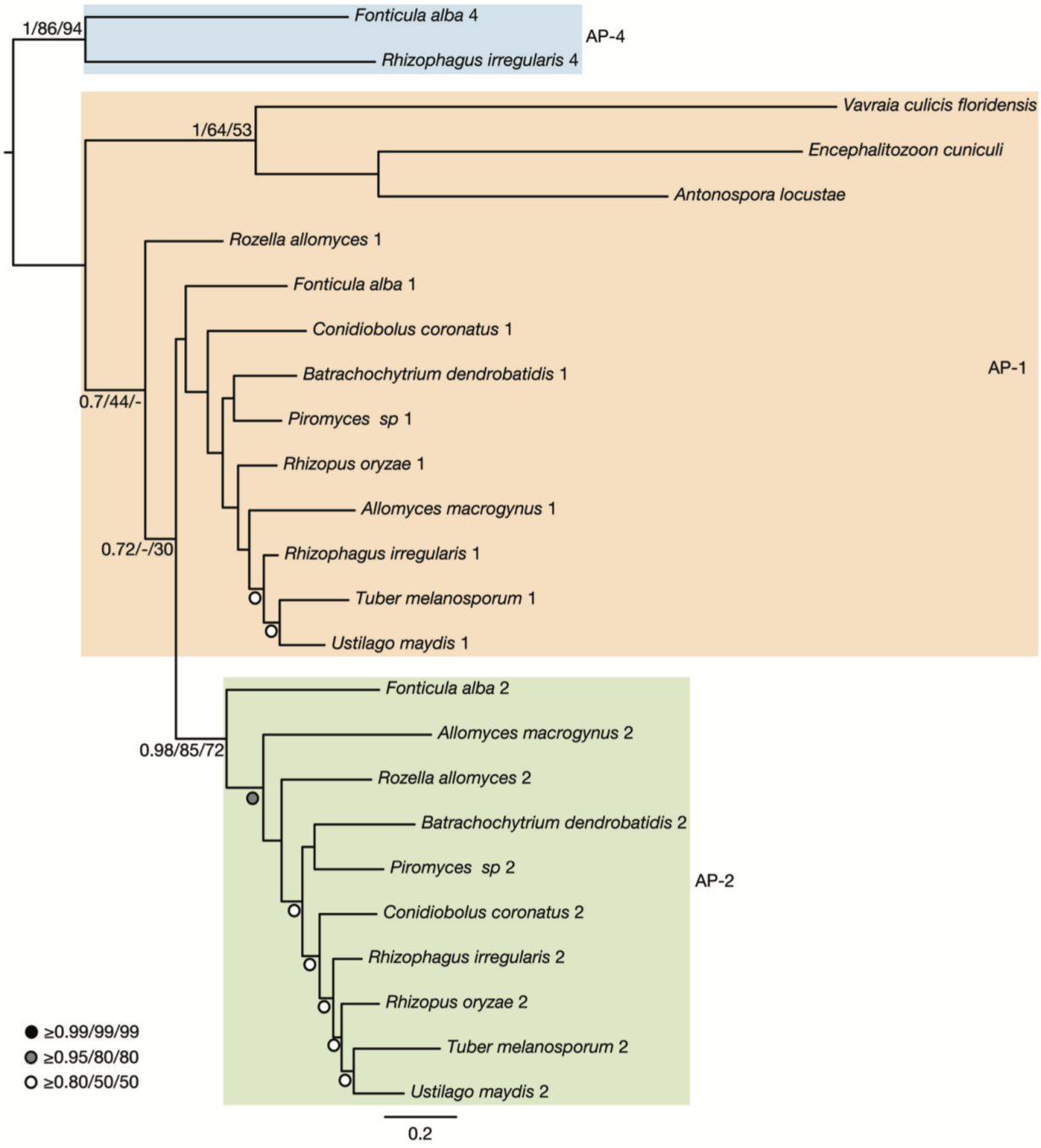

1/86/94
Fonticula alba 4
Rhizophagus irregularis 4
AP-4
1/64/53
Vavraia culicis floridensis
Encephalitozoon cuniculi
Antonospora locustae
Rozella allomyces 1
Fonticula alba 1
Conidiobolus coronatus 1
Batrachochytrium dendrobatidis 1
Piromyces sp 1
Rhizopus oryzae 1
Allomyces macrogynus 1
Rhizophagus irregularis 1
Tuber melanosporum 1
Ustilago maydis 1
AP-1
0.7/44/-
0.72/-/30
0.98/85/72
Fonticula alba 2
Allomyces macrogynus 2
Rozella allomyces 2
Batrachochytrium dendrobatidis 2
Piromyces sp 2
Conidiobolus coronatus 2
Rhizophagus irregularis 2
Rhizopus oryzae 2
Tuber melanosporum 2
Ustilago maydis 2
AP-2
≥0.99/99/99
≥0.95/80/80
≥0.80/50/50
0.2

Figure 2.5: Concatenated phylogenetic analysis of holomycotan AP complex subunits. An alignment of AP-1, 2, and 4 subunit amino acid sequences was prepared with each hypothetical AP complex in each species represented by sequences of the β, γ/α/δ/ε, and σ subunits concatenated in that order (as for the analysis in Figure 2.4, but without AP-3 sequences). The MrBayes tree topology is shown, and support for the nodes defining the clades for different AP complexes (boxed) are shown in the following order: MrBayes/PhyML/RAxML. Where the topology conflicts with the consensus of bootstrap topologies for either PhyML or RAxML, this is indicated with a dash ('-'). For other nodes, the values are symbolized as inset. Significance thresholds for branch support were considered to be MrBayes prior probability of ≥ 0.80 (out of a maximum of 1.0), and bootstrap percentages of ≥ 50 (out of a maximum of 100) from bootstrap analyses using PhyML or RAxML. The topology and branch supports indicate again that the microsporidian AP complexes are related to the AP-1 and AP-2 complexes, but the lack of support found here for monophily of the AP-1 sequences to the exclusion of the microsporidian AP sequences (as is seen again for the AP-2 clade) suggests that the hypothesis that the microsporidian sequences are orthologous to the AP-1 sequences is slightly preferred. This figure is reproduced from Barlow *et al.* (2014) Figure 4B.

2.4 Conclusion

In this study, we have provided context to AP evolution in fungi that will allow for better application of knowledge gained from experimental models like *S. cerevisiae* to fungal diversity as a whole. First, we explored the genome of *F. alba* and demonstrated that the last common ancestor of the Holomycota contained all five AP complexes. Second, our analysis has shown that several independent losses of AP-4 have occurred over the course of fungal evolution. In our analysis, only the glomeromycete *Rhizophagus irregularis* was found to retain a complete AP-4 complex. Thus, this species is a candidate model organism for studying AP-4 function in fungi. Third, in addition to the general trend of loss in fungi, our results indicate that the duplication that gave rise to the β1 and β2 proteins in *S. cerevisiae* occurred very early, prior to the divergence of *F. alba* and Fungi. This means that, although functional data from β1 and β2 in *S. cerevisiae* cannot be directly compared with the similarly named genes in metazoan systems, there are direct orthologs of these genes in other fungi that can, and should, be compared. This opens new avenues of investigation of membrane-trafficking in the alternate model systems in fungi such as *Neurospora*, *Yarrowia* and more. Last, we present evidence that microsporidia contain an extremely reduced endosomal system, as they lack clathrin and retain, at most, only a single cryptic AP complex.

Our results demonstrate the ubiquity of AP-1 to 3 in fungi suggesting a general conservation of endosomal trafficking throughout this kingdom. This suggests that findings about these complexes in *S. cerevisiae* are likely generalizable to other fungi. However, presence of AP-4 subunits in several fungi highlights lineages where generalizations from *S. cerevisiae* are not adequate. Furthermore, in highly reduced organisms like the Microsporidia where only a single cryptic AP complex is retained knowledge gained from model systems is not very informative. Thus systems that contain a more complex (*R. irregularis* and *F. alba*) or more reduced (Microsporidia) membrane trafficking than familiar model organisms represent important avenues for further experimentation.

2.5 Methods

2.5.1 Genome databases

Publicly available genomes analyzed in this study include[8]: from the Joint Genome Institute (Grigoriev *et al.*, 2012): *Aspergillus nidulans*, *Tuber melanosporum*, *Yarrowia lipolytica*, *Rhodotorula graminis*, *Atractiellales* sp., *Cryptococcus neoformans*, *Ustilago maydis*, *Rhizophagus irregularis*, *Rhizopus oryzae*, *Conidiobolus coronatus*, *Coemansia reversa*, *Catenaria anguillulae*, *Batrachochytrium dendrobatidis*, *Piromyces* sp., *Rozella allomyces*, *Antonospora locustae*;

[8]These genomes were selected to represent the taxonomic diversity of Fungal lineages.

from the Broad institute Microsporidia Comparative Sequencing Project, Broad Institute of Harvard and MIT (`http://www.broadinstitute.org/`): *Fonticula alba*, *Spizellomyces punctatus*, *Allomyces macrogynus*, *Nematocida parisii*, *Vavraiaculicis floridensis*, *Vittaforma corneae*, *Edhazardia aedis*, *Nosema ceranae*, *Encephalitozoon cuniculi*, *Encephalitozoon intestinalis*; from NCBI: *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Spraguea lophii*, *Enterocytozoon bieneusi*, *Nosema bombycis*, *Nosema apis*.

2.5.2 Homology searching

Putative AP orthologs were identified using a modified reciprocal best hit method. First, previously identified opisthokont AP-1 to 4 subunits were used to construct Hidden Markov Models (HMMs) of all four complex subunits. AP-5 subunit HMMs were constructed separately using a set of previously validated sequences found by Hirst *et al.* (2011). These HMMs were then used to search for homologs in various holomycotan genomes[9] using HMMer (`http://hmmer.janelia.org/`). A reciprocal pHMMer search into the human genome was then performed using the putative fungal homologs as queries. Putative fungal adaptin subunits that retrieved human AP subunits as first hits in the reciprocal analysis were preliminarily annotated as orthologous to the human AP subunits. AP-1 to 4 subunit annotations were verified by phylogenetic analysis (not shown). Some fungal AP subunits that were not found in predicted protein databases were reconstructed from genomic assembly sequences available on NCBI. Sequences retrieved in this study are listed in Appendix A chapter 2 Table S1.

When searching for AP subunits in microsporidia, in addition to searching the predicted protein sets, we also searched the NCBI non-redundant nucleotide database as well as all relevant whole-genome shotgun assemblies for missing adaptin subunits; however, no additional AP subunits could be identified. The only putative AP subunit identified in this manner was shown to be the Nematocida ζ-COP ortholog. Coulson plots were generated using the Coulson plot generator (Field *et al.*, 2013).

2.5.3 Phylogenetic analyses

Each of the four sets of paralogous sequences (γ1, α2, δ3, ε4, ζ5; β15; μ15; σ15) were aligned using MUSCLE (Edgar, 2004) v.3.8.31, and manually adjusted as needed using MacClade v.4.08 (`http://macclade.org/index.html`) and/or Mesquite v.2.75 (Maddison, W. P. and D.R. Maddison. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 `http://mesquiteproject.org`). Since AP-5 was only found in *F. alba* and AP-5 proteins are

[9]Specifically, peptide sequences predicted from these genomes were searched

always long-branching (Hirst *et al.*, 2011), these sequences were left out of our phylogenetic analyses. Model testing was performed using ProtTest v1.3 with a Gamma rate distribution and accounting for invariant sites as appropriate (Abascal *et al.*, 2005). Phylogenetic tree reconstructions were performed using MrBayes v3.2.2 for Bayesian analysis (Ronquist and Huelsenbeck, 2003). MrBayes analyses were run with the following parameters; prset aamodelpr = fixed(WAG); mcmcngen = 10,000,000; samplefreq = 1000; nchains = 4; startingtree = random; sump burnin = 2500; sumt burnin 2500. Each of the MrBayes analyses reached split frequencies with average standard deviations of < 0.012, indicating convergence. Posterior probabilities were used to measure node support, and values ≥ 0.80 were considered significant. Maximum likelihood bootstrap values (100 pseudoreplicates) were obtained using PhyML v.3.0 (Guindon and Gascuel, 2003; Guindon *et al.*, 2010) and RAxML v.7.2.6 (Stamatakis, 2006) with the LG (Le and Gascuel, 2008) model. Bootstrap values ≥ 50 were considered significant. MrBayes and RAxML analyses were run on the CIPRES server (Miller *et al.*, 2010). Divergent sequences representing long branches were removed from the alignments of initial data sets (data not shown) in order to limit the effects of long-branch attraction. The microsporidian sequences were aligned separately from those of other fungi in the final data sets, in order to reduce the effect of random alignment of unconserved positions. The number of taxa and amino acid positions in the alignments for all figures are shown in Appendix A chapter 2 Table S2, and all alignment files are available by request.

2.6 Afterword

The evolution of AP complexes in Fungi shows numerous gene losses compared to many other eukaryotic lineages (see Figures 2.1 and 2.3). Do these result in organelle losses that might be predicted by the OPH? The most extensive losses occur very early in the Microsporidia lineage of Fungi (Figure 2.3), which may only have one functional AP complex. Interestingly, the Microsporidia may be the single most highly reduced eukaryotic lineage with respect to their cellular features (Vávra and Lukeš, 2013). Early losses of AP complexes in this lineage then seem to be correlated with establishment of the reduced endomembrane system of microsporidia. Whether there is indeed a causal relationship here, however, remains unknown especially considering that little is known about the membrane trafficking system of microsporidia and the precise organellar complement they contain.

Notable losses in the remainder of Fungi are mainly losses of the AP-4 and AP-5 complex subunits. Do these correlate with reductive evolution of the endomembrane system as well? Cell biological work with the model yeast *Saccharomyces cerevisiae* has previously revealed that the yeast

endomembrane system does appear less complex than the mammalian endomembrane system, and specifically may lack early endosomes that are distinct from the *trans*-Golgi network (TGN) (Day *et al.*, 2018). However, even if Fungi including *S. cerevisiae* have not lost any relevant organelles, the results here may still be consistent with the OPH. Loss of AP-5 is quite common among different eukaryotic lineages (Hirst *et al.*, 2011, 2014), and its function in transporting cargo from the late endosome to the Golgi may be somewhat redundant with the function of the Retromer complex in extant eukaryotes anyway (Hirst *et al.*, 2018). In the case of AP-4 losses, the results here show that these occurred multiple times independently (Figure 2.1). This indicates that losses of AP-4 were relatively recent events that could not have play a role in defining the ancestral fungal endomembrane system. The function of AP-4 is less well-understood than other adaptin complexes, but after the work presented in this chapter was completed work by (Ramirez-Macias *et al.*, 2018) (which I contributed to as a co-author) showed that in yeast evolution the loss of AP-4 occurs at a similar time point to the origin of the Exomer complex, which is a fungal-specific vesicle coat that also acts at the TGN. With these considerations regarding AP-5 and AP-4, as well as the overall pattern of conservation of the other complexes, it seems that the evolution of AP complexes in Fungi is generally consistent with the OPH.

Chapter 3

Recent gene duplications dominate evolutionary dynamics of adaptor protein complex subunits in embryophytes[10]

3.1 Preface

The mechanism of evolution described by the OPH involves protein duplication and the origins of novel organelles occuring coordinately in evolution (Dacks and Field, 2007; Dacks *et al.*, 2008). So, if a relevant membrane trafficking protein family underwent duplications in a eukaryotic lineage, then the endomembrane system of that lineage would be expected to eventually reflect those duplications in expansion of the number of organelles (or perhaps trafficking pathways) in cells of that lineage. The embryophytes (land plants) are notorious for possessing large numbers of duplicated genes (Panchy *et al.*, 2016), and AP complex subunits are no exception (Hirst *et al.*, 2014). Yet, plants have a remarkably conserved endomembrane system overall compared to several other more divergent eukaryotic lineages (Barlow and Dacks, 2018). Plants lack novel lineage-specific endomembrane organelles that would rival those such as the alveoli of alveolates including ciliates (Stelly *et al.*, 1991). Although plant cells contain an endomembrane compartment termed the cell plate during cell division, this is a nascent plasma membrane rather than a persistent organelle (Barlow and Dacks, 2018). Considering this, a phylogenetic prediction based on the OPH would be that these duplications of adaptins in plants occurred recently in specific plant lineages, and thus have not been established for long enough to result in origin of complexes with specificities for novel organelles. Thus, the analysis performed in this chapter is relevant for understanding whether duplications of AP complexes (simultaneous duplications of their subunits) is associated with increases in the number of organelles as would be predicted by the OPH.

[10]The contents of this chapter, exclusive of preface and afterword, have been previously published as: Larson, R.T., Dacks, J.B., and Barlow, L.D. (2019). Recent gene duplications dominate evolutionary dynamics of adaptor protein complex subunits in embryophytes. *Traffic*, in press doi:10.1111/tra.12698.

3.2 Introduction

Plants are key sources of food, medicine, and shelter. In plant cells, carbohydrate and medicinal molecules that compose the products we rely on must be trafficked to specific locations in the cell to be processed, secreted, or degraded. As in other eukaryotic cells, this trafficking occurs via a system of organelles and vesicle trafficking pathways, termed the membrane trafficking system (Bonifacino and Glick, 2004). This system in plants, as revealed mainly from work with the model plant *Arabidopsis thaliana* (Koornneef and Meinke, 2010; Singh and Jürgens, 2018) has maintained and specialized ancient trafficking components (reviewed by Barlow and Dacks (2018)). Moreover, embryophytes comprise a remarkable diversity of species, with over 300,000 described species of angiosperms alone (Jiao *et al.*, 2011). Among this diversity, specialization of cellular machinery, including membrane trafficking proteins, may occur in different plant lineages, allowing a means of cellular adaptation to different environments.

Gene duplication and loss are key mechanisms involved in adaptation, increases in complexity, and specialization in the evolution of eukaryotic cells including plant cells (Conant and Wolfe, 2008). Gene duplication is specifically implicated in the origin of new membrane trafficking organelles and pathways via progressive duplication and divergence of genes encoding membrane trafficking proteins (Dacks and Field, 2007; Elias *et al.*, 2012; Mast *et al.*, 2014). There are several potential examples of membrane trafficking protein paralogues underlying plant-specific trafficking pathways. The *A. thaliana* proteins KNOLLE, PEN1, and Phytolongins are the result of duplications of genes encoding SNARE proteins that occurred during the origin of embryophytes (Sanderfoot, 2007; Vedovato *et al.*, 2009). At similar time points, two Rab11 paralogues originated and orthologues of the human Rab proteins 24, 28, 34, RTW and IFT27 were lost (Petrželková and Eliáš, 2014). The plant Rab5 paralogue RabF1 originated earlier, prior to the divergence of the embryophyte lineage from charophyte algae (Petrželková and Eliáš, 2014). Moreover, nine paralogues of the Exo70 subunit of the exocyst tethering complex were present in the ancestor of angiosperms, while only one is present in green algae (Synek *et al.*, 2006).

Yet to be explored in plants is the evolution of a paralogous family of protein complexes which includes the Adaptor Protein (AP) complexes (AP-1 to 5), COat Protein complex I (COPI), and TSET (Figure 3.1) (Dacks and Robinson, 2017; Hirst *et al.*, 2014). These complexes have conserved organelle-specific roles in vesicle formation in eukaryotes (Dacks and Robinson, 2017) (Figure 3.1). The AP complexes are involved in the endocytic and secretory pathways (Figure 3.1) (Markus Boehm and Juan S. Bonifacino, 2001). Each AP complex is a heterotetramer comprising a small sigma (σ) subunit, medium mu (μ) subunit, large beta (β) subunit, and a second large

subunit, either AP-1 gamma (γ), AP-2 alpha (α), AP-3 delta (δ), AP-4 epsilon (ε), or AP-5 zeta (ζ). In addition, AP-1 and AP-2 interact with the vesicle coat proteins clathrin light chain (LC) and clathrin heavy chain (HC). The evolution of APs and related complexes has been explored extensively in lineages including Fungi (Barlow *et al.*, 2014), haptophytes (Lee *et al.*, 2015a), trypanosomatids (Manna *et al.*, 2013), and apicomplexans (Kibria *et al.*, 2019; Nevin and Dacks, 2009; Woo *et al.*, 2015). Pan-eukaryotic analyses of APs and related complexes include sparse sampling of plant genomes, but only indicate general retention of these complexes with loss of AP-5 sigma subunit in the taxonomic supergroup Archaeplastida, which includes red algae, green algae, and embryophytes (Adl *et al.*, 2018; Hirst *et al.*, 2014, 2011). Filling this knowledge gap will be important for understanding the relevance of functional characterizations of AP, COPI, and TSET subunits in *A. thaliana* to those of orthologous subunits in other plants.

Specifically, while *A. thaliana* contains multiple paralogues of several subunits (Hirst *et al.*, 2014; Lee and Hwang, 2014), without the incorporation of plant species from across embryophyta, it is unclear whether the paralogues present in *A. thaliana* are the result of ancient or recent duplication events. In other words, whether these are outparalogues or inparalogues relative to *A. thaliana* or its close relatives is unknown (Sonnhammer and Koonin, 2002). These alternative scenarios have different implications for translating experimental characterizations of AP, COPI, or TSET subunit homologues among plant species: An ancient duplication of a gene occurring in an embryophyte ancestor may indicate a shared or similar function of the subsequently retained paralogues among plants, allowing for experimental characterizations to extend among plants retaining a specific paralogue of the gene. On the other hand, if the paralogues resulted from independent recent duplication events, then this would preclude inference of differential function of paralogues in a given plant species based on differential function of paralogues in another species, such as *A. thaliana*.

To determine the timing of the origin of AP, COPI, and TSET subunit paralogues, we analyzed genomic data from across the taxonomic diversity of embryophytes and close algal outgroups. This involved standard phylogenetic approaches but also the generation of a novel bioinformatics script toolkit which may be of more general use for analyses of paralogous protein family evolution using plant genomic data. Overall, we find that, in contrast to the evolutionary dynamics observed for other membrane-trafficking components such as SNAREs and Rabs, genes encoding AP, COPI, and TSET subunits in embryophytes are generally retained among lineages and have frequently undergone recent duplications, but rarely ancient duplications.

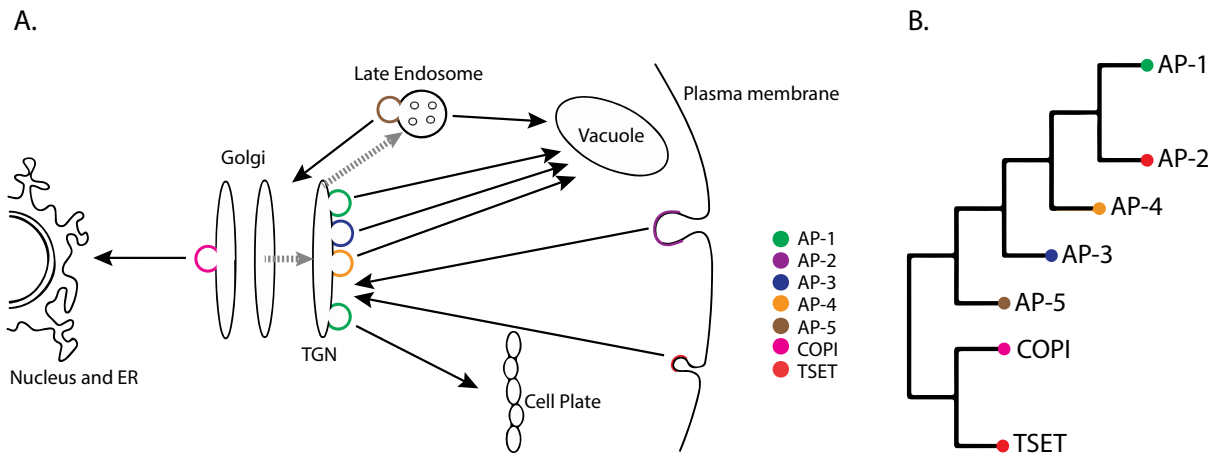


Figure 3.1: Functions and localization of the Adaptor Protein (AP), COatomer Protein I (COPI), and TSET complexes in plant cells, and their evolutionary relationships. (A) Function and localization of these complexes. With the exception of AP-5, functions and localization adapted from review by Singh and Jürgens (2018). The role of AP-5 is inferred from its function in mammalian cells (Hirst *et al.*, 2018). Each complex is denoted with a single colour corresponding to the adjacent figure legend. Black arrows indicate hypothetical routes of transport mediated by specific complexes (or fusion of late endosomes with vacuoles), and grey dashed arrows indicate maturation[11]. (B) Evolutionary history of APs, COPI, and TSET adapted from (Hirst *et al.*, 2014; Klinger *et al.*, 2016b). This figure is reproduced from Larson *et al.* (2019) Figure 1.

[11] Although not indicated here, the COPI complex is involved in transport between Golgi cisternae in addition to transport from the Golgi to the ER.

3.3 Results

3.3.1 AP, COPI, and TSET subunits are generally retained in embryophytes, and are often encoded by multiple gene inparalogues

To determine the presence and number of paralogues among land plants, we initially performed similarity searches using *A. thaliana* peptide sequences as BLASTP and TBLASTN queries to search in a representative sampling of land plant genomes, and applied a reciprocal-best hit criterion. We then generated alignments of positive BLAST hits, and used these alignments as queries in profile searches using HMMer.

This approach allowed the identification of numerous orthologous sequences in each genome analyzed (Appendix A chapter 3 Table S3). No orthologues of the AP-5 sigma subunit were identified in any genome, even when using a Hidden Markov model (HMM) of previously identified AP-5 sigma subunits from a diversity of eukaryotes (Hirst *et al.*, 2014, 2011). However, the remaining AP, COPI, and TSET subunits are highly conserved among embryophytes. This is consistent with previous results from analyses that included the *A. thaliana* and *Physcomitrella* genomes (Hirst *et al.*, 2014, 2011).

3.3.2 A novel approach to sequence comparison reveals distinct paralogues by applying stringent criteria

In the above analysis, the average number of positive hits identified per subunit among all 35 AP, COPI, and TSET subunits (not including AP-5 sigma, or any hits with redundant IDs or those encoded at the same locus) among all 15 genomes is 6.95. Given this apparent number of paralogues, the high identity score between many of the sequences, and in some cases the short sequence length, the criteria applied in the above analysis likely result in over-estimate of the true number of paralogues in each genome. For example, while the *A. thaliana* genome is assembled to the chromosome level, only scaffolds are available for the gymnosperms *Ginkgo biloba* (Guan *et al.*, 2016; Lamesch *et al.*, 2012) and *Picea abies* (Nystedt *et al.*, 2013). This, combined with the long intron length documented for these genomes (Guan *et al.*, 2016; Nystedt *et al.*, 2013), means that some scaffolds appear to only contain coding sequence for a single exon. This is one condition that could lead to errors regarding quantifying the number of paralogous genes in a comparative genomic analysis, if for example each scaffold identified in TBLASTN similarity searches were assumed to represent a separate gene locus. Also, genome assembly errors can sometimes introduce false segmental duplications resulting in duplicate gene models that represent alleles with high sequence identity rather than separate genes (Kelley and Salzberg,

2010). Therefore, further analysis was required to more accurately determine the sets of sequences that are likely to represent considerably different paralogues.

To circumvent these potential sources of error, we applied the following criteria to define positive hits for distinct paralogues[12]. 1) Minimum sequence length of 55 amino acid residues and minimum 15% of the query length. 2) Maximum proximity of 10,000 base pairs between High-scoring Segment Pairs (HSPs) in TBLASTN hits for inclusion in a single gene locus hit, which allows inclusion of multiple hits on the same nucleotide sequence. 3) Overlap of the sequence with higher-ranking hit(s) in a multiple sequence alignment of orthologues, which excludes gene fragments (for example, TBLASTN hits representing different exons of the same gene on different scaffolds). 4) Maximum percent identity of 98% with higher-ranking hits among identified similar sequences in the same genome, which excludes hits that are possibly redundant.

To minimize human error and increase the reproducibility of this analysis, we wrote scripts to perform the analysis (`https://github.com/laelbarlow/amoebae`, `https://zenodo.org/badge/latestdoi/186658027`). While existing software packages facilitate parsing sequence and annotation files, similarity search results, and multiple sequence alignments (Cock *et al.*, 2009), to our knowledge no available software allows the above criteria to be applied. Moreover, the number of identified sequences to which these specific criteria must be applied (approximately 3,647 sequences) makes these scripts a practical necessity.

Using this approach, we identified an exclusive set of sequences that likely represent distinct paralogues (Figure 3.2 and Appendix A chapter 3 Table S2). Inevitably in some cases this approach may exclude distinct paralogues of potential interest. For example, the well-annotated genome of *A. thaliana* contains two paralogues of the AP-2 alpha subunit and of the COPI beta subunit at separate loci, but these paralogues are more than 98% identical (98.2% and 98.9% respectively), and therefore not counted as separate paralogues when our stringent criteria were applied (Appendix A chapter 3 Table S3). These cases may be examples of very recent gene duplications occurring in an ancestor of *A. thaliana* but not prior to its common ancestor with the closely related brassicalid *Capsella grandiflora*, which possesses two paralogues of AP-2 alpha and only one orthologue of COPI beta (Appendix A chapter 3 Figure S2-34 and S2-35). However, the AP-2 alpha paralogues are directly adjacent on *A. thaliana* chromosome 5 and the COPI beta paralogues are directly adjacent on chromosome 4, which suggests some possibility of false segmental duplications in the genome assembly (Kelley and Salzberg, 2010). While resolving uncertainty in genome assembly is beyond the scope of the current study, our maximum 98% identity cutoff reduces the risk of

[12] See afterword to this chapter for further discussion.

propagating assembly errors in our analysis. Overall, applying these criteria brought the average paralogue count per subunit (across all sampled genomes) down from 6.95 to 2.40 (see Appendix A chapter 3 Figure S1 for visualization of these differences). However, there remains considerable variability in the number of paralogues of different subunits in different genomes. For example AP5 zeta is present in fifteen duplicates in *S. moellendorffii*, in contrast with AP5 mu and beta present in one and three copies respectively (Figure 3.2). Also, this more conservative estimate of the number of paralogues still implies that a considerable number of AP, COPI, and TSET subunit duplications occurred in embryophyte evolution, the timing of which was unknown, leading us to investigate further.

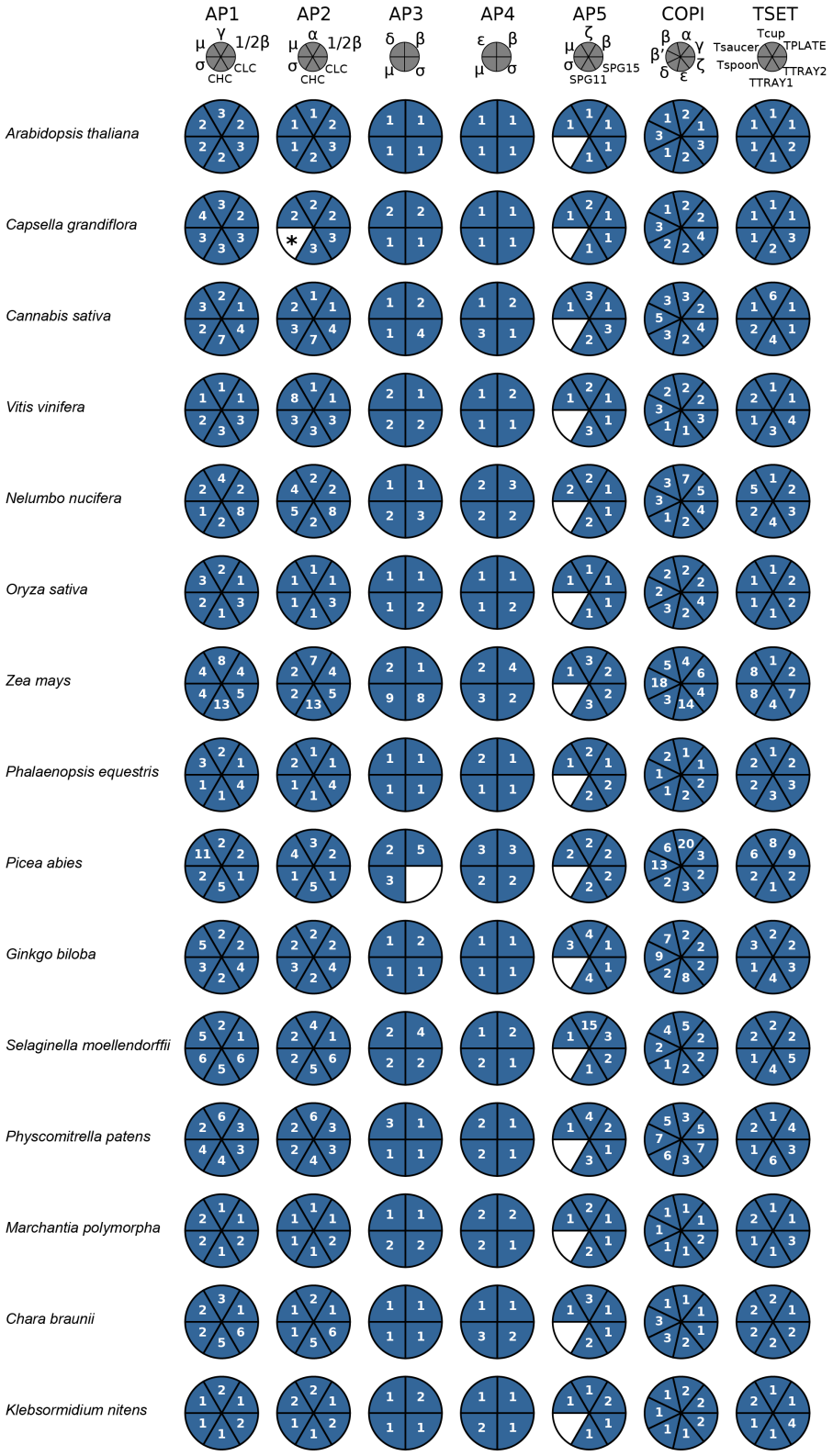

AP1
γ
μ
1/2β
σ
CLC
CHC
AP2
α
μ
1/2β
σ
CLC
CHC
AP3
δ
β
μ
σ
AP4
ε
β
μ
σ
AP5
ζ
μ
β
σ
SPG15
SPG11
COPI
β
α
β'
γ
δ
ε
ζ
TSET
Tcup
Tsaucer
TPLATE
Tspoon
TTRAY2
TTRAY1
Arabidopsis thaliana
Capsella grandiflora
Cannabis sativa
Vitis vinifera
Nelumbo nucifera
Oryza sativa
Zea mays
Phalaenopsis equestris
Picea abies
Ginkgo biloba
Selaginella moellendorffii
Physcomitrella patens
Marchantia polymorpha
Chara braunii
Klebsormidium nitens

Figure 3.2: Summary of search results for each the Adaptor Protein (AP), COatomer Protein I (COPI), and TSET complex subunit in each embryophyte genome sampled. Each subplot of the coulson plot indicates how many paralogues were identified for a subunit of the complex indicated in the legend at the top, specifically in the genome for the species indicated on the left. Greek symbols are used to indicate each AP and COPI subunit in the legend at the top as follows: alpha (α), beta (β), beta prime (β), delta (δ), epsilon (ε), gamma (γ), mu (μ), sigma (σ), and zeta (ζ). Alternative names for relevant subunits of the TSET complex (also known as the TPLATE complex or TPC) in *A. thaliana* are as follows: Tcup is TPLATE complex Muniscin-Like (TML), Tsaucer is TPLATE-associated SH3 domain containing protein (TASH3), Tspoon is Longin-like protein Interacting with TPLATE Adaptor (LOLITA), TTRAY1 is TWD40-1, and TTRAY2 is TWD40-1 (Gadeyne *et al.*, 2014; Hirst *et al.*, 2014). Empty sectors indicate that no orthologues were identified. As described in the Methods, the paralogue counts shown are conservative and exclude short sequences, and those that have 98% or more sequence identity with those sequences counted here. Note that AP-1 beta (AP-1/2 beta), Clathrin Heavy Chain (CHC), and Clathrin Light Chain (CLC) results are shown redundantly in the subplots corresponding to both AP-1 and AP-2, as they are components of both complexes. See Appendix A chapter 3 Table S3 for a detailed summary of all search results and sequence comparisons, and see Appendix A chapter 3 Figure S1 for a similar plot comparing these results to those retrieved with less stringent criteria. *We did not identify any AP2 sigma homologs in *Capsella grandiflora* despite two homologs that are readily identified in the *Capsella rubella* genome on the Phytozome database with the accessions Carubv10010593m and Carubv10011582m. This figure is reproduced from Larson *et al.* (2019) Figure 2.

3.3.3 Evolution of AP-1 and AP-2 in embryophytes

Each AP complex acts at specific organelles or pathways within eukaryotic cells (Figure 3.1A). AP-1 functions in trafficking between the TGN and early endosomes, as well as the plasma membrane in mammalian cells (Markus Boehm and Juan S. Bonifacino, 2001). AP-1 localizes to the TGN in plant cells and appears to function in export from the TGN (Park *et al.*, 2013). AP-1 mu is essential for plant development (Teh *et al.*, 2013), and the AP-1 complex is important for polarized sorting in plant cells (Shimada *et al.*, 2018). Mammalian AP-2 functions in clathrin-mediated endocytosis (Markus Boehm and Juan S. Bonifacino, 2001). In *A. thaliana*, AP-2 is essential for the development of the male reproductive organ and the import of cellulose synthetases (Kim *et al.*, 2013).

While AP-1 and AP-2 shared a common beta subunit in the last eukaryotic common ancestor, vertebrate-specific paralogues of the ancestral AP-1/2 beta subunit are more readily incorporated into either the AP-1 or AP-2 complex (Dacks *et al.*, 2008; Markus Boehm and Juan S. Bonifacino, 2001). However, AP-1 beta can functionally substitute for AP-2 beta in mice to some extent (Li *et al.*, 2010). In *A. thaliana*, there is no evidence of complex-specificity of AP-1/2 beta subunit paralogues (Markus Boehm and Juan S. Bonifacino, 2001), and both paralogues (phytozome accessions AT4G23460.1 and AT4G11380.2) have been identified as components of both the AP-1 and AP-2 complexes in *A. thaliana* (Teh *et al.*, 2013; Yamaoka *et al.*, 2013). Nevertheless, the timing of the duplications producing AP-1/2 beta inparalogues in embryophytes including *A. thaliana* remains unknown. To address this question, we performed a phylogenetic analysis of embryophyte AP-1/2 beta homologues (Figure 3.3). This did not reveal any duplications predating the divergence of the major plant lineages, and instead the *A. thaliana* paralogues are mainly brassicalid-specific inparalogues shared with *Capsella grandiflora*. This analysis thus supports the hypothesis that AP-1/2 beta subunits are shared between both complexes in plants.

Similarly to the AP-1/2 beta result, recent duplications were detected in AP-1 gamma, AP-1 mu, AP-1 sigma, and clathrin LC in brassicalids (Appendix A chapter 3 Figure S2-1 to S2-3 and S2-19). Phylogenetic analyses of non-brassicalid AP-1 subunits revealed poaceae (grass)-specific duplications of AP-1 gamma and sigma subunits, with representatives identified in *Oryza sativa* and *Zea mays* (Appendix A chapter 3 Figure S2-1 and S2-3). In contrast to AP-1, no ancient duplications of AP-2 subunits were identified (Appendix A chapter 3 Figure S2-4 to S2-6). There are however inparalogues in specific genomes for AP-2 subunits, including at least eight duplicates in the case of grapevine, *Vitis vinifera* (Figure 3.2). Also, our methods failed to identify any AP-2 sigma subunit homologues in the *C. grandiflora* genome (Figure 3.2). However,

two homologues are easily identifiable in the *Capsella rubella* genome on phytozome (sequence IDs Carubv10010593m and Carubv10011582m), so this may be an issue with the *C. grandiflora* genome assembly.

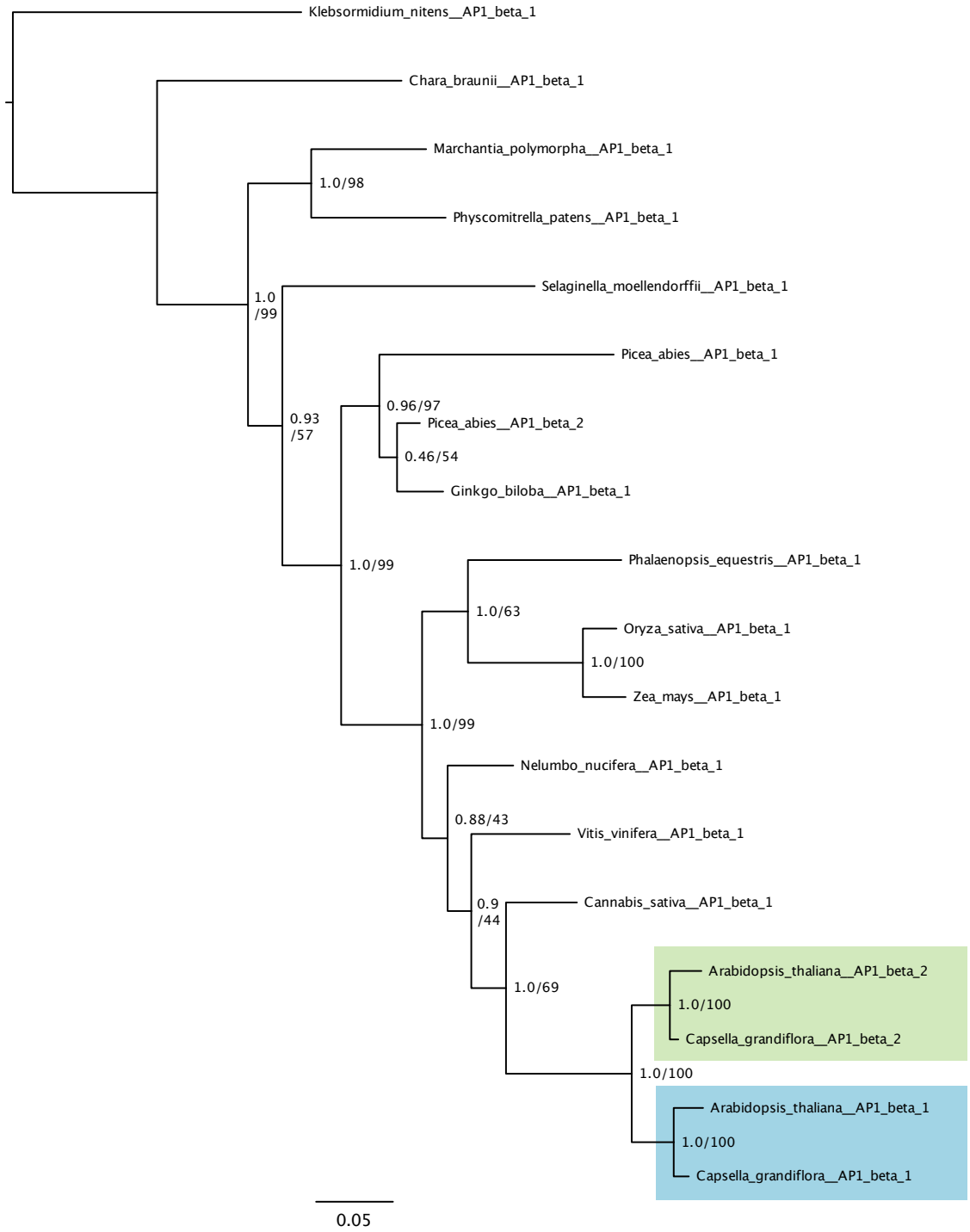


Figure 3.3: Phylogenetic tree of embryophyte Adaptor Protein (AP) 1/2 beta subunit orthologues. Node supports indicate MrBayes Bayesian posterior probabilities and IQ-TREE bootstrap support values. A Bayesian posterior probability of 0.8 or larger and an IQ-TREE bootstrap percentage of 50 or larger indicated significant clade support. A Brassicales specific duplication is highlighted by a green and blue box. This figure is reproduced from Larson *et al.* (2019) Figure 3.

3.3.4 Evolution of AP-3, AP-4, and AP-5 in embryophytes

The AP-3, AP-4, and AP-5 complexes function mainly in TGN/endosomal sorting. In mammalian cells, AP-3 functions in export from early endosomes (Peden *et al.*, 2004), and is important for the biogenesis of lysosome-related organelles (Dell'Angelica, 2009). AP-3 may interact with clathrin in mammalian cells, but this is still a matter of debate (Robinson, 2015). In *A. thaliana*, AP-3 has been shown to function in the biogenesis of lytic vacuoles, but not storage vacuoles (Feraru *et al.*, 2010; Zwiewka *et al.*, 2011). In our analysis, AP-3 exhibits relatively low subunit paralogue numbers compared to the other protein complexes (Figure 3.2) and no duplications predating divergence of the sampled plants (Appendix A chapter 3 Figure S2-7 to S2-10).

In mammalian cells, AP-4 localizes to the TGN (Hirst *et al.*, 1999), and may have functions related to autophagy (Mattera *et al.*, 2017). In *A. thaliana*, AP-4 appears to function in vacuolar protein sorting at the TGN/EE (Fuji *et al.*, 2016). While multiple paralogues were detectable for AP4 subunits (Figure 3.2), none of duplications that gave rise to them predated divergence of any of the plants analyzed (Appendix A chapter 3 Figure S2-11 to S2-14).

Finally, in mammalian cells, AP-5 functions in trafficking from late endosomes to the Golgi (Hirst *et al.*, 2018), but the orthologous complex remains entirely uncharacterized in plants. Overall, the pattern of evolution of AP5 in embryophytes is similar to AP-3 and AP-4. While multiple paralogues of AP-5 subunits were identified, these are all inparalogues of the genomes that we sampled (Appendix A chapter 3 Figures S2-15 to S2-17). An exception to the low number of identified paralogues of AP-5 subunits is the fifteen paralogues identified for AP-5 zeta in the genome of *S. moellendorffii* (Figure 3.2). This is in contrast with AP5 mu and beta subunits present in one and three copies respectively in *S. moellendorffii*.

3.3.5 Evolution of the COPI and TSET complexes in embryophytes

The COPI and TSET complexes perform comparable functions to the adaptor protein complexes in trafficking (Hirst *et al.*, 2014). Moreover, COPI and TSET are likely more closely related to each other than to any of the AP complexes (Klinger *et al.*, 2016b) (Figure 3.1B). COPI is highly conserved among eukaryotes, and has a well-characterized function in vesicle formation at the Golgi (Markus Boehm and Juan S. Bonifacino, 2001). The seven subunits of COPI are alpha, beta, beta prime (β), gamma, delta, epsilon, and zeta. Phylogenetic analysis of embryophyte orthologues of these subunits revealed relatively numerous duplications of COPI subunits at relatively early time points in embryophyte evolution. The COPI alpha and epsilon subunits were duplicated in the ancestor of gymnosperms (Appendix A chapter 3 Figure S2-20 and S2-24). COPI beta

prime, gamma, epsilon, and zeta subunits were duplicated in the ancestor of grasses (Appendix A chapter 3 Figure S2-22 and S2-24 to S2-26). COPI beta prime was duplicated in the ancestor of *Vitis*, *Cannabis*, *Capsella*, and *Arabidopsis*, but one of these paralogues was subsequently lost in the brassicalids (Appendix A chapter 3 Figure S2-22). Also, duplications of COPI alpha, beta prime, epsilon, and zeta occurred in an ancestor of *A. thaliana* and *Capsella* (Appendix A chapter 3 Figure S2-20, S2-22, S2-24, and S2-26).

The TSET complex was only recently discovered and characterized in *A. thaliana* (Gadeyne *et al.*, 2014) and the amoeba *Dictyostelium* (Hirst *et al.*, 2014). Unlike COPI, which is highly conserved among eukaryotes, TSET is frequently not identified and appears to have been lost in multiple lineages (Hirst *et al.*, 2014). TSET (also known as the TPLATE complex in *A. thaliana*) is involved in clathrin-mediated endocytosis along with AP-2 in *A. thaliana* (Bashline *et al.*, 2015), and is essential for growth and development of *A. thaliana* (Gadeyne *et al.*, 2014). The six subunits of the TSET complex are TPLATE (homologous to beta subunits), TSAUCER (homologous to AP-1 gamma and corresponding subunits in the other complexes), TCUP (homologous to mu subunits), TSPOON (homologous to sigma subunits), and TTRAY1 and TTRAY2 (possibly scaffolding proteins analogous to clathrin) (Hirst *et al.*, 2014). Phylogenetic analysis of each of these subunits in embryophytes revealed a duplication of TTRAY2 predating embryophytes (and multicellular green algae) (Figure 3.4). This supports and extends the initial result found by Hirst et al. which showed support for duplication of TTRAY2 at some point in embryophytes based only on analysis of *A. thaliana* and *Physcomitrella patens* sequences (Hirst *et al.*, 2014). This is the only TSET subunit duplication that we identified at this early time point.

Overall, among subunits of all the AP, COPI, and TSET complexes, we identified only 20 duplications that occurred prior to the divergence of the species for which we analyzed genomes (Appendix A chapter 3 Figure S2). Considering this and the number of paralogues identified here for each of the subunits (Figure 3.2 and Appendix A chapter 3 Table S3), there were at least 676 gene duplication events that occurred after the divergence of these species (*e.g.*, producing inparalogues in *A. thaliana* that are not represented in *C. grandiflora*). Thus at least 97.12% of the relevant duplications occurred recently in plant evolution.

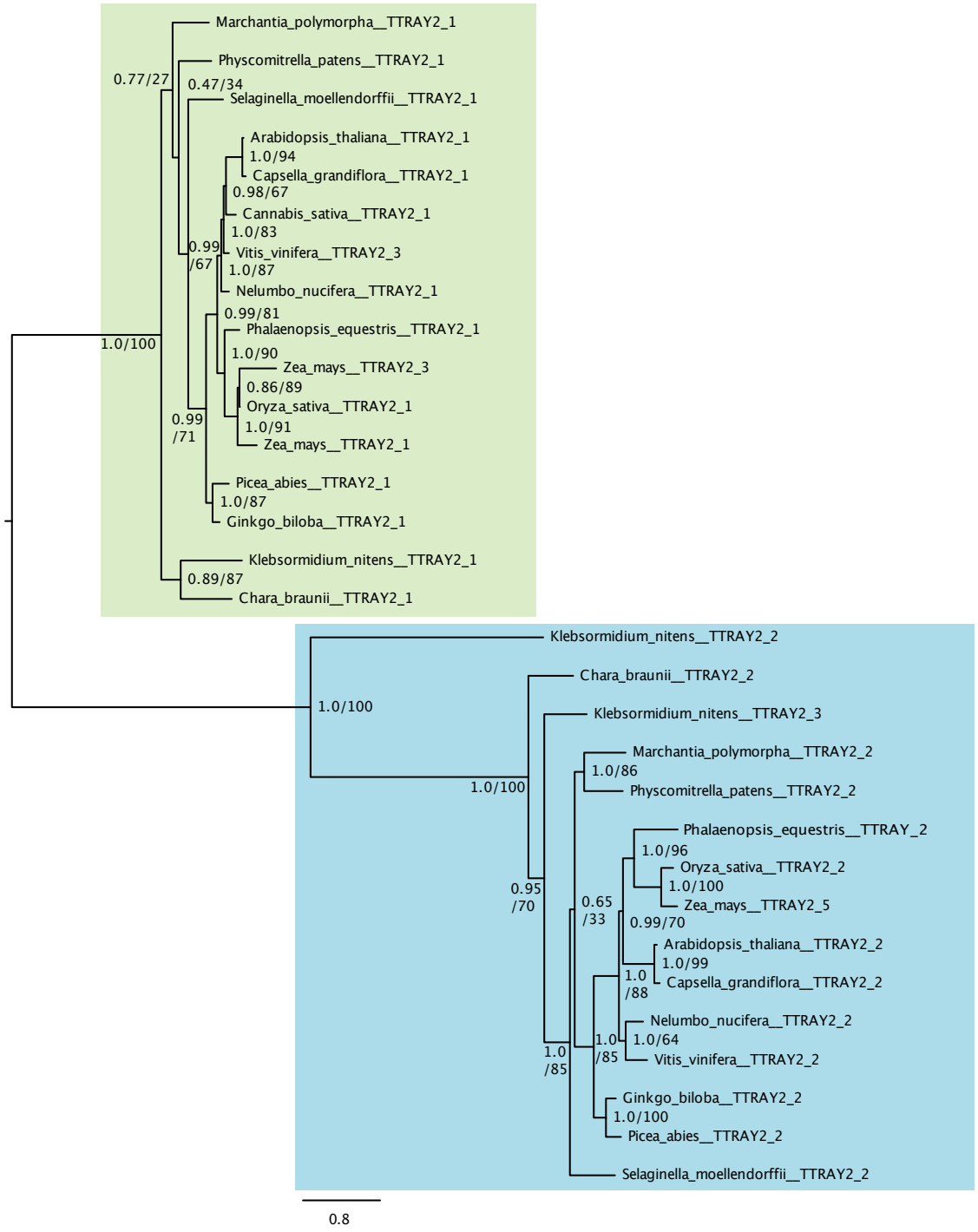


Figure 3.4: Phylogenetic tree of embryophyte TTRAY2 orthologues. Node supports indicate MrBayes Bayesian posterior probabilities and IQ-TREE bootstrap support values. A Bayesian posterior probability of 0.8 or larger and an IQ-TREE bootstrap percentage of 50 or larger indicated significant clade support. An early duplication of TTRAY2 (occurring before the divergence of *Klebsormidium nitens*) is highlighted by a green and blue box. This figure is reproduced from Larson *et al.* (2019) Figure 5.

3.4 Discussion

The AP, COPI, and TSET complexes are fundamental components of membrane trafficking machinery in plant cells, and predictions regarding specialized function of their subunits hinge on the timing of their duplications in embryophyte evolution. To investigate the evolution of these complexes in embryophytes, we performed searches for homologues of their subunits and performed phylogenetic analyses of each subunit of each protein complex. We applied these methods to a representative sampling of 13 embryophytes including 3 non-seed plants, which are often neglected in analyses of plant genomes (Rensing, 2017). We also analyzed the genomes of the multicellular green algae *Klebsormidium nitens* and *Chara braunii*, which are closely related to embryophytes (Wickett *et al.*, 2014), providing information about the earliest stages of embryophyte evolution. The results show overall that this evolution involved few ancient gene duplications (Figure 3.5) which were limited to subunits of the AP-1, COPI, and TSET complexes, but at least thirty times more duplications producing genome-specific inparalogues among all complexes (Appendix A chapter 3 Figure 3.2 and S2). This suggests that either the last common ancestor of embryophytes contained only a single copy of most AP, COPI, and TSET subunits, or that if duplicates were present, only one has been conserved in extant taxa.

In general, genomes of embryophytes exhibit a relatively large number of gene duplications, in many cases arising from whole genome duplications at both early and late time points in embryophyte evolution (Panchy *et al.*, 2016). Recent whole-genome duplications (WGDs) are important for adaptations of crops (Salman-Minkov *et al.*, 2016), while more ancient WGDs are associated with key events in the evolution of plants. The acquisition of seeds in vascular plants was coincident with a WGD (Li *et al.*, 2010). Angiosperms (flowering plants) diverged in the monocot-eudicot split (Li *et al.*, 2010; Shu-Miaw Chaw *et al.*, 2004), after which eudicots underwent a genome triplication before a period of rapid radiation. Similar to eudicots, there have been 3 common WGDs in the monocot lineage, the first occurring soon after monocot divergence, while the remaining two occurred in the Poaceae (grasses such as wheat, corn, and rice) (Jiao *et al.*, 2014). In our results, only 11 duplications occurred at similar timepoints to known early WGDs (Figure 3.5), possibly suggesting a role for segmental duplications in the evolution of APs, COPI, and TSET.

Analyses aimed at tracing the evolution of gene paralogues in plant genomes are complicated by these gene duplications, which result in more sequences requiring analysis, and by the varying quality and completeness of the genome assemblies used. We applied methods designed to mitigate these sources of error, and the script toolkit developed here to implement these methods may be

useful for future analyses of protein family evolution in embryophytes. To avoid under-estimating the age of paralogues due to their not being represented in the genomic assembly for specific genomes, taxa were sampled broadly throughout the embryophyte lineage. Also, to reduce false positives for paralogues due to the presence of inaccurate or redundant gene models, we did not count very short sequences or sequences with very high identity (over 98%) as separate paralogues. A caveat to these results then is that if improved genome assemblies are released in the future, the number of distinct paralogues that are identifiable may increase.

Our methods may have excluded some nearly identical duplicates that may be functionally distinct due to very few amino acid substitutions, or differences in expression patterns. Nevertheless, these duplicates would likely be the result of recent duplications and not ancient duplications. Moreover, true paralogues of adaptor protein complex subunits tend to share identities much lower than our cutoff of maximum 98% identity. For example, the percent identity between the two human AP-2 alpha paralogues (NP_055018.2 and NP_001229766.1) calculated by BLASTP is 80.12%. Similarly, the percent identity between human AP-1 beta (NP_001118.3) and AP-2 beta (NP_001025177.1) (which resulted from a duplication in vertebrate evolution) is 82.86%. Even the relatively recent brassicalid-specific AP-1/2 beta paralogues in *A. thaliana* (Figure 3.3) share only 92.9% identity (Appendix A chapter 3 Table S3).

Evolution of APs and related complexes plays a significant role in the cellular evolution of eukaryotic lineages. Due to the integral role of these complexes in membrane trafficking, the evolution of genes encoding the constituent protein subunits is indicative of specialization in the membrane trafficking system. Evolution of APs and TSET in other eukaryotic lineages often involves loss of entire complexes. For example, the membrane trafficking system in fungal cells may lack counterparts to mammalian early endosomes and recycling endosomes as distinct compartments from the trans-Golgi network (TGN) (Day *et al.*, 2018), and intriguingly all subunits of the AP-4 complex, which localizes to the TGN, are absent in yeast and many other fungi (Barlow *et al.*, 2014), suggesting a role for AP-4 in maintaining distinct endocytic compartments. In *Trypanosoma brucei*, the AP-2 complex is absent, and this loss may be connected to the origin of the variable surface glycoprotein coat that functions in immune evasion by this parasite (Manna *et al.*, 2013). Also, certain apicomplexan parasites have lost the AP-3 complex (Nevin and Dacks, 2009). Examples of differential function of paralogues include paralogues of the AP-1 mu subunit in mammalian cells, which have different functions, mediating traffic of different sets of cargo in epithelial cells (Fölsch *et al.*, 2003; Guo *et al.*, 2013), suggesting importance for evolution of distinct cell types in vertebrates. Also, the two AP-3 beta subunit paralogues in humans have

tissue-specific expression, and subsequently different disease phenotypes when mutated (Assoum *et al.*, 2016).

The scarcity of AP, COPI, and TSET subunit duplications occurring after the last embryophyte common ancestor, and before the divergence of two or more embryophyte species indicates that duplication and subfunctionalization of these proteins was not instrumental in establishing emb ryophyte-specific trafficking pathways that may be conserved across the diversity of extant embryophytes. Altogether 22 AP, COPI, or TSET subunit duplications occurred after the ancestor of the sampled embryophytes and related green algae and prior to the diversification of two or more sampled embryophyte species (Figure 3.5). With the exception of TTRAY2, clathrin LC, and COPI zeta, all of these duplications occurred in an ancestor of a specific lineage of embryophytes such as gymnosperms, Poaceae (grasses), or brassicalids. Within the order Brassicales, *C. grandiflora* and *A. thaliana* diverged approximately 11-14 million years ago (Koch *et al.*, 2001), which is recent in the context of the evolution of embryophytes, which originated at least 470 million years ago (Morris *et al.*, 2018). Nevertheless, the majority of distinct paralogues identified are accounted for by relatively recent (species-specific) duplications.

This result indicates similarity of the embryophyte ancestor to the last eukaryotic common ancestor with respect to APs, COPI, and TSET. This similarity is in contrast with other membrane trafficking protein families present in the embryophyte common ancestor such as the SNAREs KNOLLE (Sanderfoot, 2007), PEN1 (Sanderfoot, 2007), and phytolongins (Vedovato *et al.*, 2009), the Rabs novel plant Rab11 (Petrželková and Eliáš, 2014) and RabF1 (Petrželková and Eliáš, 2014), and the Exocyst tethering complex subunit Exo70 (Synek *et al.*, 2006).

Recent AP, COPI, and TSET subunit duplications are abundant in embryophytes, but the function of the resulting paralogues remain uncertain. Genes encoding 32 of the 36 subunits investigated exhibit species-specific duplications. This result leads to the question of why the paralogues identified in this study were retained. One possibility is that multiple copies of a gene may be retained simply because their promoters lose effectiveness in different cell types or developmental stages (Khoriaty *et al.*, 2018). For example, the two paralogues of the Sec23 subunit of COat Protein complex II (COPII) in mammalian cells, which originated from a duplication occurring early in vertebrate evolution, have different tissue-specific expression patterns, but indistinguishable interactomes (Khoriaty *et al.*, 2018). A corollary of this is that although there may be multiple paralogues present for a given complex subunit, they may not be expressed simultaneously in the same cells, therefore perhaps contributing little to complexity at the cellular level. Also, in *A. thaliana* the function of the two AP-1 mu paralogues so far appears to be redundant, although the

second paralogue (AP1M2) is more highly expressed and thus plays a more important role (Park *et al.*, 2013). In addition, paralogues resulting from recent duplications may have acquired distinct functions through either subfunctionalization or neofunctionalization (Conant and Wolfe, 2008). However, experimental characterization of AP, COPI, and TSET subunit duplicates identified in this study will be necessary to investigate the role, if any, of these duplications in adaptation of membrane trafficking (*e.g.*, to stresses).

The classification of AP, COPI, and TSET subunits in the broad sampling of plant taxa used in this study will be a resource for future studies, providing a foundation for both the generalization of functional characterizations of these complexes across embryophyte and broader eukaryotic model systems, as well as hypothesis generation regarding potential species-specific gene functions among embryophytes. The script toolkit developed to perform this study may also be of use for future investigations of protein family evolution, particularly where numerous duplicate genes are present as in plants. This study highlights several subunits as potentially relevant to early embryophyte evolution, and thus of interest for investigation of plant-specific membrane trafficking capacities. However, this study suggests that the majority AP, COPI, and TSET subunits in embryophytes, including the standard model plant *A. thaliana*, are likely to be representative of the function of their orthologues among eukaryotes and in the last eukaryotic common ancestor despite any unique features of the plant membrane trafficking system and the presence of multiple paralogues. Therefore, future investigations of AP, COPI, and TSET function in plant cells will allow for an increased understanding of membrane trafficking in both plants and other eukaryotes.

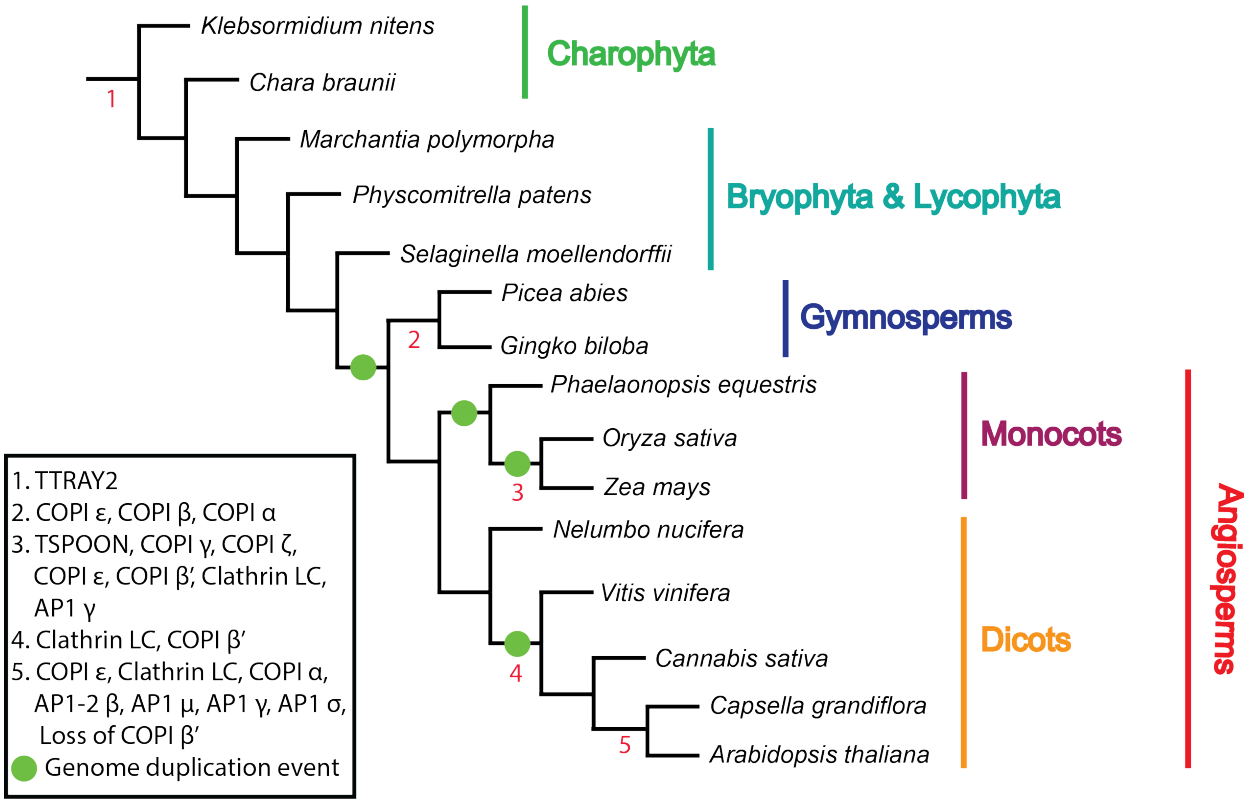


Figure 3.5: Timing of duplications of subunits of the Adaptor Protein (AP), COatomer Protein I (COPI), and TSET complexes mapped onto a species phylogeny. Numbers correspond with duplication events listed in the inset legend. Major known genome duplication events that occurred prior to divergence of the sampled species are indicated by green circles (Jiao *et al.*, 2014; Li *et al.*, 2015). The clathrin Light Chain (LC) duplication occurred twice at the third duplication point. The tree topology is based on that reported previously (Moore *et al.*, 2010; Puttick *et al.*, 2018; Zeng *et al.*, 2017).

3.5 Methods

3.5.1 Taxonomic sampling

To investigate the evolution of AP, COPI, and TSET subunits in embryophytes, we selected not only flagship angiosperm genomes, but embryophytes with sequenced genomes representing gymnosperms, lycophytes, and bryophytes, as well as closely related streptophyte algae (Appendix A chapter 3 Table S1). *Capsella grandiflora*, a close relative of *A. thaliana*, in the order Brassicales was included to more precisely time the relatively recent origins of *A. thaliana* AP, COPI, and TSET subunit paralogues.

3.5.2 Similarity searching and identification of orthologues

To identify orthologues of AP, COPI, and TSET subunits, similarity searches were performed with protein sequences from *A. thaliana* (TAIR10 from phytozome (Goodstein *et al.*, 2012) as queries (see Appendix A chapter 3 Table S2 for sequence identifiers), using the Basic Local Alignment Search Tool (BLAST)+ software package (Camacho *et al.*, 2009). Specifically, BLASTP was used to search in predicted peptide sequences, while TBLASTN was used to search in nucleotide sequences (chromosomes or scaffolds). WD40 repeats were removed from query sequences where present, to reduce the number of false-positives. Hits retrieved with an E-value of 0.0005 or lower were considered potential positive hits. For the TBLASTN hits, High-Scoring Pairs (HSPs) within 10,000 bp of each other, and on the same strand, were considered to be representative of a single gene locus. This allowed identification of multiple gene loci in larger nucleotide sequences, such as chromosomes. Translations of non-overlapping sets of the subject subsequences for HSPs clustered at each such locus were concatenated, taking into account the relative positions of the relevant subsequences in both the subject and query sequences.

These sequences constructed from TBLASTN results, as well as the subsequences of peptide sequences identified using BLASTP were then used as queries in further BLASTP searches back into the *A. thaliana* peptide sequences. Sequences that retrieved the original *A. thaliana* query sequences with an E-value of 0.05 or lower were considered potential positive hits for further validation. Moreover, the *A. thaliana* query and the identified sequence were required to be reciprocal best hits, with the identified sequence retrieving the query (or an equivalent sequence) with an E-value at least two orders of magnitude lower than that for any other sequence.

To identify peptide sequences for more potential divergent homologues, profile-based searches were performed using the HMMer3 software package version 3.1b1 (`https://hmmer.org`) (Eddy, 1998). Queries were prepared by aligning amino acid sequences for positive hits from BLAST

searches using MUSCLE v3.8.31 (Edgar, 2004) with default parameters. Again, similar E-value and reciprocal best hit criteria were applied.

BLASTP and HMMer hits with redundant sequence IDs were removed. Then any TBLASTN hits corresponding to the same locus as one of the peptide hits were excluded by searching for overlapping gene loci in the annotation (GFF3) files when available.

To facilitate the application of these search methods, scripts were written in the Python language, and these made use of libraries including Biopython (Cock *et al.*, 2009) and Matplotlib (Hunter, 2007). An archived version of these scripts is available at `https://zenodo.org/badge/latestdoi/186658027`, and the latest versions are available on GitHub (`https://github.com/laelbarlow/amoebae`).

3.5.3 Identification of distinct sequences among embryophyte inparalogues

First, for each protein of interest, BLASTP, TBLASTN, and HMMer hits were ranked by ascending E-value, and any hits with sequences shorter than 55 amino acid residues or 15% of the query length were excluded as these are not likely to be classifiable by phylogenetic analysis (see below). The resulting non-redundant sequences were compared by aligning the top ranked hit to a multiple sequence alignment of orthologues, and then aligning each of the subsequent hits separately to this alignment. From these alignments, several measures of similarity were calculated: 1) Whether the two sequences of interest overlap[13] at all in the alignment (indicating whether they correspond to an overlapping portion of the full length query), and 2) Percent identity over all the pairs of aligning residues between the two sequences of interest. Subsequent hits with no significant overlap with the top hit were excluded as it is impossible to determine whether these represent a distinct paralogue based on sequence similarity with the top hit. Also, subsequent hits with more than 98% identity with the top hit were excluded as redundant with the top hit. The top hit was then assumed to represent a distinct paralogue, and the remaining subsequent hits were processed in the same way until only likely distinct paralogues remained. The 98% identity cutoff was applied to reduce the possibility of counting redundant gene models that result from false segmental duplications that can occur when assembling diploid (or possibly polyploid) genomes (Kelley and Salzberg, 2010). Again, these methods were performed using custom Python scripts (`https://github.com/laelbarlow/amoebae`, `https://zenodo.org/badge/latestdoi/186658027`).

[13] See afterword (section 3.6) for further details.

3.5.4 Phylogenetic Analysis

Identified homologs were aligned using MUSCLE v3.8.31 (Edgar, 2004). The resulting alignments were then masked and trimmed to include only positions displaying homology. Appropriate amino acid substitution models were determined for each masked and trimmed alignment using ModelFinder (Kalyaanamoorthy *et al.*, 2017).

Phylogenetic trees were constructed using both IQ-TREE version 1.6.5 (Nguyen *et al.*, 2015) and MrBayes version 3.2.6 (Ronquist and Huelsenbeck, 2003) on the CIPRES server (Miller *et al.*, 2010) or on clusters provided by Compute Canada (`www.computecanada.ca`). IQ-TREE was used to construct maximum likelihood trees with bootstrap values calculated using 100 non-parametric bootstrap pseudoreplicates or 1000 ultrafast bootstraps. Bootstrap percentages of 50 or higher were interpreted as significant clade support for non-parametric bootstrapping and percentages of 95 or higher were considered significant for ultrafast bootstrapping. Posterior probabilities calculated in Bayesian analysis of 0.8 or higher were interpreted as significant clade support. Only MrBayes analyses run to convergence as indicated by an average standard deviation of split frequencies of less than or equal to 0.01 were considered acceptable.

To optimize phylogenetic resolution and determine the timing of relatively early gene duplications several iterative rounds of analysis with IQ-TREE using ultrafast bootstrapping were performed. In these analyses, sequences represented by relatively long branches in the phylogenetic trees or which branched (with significant support) as sister to other sequences from the same genome with shorter branches (therefore resulting from recent lineage-specific duplications) were removed from the alignments before performing the subsequent analyses. Specifically, long branches were identified as those with branch lengths over 1.5 interquartile ranges above the third quartile of all terminal branch lengths in the phylogeny. However, informative sequences that did not meet these criteria were retained after manual inspection of the phylogenetic analysis results. The resulting sequence alignments were then analyzed with MrBayes, and IQ-TREE non-parametric bootstrap percentages were mapped onto the MrBayes topologies.

3.6 Afterword

The results presented in this chapter clearly exclude the possibility of the origin of novel AP, COPI, or TSET complexes composed mostly or entirely of plant inparalogues of ancestral eukaryotic subunits originating early in plant evolution. Such a scenario would have conflicted with predictions of the OPH, or at least relevance of AP, COPI, and TSET complexes to understanding organelle evolution in terms of the OPH, given the lack of novel endomembrane organelles in plant

cells (Barlow and Dacks, 2018). Moreover, this pattern of ancestral conservation in embryophytes is further highlighted by comparison of the results for the AP-1/2 β subunit in the analysis of holomycotan (including fungal) sequences in Chapter 2 (see Figure 2.2 and Barlow *et al.* (2014) Figure 2) and of embryophyte sequences in this chapter (see Figure 3.3 and Larson *et al.* (2019) Figure 3). These show that while the AP-1 and AP-2 complexes of the model yeast *Saccharomyces cerevisiae* include dedicated β subunits (Rad *et al.*, 1995; Yeung *et al.*, 1999) which are derived from a duplication event pre-dating the last common ancestor of holomycota, duplicates of the ancestral AP-1/2 β subunit in the model embryophyte *Arabidopsis thaliana* arose very recently, consistent with an apparent lack of functional divergence (Teh *et al.*, 2013; Yamaoka *et al.*, 2013). Thus, the evolution of these complexes in plants is consistent with the OPH.

An issue that requires further discussion is that description of the similarity searching and sequence comparison methods employed by Larson *et al.* (2019) in section 3.5.3 above lacks certain details important for interpreting the results. First, section 3.5.2 describes the criterion that the sequences of secondary hits (among HMMer, BLASTP, and TBLASTN results) for a query (*e.g.*, the AP-1 gamma subunit) must overlap with a higher-ranking hit. However, it was not explained to what extent the sequences must overlap for the overlap to be considered sufficient to compare the similarity of the sequences based on that evident in the overlapping (aligned) regions. These minimum criteria were applied in combination: 1) 20 residues of each of the two sequences must be aligned, 2) 15 of these aligned residues must be similar (including identical) according to the Blosum62 scoring matrix, and 3) 10 of these aligned residues must be identical. These criteria were selected due to observation of some sequences which exhibited such low levels of alignment overlap, yet in the overlapping region were highly similar. In the absence of automated methods, these would most likely have been counted as potentially representing distinct genes upon visual inspection alone.

This combination of criteria may be problematic, however, because they could in principle be met even if one of the two sequences did not include homologous amino acid residues but non-homologous residues aligned spuriously. Also, short alignments on the order of 20 residues would yield very unreliable estimates of percent identity used as a basis to exclude some sequences as potential redundant sequences (such as alleles retained in the genome assembly). Both of these sources of error could be exacerbated by errors introduced to translations of nucleotide sequences by TBLASTN, because TBLASTN does not predict intron-exon boundaries and may include translations of non-coding regions. As a consequence, there is likely a minority of sequences identified as potentially representing distinct gene loci in this study, which may not represent separate gene

loci. In many cases, whether this is the case or not may be impossible to determine given the nature of the sequence data available. While some rate of false-positives is inevitable in studies of this nature, the methods employed here may be improved considerably to exclude more potential false-positive results. This is an important future direction which will involve further development of the AMOEBAE script toolkit, and possibly further evaluation of the analysis published by Larson *et al.* (2019).

Nevertheless, the main conclusion of this study is robust: There are far more recent duplications of genes encoding subunits of Adaptin, COPI, and TSET complexes, than duplications occuring early in embryophyte evolution. Using phylogenetic analysis, we only identified 20 gene duplications predating the divergences of species included in the analysis. While we determined that at least 676 duplications occured at later time points, even if this estimate were reduced by 50% this would overshadow the 20 early gene duplications identified.

Chapter 4

A sophisticated, differentiated Golgi in the ancestor of eukaryotes[14]

4.1 Preface

The Golgi was present in the LECA and existed most likely as a complex stack of cisternae, but has been dynamically re-organized in many independent lineages (Mowbrey and Dacks, 2009). Experimental work in several model systems has revealed functional differentiation between different cisternae that compose the Golgi. The OPH may be relevant to explaining the evolution of this differentiation, as the OPH has been suggested to be involved not only in organelle differentiation, but in differentiation of organelle sub-domains as well (Dacks and Field, 2007). Specifically, it was suggested that the evolutionary divergence of the COPI coat complex (with *cis* and *medial*-Golgi-specific function) from an ancestral AP complex (with *trans*-Golgi/early endosome function) might have been involved in the differentiation of the Golgi organelle into *cis* and *trans* Golgi. However, the Golgi of animals, fungi, and plants comprise three subtypes of cisternae: The *cis*, *medial*, and *trans* Golgi cisternae (Brigance *et al.*, 2000; Schoberer and Strasser, 2011). In mammalian cells the *medial* Golgi cisternae serve distinct functions in carbohydrate synthesis when compared to *cis* and *trans* cisternae (Day *et al.*, 2013). This raises the question of whether the OPH might yield insights into the differentiation of the Golgi into not two, but at least three sub-domains.

The golgin proteins have been identified as key players in organizing membrane trafficking at the Golgi, as well as in maintaining Golgi structure (Munro, 2011). Formally, golgins are a collection of 11 proteins in mammalian cells defined by the presence of coiled-coil domains (involved in forming golgin homodimers), attachment to Golgi membranes near their C-termini (either by tail-anchor transmembrane domains or through binding to small GTPases), and functions that include tethering and scaffolding (Munro, 2011). While multi-subunit tethering factors are not relevant

[14]The contents of this chapter, exclusive of preface and afterword, have been previously published as: Barlow, L.D., Nỳvltovà, E., Aguilar, M., Tachezy, J., and Dacks, J.B. (2018). A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biology*, 16(1). doi:10.1186/s12915-018-0492-9.

to understanding organelle evolution in terms of the OPH due to their apparent independent (non-paralogous) evolution (Dacks and Field, 2007; Koumandou *et al.*, 2007), the golgins are potentially relevant at least with respect to the question of differentiation between Golgi cisternae. Importantly, different mammalian golgins are involved in tethering distinct transport vesicles destined for specific Golgi cisternae (Wong and Munro, 2014; Wong *et al.*, 2017), and exhibit specificity for either the *cis*-Golgi, *medial*-Golgi rims, or *trans*-Golgi/*trans*-Golgi Network (TGN). For example, golgin-84 and TMF tether intra-Golgi transport vesicles along the cisternal rims (Wong and Munro, 2014). Also, some golgins are already known to be present in non-animal eukaryotes, with several having been localized in plant cells and shown to have similar localizations as compared to their mammalian orthologues (Latijnhouwers *et al.*, 2005; Renna *et al.*, 2005; Latijnhouwers *et al.*, 2007; Osterrieder, 2012). The OPH would predict that reorganization of the Golgi in independent lineages of eukaryotes would be associated with losses of golgins, while those with more canonical Golgi organization would generally retain an essential complement of golgins. Moreover, the evolutionary relationships between golgins, if any, were unclear, but the OPH would predict paralogous evolution of golgins with specificity to different cisternae. The analysis of golgins (and other Golgi proteins) in this chapter is thus relevant to understanding whether OPH-type protein evolution may explain the evolution of organelle sub-domains.

4.2 Background

At the intersection of the secretory and endocytic membrane-trafficking pathways in eukaryotes lies the Golgi. This organelle comprises a series of compartments termed cisternae, providing a platform for protein transport, glycosylation, and targeting. The Golgi is crucially important for normal cellular function, as demonstrated by the myriad diseases that result when genes associated with it are mutated (Bexiga and Simpson, 2013). The most salient hallmark of Golgi structure is the presence of multiple membraneous compartments, differentiated into *cis*, *medial*, and *trans*-Golgi, and organized into flattened stacks, which facilitates many key Golgi functions in mammalian cells (Zhang and Wang, 2016). In mammalian cells, numerous proteins are involved in maintaining the structure and positioning of the Golgi, as well as the specificity of membrane trafficking pathways to and from the Golgi (Munro, 2011), although the precise mechanism of Golgi stacking is unknown.

Golgins and Golgi reassembly and stacking proteins (GRASPs) are the main factors implicated in Golgi organization and stacking, as reviewed previously (Ramirez and Lowe, 2009). The golgins are a collection of 11 proteins in mammalian cells defined by the presence of coiled-coil domains,

attachment to Golgi membranes near their C-termini (either by tail-anchor transmembrane domains or through binding to small GTPases), and functions that include tethering/scaffolding (Munro, 2011; Witkos and Lowe, 2016). The domain topology and functions of mammalian golgins have been reviewed extensively elsewhere (Gillingham, 2017; Munro, 2011). Striking evidence for a role of GRASP 55 and 65, GM130, and golgin-45 in stacking was shown by a knock-sideways experiment demonstrating that ectopic expression of GRASP55 on mitochondria is sufficient for stacking of mitochondrial and Golgi membranes together (Lee and Hwang, 2014). A similar ectopic expression of golgin-84 on mitochondrial membranes also caused stacking of mitochondria (Wong and Munro, 2014). In addition to apparent roles in stacking, golgins including GM130 and golgin-84 are involved in tethering specific transport vesicles destined for different regions of the Golgi (Wong and Munro, 2014). Furthermore, several golgins including GM130 are involved in connecting the Golgi to the cytoskeleton (Kodani and Sutterlin, 2008; Rivero *et al.*, 2009). Various additional proteins have also been suggested to be involved in Golgi structure and organization (Appendix A chapter 3 Additional file 1: Table S1).

The integral role of golgins and other implicated structural proteins at the Golgi makes their evolutionary histories essential to reconstructing both the nature of the Golgi in the last eukaryotic common ancestor (LECA) 1.5 billion years ago (Eme *et al.*, 2014), and to tracing the subsequent changes that have occurred in the evolution of diverse eukaryotic lineages. While it has been inferred that the LECA possessed a stacked Golgi (Mowbrey and Dacks, 2009), whether there are pan-eukaryotic proteins (*e.g.*, golgins) that may have conserved roles in Golgi stacking remains unknown. Furthermore, the extent and details of golgin-mediated vesicle trafficking in the diversity of eukaryotes as compared with mammalian cells is also an open question.

Intriguingly, while Golgi stacking is observed in most organisms across eukaryotic diversity, there are a few lineages of microbial eukaryotes that lack stacked Golgi, as reviewed previously (Mowbrey and Dacks, 2009). In the absence of a morphologically recognizable Golgi, the question arose, for each of these lineages, as to whether the organelle i) was ever present, ii) is present but is no longer a feature of the cellular configuration, or iii) is present but has been shifted to an unrecognizable morphology.

Phylogenetic analysis to determine the evolutionary relationships of these organisms has placed them as embedded within various different eukaryotic groups, in almost every case having relatives with canonical stacked Golgi, rather than related to other organisms lacking stacks (James *et al.*, 2013; Janouškovec *et al.*, 2013; Karnkowska *et al.*, 2016; Tekle *et al.*, 2016). Furthermore, in every case yet examined, when genome-scale data became available, genes were identified that encode

orthologues of proteins that function at the Golgi in mammalian and yeast systems (Fritz-Laylin and Cande, 2010; Karnkowska *et al.*, 2016; Katinka *et al.*, 2001; Marti *et al.*, 2003). Localization data and functional assays have also confirmed that these proteins are expressed and indeed have shown that discrete Golgi, of morphologies other than stacked cisternae, exist in several of these lineages (Ghosh *et al.*, 1999; Marti and Hehl, 2003; Struck *et al.*, 2005). Recent genomic data for diverse eukaryotes, including from additional organisms with evidence for unstacked Golgi, therefore present the opportunity to understanding the evolution of Golgi structure across the broadest span of eukaryotes and organelle morphologies.

Herein we report an analysis of golgins and other Golgi-structure associated proteins across eukaryotes, using genomics, molecular cell biology, and bioinformatics techniques to address evolutionary cell biology of the Golgi in eukaryotes.

4.3 Results

4.3.1 The genome of the Golgi-less amoeba *M. balamuthi* encodes Golgi proteins

Genome sequences exist for 11 microbial eukaryotes with evidence for the presence of a Golgi, but presumably in an unstacked morphology. These organisms are spread throughout the diversity of eukaryotes (Appendix A chapter 3 Additional file 2: Figure S1), but in the supergroup Amoebozoa only one genus, the parasitic *Entamoeba*, has an unstacked Golgi which has been characterized to some extent (Ghosh *et al.*, 1999). *M. balamuthi* is a free-living anaerobic amoeba, related to Entamoeba, that lacks an identifiable stacked Golgi and that was at one time proposed to be lacking the organelle (Cavalier-Smith, 1987). To expand our sampling of eukaryotic genomes for this comparative analysis, particularly to increase taxon sampling in the Amoebozoa by adding a non-parasitic representative, we searched within the draft genome of *M. balamuthi* (see Methods) for genes that might indicate presence of a Golgi. A set of Golgi marker genes has been previously established to have been present in the LECA (Klute *et al.*, 2011), and also as present in the genomes of organisms that lack Golgi stacking (Dacks *et al.*, 2003; Fritz-Laylin and Cande, 2010; Karnkowska *et al.*, 2016; Katinka *et al.*, 2001; Marti and Hehl, 2003; Mowbrey and Dacks, 2009). Previously seven such proteins were reported for *M. balamuthi*, based on individual gene studies (Dacks *et al.*, 2003; Mowbrey and Dacks, 2009). We were able to expand this list to 22 proteins total (Fig. 4.1; Appendix A chapter 3 Additional file 3: Table S2), including the soluble *N*-ethylmaleimide sensitive fusion protein attachment protein receptor (SNARE) proteins Syn5, Syn16, and Sec22, the Retromer complex component Vps35, and the components of the multisubunit tethering complexes that act at the Golgi, COG and TRAPPII. This list also includes the

genes encoding the large subunits of the Adaptin 1, 3, and 4 complexes involved in transport from the *trans*-Golgi network (TGN), and the β-subunit of the COPI coat complex involved in intra-Golgi transport and traffic from the Golgi back to the Endoplasmic Reticulum (ER).

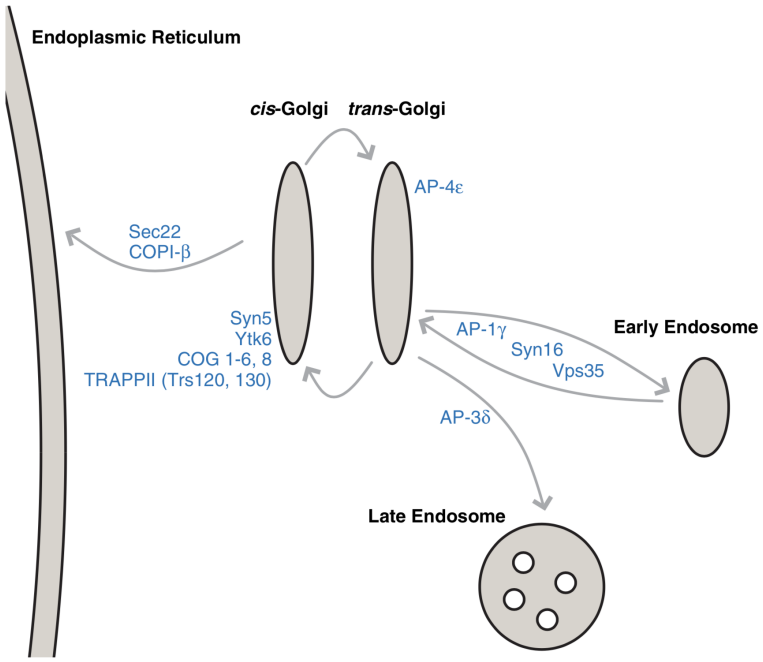


Figure 4.1: Diagram showing the Golgi marker genes found in *M. balamuthi* and their location in a generalized eukaryotic cell (see Appendix A chapter 3 Additional file 3: Table S2 for further details). Notably, we identified proteins with roles in vesicle fusion and formation, transport to and from the Golgi, and whose orthologues act at both the cis and trans faces of the organelle in other eukaryotes. Arrows indicate some membrane trafficking pathways that are reconstructed as likely present in the membrane trafficking system of *M. balamuthi*. This figure is reproduced from Barlow *et al.* (2018) Figure 1.

4.3.2 Golgi-like compartments in *M. balamuthi* are dispersed and punctate

To validate our genomic and informatics findings, we took a molecular cell biological approach. After further confirming orthology of the COPI-β orthologue in *M. balamuthi* by phylogenetic analysis (Appendix A chapter 3 Additional file 4: Figure S2), a specific antibody was raised and validated (Appendix A chapter 3 Additional file 5: Figure S3), and used for immunofluorescence light microscopy. This showed localization to discrete punctate structures scattered throughout the *M. balamuthi* cytosol, confirming expression of the protein and indicating a vesicular form of the organelle (Fig. 4.2-bottom row). We did not observe any association of Golgi with cytoskeletal structures of the microtubular conus around the cells multiple nuclei and microtubular fibres. The COPI complex mediates traffic from the Golgi to the ER in eukaryotic cells, and therefore the ER would be a likely location for the COPI complex, were a Golgi not present. To ensure that this was not the case, we co-localized the COPI-β with Protein Disulfide-Isomerase (PDI), a well-known ER marker. This showed PDI signal present in tubular structures close to nuclei as well as in numerous vesicles in the endoplasm, but little overlap with the COPI-β signal (Fig. 4.2-top row). Furthermore, since hydrogenosomes, the mitochondria-derived organelles in *M. balamuthi*, can also take the form of small discrete punctae (Nývltová *et al.*, 2013), co-localization experiments were performed (Fig. 4.2-middle row) showing no overlap between COPI-β and the hydrogenosomal marker Malate Dehydrogenase. Together these informatics and microscopy results are most consistent with the presence of a cryptic unstacked Golgi in *M. balamuthi*, and validate the inclusion of genomic information from this organism in our subsequent searches.

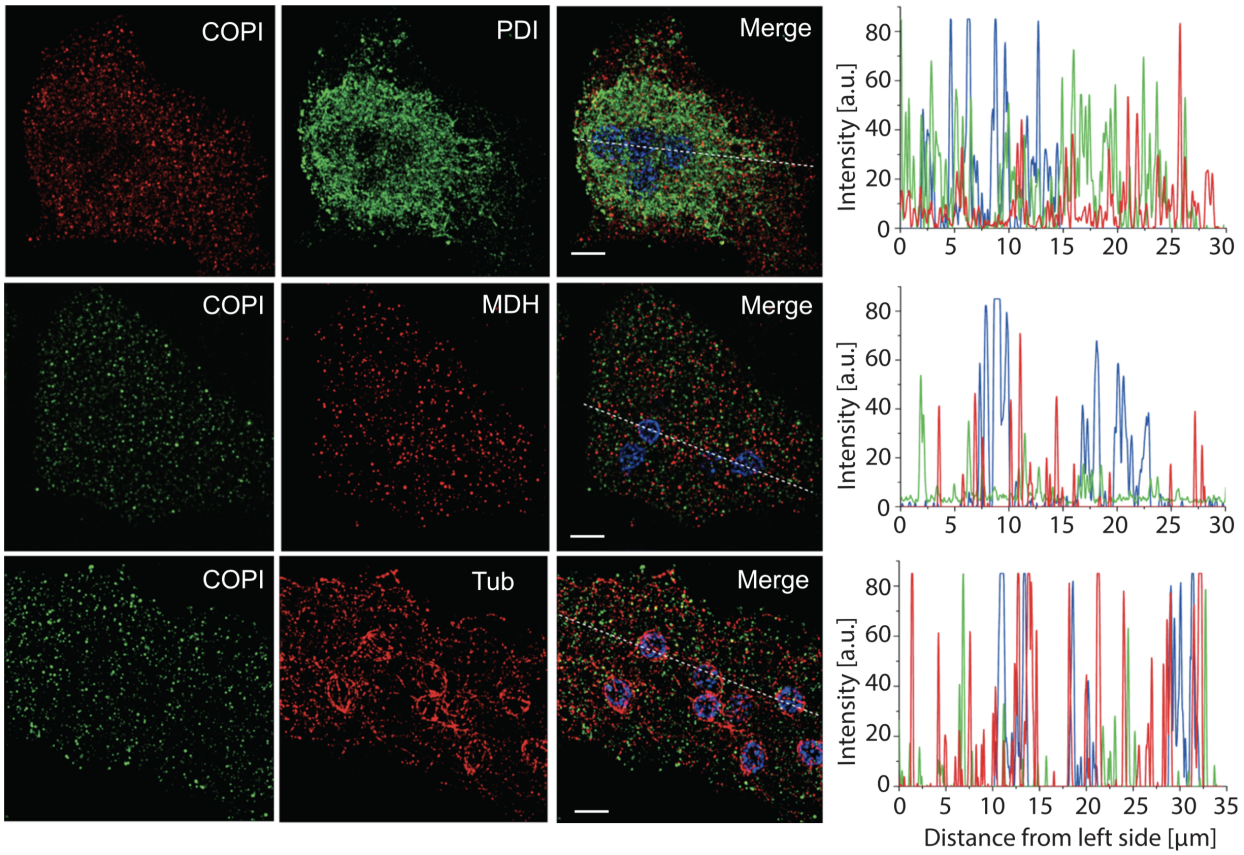


Figure 4.2: Localization for *M. balamuthi* COPI-β. Structured illumination microscopy of *M. balamuthi* labelled with antibodies against COPI and PDI (top row, ER structure), MDH (middle row, hydrogenosomes), and αtubulin (bottom row). The COPI signal is observed in numerous vesicles scattered within the *M. balamuthi* cells. α tubulin antibody labelled the tubular conus around nuclei and network of fibers. Signal for PDI network is concentrated around multiple nuclei. Graphs show line scans for fluorescence intensities corresponding to the dotted lines in merged images. Scale bar, 5 μm. This figure is reproduced from Barlow *et al.* (2018) Figure 2.

4.3.3 Evolution of the interacting Golgi structural proteins GM130, golgin-45, GRASP55, and GRASP65

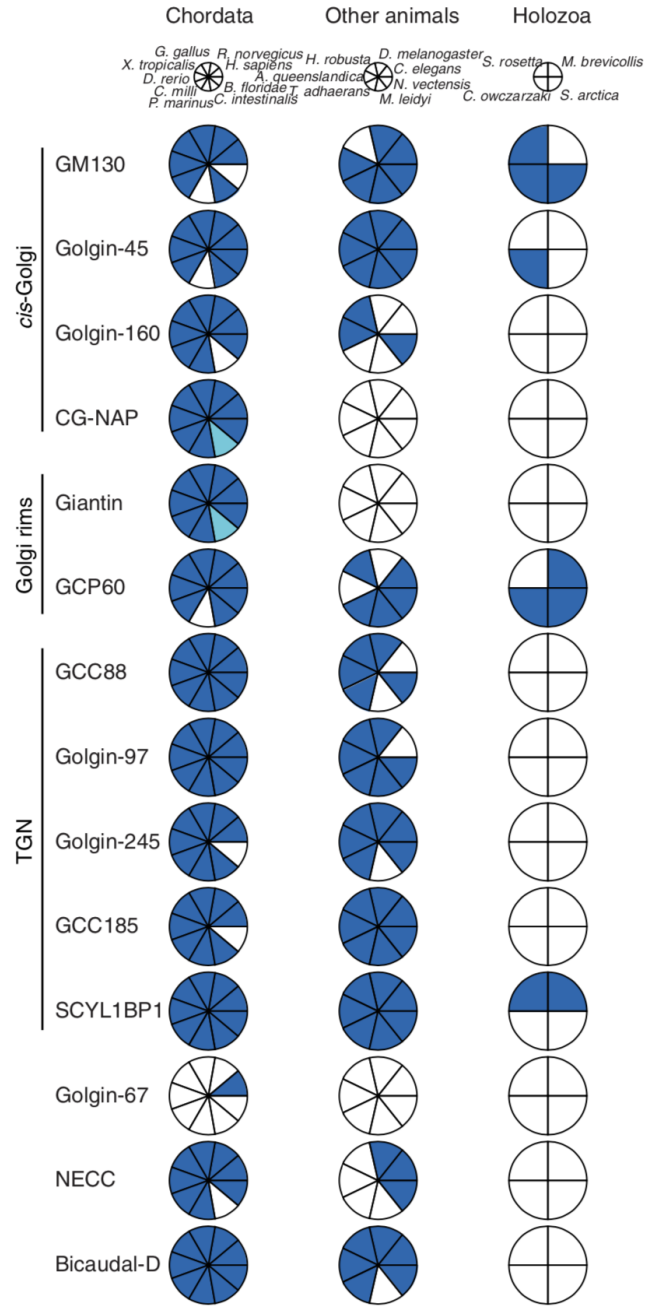
To understand the distribution and evolution of proteins with putative roles in Golgi stacking, we performed comparative genomic searches to assess the taxonomic distribution of mammalian golgins, as well as other Golgi proteins which are either golgin-like (*e.g.*, golgin-45), golgin-associated (*e.g.*, ZFPL1), or GRASPs (Appendix A chapter 3 Additional file 1: Table S1).

GM130, golgin-45, and GRASP 55 and 65 play key roles in Golgi stacking in mammalian cells (Lee and Hwang, 2014; Ramirez and Lowe, 2009). GM130 binds to GRASP65 at the *cis*-Golgi, while golgin-45 binds to GRASP55 at the *medial*-Golgi cisternae of mammalian cells (Barr *et al.*, 1998; Short *et al.*, 2001). Searches for GM130 and golgin-45 (Fig. 4.3A; Appendix A chapter 3 Additional file 2: Figure S1; Additional file 6: Table S3) revealed no homologues outside of animals and their single celled relatives (Holozoa). Consistent with previous efforts, our analysis did not identify the GM130 analogue Bug1p as a homologue of GM130 in *Saccharomyces* based on sequence similarity (Behnia *et al.*, 2007). Homologues of GRASP 55 and 65 have previously been identified in diverse eukaryotes and functionally studied in organisms both with canonical stacked Golgi (Ho *et al.*, 2006) and with unusual morphologies (Struck *et al.*, 2005). Consistent with this result and expanding upon it, we found that the duplication into GRASP 55 and 65 is a metazoan trait, predating the evolution of jawed fish (Appendix A chapter 3 Additional file 7: Figure S4), which means that all GRASP proteins in other eukaryotes are pre-duplicates of these two proteins. Also consistent with previous analyses (Kinseth *et al.*, 2007; Klute *et al.*, 2011), GRASP was found across eukaryotes (Fig. 4.4A; Appendix A chapter 3 Additional file 2: Figure S1; Additional file 6: Table S3) implying its presence in the LECA.

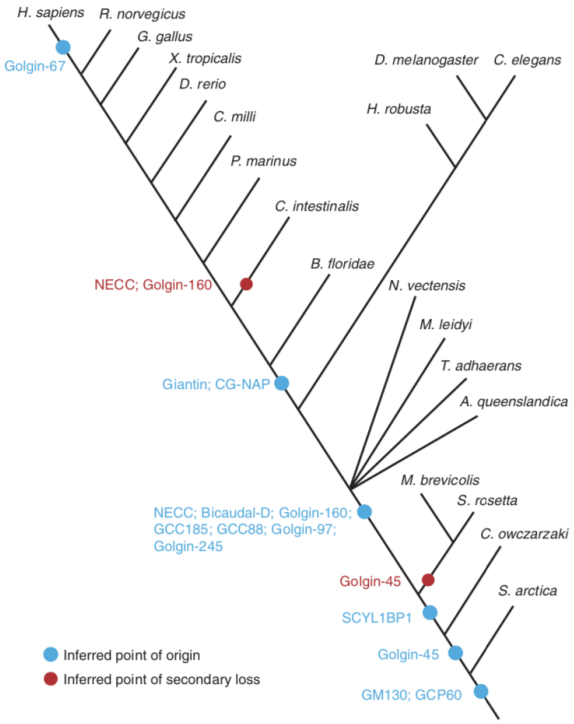
GRASP was not identified in many cases, most prominently in Embryophyta as previously noted (Kinseth *et al.*, 2007), and independently in Cryptophyta (*Guillardia theta*) as well as Rhizaria and Metamonada 4.4). Although we did not identify any GRASP homologues in the red alga *Cyanidioschyzon merolae*, we identified the following sequences in data from other red algae using the same methods: *Chondrus crispus* (XP_005713669.1), *Galdieria sulphuraria* (XP_005704721.1 and XP_005704722.1), *Porphyra umbilicalis* (OSX69770.1), and *Porphyridium purpureum* (ev m.model.contig_2019.4 from `http://cyanophora.rutgers.edu/porphyridium/`). Therefore, the ancestor of Archaeplastida plus Cryptophyta likely possessed a GRASP homologue, and multiple losses likely occurred, including in cryptophytes, glaucophytes, and *Cyanidioschyzon*. However, because cryptophytes and glaucophytes are represented in the analysis only by one exemplar genome per lineage, loss of a GRASP gene cannot be strongly inferred (and is thus not shown in

Fig. 4.4B).

The above observations suggest that the origin of both GM130 and golgin-45 predates the duplication that produced separate GRASP55 and GRASP65 paralogues, rather than coordinately appearing with them. Recent structural studies have elucidated the interaction between GRASP65 and GM130 (Hu *et al.*, 2015a), and between GRASP55 and golgin-45 (Zhao *et al.*, 2017). These suggest that these binding interactions involve specific residues near the C-terminus of GM130 and golgin-45 interacting with specific residues of GRASP65 and GRASP55, respectively. Evaluation of the conservation of these residues in vertebrates and non-vertebrate holozoan GM130 homologues reveals that residues near the C-termini that are important for binding to GRASP65 are contained in an extended region acquired in a vertebrate ancestor (Appendix A chapter 3 Additional file 8: Figure S5A). These residues include F975 and I990 of the human orthologue, which have been shown experimentally to be important for binding of GM130 to GRASP65 (Hu *et al.*, 2015b). GRASP65 may have become specialized for interaction with GM130 in vertebrates through corresponding amino acid substitutions. For example, M164 of GRASP65 is one of several residues that form a hydrophobic cleft occupied by the C-terminus of GM130 (Hu *et al.*, 2015b). However, while GRASP65 orthologues have either methionine or leucine residues at the position corresponding to M164, GRASP55 orthologues and pre-duplicate GRASP have tyrosine or phenylalanine residues (Appendix A chapter 3 Additional file 8: Figure S5B). Understanding whether GM130 interacts with preduplicate GRASP proteins in non-vertebrate metazoans will be an important point to resolve to understand both the evolution of Golgi and biology in species of ecological and agricultural importance.


Chordata
Other animals
Holozoa
G. gallus
R. norvegicus
X. tropicalis
H. sapiens
D. rerio
C. milli
P. marinus
B. floridae
C. intestinalis
H. robusta
D. melanogaster
C. elegans
A. queenslandica
N. vectensis
T. adhaerans
M. leidyi
S. rosetta
M. brevicollis
C. owczarzaki
S. arctica
cis-Golgi
Golgi rims
TGN
GM130
Golgin-45
Golgin-160
CG-NAP
Giantin
GCP60
GCC88
Golgin-97
Golgin-245
GCC185
SCYL1BP1
Golgin-67
NECC
Bicaudal-D

a


H. sapiens
R. norvegicus
G. gallus
X. tropicalis
D. rerio
C. milli
P. marinus
C. intestinalis
B. floridae
N. vectensis
M. leidyi
T. adhaerans
A. queenslandica
M. brevicollis
S. rosetta
C. owczarzaki
S. arctica
D. melanogaster
C. elegans
H. robusta
Golgin-67
NECC; Golgin-160
Giantin; CG-NAP
NECC; Bicaudal-D; Golgin-160; GCC185; GCC88; Golgin-97; Golgin-245
Golgin-45
SCYL1BP1
Golgin-45
GM130; GCP60
Inferred point of origin
Inferred point of secondary loss

b

Figure 4.3: Metazoa-specific golgin evolution. **a**) Coulson plot of Metazoa-specific golgin complement. Note that, for this figure and Fig. 4.4, filled pie sectors represent the positive identification of at least one orthologue (paralogue numbers are not shown). Light blue sectors indicate instances where an orthologue was not found in *Ciona intestinalis* but was found in the genome of a closely related ascidian. This representation is based on data shown in Appendix A chapter 3 Additional file 2: Figure S1 and Additional file 6: Table S3. **b**) Schematic showing timing of gains and losses of metazoan golgin genes. Note that, here and for Fig. 4.4, gene duplications yielding expanded complements are not tracked and losses are only inferred when a factor was not identified in more than one representative of a taxonomic group. This figure is reproduced from Barlow *et al.* (2018) Figure 3.

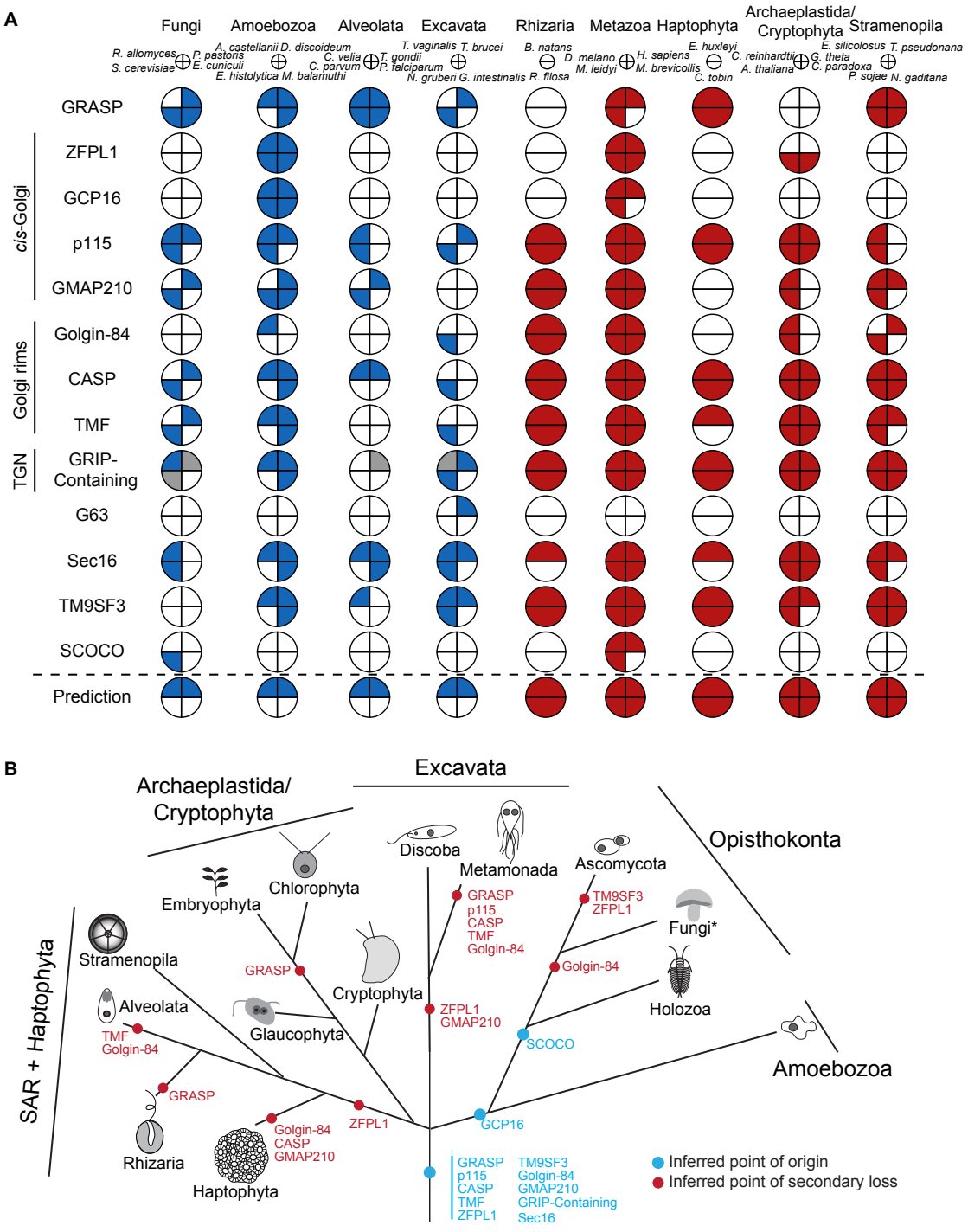

A
Fungi
Amoebozoa
Alveolata
Excavata
Rhizaria
Metazoa
Haptophyta
Archaeplastida/ Cryptophyta
Stramenopila
R. allomyces
S. cerevisiae
P. pastoris
E. cuniculi
E. histolytica
A. castellanii
D. discoideum
M. balamuthi
C. velia
C. parvum
T. vaginalis
T. gondii
P. falciparum
T. brucei
N. gruberi
G. intestinalis
B. natans
R. filosa
D. melano.
M. leidyi
H. sapiens
M. brevicollis
E. huxleyi
C. tobin
C. reinhardtii
A. thaliana
G. theta
C. paradoxa
E. silicolosus
P. sojae
T. pseudonana
N. gaditana
GRASP
ZFPL1
GCP16
p115
GMAP210
Golgin-84
CASP
TMF
GRIP-Containing
G63
Sec16
TM9SF3
SCOCO
Prediction
cis-Golgi
Golgi rims
TGN
B
Archaeplastida/ Cryptophyta
Excavata
Opisthokonta
Amoebozoa
SAR + Haptophyta
Embryophyta
Chlorophyta
Discoba
Metamonada
Ascomycota
Fungi*
Holozoa
Stramenopila
Alveolata
Glaucophyta
Cryptophyta
Rhizaria
Haptophyta
GRASP
p115
CASP
TMF
Golgin-84
ZFPL1
GMAP210
TM9SF3
SCOCO
GCP16
GRIP-Containing
Sec16
Inferred point of origin
Inferred point of secondary loss

Figure 4.4: Pan-eukaryotic Golgi protein evolution. **a**) Coulson plot of Golgi proteins found outside the Metazoa. Most importantly, while these represent ancient proteins, none show the phylogenetic pattern that would be expected for a necessary stacking factor, illustrated in the Prediction row. To clarify the patterns of presence and absence in organisms with stacked and unstacked cisternae, only selected genomes are shown here. The full data are given in Appendix A chapter 3 Additional file 2: Figure S1 and Additional file 6: Table S3. The first four columns (blue) show genes identified in organisms with unstacked Golgi, and closely related organisms with stacked Golgi, while remaining columns (red) indicate genes identified in representatives of taxonomic groups with stacked Golgi. Gray sectors indicate sequences identified using alternative methods (Appendix A chapter 3 Additional file 2: Figure S1). **b**) Schematic showing the timing of gains and losses of the proteins across eukaryotic evolution. Note that, if a single member of the taxonomic group possesses an orthologue of the protein, it is inferred as present in that group. Relationships between eukaryotes are based on recent concatenated phylogenetic results (Brown *et al.*, 2013; Burki *et al.*, 2016). To highlight losses in the Ascomycota, they are broken out to the exclusion of the paraphyletic remaining Fungi (denoted by the asterisk). This figure is reproduced from Barlow *et al.* (2018) Figure 4.

4.3.4 Evolution of *cis*-Golgi golgins

The *cis*-Golgi receives material through anterograde vesicle transport from the ER and in a retrograde fashion from the *medial*-Golgi and *trans*-Golgi/TGN. Multiple golgins are involved in tethering incoming vesicles at cis-Golgi cisternae. Although GM130 is Holozoa-specific, one of its interactors ZFPL1 (Chiu *et al.*, 2008) is more widely conserved and likely present in the LECA (Fig. 4.4A), consistent with previous identification of a homologue in Arabidopsis, which localizes to the cis-Golgi (Osterrieder, 2012). Similar to GM130, golgin-160 appears restricted to Metazoa, and was present in the earliest metazoans, despite being absent in *Drosophila* and *Caenorhabditis* (Fig. 4.3A). By contrast its binding partner GCP16 appears to be a more ancient invention, being found in opisthokonts and Amoebozoa (Fig. 4.4). Even more ancient still are p115 and GMAP210, the homologues of which are found across the diversity of eukaryotes and thus were likely present in the LECA.

Mammalian GMAP210 contains an N-terminal amphipathic alpha helix (ALPS domain), which is important for tethering ER-derived vesicles to the cis-Golgi (Drin *et al.*, 2007). Using the HeliQuest web service (Gautier *et al.*, 2008), we did not identify any such helices in the first 80 residues of GMAP210 sequences from non-vertebrates, suggesting that this is a lineage-specific mechanism for recognition of vesicles by GMAP210, consistent with previous observations (Wong *et al.*, 2017). Also, GMAP210 orthologues from non-holozoans do not share the N-terminal tryptophan-containing motif also shown to be involved in recognizing vesicles for tethering to the *cis*-Golgi (Wong *et al.*, 2017) (Appendix A chapter 3 Additional file 8: Figure S5C). This motif was previously shown to be necessary for tethering vesicles containing GalNAc-T2 and giantin, but not those containing golgin-84 instead (Wong *et al.*, 2017), which may indicate lineage-specific trafficking mechanisms as giantin is specific to chordates (Fig. 4.3B). Increased complexity of GMAP210-mediated trafficking pathways may be due to the presence of an ER-Golgi intermediate compartment (ERGIC) in metazoan cells, as GMAP210 has been shown to be involved in trafficking to both ERGIC and the *cis*-Golgi (Roboti *et al.*, 2015). In contrast to the N-terminal motifs, the Arf-binding GRAB domain of GMAP210 (Gillingham *et al.*, 2004) is conserved in orthologues across eukaryotes (Appendix A chapter 3 Additional file 8: Figure S5D).

4.3.5 Evolution of cisternal rim golgins

At least four golgins localize to the rims of Golgi cisternae (including *medial*-Golgi cisternae) in mammalian cells: golgin-84, CASP, TMF, and giantin. TMF and golgin-84 have direct roles in vesicle tethering, while giantin appears to be important for organizing Golgi cisternae (Koreishi

et al., 2013). Giantin is the most recently evolved, appearing in the chordates (Fig. 4.3). In contrast to previous suggestions that the *Drosophila* protein lava lamp is a giantin homologue (Kondylis and Rabouille, 2009), no homologues of giantin were identified in *Drosophila*. However, the origin of the giantin-interacting protein GCP60 (ACBD3) (Sohda *et al.*, 2001) (Appendix A chapter 3 Additional file 1: Table S1) predates that of giantin, having originated prior to the common ancestor of extant holozoans. Both CASP and golgin-84, however, appear to have been present in the LECA as they can be identified in taxonomically diverse eukaryotic genomes (Fig. 4.4A; Appendix A chapter 3 Additional file 2: Figure S1). While golgin-84 and CASP have been identified previously in plants (Latijnhouwers *et al.*, 2007; Renna *et al.*, 2005), we also identify orthologues of golgin-84 in Excavata, rhizarians, amoebozoans, and a basal opisthokont, and identify CASP in even more numerous taxa (Fig. 4.4; Appendix A chapter 3 Additional file 2: Figure S1).

Golgin-84, CASP, and giantin are anchored to the Golgi rims by transmembrane domains of similar length that share sequence similarity, even among mammalian and plant homologues (Gillingham *et al.*, 2002). Mutation of a conserved tyrosine in the transmembrane domain (TMD) of mammalian CASP prevents export from the ER, suggesting similar importance for this residue in the TMDs of golgin-84 and giantin (Gillingham *et al.*, 2002). In addition, residues within 100 residues immediately upstream of the TMD of mammalian golgin-84 and giantin, although dissimilar to each other, were shown to be involved in localization of these proteins to the Golgi (Misumi *et al.*, 2001). The TMD and 100 residues on the cytoplasmic side are sufficient for Golgi localization of the *Arabidopsis* orthologues of both golgin-84 (Latijnhouwers *et al.*, 2007) and CASP (Renna *et al.*, 2005). Here we confirm that the TMD and upstream cytoplasmic region of CASP and golgin-84 orthologues are conserved across eukaryotes, including Excavata (Appendix A chapter 3 Additional file 8: Figure S5E). These observations are consistent with conserved mechanisms of localization of golgin-84 and CASP within the Golgi, which would also have occurred in the LECAs Golgi.

Mammalian golgin-84 and TMF have previously been shown to contain tryptophan-containing N-terminal motifs similar to that of GMAP210 (Wong *et al.*, 2017). Like GMAP210, TMF does not show conservation of this motif outside of metazoans. In contrast, golgin-84 orthologues across eukaryotes contain comparable N-terminal motifs (Appendix A chapter 3 Additional file 8: Figure S5F). TMF shows conservation within the coiled-coil region that is thought to function in vesicle capture (Wong *et al.*, 2017) (Appendix A chapter 3 Additional file 9), as well as its C-terminal Rab6-binding domain (Fridmann-Sirkis *et al.*, 2004) (Appendix A chapter 3 Additional file 8: Figure S5G).

4.3.6 Evolution of *trans*-Golgi/TGN golgins

Mammalian GRIP (Golgin-97, RanBP2alpha, Imh1p and P230/golgin-245) domain-containing golgins at the trans-Golgi/TGN receive vesicles from various endosomal sources (GCC88, golgin-97, and golgin-245) (Cheung and Pfeffer, 2016; Wong and Munro, 2014). The presence of four distinct GRIP golgins in mammalian cells suggests that there might be multiple ancient GRIP golgin paralogues. This is, however, not what we observe. All four of the human GRIP golgins (the vesicle tethers and GCC185) appear to be restricted to metazoa (Fig. 4.3). Non-mammalian GRIP domain-containing proteins include the previously identified and characterized golgins Saccharomyces Imh1p (Munro and Nichols, 1999), *Arabidopsis* AtGRIP (Gilson *et al.*, 2004), and *Trypanosoma* TbGRIP (McConville *et al.*, 2002). We find GRIP domain-containing proteins across all supergroups (Fig. 4.4A; Appendix A chapter 3 Additional file 2: Figure S1).

Also, the coiled-coil domain-containing protein SCY1-like 1 binding protein 1 (SCYL1BP1) binds Rab6 at the *trans*-Golgi in mammalian cells, but has unknown function (Hennies *et al.*, 2008). The origin of SCYL1BP1 predated that of the choanoflagellate lineage of Holozoa (Fig. 4.3). A potential Arabidopsis homologue has been noted previously (Al-Dosari and Alkuraya, 2009). This protein was identified but did not meet the criteria for inclusion, whereas proteins that met the E-value cutoffs were identified here in *Guillardia* and *Bigelowiella* (Appendix A chapter 3 Additional file 6: Table S3). Nevertheless, whether these are true homologues remains ambiguous considering the short length of similar sequence regions, as well as the numerous independent gene losses implied by such a patchy distribution of homologues. Should these be true orthologues, then SCYL1BP1 would be deduced to have a much earlier evolutionary origin than stated. However, we suggest that conclusions regarding homology be reserved until functional characterization is available.

4.3.7 Evolution of additional proteins implicated in Golgi structure

Three golgin-like proteins with functions that have not been assigned to specific Golgi regions were also included in the analysis, and appear to have originated within the Holozoa or Opisthokonta. First, CG-NAP, a protein with function at both the Golgi and the centrosome (Takahashi *et al.*, 1999) (Appendix A chapter 3 Additional file 1: Table S1), originated prior to the divergence of Branchiostoma from other chordates. Second, homologues of NECC1/NECC2 were found to have an earlier origin, with identification of a homologue in *Nematostella* indicating that the origin possibly predated the diversification of the deepest-branching animal lineages (Fig. 4.3). Third, SCOCO, an Arl1/Arl3-binding protein of unknown function (Van Valkenburgh *et al.*, 2001;

Panic *et al.*, 2003), appears to be opisthokont-specific, with homologues only identified in fungi and Holozoa (Fig. 4.4; Appendix A chapter 3 Additional file 2: Figure S1).

Finally, three proteins of interest are relevant to evolutionary investigation of Golgi structure. First, the existence of metazoan-specific golgins suggested that lineage-specific golgin-like proteins may be present in other eukaryotic lineages as well. One such protein has already been identified in kinetoplastids, and the homologue in *Trypanosoma brucei* (TbG63) has been implicated in Golgi organization (Ramirez *et al.*, 2008). Our analyses found that this protein is present in the genome of *Bodo saltans*, the sister lineage to kinetoplastids, but not in any non-kinetoplastids (Appendix A chapter 3 Additional file 2: Figure S1). Second, although not localized to the Golgi, Sec16 has been shown to be widely conserved (Schlacht and Dacks, 2015) and important for Golgi stacking in the yeast *Pichia pastoris*, through its function in regulating COPII coat components at tER exit sites (Bharucha *et al.*, 2013; Connerly *et al.*, 2005). We recapitulate this finding, albeit with increased sampling. Finally, TM9SF3 is one of four widely conserved TM9 superfamily proteins (or nonaspanins)(Chluba-de Tapia *et al.*, 1997). It is not orthologous to EMP70 in *Saccharomyces*, but is instead most similar to human TM9SF4. Based on its exclusive Golgi localization and its loss of expression correlated with Golgi fragmentation in mammalian spermatids, TM9SF3 has been implicated in Golgi structure (Au *et al.*, 2015). Our analyses demonstrated that TM9SF3 is found across the span of eukaryotes though is not found in several taxonomically coherent groups, including ascomycete and basidiomycete fungi, ciliates, and apicomplexans (Fig. 4.4; Appendix A chapter 3 Additional file 2: Figure S1).

4.4 Discussion

4.4.1 *M. balamuthi* contains a cryptic Golgi

M. balamuthi was one of the organisms originally proposed to lack a Golgi, consistent with the idea at the time that it had diverged prior to the evolutionary emergence of the organelle (Cavalier-Smith, 1987). This idea of primitive Golgi absence has been fully disproven (Dacks *et al.*, 2003), and ultrastructural work has identified compartments proposed as candidate unstacked Golgi cisternae in some *Mastigamoeba* species (*M. balamuthi* was not imaged) (Walker *et al.*, 2001). Nevertheless, the possibility of complete absence of this organelle in any given organism remains viable, as was recently demonstrated for mitochondria (Karnkowska *et al.*, 2016). Our genomic and immunomicroscopy data suggests that *M. balamuthi* possesses a cryptic Golgi, possibly composed of distributed vesicles[15]. The precise form and dynamics of the organelle remain interesting open

[15]Here we use the term "Golgi" to mean a population of organelles physically separated from other organelles in the cell and being homologous to the described Golgi of other eukaryotes, retaining early secretory functions of Golgi

questions, ones that must await the technological development of better tools for molecular cell biology in this organism.

4.4.2 Holozoa-specific golgins reflect lineage-specific increases in trafficking complexity

Our comparative analyses identified a set of Golgi proteins that appear to have originated within Holozoa and which may reflect increased complexity of both vesicle traffic at the Golgi and connection to the cytoskeleton, relative to a pre-holozoan ancestor. N-terminal vesicle recognition motifs present in mammalian orthologues of GMAP210, TMF, and GRIP golgins, but absent outside of Holozoa suggest potential gain of tethering functions in these proteins relative to the ancestral sequences. As well, several of the proteins originating within Holozoa, for which functional information is available have roles in tethering the Golgi to the cytoskeleton: golgin-160 (Yadav *et al.*, 2012), GM130 (Rivero *et al.*, 2009), GCC185 (Efimov *et al.*, 2007), CG-NAP (Rivero *et al.*, 2009), and bicaudal-D (Hoogenraad *et al.*, 2003). Cytoskeleton-dependent Golgi positioning along microtubules is important for cellular functions that are essential to metazoan multicellularity including wound healing (Yadav *et al.*, 2009). This may explain the relatively recent origin of some of these factors. Despite animal-specific gains in complexity, other eukaryotes may also exhibit comparably complex Golgi. One possibility is that proteins such as TbG63 as well as undiscovered Golgi proteins in other eukaryotic lineages reflect parallel increases in complexity, which cannot be inferred by characterization of homologues of human Golgi proteins.

4.4.3 Conservation of golgins suggests differentiated Golgi compartments were present in the LECA

Counter to the intuitive idea that the ancient ancestor of eukaryotes was simple, molecular evolutionary reconstruction of the LECA has revealed a complement of cell biological machinery that is consistent with a highly complex cell. This applies not only to membrane-trafficking but also to nuclear proteins, the cytoskeleton, mitochondria, and metabolism (Koumandou *et al.*, 2013). The set of pan-eukaryotic Golgi-structural proteins that can be deemed as ancient, which we identify here, adds to this ancestral complexity. This has important implications for the complexity and organization of the Golgi in diverse eukaryotes and in the LECA. The presence of proteins such as p115 and ZFPL1 in non-metazoan eukaryotes raises important questions about Golgi function to be explored in those organisms, given that known binding partners of those proteins are

such as glycosylation, and being the active site of homologues of Golgi proteins such as the COPI complex. This is in contrast to the morphological definition of being composed of flattened, stacked cisternae, which was used by authors such as Walker *et al.* (2001).

metazoa-specific. Evolutionarily, although homologues of p115, GMAP210, golgin-84, CASP, TMF, ZFPL1, and GRIP-containing golgins have been previously identified and localized in plant cells (Latijnhouwers *et al.*, 2005, 2007; Osterrieder, 2012; Renna *et al.*, 2005), identification of homologues in the extensive taxonomic sampling used here confirms that these were present in the LECA for two reasons. First, it makes the possibility of lateral gene transfer even less likely. Second, identification of CASP, golgin-84, TMF, p115, and TM9SF3 in excavates (*Naegleria gruberi* in particular) provides evidence that they were present in the LECA regardless of uncertainty in the rooting of the eukaryotic tree (Burki *et al.*, 2016; Derelle *et al.*, 2015; He *et al.*, 2014).

Based on the data collected in metazoan model organisms, and assuming functional homology, the presence of at least four factors at the *cis*-Golgi (p115, GRASP, ZFPL1, and GMAP210) and three at the Golgi rims of successively later cisternae (golgin-84, CASP, and TMF) suggests that the Golgi had differentiated into at least three regions (Fig. 4.5). And, the conservation of specific sequence motifs provides further evidence for this. The presence of Sec16 which is involved in vesicle formation at ER exit sites and GMAP210 which receives vesicles from the ER, together with the well-established ancient nature of the COPII coat (Schlacht and Dacks, 2015), provides detail of the anterograde trafficking pathways coming into the *cis*-Golgi (Fig 5). Conservation of the Arf binding GRAB domain in GMAP210 (Appendix A chapter 3 Additional file 8: Figure S5D) and the previously identified conservation of Arf in eukaryotes, including representatives of Excavata (Li *et al.*, 2004), and localization of GMAP210 to the Golgi in *Arabidopsis* (Latijnhouwers *et al.*, 2007) are consistent with conservation of GMAP210 function from the LECA. Tryptophan-containing N-terminal motifs in golgin-84 orthologues from across eukaryotes, and in key residues in its transmembrane domain suggest a widely conserved role in intra-Golgi vesicle traffic to the Golgi rims. Similarly conservation of likely vesicle tethering motifs in TMF suggests a vesicle tethering role for TMF at rims of cisternae closer to the *trans*-Golgi. Again, conservation of Rab6 (Elias *et al.*, 2012), and the Rab6 binding domain of TMF is consistent with this as well (Appendix A chapter 3 Additional file 8: Figure S5G).

With respect to established TGN compartments, the only inferred LECA golgin at the TGN is a GRIP domain-containing golgin, which acts to receive vesicles from endosomes. The presence of a GRIP domain in proteins across eukaryotic diversity, and the localization of these GRIP domain-containing proteins at the TGN in yeast, plants, and trypanosomes (Latijnhouwers *et al.*, 2005; McConville *et al.*, 2002; Munro and Nichols, 1999) suggests some conserved TGN function from the LECA. The previously identified conservation of Arl1 in eukaryotes, including the representatives of the Excavata, is consistent with conserved function of GRIP golgins (Li *et al.*, 2004).

However, the lack of clear conservation of multiple TGN golgins suggests that vesicle traffic to the *trans*-Golgi in non-metazoan cells, and in the LECA, involves fewer specialized tethers, and possibly fewer types of transport vesicles. This could also be reflective of the variation of TGN organelles across eukaryotes.

Previous reconstruction of trafficking pathways as present in the LECA, for example via analysis of COPI, COPII, Retromer and AP1,4 complexes, as well as Golgi-specific SNARE proteins (Dacks and Doolittle, 2001, 2004), had suggested potential differentiation of Golgi compartments to some degree. However, these did not indicate whether the ancestral Golgi was a single compartment with specialized domains, or was composed of differentiated cisternae. The presence of at least 8 ancient proteins implicated in Golgi structure at cis-Golgi, cisternal rims, or *trans*-Golgi/TGN, along with conservation of several functional motifs that mediate interactions with bindings partners (*e.g.*, Rab6, Arl1, Arf) also reconstructed as present in the LECA, shows that the LECA Golgi was much more complicated than it has been previously possible to infer (Fig. 4.5). Conservation of golgin-84 and TMF is particularly relevant, as they are specific to intra-Golgi vesicle traffic, which would be arguably unnecessary if Golgi cisternae were not differentiated.

4.4.4 Golgi stacking is likely an ancient, emergent property

Our analyses also speak to the cell biological question of how Golgi stacking takes place today which, despite its importance and the apparent conservation of the stacked morphology of the organelle, remains a matter of significant debate (Zhang and Wang, 2016). The predominant paradigm is that one or more Golgi-localized proteins are necessary for the morphology. Given the presence of Golgi stacking across eukaryotes, such a protein could well be predicted to be universal. However, it is not known which proteins, if any, may be necessary for a conserved pan-eukaryotic mechanism of stacking.

By contrast with this paradigm, other suggestions have been put forward to explain Golgi stacking as a morphological property based on several combined factors. This idea has most explicitly been laid out by the cisternal adhesion model of Lee and Hwang (2014), whereby one or more proteins with adhesive functions have a stacking effect when present in sufficient quantities. Stacking could also involve regulation of membrane flux through the Golgi, with insufficient input or replenishment as compared to output causing dissolution of stacks (Kühnle *et al.*, 2010). A model of additive effects of redundant proteins or membrane flux is also consistent with the phenotypes observed in knockouts of retromer components that result in depleted retrograde trafficking from the endosomes to the TGN and fragmentation of the Golgi (Koumandou *et al.*, 2011; Seaman,

2004). The idea that properties of organelles, including Golgi stacking, are dependent on systems-level properties is gaining traction as a viable alternative to exclusively genetic explanations (Mani and Thattai, 2016). We denote these hypotheses collectively as Golgi stacking being an emergent property. Overall, the question of how the hallmark morphology of the organelle is established and maintained remains an open one.

Under the paradigm of a protein with a conserved necessary function in Golgi stacking, it would be likely present in all genomes of organisms showing Golgi stacking, and likely absent from the genomes of those organisms without (*i.e.*, the taxonomic distribution of stacking factors should match that of Golgi stacking). Such a pattern, of presence directly correlating with function, has been observed for protein complexes responsible for cristae formation in mitochondria (Munoz-Gomez *et al.*, 2014), and this phylogenetic screening approach has successfully identified proteins involved in flagellar function (Avidor-Reiss *et al.*, 2004; Carvalho-Santos *et al.*, 2011). The evolutionary analyses performed here across 75 taxa with stacked Golgi and 12 without showed that none of the 27 putative stacking factors that we examined matched this pattern.

There are several caveats to our results. First, individual false positives, or false negatives, are always possible in comparative genomic analyses. Nonetheless, we have used the most accurate homology searching methods, examined datasets of alternate protein models for genomes when relevant and have manually curated the gene assignments. Second, it is conceivable that a universal and necessary stacking gene could exist that possesses multiple functions and so had lost the relevant Golgi function in organisms with unstacked Golgi. However, the fact that every candidate protein examined was apparently absent in multiple genomes of organisms that possess Golgi stacks renders this possibility incompatible with our observations. Finally, it is possible that there exists a protein that is a necessary stacking factor, that has not yet been reported, and therefore for which we did not search. Proteomics technology allowing distinction between the proteomes of organelles with similar densities, such as the plant ER and Golgi, and even the unique proteomes of organelle sub-compartments (Parsons and Lilley, 2017) may identify previously uncharacterized Golgi proteins that could be candidates for such a necessary stacking factor.

However, accepting these caveats, our results are inconsistent with the hypothesis that any one of the proteins participates in a pan-eukaryotic mechanism of Golgi stacking. This does not discount the importance of lineage-specific functions. Nonetheless, our data are most consistent with Golgi stacking being dependent on additive, redundant function of non-homologous proteins, *i.e.* the emergent property hypotheses.

An emergent property could rely on ancient redundant proteins, or could rely upon recently

evolved, lineage-specific ones that replace ancient factors. With 14 recently evolved proteins identified within the Holozoa (Fig. 4.3), it is tempting to speculate that additional lineage-specific proteins are also present in other eukaryotes, and may have stacking functions. The presence of a kinetoplastid and bodonid-specific protein (TbG63) is consistent with this scenario, and certainly searches for lineage-specific membrane-trafficking factors associated with clathrin-mediated endocytosis (Adung'a *et al.*, 2013) and the sortilin system (Briguglio *et al.*, 2013) have been fruitful and illuminating. This will be exciting to pursue in order to understand the mechanisms of Golgi trafficking and stacking, particularly as more genetic and molecular biological tools become available for non-opisthokont model organisms.

Overall, our data do not rule out the existence of a widely conserved necessary stacking factor, but rather support the idea that Golgi stacking as an emergent property needs to be more extensively explored. This may well be the key to understanding one of the most prominent eukaryotic cellular features.

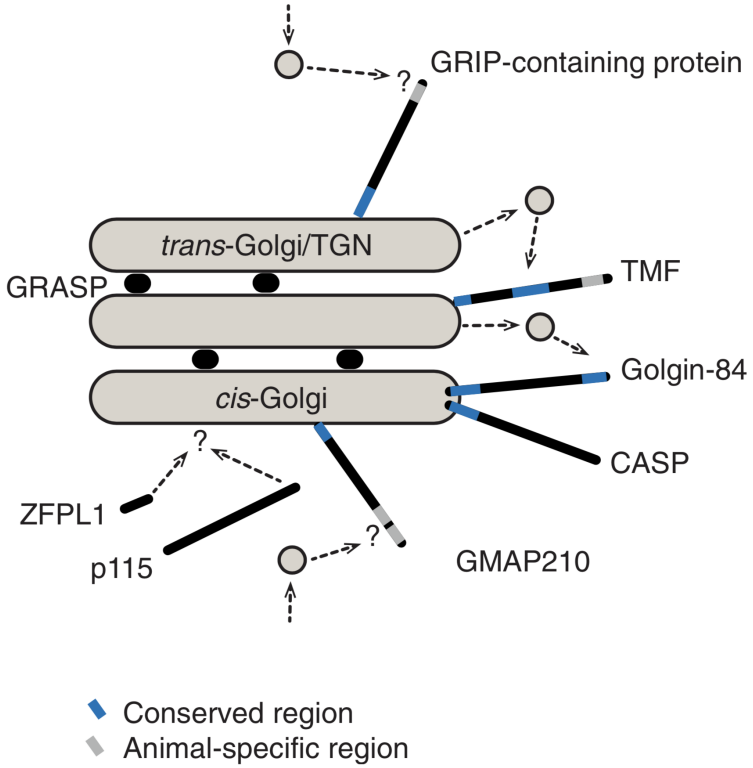


Figure 4.5: Golgi structure proteins inferred to be present in the LECA. Functional domains and motifs conserved in animals or conserved in the LECA are color coded as inset, and inferred membrane trafficking pathways are shown. Other Golgi proteins were also identified as present in the LECA: TM9SF3 and Sec16. However, their role, if any, in differentiating separate Golgi compartments is unknown. This figure is reproduced from Barlow *et al.* (2018) Figure 5.

4.5 Methods

4.5.1 Cell cultivation

M. balamuthi strain (ATCC 30984) was maintained axenically in PYGC medium at 24C in 50 ml culture tissue flask (Chavez *et al.*, 1986). For immunofluorescence microscopy, *M. balamuthi* cells were fixed in 1% formaldehyde for 30 min, washed, and treated in 1% Triton TX-100 for 10min. Fixed cells were stained using polyclonal rat α COPI β subunit, rabbit α PDI, rabbit α MDH (Nývltová *et al.*, 2013) Abs, and monoclonal mouse α tubulin (Sigma) Ab. Alexa Fluor 488 (or 594) donkey α-rabbit, Alexa Fluor 594 (or 488) donkey α-rat, and Alexa Fluor 594 donkey α mouse Abs (Life Technologies) were used as secondary antibodies. Structured illumination microscopy (SIM) was performed using a commercial 3D N-SIM microscope (inverted Nikon Eclipse Ti-E, Nikon) equipped with a Nikon CFI SR Apo TIRF objective (100x oil, NA 1.49). A structured illumination pattern projected into the sample plane was created on a diffraction grating block (100 EX V-R 3D-SIM) for laser wavelengths 488, 561 nm. Excitation and emission light was separated by filter cubes with appropriate filter sets SIM488 (ex. 470-490,em. 500-545), and SIM561 (556-566, 570-640). Emission light was projected through a 2.5x relay lens onto the chip of an EM CCD camera (AndoriXon Ultra DU897, 10 MHz at 14-bit, 512x512 pixels). Three-color z-stacks (z-step: 120 nm) were acquired in NIS-Elements AR software (Laboratory Imaging). Laser intensity, EM gain and camera exposure time were set independently for each excitation wavelength. Intensity of fluorescence signal was held within the linear range of the camera. Fifteen images (3 rotations and 5 phase shifts) were recorded for every plane and color. SIM data were processed in NIS-Elements AR. Before sample measurement, the symmetry of point spread function was checked with 100 nm red fluorescent beads (580/605, Carboxylate-Modified Microspheres, Life Technologies) mounted in Prolong Diamond Antiface Mountant (Life Technologies), and optimized by adjusting objective correction collar. Signal for 4,6-Diamidine-2-phenylindole dihydrochloride (DAPI) was observed in widefield mode.

4.5.2 Preparation of antibodies

To obtain complete and partial recombinant PDI and COPIβ proteins, respectively, the corresponding gene sequences were amplified by PCR (Primers: COPIβ-forward: CATATGAAGAAC-CTCGAGCACAGG, COPIβ-reverse: AAGCTTCGCGTCGGCCTTGA; PDI-forward: CATAT-GAAGTGGCAGTACATCG, PDI-reverse: AAGCTTGAGCTCCTTCTTCTCCCC) using *M. bal-amuthi* cDNA as template. The PCR products were subcloned into the pET42b+ vector (Novagen), and expressed with a 6xHis tag in *Escherichia coli* BL21 (DE3). The proteins were purified by

affinity chromatography under denaturing conditions according to the manufacturers protocol (Qiagen) and used to immunize rats (COPIβ) or rabbits (PDI).

4.5.3 Similarity searches

The genomic databases used for bioinformatics searches are listed in Appendix A chapter 3 Additional file 10: Table S4. Of note, both the filtered and unfiltered gene model databases at JGI were searched (unfiltered datasets include any redundant gene models for the same gene loci). Additionally, the draft genome of *M. balamuthi*, produced as part of an ongoing project was searched for conserved Golgi marker and putative stacking factor genes. The draft genome sequence is available at (`http://www.ebi.ac.uk/ena/data/view/CBKX00000000` (deposited January 22, 2015)). The identified gene sequences are detailed and made available in Appendix A chapter 3 Additional file 3: Table S2.

Basic Local Alignment Search Tool (BLAST 2.2.29+) (Camacho *et al.*, 2009) was used to search for homologues of proteins of interest in *M. balamuthi* predicted proteins. A bidirectional best-hit criterion was applied with an E-value cut-off of 0.05 for both forward and reverse searches. Also, identified sequences were required to retrieve the original query in the reverse search with an E-value of at least two orders of magnitude lower than other sequences. Initial queries are either from the *H. sapiens* or *S. cerevisiae* genomes, or are from other eukaryotes as identified in previous studies (Hirst *et al.*, 2011, 2014; Koumandou *et al.*, 2011; Murungi *et al.*, 2014), and multiple queries were used.

For searches to identify orthologues of Golgi-structure associated proteins of interest, a multiphase approach was taken. BLAST was run locally to search protein sequence databases from a large sampling of eukaryotes (Appendix A chapter 3 Additional file 10: Table S4). To identify highly similar homologues, reciprocal best hit BLASTP searches were performed using *H. sapiens* query sequences and with the following criteria: E-value of 1e-20 or lower for forward search, E-value of 0.05 or lower for reverse search, and a minimum E-value difference of two orders of magnitude, in the reverse BLAST results, between the hit(s) corresponding to the original query, and the first negative hit.

HMMER 3.1b1 was then used to perform searches in the same protein sequence databases (`http://hmmer.org`) (Eddy, 1998). For this, positive hits from BLAST searches were used to build initial Hidden Markov Models (HMMs). Sequences were aligned using MUSCLE v3.8.31 (Edgar, 2004) with default parameters. For these searches, the following criteria were applied to define positive hits: E-value of 1e-10 or lower for forward (HMMer) search, E-value of 0.05 or

lower for reverse (BLASTP) search. After each HMMer search, positive hits, if identified, were aligned and viewed manually before inclusion in HMMs for subsequent searches. This process was repeated until no more positive hits were identified. An exception to these methods was made in the case of the GRIP domain-containing proteins in taxa outside of Metazoa, which were identified using HMMs including only the subsequence of proteins corresponding to the GRIP domain, because no proteins with sequence similarity to individual human GRIP containing proteins outside the GRIP domain were identified outside metazoan taxa. In addition to the above methods, for these non-metazoan GRIP golgins, due to the short length and high sequence conservation of the GRIP domain, a bit score of 25 was used as a cutoff to identify positive hits, and criteria based on reverse search results were not applied. Results of the final searches, including accessions and E-values are summarized in Appendix A chapter 3 Additional file 6: Table S3. Alignments used for constructing HMMs are found in Appendix A chapter 3 Additional file 9.

Finally, false negatives could be due to the divergence of a candidate from the experimentally validated *H. sapiens* query. In order to mitigate this possibility, HMMer searches were repeated with the same E-value cutoffs, but using protein databases of different taxa for reciprocal BLAST analysis. These taxa were selected from those taxa for which positive hits were validated in the previous HMMer searches, and which are included in the same supergroup as the taxa queried. For example, a CASP orthologue was identified in *Neospora caninum* using the closely related taxon *Toxoplasma gondii* for reverse BLAST searches, but not using *H. sapiens* (Appendix A chapter 3 Additional file 6: Table S3). Also, BLAST was used to search nucleotide scaffold sequences in the case of one protein of interest (Sec16) in *Pichia pastoris*, because it could not be found in the protein sequence database for this organism, and the protein database for the very closely related yeast *Komagataella phaffii* (which does contain a Sec16 sequence) was included in the analyses as well.

4.5.4 Phylogenetic analyses

For phylogenetic analyses, sequences were aligned using MUSCLE v3.8.31 (Edgar, 2004) with default parameters, and manually trimmed to retain only regions of clear homology. Alignments used for phylogenetic analyses are found in Appendix A chapter 3 Additional file 11 and Additional file 12. RAxML version 8.2.8 (Stamatakis, 2014) was used for Maximum Likelihood analysis. For RAxML analyses, the PROTGAMMALG4X model was used, and 100 non-parametric bootstraps were performed using the default faster hill climbing method (-f b, b, N 100). MrBayes version 3.2.6 (Ronquist and Huelsenbeck, 2003) was used for Bayesian analysis. For MrBayes analyses,

over four million Markov chain Monte Carlo generations were run under the Mixed model with a burnin of 25% to average standard deviations of splits frequencies of 0.01 or lower, indicating convergence. Both RAxML and MrBayes analyses were run using the CIPRES webservice (Miller *et al.*, 2010). In the case of the GRASP proteins, several consecutive analyses were required with removal of divergent sequences to resolve phylogenetic relationships.

4.5.5 Afterword

While the results reported in this chapter support the hypothesis that the LECA possessed a Golgi with cisternae already differentiated into at least three subtypes defined in part by specific golgins, whether golgins are paralogous and hence relevant to understanding Golgi differentiation in terms of the OPH remains unknown. Alignments of identified golgin sequences from diverse eukaryotes revealed no evidence of homology among golgins with specificity to different sub-domains the the Golgi, consistent with previous observations of smaller sequence samples (Munro, 2011). Thus there is no evidence to say whether they are related. It is possible that during pre-LECA evolution of the Golgi golgins with specificity to different Golgi cisternae originated entirely independently. This would be comparable to the evolutionary history inferred for multi-subunit tethering complexes (Dacks and Field, 2007; Koumandou *et al.*, 2007). However, it is also possible that they originated through paralogous evolution from a single original golgin, but that sequence similarity that would be needed to infer paralogy and even homology was subsequently erased by rapid sequence evolution.

Chapter 5

Phylogenetic analysis of SNARE proteins elucidates origins of eukaryotic endomembrane organelles

5.1 Preface

While the previous chapters are relevant to the general question of whether membrane trafficking protein families and organelles evolve in a manner consistent with the OPH, work presented in this chapter focuses on one particularly relevant superfamily of membrane trafficking proteins, the SNAREs, and attempts to dig deeper into the phylogenetic relationships between eukaryotic outparalogues to evaluate specific alternative scenarios for early endomembrane organelle evolution.

Of course, for this it is important to first establish what SNAREs represent eukaryotic outparalogues. During my graduate studies I performed two analyses of SNARE protein families which were instrumental in laying this groundwork, and which have already been published along with the work of collaborators, with myself as co-author on both manuscripts. That work is not reproduced in full in this thesis, because I did not play leading roles in the overall projects that were published. Therefore it is important to briefly describe that work here.

First, I analyzed the evolution of an unusual Qa-SNARE, Syntaxin 17. This SNARE was previously thought to be a metazoan inparalogue of endosomal Qa-SNAREs, related to Syntaxin 7 and 20 (Kloepper *et al.*, 2008; Kienle *et al.*, 2009). However, my analysis revealed orthologues of Syntaxin 17 in a sparse yet diverse collection of eukaryotes including the cryptophyte *Guillardia theta*, the ciliate *Tetrahymena thermophila*, the rhizarian alga *Bigelowiella natans*, heterolobosean *Naegleria gruberi*, and the apusozoan *Thecamonas trahens*. These results were published in Arasaki *et al.* (2015) (relevant figures in that publication are Figure 1 and Figure S1).

Second, I worked along with an undergraduate student Nerissa N. Nankisoor in the Dacks laboratory to perform phylogenetic analyses of putative orthologues of the Qb-SNARE Novel Plant SyNtaxin (NPSN) and the Qc-SNARE Syp7. Confirming earlier results based on similarity

searching methods (Sanderfoot, 2007; Yoshizawa *et al.*, 2006; Kloepper *et al.*, 2007; Kienle *et al.*, 2009), we found significant support in our phylogenetic analyses for orthology of NPSN and Syp7 SNAREs across eukaryotic diversity, with independent losses of both SNAREs in both Metazoa and Fungi. Because of the relevance of these findings to work of our collaborators on membrane trafficking proteins in trypanosomatid parasites, we published these results in Venkatesh *et al.* (2017) (relevant figures in that publication are Figure 1, Figure S2 F and G, and the relevant tables are Table S3 and S4). Together, these results added to the known complement of SNARE proteins that were present in the LECA, thus adding new pieces to the puzzle of pre-LECA evolution of SNAREs and membrane trafficking organelles.

5.2 Introduction

5.2.1 General SNARE structure and function

The Soluble *N*-ethylmaleimide sensitive factor Attachment protein REceptor (SNARE) protein superfamily comprises a diverse and highly conserved set of homologous membrane trafficking proteins that are integral to the function of the eukaryotic membrane trafficking system (Jahn and Scheller, 2006). All membrane trafficking pathways end in membrane fusion events. SNAREs are necessary to catalyze the process of fusion between adjacent membranes following their close apposition by tethering factors (Figure 1.2) (Bonifacino and Glick, 2004). The foundational roles of SNAREs in eukaryotic cell biology are highlighted by the many cases of SNARE targeting by factors secreted by bacterial pathogens, which can inhibit specific processes such as synaptic transmission (Simpson, 2004b), phagocytosis (Shi *et al.*, 2016), and autophagy (Arasaki *et al.*, 2017).

Each functional SNARE complex forms a core heterotetrameric coiled-coil structure composed of four SNARE domains within SNARE proteins (Sutton *et al.*, 1998) (Figure 5.1A). SNAREs are classified as either R-SNAREs or Q-SNAREs based on the presence of an arginine (R) or glutamine (Q) at the central residue of the SNARE domain coiled-coil structure (Fasshauer *et al.*, 1998). Q-SNAREs are further classified into Qa, Qb, and Qc SNAREs, defined by overall sequences similarity to other Q-SNAREs of the same class as well as their functional contribution to SNARE complexes, and each SNARE complex requires one of each type of SNARE domain (Qa, Qb, Qc, and R) (Weimbs *et al.*, 1998; Bock *et al.*, 2001). SNAREs are typically tail-anchored proteins, which are post-translationally inserted into the ER membrane at their C-terminus and are either retained in the ER or subsequently transported to other locations in the cell via membrane trafficking (Borgese *et al.*, 2003). SNARE complexes form between SNAREs on opposing

membranes, typically with an R-SNARE on a vesicle membrane and Q-SNAREs on the target membrane. The formation, or "zippering", of these SNARE complexes then pulls the membranes together promoting fusion (Lin and Scheller, 1997; Hanson *et al.*, 1997; McNew *et al.*, 2000b). Formation of specific SNARE complexes are coordinated in part by various tethering complexes (Hong and Lev, 2014) and proteins of the Sec1/Munc18 (SM) family (Toonen and Verhage, 2003). Once membrane fusion is complete, *N*-ethylmaleimide sensitive fusion protein (NSF) and α Soluble NSF Attachment Protein (αSNAP) mediate disassembly of SNARE complexes in preparation for subsequent fusion events (Söllner *et al.*, 1993a,b; Ungermann *et al.*, 1998).

In addition to a SNARE domain and a C-terminal transmembrane domain, SNAREs often contain N-terminal domains with function in regulation (Dietrich *et al.*, 2003). The R-SNAREs typically contain an N-terminal longin domain which functions in regulating localization or function (Filippini *et al.*, 2001; Gonzalez *et al.*, 2001; Tochio, 2001; Rossi *et al.*, 2004). The longin domain of a mammalian R-SNARE Vesicle Associated Membrane Protein (VAMP) 7 was shown to mediate interaction with the ArfGAP AGFG/Hrb for inclusion in endocytic vesicles budding from the plasma membrane (Schlacht *et al.*, 2013; Pryor *et al.*, 2008), and also interaction with the δ subunit of the Adaptor Protein (AP)-3 complex for trafficking to late endosomes (Martinez-Arca *et al.*, 2003). Similar findings implicate the longin domain of the yeast homologue of VAMP7, Nyv1, in AP-3-dependent sorting to the vacuole (Wen *et al.*, 2006), and also the *Arabidopsis thaliana* vacuole-localized VAMP7 orthologues (Uemura *et al.*, 2005). Also, several of the Q-SNAREs contain N-terminal domains composed of a bundle of three alpha helices termed the Habc domain (for alpha Helices a, b, and c) (Dietrich *et al.*, 2003). Examples include the Qa-SNARE Syntaxin 1A (Fernandez *et al.*, 1998), Qb-SNARE Vti1 (Gossing *et al.*, 2013), and Qc-SNARE Syntaxin 6 (Misura *et al.*, 2002). These Habc domains appear to function in regulation, and in the case of Qa-SNAREs in binding to SM proteins (Yamaguchi *et al.*, 2002; Dulubova, 2002).

Several SNAREs deviate considerably from the typical domain topology and/or functions of SNARE proteins. The Qbc SNAREs (termed Synaptosomal-associated proteins (SNAPs) in human) contain both a Qb and a Qc domain (Figure 5.1B). Thus SNARE complexes containing Qbc SNAREs only contain three SNARE peptides, which is the exception rather than the rule among SNARE complexes (Fukuda *et al.*, 2000). Qbc SNAREs also do not contain a transmembrane domain, although they are lipidated in some cases such as the human proteins SNAP-25 and SNAP-23 (Veit *et al.*, 1996), but not SNAP-29 or SNAP-47 (Holt *et al.*, 2006). Also, the R-SNARE Ykt6 similarly lacks a transmembrane domain and associates with membranes via a palmitoyl group (Fukasawa *et al.*, 2004). SNARE domain-containing proteins which apparently lack transmem-

brane domains also include the R-SNARE domain-containing regulatory protein tomosyn (Fujita *et al.*, 1998; Pobbati *et al.*, 2004) as well as some trypanosome-specific Qa-SNAREs (Venkatesh *et al.*, 2017). In addition, certain R-SNAREs in vertebrates and other eukaryotes lack longin domains (Filippini *et al.*, 2001). These have been termed brevins, and are discussed further below. The opposite transition appears to have occurred in the case of the plant-specific phytolongins, which are related to R-SNAREs, but only contain the longin domain (Vedovato *et al.*, 2009).

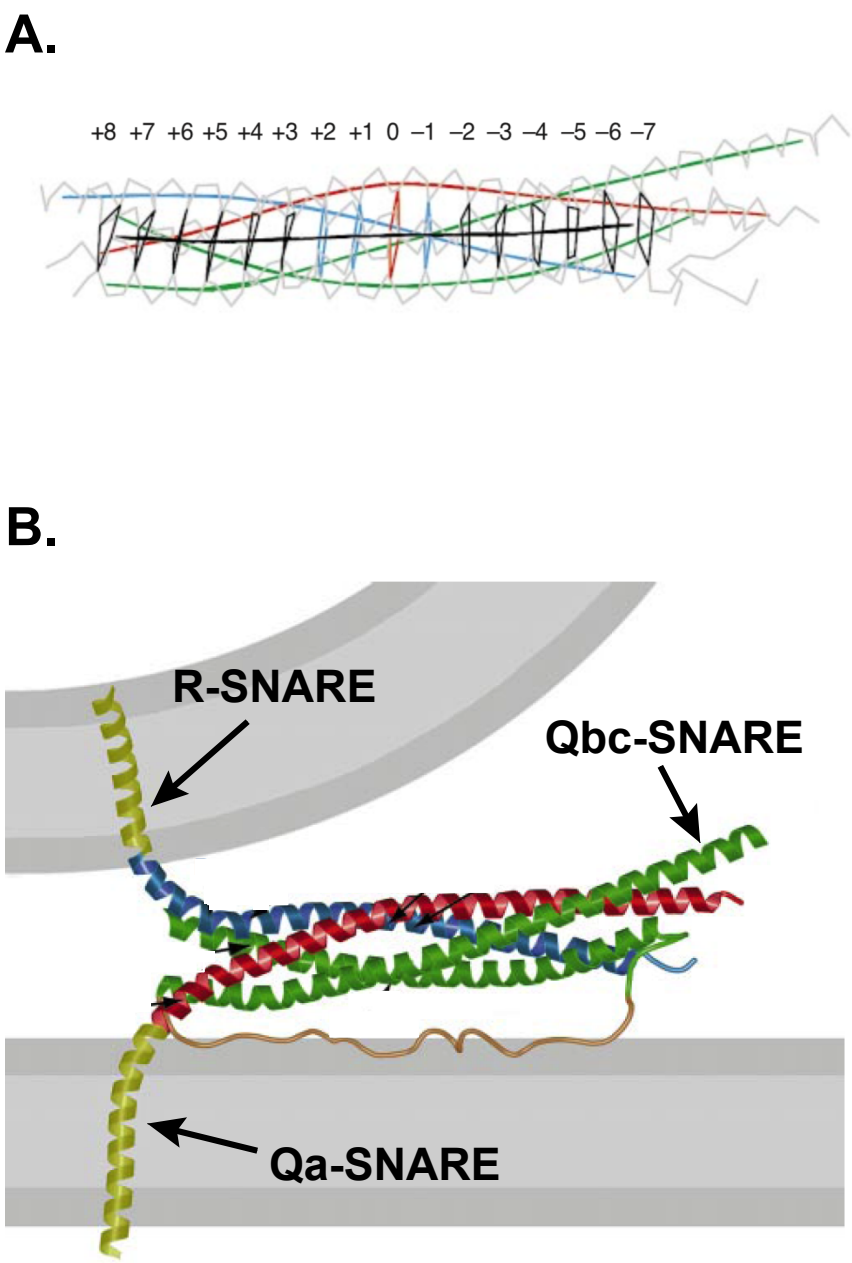


Figure 5.1: The structure of SNARE protein complexes. (*A*) The crystal structure of the synaptosomal SNARE complex reported by Sutton *et al.* (1998) revealed key conserved amino acid residues at the center ("zero") layer of the heterotetrameric coiled-coil structure: R (arginine) in the case of VAMP2, and Q (glutamine) for the remaining SNARE domains. (*B*) The structure of the synaptosomal (exocytic) SNARE complex showing the domain topology of the constituent SNAREs (Sutton *et al.*, 1998). The C-terminal transmembrane domain of the R-SNARE VAMP2 (blue) anchors it in the synaptic vesicle membrane, while that of the Qa-SNARE Syntaxin1A (red) anchors it to the opposing presynaptic membrane (plasma membrane). The Qbc-SNARE SNAP-25 (green) contributes both a Qb and Qc SNARE domain to the SNARE complex, which are connected by a flexible linker. This figure is adapted from Sutton *et al.* (1998) Figures 2C and 5.

5.2.2 Evolution and organelle-specificity of SNAREs

Genes encoding multiple SNAREs belonging to each of the four SNARE families are found in organisms representing every taxonomic supergroup of extant eukaryotes (Yoshizawa *et al.*, 2006; Kloepper *et al.*, 2007), but none have been identified in prokaryotes including closely related archaeal lineages (Klinger *et al.*, 2016b). Soon after the discovery of SNARE proteins, it was recognized, first regarding the Qa-SNAREs, that different SNAREs are specific to different target membranes of transport vesicles in the cell, and that sequence similarity seemed to predict functional similarity among SNARE homologues in mammalian and yeast cells (Bennett *et al.*, 1993). Parsimony-based analysis as well as clustering according to pairwise sequence similarities first indicated phylogenetic evidence for the existence of organelle-specific eukaryotic outparalogues of Qa-SNARE family members including endosomal syntaxins (Wang *et al.*, 1997; Bogdanovic *et al.*, 2000; Sanderfoot *et al.*, 2000). Shortly thereafter maximum likelihood (ML)-based phylogenetic analyses revealed the existence of five Qa-SNAREs conserved across diverse eukaryotes (Dacks and Doolittle, 2002, 2004).

Subsequent work to date reveals that each of the four SNARE families comprise several eukaryotic outparalogues, which are conserved in diverse extant eukaryotes (Arasaki *et al.*, 2015; Venkatesh *et al.*, 2017; Murungi *et al.*, 2014)[16]. Members of each family appear to share functional conservation not only between yeast and mammals, but among diverse eukaryotic lineages including plants (Uemura *et al.*, 2004), ciliates (Plattner, 2010), and trypanosomatid parasites (Murungi *et al.*, 2014; Besteiro *et al.*, 2006). This is consistent with general conservation of function among membrane trafficking protein orthologues (Klinger *et al.*, 2016a). Moreover, SNAREs within each family may generally be categorized into those acting at one of four membranes/organelles in the cell: the ER, Golgi, endosomes, or PM (Figure 5.2 and Table 5.1). For simplicity, herein individual members of each SNARE family in any given eukaryote will be referred to by the most commonly used term for their orthologue in mammalian or yeast cells, and if there are not such orthologues then the name of the orthologue in plant cells will be used. However, alternative terms may be provided for clarification. Conservation of localization among SNARE orthologues is consistent with their function being organelle-specific in the LECA, and with their evolution taking place in concert with that of those organelles.

[16] I am a co-author of each of these three articles

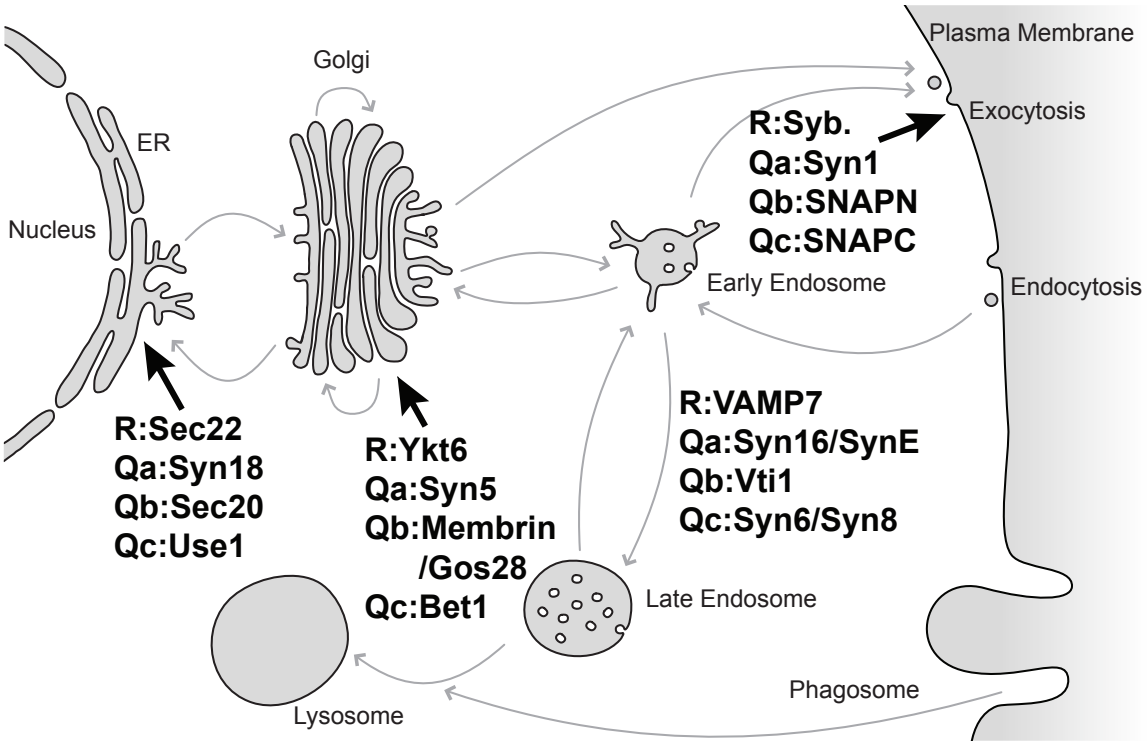


Figure 5.2: Organelle specificity of SNAREs to membrane trafficking organelles in a generalized eukaryotic cell. SNARE protein names are shown for mammalian SNARE representing eukaryotic outparalogues. Synaptobrevin 2/VAMP2 is abbreviated to "Syb.". SNAPN and SNAPC indicate the Qb and Qc SNARE domains, respectively, of the Qbc-SNARE SNAP-25. See main text for supporting references, and Table 5.1 for names of comparable SNAREs in yeast and plant cells.

Table 5.1: Conserved organelle-specific SNAREs in eukaryotes. Common names for representative orthologous SNAREs in mammalian, yeast, and plant cells and their organelle specificities are shown. See main text for supporting references. *The *Arabidopsis thaliana* R-SNARE VAMP72 is not orthologous to the human protein VAMP2 or yeast protein Snc2 (Sanderfoot, 2007) (see section 5.3.3).

Organelle-specificity	Taxon	R-SNARE	Qa-SNARE	Qb-SNARE	Qc-SNARE
PM	Mammals	VAMP2	Syntaxin 1	SNAP-25 N-term.	SNAP-25 C-term
	Yeast	Snc2	Sso1	Sec9 N-term.	Sec9 C-term.
	Plants	VAMP72*	Syp1	SNAP-33 N-term., NPSN	SNAP-33 C-term., Syp7
Endosomes and/or TGN	Mammals	VAMP7	Syntaxin 16	Vti1B	Syntaxin 6
	Yeast	Nyv1	Tlg2	Vti1	Tlg1
	Plants	VAMP71	Syp2	Vti1	Syp5
Golgi	Mammals	Ykt6	Syntaxin 5	Membrin	Bet1
	Yeast	Ykt6	Sed5	Bos1	Sft1
	Plants	Ykt6	Syp3	Membrin	Bet1
ER	Mammals	Sec22	Syntaxin 18	Sec20	Use1
	Yeast	Sec22	Ufe1	Sec20	Use1
	Plants	Sec22	Syp8	Sec20	Use1

5.2.2.1 SNAREs that function in exocytosis at the plasma membrane

In mammalian cells, the most well-studied SNARE complex is that which functions in exocytosis at the presynaptic membrane of neuronal synapses. The crystal structure of the exocytic SNARE complex of mammalian neuronal cells composed of the Qa-SNARE Syntaxin 1A, the Qbc-SNARE SNAP-25B, and the R-SNARE Synaptobrevin 2 (Sutton *et al.*, 1998) (Figure 5.1). Similar findings were reported for the exocytic SNARE complex of yeast cells (Katz *et al.*, 1998) composed of the Syntaxin 1 orthologues Sso1/2 (Aalto *et al.*, 1993), the SNAP-25 orthologue Sec9 (Brennwald, 1994), and the putative Synaptobrevin orthologue SNC1/2 (Protopopov *et al.*, 1993). The counterparts of these exocytic SNAREs in plant cells appear to be the Qa-SNARE PEN1 (Syp121) (Collins *et al.*, 2003), the Qbc-SNARE SNAP-33, and the R-SNARE VAMP72 (Kwon *et al.*, 2008). In plant cell cytokinesis, the role of PEN1 in the exocytic SNARE complex is filled by the cytokinesis-specific inparalogue KNOLLE (SYP111) (Lauber *et al.*, 1997). For simplicity, herein orthologous plasma membrane-specific Qa-SNAREs will be referred to as SynPM, as previously (Dacks and Doolittle, 2004). Also, the Qb-SNARE NPSN and Qc-SNARE Syp7 in combination are functionally similar to the Qbc-SNARE SNAP-33 (El Kasmi *et al.*, 2013) (see below for further discussion).

The exocytic R-SNAREs are of particular interest here, due to their unclear origins which are addressed in this study. In mammalian cells, Vesicle Associated Membrane Proteins (VAMP)s 1 to 5 appear to be entirely or partially specific to exocytosis, and these are all brevins lacking the Longin domain present in the canonical R-SNAREs Sec22, Ykt6, and VAMP7 (Filippini *et al.*, 2001). The synaptobrevins VAMP1 and VAMP2 are specific to regulated secretion in neuronal cells, having been first identified on synaptic vesicles (Trimble *et al.*, 1988; Baumert *et al.*, 1989; Elferink *et al.*, 1989). Similar specificity was seen for a homologue in *Drosophila* (Sudhof *et al.*, 1989). Also, VAMP3 (cellubrevin) was shown to function similarly to VAMP1 and 2 but in non-neuronal cells (McMahon *et al.*, 1993). Specificity of VAMP3 for function in exocytosis was demonstrated, despite trafficking through endosomes (Link *et al.*, 1993). VAMP4 localizes to and functions at the TGN/EE in addition to synaptic vesicles due to a dileucine motif that mediates interaction with the Adaptor Protein 1 complex for inclusion in clathrin-coated vesicles (Bethani *et al.*, 2009; Steegmaier *et al.*, 1999; Peden *et al.*, 2001). However, VAMP4 does play a role in exocytosis in neural cells (Raingo *et al.*, 2012). VAMP5 is somewhat specific to muscle tissue, and localizes to the PM (Zeng *et al.*, 1998). The exocytosis-specific brevin R-SNAREs in yeast SNC1 and SNC2 are putative fungal orthologues of mammalian VAMPs 1 to 5 (referred to herein collectively as synaptobrevin) (Protopopov *et al.*, 1993). However, the exocytic R-SNARE of plants, VAMP72,

is not orthologous to the exocytic R-SNAREs of metazoa and fungi (synaptobrevins) (Sanderfoot, 2007), and this is also discussed in greater detail below.

5.2.2.2 SNAREs that function at the *trans*-Golgi network, endosomes, or lysosomes

In yeast cells, SNARE complexes mediating fusion at early endosomal/TGN membranes are composed of the Qa-SNARE Tlg2 (an orthologue of Syntaxin 16), the Qb-SNARE Vti1, the Qc-SNARE Tlg1 (orthologue of Syntaxin 6) (von Mollard *et al.*, 1997; Abeliovich *et al.*, 1998; Holthuis, 1998; Brickner *et al.*, 2001). Endosomal SNARE complexes in yeast may alternatively contain the Qa-SNAREs Pep12 or Vam3 in place of Tlg2, which are specific to the prevacuolar compartment (PVC) and vacuole, respectively (Becherer and Jonest, 1996; Wada *et al.*, 1997). Also, Syn8 (a Syntaxin 8 orthologue) is functionally redundant with Tlg1 (Lewis and Pelham, 2002) while Vam7 (another Syntaxin 8 orthologue) acts in place of Tlg1 at the vacuole (Wada and Anraku, 1992; Sato *et al.*, 1998; Cheever *et al.*, 2001). The R-SNARE Nyv1 (orthologue of VAMP7) functions in fusion at the vacuole in yeast (Nichols *et al.*, 1997; Reggiori *et al.*, 2000), and the exocytic R-SNARE Snc1/2 can function in fusion at early endosomal/TGN membranes (Holthuis, 1998; Reggiori *et al.*, 2000) as it cycles continuously between these membranes and the PM (Lewis *et al.*, 2000).

In mammalian cells, the situation is more complex due to the abundance of metazoa-specific inparalogues of endosomal SNAREs (Kienle *et al.*, 2009) which results in many SNARE complexes of different and often overlapping composition. SNARE complexes mediating fusion at the TGN and early endosomes contain the Qa-SNARE Syntaxin 16 (Syn16), the Qb-SNARE Vti1A, the Qc-SNARE Syntaxin 6 and one of several R-SNAREs including VAMP3 and VAMP4 (Bock *et al.*, 1997; Mallard *et al.*, 2002; Kreykenbohm, 2002). While VAMP3 and VAMP4 may have some activity at the early endosomes, in general VAMPs 1 to 5 are more specific for exocytosis as described above (section 5.2.2.1). In contrast, VAMP7 (also known as tetanus-insensitive VAMP) and VAMP8 function primarily at endosomes; VAMP7 localizes to the TGN/EE and late endosomes, and is necessary for traffic from the late endosome to lysosome (Advani *et al.*, 1999). SNARE complexes mediating fusion at the late endosome and lysosome contain the Qa-SNARE Syntaxin 7, the Qb-SNARE Vti1B, the Qc-SNARE Syntaxin 8, and the R-SNARE VAMP8 (Mullock *et al.*, 2000; Wang *et al.*, 1997; Antonin *et al.*, 2000). A crystal structure of this second endosomal SNARE complex was reported (Antonin *et al.*, 2002), and it confirmed overall similarity/homology to the crystal structure of the exocytic SNARE complex that was reported earlier (Sutton *et al.*, 1998). For simplicity, herein orthologues of the endosomal Qa-SNARE, such as

Syntaxin 7 and Vam3, will be referred to as SynE, as previously (Dacks and Doolittle, 2004).

Endosomal SNAREs in non-opisthokont eukaryotes are similar to those in yeast and mammals. In the amoebozoan *Dictyostelium discoideum*, a SNARE complex composed of SynE, Vti1, Syn8, and VAMP7 mediates fusion at endosomal organelles including macropinosomes (Bogdanovic *et al.*, 2000, 2002; Bennett *et al.*, 2008). At the TGN/early endosome of *Arabdiopsis thaliana* cells, SNARE complexes are composed of the Qa-SNARE Syp41 (a Syn16 orthologue), the Qb-SNARE Vti12, the Qc-SNARE Syp61 (a Syn6 orthologue), and the R-SNARE VAMP71 (Bassham *et al.*, 2000; Sanderfoot *et al.*, 2001; Sanderfoot, 2007; Drakakaki *et al.*, 2012). At the vacuole of *A. thaliana*, a complex has been identified composed of SYP22 (a SynE orthologue), VTI11, SYP51 (a Syn8 orthologue), and VAMP727 (Sato *et al.*, 1997; Ebine *et al.*, 2008; Sanderfoot, 2007; Uemura *et al.*, 2010). VAMP727 is a seed plant-specific member of the VAMP72 group of VAMP7 orthologues in *A. thaliana*, as defined by Sanderfoot (2007). Other members of this VAMP72 group function in secretion in *A. thaliana* (Ebine *et al.*, 2008), and so the role of VAMP727 at the vacuole in seed plants appears to be an exception among embryophytes, especially as the VAMP72 orthologue VAMP72B in the liverwort *Marchantia polymorpha* localizes to the plasma membrane (and partially to the TGN/EE) (Kanazawa *et al.*, 2016). On the other hand, VAMP71 paralogues are specific to the vacuole in both *A. thaliana* and *M. polymorpha* (Carter *et al.*, 2004; Uemura *et al.*, 2004, 2005; Kanazawa *et al.*, 2016).

Importantly, herein *trans*-Golgi networks, early endosomes, late endosomes, and lysosomes are combined in a single "endosomal" SNARE category (Table 5.1). The rationale for this is based on extensive overlap between sets of SNAREs that function at these different organelles. In addition, as reviewed previously by Lam *et al.* (2007), plant cells appear to have a combined *trans*-Golgi network/early endosome (TGN/EE). The key observation supporting this is that endocytic vesicles appear to be delivered directly to the *trans*-Golgi region (Viotti *et al.*, 2010). Similar observations have been made in the budding yeast *Saccharomyces cerevisiae* (Day *et al.*, 2018) and in the apicomplexan parasite *Toxoplasma gondii* (Jackson *et al.*, 2013). Even in mammalian cells, which appear to have somewhat distinct EE and TGN, an orthologue of the Qb-SNARE Vti1 (Vti1A) (Mallard *et al.*, 2002) and the Qc-SNARE Syn6 (Simonsen *et al.*, 1999) localize to and function at both organelles. It is important to consider that metazoa exhibit a greatly expanded set of SNAREs (metazoan inparalogues) particularly endosomal SNAREs (Kloepper *et al.*, 2008). With this in mind, specificity of metazoan inparalogues of ancient SNARE family members for either the EE or TGN may indicate that the distinction between EE and TGN is a lineage-specific elaboration of the membrane trafficking system. For example, there are inparalogues of SynE, Vti1, Syn6, and

VAMP7 in both metazoa and embryophytes (Sanderfoot, 2007; Dacks *et al.*, 2008). Consequently SNARE evolution is unlikely to be informative for distinguishing between evolutionary scenarios including distinctions between the TGN and EE that may have been present in the LECA or prior to the LECA.

Distinctions between SNARE complexes that function at TGN/EEs and recycling endosomes (REs), late endosomes (LEs), or lysosomes are also not of particular interest for the current study, because there do not appear to be eukaryotic outparalogues within every SNARE family that are specific to these organelles. The Qa family contains SynE, which appears to be more specific to late endosomes and lysosomes in contrast to the TGN/EE-specific Qa-SNARE Syn16 (Becherer and Jonest, 1996; Wada *et al.*, 1997; Mullock *et al.*, 2000; Wang *et al.*, 1997; Sato *et al.*, 1997). Similarly, the Qc family contains Syn8, which appears to be specific to late endosomes and lysosomes as well in contrast to the TGN/EE-specific Qc-SNARE Syn6 (Wada and Anraku, 1992; Sato *et al.*, 1998; Antonin *et al.*, 2002). However, there does not appear to be pairs of eukaryotic R-SNARE or Qb-SNARE outparalogues with comparable differentiation in specificity for early vs. late endosomes/lysosomes. Instead, VAMP7-related R-SNAREs seem to fulfil roles at both late and early endosomes and in exocytosis at the plasma membrane, although different paralogues play distinct roles in specific eukaryotic lineages (this is discussed further below). Also, orthologues of the Qb-SNARE Vti1 clearly function at both the TGN/EE and late endosomes and lysosomes/vacuoles (von Mollard *et al.*, 1997; Holthuis, 1998; Mallard *et al.*, 2002; Antonin *et al.*, 2002; Sanderfoot *et al.*, 2001). To our knowledge, no SNAREs have been found to be specific to REs, also sometimes referred to as the Endosomal Recycling Compartment (ERC). Also, the evolutionary conservation and origins of REs is somewhat unclear, with some eukaryotes including plants and budding yeast apparently lacking distinguishable REs (Singh and Jürgens, 2018; Day *et al.*, 2018).

Moreover, the specificity of SynE and Syn8 orthologues for late endosomes may be questioned. Syntaxin 13 a metazoan inparalogue of SynE related to Syntaxin 7 does not localize to late endosomes, but is instead restricted to early and recycling endosomes (Prekeris *et al.*, 1998), thus obscuring the specificity of SynE orthologues for late endosomes or lysosomes. Also, while Syntaxin 8 has been identified as a component of late endosomal SNARE complexes (Antonin *et al.*, 2002), it also appears to localize to both early and late endosomes in mammalian cells (Prekeris *et al.*, 1999; Subramaniam *et al.*, 2000). Similarly, the yeast orthologue of Syntaxin 8 appears to be functionally redundant with the early endosome/TGN Qc-SNARE Syn6 (Tlg1) (Lewis and Pelham, 2002). And, the plant orthologue of Syntaxin 8, Syp51, appears to function at both the TGN/EE and at the prevacuolar compartment (Sanderfoot *et al.*, 2001). These observations indicate a close

evolutionary relationship between these organelles that perhaps fits well with the observed maturation of late endosomes and subsequently lysosomes from TGN/EE in diverse eukaryotes (Murphy, 1991; Stoorvogel *et al.*, 1991; Maniak, 2003; Scheuring *et al.*, 2011).

5.2.2.3 SNAREs that function at the Golgi and endoplasmic reticulum

In contrast to the complexity of endosomal SNAREs, the Golgi and ER SNAREs in different eukaryotes appear to be less complex, with fewer inparalogues identified in metazoa and embryophytes (Kienle *et al.*, 2009; Sanderfoot, 2007). In mammals, yeast, and plants, the Golgi SNARE complex is composed of orthologues of the Qa-SNARE Syn5, the Qb-SNAREs Membrin or Gos28, the Qc-SNARE Bet1, and the R-SNARE Ykt6 (Newman *et al.*, 1990; Banfield *et al.*, 1995; Subramaniam *et al.*, 1996; Lupashin *et al.*, 1997; McNew *et al.*, 1997; Hay *et al.*, 1998; Zhang and Hong, 2001; Shorter *et al.*, 2002; Xu *et al.*, 2002; Volchuk *et al.*, 2004; Tai *et al.*, 2004; Uemura *et al.*, 2004; Kanazawa *et al.*, 2016). Similarly, the ER SNARE complex is composed of orthologues of the Qa-SNARE Syn18, the Qb-SNARE Sec20, the Qc-SNARE Use1, and the R-SNARE Sec22 (Lewis, 1997; Hay *et al.*, 1998; Hatsuzawa *et al.*, 2000; Uemura *et al.*, 2004; Kanazawa *et al.*, 2016). The ER R-SNARE Sec22 can function in ER-Golgi transport, fusing with Golgi-localized Q-SNAREs (Newman *et al.*, 1990; Paek *et al.*, 1997; Parlati *et al.*, 2000), but also functions in homotypic fusion of ER membranes (Lee *et al.*, 2015b).

5.2.2.4 Challenges to organelle-specificity of SNAREs

The relevance of SNARE proteins to organelle evolution with reference to an OPH mode of evolution is clearly predicated on consistent maintenance of organelle-specificity among SNARE family members throughout eukaryotic evolution. Thus, examples of SNARE promiscuity pose a potential challenge to this relevance. In the original formulation of the OPH, relevant protein families were explicitly identified as those that are specificity-encoding (Dacks and Field, 2007; Dacks *et al.*, 2008).

A spectrum of opinions exists in the field of membrane trafficking research regarding the role of SNARE proteins in encoding organelle specificity of membrane traffic. This is a key issue, because the existence of organelles in the membrane trafficking system depends on some proteins but not others reaching them through the trafficking system. The multiplicity, functional similarity, and organelle-specific localizations of SNAREs is consistent with the view that specificity of SNARE-SNARE interactions are sufficient to govern specificity of membrane traffic. This view has been termed the SNARE hypothesis (Söllner *et al.*, 1993b; Rothman, 1994; Rothman and Warren, 1994), and is supported by *in vitro* assays which examined which combinations of yeast SNAREs are suf-

ficient to mediate fusion of proteoliposomes (McNew *et al.*, 2000a; Paumet *et al.*, 2004) as well as some *in vivo* experiments (Yang *et al.*, 2008). On the other hand, some researchers argue that SNAREs themselves do not confer any specificity to membrane fusion (Brandhorst *et al.*, 2006). This view is supported by studies examining the interaction of SNARE domains *in vitro*, which suggest general promiscuity of SNARE-SNARE iteractions (Fasshauer *et al.*, 1999; Yang *et al.*, 1999). Also, some studies that examined the fusion propensities of proteoliposomes with various SNARE protein compositions suggest at least partial promiscuity as well (Brandhorst *et al.*, 2006; Furukawa and Mima, 2014). Moreover, mammalian endosomal and exocytic SNAREs have been known to mix in endosomal membranes and bind non-specifically *in vivo*, only forming specific interactions with the organization of membrane contact sites (Bethani *et al.*, 2007). Indeed, as noted above, several SNAREs especially of the R-SNARE family are known to function in multiple different SNARE complexes *in vivo*. For example, the Golgi R-SNARE Ykt6 can compensate for the loss of the ER SNARE Sec22 in yeast, as demonstrated by *in vivo* experiments (Liu and Barlowe, 2002). In fruit fly cells, Ykt6 plays a role in membrane fusion at the PM as well (Gordon *et al.*, 2017). Specificity of membrane traffic *in vivo* likely involves coordination of a variety of factors of which specificity of SNARE-SNARE interactions may be just one. SM proteins and vesicle tethers including golgins and multisubunit tethering complexes appear to encode specificity in combination with SNAREs (Shorter *et al.*, 2002; Shen *et al.*, 2007; Laufman *et al.*, 2013; Kraynack *et al.*, 2005; Koike and Jahn, 2019; Koumandou *et al.*, 2007).

While SNAREs appear to be only partially responsible for encoding organelle-specificity, their overall conservation nevertheless supports their relevance to investigation of organelles via an OPH mechanism. SNAREs do not exhibit a pattern of frequent neofunctionalization, with SNAREs switching specificity to different organelles. Instead, the literature to date supports at least a general conservation of SNAREs specific to the ER, Golgi, endosomes, or PM, as discussed above (sections 5.2.2.1, 5.2.2.2, and 5.2.2.3). For example, there are no orthologues of ER or Golgi SNAREs that have acquired specific localization and function at the PM or endosomes, or vice versa. While some SNAREs, especially R-SNAREs have trafficking itineraries that may span multiple organelles, their primary organelle specificities tend to be conserved among different eukaryotes.

One reason for the maintenance of multiple organelle-specific SNAREs in most eukaryotes, despite the potential interchangeability of SNAREs in fusion complexes may simply be that additional features are necessary for maintaining localization to different organelles for organelle-specific functions, making existence of organelle-specific SNAREs integral to maintenance of distinct organelle identities. It may be that the diversity of SNAREs exist more due to the need for

differential sorting of the SNAREs themselves. While SNAREs rarely exhibit cytosolic sorting motifs (Jahn and Scheller, 2006), properties of the tail-anchors of Q-SNAREs have been implicated in mediating differential sorting within the membrane trafficking system (González Montoro *et al.*, 2017; Quiroga *et al.*, 2013; Rayner and Pelham, 1997; Watson and Pessin, 2001). The longin domains of R-SNAREs are responsible for differential sorting as well (Martinez-Arca *et al.*, 2003; Uemura *et al.*, 2005). An alternative possibility is that organelle-specific SNAREs are required due to their differing abilities to bind organelle-specific tethering complexes. In any case, the reason for the existence of organelle-specific SNARE paralogues is not of primary importance for relevance to the OPH, and conservation of organelle-specific localization/function among orthologues is all that is required as a premise for this study as discussed further below.

5.2.3 Unresolved questions regarding the early evolution of SNAREs

Considering the overall conservation of function among SNARE orthologues of different eukaryotes, SNARE proteins offer potential for phylogenetic trees that are particularly informative for organelle evolution under an OPH paradigm (see Figure 1.5B). However, SNARE evolution has been incompletely reconstructed. While extensive phylogenetic analyses have been performed for the Adaptor complex subunits (Hirst *et al.*, 2014), and the Rab family of GTPases (Elias *et al.*, 2012), investigation of the SNARE superfamily has been more limited.

SNARE protein phylogenetics poses several challenges that must be approached with caution to avoid placing too much confidence in reconstructed phylogenetic tree topologies. *A priori*, the structure of SNARE proteins inherently limits the phylogenetic information that the sequences may contain. SNAREs are relatively short proteins (although not as short as Rab GTPases), with SNARE domains being approximately 60 amino acids long (Fasshauer *et al.*, 1998). Also, SNARE domains are coiled-coil domains, and as such may contain less phylogenetic signal, as considerable similarity is observed even between non-homologous coiled-coil domains. Also, considering the uncertain and potentially rapid origin of the eukaryotic membrane trafficking system given a lack of extant organisms representing intermediate states, the time between gene duplications giving rise to organelle-specific SNAREs of extant eukaryotes may have occurred in rapid succession, meaning that most of the sequence evolution would have occurred after diversification of SNARES, thereby obscuring the deep relationships of interest. However, the considerable branch support found in previous studies indicating evolutionary relationships between eukaryotic outparalogues within the Qa, Qb, Qc, and R families (Murungi *et al.*, 2014; Arasaki *et al.*, 2015; Venkatesh *et al.*, 2017) does suggest that inferences regarding pre-LECA SNARE evolution may be possible.

Prerequisites to such inferences include establishing a complete set of eukaryotic outparalogues for each SNARE family, and determining the relationships between these SNAREs by analysis of sequences from a diverse sampling of eukaryotes. Recent advances in SNARE phylogenetics include identification of the Qa-SNARE Syntaxin 17 as an additional eukaryotic outparalogue (Arasaki *et al.*, 2015), and confirmation of the Qb-SNARE NPSN and Qc-SNARE Syp7 as eukaryotic outparalogues as well (Venkatesh *et al.*, 2017). However, open questions remain, particularly regarding the evolution of exocytic SNAREs which appear to have evolved dynamically compared to other SNAREs. First, the relationships between the Qb and Qc SNARE domains and members of the Qb and Qc SNARE families, respectively, are unclear. Second, several studies have suggested the existence of an additional eukaryotic R-SNARE outparalogue with a role in exocytosis represented by the mammalian R-SNARE synaptobrevin (Kloepper *et al.*, 2007), but no phylogenetic evidence for this has been found.

Therefore, herein we attempt to determine the origin of the Qb and Qc SNARE domains of Qbc SNAREs, determining that the phylogenetic placement of their SNARE domains within the Qb and Qc SNARE families are likely as sister to the NPSN and Syp7 clades, respectively. This supports the hypothesis that NPSN and Syp7 have ancestral functions in exocytosis, similar to Qbc SNAREs. Next we examine the evolution of putative orthologues of vertebrate synaptobrevins, and find significant phylogenetic evidence for orthology of sequences from diverse eukaryotes with synaptobrevins.

Building upon these findings, we consider possible implications of SNARE evolution, for the early evolution of the eukaryotic endomembrane system. Cell biologists have hypothesized scenarios for the evolution of the endomembrane system of eukaryotic cells for over half a century, with the hope of providing a unifying theory for the cell biology of the endomembrane organelles of extant eukaryotes. However, it was aptly noted early on by de Duve and Wattiaux (1966) that, "It is obviously not possible to marshal any meaningful evidence either for or against a theory of this sort." This view has remained valid until now, despite an abundance of hypotheses being proposed over subsequent decades (de Duve, 1969; Blobel, 1980; Jékely, 2003; de Duve, 2007; Cavalier-Smith, 2009; Gould *et al.*, 2016). The primary reason for this is the absence of any extant organisms providing information on intermediate stages between complex eukaryotic cells and much simpler prokaryotic relatives. Nevertheless, evolution of the SNARE protein superfamily promises potential for differentiating between scenarios of endomembrane evolution, and we explore this possibility by attempting to determine the relative timing of the gene duplications giving rise to organelle-specific eukaryotic SNARE outparalogues within each SNARE family.

5.3 Results and discussion

5.3.1 SNARE domains of Qbc SNAREs are monophyletic and are closely related to NPSN and Syp7

Qbc-SNAREs have long been thought to be derived through fusion(s) of a Qb SNARE with a Qc SNARE (Fukuda *et al.*, 2000), but the origins of the SNARE genes that were fused remain unclear. As noted by Sanderfoot (2007), there is some possibility that Qbc-SNAREs originated independently in Opisthokonts and Archaeplastida. This suggestion was based on the findings of Dacks and Doolittle (2001), who performed similarity searches for SNAP-25-like Qbc-SNARES in a small sampling of eukaryotes and did not identify any Qbc-SNAREs in eukaryotes other than *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. Such a distribution would imply that if the LECA possessed an ancestral Qbc-SNARE, then it was lost at least five times independently (depending on the phylogenetic placement of Discoba and Metamonada), accounting for apparent absences in *Giardia*, *Trypanosoma*, *Dictyostelium*, *Chlamydomonas*, *Porphyra*, *Plasmodium*, *Theileria*, *Phytophthora*, and *Entamoeba* which were reported by Dacks and Doolittle (2001). Given these results, and assuming that fusion of a Qb SNARE with a Qc-SNARE to form a Qbc SNARE capable of functioning in membrane fusion is not far less probable than a gene loss event, a scenario involving independent origins of Qbc-SNAREs would indeed be more parsimonious as it would involve three evolutionary events instead of five. In any case, we address both the question of Qbc-SNARE monophyly and also the question of which Qb and Qc SNAREs the respective SNARE domains of Qbc-SNAREs may be most closely related.

As explained, the taxonomic distribution of Qbc-SNAREs is relevant to inferring their monophyly or independent origins. Since the analysis of Dacks and Doolittle (2001) many more genomic sequence data have become available, and additional proteins with the Qbc-SNARE topology have been identified. First, both (Yoshizawa *et al.*, 2006) and (Kloepper *et al.*, 2007) independently reported identification of Qbc-SNAREs in diverse eukaryotes including *Trypanosoma cruzi*, *Dictyostelium discoideum*, *Cryptosporidium*, and *Plasmodium*. Also, both Kloepper *et al.* (2007) and Schilde *et al.* (2008) independently identified Qbc-SNAREs in the ciliates *Paramecium tetraurelia* and *Tetrahymena thermophila*. Kloepper *et al.* (2007) further identified Qbc-SNAREs in *Theileria* and *Ostereococcus*. Sanderfoot (2007) identified Qbc-SNAREs in the chlorophyte algae *Chlamydamonas* and *Volvox*. Also, Venkatesh *et al.* (2017) identified Qbc-SNAREs in the genomes of the kinetoplastids, *Trypanosoma grayi* and *Bodo saltans*, in addition to *Trypanosoma cruzi*. Considering this abundance of SNAREs with the Qbc-SNARE topology identified by their similarity to Qbc-SNAREs of opisthokonts and plants, independent origins of Qbc-SNAREs as

suggested by Sanderfoot (2007) becomes considerably less parsimonious even without support from phylogenetic analysis.

To identify Qbc SNAREs in our sampling of genomic and transcriptomic data, we scanned SNARE domain-containing sequences against a database of SNARE family-specific Hidden Markov Models (HMMs), retaining those that retrieved an HMM constructed from selected reference Qbc SNAREs with the highest score (see methods). Next, we aligned positive hits for Qbc SNAREs and visually confirmed alignment of both SNARE domains among the sequences. These methods were successful in retrieving Qbc-SNAREs from taxonomic groups in which they were previously identified where our taxonomic sampling overlapped. We also identified Qbc-SNAREs in additional taxa including *Phytophthora sojae*, *Porphyra umbilicalis*, *Naegleria gruberi*, *Bigelowiella natans*, and *Emiliania huxleyi* (see Appendix A chapter 5 Alignment 1). We did not identify any Qbc-SNAREs *Entamoeba*, and did not search *Giardia*, so Qbc-SNAREs remain unidentified in these parasites, as reported by Dacks and Doolittle (2001). These observations further support the hypothesis that Qbc-SNAREs, where present, are conserved from the LECA and not independently derived. However, there appears to be high levels of divergence among the identified Qbc-SNARE sequences, perhaps leaving the question of their origins somewhat unresolved.

To determine the evolutionary relationships of constituent SNARE domains of Qbc-SNAREs among diverse eukaryotes to each other and to Qb and Qc SNARE family members, we aimed to extract representative N-terminal Qb and C-terminal Qc SNARE domains from Qbc SNAREs (termed SNAPN and SNAPC, respectively (Kloepper *et al.*, 2007)) and analyze them separately in the context of their respective SNARE family phylogenies. Previously published phylogenetic analyses of the Qb and Qc SNARE families include NPSN and Syp7 clades (Venkatesh *et al.*, 2017), establishing an important prerequisite to exploring all possible phylogenetic relationships of the SNAPN and SNAPC domains within these families. Based on sequence similarity, NPSN has been suggested to be closely related to the SNAPN domain of Qbc SNAREs and also to Vti1 (Kloepper *et al.*, 2007). Functional equivalence or similarity between NPSN/Syp7 and Qbc SNAREs has even been proposed previously based on apparent conservation of NPSN and Syp7 in eukaryotes without obvious Qbc homologues (Yoshizawa *et al.*, 2006; Sanderfoot, 2007). A strong functional basis for predicting a close evolutionary relationship between NPSN and SNAPN, and between Syp7 and SNAPC exists as well and are discussed below (5.3.2). However, to our knowledge, no evidence from phylogenetic analyses has been reported.

To perform phylogenetic analysis of SNAPN and SNAPC sequences, we first aligned Qbc SNARE sequences to visually identify and extract both the SNAPN and SNAPC subsequences

separately. These Qbc sequences included metazoan homologues of human SNAP-25, archaeplastid homologues of *Arabidopsis thaliana* SNAP-33, as well as Qbc SNAREs identified by Venkatesh *et al.* (2017) in kinetoplastids. The SNAPN and SNAPC sequences were then respectively aligned to previous alignments of representative sequences of Qb and Qc SNAREs (Venkatesh *et al.*, 2017). These two alignments were then analyzed similarly for the alignments without the SNAPN or SNAPC sequences (Venkatesh *et al.*, 2017). Also, to find the best trees that are consistent with the prior classification of the canonical Qb and Qc SNAREs, we constrained clades containing the previously identified orthologues. We performed several rounds of phylogenetic analysis, selecting for relatively short-branching sequences from a broad taxonomic diversity.

The resulting tree topologies indicate that SNAPN and SNAPC domains are closely related to NPSN and Syp7, respectively, as predicted (Figures 5.3 and 5.4). The Qb analysis supports the monophyly of SNAPN sequences with significant prior probability and bootstrap support (Figure 5.3). In an unconstrained analysis, the monophyly of SNAPN sequences was supported by a prior probability of 0.93, but a bootstrap percentage of only 26 (see Appendix A chapter 5 Figure S1). Prior probability branch support for the clade containing NPSN and SNAPN is low (0.58 out of 1.0) and bootstrap support is marginal at 49% (Figure 5.3). In contrast, support for the clade containing NPSN, SNAPN, and Vti1 is much more robust, thus we can confidently place SNAPN within the clade containing NPSN and Vti1, and there may be two potentially viable alternatives involving placement of SNAPN as sister to Vti1 or as sister to the clade containing Vti1 and NPSN.

In the Qc analysis, support for monophyly of all SNAPC sequences included was below the significance threshold (only a prior probability of 0.56, and bootstrap percentage of 42) (Figure 5.4). However, a strong prior probability of 0.92 supported the monophyly of SNAPC sequences from metazoans, embryophytes, and the cryptophyte *Guillardia theta*. This monophyletic relationship was supported by a prior probability of 0.98 in in an unconstrained analysis as well (see Appendix A chapter 5 Figure S2). Full prior probability support for the clade containing Syp7 and SNAPC is seen (prior probability of 1.0) and the bootstrap support is significant as well (73%) (Figure 5.4).

Taken together, these phylogenetic results for SNAPN and SNAPC domains suggest that Qbc-SNAREs identified across eukaryotic diversity are monophyletic, with the SNAPC domain clearly showing a close relationship to Syp7, and the SNAPN domain clearly showing a relationship to NPSN and Vti1, but somewhat higher affinity for NPSN. A caveat to these results is that both of these phylogenies are arbitrarily rooted. So, if the true placement of the roots were on the NPSN or Syp7 clades, for example, then SNAPN and SNAPC would be no more closely related to NPSN and Syp7 than any other Qb or Qc SNAREs. The evolution of Qbc-SNAREs is discussed further

in the following section.

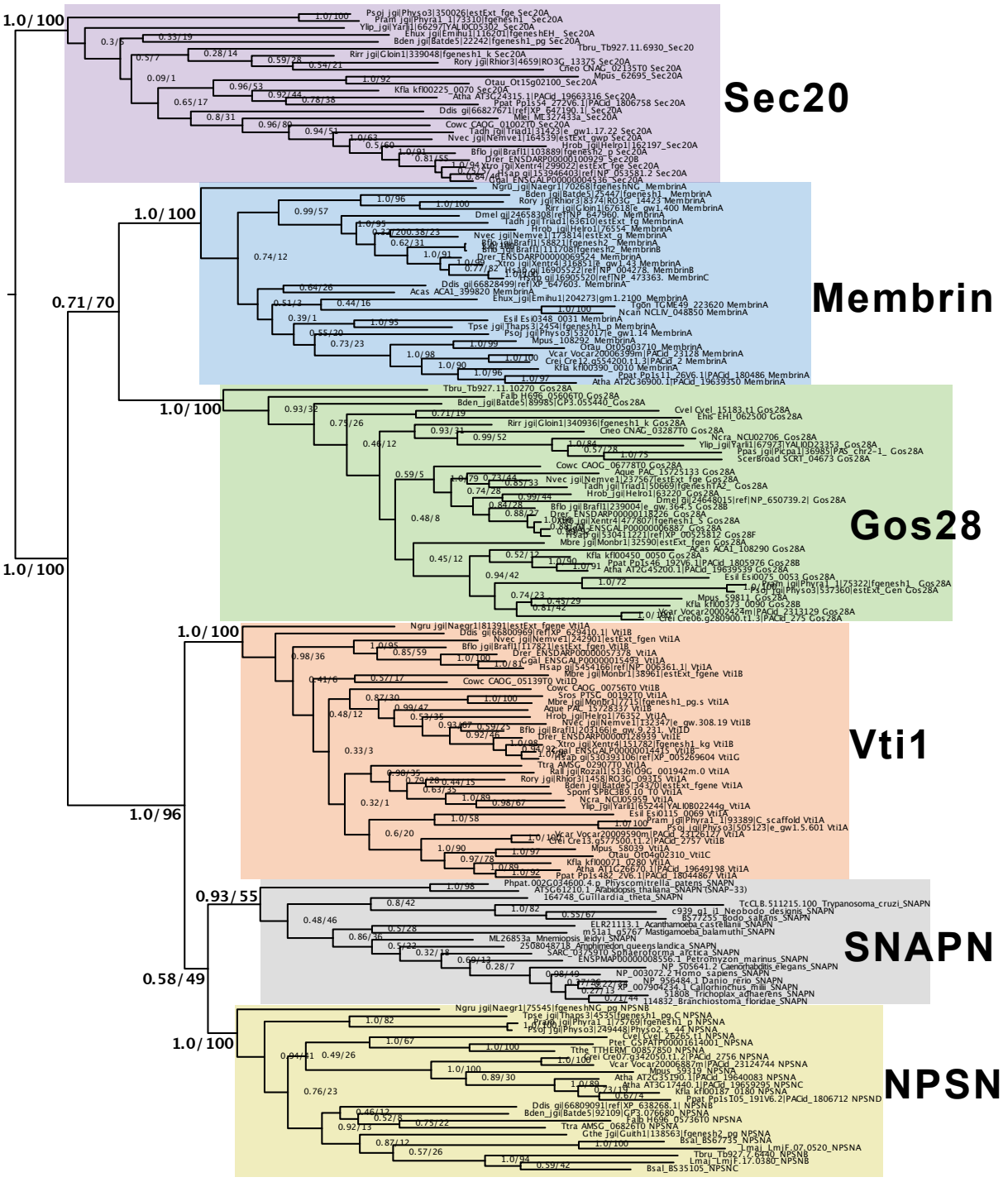

Sec20
Membrin
Gos28
Vti1
SNAPN
NPSN
0.5

Figure 5.3: Phylogenetic analysis of Qb SNAREs and the N-terminal (Qb) SNARE domain of identified Qbc SNAREs (SNAPN). SNAPN sequences were appended to a previous alignment of Qb SNAREs from from a broad sampling of eukaryotes (Venkatesh *et al.*, 2017) (see Appendix A chapter 5 Alignment 2). Branch support values are from MrBayes (posterior probability, maximum 1.0 with 0.8 generally being considered the minimum significant value) and RAxML (nonparametric bootstrap values, maximum 100 with 50 generally being considered the minimum significant value). Clades containing previously classified orthologues (Venkatesh *et al.*, 2017) were constrained in this analysis, and the branch supports for constrained clades are shown in parentheses. The MrBayes topology is shown, arbitrarily rooted on the clade containing orthologues of Sec20. The amino acid substitution models specified for this analysis were Mixed for MrBayes, and LG4X for RAxML. Support for grouping of N-terminal Qb domains (SNAPN) of Qbc SNAREs with Vti1 and NPSN is strong. The prior probability for the clade containing NPSN and SNAPN is low, while the bootstrap percentage is marginal. SNAPN sequences here include representatives from Metazoa, Archaeplastida, and Kinetoplastida.

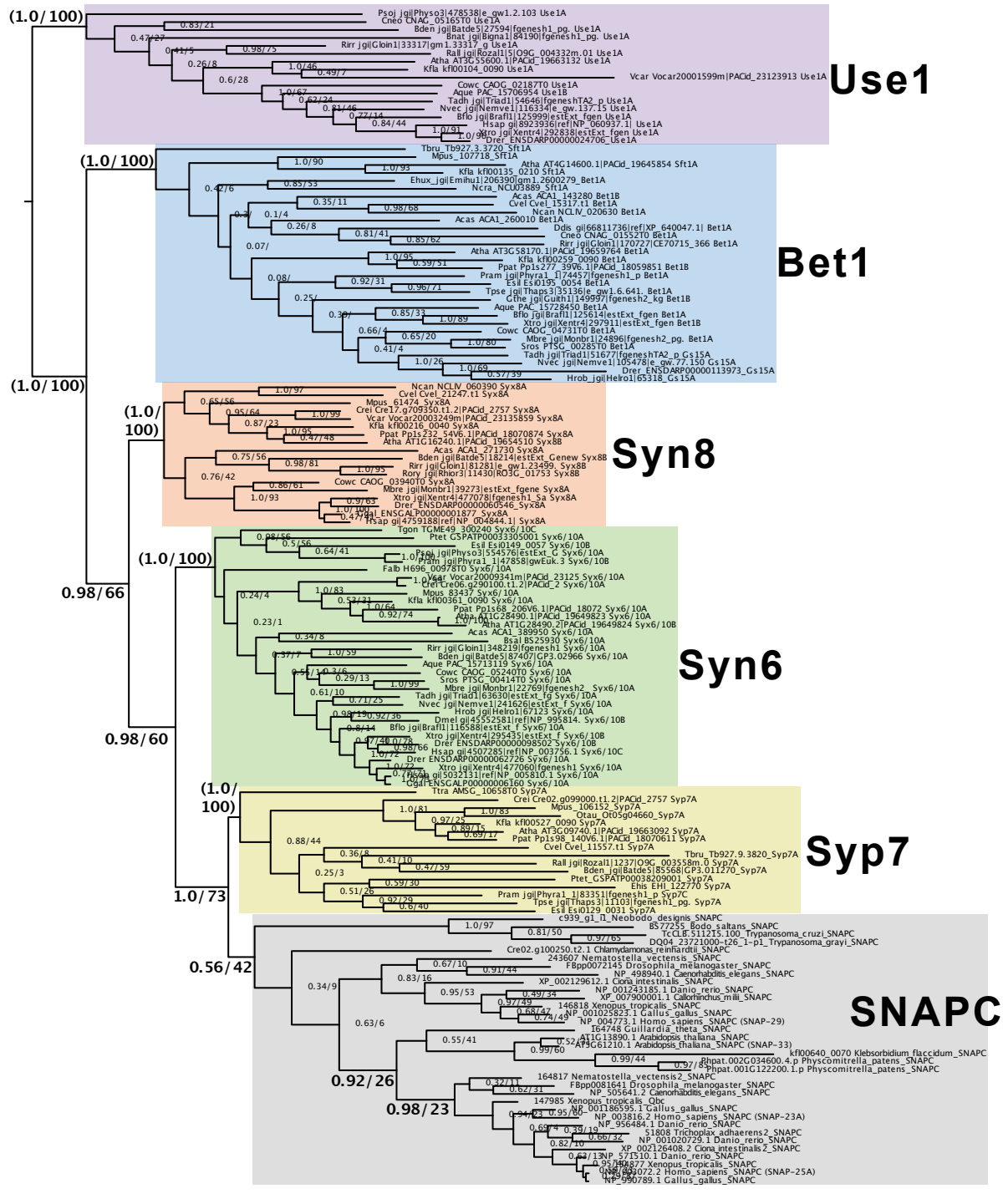

Use1
Bet1
Syn8
Syn6
Syp7
SNAPC
0.6

Figure 5.4: Phylogenetic analysis of Qc SNAREs and the C-terminal (Qc) SNARE domain of identified Qbc SNAREs (SNAPC). SNAPC sequences were appended to a previous alignment of Qc SNAREs from from a broad sampling of eukaryotes (Venkatesh *et al.*, 2017) (see Appendix A chapter 5 Alignment 3). Branch support values are from MrBayes (posterior probability, maximum 1.0 with 0.8 generally being considered the minimum significant value) and RAxML (nonparametric bootstrap values, maximum 100 with 50 generally being considered the minimum significant value). Clades containing previously classified orthologues (Venkatesh *et al.*, 2017) were constrained in this analysis, and the branch supports for constrained clades are shown in parentheses. The MrBayes topology is shown, arbitrarily rooted on the clade containing orthologues of Use1. The amino acid substitution models specified for this analysis were Mixed for MrBayes, and LG4X for RAxML. Prior probability support for grouping of C-terminal Qc domains (SNAPC) of Qbc SNAREs with Syp7 is strong (full support), and the bootstrap percentage is significant. SNAPC sequences here include representatives from Metazoa, Archaeplastida, and Kinetoplastida.

5.3.2 Ancestral NPSN, Syp7, and Qbc SNAREs most likely functioned in exocytosis at the plasma membrane

Previous characterization of NPSN and Syp7 orthologues of plants strongly suggests they have ancestral roles in exocytosis at the PM similar to the Qbc SNAREs of plant, mammalian, and yeast cells. First, it was suggested based on sequence similarity of NPSN and Syp7 to Qbc SNAREs that these are ancestral PM SNAREs (Sanderfoot, 2007). Subsequent experimental work with *Arabidopsis* cells confirmed that NPSN11 and Syp71 (plant-specific inparalogues of NPSN and Syp7) form a complex together with a PM syntaxin-related Qa-SNARE, KNOLLE, and together are functionally redundant with a Qbc-SNARE, AtSNAP-33 (Zheng *et al.*, 2002; El Kasmi *et al.*, 2013). These functions appear to be important for formation of the cell plate during plant cell division. Also, in *Arabidopsis* NPSN localizes to the PM, while Syp71 appears to localize to both the ER and PM (Uemura *et al.*, 2004; Suwastika *et al.*, 2008), although expression by the endogenous promoter rather than transient over-expression results in primarily PM localization (Suwastika *et al.*, 2008). Also, it would be very un-parsimonious for NPSN and Syp7 to have an ancestral function at a non-PM organelle and then simultaneously acquire a coordinate functional redundancy with AtSNAP-33 at the PM/cell plate in plants.

The phylogenetic relationships revealed between NPSN and the SNAPN domain (Figure 5.3), and between Syp7 and the SNAPC domain (Figure 5.4) further suggest that an ancestral function for these SNAREs similar to the plant orthologues. However, hypotheses regarding conservation of NPSN and Syp7 localization and function among extant eukaryotes may only be tested by experimentation. Characterization of SNAREs in mammalian and yeast model systems has not provided any insight into this, due to the loss of both NPSN and Syp7 in both Metazoa and Fungi (Venkatesh *et al.*, 2017). To our knowledge, the only relevant previously reported results are a localization of a Syp7 orthologue in *Trypanosoma brucei*, which indicates localization near the posterior face of the nucleus, however the authors described this result as "equivocal" (Murungi *et al.*, 2014). Further work in a different model system would provide further justification for inference of an ancestral function of NPSN and Syp7 at the PM.

Therefore, to investigate the ancestral function of NPSN and Syp7 further, I worked with Prof. Robert Kay and Dr. Gareth Bloomfield at the MRC-Laboratory of Molecular Biology (for six weeks) to attempt a preliminary characterization of NPSN and Syp7 in the amoebozoan *Dictyostelium discoideum*. *D. discoideum* is a tractable and relevant non-plant cell biological model system in which to function of NPSN and Syp7 for the following reasons: First, it possesses both a homologue of NPSN (Dictybase ID DDB_G0285365) and of Syp7 (DDB_G0289063) (in contrast

to animals and fungi, as well as other ancestral SNAREs and trafficking proteins (Venkatesh *et al.*, 2017). Correspondingly, it also possesses a complete array of ancestral organelles (with the exception of the flagellum). Second, it is a member of the supergroup Amoebozoa, which is distantly related to plants, and sister to the Opisthokonta (see Figure 1.3).

To examine the localization of NPSN and Syp7 in *D. discoideum*, we overexpressed chimeras of both of these proteins tagged at the N-termini with either Green Fluorescence Protein (GFP) or monomeric red fluorescence protein (mCherry). The results for NPSN are somewhat consistent with a plasma membrane localization (Figure 5.5), while the results for Syp7 show retention on internal membranes, potentially ER membranes (Figure 5.6). These results are somewhat comparable to localizations of overexpressed GFP-tagged NPSN and Syp7 in plant cells (Uemura *et al.*, 2004). Interpretation of these results is problematic due to likely issues arising from overexpression of the reporter constructs and the relatively large size of the fluorescent protein tags compared to the SNARE proteins (both approximately 25-30 kD). Thus, these results are inconclusive so far, particularly in the case of *D. discoideum* Syp7.

Further work is needed to follow up on these results, and is possible using existing tools for experimentation with *D. discoideum*. Localization of endogenous NPSN and Syp7 could be investigated by tagging the endogenous loci with an epitope tag such as Hemagglutinin Antigen (HA) using existing integration vectors for knock-in (Mukai *et al.*, 2016). Or, better yet, by growing anti-serum against the cytosolic domains of the endogenous proteins. To confirm ER localization, co-localization immunofluorescence analysis could be performed with ER markers such as GFP-tagged Protein Disulphide Isomerase, as previously (Monnat *et al.*, 2000). Co-precipitation experiments could also be done to identify which SNAREs form heterotetrameric complexes with NPSN or Syp7, as has been done for *Arabidopsis thaliana* cells (El Kasmi *et al.*, 2013). In this case the prediction would be that NPSN and Syp7 would be found in the same SNARE complex, and that a PM Qa-SNARE (DDB0233589 and/or DDB0233590) would also be a member of the same complex. Another approach would be gene knockdown. Blockage of SNARE-mediated exocytosis at the plasma membrane in *D. discoideum* results in a readily observable phenotype with enlarged contractile vacuoles (Zanchi *et al.*, 2010), and methods are available for RNAi gene knockdown (Martens *et al.*, 2002; Friedrich *et al.*, 2015). So, knockdown of NPSN or Syp7 could be performed to test the prediction of exocytic function. However, *Dictyostelium discoideum* possesses two Qbc-SNAREs (DDB0237970 and DDB0214988), which could perhaps compensate for the loss of NPSN or Syp7. One of these genes has been named cspA (which encodes for culmination specific protein 37D) (Loughran *et al.*, 2000), although the function of neither of these genes has

been experimentally characterized. For this reason simultaneous knockdowns of NPSN or Syp7 and these Qbc-SNAREs might have to be performed to see a clear phenotype, as was the case for NPSN11, Syp71, and SNAP-33 in *A. thaliana* cells (El Kasmi *et al.*, 2013).

While it is possible that NPSN and Syp7 orthologues in *D. discoideum* or other eukaryotes have different localizations and functions than those in *Arabidopsis thaliana*, the leading hypothesis remains that the functions of these proteins in *A. thaliana* represent ancestral, conserved pan-eukaryotic functions. Additional suggestive evidence for this is that the expression profile of NPSN and Syp7 genes over the lifecycle (development of aggregative multicellular stalks) of both *D. discoideum* and *Dictyostelium purpureum* are comparable (Parikh *et al.*, 2010). Also, the mechanism by which Syp7 is exported from the ER in *A. thaliana* cells suggests an additional reason why the overexpressed GFP-tagged Syp7 localized primarily to the ER in *A. thaliana* cells: Endogenous Syp7 forms a complex with NPSN11 and the Qa-SNARE KNOLLE in the ER before export to the Golgi (Karnahl *et al.*, 2017). If a similar mechanism occurs in *D. discoideum*, then this export mechanism could be oversaturated by the overexpression of Syp7 chimeras, resulting in a small proportion successfully exiting the ER, as was observed when GFP-Syp71 was overexpressed in *A. thaliana* (Uemura *et al.*, 2004).

Considering work on the localization and function of NPSN and Syp7 orthologues to date, together with the findings in Section 5.3.1, the simplest hypothesis for the evolution of NPSN, Syp7, and Qbc SNAREs involves functional redundancy of NPSN/Syp7 and Qbc SNAREs followed by replacement of NPSN/Syp7 function entirely by Qbc SNAREs in some lineages, as previously suggested (Sanderfoot, 2007). The phylogenetic results herein suggest a more detailed scenario: prior to the divergence of extant eukaryotic lineages (*i.e.*, at a time point predating the LECA) a duplicate of an NPSN-encoding gene fused with a duplicate of a Syp7-encoding gene, resulting in a cytosolic Qbc SNARE capable of mediating fusion through binding the same Qa and R SNARE bound by NPSN and Syp7 (as is the case in plants (El Kasmi *et al.*, 2013)). Also, redundancy of Qbc SNAREs with NPSN/Syp7 may partially explain the coordinated loss of both NPSN and Syp7 independently in Metazoa and Fungi (Sanderfoot, 2007; Kienle *et al.*, 2009; Venkatesh *et al.*, 2017).

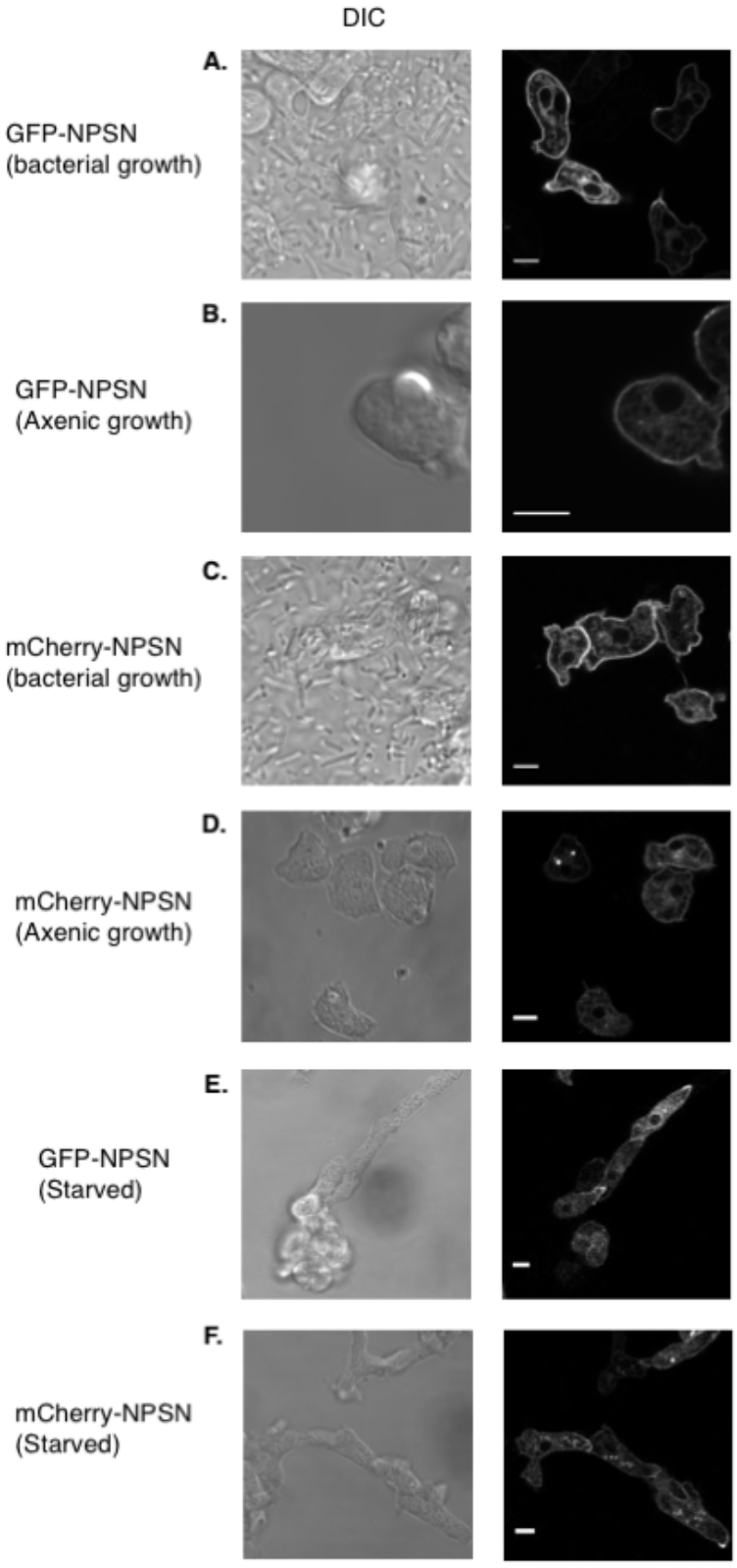

DIC
A.
GFP-NPSN
(bacterial growth)
B.
GFP-NPSN
(Axenic growth)
C.
mCherry-NPSN
(bacterial growth)
D.
mCherry-NPSN
(Axenic growth)
E.
GFP-NPSN
(Starved)
F.
mCherry-NPSN
(Starved)

Figure 5.5: NPSN reporters localize primarily to the plasma membrane. *Dictyostelium discoideum* AX2 cells over-expressing GFP or mCherry reporter constructs were generated using standard cloning methods, extrachromosomal vectors with G418 resistance markers, and transformation by electroporation. Cells were visualized using a Zeiss 710 confocal microscope. Cells were grown in either axenic media, buffer containing *Klebsiella pneumoniae* bacteria, or nutrient-free starvation buffer for 2 hours. Starved cells have entered the aggregation stage of the *Dictyostelium* lifecycle, and the end-to-end streaming activity shows concentration of fluorescent signal at the uropod. Scale bars are 5μm.

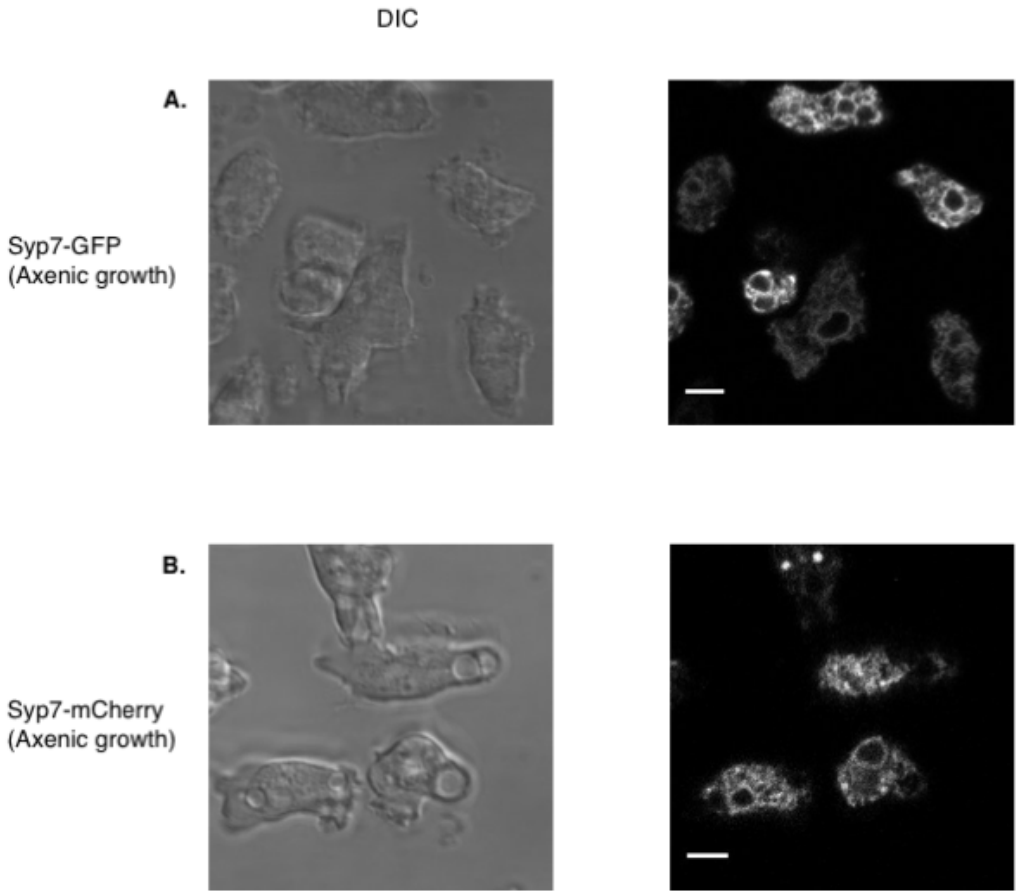


Figure 5.6: Syp7 reporters localize primarily to intracellular membranes, potentially including the endoplasmic reticulum. *Dictyostelium discoideum* AX2 strains over-expressing GFP or mCherry reporter constructs were generated using standard cloning methods, extrachromosomal vectors with G418 resistance markers, and transformation by electroporation. Cells were visualized using a Zeiss 710 confocal microscope. Scale bars are 5μm.

5.3.3 Phylogenetic analysis confirms that the complete set of ancestral SNAREs includes an orthologue of the R-SNARE Synaptobrevin

As mentioned above, in mammalian, yeast, and plant cells, certain R-SNAREs are more specific to exocytosis while others are more specific to endosomes (see section 5.2.2.1 and 5.2.2.2). However, Sanderfoot (2007) showed that the exocytosis-specific (VAMP72) and endosome-specific (VAMP71) R-SNAREs in plants are inparalogues of an ancestral VAMP7-like R-SNARE. This raises the possibility that the counterparts of these VAMP7-related SNAREs in mammalian and yeast cells are also inparalogues, and that the LECA possessed a single R-SNARE for fusion at both the PM and endosomes. However, previous similarity search-based analyses suggested that the exocytosis-specific synaptobrevin-like R-SNAREs of mammalian cells (VAMP1 to 5) and yeast cells (Snc1 and 2) have orthologues in diverse eukaryotes including Amoebozoa and ciliates, but not plants (Yoshizawa *et al.*, 2006; Kloepper *et al.*, 2007). Putative orthologues of this R-SNARE were identified in genomes representing Holozoa, Fungi, Amoebozoa, and ciliates (Kloepper *et al.*, 2007). However, no phylogenetic analyses supported these findings.

The function of putative synaptobrevin orthologues outside of opisthokonts is unknown, however work has been done to localize some of these proteins. A PhD thesis by Margaret Clotworthy from the laboratory of Robert Kay reported localization of a putative synaptobrevin as well as a VAMP7 orthologue in *Dictyostelium discoideum* using overexpression of N-terminally tagged chimeric proteins (Clotworthy, 2005). The results therein suggest localization of synaptobrevin to small vesicles and the plasma membrane and exclusion of VAMP7 from the plasma membrane and localization instead to larger internal organelles (Clotworthy (2005) figure 21). The VAMP7 localization is consistent with another study showing localization of VAMP7 to endocytic organelles (Bennett *et al.*, 2008). Moreover, the R-SNAREs previously identified as synaptobrevin orthologues in the ciliate *Paramecium tetraurelia* (Kloepper *et al.*, 2007), Syb8-1, Syb9-1, Syb10-1, and Syb11-1, generally appear to localize to either the plasma membrane or early food vacuoles/phagosomes, but not to contractile vacuoles (Schilde *et al.*, 2010, 2006; Plattner, 2010; Kissmehl *et al.*, 2007). This is perhaps consistent with conservation of Synaptobrevin orthologues in the Diaphoretickes branch of the eukaryotic tree, with comparable functions to those in the Amorphea, which includes opisthokonts and amoebozoans.

Nevertheless, strong (phylogenetic) evidence for orthology of these secretory R-SNAREs has remained elusive, and it is possible that these putative Synaptobrevin orthologues are simply lineage-specific paralogues of VAMP7. In any case, such a scenario appears to have occurred in the evolution of plant SNAREs. Analyses of archaeplastid SNAREs suggests that the plant R-SNARE

VAMP72 represents a plant-specific paralogue of VAMP7 (Sanderfoot, 2007), and *Arabidopsis thaliana* inparalogues localize to the plasma membrane (Uemura *et al.*, 2004) and are involved in secretion, forming a complex with a plasma membrane Qa-SNARE, and the Qbc SNARE SNAP33 or a combination of NPSN and Syp7 (Kwon *et al.*, 2008; El Kasmi *et al.*, 2013). It has been suggested that this function is somewhat analogous to the function of Synaptobrevin (Kwon *et al.*, 2008). The latest attempt to classify R-SNAREs lacking longin domains from diverse eukaryotes resulted in little resolution of this issue, but did show some evidence for grouping of the human brevins VAMP1, 2, 3, and 5 with *Saccharomyces* SNC1 and SNC2 (Venkatesh *et al.*, 2017).

To address the question of whether synaptobrevin originated prior to the divergence of eukaryotes we first analyzed the putative Synaptobrevin orthologues identified by Kloepper *et al.* (2007). The topology identified by bayesian analysis places putative VAMP7 and synaptobrevin orthologues in separate clades (Figure 5.7), consistent with the classifications reported by (Kloepper *et al.*, 2007). However, the prior probability for this bifurcation is marginal at 0.77 out of 1.0, and the bootstrap percentage was only 22.

The ambiguous results of this analysis prompts further investigation with different methods. Topology testing offers a means to test whether alternative phylogenetic hypotheses may be rejected as less likely than others, with measures of statistical significance (Goldman *et al.*, 2000). An advantage of topology testing is that *a priori* hypotheses, such as the pre-LECA versus lineage-specific origins of putative synaptobrevin orthologues, may be explicitly tested. Various topology tests including the Approximately Unbiased (AU) test (Shimodaira, 2002) are widely used in many types of phylogenetic analyses. Examples of applications include testing for mitochondrial vs. non-mitochondrial origins of eukaryotic genes (Hug *et al.*, 2010), testing alternative hypotheses for the branching order of early metazoan lineages (Whelan *et al.*, 2015), and testing the hypothesis that peroxisomes have a bacterial (endosymbiotic) origin (Gabaldón and Capella-Gutiérrez, 2010).

Thus, to specifically test the hypotheses that these putative synaptobrevin orthologues have a common pre-LECA origin distinct from that of VAMP7, or are lineage-specific inparalogues of VAMP7, we performed AU tests to compare the best ML tree against the best ML trees found when constraints were applied (as indicated in Figure 5.7 C and D). The best ML tree for independent origins of the putative synaptobrevin orthologues in ciliates, Metazoa, Fungi, and Amoebozoa was rejected with a P-value below a significance threshold of 0.05. However, the best ML tree involving independent origins of synaptobrevins in ciliates and amorphea (*i.e.*, an ancestor of Metazoa, Fungi, and Amoebozoa, but not ciliates) was not rejected, thus leaving open the possibility of

independent origins of putative synaptobrevin orthologues.

Also, we noticed that a defining feature of Synaptobrevin orthologues outside of Opisthokonta seems to be a residue other than Arginine (R) at the zero layer position in the domain, usually an Aspargine (N) (Figure 5.8). This pattern could contribute to uncertainty in distinguishing opisthokont Synaptobrevin orthologues as orthologous to non-opisthokont Synaptobrevin orthologues (assuming they are in fact orthologous). The arginine residue of Snc does not appear to be essential in a yeast SNARE complex (Katz and Brennwald, 2000). Also, the *P. tetraurelia* Synaptobrevin orthologues contain longin domains, and are thus not "brevins". This suggests that absence of a longin domain may not be a shared ancestral feature of putative synaptobrevin orthologues. Instead, the longin domain may be lost several times independently, as suggested previously (Venkatesh *et al.*, 2017).

To further investigate the origins of putative synaptobrevin orthologues, we analyzed amorphean R-SNAREs and pan-eukaryotic R-SNAREs. An alignment of representative amorphean R-SNAREs yielded support for distinct VAMP7 and synaptobrevin clades both including sequences representing the amoebozoan *Dictyostelium purpureum*, the holomycotan *Fonticula alba*, as well as metazoa and fungi (Figure 5.9). The synaptobrevin clade includes the human proteins VAMP1 to 4. This result further refutes the hypothesis of independent origins of synaptobrevin-like R-SNAREs within Amorphea. An alignment of representative R-SNAREs from pan-eukaryotic taxa similarly yielded prior probability support for distinct VAMP7 and synaptobrevin clades (although bootstrap support is marginal) (Figure 5.10). The synaptobrevin clade includes sequences representing the additional taxonomic groups Archaeplastida, SAR, Haptophyta, Cryptophyta, and Metamonada, but not Discoba. However, support for in inclusion of opisthokont sequences in this clade was only seen when all opisthokont synaptobrevin sequences but those from the holomycotans *F. alba*, *Rozella allomyces*, and *Paraphelidium tribonemae* were removed (also seen in Figure 5.9). This perhaps reflects divergence of opisthokont synaptobrevin orthologues.

The lack of conservation of the central arginine residue seen in putative non-opisthokont synaptobrevin orthologues (Figure 5.8) is consistent among those observed in more inclusive alignments of Syb sequences classified with high confidence (Figure 5.10 and Appendix A chapter 5 Alignment 6). Also, most of the identified Synaptobrevin orthologues from across eukaryotes do contain longin domains detectable using InterProScan `https://www.ebi.ac.uk/interpro/search/sequence-search` (Figure 5.10). This indicates that synaptobrevin orthologues are not all brevins, confirming that this is not a defining feature of such orthologues. However, inspection of the sequence alignments reveals much lower levels of conservation in the longin domain

of synaptobrevin orthologues in comparison to VAMP7 orthologues. These considerations, loss of longin domains and conservation or secondary acquisition of a central arginine residue in opisthokont synaptobrevin orthologues, may account for the relatively low phylogenetic affinity of opisthokont synaptobrevin orthologues for non-opisthokont orthologues.

The above analyses included all R-SNAREs with the exception of Tomosyn (Figures 5.9 and 5.10). As mentioned above, tomosyn is a non-canonical R-SNARE domain containing protein. In addition to the SNARE domain, tomosyn contains a characteristic large domain on the N-terminal side of the SNARE domain, which makes tomosyn much larger than canonical SNAREs (Pobbati *et al.*, 2004). In mammalian cells tomosyn forms non-fusogenic SNARE complexes with plasma membrane Qa-SNAREs and SNAP-25, forming a tetrameric coiled-coil structure highly similar to that of the functional exocytic SNARE complex (Hatsuzawa *et al.*, 2003; Pobbati *et al.*, 2004). The role of tomosyn appears to be in regulating exocytosis (Hatsuzawa *et al.*, 2003; Pobbati *et al.*, 2004; Williams *et al.*, 2011). N-terminal domain of tomosyn is involved in binding to the syntaxin N-terminus, while the C-terminal R-SNARE domain binds to the Q-SNARE domains inhibiting formation of fusogenic SNARE complexes (Yu *et al.*, 2014). In addition, Amysin is a vertebrate-specific paralogue of Tomosyn (Scales *et al.*, 2002). Both Tomosyn and Amysin bind Syntaxin 1a and SNAP-25 forming non-fusogenic SNARE complexes in place of VAMP2 (synaptobrevin 2) (Scales *et al.*, 2002).

The interaction of tomosyn and amysin with exocytic SNAREs is consistent with and indicative of a close evolutionary relationship between these proteins and Synaptobrevin (specifically VAMP2). Also, Tomosyn orthologues have previously been readily identified in diverse eukaryotes (Yoshizawa *et al.*, 2006; Sanderfoot, 2007; Kloepper *et al.*, 2007). Thus, to investigate the potential relationship between synaptobrevin and tomosyn, we first identified positive hits for tomosyn among our similarity search results (as for synaptobrevin and Qbc orthologues). We aligned these sequences and extracted the SNARE domains using a reference SNARE domain alignment as a guide (Fasshauer *et al.*, 1998). We then added these tomosyn SNARE domain sequences to the alignment used for the analysis shown in Figure 5.10, and performed phylogenetic analyses as before. The resulting topology, arbitrarily rooted on the clade containing orthologues of Ykt6 and Sec22, indicates that Tomosyn orthologues branched sister to synaptobrevin orthologues in the tree of pan-eukaryotic R-SNAREs (Figure 5.11). Non-parametric bootstrap support for this relationship is insignificant, similar to the results shown above for phylogenetic placement of the SNAPN and SNAPC domains (see Figures 5.3 and 5.4). This is likely due in part to the absence of homologous positions extending beyond the SNARE domains of tomosyn sequences in the align-

ment.

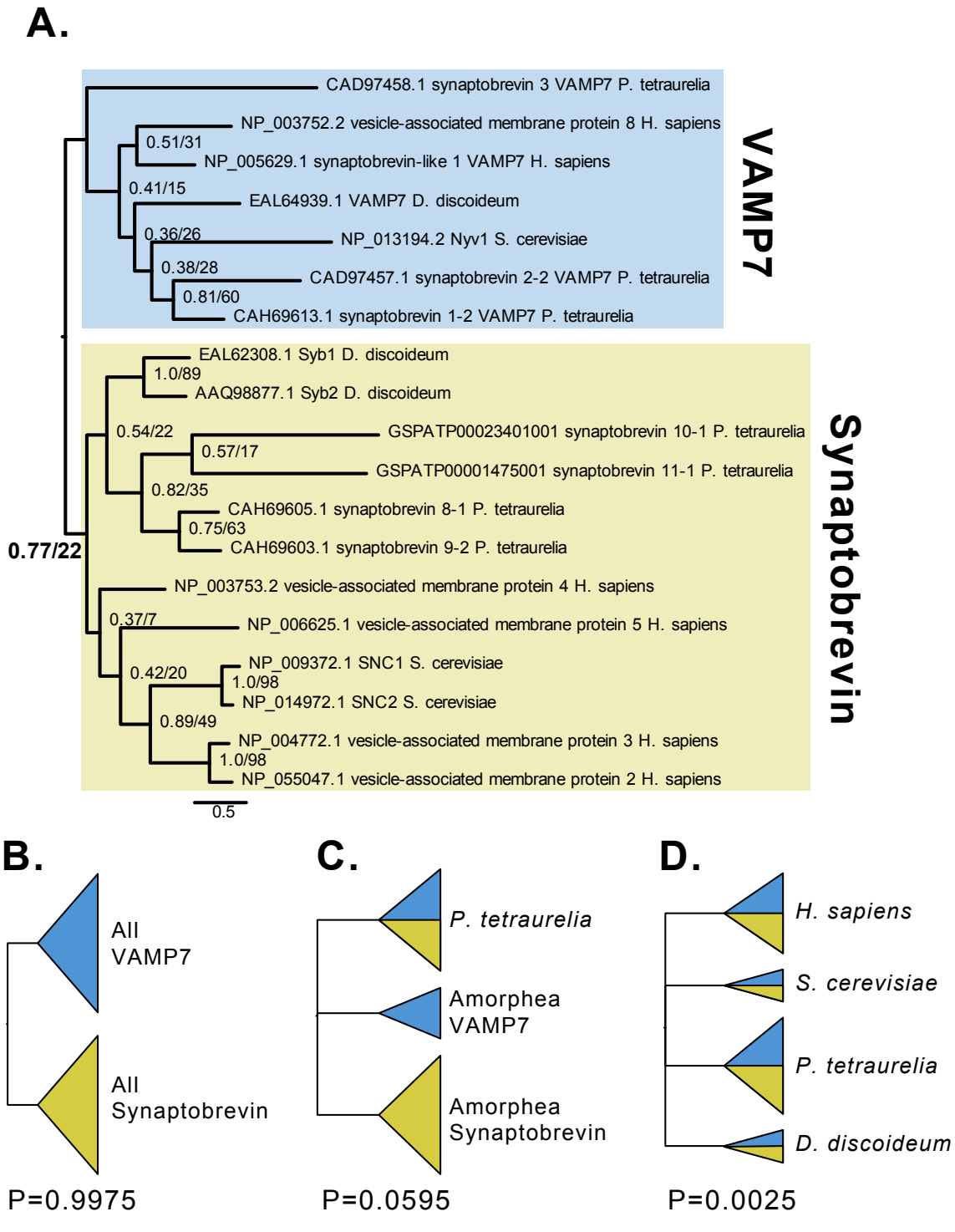

A.
CAD97458.1 synaptobrevin 3 VAMP7 P. tetraurelia
NP_003752.2 vesicle-associated membrane protein 8 H. sapiens
0.51/31
NP_005629.1 synaptobrevin-like 1 VAMP7 H. sapiens
0.41/15
EAL64939.1 VAMP7 D. discoideum
0.36/26
NP_013194.2 Nyv1 S. cerevisiae
0.38/28
CAD97457.1 synaptobrevin 2-2 VAMP7 P. tetraurelia
0.81/60
CAH69613.1 synaptobrevin 1-2 VAMP7 P. tetraurelia
VAMP7
EAL62308.1 Syb1 D. discoideum
1.0/89
AAQ98877.1 Syb2 D. discoideum
0.54/22
GSPATP00023401001 synaptobrevin 10-1 P. tetraurelia
0.57/17
GSPATP00001475001 synaptobrevin 11-1 P. tetraurelia
0.82/35
CAH69605.1 synaptobrevin 8-1 P. tetraurelia
0.75/63
CAH69603.1 synaptobrevin 9-2 P. tetraurelia
0.77/22
NP_003753.2 vesicle-associated membrane protein 4 H. sapiens
0.37/7
NP_006625.1 vesicle-associated membrane protein 5 H. sapiens
0.42/20
NP_009372.1 SNC1 S. cerevisiae
1.0/98
NP_014972.1 SNC2 S. cerevisiae
0.89/49
NP_004772.1 vesicle-associated membrane protein 3 H. sapiens
1.0/98
NP_055047.1 vesicle-associated membrane protein 2 H. sapiens
Synaptobrevin
0.5
B.
All VAMP7
All Synaptobrevin
P=0.9975
C.
P. tetraurelia
Amorphea VAMP7
Amorphea Synaptobrevin
P=0.0595
D.
H. sapiens
S. cerevisiae
P. tetraurelia
D. discoideum
P=0.0025

Figure 5.7: Phylogenetic analysis of previously reported Synaptobrevin and VAMP7 orthologues. Previously reported synaptobrevin and VAMP7 orthologues (Kloepper *et al.*, 2007) were aligned for phylogenetic analysis (see Appendix A chapter 5 Alignment 4). (**A**) MrBayes topology showing both posterior probabilities and bootstrap percentages from IQ-tree nonparametric bootstrap analysis with 100 pseudoreplicates. The prior probability of 0.77 for monophyly of the putative Synaptobrevin orthologues is marginal, while the bootstrap percentage of 22 is insignificant. (**B**) Diagram of the phylogenetic hypothesis suggested by the topology found in (A), and suggested previously(Kloepper *et al.*, 2007), that the putative Synaptobrevin orthologues form a monophyletic clade to the exclusion of VAMP7 orthologues. (**C**) Diagram of an alternative hypothesis in which Synaptobrevin orthologues are lineage-specific inparalogues of VAMP7 arising independently in Amorphea and ciliates (or some ancestor of *P. tetraurelia* that is not an ancestor of Amorphea). (**D**) Diagram of an alternative hypothesis in which putative Synaptobrevin orthologues are lineage-specific inparalogues of VAMP7 in Metazoa, Fungi, Amoebozoa, and ciliates. The hypotheses represented by (C) and (D) were tested against the topology shown in (A) and (B) using the same sequence alignment as for (A) with an Approximately Unbiased (AU) test. The hypothesis represented in (D) was rejected with a p-value of 0.0025, which is below the standard significance level of 0.05, suggesting that the putative Synaptobrevin orthologues are not all lineage-specific inparalogues of VAMP7. However, the hypothesis shown in (C) was not rejected, as the p-value was a marginal 0.0595, leaving the question of whether Synaptobrevin originated in a common ancestor of Amorphea and ciliates unresolved.

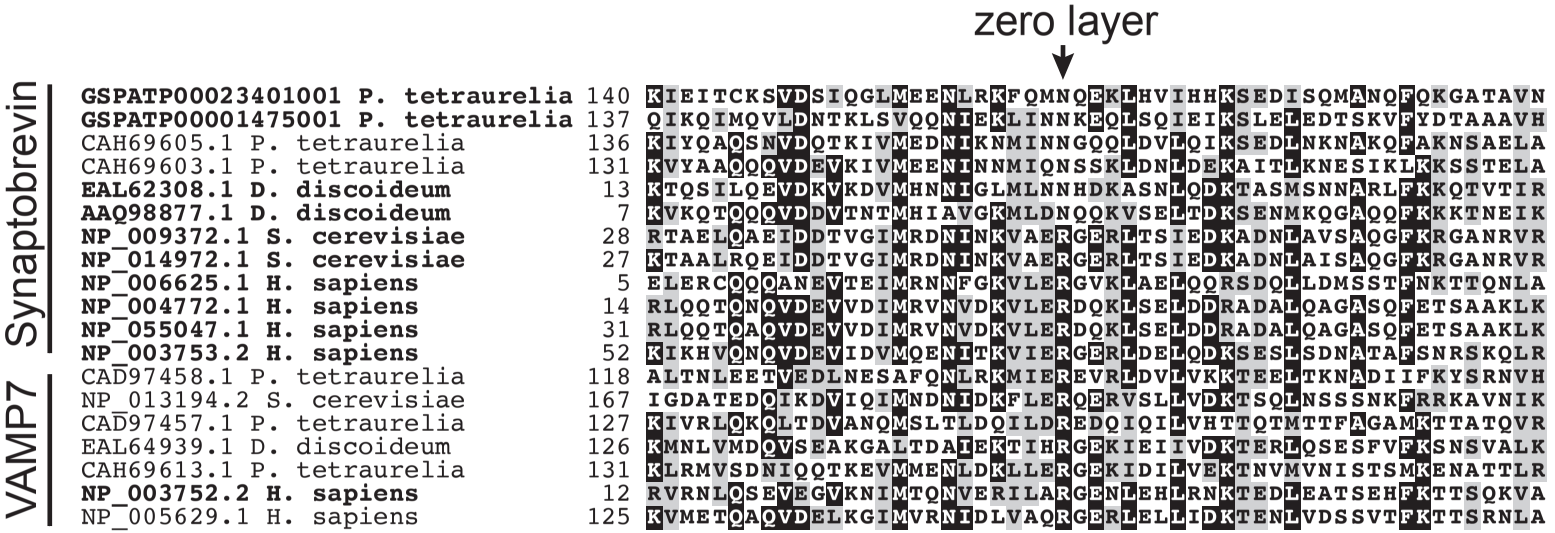


Figure 5.8: Alignment of VAMP7 and Synaptobrevin SNARE domain amino acid sequences from the same dataset as for Figure 5.7. Numbers immediately to the left of each sequence indicate the residue number of the first residue shown for that sequence. This shows that the *P. tetraurelia* sequences have more residues on the N-terminal side of the SNARE domain compared to the Synaptobrevin orthologues from the amorphean species (*D. disoideum*, *S. cerevisiae*, and *H. sapiens*). Comparison of residues at the indicated position representing the zero layer of the SNARE domains, as defined previously (Fasshauer *et al.*, 1998), shows that the *P. tetraurelia* and *D. discoideum* sequences have Aspargine (N) amino acids at this position, instead of the Arginine (R) amino acid that is characteristic of R-SNAREs including VAMP7 orthologues and opisthokont orthologues of Synaptobrevin. Sequences with names shown in bold font do not contain an N-terminal longin domain identifiable by InterProScan `https://www.ebi.ac.uk/interpro/search/sequence-search` (Hunter *et al.*, 2009). Conservation of a longin domain in at least two putative *P. tetraurelia* Synaptobrevin orthologues suggests absence of this domain may not be a defining feature of Synaptobrevin orthologues. Figure generated using BoxShade `https://embnet.vital-it.ch/software/BOX_form.html`.

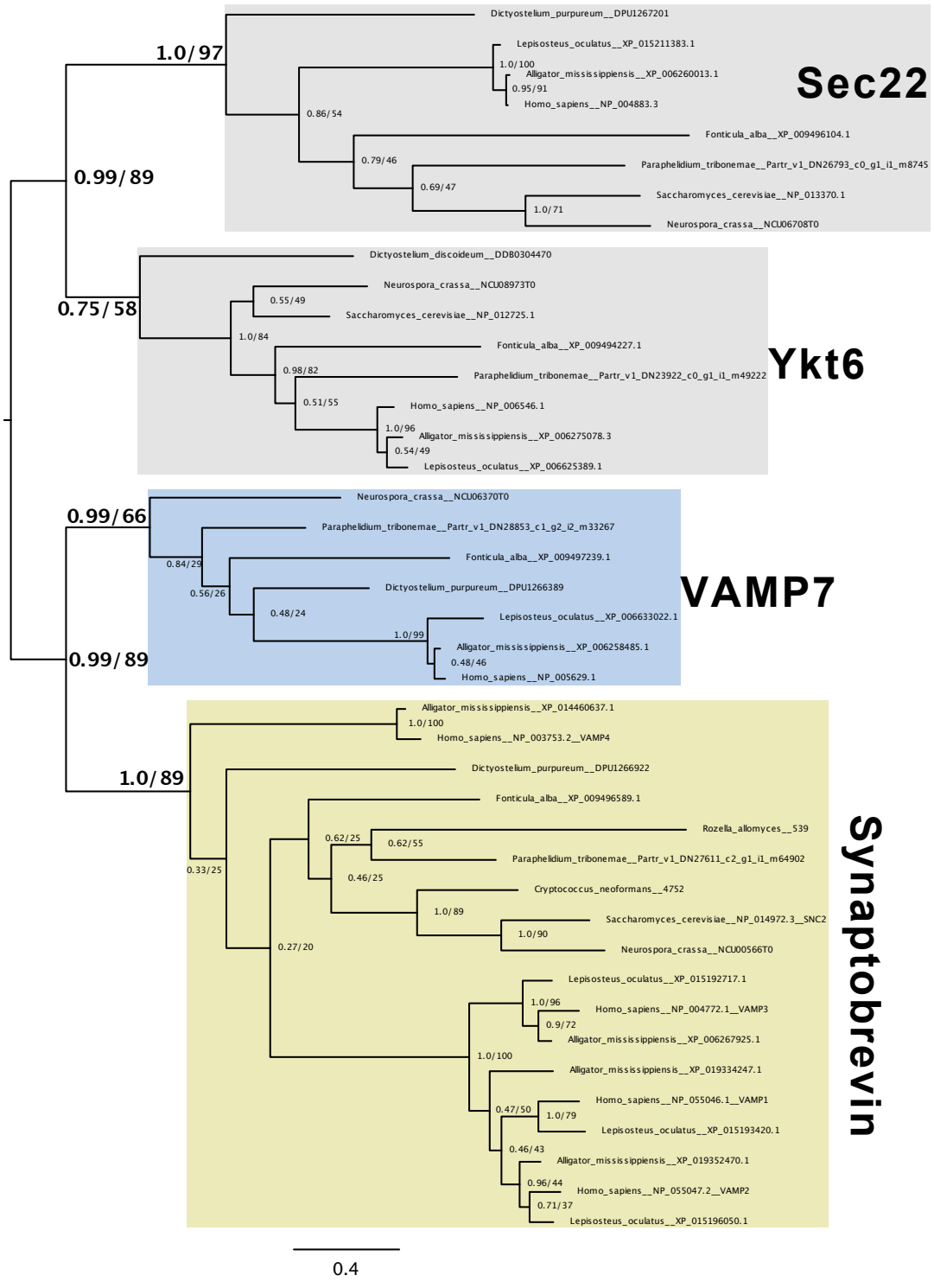

Sec22
Dictyostelium_purpureum__DPU1267201
Lepisosteus_oculatus__XP_015211383.1
Alligator_mississippiensis__XP_006260013.1
Homo_sapiens__NP_004883.3
Fonticula_alba__XP_009496104.1
Paraphelidium_tribonemae__Partr_v1_DN26793_c0_g1_i1_m8745
Saccharomyces_cerevisiae__NP_013370.1
Neurospora_crassa__NCU06708T0
1.0/97
1.0/100
0.95/91
0.86/54
0.79/46
0.69/47
1.0/71
0.99/89
Ykt6
Dictyostelium_discoideum__DDB0304470
Neurospora_crassa__NCU08973T0
Saccharomyces_cerevisiae__NP_012725.1
Fonticula_alba__XP_009494227.1
Paraphelidium_tribonemae__Partr_v1_DN23922_c0_g1_i1_m49222
Homo_sapiens__NP_006546.1
Alligator_mississippiensis__XP_006275078.3
Lepisosteus_oculatus__XP_006625389.1
0.75/58
0.55/49
1.0/84
0.98/82
0.51/55
1.0/96
0.54/49
VAMP7
Neurospora_crassa__NCU06370T0
Paraphelidium_tribonemae__Partr_v1_DN28853_c1_g2_i2_m33267
Fonticula_alba__XP_009497239.1
Dictyostelium_purpureum__DPU1266389
Lepisosteus_oculatus__XP_006633022.1
Alligator_mississippiensis__XP_006258485.1
Homo_sapiens__NP_005629.1
0.99/66
0.84/29
0.56/26
0.48/24
1.0/99
0.48/46
0.99/89
Synaptobrevin
Alligator_mississippiensis__XP_014460637.1
Homo_sapiens__NP_003753.2__VAMP4
Dictyostelium_purpureum__DPU1266922
Fonticula_alba__XP_009496589.1
Rozella_allomyces__539
Paraphelidium_tribonemae__Partr_v1_DN27611_c2_g1_i1_m64902
Cryptococcus_neoformans__4752
Saccharomyces_cerevisiae__NP_014972.3__SNC2
Neurospora_crassa__NCU00566T0
Lepisosteus_oculatus__XP_015192717.1
Homo_sapiens__NP_004772.1__VAMP3
Alligator_mississippiensis__XP_006267925.1
Alligator_mississippiensis__XP_019334247.1
Homo_sapiens__NP_055046.1__VAMP1
Lepisosteus_oculatus__XP_015193420.1
Alligator_mississippiensis__XP_019352470.1
Homo_sapiens__NP_055047.2__VAMP2
Lepisosteus_oculatus__XP_015196050.1
1.0/89
1.0/100
0.62/25
0.62/55
0.33/25
0.46/25
1.0/89
1.0/90
0.27/20
1.0/96
0.9/72
1.0/100
0.47/50
1.0/79
0.46/43
0.96/44
0.71/37
0.4

Figure 5.9: Phylogenetic analysis of amorphean R-SNAREs supports monophyly of putative Synaptobrevin orthologues. (see Appendix A chapter 5 Alignment 5) Node support values are from MrBayes (posterior probability, maximum 1.0 with 0.8 generally being considered the minimum significant value) and IQ-TREE (nonparametric bootstrap values, maximum 100 with 50 being the minimum significant value). The MrBayes topology is shown, arbitrarily rooted on the clade containing orthologues of Sec22 and Ykt6. The amino acid substitution model used for this analysis is LG+G4. Significant support was found for a clade containing the *Homo sapiens* proteins VAMP1, 2, 3, and 4, the *Saccharomyces cerevisiae* protein SNC2, as well as putative orthologues of synaptobrevin from other amorphean taxa including the amoebozoan *Dictyostelium discoideum*.

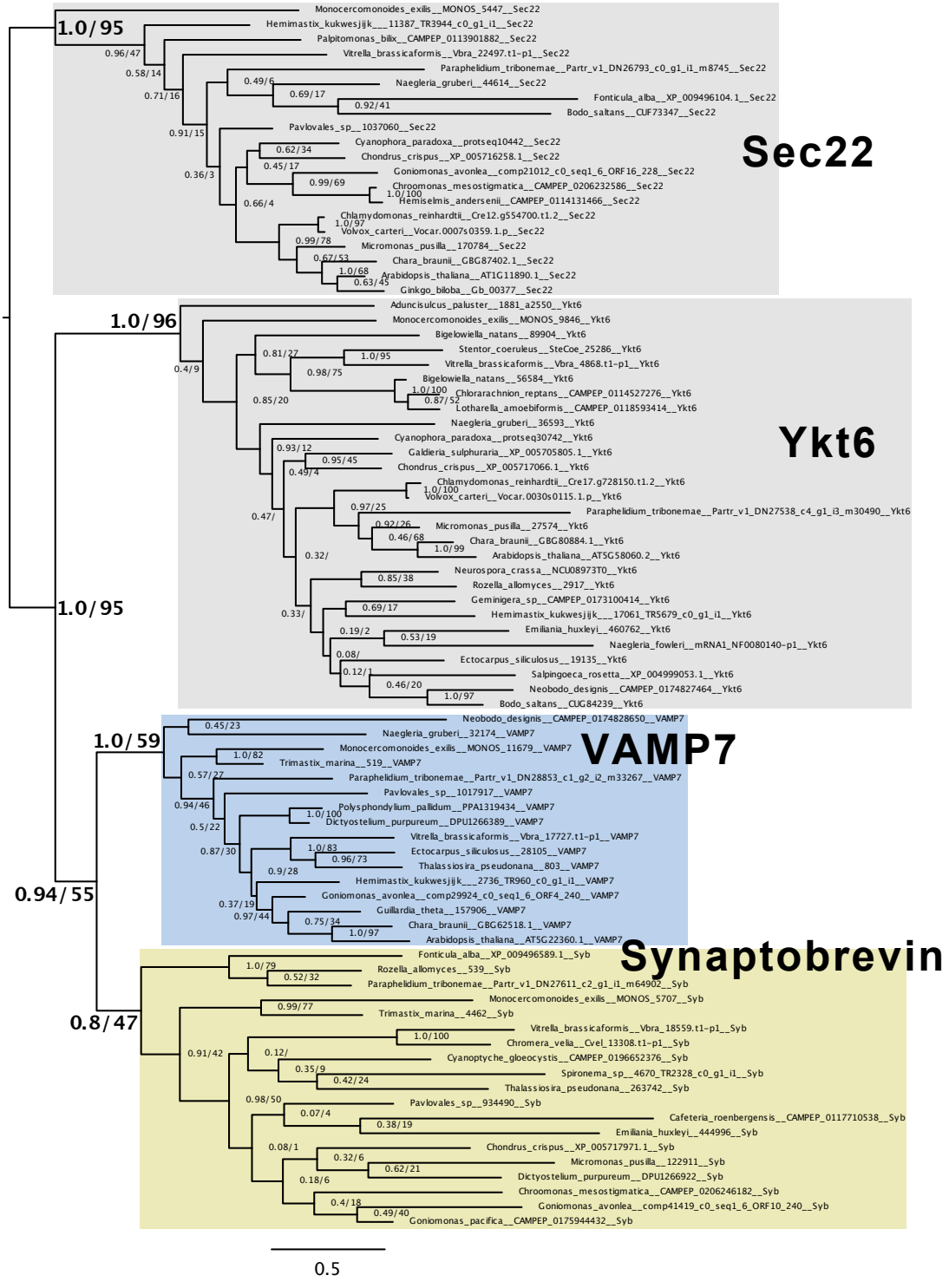

Sec22
Monocercomonoides_exilis__MONOS_5447__Sec22
Hemimastix_kukwesjijk___11387_TR3944_c0_g1_i1__Sec22
Palpitomonas_bilix__CAMPEP_0113901882__Sec22
Vitrella_brassicaformis__Vbra_22497.t1-p1__Sec22
Paraphelidium_tribonemae__Partr_v1_DN26793_c0_g1_i1_m8745__Sec22
Naegleria_gruberi__44614__Sec22
Fonticula_alba__XP_009496104.1__Sec22
Bodo_saltans__CUF73347__Sec22
Pavlovales_sp__1037060__Sec22
Cyanophora_paradoxa__protseq10442__Sec22
Chondrus_crispus__XP_005716258.1__Sec22
Goniomonas_avonlea__comp21012_c0_seq1_6_ORF16_228__Sec22
Chroomonas_mesostigmatica__CAMPEP_0206232586__Sec22
Hemiselmis_andersenii__CAMPEP_0114131466__Sec22
Chlamydomonas_reinhardtii__Cre12.g554700.t1.2__Sec22
Volvox_carteri__Vocar.0007s0359.1.p__Sec22
Micromonas_pusilla__170784__Sec22
Chara_braunii__GBG87402.1__Sec22
Arabidopsis_thaliana__AT1G11890.1__Sec22
Ginkgo_biloba__Gb_00377__Sec22
Ykt6
Aduncisulcus_paluster__1881_a2550__Ykt6
Monocercomonoides_exilis__MONOS_9846__Ykt6
Bigelowiella_natans__89904__Ykt6
Stentor_coeruleus__SteCoe_25286__Ykt6
Vitrella_brassicaformis__Vbra_4868.t1-p1__Ykt6
Bigelowiella_natans__56584__Ykt6
Chlorarachnion_reptans__CAMPEP_0114527276__Ykt6
Lotharella_amoebiformis__CAMPEP_0118593414__Ykt6
Naegleria_gruberi__36593__Ykt6
Cyanophora_paradoxa__protseq30742__Ykt6
Galdieria_sulphuraria__XP_005705805.1__Ykt6
Chondrus_crispus__XP_005717066.1__Ykt6
Chlamydomonas_reinhardtii__Cre17.g728150.t1.2__Ykt6
Volvox_carteri__Vocar.0030s0115.1.p__Ykt6
Paraphelidium_tribonemae__Partr_v1_DN27538_c4_g1_i3_m30490__Ykt6
Micromonas_pusilla__27574__Ykt6
Chara_braunii__GBG80884.1__Ykt6
Arabidopsis_thaliana__AT5G58060.2__Ykt6
Neurospora_crassa__NCU08973T0__Ykt6
Rozella_allomyces__2917__Ykt6
Geminigera_sp__CAMPEP_0173100414__Ykt6
Hemimastix_kukwesjijk___17061_TR5679_c0_g1_i1__Ykt6
Emiliania_huxleyi__460762__Ykt6
Naegleria_fowleri__mRNA1_NF0080140-p1__Ykt6
Ectocarpus_siliculosus__19135__Ykt6
Salpingoeca_rosetta__XP_004999053.1__Ykt6
Neobodo_designis__CAMPEP_0174827464__Ykt6
Bodo_saltans__CUG84239__Ykt6
VAMP7
Neobodo_designis__CAMPEP_0174828650__VAMP7
Naegleria_gruberi__32174__VAMP7
Monocercomonoides_exilis__MONOS_11679__VAMP7
Trimastix_marina__519__VAMP7
Paraphelidium_tribonemae__Partr_v1_DN28853_c1_g2_i2_m33267__VAMP7
Pavlovales_sp__1017917__VAMP7
Polysphondylium_pallidum__PPA1319434__VAMP7
Dictyostelium_purpureum__DPU1266389__VAMP7
Vitrella_brassicaformis__Vbra_17727.t1-p1__VAMP7
Ectocarpus_siliculosus__28105__VAMP7
Thalassiosira_pseudonana__803__VAMP7
Hemimastix_kukwesjijk___2736_TR960_c0_g1_i1__VAMP7
Goniomonas_avonlea__comp29924_c0_seq1_6_ORF4_240__VAMP7
Guillardia_theta__157906__VAMP7
Chara_braunii__GBG62518.1__VAMP7
Arabidopsis_thaliana__AT5G22360.1__VAMP7
Synaptobrevin
Fonticula_alba__XP_009496589.1__Syb
Rozella_allomyces__539__Syb
Paraphelidium_tribonemae__Partr_v1_DN27611_c2_g1_i1_m64902__Syb
Monocercomonoides_exilis__MONOS_5707__Syb
Trimastix_marina__4462__Syb
Vitrella_brassicaformis__Vbra_18559.t1-p1__Syb
Chromera_velia__Cvel_13308.t1-p1__Syb
Cyanoptyche_gloeocystis__CAMPEP_0196652376__Syb
Spironema_sp__4670_TR2328_c0_g1_i1__Syb
Thalassiosira_pseudonana__263742__Syb
Pavlovales_sp__934490__Syb
Cafeteria_roenbergensis__CAMPEP_0117710538__Syb
Emiliania_huxleyi__444996__Syb
Chondrus_crispus__XP_005717971.1__Syb
Micromonas_pusilla__122911__Syb
Dictyostelium_purpureum__DPU1266922__Syb
Chroomonas_mesostigmatica__CAMPEP_0206246182__Syb
Goniomonas_avonlea__comp41419_c0_seq1_6_ORF10_240__Syb
Goniomonas_pacifica__CAMPEP_0175944432__Syb
0.5

Figure 5.10: Phylogenetic analysis of R-SNAREs shows that a Synaptobrevin-like SNARE was present in the ancestor of eukaryotes. (see Appendix A chapter 5 Alignment 6) Node support values are from MrBayes (posterior probability, maximum 1.0 with 0.8 generally being considered the minimum significant value) and IQ-TREE (nonparametric bootstrap values, maximum 100 with 50 being the minimum significant value). The MrBayes topology is shown, arbitrarily rooted on the clade containing orthologues of Sec22. The amino acid substitution model used for this analysis is LG+G4. Significant support was found for the monophyly of sequences from across eukaryotes falling into four clades of canonical R-SNAREs including one composed of orthologues of Synaptobrevin. Due to difficulty in classification, metazoan and yeast Synaptobrevin orthologues were not included in the alignment for this analysis. Nevertheless, Synaptobrevin orthologues from the holomycotans *Fonticula alba*, *Rozella allomyces*, and *Paraphelidium tribonemae*, also included in Figure 5.9, are included in this supported pan-eukaryotic clade of synaptobrevin-orthologues.

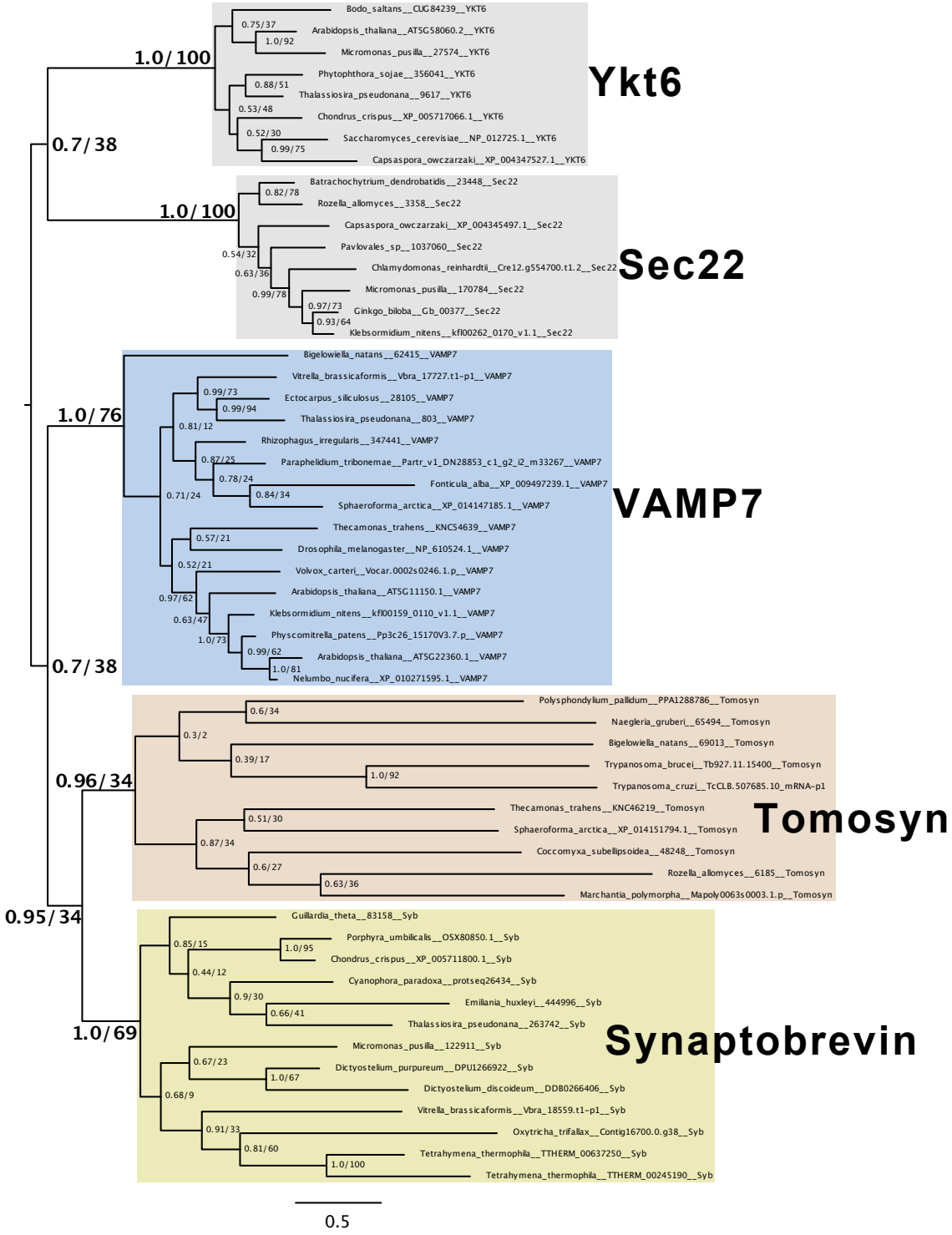

Ykt6
Sec22
VAMP7
Tomosyn
Synaptobrevin
1.0/100
0.7/38
1.0/100
1.0/76
0.7/38
0.96/34
0.95/34
1.0/69
Bodo_saltans__CUG84239__YKT6
Arabidopsis_thaliana__AT5G58060.2__YKT6
Micromonas_pusilla__27574__YKT6
Phytophthora_sojae__356041__YKT6
Thalassiosira_pseudonana__9617__YKT6
Chondrus_crispus__XP_005717066.1__YKT6
Saccharomyces_cerevisiae__NP_012725.1__YKT6
Capsaspora_owczarzaki__XP_004347527.1__YKT6
Batrachochytrium_dendrobatidis__23448__Sec22
Rozella_allomyces__3358__Sec22
Capsaspora_owczarzaki__XP_004345497.1__Sec22
Pavlovales_sp__1037060__Sec22
Chlamydomonas_reinhardtii__Cre12.g554700.t1.2__Sec22
Micromonas_pusilla__170784__Sec22
Ginkgo_biloba__Gb_00377__Sec22
Klebsormidium_nitens__kfl00262_0170_v1.1__Sec22
Bigelowiella_natans__62415__VAMP7
Vitrella_brassicaformis__Vbra_17727.t1-p1__VAMP7
Ectocarpus_siliculosus__28105__VAMP7
Thalassiosira_pseudonana__803__VAMP7
Rhizophagus_irregularis__347441__VAMP7
Paraphelidium_tribonemae__Partr_v1_DN28853_c1_g2_i2_m33267__VAMP7
Fonticula_alba__XP_009497239.1__VAMP7
Sphaeroforma_arctica__XP_014147185.1__VAMP7
Thecamonas_trahens__KNC54639__VAMP7
Drosophila_melanogaster__NP_610524.1__VAMP7
Volvox_carteri__Vocar.0002s0246.1.p__VAMP7
Arabidopsis_thaliana__AT5G11150.1__VAMP7
Klebsormidium_nitens__kfl00159_0110_v1.1__VAMP7
Physcomitrella_patens__Pp3c26_15170V3.7.p__VAMP7
Arabidopsis_thaliana__AT5G22360.1__VAMP7
Nelumbo_nucifera__XP_010271595.1__VAMP7
Polysphondylium_pallidum__PPA1288786__Tomosyn
Naegleria_gruberi__65494__Tomosyn
Bigelowiella_natans__69013__Tomosyn
Trypanosoma_brucei__Tb927.11.15400__Tomosyn
Trypanosoma_cruzi__TcCLB.507685.10_mRNA-p1
Thecamonas_trahens__KNC46219__Tomosyn
Sphaeroforma_arctica__XP_014151794.1__Tomosyn
Coccomyxa_subellipsoidea__48248__Tomosyn
Rozella_allomyces__6185__Tomosyn
Marchantia_polymorpha__Mapoly0063s0003.1.p__Tomosyn
Guillardia_theta__83158__Syb
Porphyra_umbilicalis__OSX80850.1__Syb
Chondrus_crispus__XP_005711800.1__Syb
Cyanophora_paradoxa__protseq26434__Syb
Emiliania_huxleyi__444996__Syb
Thalassiosira_pseudonana__263742__Syb
Micromonas_pusilla__122911__Syb
Dictyostelium_purpureum__DPU1266922__Syb
Dictyostelium_discoideum__DDB0266406__Syb
Vitrella_brassicaformis__Vbra_18559.t1-p1__Syb
Oxytricha_trifallax__Contig16700.0.g38__Syb
Tetrahymena_thermophila__TTHERM_00637250__Syb
Tetrahymena_thermophila__TTHERM_00245190__Syb
0.5

Figure 5.11: Phylogenetic analysis of pan-eukaryotic R-SNAREs suggests that Synaptobrevin-like SNAREs are related to the regulatory R-SNARE domain-containing protein Tomosyn. (see Appendix A chapter 5 Alignment 7). Node support values are from MrBayes (posterior probability, maximum 1.0 with 0.8 generally being considered the minimum significant value) and IQ-TREE (nonparametric bootstrap values, maximum 100 with 50 being the minimum significant value). The MrBayes topology is shown, arbitrarily rooted on the clade containing orthologues of Sec22 and Ykt6. The amino acid substitution model used for this analysis is LG+G4. Significant prior probability support was found for the placement of Tomosyn as sister to synaptobrevin.

5.3.4 Distibution of synaptobrevin orthologues across the eukaryotic tree

To further explore the extent of conservation of synaptobrevin orthologues among eukaryotes we attempted to specifically identify synaptobrevin orthologues in predicted peptide sequences for all genomes sampled using similarity searching methods. However, synaptobrevin orthologues are in many cases not readily classifiable by pairwise or HMM-based similarity searching methods. For example, many of the sequences classified as Synaptobrevin orthologues in Figure 5.11 do not retrieve either *Homo sapiens* or *Saccharomyces cerevisiae* synaptobrevin orthologues as top BLASTP hits among predicted proteins for those organisms: *Monocercomonoides exilis* MONOS_5707, *Cyanoptyche gloeocystis* CAMPEP_0196652376, *Cafeteria roenbergensis* CAMPEP_0117710538, *Chondrus crispus* XP_005717971.1, and *Goniomonas avonlea* comp4141 9_c0_seq1_6_ORF10_240 (Table S2). These inconsistent results may be due to the presence of longin domains in non-amorphean synaptobrevin sequences or overall sequence divergence. Next, we attempted to classify synaptobrevin orthologues using hidden markov models (HMMs). Nine out of the nineteen Syb sequences from the highly representative dataset for the analysis in Figure 5.10 were classified as VAMP7 instead of synaptobrevin by the online HMM-based SNARE domain classification tool on the SNAREDB website provided by Kloepper *et al.* (2007). These include *Micromonas pusilla* 122911, *Thalassiosira pseudonana* 263742, *Pavlovales* sp. 934490, and *Trimastix marina* 4462. We were also unable to produce HMMs that distinguished reliably between Syb and VAMP7 orthologues among diverse eukaryotes (not shown). Thus, although HMM-based classification may be accurate in the case of most SNARE sequences, distinguishing between orthologues of synaptobrevin and VAMP7 appears to be problematic. It may be that synaptobrevins in Metazoa and Fungi are difficult to classify as orthologues of the other non-VAMP7 R-SNAREs, because synaptobrevins have diverged in these lineages. Such divergence might be due in part to changing selective pressures associated with the loss of NPSN and Syp7.

While similarity searching is largely inconclusive, the phylogenetic analyses performed here identify synaptobrevin orthologues in a diverse sampling of taxa in which they have never been suggested to exist including Holomycota (*e.g.*, *Fonticula alba*), stramenopiles (*e.g.*, *Thalassiosira pseudonana*), metamonads (*e.g.*, *Monocercomonoides exilis*), chromerids (*e.g.*, *Chromera velia*), red algae (*e.g.*, *Chondrus crispus*), green algae (*e.g.*, *Micromonas pusilla*), glaucophytes (*e.g.*, *Cyanoptyche gloeocystis*), cryptophytes (*e.g.*, *Goniomonas avonlea*), haptophytes (*e.g.*, *Pavlovales* sp.), and hemimastigophorids (*e.g.*, *Spironema* sp.) (Figures 5.9 and 5.11). These findings confirm that synaptobrevin orthologues originated prior to the LECA.

We did not identify synaptobrevin orthologues in all genomes or transcriptomes sampled. How-

ever, it is difficult to confirm losses of synaptobrevin orthologues, as they may simply be difficult to classify as in the case of metazoan and fungal orthologues. However, potential independent losses may have occurred conspicuously in ancestors of Embryophyta and of Discoba. Interestingly, inparalogues of VAMP7 arising in the early evolution of embryophytes and in kinetoplastids, a subgroup of discobids, have been identified previously (Sanderfoot, 2007; Murungi *et al.*, 2014). It may be that these inparalogues accomodate for the loss of synaptobrevin orthologues.

5.3.5 Unrooted SNARE family phylogenies favour a limited set of scenarios for organelle evolution

The OPH posits essentially that the evolutionary history of endomembrane organelles in the membrane trafficking system may be inferred from phylogenies of membrane trafficking protein families with organelle-specific family members (Dacks and Field, 2007; Dacks *et al.*, 2008). As discussed above, each of the four SNARE families are relevant for making such inferences, due to the organelle specificity of their members. In this section we consider whether the unrooted SNARE family phylogenies are consistent with basic predictions of the OPH, and find that they are overall. Therefore, we consider in detail what specific scenarios might be distinguishable based on evidence from SNARE phylogenetics (including both unrooted and rooted phylogenies) and then how these scenarios may relate to scenarios previously proposed in the literature.

A preliminary prediction of the OPH is that the topologies of relevant membrane trafficking protein family phylogenies will be essentially topologically similar with respect to the organelle specificities of the family members. The four unrooted pan-eukaryotic SNARE family trees, now including R-SNAREs, are comparable in that each contains clades of SNAREs specific to either the ER, Golgi, endosomes, or PM. While a phylogenetic tree with two or three clades of interest can only yield one topology of the branches connecting these clades of interest, a tree with four clades may yield one of three alternative topologies. Therefore, a SNARE family tree with clades representing four or more family members may yield three or more distinct topologies that describe the deep evolutionary relationships between the family members, suggesting the possibility of conflict between the SNARE family topologies. When compared in this way, however, the four SNARE family phylogenies presented in the preceding sections are consistent with respect to the organelle-specificities of the family members (see Figure 5.12), which is consistent with a preliminary prediction of the OPH with respect to protein families involved in the origin of novel organelles 1.5. In each case the (putative) PM and endosome SNAREs (Qa: SynPM and SynE; Qb: NPSN, SNAPN, and VTI1; Qc: Syp7, SNAPC, Syn8, and Syn6; R: Tomosyn, Synaptobrevin,

and VAMP7) branch together to the exclusion of ER and Golgi SNAREs.

Moreover, in the case of the R and Qb phylogenies, SNAREs functioning at each of the four membranes form monophyletic clades, with clades representing the Golgi SNAREs Membrin and Gos28 being sister clades, and the synaptobrevin and tomosyn clades branching sister to each other as well. In the case of the Qa and Qc families, however, clades containing orthologues of eukaryotic outparalogues that function at endosomes do not form monophyletic clades. Instead, clades representing the Qa SNAREs Syn16 and SynE are paraphyletic, and the same is true for the Qc SNAREs Syn6 and Syn8. Furthermore, the paraphyletic branching orders of these endosomal Qa and Qc SNAREs are inconsistent with respect to their specificities to sub-populations of endosomes. That is, the relative placements of the early endosomal SNAREs (Syn16, and Syn6) and late endosomal SNAREs (SynE and Syn8) are reversed in their respective family trees. These inconsistencies may challenge interpretation of SNARE evolution in terms of the OPH. Nevertheless, considering the overall similarities between the SNARE family topologies, and that these inconsistencies were not found in the Qb and R SNARE families, we determined to continue further investigation.

While these results so far are mostly consistent with potential intepretation of SNARE evolution via OPH-type scenarios, these SNARE family trees are unrooted (arbitrarily rooted in the figures shown above: Figures 5.3, 5.4, 5.10) as they have no outgroup. This means that when a set of outgroup sequences is added to the tree the branch leading to the resulting additional clade could be placed at any of the five deep branches in the unrooted SNARE family tree. Moreover, if the different SNARE families have different placements for their roots, then they would be topologically different, making interpretation in terms of the OPH difficult if not impossible. Therefore, placing the root of each SNARE family tree is critical for drawing inferences to the evolution of membrane trafficking organelles.

Given the split separating ER and Golgi SNAREs from PM and endosomal SNAREs in each of the unrooted SNARE family trees, five basic possibilities for the rooted tree of any SNARE family. Figure (5.13) illustrates how these five alternative rootings might relate to different scenarios of organelle evolution. In generating the scenarios shown in Figure 5.13, a minimal set of assumptions were made, which are necessary to make meaningful inferences regarding early endomembrane evolution from SNARE family phylogenies. These are implied by the OPH as it was originally conceived (Dacks and Field, 2007; Dacks *et al.*, 2008), but are considered explicitly here. Each of these assumptions is essentially based on the principle of parsimony. To our knowledge, with the partial exception of the fourth assumption as discussed below, there is currently no evidence that conclusively invalidates these assumptions. Violations of these assumptions would render any

alternative SNARE tree topology consistent with all alternative scenarios for early endomembrane evolution, or at least a much larger proportion of alternative scenarios.

These minimal assumptions are as follows:

1. No types of membrane fusion events among vesicles or organelles in the membrane trafficking system (such as fusion of an endomembrane with the plasma membrane) occurred by a non-SNARE-mediated process after the origin of the first fusogenic SNARE complex. That is, SNARE complexes did not diversify only to supplant an already complex set of fusion machineries.

2. The function or functions of the first fusogenic SNARE complex included membrane fusion at the plasma membrane.

3. In cases where SNARE complexes differentiate from a PM-localized SNARE complex, the ancestral SNARE complex functioned at the PM rather than at an endomembrane organelle. This is why, for example, in Figure 5.13C, D, and E a second endomembrane organelle does not appear until the second of the three stages shown.

4. Following the first differentiation of SNARE complexes, the evolutionary histories of both SNARE-encoding genes and the organelles at which they act can be accurately described by topologically identical strictly bifurcating trees.

5. Complete SNARE complex differentiation is required for differentiation of membranes/organelles. That is, to be considered a differentiated and distinct from other organelles/membranes an organelle/membrane must have a dedicated complete set of four SNARE proteins that are specific to it.

We note, however, that assumption 4 cannot be strictly valid in fact, because the four SNARE families each contain more than four eukaryotic out-paralogues, and only four types of organelles are considered (including the PM). Also, the inconsistent, paraphyletic branching of endosomal SNAREs in the Qa and Qc SNARE families, described above (section 5.3.5), implies further violation of this assumption. Nevertheless, we argue that resulting discrepancies between the phylogeny of SNAREs and the phylogeny of the organelles considered may be explained by exceptional cases of neofunctionalization (*e.g.*, in the case of Tomosyn) or subfunctionalization among SNAREs which function at different sub-structures or sub-populations of organelles (*e.g.*, in the case of the Golgi-localized Qb-SNAREs Membrin and Gos28).

Another possible exception to assumption 4 would be co-option of SNARE paralogues to the outer membrane of endosymbiotic organelles for fusion of vesicles with these organelles or fusion of endosymbiont-derived vesicles with membranes of the endomembrane system. Perhaps the best characterized example of this is the localization of the Qa-SNARE Syntaxin 17 (Syn17) to the mitochondrial outer membrane, concentration in mitochondria-derived vesicles (MDVs), and role in fusion of these vesicles with late endosomes/lysosomes in mammalian cells (McLelland *et al.*, 2016). While MDVs perform important functions (Sugiura *et al.*, 2014) and an orthologue of Syn17 was present in the LECA (Arasaki *et al.*, 2015), Syn17 performs a diversity of functions in mammalian cells including autophagosome-lysosome fusion (Itakura *et al.*, 2012) and regulation of mitochondrial fission (Arasaki *et al.*, 2015), and neither the localization nor function of Syn17 orthologues have been characterized in other eukaryotes. Thus it is currently unclear whether the role in MDV fusion was acquired during the early evolution of the eukaryotic endomembrane system. Another potential example of post-LECA co-option of a SNARE to an endosymbiotic organelle is an *Oryza sativa* (rice) VAMP7 orthologue R-SNARE VAMP714, which has been suggested to partially localize to chloroplasts (Sugano *et al.*, 2016).

Assumption 4 listed above also excludes two possibilities for organelle origin of which there are no confirmed instances to our knowledge. First is organelle mergence. This type of event becomes potentially relevant when considering the evolution of eukaryotes possessing membrane trafficking systems with apparently secondarily reduced complexity. However, in the absence of evidence to the contrary, it is more parsimonious to assume that FECA to LECA evolution involved only increases in membrane trafficking complexity, which is a corollary to assumption 4 above. Second is the origin of organelles as intermediates between two progenitor organelles, with the resulting novel organelle being equally closely related to the two original organelles. Such an event would perhaps be biologically plausible as an elaboration on a process of maturation from one organelle type to another, but phylogenetic relationships between the resulting organelles would be best represented by a multifurcating tree topology, or perhaps a network.

All the possible scenarios consistent with the above assumptions and with the supported unrooted SNARE family topologies observed are shown in Figure 5.13. This, again, excludes two thirds of all possible topologies for the relationships between the endomembranes in question. For example, because there is no Golgi plus endosomal SNARE clade in the unrooted topologies, none of the scenarios shown in Figure 5.13 involve differentiation of endosomes and Golgi from an ancestral endosomal/Golgi organelle. Thus, the unrooted topologies alone already refute a considerable number (10) of distinct scenarios that would be consistent with the above assumptions.

While the minimal assumptions listed above are necessary to draw meaningful inferences regarding the evolutionary relationships between organelles from SNARE protein phylogenies, they are not sufficient to determine whether any of the scenarios in Figure 5.13 favours or refutes various previously proposed scenarios for the early evolution of the endomembane system and the overall process of eukaryogenesis which were mentioned in the introduction (see section 1.4 and discussion below). No assumptions are made regarding the function or origin of the first internal membranes to be fused in a SNARE-dependent manner to the PM, or of any evolutionary intermediate internal organelles. This means that the scenarios proposed cannot be strictly consistent or inconsistent with the alternative hypotheses regarding the origin of the first endomembrane organelle, even hypotheses of endosymbiotic origins of such an organelle. In particular, the first and fourth assumptions do not to exclude such possibilities *a priori*. An important corollary to this is that the scenarios described above do not necessarily allow inference of the function of various progenitor organelles by assigning them functions equivalent to the extant organelles that may have been derived from them. For example, the "ER/Golgi/Endosome" in Figure 5.13A may have performed a function similar to the ER or to endosomes, or perhaps some combination.

However, additional considerations regarding the number or magnitude of evolutionary steps required for alternative hypotheses to accommodate the evolutionary relationships between organelles shown in Figure 5.13 may provide insight into these issues. It is more parsimonious to assume lower rates of evolution of organelle functions. For example, shifting of protein machinery for membrane protein synthesis and insertion (which acts at the ER in extant cells), machinery for post-translational protein glycosylation (which acts at the Golgi in extant cells), or machinery for intralumenal acidification and protein degradation (which acts at late endosomes and lysosomes in extant cells) from one already differentiated organelle to another would be major events which could be expected to occur very few times if any during the course of endomembrane evolution. For this reason, it is simplest to assume that intermediate organelles that pre-dated differentiation events performed at least some of the functions performed by the resulting differentiated organelles.

With the above considerations in mind, comparisons between the alternative scenarios shown in Figure 5.13 and previously proposed scenarios for the early evolution of eukaryotes and of their endomembrane system may be considered. One of the major differences between prokaryotes and eukaryotes, beyond the prominence of endomembranes is that translocons and associated ribosomes are not localized to the plasma membrane (cytosolic membrane in the case of gram-negative bacteria) but on internal membranes (the ER in extant eukaryotes). Most previously proposed scenarios suggest that the first endomembrane organelle possessed translocons (and associated ri-

bosomes) on the cytosolic face of its membrane (de Duve and Wattiaux, 1966; de Duve, 1969; Blobel, 1980; Jékely, 2007; Cavalier-Smith, 2009; Gould *et al.*, 2016). Even in the case of the scenario proposed by Gould *et al.* (2016), that the first endomembrane organelle was formed from mitochondrial-derived vesicles, localization of translocons to an internal organelle is one of the earliest steps. This would be most consistent with scenarios A and B (and potentially E) (Figure 5.13A, B, and E), as these scenarios involve the ER of extant eukaryotes being derived from the initial endomembrane organelle, and thus not requiring relocation from the plasma membrane after diversification of endomembranes.

It was noted first by Blobel (1980) that, assuming the initial endomembrane organelle was ER-like, endosomes could be subsequently derived from either this first organelle or independently from the PM. These alternatives are represented by the scenarios in Figure 5.13 A and B. Phagotrophic origins of eukaryotic endomembranes were proposed by de Duve (1969) and Cavalier-Smith (2009). Cavalier-Smith (2009) further posited a phagocytic vacuole with translocons as the initial organelle followed by differentiation into a dedicated phagosome/endosome and a nuclear envelope/ER-like secretory organelle, which is most consistent with the scenario in Figure 5.13A and not B.

Hypotheses of early endomembrane evolution most consistent with remaining scenarios, C, D, and E, depicted in Figure 5.13 have received little if any attention, and are perhaps less intuitively appealing. Moreover, the Golgi-first scenario may be biologically implausible or unparsimonious considering the flow of the secretory pathway from the ER through the Golgi to endosomes and the PM. However, all of these scenarios are viable alternatives in the sense that they may be supported or refuted by rooted SNARE family phylogenies given the assumptions described above. Thus SNAREs offer a means to corroborate aspects of the most prominent scenarios for early endomembrane evolution by evaluating them against alternatives.

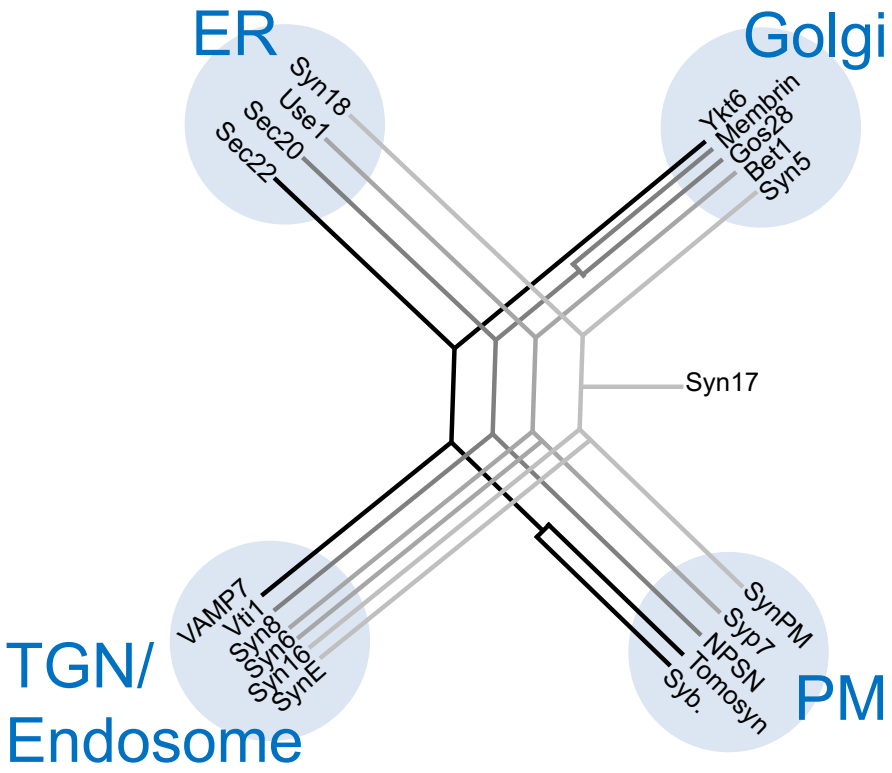


Figure 5.12: The SNARE trees are all consistent with the same unrooted topology of the tree with respect to the organelles at which the proteins act. The topologies of the SNARE family trees overlaid here are based on those from phylogenetic analyses of individual SNARE families (see Figure 5.10 and Appendix A chapter 5 Figures S3-S5), and the hypothetical ancestral functions discussed in the introduction (5.2.2).

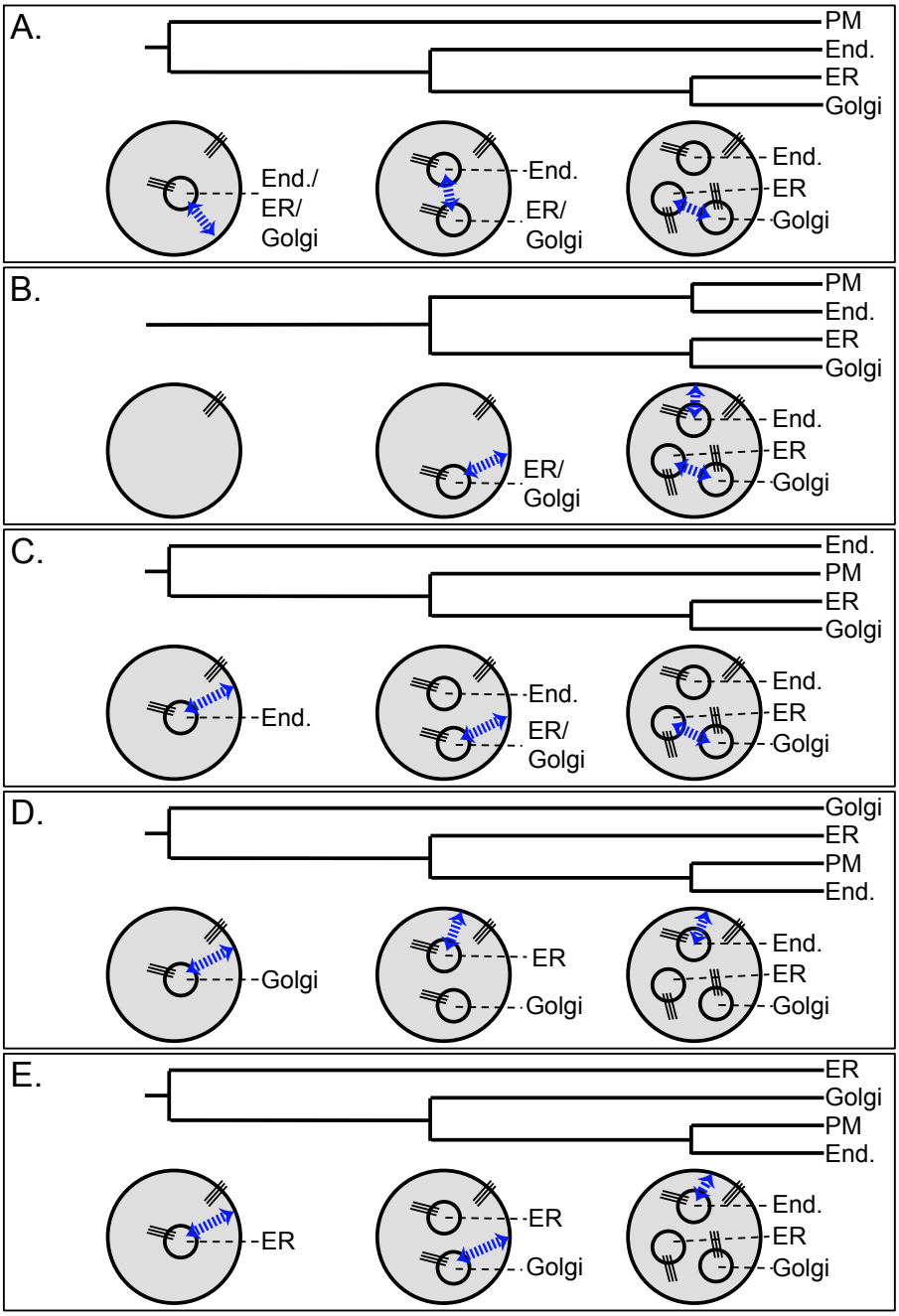

A.
PM
End.
ER
Golgi
End./
ER/
Golgi
End.
ER/
Golgi
End.
ER
Golgi
B.
PM
End.
ER
Golgi
ER/
Golgi
End.
ER
Golgi
C.
End.
PM
ER
Golgi
End.
End.
ER/
Golgi
End.
ER
Golgi
D.
Golgi
ER
PM
End.
Golgi
ER
Golgi
End.
ER
Golgi
E.
ER
Golgi
PM
End.
ER
ER
Golgi
End.
ER
Golgi

Figure 5.13: Scenarios of early eukaryotic endomembrane organelle evolution implied by alternative rootings of the SNARE family phylogenies. The top portion of each panel shows one alternative rooting of the SNARE family phylogenies with respect to the organelle specificities of the SNAREs represented by each clade (endosome is abbreviated as "End."). The bottom portion of each panel shows three consecutive intermediate stages of endomembrane organelle evolution. The first stage shown is the first differentiation of an original ancestral SNARE complex (except for panel B, because two subsequent duplications may have occurred simultaneously in this case). The second and third stages represent events occurring with subsequent SNARE complex duplications. Large circles represent the PM, while small circles represent endomembrane organelles. Distinct SNARE complexes are indicated by the groups of short parallel lines. Dashed double-ended blue arrows indicate endomembrane/SNARE complex differentiation events. See main text for discussion of the assumptions upon which these inferences are based. **A**. One primordial organelle differentiates to produce entire endomembrane system. **B**. Endosomes derived independently from the PM after the ER/Golgi. **C**. Endosomes derived independently from the PM before the ER/Golgi. **D**. Golgi derived first and independently from the PM. **E**. ER first, then Golgi, then endosomal SNAREs derived from plasma membrane SNAREs.

5.3.6 Placing the root of each SNARE family phylogeny

The unrooted phylogenies of SNARE families do not indicate the relative order/timing of gene duplications giving rise to eukaryotic outparalogues, because rooting on different clades would imply conflicting scenarios (Figure 5.13). Rooting of phylogenetic trees with confidence requires an appropriate outgroup of homologous sequences. Each of the SNARE families could be rooted on any combination of sequences of SNAREs from any one or more of the remaining three SNARE families, due to their homology and uncertain relationships between SNARE families. However, ideally the rooting of each SNARE family should be evident in the topology of a SNARE superfamily phylogeny. Previous attempts to determine the topology of the SNARE superfamily, including rootings for each family, have yielded little evidence for the rooting of any SNARE family. Kloepper *et al.* (2007) suggest that secretory (PM) SNAREs represent the earliest branches in each SNARE family. However, their phylogenetic analysis of Qa, Qb, Qc, and R SNARE sequences from 11 eukaryotes do not support this hypothesis (see Kloepper *et al.* (2007) supplementary information). The more extensive sequence data sampled in this study, and the advances in phylogenetics of Qbc SNAREs and synaptobrevins detailed above suggest potential for placing the root of SNARE families with more confidence.

To investigate the rooting of each SNARE family tree, we analyzed the SNARE superfamily phylogeny. To construct a SNARE superfamily sequence alignment, we assembled alignments of representative sequences that yielded relatively well-resolved unrooted phylogenetic trees for each SNARE family. This included Qa, Qb, and Qc SNARE alignments which yielded similar phylogenetic results to previous work, but with updated taxon sampling (see Appendix A chapter 5 Figures S3-S5, and Alignments 8-10) (Arasaki *et al.*, 2015; Venkatesh *et al.*, 2017). We excluded the non-canonical SNAREs from each family: Tomosyn, Syn17, and Qbc SNARE domains SNAPN and SNAPC. It is most parsimonious to assume that these originated relatively late in SNARE evolution, and these are divergent so not as informative. Altogether 364 SNARE sequences were included in the full superfamily alignment (76 Qa, 108 Qb, 97 Qc, and 83 R SNAREs). The alignment was trimmed to include only positions representing the SNARE domains, using a reference alignment as a guide (Fasshauer *et al.*, 1998). Homology of positions outside this region was not obvious. ML analysis was performed (with 100 non-parametric bootstraps) using IQ-TREE (Nguyen *et al.*, 2015). Branches within clades of orthologous pan-eukaryotic SNARE family members recovered previously using more inclusive trimming (see Figure 5.10 and Appendix A chapter 5 Figures S3-S5) were constrained to reduce the tree space and also prevent introduction of long-branch attraction artefacts. The resulting SNARE superfamily phylogeny indicates rooting

of the R, Qb, and Qc families on the respective clades representing PM SNAREs (synaptobrevin, NPSN, and Syp7) (Figure 5.14). In contrast, the Qa family is rooted on the branch separating ER and Golgi SNAREs from PM and endosomal SNAREs, but bootstrap support for this rooting is marginal.

While the topology shown in Figure (5.14) favours rooting of at least three out of the four SNARE families on PM SNAREs, support for rooting of the Qa-SNARE family is insignificant, suggesting that this family may be relatively less informative for this analysis, and/or that the Qa-SNARE family has a different evolutionary history than the other SNARE families. Thus we performed a similar analysis excluding Qa SNAREs and updating the best-fit amino acid substitution model. The resulting consensus of bootstrap topologies shows the same topology of the Qb, Qc, and R families, rooted on their respective PM family members (Figure 5.15). In this analysis the bootstrap percentage for rooting of the Qc family, 40%, is insignificant, and is lower than 58% found in the previous analysis (Figure 5.14). This discrepancy may indicate sensitivity to model selection and/or conflicting phylogenetic signals present in the selected Syp7 sequences. However, support for the Qb and R family rootings is relatively strong. The relatively higher bootstrap support for the rooting of the Qb and R SNARE families in both Figures 5.14 and 5.15 indicate that these rootings are more likely to represent the true rooting than the less well-supported rootings for the Qa and Qc families.

The somewhat ambiguous results of these phylogenetic analyses prompted us to explicitly test *a priori* rooting hypotheses using a topology testing approach. As discussed in the previous section (5.3.5), five alternative rootings of each SNARE family are possible, and hypotheses involving consistent rootings of all four SNARE families (with respect to the organelle-specificities of SNAREs on either side of the root) are of particular interest. For topology testing, we used the same sequence alignments as for the analyses shown in Figures 5.14 and 5.15. We constructed tree topologies for constraining ML searches to find the ML trees representing each of the hypotheses shown in Figure 5.13 (for all four SNARE families simultaneously), as well as the ML tree without rooting constraints. Importantly, as noted in the previous section, the Qa and Qc SNARE phylogenies include paraphyletic endosomal SNARE clades, which implies that there are more possible rootings for these families. However, for simplicity, in the constraint trees for endosomal rooting hypotheses we made these paraphyletic clades monophyletic in contrast to the topologies found in previous analyses. Again, in each constraint tree branches within clades of orthologous pan-eukaryotic SNARE family members recovered previously using more inclusive trimming (see Figure 5.10 and Appendix A chapter 5 Figures S3-S5) were constrained to reduce the tree space

and also prevent long-branch attraction.

In the topology test on the alignment including all four SNARE families (Table 5.2 Test 1), the ML tree topology without rooting constraints was the same as shown in Figure 5.14 (Qa family rooted on branch separating ER and Golgi SNAREs from PM and endosomal SNAREs), but the tree with the highest log likelihood was the ML tree obtained with the ER root constraints. The difference in log likelihood between these two trees, however, was small. The AU test P-values for Test 1 suggest that the hypotheses involving rooting all the SNARE families on endosomal SNAREs or all on Golgi SNAREs may be rejected, as less likely than the ER rooting hypothesis. Also, the ELW confidence values indicate 96% confidence that the true topology is represented by the ER or PM rooting hypotheses or the ML tree with no rooting constraints, which indicates rejection of the middle, Golgi, and endosome rooting hypotheses by the ELW analysis.

Somewhat different results were observed in the topology test on the alignment without Qa-SNAREs (Table 5.2 Test 1). In this case, the ML tree topology without rooting constraints differed from the consensus of bootstrap topologies shown in Figure 5.15, as the Qc family was rooted on the clade containing orthologues of the ER SNARE Use1. The AU test P-values for Test 2 suggest that the hypotheses involving rooting all the SNARE families on endosomal SNAREs, all on ER SNAREs, or all on Golgi SNAREs may be rejected as less likely than hypothesis represented by the rootings in the ML topology without rooting constraints. The ELW confidence value in Test 2 for the middle rooting hypothesis is only 0.0378, again suggesting rejection of this hypothesis by the ELW method but not the AU test.

Taken together, the results of these SNARE superfamily analyses suggest that the Qa-SNARE family may have a different rooting compared to the other families, and placement of the roots may be sensitive to outgroup selection and/or model selection. To further investigate these possibilities, we analyzed alignments composed of all combinations of two SNARE families. For this analysis, we subdivided the SNARE superfamily alignment used for the analyses above into separate alignments for the six SNARE family pairings. We analyzed these alignments using both bayesian and maximum likelihood approaches. Also, we constrained the monophyly of previously recovered clades representing orthologues of eukaryotic outparalogues (though not the internal branches of these clades). The phylogenetic tree for the analysis of Qa-SNAREs with Qb-SNAREs is shown in Figure 5.16, and the results of all of these phylogenetic analyses are summarized in Table 5.3 (also see Appendix A chapter 5 Figures S6-S11).

The results of these pairwise rooting analyses do not refute the root placements suggested by the previous analyses for the Qb, Qc, or R families. Significant bootstrap support is found for

rooting of Qb-SNAREs on NPSN and of Qc-SNAREs on Syp7, and alternative rootings for the Qc-SNAREs and R-SNAREs were not supported (Table 5.3). However, in this analysis significant prior probability and bootstrap percentage support was found for rooting of Qa-SNAREs on the ER Qa-SNARE Syn18 (Figure 5.16 and Table 5.3). This contrasts with the rooting on the clade containing both Syn18 and Syn5 in Figure 5.14. However, rooting of Qa-SNAREs on Syn18 is consistent with the results of the topology test for the alignment shown in Figure 5.14, Test 1 in Table 5.2, in which the hypothesis involving rooting all the SNARE families on ER-specific family members was not rejected.

There are multiple ways in which these analyses might be further optimized for resolution of SNARE family rootings. A future direction for this work would be to apply the AU test without RELL approximation of boostraps, as this approximation may cause increased sensitivity of the results to low amounts of data and model misspecification(Hasegawa and Kishino, 1994; Goldman *et al.*, 2000). This may be done using the CONSEL program(Shimodaira and Hasegawa, 2001). Another approach may be selection and use of amino acid substitution models which may more accurately model the evolution of SNARE domains. The differences in ML topologies and bootstrap support for the Qc family in analyses with different models may suggest sensitivity to model selection. Development of more accurate amino acid substitution models is an active area of research, with new models continuing to be published (Braun, 2018; Chi *et al.*, 2018). SNARE domains are a type of coiled-coil-forming domain, and coiled-coils have been found to evolve differently than other protein domains due to unique evolutionary constraints on structure, suggesting the need for non-standard models (Surkont and Pereira-Leal, 2015). In addition, there is potential for use of substitution models optimized for rooting phylogenetic trees (Williams *et al.*, 2015).

Sequence selection is another aspect to this analysis that may be optimized further potentially allowing more confident phylogenetic inferences. A minority of the sequences in the analyses shown in both Figure 5.14 and 5.15 had exceptionally long branches. This is may be due to divergence within the SNARE domain sequences that was not evident in the single-family SNARE analyses. The existence of these long branches in the trees may make likelihood estimations less accurate. Also, we attempted to maintain a high-level of pan-eukaryotic taxonomic representation in the dataset, but this may be less important for examining relationships between eukaryotic outparalogues than for establishing pan-eukaryotic orthology, which was an important goal of the single-family SNARE analyses (Figures 5.3, 5.4, and 5.10). Accordingly, since all established eukaryotic out-paralogues are represented among Amorphea and Diaphoretickes, and considering that many sequences from Discoba and Metamonada appear to be divergent, a reduced dataset with

only sequences from Amorphea and Diaphoretickes may contain higher phylogenetic signal and thus yield more informative results.

Taken together, the results of the phylogenetic analyses (Figures 5.14, 5.15, and 5.3) and the topology tests (Table 5.2) favour the hypothesis that divergence of PM SNAREs from SNAREs that eventually differentiated into endosomal, Golgi, and ER SNAREs was the first major event in the evolution of the SNARE superfamily after the origin of the first functional SNARE complex (as shown in Figure 5.13A). A consensus of the supported SNARE family rootings in these analyses indicates the PM rooting, as the Qb, Qc, and R SNAREs are consistent and only the Qa SNARE family suggests a rooting on the ER clade. Other possibilities are not supported so far. Considering the topology test results, the alternative hypothesis that each SNARE family is rooted on the "middle" branch (Figure 5.13B) is not rejected here. Hypotheses involving contrasting rootings of the Qa and/or Qc families are also not rejected. In addition, we found no support for rooting of any of the SNARE families on Golgi or endosomal SNAREs, and both of these hypotheses were rejected by both topology tests, so at the very least the hypotheses represented in Figure 5.13C and D are rejected here.

The rooting of the Qa SNAREs on the branch separating endosomal and PM SNAREs from ER and Golgi SNAREs in the tree without rooting constraints contrasts with the rooting of the other SNARE families on PM-localized SNAREs (Figure 5.14). The hypotheses involving rooting of all SNARE families on either the PM representatives or on the "middle" branches separating PM and endosomal SNAREs from ER and Golgi SNAREs would be simpler to interpret (Figure 5.13 A and B), and were not rejected by the topology tests. However, a coherent interpretation of implications for endomembrane evolution may be possible, even assuming the rootings shown in Figure 5.14 are correct. The overall scenario may be as shown in Figure 5.13A, as suggested by the supported rootings of the Qb, Qc, and R SNARE families, and the contrasting rooting of the Qa SNARE family may be explained by a single Qa-SNARE functioning at both the PM and an endomembrane organelle until after the first differentiation of endomembrane organelle. This would perhaps not be surprising considering several examples of homologous membrane trafficking protein complexes with different functions but partially overlapping sets of subunits. For example, in most eukaryotes the β subunits of Adaptor Protein (AP) complexes AP-1, which functions at endosomes, and AP-2, which functions at the PM (Markus Boehm and Juan S. Bonifacino, 2001; Dacks *et al.*, 2008; Teh *et al.*, 2013; Yamaoka *et al.*, 2013). However, this would perhaps require relaxing the fifth assumption listed in section 5.3.5.

Another challenge in interpreting these results, which was noted in the preceding section as

well, is that in the case of the Qa and Qc families, clades containing orthologues of eukaryotic outparalogues that function at endosomes do not form monophyletic clades (Figure 5.12). Considering that the analyses performed here seem to suggest overall that PM rootings are likely, it may be interesting to consider what the paraphyletic branching of endosomal SNAREs in the Qa and Qc families might indicate assuming this is the case. One possibility is that late endosomes and lysosomes began to differentiate from an initial endomembrane organelle prior to the differentiation of an ER/Golgi organelle. However, further work is needed to investigate this possibility.

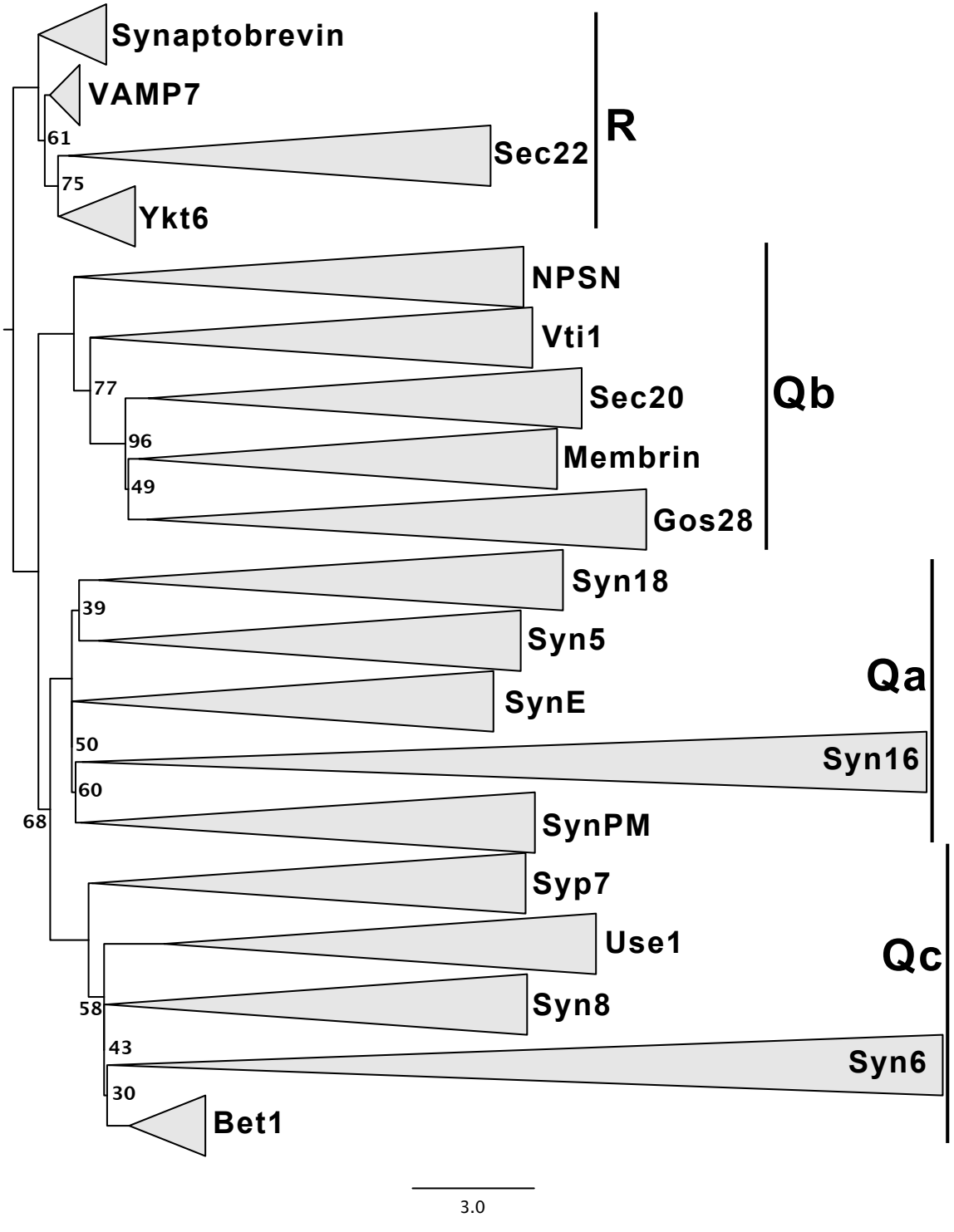

Synaptobrevin
VAMP7
61
Sec22
R
75
Ykt6
NPSN
Vti1
77
Sec20
Qb
96
Membrin
49
Gos28
Syn18
39
Syn5
Qa
SynE
50
Syn16
60
68
SynPM
Syp7
Use1
Qc
58
Syn8
43
Syn6
30
Bet1
3.0

Figure 5.14: Phylogenetic analysis of the SNARE superfamily suggests all but the Qa-SNAREs are rooted on clades representing plasma membrane-localized family members. Alignments of representative sequences from the Qa, Qb, Qc, and R SNARE families were assembled into a SNARE superfamily alignment (see Appendix A chapter 5 Alignment 11). The non-canonical SNAREs Syn17, Tomosyn, and Qbc SNAREs were excluded. Positions in the alignment representing amino acid residues outside of the homologous SNARE domains were trimmed out of the alignment. Using this alignment, IQ-TREE was used to perform ML searches on 100 bootstrap pseudoreplicates, and the consensus of bootstrap tree topologies is shown. Branches within clades representing eukaryotic outparalogues identified in previous family-specific analyses (see Figure 5.10 and Appendix A chapter 5 Figures S3-S5) were constrained to reduce the tree search space to informative topologies. In the resulting topology the subtrees representing the R, Qb, and Qc SNARE families are rooted on the plasma membrane-localized family member: Synaptobrevin, NPSN, and Syp7, respectively. In contrast, the Qa subtree is rooted on the branch separating the ER and Golgi Qa-SNARE clades (Syn18 and Syn5) from the PM and endosomal Qa-SNARE clades (SynPM, SynE, Syn16). Support values for unconstrained branches are shown as bootstrap percentages. The bootstrap support for the rooting of the Qa subtree is marginal (50 is generally considered the minimum threshold for signficance). See Table 5.2 Test 1 for testing of alternative rooting hypotheses.

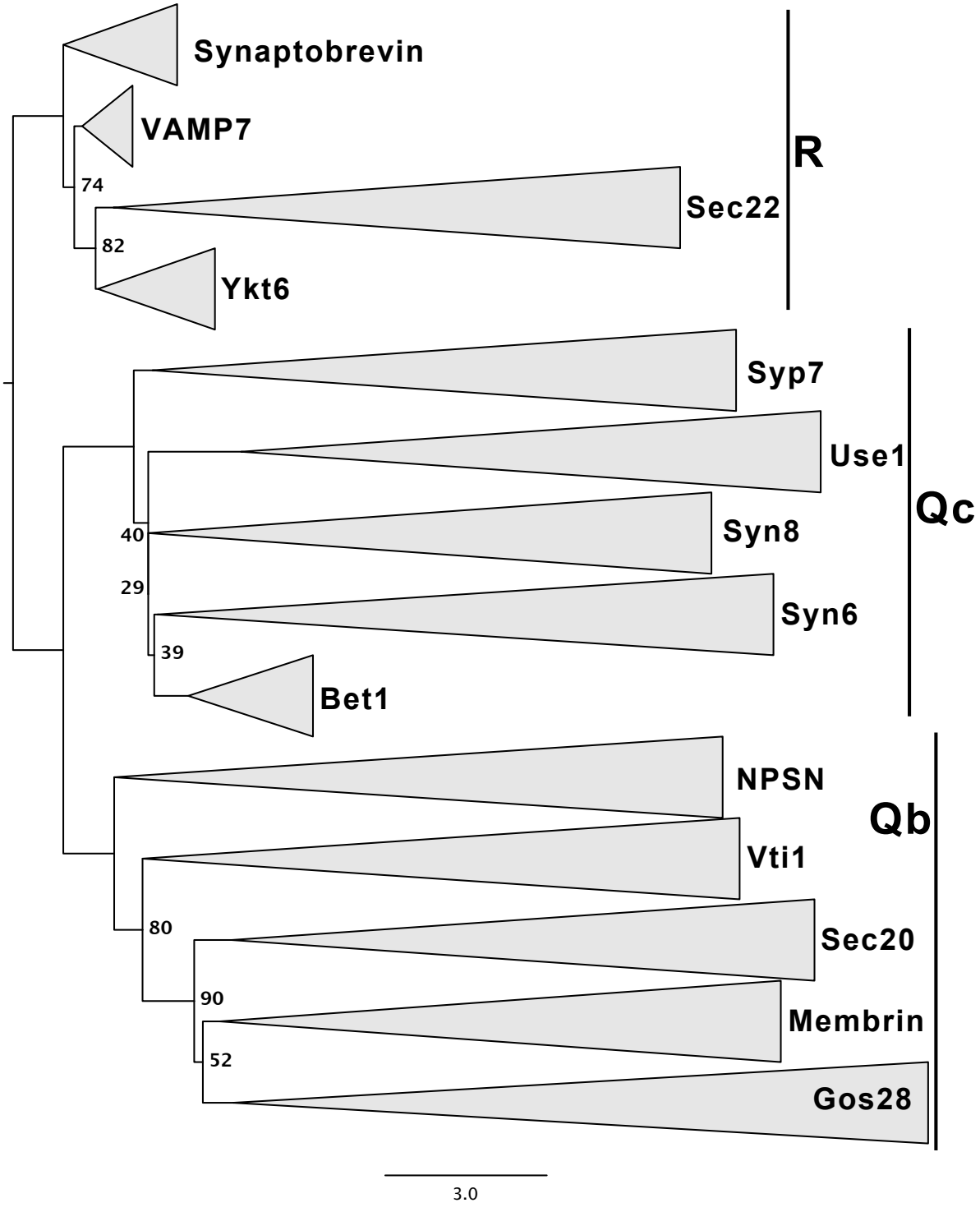

Synaptobrevin
VAMP7
74
Sec22
82
Ykt6
R
Syp7
Use1
40
Syn8
29
Syn6
39
Bet1
Qc
NPSN
Vti1
80
Sec20
90
Membrin
52
Gos28
Qb
3.0

Figure 5.15: Phylogenetic analysis of the R, Qb, and Qc SNARE families suggests rooted on clades representing plasma membrane-localized family members. As for Figure 5.14, alignments of representative sequences from the Qb, Qc, and R SNARE families were assembled into a SNARE superfamily alignment. The non-canonical SNAREs Syn17, Tomosyn, and Qbc SNAREs were excluded. Positions in the alignment representing amino acid residues outside of the homologous SNARE domains were trimmed out of the alignment. Using this alignment, IQ-tree was used to perform ML searches on 100 bootstrap pseudoreplicates, and the consensus of bootstrap tree topologies is shown. Branches within clades representing eukaryotic outparalogues identified in previous family-specific analyses (Figure 5.10 and Appendix A chapter 5 Figures S3-S5) were constrained to reduce the tree search space to informative topologies. In the resulting topology the subtrees representing the R, Qb, and Qc SNARE families are rooted on the plasma membrane-localized family member: Synaptobrevin, NPSN, and Syp7, respectively. Support values for unconstrained branches are shown as bootstrap percentages. The bootstrap support of 40% for the rooting of the Qc subtree on the Syp7 clade is insignificant (50 is generally considered the minimum threshold for significance). See Table 5.2 Test 2 for testing of alternative rooting hypotheses for this dataset.

Table 5.2: Topology tests of alternative root placements for SNARE family phylogenies. The same alignments as for Figures 5.14 and 5.15 was used as input for topology tests as described in the Methods. Constraint tree topologies were constructed to represent alternative rooting hypotheses for the SNARE family trees within the superfamily tree as indicated (*e.g.*, all SNARE families rooted on the respective plasma membrane (PM)-localized family members). Branches within clades representing eukaryotic outparalogues identified in previous family-specific analyses were constrained to reduce the tree space. Maximum Likelihood (ML) trees obtained for each of these alternative topology constraints were compared and tested separately for the two datasets using the best-fit substitution models (LG+F+G4 for Test 1 and LG+F+R4 for Test 2). Approximately Unbiased (AU) tests (Shimodaira, 2002) with RELL bootstrap approximation (Kishino *et al.*, 1990) as well as calculation of Expected Likelihood Weights (ELW) among the alternative hypotheses (Strimmer and Rambaut, 2002) were performed using IQ-TREE. These confidence values are distributed among the alternatives for each test and sum to a total of 1.0. In **Test 1**, the ML tree topology without rooting constraints was the same as shown in Figure 5.14 (Qa family rooted on branch separating ER and Golgi SNAREs from PM and endosomal SNAREs), but the tree with the highest log likelihood (logL) was the ML tree obtained with the ER root constraints. The difference in logL between these two trees (ΔL) was small. The AU test P-values (p-AU) for Test 1 suggest that the hypotheses involving rooting all the SNARE families on endosomal SNAREs or all on Golgi SNAREs may be rejected, as less likely than the ER rooting hypothesis, with the standard significance threshold of $\leq$0.05. The ELW confidence values (c-ELW) indicate 96% confidence that the true topology is represented by the ER or PM rooting hypotheses or the ML tree with no rooting constraints. In **Test 2**, the ML tree topology without rooting constraints differed from the consensus of bootstrap topologies shown in Figure 5.15, as the Qc family was rooted on the clade containing orthologues of the ER SNARE Use1. The AU test P-values for Test 2 suggest that the hypotheses involving rooting all the SNARE families on endosomal SNAREs, all on ER SNAREs, or all on Golgi SNAREs may be rejected, as less likely than the ML topology without rooting constraints. The ELW confidence values indicate in this case 96% confidence that the true topology is represented by the ML tree with no rooting constraints or the PM or middle rooting hypotheses, but the confidence for the middle rooting hypothesis is less than 0.05.

Test 1 (Qa, Qb, Qc, and R)

Root constraints	logL	ΔL	c-ELW	p-AU	Result
ER	-35648.619	-	0.5513	0.5998	Accept
None	-35651.041	2.422	0.3150	0.6216	Accept
PM	-35653.126	4.508	0.0973	0.1986	Accept
Middle	-35660.198	11.579	0.0237	0.1162	Accept
Endosome	-35663.885	15.266	0.0055	0.0386	Reject
Golgi	-35667.821	19.202	0.0072	0.0276	Reject

Test 2 (Qb, Qc, and R)

Root constraints	logL	ΔL	c-ELW	p-AU	Result
None	-27837.959	-	0.4607	0.6895	Accept
PM	-27838.067	0.108	0.4596	0.5813	Accept
Middle	-27843.583	5.624	0.0378	0.0664	Accept
Endosome	-27844.674	6.715	0.0204	0.0483	Reject
ER	-27849.143	11.184	0.0046	0.0113	Reject
Golgi	-27865.980	28.021	0.0170	0.0238	Reject

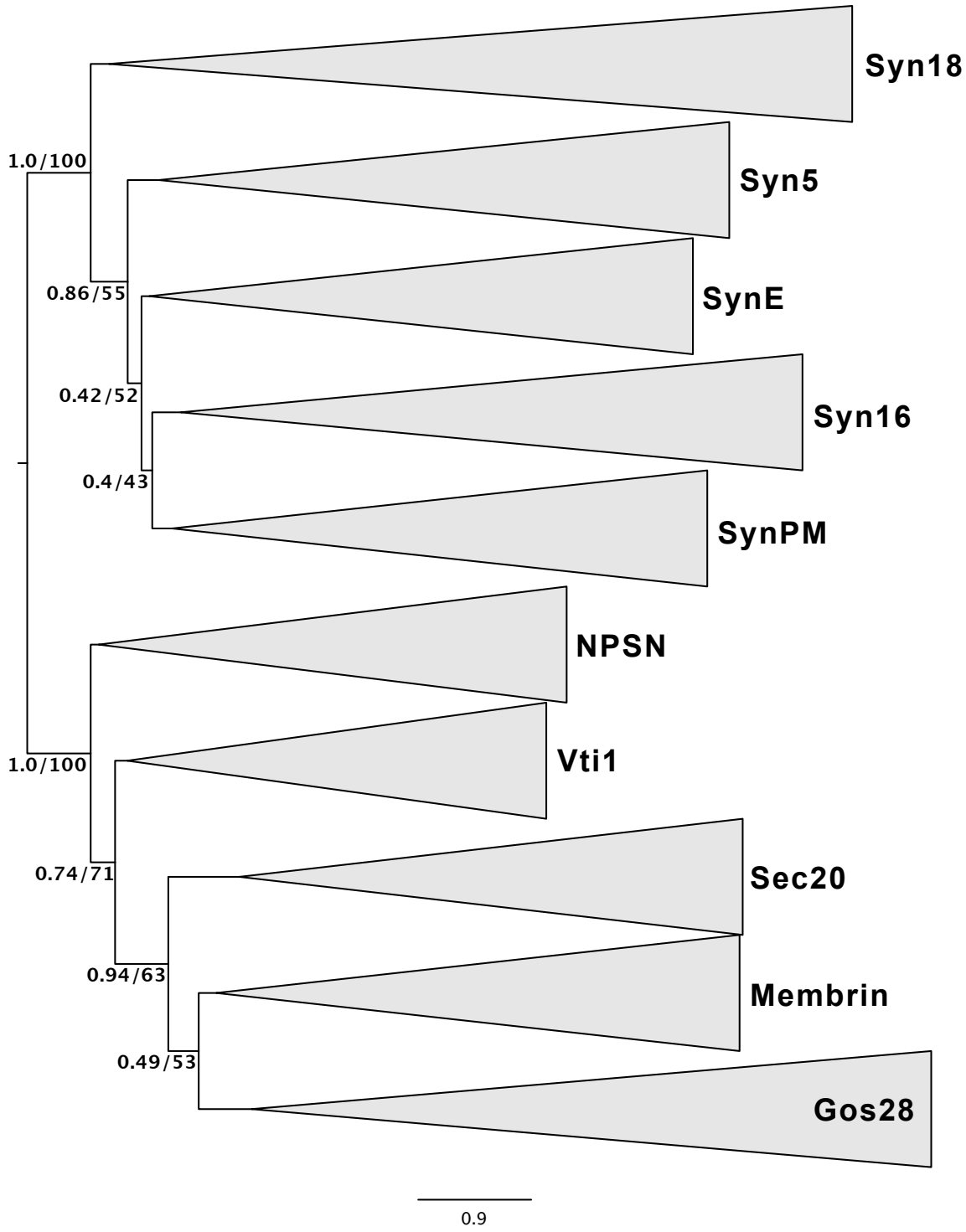

Syn18
1.0/100
Syn5
0.86/55
SynE
0.42/52
Syn16
0.4/43
SynPM
NPSN
1.0/100
Vti1
0.74/71
Sec20
0.94/63
Membrin
0.49/53
Gos28
0.9

Figure 5.16: Phylogenetic analysis of the Qa and Qb SNARE families. An alignment was generated by removal of Qc-SNARE and R-SNARE sequences from the superfamily alignment used in Figure 5.14. Constraint tree topologies were constructed to only constrain the monophyly of previously identified orthologous sequences within the Qa-SNARE and Qb-SNARE families (see Appendix A chapter 5 Figures S3 and S4). The best-fit substitution model used for this analysis was LG+F+R5. Consistent with the analyses shown in Figures 5.14 and 5.15, significant support is found for rooting of the Qb-SNARE family on the NPSN clade. However, in this analysis significant support is also found for rooting the Qa-SNARE family on the Syn18 clade, in contrast to the topology shown in Figure 5.14. See Table 5.3 for results of analyses with alternative combinations of SNARE family sequences.

Table 5.3: Summary of phylogenetic analyses of SNARE family pairs. Alignments were generated by removal of sequences from the superfamily alignment used in Figure 5.14. Constraint tree topologies were constructed to only constrain the monophyly of previously identified orthologous sequences. For each of the six alignments, the best-fit substitution model was used, as indicated, for ML analyses, and the LG+G4 model was used for bayesian analysis. Branch supports are shown for the branches separating the deepest branching clade within each SNARE family from the remaining SNARE clades of the same family (which defines support for the rooting). The organelle specificity of the SNAREs in the root clade(s) are indicated, as detailed in the introduction. Significant support is only found for rooting of Qb-SNAREs on NPSN, Qc-SNAREs on Syp7, and Qa-SNAREs on Syn18. Support values that indicate significant support for a root placement are shown in bold font. Where the root is placed on a branch with more than one clade on either side, both support values are shown separated by a semicolon. For tree figures see (see Appendix A chapter 5 Figures S6-S11).

SNARE family	Alignment	Substitution model	Root clade	Organelle specificity	Prior probability	Bootstrap percentage
Qa	Qa+Qb	LG+F+R5	Syn18	ER	**0.86**	**55**
Qa	Qa+Qc	LG+F+G4	Syn18+Syn5	ER/Golgi	0.35; 0.26	33; 20
Qa	Qa+R	LG+F+G4	Syn18	ER	0.79	40
Qb	Qa+Qb	LG+F+R5	NPSN	PM	0.74	**71**
Qb	Qb+Qc	LG+F+R5	NPSN	PM	0.69	**50**
Qb	Qb+R	LG+F+R4	NPSN	PM	0.54	**55**
Qc	Qa+Qc	LG+F+G4	Syp7	PM	0.75	**67**
Qc	Qb+Qc	LG+F+R5	Use1	ER	0.31	22
Qc	Qc+R	LG+F+G4	Bet1	Golgi	0.44	27
R	Qa+R	LG+F+G4	Syb+VAMP7	PM/Endosomal	0.62; 0.48	30; 59
R	Qb+R	LG+F+R4	Sec22	ER	0.51	13
R	Qc+R	LG+F+G4	Syb+VAMP7	PM/Endosomal	0.55; 0.58	32; 57

5.4 Conclusions and future directions

Our phylogenetic analyses of Qbc SNAREs indicate a close relationship with the SNAREs NPSN and Syp7, favouring an ancestral function of these canonical SNAREs in exocytosis. Also, phylogenetic analysis of the R-SNARE family confirms that synaptobrevin orthologues in extant eukaryotes are inherited from the last common ancestor of eukaryotes. Building upon these results, phylogenetic analysis of the SNARE superfamily so far favours scenarios of early endomembrane evolution involving origin of the widely conserved endomembrane organelles of extant eukaryotes from a single progenitor organelle which may have been ER- and/or endosome-like. The evidence also favours endosomes originating prior to differentiation of the Golgi and ER. While some alternative rooting hypotheses were not conclusively rejected here, further analysis of the SNARE superfamily involving more aggressive selection of sequences to optimize for higher phylogenetic signal may yield more conclusive results. In addition, further phylogenetic analyses of additional protein families with paralogues specific to the ER, Golgi, endosomes, and PM may refute or confirm the results herein if they yield contrasting or consistent topologies with respect to the organelle-specificity of paralogues. Further work is also needed to clarify the timing of the differentiation between the Golgi and ER relative to the differentiation between early and late endosomes/lysosomes, for which we argue SNAREs provide little information.

5.5 Methods

5.5.1 Taxonomic sampling and genomic data sources

A broad sampling of 112 genomes and transcriptomes from eukaryotes was assembled from varous public databases (See Table S1). Lineages represented include the most species-rich taxonomic groups of eukaryotes Metazoa, Holozoa, Amoebozoa, Metamonada, Discoba, Stramenopiles Alveolates and Rhizaria (SAR), Haptophyta, Cryptista, and Archaeplastida. Emphasis was placed on less divergent, more complete representative species/genomes for example, parasitic lineages were not sampled extensively, but genomes that are known to have evolved relatively slowly such as spotted gar and alligator were included (Braasch *et al.*, 2016; Green *et al.*, 2014). Transcriptomes from the MMETSP project were used to increase sampling particularly from among amoebozoans and cryptophytes (Keeling *et al.*, 2014). In addition, previously classified Synaptobrevin and VAMP7 sequences were downloaded from the SNAREDB website `http://bioinformatics.mpibpc.mpg.de/snare/snareMainPage.jsp` (Kloepper *et al.*, 2007).

5.5.2 Similarity searching

Predicted peptide sequences for each sampled genome or transcriptome were searched with HMMER 3.1b1 (hmmsearch) (`http://hmmer.org` Eddy (1998)) using a general SNARE domain HMM constructed from reference SNARE sequences, including those from previous studies for Qa SNAREs (Arasaki *et al.*, 2015) as well as Qb and Qc SNAREs (Venkatesh *et al.*, 2017). Alignments for constructing HMMs were generated using MUSCLE v3.8.31 (Edgar, 2004) with default parameters. Hits retrieved with an E-value less than or equal to 0.05 were searched against a database of HMMs with specific HMMs for Qa-SNAREs, Qc-SNAREs, Qb-SNAREs, Qbc-SNAREs, R-SNAREs, and Tomosyn using HMMer (using the hmmscan program in the HMMer3 package). The top hit HMM was used to classify sequences into one of these categories. Sequences that did not retrieve a family-specific SNARE HMM with an evalue less than or equal to 0.05 were discarded. Also, each positive hit was used as a query to search in the predicted proteins of *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana* using the BLASTP algorithm in the Basic Local Alignment Search Tool (BLAST)+ software package (Camacho *et al.*, 2009). A detailed summary of the results of all of these sequence similarity searches is provided in (Table S2). Scripts that were used for running these similarity searches are available here: `https://github.com/laelbarlow/amoebae` (Larson *et al.*, 2019). These scripts make use of Python libraries including Biopython (Cock *et al.*, 2009).

5.5.3 Phylogenetic analysis

Sequence sets for phylogenetic analysis of each SNARE family were generated initially from top BLASTP hits for sequences from previous reference alignments of Q SNAREs (Arasaki *et al.*, 2015; Venkatesh *et al.*, 2017) and from previously annotated reference sequences for R-SNAREs. Alignments were constructed using MUSCLE v3.8.31 (Edgar, 2004) then positions/sites were trimmed to include only positions displaying potential homology. Amino acid substitution models were selected for each trimmed alignment using ModelFinder (Kalyaanamoorthy *et al.*, 2017), and the best-fit model according to the Bayesian Information Criterion (BIC) (Posada and Buckley, 2004) was used for phylogenetic analyses (by default). Initially, several rounds of Maximum Likelihood (ML) phylogenetic analysis were performed using IQ-TREE version 1.6 (Nguyen *et al.*, 2015) specifying 1000 ultrafast bootstraps to estimate branch support (Minh *et al.*, 2013). Between subsequent rounds of analysis, divergent sequences were manually identified for removal to improve resolution and in some cases alternative representative sequences added to increase taxonomic representation.

Final ML phylogenetic analyses were run with IQ-TREE specifying 100 non-parametric bootstrap pseudoreplicates to quantify bootstrap percentage support for branches, with 50% generally being considered the minimum threshold for significance. Also, for more thorough tree-searching and calculation of posterior probabilities for branches, MCMC tree searching and bayesian analysis were was performed using MrBayes version 3.2.6 (Ronquist and Huelsenbeck, 2003). For all bayesian analyses, two MCMCs were run for 10 million generations or until the standard deviation of splits frequencies reached 0.01, indicating convergence between the two chains, and a burnin percentage of 25 was used to exclude highly unlikely topologies from the bayesian analysis. A bayesian posterior probability of 0.8 (out of 1.0) was considered the threshold for significance. Phylogenetic analyses were run either on clusters provided by Compute Canada (`www.computecanada.ca`) or on the CIPRES webserver (Miller *et al.*, 2010).

The alignment for the SNARE superfamily analyses was constructed by aligning SNARE family-specific sequence alignments to each other by specifying the '-profile' option when running MUSCLE v3.8.31 (Edgar, 2004). This preserved the quality of each alignment, preventing introduction of alignment errors. It was ensured that the positions representing the zero layer of the SNARE domains in each sub-alignment were aligned at the same position. An alignment identical to a definitive SNARE domain reference alignment (Fasshauer *et al.*, 1998) was used as a guide for trimming all positions representing residues outside of the SNARE domains.

For topology testing, alternative topologies with polytomies were constructed as described in the Results and Discussion (5.3) to constrain IQ-tree tree searches (with the same input alignment) to find the best ML trees representing hypotheses of interest. Scripts were written for manipulation of tree topologies to construct constraint trees. These scripts make use of Python libraries including the Environment for Tree Exploration (Huerta-Cepas *et al.*, 2010, 2016), and are available here: `https://github.com/laelbarlow/amoebae`. IQ-TREE (Nguyen *et al.*, 2015) was used for ML searches with the '-allnni' option for more thorough tree searching, and the best ML tree of five runs was used in each case (specified using the '–runs' option). Approximately Unbiased (AU) tests (Shimodaira, 2002) were performed using IQ-TREE (Nguyen *et al.*, 2015). In addition, the Expected Likelihood Weights method was used to calculate confidence in the alternative hypotheses using IQ-TREE, as this approach may be relatively robust against model miss-specification (Strimmer and Rambaut, 2002). For both of these tests, 10,000 resamplings were performed using the Resampling Estimated Log-Likelihood (RELL) method (Kishino *et al.*, 1990) as implemented in IQ-TREE.

5.5.4 Plasmid construction

The NPSN gene (DDB_G0285365) was amplified from *Dictyostelium discoideum* AX2 genomic DNA using PCR. The forward (5') primer (5'-GCAGTCAGATCTATGGCCGATTTACA AGAAAATG-3') contained an additonal *Bgl*II restriction site, and the reverse (3') primer (5'-CT GACGACTAGTTTATTGGGTTGAACCAGTAACAT-3') contained a *Spe*I restriction site. Due to an apparent lack of complexity in the flanking regions of the Syp7 gene (DDB_G0289063), this gene could not be amplified by PCR so the coding sequence was synthesized by Invitrogen. These products were ligated into into non-integrating (extra-chromosomal) expression vectors Veltman *et al.* (2009), amplified using *Escherichia coli* DH5α, and verified by sanger sequencing. Specifically, we used the plasmids pDM1207 (for N-terminal Green Fluorescent Protein chimeras) and pDM1208 (for N-terminal mCherry chimeras), which drive overexpression of chimeras with the actin15 promoter and drive expression of a G418 resistance cassette with the cofillin A promoter.

5.5.5 *Dictyostelium* AX2 cell culture and transformation

Cells were cultured at 22°C suspended in liquid HL5 media Ashworth and Watts (1970) in an orbital shaker at 180 revolutions per minute. Amoebae were transformed via electroporation as described previously (Gaudet *et al.*, 2007). For this, $3x10^7$ cells were harvested from liquid culture at a density of approximately $1\text{-}5x10^6$ cells/ml (log phase) by centrifuging at 300g for 3 minutes, resuspended in H50 electroporation buffer and put on ice. For each transformation, 5μg of plasmid DNA was pipetted into a sterile 1mm electroporation cuvette (at 4°C), to which 100μL of cell suspension (also at 4°C) was added. This mixture was electroporated using a Bio-Rad gene pulser electroporator with the following settings: Exponential voltage curve, 750 volts, 25 micro Farads, and ∞ Ω resistance. Two pulses were applied to the cells with 5 seconds between pulses. The cuvette was then set on ice for 5 minutes, and then the electroporated cells were incubated in HL5 media in petri dishes at 22°C overnight. Transformants were then selected using G418 at 10μg/mL.

Chapter 6

General discussion

6.1 Potential lack of phylogenetic signal

Attempts to reconstruct evolutionary relationships between genes that diverged early in eukaryotic evolution, including those detailed in the preceding chapters, vary considerably in the extent to which they achieve robust answers. Gene homology provides no guarantee whatsoever of sufficient conservation of phylogenetic signal for reconstruction of all deep relationships between homologues. Despite this, phylogenetic analyses of membrane trafficking proteins with organelle-specific functions is often successful at classifying homologous sequences as orthologues of specific eukaryotic outparalogues. This makes sense given the relatively recent shared ancestry of these orthologues, as well as frequent conservation of function from the LECA to extant eukaryotes. The analyses of the synaptobrevin-related R-SNAREs and the SNAPN and SNAPC domains of Qbc-SNAREs presented in Chapter 5 are examples of low but arguably sufficient phylogenetic signal for determining orthology of paralogues from diverse eukaryotes (see Figures 5.3 and 5.4). Moreover, it is not always necessary to perform a phylogenetic analysis, as similarity searches and/or manual inspection of sequence alignments often allows conclusive classification of orthologues. For example, phylogenetic analysis is often not necessary to classify adaptins as orthologues of specific Adaptor Protein complex subunits, as seen in Chapter 3 (see Figure 3.2 and Larson *et al.* (2019) Figure 2). Also, phylogenetic analysis is generally not necessary to distinguish between SNAREs of different families (*e.g.*, R-SNAREs vs. Qa-SNAREs) (see Appendix A chapter 5 Table S2).

Although to my knowledge there are no established methods to conclusively determine *a priori* whether a sequence alignment contains enough information to determine the true topology of relationships between eukaryotic outparalogues, there is reason to suspect that this is unlikely for any given membrane trafficking gene/protein family. One reason is that analyses including phylogenomic analyses have been unable to reach a consensus regarding the relationships between

the major lineages of eukaryotes: Amorphea, Metamonada, Discoba, and Diaphoretickes. And, these lineages diverged at a later time point than did paralogues of interest in the SNARE protein families, suggesting higher levels of phylogenetic signal should be present. An approach based on minimizing gene duplication and loss events (gene tree parsimony) taken by Katz *et al.* (2012) suggested that the tree of eukaryotes is rooted on Opisthokonta. A phylogenomic analysis of mitochondrial and related alphaproteobacterial genes found support for placement of the root between Amorphea and a clade containing Discoba and Diaphoretickes (although Metamonada was not represented) (Derelle and Lang, 2012). He *et al.* (2014) found support for rooting the eukaryotic tree on Discoba (although Metamonada was not represented). Perhaps the best justified analysis inferring a root of the eukaryotic tree to date is that of Derelle *et al.* (2015), who performed a re-analysis of the dataset used by He *et al.* (2014) as well as additional mitochondrial genes and finding support for a root between Amorphea and a clade containing Diaphoretickes and Discoba (similar to Derelle and Lang (2012)), but again did not include sequences from Metamonada. More recent analyses support the monophyly of the Diaphoretickes, Discoba, Metamonada, and Amorphea clades, but does not provide evidence for a root placement (Brown *et al.*, 2018; Lax *et al.*, 2018). Another reason to predict low levels of phylogenetic signal is the potential for very short timespans between consecutive gene duplications that produced paralogous gene/protein families. This uncertainty is supported by the absence of most membrane trafficking proteins, such as SNAREs, in the closest archaeal relatives of eukaryotes Klinger *et al.* (2016b), as well as a lack of extant organisms representing intermediate stages of membrane trafficking evolution.

6.2 Alignment quality and sequence selection

If phylogenetic analysis is to be performed, the strength of phylogenetic signal depends on the quality of the input multiple sequence alignment, and selection of homologous sequences for inclusion in alignments can affect the tree topology inferred and branch supports calculated. The analyses including sequences from the Qc-SNARE family presented in Chapter 5 of this thesis illustrates a lack of consistency between the topologies inferred from analyses with different selections of outgroup sequences. For example, in the analyses of the Qc-SNARE family alone, the Syp7 (and SNAPC) clade was sister to the Syn6 clade (Figure 5.4 and Appendix A chapter 5 Figures S2 and S5). However, the topologies of the Qc-SNARE subtrees in all of the analyses where sequences from other families were included were incompatible with this topology, and all suggest instead that Syp7 is sister to the ER SNARE Use1 if the Qc-SNARE subtree were unrooted (Figures 5.14 and 5.14 and Appendix A chapter 5 Figures S7, S9, and S11). Significant support for this

alternative topology was not found, but this perhaps suggests the presence of conflicting phylogenetic signal in the SNARE domain subsequences compared to the additional sequence positions included in the Qc-SNARE-only analyses.

Phylogenetic analyses of AP, COPI, and TSET complexes shows inconsistencies among the topology of relationships between eukaryotic outparalogues revealed by alternative sequence selections as well. The topologies for the phylogenetic tree of the AP (1-5) and COPI complexes reported by Hirst *et al.* (2011) (their figure 6) and a very similar analysis by Hirst *et al.* (2014) (their figure 4A) that included subunits of the TSET complex both show significant support for all internal branches. Thus both of these studies might be interpreted in terms of the OPH in an attempt to make inferences regarding organelle evolution. However, these two analyses resulted in incompatible topologies: The analysis by Hirst *et al.* (2011) shows the AP-3 complex as sister to the clade comprising AP-1, 2, and 4, while the analysis by Hirst *et al.* (2014) shows AP-5 instead as sister to this clade (in other words the branches leading to AP-3 and AP-5 switched positions in the tree). Hirst *et al.* (2014) did find considerably higher posterior probability and bootstrap support for their topology, but Hirst *et al.* (2011) found significant support as well (prior probability 0.93 and bootstrap percentage $\geq$57).

Another example is the analysis of holomycotan AP complexes in Chapter 2 of this thesis. In that case, analysis of AP complexes 1 to 4 resulted in highly significant exclusion of microsporidian sequences from both the holomycotan AP-1 and AP-2 clades, which if interpreted literally would imply that these sequences are neither AP-1 nor AP-2 sequences (see Figure 2.4 and Barlow *et al.* (2014) Figure 4A). When AP-3 sequences were removed, however, then the microsporidian sequences were only significantly excluded from the AP-2 clade (see Figure 2.5 and Barlow *et al.* (2014) Figure 4B), and in this case the authors agreed that the dataset with fewer sequences yielded the more informative results (although this is not conclusive). So, phylogenetics of AP, COPI, and TSET complexes illustrates that in some cases addition of a new clade of homologous sequences, as well as a few minor changes to the taxonomic sampling from which homologues were used, may be all that is required to drastically change what deep phylogenetic relationships seem to be supported. This also illustrates that posterior probabilities and bootstrap percentages may not be sufficient to indicate the robustness phylogenetic results for relationships between eukaryotic outparalogues.

These examples of the effect of sequence selection in determining both the level of branch support and topology of phylogenetic results suggests the need for analysis of multiple datasets to identify sensitivity to outgroup selection (the approach taken in Chapter 5 of this thesis), as well as

systematic approaches to sequence selection. In any case, some subset of homologous sequences must be selected, because alignments with large numbers of protein sequences tend to be poor quality due to accumulation of alignment errors (Sievers *et al.*, 2013). The challenge of selecting sequence data is an emerging problem in this area of research that has arisen due to increases in computational resources making analysis of many alternative datasets (alignments) somewhat more feasible, and increases in the number of sequenced genomes and thus homologous sequences available. In an analysis of eukaryotic outparalogues with pan-eukaryotic taxon sampling the number of homologous sequences identified can be quite large. For example, the similarity searches performed in Chapter 5 identified 4,504 SNARE sequences among the 113 genomes and transcriptomes queried (Appendix A chapter 5 Table S2). While the approach taken in Chapter 5 was based on manual sequence selection over several iterations of phylogenetic analysis, sequence selection is a formidable task and there are several methods of systematic sequence selection that may be useful for further work.

Random selection and analysis of multiple subsamples of sequences is one approach that has been implemented previously with some success (Sloutsky and Naegle, 2016). This would provide some indication of the extent to which support for certain backbone topologies is dependent on sequence selection, and whether the majority of possible subsamples result in different topologies than that resulting from a manually selected subsample (or a systematically selected sample of short-branching sequences). A drawback of this approach is that it is computationally intensive, due to the requirement for analysis of potentially thousands of alternative alignments. For example, in Chapter 5 the alignment of R-SNARE sequences that yielded support for a pan-eukaryotic synaptobrevin clade included 83 sequences, and the number of canonical R-SNARE sequences identified in the similarity searches was 1,017 (see Figure 5.10 and Appendix A chapter 5 Table S2). Given that the formula to calculate the number of k-element subsets of an n-element set is $\frac{n!}{k!(n-k)!}$ (see Mazur (2010) pg. 9), the number of possible selections of 83 sequences from 1,017 total sequences is $3.29\text{x}10^{123}$. This is rather astronomical, especially considering that the optimal number of sequences to select might be in a range of numbers. Also, random selection of sequences does not take into consideration that highly divergent sequences may be less informative for phylogenetic analysis than other sequences.

Another approach to sequence selection, which has been applied to inferring relationships between eukaryotic outparalogues is a method based on pairwise maximum likelihood distances calculated from a multiple sequence alignment (Elias *et al.*, 2012). This method has been termed Scrollsaw, and was first applied to analysis of the Rab family of small GTPases (Elias *et al.*, 2012).

Scrollsaw is intended to select the least divergent, most informative representative sequences for analysis, and works by identifying sets of sequences among eukaryotic supergroups that show shortest distances (most sequence conservation) from a matrix of pairwise distances between sequences. These pairwise distances are measures of the evolutionary distance between each unique pair of sequences in an alignment given a model of amino acid substitution, and are calculated by phylogenetics software packages such as RAxML and IQ-TREE for generating initial starting trees for maximum likelihood analysis (Stamatakis, 2014; Nguyen *et al.*, 2015). This method was successful in identifying subsamples of Rab protein sequences that result in relatively higher bootstrap proportion and posterior probability values supporting some deep branches of the Rab family tree including support for a clade containing several Rabs with endocytic functions, and a clade containing several Rabs with exocytic functions (Elias *et al.*, 2012). A very similar approach was taken by (Gabernet-Castello *et al.*, 2013) in analysis of the Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. In this case, however, more highly conserved sequences were identified by their higher rank among BLAST search results using representative sequences from model systems as queries. This was successful in finding phylogenetic support for several clades containing multiple eukaryotic TBC outparalogues. The main advantages of these Scrollsaw methods is that selection of sequences is systematic and reproducible, in principle, and computationally efficient due to reliance on distance-based methods (Pardi and Gascuel, 2016).

However, in both of these studies that used Scrollsaw, multiple rounds of phylogenetic analysis and visual/manual selection of divergent proteins for removal were still required to resolve the phylogenies (Elias *et al.*, 2012; Gabernet-Castello *et al.*, 2013). Also, while Scrollsaw may be advantageous for many analyses, an inherent limitation of Scrollsaw is its dependence on information yielded from pairwise distances. For example, in the analysis of R-SNAREs in Chapter 5, I detailed extensive differences between similarity search results (both pairwise BLAST searches and profile methods) and the results of my phylogenetic analyses, which are based on searches for tree topologies with high likelihood or probability given models of sequence evolution and considering all the sequences in an alignment. Also, trees generated using the same R-SNARE alignment (as for the analysis in Figure 5.10) as input but using a neighbor-joining algorithm, BIONJ implemented in IQ-tree (Gascuel, 1997; Nguyen *et al.*, 2015), that clusters sequences based on pairwise ML distances did not yield a topology indicating a relationship between Opisthokont synaptobrevin orthologues and non-opisthokont synaptobrevin orthologues. Thus Scrollsaw would likely not be useful for identification of sequence sets representing synaptobrevin orthologues in both opisthokonts and non-opisthokonts.

Another alternative approach to sequence selection which is systematic and selects for more informative sequences is to perform iterative rounds of phylogenetic analysis and apply specific criteria for identification of sequences for removal from subsequent rounds of analysis based on phylogenetic results. This approach was taken in Chapter 3 for identification of duplication events in the individual AP, COPI, and TSET subunits (see Figures 3.3 and 3.4 as well as Larson *et al.* (2019) Figures 3 and 4). This was successful particularly in identifying redundant and highly divergent sequences for removal. This method was implemented in the AMOEBAE script toolkit, as described in Chapter 3, and can be applied via the 'amoebae auto_prune' command. Conveniently, this allows identification of long-branching and taxonomically redundant sequences based on quantitative criteria and generation of alignments with these sequences removed in a single automated step. While the limited scope of the analyses performed in Chapter 3 may partially account for success of this method, it may be useful for future analyses aimed at resolving relationships between eukaryotic outparalogues.

6.3 Efficient classification of additional sequences

Phylogenetic classification of particular protein sequences within the SNARE superfamily and other protein families is of practical value for formulating hypotheses about their roles in the functioning of particular eukaryotic cells. For example, it might be useful to know whether a VAMP7-like R-SNARE in the genome of a particular parasitic eukaryote is orthologous to VAMP7 or synaptobrevin, as this might lead to different predictions for experimental work. However, even the most comprehensive phylogenetic analyses do not include all available genomes, and comprehensive re-analysis of protein families with additional sequences is a considerable undertaking. Also, there is a continual need to accurately classify genes encoding membrane trafficking proteins in newly sequenced genomes. Ideally, a full re-analysis of protein family trees need not be done in every study of genes/proteins that have not been previously classified. Several possibilities exist using previous phylogenetic result efficiently, but these do not always yield accurate results for distinguishing between orthologues of different eukaryotic outparalogues.

Due to the need for simplicity and computational efficiency, automated pipelines for protein classification usually rely on pairwise sequence comparison or profile-based methods such as using Hidden Markov Models (HMMs) built from alignments of representative sequences. Such a utility for classifying SNARE proteins was constructed and made available by Kloepper *et al.* (2007) (`http://bioinformatics.mpibpc.mpg.de/snare/index.jsp`). This utility is capable of accurately classifying many if not most SNARE protein sequences. However, my subsequent

analyses revealed that this utility is not sufficient to classify all SNARE proteins. First, no HMM was included to represent orthologues of the Qa-SNARE outparalogue Syntaxin 17, which I identified more recently (Arasaki *et al.*, 2015). Moreover, (Kloepper *et al.*, 2007) used HMMs of only the SNARE domain sequences, but the transmembrane domains of Syntaxin 17 are more informative for classification (Arasaki *et al.*, 2015). Also, as detailed in Chapter 5, this utility does not yield the same classifications of synaptobrevin orthologues that I identified using phylogenetic analysis (see section 5.3.4), and this may be due to inherent limitations of HMM-based sequence classification methods. Other similar pipelines include a Rab protein classifier constructed by Diekmann *et al.* (2011), which allows rapid classification of Rab sequences, but again not to the level of detail provided by comprehensive phylogenetic analysis (Elias *et al.*, 2012).

To efficiently apply results of comprehensive phylogenetic analyses of eukaryotic outparalogues of membrane trafficking proteins, future work should include development of automated pipelines that yield more accurate and detailed results. It would be necessary for such methods to be based on phylogenetic analysis rather than pairwise or profile-based comparisons, for the reasons described above. One possibility would be to construct a pipeline that adds sequences to a reference alignment and performs a partially constrained phylogenetic analysis to increase efficiency by reducing the number of possible tree topologies to be considered (the tree space).

6.4 Interpretion of the evolution of the dissimilar but functionally related golgin proteins

As argued above, homology of genes alone provides no guarantee of sufficient phylogenetic signal for accurately inferring their phylogenetic relationships based on their primary DNA structure or the primary structure of the peptides they may encode. To take this idea one step further, higher levels of sequence divergence could be expected to obscure evidence of homology, as indicated by measures of sequence similarity, as well. As noted in the afterword of Chapter 4, the golgin coiled-coil tethers with specificity for different subdomains of the Golgi do not show similarity in their amino acid sequences, suggesting the possibility that they originated independently in evolution and not from a single golgin that may have functioned in a primordial undifferentiated Golgi. This contrasts with what might be predicted by the OPH (Dacks and Field, 2007), and what is implied by the term "family" commonly being applied to this collection of proteins as a whole (Witkos and Lowe, 2016). Indeed, origin from a single ancestral golgin would be more parsimonious than a scenario involving multiple independent origins of proteins with such comparable structure and function.

Multiple golgin researchers have previously suggested that the lack of sequence conservation

among golgins may be due to a lack of evolutionary constraints on their structure (Munro, 2011). The coiled-coil domains of golgins appear to serve mostly as semi-flexible linkers between a motif that anchors one end to a Golgi membrane, and a motif at the opposite end that binds to specific vesicles (Oas and Endow, 1994; Yamakawa *et al.*, 1996; Cheung *et al.*, 2015). Considering this, then function of golgins would be independent of sequence in the coiled-coil (except to the extent that propensity for coiled-coil formation must be maintained). And, the N- and C-termini would also be expected to be very different due to their role in attaching to specific proteins/membranes.

It is certainly more difficult to trace evolutionary history of proteins that share little or no sequence similarity, but may not be impossible. A relevant example of remarkable absence of sequence similarity has been observed in the vertebrate-specific coiled-coil tether protein Tetherin, which tethers exosomes and viruses to the cell surface of vertebrate cells (Blanco-Melo *et al.*, 2016; Edgar *et al.*, 2016). Orthologues of this protein share no evident sequence similarity among vertebrates, but could be identified based on gene synteny and very general structural features (an N-terminal transmembrane domain followed by a coiled-coil domain followed by a GPI anchor). Similar investigations may be an interesting avenue of future research on golgin evolution, and may aid in identification of novel lineage-specific golgins (or golgin-like proteins).

Although not always present, high levels of protein sequence conservation represent the best-case scenario in terms of the impact of studies in the field of cell biology, allowing inference of more direct relevance between model systems. Thus it makes sense to focus first on highly conserved proteins. However, many examples exist of unstructured protein domains with important functions (Tompa, 2002), and this is worth considering regarding the evolution and function of golgins. If the function of proteins depends on being fundamentally disordered, then they could be expected to have less highly conserved sequences among species. As mentioned above, golgins appear to be at least partially flexible/disordered. While this may be important for the tethering functions of golgins, other functions may be possible as well. Golgins form a dense protein matrix surrounding each Golgi cisterna, at least in mammalian cells, and this matrix has been implicated in maintaining the structure of the Golgi (Cluett and Brown, 1992), although as discussed in Chapter 4 no particular golgins have been identified as having a conserved role in maintaining Golgi structure (see Figure 4.4 and Barlow *et al.* (2018) Figure 4). A recent opinion piece proposed a new model for explaining Golgi cisternal adhesion based on liquid phase separation generated by the disordered domains of golgins (Rothman, 2019). This would be analogous to the liquid phase separation observed within the nucleus between the nucleolus and the remaining nucleoplasm (Feric *et al.*, 2016).

This suggestion by Rothman (2019) of a primary role for disordered domains and numerous very low-affinity protein-protein interactions in stacking the Golgi and mediating vesicle tethering is coincidentally consistent with the findings regarding the evolution of golgins which we reported in Chapter 4 of this thesis (see Figure 4.4 and Barlow *et al.* (2018) Figure 4). Also, this is consistent with experimental work demonstrating that higher expression levels of certain golgins can compensate for the loss of other golgins (Lee *et al.*, 2014). Considering the evolution of golgins and these possible functional mechanisms, a clear evolutionary prediction would be that all eukaryotes with the typical complex stacked Golgi morphology (Mowbrey and Dacks, 2009) will possess abundant semi-disordered coiled-coil proteins analogous (if not homologous) to mammalian golgins. So far, a protein matrix similar to the mammalian Golgi matrix has been suggested to maintain the structure of the plant Golgi (Hawes, 2004), but is less well-characterized.

6.5 General conclusions

The membrane trafficking system is an intricate, dynamic, and only partially understood component of cell biology. It is essential to human health and the virulence of some of the most terrifying pathogens. Overall, the results of the work presented herein contribute to development of testable theories for the early evolution of the membrane trafficking system which are important for unifying the cell biology of distantly related eukaryotes. These results indicate that the adaptin, golgin, and SNARE proteins are each sources of information for deciphering organelle evolution. The adaptins exhibit evolutionary patterns consistent with a close connection between adaptin evolution and membrane trafficking evolution. The golgins reveal conservation reflecting inferred ancient complexity of the Golgi apparatus. Analysis of SNARE proteins provided the first phylogenetic tests of hypotheses that have been pondered in the field of cell biology for over half a century. As more organelle-specific proteins are characterized experimentally as well as subjected to evolutionary analysis, accumulating evidence will likely allow testing of more detailed hypotheses to describe the evolutionary history of organelles and pathways in the membrane trafficking system.

Bibliography

Aalto, M., Ronne, H., and Keränen, S. (1993). Yeast syntaxins Sso1p and Sso2p belong to a family of related membrane proteins that function in vesicular transport. *The EMBO Journal*, 12(11):4095–4104. doi: 10.1002/j.1460-2075.1993.tb06093.x.

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*, 21(9):2104–2105. doi:10.1093/bioinformatics/bti263.

Abeliovich, H., Grote, E., Novick, P., and Ferro-Novick, S. (1998). Tlg2p, a Yeast Syntaxin Homolog That Resides on the Golgi and Endocytic Structures. *Journal of Biological Chemistry*, 273(19):11719–11727. doi: 10.1074/jbc.273.19.11719.

Abou Jamra, R., Philippe, O., Raas-Rothschild, A., Eck, S.H., Graf, E., Buchert, R., Borck, G., Ekici, A., Brockschmidt, F.F., Nöthen, M.M., Munnich, A., Strom, T.M., Reis, A., and Colleaux, L. (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *American Journal of Human Genetics*, 88(6):788–795. doi:10.1016/j.ajhg.2011.04.019.

Acehan, D., Santarella-Mellwig, R., and Devos, D.P. (2014). A bacterial tubulovesicular network. *Journal of Cell Science*, 127(2):277–280. doi:10.1242/jcs.137596.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana i Molera, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella i Cortes, G., Youssef, N., Zlatogursky, V., and Zhang, Q. (2018). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*. doi:10.1111/jeu.12691.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., Mccourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., and Taylor, M.F.J.R. (2005). The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *The Journal of Eukaryotic Microbiology*, 52(5):399–451. doi: 10.1111/j.1550-7408.2005.00053.x.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C.L., Smirnov, A., and Spiegel, F.W. (2012). The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology*, 59(5):429–514. doi:10.1111/j.1550-7408.2012.00644.x.

Adung'a, V.O., Gadelha, C., and Field, M.C. (2013). Proteomic analysis of clathrin interactions in trypanosomes reveals dynamic evolution of endocytosis. *Traffic*, 14(4):440–57. doi:10.1111/tra.12040.

Advani, R.J., Yang, B., Prekeris, R., Lee, K.C., Klumperman, J., and Scheller, R.H. (1999). Vamp-7 Mediates Vesicular Transport from Endosomes to Lysosomes. *The Journal of Cell Biology*, 146(4):765–776. doi: 10.1083/jcb.146.4.765.

Al-Dosari, M. and Alkuraya, F.S. (2009). A novel missense mutation in SCYL1BP1 produces geroderma osteodysplastica phenotype indistinguishable from that caused by nullimorphic mutations. *Am. J. Med. Genet. Part A*, 149(10):2093–2098. doi:10.1002/ajmg.a.32996.

Allan, B.B., Moyer, B.D., and Balch, W.E. (2000). Rab1 Recruitment of p115 into a cis-SNARE Complex: Programming Budding COPII Vesicles for Fusion. *Science*, 289(July):444–448.

Alonso, A. and Portillo, F.G.d. (2004). INTERNATIONAL MICROBIOLOGY (2004) 7:181–191 www.im.microbios.org. *International Microbiology*, 7:181–191.

Andrews, N.W. (2000). Regulated secretion of conventional lysosomes. *Trends in Cell Biology*, 10(8):316–321. doi:10.1016/S0962-8924(00)01794-3.

Antonin, W., Fasshauer, D., Becker, S., Jahn, R., and Schneider, T.R. (2002). Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. *Nature Structural Biology*, 9(2):107–111. doi:10.1038/nsb746.

Antonin, W., Holroyd, C., Tikkanen, R., Höning, S., and Jahn, R. (2000). The R-SNARE Endobrevin/VAMP-8 Mediates Homotypic Fusion of Early Endosomes and Late Endosomes. *Molecular Biology of the Cell*, 11(10):3289–3298. doi:10.1091/mbc.11.10.3289.

Aoki, M., Ehara, M., Saito, Y., Yoshioka, H., Miyazaki, M., Saito, Y., Miyashita, A., Kawakami, S., Yamaguchi, T., Ohashi, A., Nunoura, T., Takai, K., and Imachi, H. (2014). A Long-Term Cultivation of an Anaerobic Methane-Oxidizing Microbial Community from Deep-Sea Methane-Seep Sediment Using a Continuous-Flow Bioreactor. *PLoS ONE*, 9(8):e105356. doi:10.1371/journal.pone.0105356.

Arasaki, K., Mikami, Y., Shames, S.R., Inoue, H., Wakana, Y., and Tagaya, M. (2017). *Legionella* effector Lpg1137 shuts down ER-mitochondria communication through cleavage of syntaxin 17. *Nature Communications*, 8(1):15406. doi:10.1038/ncomms15406.

Arasaki, K., Shimizu, H., Mogari, H., Nishida, N., Hirota, N., Furuno, A., Kudo, Y., Baba, M., Baba, N., Cheng, J., Fujimoto, T., Ishihara, N., Ortiz-Sandoval, C., Barlow, L.D., Raturi, A., Dohmae, N., Wakana, Y., Inoue, H., Tani, K., Dacks, J.B., Simmen, T., and Tagaya, M. (2015). A Role for the Ancient SNARE Syntaxin 17 in Regulating Mitochondrial Division. *Developmental Cell*, 32(3):304–317. doi:10.1016/j.devcel.2014.12.011.

Asensio, C.S., Sirkis, D.W., Maas, J.W., Egami, K., To, T.L., Brodsky, F.M., Shu, X., Cheng, Y., and Edwards, R.H. (2013). Self-assembly of VPS41 promotes sorting required for biogenesis of the regulated secretory pathway. *Developmental Cell*, 27(4):425–437. doi:10.1016/j.devcel.2013.10.007.

Ashworth, J.M. and Watts, D.J. (1970). Metabolism of the cellular slime mould *Dictyostelium discoideum* grown in axenic culture. *Biochemical Journal*, 119(2):175–182. doi:10.1042/bj1190175.

Assoum, M., Philippe, C., Isidor, B., Perrin, L., Makrythanasis, P., Sondheimer, N., Paris, C., Douglas, J., Lesca, G., Antonarakis, S., Hamamy, H., Jouan, T., Duffourd, Y., Auvin, S., Saunier, A., Begtrup, A., Nowak, C., Chatron, N., Ville, D., Mireskandari, K., Milani, P., Jonveaux, P., Lemeur, G., Milh, M., Amamoto, M., Kato, M., Nakashima, M., Miyake, N., Matsumoto, N., Masri, A., Thauvin-Robinet, C., Rivière, J.B., Faivre, L., and Thevenon, J. (2016). Autosomal-Recessive Mutations in AP3B2 , Adaptor-Related Protein Complex 3 Beta 2 Subunit, Cause an Early-Onset Epileptic Encephalopathy with Optic Atrophy. *The American Journal of Human Genetics*, 99(6):1368–1376. doi:10.1016/j.ajhg.2016.10.009.

Au, C.E., Hermo, L., Byrne, E., Smirle, J., Fazel, A., Simon, P.H.G., Kearney, R.E., Cameron, P.H., Smith, C.E., Vali, H., Fernandez-Rodriguez, J., Ma, K., Nilsson, T., and Bergeron, J.J.M. (2015). Expression, sorting, and segregation of Golgi proteins during germ cell differentiation in the testis. *Mol. Biol. Cell*, 26(22):4015–32. doi:10.1091/mbc.E14-12-1632.

Avidor-Reiss, T., Maer, A.M., Koundakjian, E., Polyanovsky, A., Keil, T., Subramaniam, S., and Zuker, C.S. (2004). Decoding Cilia Function: Defining Specialized Genes Required for Compartmentalized Cilia Biogenesis. *Cell*, 117(4):527–539. doi:10.1016/S0092-8674(04)00412-X.

Badolato, R. and Parolini, S. (2007). Novel insights from adaptor protein 3 complex deficiency. *Journal of Allergy and Clinical Immunology*, 120(4):735–741. doi:10.1016/j.jaci.2007.08.039.

Banfield, D.K., Lewis, M.J., and Pelham, H.R.B. (1995). A SNARE-like protein required for traffic through the Golgi complex. *Nature*, 375(6534):806–809. doi:10.1038/375806a0.

Barlow, L.D. and Dacks, J.B. (2018). Seeing the endomembrane system for the trees: Evolutionary analysis highlights the importance of plants as models for eukaryotic membrane-trafficking. *Seminars in Cell & Developmental Biology*, 80:142–152. doi:10.1016/j.semcdb.2017.09.027.

Barlow, L.D., Dacks, J.B., and Wideman, J.G. (2014). From all to (nearly) none: Tracing adaptin evolution in Fungi. *Cellular Logistics*, 4(1):e28114. doi:10.4161/cl.28114.

Barlow, L.D., Nývltová, E., Aguilar, M., Tachezy, J., and Dacks, J.B. (2018). A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biology*, 16(1). doi:10.1186/s12915-018-0492-9.

Barlowe, C., d'Enfert, C., and Schekman, R. (1993). Purification and Characterization of SAR1p, a Small GTP-binding Protein Required for Transport Vesicle Formation from the Endoplasmic Reticulum. *1993*, 268(2):873–879.

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Salama, N., Rexach, M.F., Ravazzola, M., Amherdt, M., and Schekman, R. (1994). COPII: A Membrane Coat Formed by Set Proteins That Drive Vesicle Budding from the Endoplasmic Reticulum. *Cell*, 77(June 17):895–907.

Barr, F.A., Nakamura, N., and Warren, G. (1998). Mapping the interaction between GRASP65 and GM130, components of a protein complex involved in the stacking of Golgi cisternae. *EMBO J.*, 17(12):3258–68. doi:10.1093/emboj/17.12.3258.

Bashline, L., Li, S., Zhu, X., and Gu, Y. (2015). The TWD40-2 protein and the AP2 complex cooperate in the clathrin-mediated endocytosis of cellulose synthase to regulate cellulose biosynthesis. *Proceedings of the National Academy of Sciences*, 112(41):12870–12875. doi:10.1073/pnas.1509292112.

Bassham, D.C., Sanderfoot, A.A., Kovaleva, V., Zheng, H., and Raikhel, N.V. (2000). AtVPS45 Complex Formation at the *trans* -Golgi Network. *Molecular Biology of the Cell*, 11(7):2251–2265. doi:10.1091/mbc.11.7.2251.

Baudhuin, P., Muller, M., Poole, B., and de Duve, C. (1965). Non-mitochondrial oxidizing particles (microbodies) in rat liver and kidney and kidney and in *Tetrahymena pyriformis*. *Biochemical and Biophysical Research Communications*, 20(1):53–59. doi:10.1016/0006-291x(65)90949-6.

Baumert, M., Maycox, P.R., Navone, F., De Camilli, P., and Jahn, R. (1989). Synaptobrevin: An integral membrane protein of 18,000 daltons present in small synaptic vesicles of rat brain. *The EMBO Journal*, 8(2):379–384. doi:10.1002/j.1460-2075.1989.tb03388.x.

Becherer, K.A. and Jonest, E.W. (1996). Novel Syntaxin Homologue, Pepl2p, Required for the Sorting of Lumenal Hydrolases to the Lysosome-like Vacuole in Yeast. *Molecular Biology of the Cell*, 7:16.

Beck, R., Sun, Z., Adolf, F., Rutz, C., Bassler, J., Wild, K., Sinning, I., Hurt, E., Brugger, B., Bethune, J., and Wieland, F. (2008). Membrane curvature induced by Arf1-GTP is essential for vesicle formation. *Proceedings of the National Academy of Sciences*, 105(33):11731–11736. doi:10.1073/pnas.0805182105.

Behnia, R., Barr, F.A., Flanagan, J.J., Barlowe, C., Munro, S., and Barr, A. (2007). The yeast orthologue of GRASP65 forms a complex with a coiled-coil protein that contributes to ER to Golgi traffic. *J. Cell Biol.*, 176(3):255–261.

Bennett, M.K., Garcia-Arrarás, J., Elferink, L.A., Peterson, K., Fleming, A.M., Hazuka, C.D., and Scheller, R.H. (1993). The syntaxin family of vesicular transport receptors. *Cell*, 74(5):863–873. doi:10.1016/0092-8674(93)90466-4.

Bennett, N., Letourneur, F., Ragno, M., and Louwagie, M. (2008). Sorting of the v-SNARE VAMP7 in *Dictyostelium discoideum*: A role for more than one Adaptor Protein (AP) complex. *Experimental Cell Research*, 314(15):2822–2833. doi:10.1016/j.yexcr.2008.06.019.

Besteiro, S., Coombs, G.H., and Mottram, J.C. (2006). The SNARE protein family of *Leishmania major*. *BMC Genomics*, 7(1):250. doi:10.1186/1471-2164-7-250.

Bethani, I., Lang, T., Geumann, U., Sieber, J.J., Jahn, R., and Rizzoli, S.O. (2007). The specificity of

SNARE pairing in biological membranes is mediated by both proof-reading and spatial segregation. *The EMBO Journal*, 26(17):3981–3992. doi:10.1038/sj.emboj.7601820.

Bethani, I., Werner, A., Kadian, C., Geumann, U., Jahn, R., and Rizzoli, S.O. (2009). Endosomal Fusion upon SNARE Knockdown is Maintained by Residual SNARE Activity and Enhanced Docking. *Traffic*, 10(10):1543–1559. doi:10.1111/j.1600-0854.2009.00959.x.

Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C.J., and Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology & Evolution*, 2(10):1556–1562. doi:10.1038/s41559-018-0644-x.

Bexiga, M.G. and Simpson, J.C. (2013). Human diseases associated with form and function of the Golgi complex. *Int. J. Mol. Sci.*, 14(9):18670–81. doi:10.3390/ijms140918670.

Beyer, A. (1997). Sequence analysis of the AAA protein family. *Protein Science*, 6:2043–2058. doi:10.1002/pro.5560061001.

Beznoussenko, G.V., Dolgikh, V.V., Seliverstova, E.V., Semenov, P.B., Tokarev, Y.S., Trucco, A., Micaroni, M., Di Giandomenico, D., Auinger, P., Senderskiy, I.V., Skarlato, S.O., Snigirevskaya, E.S., Komissarchik, Y.Y., Pavelka, M., a De Matteis, M., Luini, A., Sokolova, Y.Y., and a Mironov, A. (2007). Analogs of the Golgi complex in microsporidia: Structure and avesicular mechanisms of function. *J. Cell Sci.*, 120(Pt 7):1288–98. doi:10.1242/jcs.03402.

Bharucha, N., Liu, Y., Papanikou, E., McMahon, C., Esaki, M., Jeffrey, P.D., Hughson, F.M., and Glick, B.S. (2013). Sec16 influences transitional ER sites by regulating rather than organizing COPII. *Mol. Biol. Cell*, 24(21):3406–3419. doi:10.1091/mbc.E13-04-0185.

Blanco-Melo, D., Venkatesh, S., and Bieniasz, P.D. (2016). Origins and Evolution of tetherin , an Orphan Antiviral Gene. *Cell Host & Microbe*, 20(2):189–201. doi:10.1016/j.chom.2016.06.007.

Blobel, G. (1980). Intracellular protein topogenesis. *Cell Biology*, page 5.

Bock, J.B., Klumperman, J., Davanger, S., and Scheller, R.H. (1997). Syntaxin 6 functions in trans-Golgi network vesicle trafficking. *Molecular Biology of the Cell*, 8(7):1261–1271. doi:10.1091/mbc.8.7.1261.

Bock, J.B., Matern, H.T., Peden, A.A., and Scheller, R.H. (2001). A genomic perspective on membrane compartment organization. *Nature*, 409(6822):839–841. doi:10.1038/35057024.

Bogdanovic, A., Bennett, N., Kieffer, S., Louwagie, M., Morio, T., Garin, J., Satre, M., and Bruckert, F. (2002). Syntaxin 7, syntaxin 8, Vti1 and VAMP7 (vesicle-associated membrane protein 7) form an active SNARE complex for early macropinocytic compartment fusion in *Dictyostelium discoideum*. *Biochemical Journal*, 368(1):29–39. doi:10.1042/bj20020845.

Bogdanovic, A., Bruckert, F., Morio, T., and Satre, M. (2000). A Syntaxin 7 Homologue Is Present in *Dictyostelium discoideum* Endosomes and Controls Their Homotypic Fusion. *Journal of Biological Chemistry*, 275(47):36691–36697. doi:10.1074/jbc.M006710200.

Bonifacino, J.S. (2004). Insights into the Biogenesis of Lysosome-Related Organelles from the Study of the Hermansky-Pudlak Syndrome. *Annals of the New York Academy of Sciences*, 1038(1):103–114. doi:10.1196/annals.1315.018.

Bonifacino, J.S. (2014). Vesicular transport earns a Nobel. *Trends in Cell Biology*, 24(1):3–5. doi:10.1016/j.tcb.2013.11.001.

Bonifacino, J.S. and Glick, B.S. (2004). The Mechanisms of Vesicle Budding and Fusion. *Cell*, 116(2):153–166. doi:10.1016/S0092-8674(03)01079-1.

Borgese, N., Colombo, S., and Pedrazzini, E. (2003). The tale of tail-anchored proteins: Coming from the cytosol and looking for a membrane. *The Journal of Cell Biology*, 161(6):1013–1019. doi:10.1083/jcb.200303069.

Boukhris, A., Stevanin, G., Feki, I., Denis, E., Elleuch, N., Miladi, M.I., Truchetto, J., Denora, P., Belal, S., Mhiri, C., and Brice, A. (2008). Hereditary spastic paraplegia with mental impairment and thin corpus callosum

in Tunisia: SPG11, SPG15, and further genetic heterogeneity. *Archives of Neurology*, 65(3):393–402. doi:10.1001/archneur.65.3.393.

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., Berlin, A.M., Campbell, M.S., Barrell, D., Martin, K.J., Mulley, J.F., Ravi, V., Lee, A.P., Nakamura, T., Chalopin, D., Fan, S., Wcisel, D., Caestro, C., Sydes, J., Beaudry, F.E., Sun, Y., Hertel, J., Beam, M.J., Fasold, M., Ishiyama, M., Johnson, J., Kehr, S., Lara, M., Letaw, J.H., Litman, G.W., Litman, R.T., Mikami, M., Ota, T., Saha, N.R., Williams, L., Stadler, P.F., Wang, H., Taylor, J.S., Fontenot, Q., Ferrara, A., Searle, S.M., Aken, B., Yandell, M., Schneider, I., Yoder, J.A., Volff, J.N., Meyer, A., Amemiya, C.T., Venkatesh, B., Holland, P.W., Guiguen, Y., Bobe, J., Shubin, N.H., Di Palma, F., Alföldi, J., Lindblad-Toh, K., and Postlethwait, J.H. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.*, 48(4):427–437. doi:10.1038/ng.3526.

Brandhorst, D., Zwilling, D., Rizzoli, S.O., Lippert, U., Lang, T., and Jahn, R. (2006). Homotypic fusion of early endosomes: SNAREs do not determine fusion specificity. *Proceedings of the National Academy of Sciences*, 103(8):2701–2706. doi:10.1073/pnas.0511138103.

Braun, E.L. (2018). An evolutionary model motivated by physicochemical properties of amino acids reveals variation among proteins. *Bioinformatics*, 34(13):i350–i356. doi:10.1093/bioinformatics/bty261.

Brennwald, P. (1994). Sec9 is a SNAP-25-like component of a yeast SNARE complex that may be the effector of Sec4 function in exocytosis. *Cell*, 79(2):245–258. doi:10.1016/0092-8674(94)90194-5.

Brickner, J.H., Blanchette, J.M., Sipos, G., and Fuller, R.S. (2001). The Tlg SNARE complex is required for TGN homotypic fusion. *The Journal of Cell Biology*, 155(6):969–978. doi:10.1083/jcb.200104093.

Brigance, W.T., Barlowe, C., and Graham, T.R. (2000). Organization of the Yeast Golgi Complex into at Least Four Funtionally Distinct Compartments. *Molecular Biology of the Cell*, 11(1):171–182. doi:10.1091/mbc.11.1.171.

Briguglio, J.S., Kumar, S., and Turkewitz, A.P. (2013). Lysosomal sorting receptors are essential for secretory granule biogenesis in Tetrahymena. *J. Cell Biol.*, 203(3):537–550. doi:10.1083/jcb.201305086.

Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.I., Hashimoto, T., Simpson, A.G.B., and Roger, A.J. (2018). Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biology and Evolution*, 10(2):427–433. doi:10.1093/gbe/evy014.

Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G.B., and Roger, A.J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. R. Soc. B Biol. Sci.*, 280(1769):20131755. doi:10.1098/rspb.2013.1755.

Brown, M.W., Spiegel, F.W., and Silberman, J.D. (2009). Phylogeny of the "Forgotten" Cellular Slime Mold, Fonticula alba, Reveals a Key Evolutionary Branch within Opisthokonta. *Molecular Biology and Evolution*, 26(12):2699–2709. doi:10.1093/molbev/msp185.

Burgos, P.V., Mardones, G.A., Rojas, A.L., daSilva, L.L.P., Prabhu, Y., Hurley, J.H., and Bonifacino, J.S. (2010). Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Developmental Cell*, 18(3):425–436. doi:10.1016/j.devcel.2010.01.015.

Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B: Biological Sciences*, 283(1823):20152802. doi:10.1098/rspb.2015.2802.

Burki, F., Shalchian-Tabrizi, K., and Pawlowski, J. (2008). Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biology Letters*, 4(4):366–369. doi:10.1098/rsbl.2008.0224.

Burns, J., Pittis, A., and Kim, E. (2018). Gene based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat. Ecol Evol*, accepted. doi:10.1038/s41559-018-0477-7.

Butterfield, N.J. (2015). Early evolution of the Eukaryota. *Palaeontology*, 58(1):5–17. doi:10.1111/pala.12139.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1):421. doi:10.1186/1471-2105-10-421.

Capella-Gutiérrez, S., Marcet-Houben, M., and Gabaldón, T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC biology*, 10:47. doi:10.1186/1741-7007-10-47.

Carter, C., Pan, S., Zouhar, J., Avila, E.L., Girke, T., and Raikhel, N.V. (2004). The Vegetative Vacuole Proteome of *Arabidopsis thaliana* Reveals Predicted and Unexpected Proteins. *The Plant Cell*, 16(12):3285–3303. doi:10.1105/tpc.104.027078.

Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J.B., and Bettencourt-Dias, M. (2011). Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol.*, 194(2):165–175. doi:10.1083/jcb.201011152.

Carvalho-Santos, Z., Machado, P., Branco, P., Tavares-Cadete, F., Rodrigues-Martins, A., Pereira-Leal, J.B., and Bettencourt-Dias, M. (2010). Stepwise evolution of the centriole-assembly pathway. *Journal of Cell Science*, 123(9):1414–1426. doi:10.1242/jcs.064931.

Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. *Nature*, 326(6111):332–333. doi:10.1038/326332a0.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology*, 52(2):297–354. doi:10.1099/00207713-52-2-297.

Cavalier-Smith, T. (2003). Protist phylogeny and the high-level classification of Protozoa. *European Journal of Protistology*, 39(4):338–348. doi:10.1078/0932-4739-00002.

Cavalier-Smith, T. (2009). Predation and eukaryote cell origins: A coevolutionary perspective. *The International Journal of Biochemistry & Cell Biology*, 41(2):307–322. doi:10.1016/j.biocel.2008.10.002.

Chavez, L.A., Balamuth, W., and Gong, T. (1986). A Light and Electron Microscopical Study of a New, Polymorphic Free-Living Amoeba, *Phreatamoeba balamuthi* n. g., n. sp. *J. Protozool.*, 33(3):397–404. doi:10.1111/j.1550-7408.1986.tb05630.x.

Cheever, M.L., Sato, T.K., de Beer, T., Kutateladze, T.G., Emr, S.D., and Overduin, M. (2001). Phox domain interaction with PtdIns(3)P targets the Vam7 t-SNARE to vacuole membranes. *Nature Cell Biology*, 3(7):613–618. doi:10.1038/35083000.

Cheung, P.y.P., Limouse, C., Mabuchi, H., and Pfeffer, S.R. (2015). Protein flexibility is required for vesicle tethering at the Golgi. *eLife*, 4:e12790. doi:10.7554/eLife.12790.

Cheung, P.y.P. and Pfeffer, S.R. (2016). Transport Vesicle Tethering at the Trans Golgi Network: Coiled Coil Proteins in Action. *Front. Cell Dev. Biol.*, 4(March):1–10. doi:10.3389/fcell.2016.00018.

Chi, P.B., Kim, D., Lai, J.K., Bykova, N., Weber, C.C., Kubelka, J., and Liberles, D.A. (2018). A new parameter-rich structure-aware mechanistic model for amino acid substitution during evolution. *Proteins: Structure, Function, and Bioinformatics*, 86(2):218–228. doi:10.1002/prot.25429.

Chi, R.J., Liu, J., West, M., Wang, J., Odorizzi, G., and Burd, C.G. (2014). Fission of SNX-BAR–coated endosomal retrograde transport carriers is promoted by the dynamin-related protein Vps1. *The Journal of Cell Biology*, 204(5):793–806. doi:10.1083/jcb.201309084.

Chiu, C.F., Ghanekar, Y., Frost, L., Diao, A., Morrison, D., McKenzie, E., and Lowe, M. (2008). ZFPL1, a novel ring finger protein required for cis-Golgi integrity and efficient ER-to-Golgi transport. *EMBO J.*, 27(7):934–47. doi:10.1038/emboj.2008.40.

Chluba-de Tapia, J., de Tapia, M., Jäggin, V., and Eberle, a.N. (1997). Cloning of a human multispanning membrane protein cDNA: Evidence for a new protein family. *Gene*, 197:195–204.

Clotworthy, M. (2005). *Studies in Dictyostelium Discoideum on the Roll of Membranes in Amoeboid Cell Movement*. Ph.D. thesis, University of Cambridge, Cambridge, England, UK.

Cluett, E.B. and Brown, W.J. (1992). Adhesion of Golgi cisternae by proteinaceous interactions: Intercisternal bridges as putative adhesive structures. *J. Cell Sci.*, 103:773–84.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423. doi:10.1093/bioinformatics/btp163.

Collins, N.C., Thordal-Christensen, H., Lipka, V., Bau, S., Kombrink, E., Qiu, J.L., Hückelhoven, R., Stein, M., Freialdenhoven, A., Somerville, S.C., and Schulze-Lefert, P. (2003). SNARE-protein-mediated disease resistance at the plant cell wall. *Nature*, 425(6961):973–977. doi:10.1038/nature02076.

Conant, G.C. and Wolfe, K.H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, 9(12):938–950. doi:10.1038/nrg2482.

Connerly, P.L., Esaki, M., a Montegna, E., Strongin, D.E., Levi, S., Soderholm, J., and Glick, B.S. (2005). Sec16 is a determinant of transitional ER organization. *Curr. Biol.*, 15(16):1439–47. doi:10.1016/j.cub.2005.06.065.

Corradi, N., Pombert, J.F., Farinelli, L., Didier, E.S., and Keeling, P.J. (2010). The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nature Communications*, 1:77. doi:10.1038/ncomms1082.

Cowles, C.R., Odorizzi, G., Payne, G.S., and Emr, S.D. (1997). The AP-3 Adaptor Complex Is Essential for Cargo-Selective Transport to the Yeast Vacuole. *Cell*, 91(1):109–118. doi:10.1016/S0092-8674(01)80013-1.

Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., and Embley, T.M. (2008). The archaebacterial origin of eukaryotes. *Proceedings of the National Academy of Sciences*, 105(51):20356–20361. doi:10.1073/pnas.0810647105.

Dacks, J.B., Davis, L.A.M., Sjögren, A.M., Andersson, J.O., Roger, A.J., and Doolittle, W.F. (2003). Evidence for Golgi bodies in proposed 'Golgi-lacking' lineages. *Proc. Biol. Sci.*, 270 Suppl:S168–71. doi:10.1098/rsbl.2003.0058.

Dacks, J.B. and Doolittle, W. (2004). Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. *Molecular and Biochemical Parasitology*, 136(2):123–136. doi:10.1016/j.molbiopara.2004.02.014.

Dacks, J.B. and Doolittle, W.F. (2001). Reconstructing/deconstructing the earliest eukaryotes: How comparative genomics can help. *Cell*, 107(4):419–25.

Dacks, J.B. and Doolittle, W.F. (2002). Novel syntaxin gene sequences from *Giardia*, *Trypanosoma* and algae: Implications for the ancientevolution of the eukaryotic endomembrane system. *Journal of Cell Science*, 115:1635–1642.

Dacks, J.B. and Field, M.C. (2007). Evolution of the eukaryotic membrane-trafficking system: Origin, tempo and mode. *Journal of Cell Science*, 120(17):2977–2985. doi:10.1242/jcs.013250.

Dacks, J.B., Poon, P.P., and Field, M.C. (2008). Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proceedings of the National Academy of Sciences*, 105(2):588–593. doi:10.1073/pnas.0707318105.

Dacks, J.B. and Robinson, M.S. (2017). Outerwear through the ages: Evolutionary cell biology of vesicle coats. *Current Opinion in Cell Biology*, 47:108–116. doi:10.1016/j.ceb.2017.04.001.

Davidov, Y. and Jurkevitch, E. (2009). Predation between prokaryotes and the origin of eukaryotes. *BioEssays*, 31(7):748–757. doi:10.1002/bies.200900018.

Day, K.J., Casler, J.C., and Glick, B.S. (2018). Budding Yeast Has a Minimal Endomembrane System. *Developmental Cell*, 44(1):56–72.e4. doi:10.1016/j.devcel.2017.12.014.

Day, K.J., Staehelin, L.A., and Glick, B.S. (2013). A three-stage model of Golgi structure and function. *Histochemistry and Cell Biology*, 140(3):239–249. doi:10.1007/s00418-013-1128-3.

de Duve, C. (1969). EVOLUTION OF THE PEROXISOME. *Annals of the New York Academy of Sciences*, 168(2 The Nature an):369–381. doi:10.1111/j.1749-6632.1969.tb43124.x.

de Duve, C. (1982). PEROXISOMES AND RELATED PARTICLES IN HISTORICAL PERSPECTIVE. *Annals of the New York Academy of Sciences*, 386(1 Peroxisomes a):1–4. doi:10.1111/j.1749-6632.1982.tb21402.x.

de Duve, C. (2007). The origin of eukaryotes: A reappraisal. *Nature Reviews Genetics*, 8(5):395–403. doi:10.1038/nrg2071.

de Duve, C. and Baudhuin, P. (1966). Peroxisomes (microbodies and related particles). *Physiological Reviews*, 46(2):323–357. doi:10.1152/physrev.1966.46.2.323.

de Duve, C. and Wattiaux, R. (1966). Functions of Lysosomes. *Annual Review of Physiology*, 28(1):435–492. doi:10.1146/annurev.ph.28.030166.002251.

Delevoye, C., Nilges, M., Dautry-Varsat, A., and Subtil, A. (2004). Conservation of the Biochemical Properties of IncA from *Chlamydia trachomatis* and *Chlamydia caviae*: OLIGOMERIZATION OF IncA MEDIATES INTERACTION BETWEEN FACING MEMBRANES. *Journal of Biological Chemistry*, 279(45):46896–46906. doi:10.1074/jbc.M407227200.

Dell'Angelica, E.C. (2009). AP-3-dependent trafficking and disease: The first decade. *Current Opinion in Cell Biology*, 21(4):552–559. doi:10.1016/j.ceb.2009.04.014.

Dell'Angelica, E.C., Klumperman, J., Stoorvogel, W., and Bonifacino, J.S. (1998). Association of the AP-3 adaptor complex with clathrin. *Science (New York, N.Y.)*, 280(5362):431–434. doi:10.1126/science.280.5362.431.

Dell'Angelica, E.C., Mullins, C., and Bonifacino, J.S. (1999). AP-4, a novel protein complex related to clathrin adaptors. *The Journal of Biological Chemistry*, 274(11):7278–7285. doi:10.1074/jbc.274.11.7278.

Derelle, R. and Lang, B.F. (2012). Rooting the Eukaryotic Tree with Mitochondrial and Bacterial Proteins. *Molecular Biology and Evolution*, 29(4):1277–1289. doi:10.1093/molbev/msr295.

Derelle, R., López-García, P., Timpano, H., and Moreira, D. (2016). A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). *Molecular Biology and Evolution*, 33(11):2890–2898. doi:10.1093/molbev/msw168.

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., Lang, B.F., and Eliáš, M. (2015). Bacterial proteins pinpoint a single eukaryotic root. *Proceedings of the National Academy of Sciences*, 112(7):E693–E699. doi:10.1073/pnas.1420657112.

Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M.C., and Pereira-Leal, J.B. (2011). Thousands of Rab GTPases for the Cell Biologist. *PLoS Computational Biology*, 7(10):e1002217. doi:10.1371/journal.pcbi.1002217.

Dietrich, L.E., Boeddinghaus, C., LaGrassa, T.J., and Ungermann, C. (2003). Control of eukaryotic membrane fusion by N-terminal domains of SNARE proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1641(2-3):111–119. doi:10.1016/S0167-4889(03)00094-6.

Donaldson, J.G. and Cassel, D. (1992). ADP-ribosylation factor, a small GTP-binding protein, is required for binding of the coatomer protein fl-COP to Golgi membranes. *Proc. Natl. Acad. Sci. USA*, page 5.

Drakakaki, G., van de Ven, W., Pan, S., Miao, Y., Wang, J., Keinath, N.F., Weatherly, B., Jiang, L., Schumacher, K., Hicks, G., and Raikhel, N. (2012). Isolation and proteomic analysis of the SYP61 compartment reveal its role in exocytic trafficking in *Arabidopsis*. *Cell Research*, 22(2):413–424. doi:10.1038/cr.2011.129.

Drin, G., Casella, J.F., Gautier, R., Boehmer, T., Schwartz, T.U., and Antonny, B. (2007). A general amphipathic α-helical motif for sensing membrane curvature. *Nature Structural & Molecular Biology*, 14(2):138–146. doi:10.1038/nsmb1194.

Dulubova, I. (2002). How Tlg2p/syntaxin 16 'snares' Vps45. *The EMBO Journal*, 21(14):3620–3631. doi:10.1093/emboj/cdf381.

Dunn, C.W. and Munro, C. (2016). Comparative genomics and the diversity of life. *Zoologica Scripta*, 45:5–13. doi:10.1111/zsc.12211.

Ebine, K., Okatani, Y., Uemura, T., Goh, T., Shoda, K., Niihama, M., Morita, M.T., Spitzer, C., Otegui, M.S., Nakano, A., and Ueda, T. (2008). A SNARE Complex Unique to Seed Plants Is Required for Protein Storage Vacuole Biogenesis and Seed Development of *Arabidopsis thaliana*. *The Plant Cell*, 20(11):3006–3021. doi:10.1105/tpc.107.057711.

Eddy, S.R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9):755–763. doi:10.1093/bioinformatics/14.9.755.

Edgar, J.R., Manna, P.T., Nishimura, S., Banting, G., and Robinson, M.S. (2016). Tetherin is an exosomal tether. *eLife*, 5:e17180. doi:10.7554/eLife.17180.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797. doi:10.1093/nar/gkh340.

Efimov, A., Kharitonov, A., Efimova, N., Loncarek, J., Miller, P.M., Andreyeva, N., Gleeson, P., Galjart, N., Maia, A.R.R., McLeod, I.X., Yates, J.R., Maiato, H., Khodjakov, A., Akhmanova, A., and Kaverina, I. (2007). Asymmetric CLASP-Dependent Nucleation of Noncentrosomal Microtubules at the trans-Golgi Network. *Dev. Cell*, 12(6):917–930. doi:10.1016/j.devcel.2007.04.002.

El Kasmi, F., Krause, C., Hiller, U., Stierhof, Y.D., Mayer, U., Conner, L., Kong, L., Reichardt, I., Sanderfoot, A.A., and Jürgens, G. (2013). SNARE complexes of different composition jointly mediate membrane fusion in *Arabidopsis* cytokinesis. *Molecular Biology of the Cell*, 24(10):1593–1601. doi:10.1091/mbc.e13-02-0074.

Elferink, L.A., TrimbleS, W.S., and Scheller, R.H. (1989). Two Vesicle-associated Membrane Protein Genes Are Differentially Expressed in the Rat Central Nervous System. *The Journal of Biological Chemistry*, 264(July 5):11061–11064.

Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M.C., and Dacks, J.B. (2012). Sculpting the endomembrane system in deep time: High resolution phylogenetics of Rab GTPases. *Journal of Cell Science*, 125(10):2500–2508. doi:10.1242/jcs.101378.

Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J. (2014). On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Cold Spring Harbor Perspectives in Biology*, 6(8):a016139–a016139. doi:10.1101/cshperspect.a016139.

Fasshauer, D., Antonin, W., Margittai, M., Pabst, S., and Jahn, R. (1999). Mixed and Non-cognate SNARE Complexes: CHARACTERIZATION OF ASSEMBLY AND BIOPHYSICAL PROPERTIES. *Journal of Biological Chemistry*, 274(22):15440–15446. doi:10.1074/jbc.274.22.15440.

Fasshauer, D., Sutton, R.B., Brunger, A.T., and Jahn, R. (1998). Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proceedings of the National Academy of Sciences*, 95(26):15781–15786. doi:10.1073/pnas.95.26.15781.

Fast, N.M., Xue, L., Bingham, S., and Keeling, P.J. (2002). Re-examining Alveolate Evolution Using Multiple Protein Molecular Phylogenies. *The Journal of Eukaryotic Microbiology*, 49(1):30–37. doi:10.1111/j.1550-7408.2002.tb00336.x.

Fedorov, A. and Hartman, H. (2004). What does the microsporidian E. cuniculi tell us about the origin of the eukaryotic cell? *Journal of Molecular Evolution*, 59(5):695–702. doi:10.1007/s00239-003-0085-1.

Felder, S., Miller, K., Moehren, G., Ullrich, A., Schlessinger, J., and Hopkins, C. (1990). Kinase activity controls the sorting of the epidermal growth factor receptor within the multivesicular body. *Cell*, 61(4):623–634. doi:10.1016/0092-8674(90)90474-S.

Feldmann, J., Callebaut, I., Raposo, G., Certain, S., Bacq, D., Dumont, C., Lambert, N., Ouachée-Chardin, M., Chedeville, G., Tamary, H., Minard-Colin, V., Vilmer, E., Blanche, S., Le Deist, F., Fischer, A., and de Saint Basile, G. (2003). Munc13-4 Is Essential for Cytolytic Granules Fusion and Is Mutated in a Form of Familial Hemophagocytic Lymphohistiocytosis (FHL3). *Cell*, 115(4):461–473. doi:10.1016/S0092-8674(03)00855-9.

Feraru, E., Paciorek, T., Feraru, M.I., Zwiewka, M., De Groodt, R., De Rycke, R., Kleine-Vehn, J., and Friml, J. (2010). The AP-3 Adaptin Mediates the

Biogenesis and Function of Lytic Vacuoles in *Arabidopsis*. *The Plant Cell*, 22(8):2812–2824. doi:10.1105/tpc.110.075424.

Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., and Brangwynne, C.P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, 165(7):1686–1697. doi:10.1016/j.cell.2016.04.047.

Fernandez, I., Ubach, J., Dulubova, I., Zhang, X., Südhof, T.C., and Rizo, J. (1998). Three-Dimensional Structure of an Evolutionarily Conserved N-Terminal Domain of Syntaxin 1A. *Cell*, 94(6):841–849. doi:10.1016/S0092-8674(00)81742-0.

Field, H.I., Coulson, R.M.R., and Field, M.C. (2013). An automated graphics tool for comparative genomics: The Coulson plot generator. *BMC bioinformatics*, 14:141. doi:10.1186/1471-2105-14-141.

Field, M.C. and Carrington, M. (2009). The trypanosome flagellar pocket. *Nature Reviews Microbiology*, 7(11):775–786. doi:10.1038/nrmicro2221.

Field, M.C., Gabernet-Castello, C., and Dacks, J.B. (2007). Reconstructing the evolution of the endocytic system: Insights from genomics and molecular cell biology. *Advances in Experimental Medicine and Biology*, 607:84–96. doi:10.1007/978-0-387-74021-8_7.

Field, M.C. and Rout, M.P. (2019). Pore timing: The evolutionary origins of the nucleus and nuclear pore complex. *F1000Research*, 8:369. doi:10.12688/f1000research.16402.1.

Filippini, F., Rossi, V., Galli, T., Budillon, A., D'Urso, M., and D'Esposito, M. (2001). Longins: A new evolutionary conserved VAMP family sharing a novel SNARE domain. *Trends in Biochemical Sciences*, 26(7):407–409. doi:10.1016/S0968-0004(01)01861-8.

Fitch, W.M. and Margoliash, E. (1967). Construction of Phylogenetic Trees. *Science*, 155(3760):279–284.

Fitzpatrick, D.A., Creevey, C.J., and McInerney, J.O. (2006). Genome Phylogenies Indicate a Meaningful α-Proteobacterial Phylogeny and Support a Grouping of the Mitochondria with the Rickettsiales. *Molecular Biology and Evolution*, 23(1):74–85. doi:10.1093/molbev/msj009.

Fölsch, H., Pypaert, M., Maday, S., Pelletier, L., and Mellman, I. (2003). The AP-1A and AP-1B clathrin adaptor complexes define biochemically and functionally distinct membrane domains. *The Journal of Cell Biology*, 163(2):351–362. doi:10.1083/jcb.200309020.

Fridmann-Sirkis, Y., Siniossoglou, S., and Pelham, H.R.B. (2004). TMF is a golgin that binds Rab6 and influences Golgi morphology. *BMC Cell Biol.*, 5:18. doi:10.1186/1471-2121-5-18.

Friedrich, M., Meier, D., Schuster, I., and Nellen, W. (2015). A Simple Retroelement Based Knock-Down System in *Dictyostelium*: Further Insights into RNA Interference Mechanisms. *PLOS ONE*, 10(6):e0131271. doi:10.1371/journal.pone.0131271.

Fritz-Laylin, L.K. and Cande, W.Z. (2010). Ancestral centriole and flagella proteins identified by analysis of *Naegleria* differentiation. *J. Cell Sci.*, 123(Pt 23):4024–31. doi:10.1242/jcs.077453.

Fuji, K., Shirakawa, M., Shimono, Y., Kunieda, T., Fukao, Y., Koumoto, Y., Takahashi, H., Hara-Nishimura, I., and Shimada, T. (2016). The Adaptor Complex AP-4 Regulates Vacuolar Protein Sorting at the trans-Golgi Network by Interacting with VACUOLAR SORTING RECEPTOR1. *Plant Physiology*, 170(1):211–219. doi:10.1104/pp.15.00869.

Fujiki, Y. and Lazarow, P.B. (1985). PosttranslationalImport of Fatty Acyl-CoA Oxidase and Catalase into Peroxisomes of Rat Liver in Vitro. *The Journal of Biological Chemistry*, 260(9):5603–5609.

Fujita, Y., Shirataki, H., Sakisaka, T., Asakura, T., Ohya, T., Kotani, H., Yokoyama, S., Nishioka, H., Matsuura, Y., Mizoguchi, A., Scheller, R.H., and Takai, Y. (1998). Tomosyn: A Syntaxin-1–Binding Protein that Forms a Novel Complex in the Neurotransmitter Release Process. *Neuron*, 20(5):905–915. doi:10.1016/S0896-6273(00)80472-9.

Fukasawa, M., Varlamov, O., Eng, W.S., Sollner, T.H., and Rothman, J.E. (2004). Localization and activity of the SNARE Ykt6 determined by its regulatory domain and palmitoylation. *Proceedings of the National Academy of Sciences*, 101(14):4815–4820. doi:10.1073/pnas.0401183101.

Fukuda, R., McNew, J.A., Weber, T., Parlati, F., Engel, T., Nickel, W., Rothman, J.E., and Söllner, T.H. (2000). Functional architecture of an intracellular membrane t-SNARE. *Nature*, 407(6801):198–202. doi:10.1038/35025084.

Furukawa, N. and Mima, J. (2014). Multiple and distinct strategies of yeast SNAREs to confer the specificity of membrane fusion. *Scientific Reports*, 4(1):4277. doi:10.1038/srep04277.

Futter, C.E., Pearse, A., Hewlett, L.J., and Hopkins, C.R. (1996). Multivesicular endosomes containing internalized EGF-EGF receptor complexes mature and then fuse directly with lysosomes. *The Journal of Cell Biology*, 132(6):1011–1023. doi:10.1083/jcb.132.6.1011.

Gabaldón, T. and Capella-Gutiérrez, S. (2010). Lack of phylogenetic support for a supposed actinobacterial origin of peroxisomes. *Gene*, 465(1-2):61–65. doi:10.1016/j.gene.2010.06.004.

Gabaldón, T., Ginger, M.L., and Michels, P.A. (2016). Peroxisomes in parasitic protists. *Molecular and Biochemical Parasitology*, 209(1-2):35–45. doi:10.1016/j.molbiopara.2016.02.005.

Gabaldón, T. and Pittis, A.A. (2015). Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie*, 119:262–268. doi:10.1016/j.biochi.2015.03.021.

Gabaldón, T., Snel, B., van Zimmeren, F., Hemrika, W., Tabak, H., and Huynen, M.A. (2006). Origin and evolution of the peroxisomal proteome. *Biology Direct*, page 14.

Gabernet-Castello, C., O'Reilly, A.J., Dacks, J.B., and Field, M.C. (2013). Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. *Molecular Biology of the Cell*, 24(10):1574–1583. doi:10.1091/mbc.e12-07-0557.

Gadeyne, A., Sánchez-Rodríguez, C., Vanneste, S., Di Rubbo, S., Zauber, H., Vanneste, K., Van Leene, J., De Winne, N., Eeckhout, D., Persiau, G., Van De Slijke, E., Cannoot, B., Vercruysse, L., Mayers, J.R., Adamowski, M., Kania, U., Ehrlich, M., Schweighofer, A., Ketelaar, T., Maere, S., Bednarek, S.Y., Friml, J., Gevaert, K., Witters, E., Russinova, E., Persson, S., De Jaeger, G., and Van Damme, D. (2014). The TPLATE Adaptor Complex Drives Clathrin-Mediated Endocytosis in Plants. *Cell*, 156(4):691–704. doi:10.1016/j.cell.2014.01.039.

Gascuel, O. (1997). BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695. doi:10.1093/oxfordjournals.molbev.a025808.

Gaudet, P., Pilcher, K.E., Fey, P., and Chisholm, R.L. (2007). Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nature Protocols*, 2(6):1317–1324. doi:10.1038/nprot.2007.179.

Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: A web server to screen sequences with specific α-helical properties. *Bioinformatics*, 24(18):2101–2102. doi:10.1093/bioinformatics/btn392.

Gawryluk, R.M.R., Tikhonenkov, D.V., Hehenberger, E., Husnik, F., Mylnikov, A.P., and Keeling, P.J. (2019). Non-photosynthetic predators are sister to red algae. *Nature*, 572(7768):240–243. doi:10.1038/s41586-019-1398-6.

Ghosh, S.K., Field, J., Frisardi, M., Rosenthal, B., Mai, Z., Rogers, R., and Samuelson, J. (1999). Chitinase secretion by encysting *Entamoeba invadens* and transfected *Entamoeba histolytica* trophozoites: Localization of secretory vesicles, endoplasmic reticulum, and Golgi apparatus. *Infect. Immun.*, 67(6):3073–3081.

Gibbons, I.R. and Rowe, A.J. (1965). Dynein: A Protein with Adenosine Triphosphatase Activity from Cilia. *Science*, 149(3682):424–426. doi:10.1126/science.149.3682.424.

Gill, E.E., Diaz-Triviño, S., Barberà, M.J., Silberman, J.D., Stechmann, A., Gaston, D., Tamas, I., and Roger, A.J. (2007). Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*: Novel mitochondrion-related organelles in *M. balamuthi*. *Molecular Microbiology*, 66(6):1306–1320. doi:10.1111/j.1365-2958.2007.05979.x.

Gillingham, A.K. (2017). At the ends of their tethers! How coiled-coil proteins capture vesicles at the Golgi. *Biochem. Soc. Trans.* doi:10.1042/BST20170188.

Gillingham, A.K. and Munro, S. (2007). The Small G Proteins of the Arf Family and Their Regulators. *Annual Review of Cell and Developmental Biology*, 23(1):579–611. doi:10.1146/annurev.cellbio.23.090506.123209.

Gillingham, A.K., Pfeifer, A.C., and Munro, S. (2002). CASP, the Alternatively Spliced Product of the Gene Encoding the CCAAT-Displacement Protein Transcription Factor, Is a Golgi Membrane Protein Related to Giantin. *Molecular Biology of the Cell*, 13(11):3761–3774. doi:10.1091/mbc.e02-06-0349.

Gillingham, A.K., Tong, A.H.Y., Boone, C., and Munro, S. (2004). The GTPase Arf1p and the ER to Golgi cargo receptor Erv14p cooperate to recruit the golgin Rud3p to the cis-Golgi. *J. Cell Biol.*, 167(2):281–292. doi:10.1083/jcb.200407088.

Gilson, P.R., Vergara, C.E., Kjer-Nielsen, L., Teasdale, R.D., Bacic, A., and Gleeson, P.A. (2004). Identification of a Golgi-localised GRIP domain protein from Arabidopsis thaliana. *Planta*, 219(6):1050–1056. doi:10.1007/s00425-004-1311-9.

Gissen, P., Johnson, C.A., Morgan, N.V., Stapelbroek, J.M., Forshew, T., Cooper, W.N., McKiernan, P.J., Klomp, L.W.J., Morris, A.A.M., Wraith, J.E., McClean, P., Lynch, S.A., Thompson, R.J., Lo, B., Quarrell, O.W., Di Rocco, M., Trembath, R.C., Mandel, H., Wali, S., Karet, F.E., Knisely, A.S., Houwen, R.H.J., Kelly, D.A., and Maher, E.R. (2004). Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis–renal dysfunction–cholestasis (ARC) syndrome. *Nature Genetics*, 36(4):400–404. doi:10.1038/ng1325.

Glick, B.S., Elston, T., and Oster, G. (1997). A cisternal maturation mechanism can explain the asymmetry of the Golgi stack. *FEBS Letters*, 414(2):177–181. doi:10.1016/S0014-5793(97)00984-8.

Glick, B.S. and Nakano, A. (2009). Membrane traffic within the Golgi apparatus. *Annu. Rev. Cell Dev. Biol.*, 25:113–32. doi:10.1146/annurev.cellbio.24.110707.175421.

Goldman, N., Anderson, J.P., and Rodrigo, A.G. (2000). Likelihood-Based Tests of Topologies in Phylogenetics. *Systematic Biology*, 49(4):652–670. doi:10.1080/106351500750049752.

Gonzalez, L.C., Weis, W.I., and Scheller, R.H. (2001). A Novel SNARE N-terminal Domain Revealed by the Crystal Structure of Sec22b. *Journal of Biological Chemistry*, 276(26):24203–24211. doi:10.1074/jbc.M101584200.

González Montoro, A., Bigliani, G., and Valdez Taubas, J. (2017). The shape of the transmembrane domain is a novel endocytosis signal for single-spanning membrane proteins. *Journal of Cell Science*, 130(22):3829–3838. doi:10.1242/jcs.202937.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1):D1178–D1186. doi:10.1093/nar/gkr944.

Gordon, D.E., Chia, J., Jayawardena, K., Antrobus, R., Bard, F., and Peden, A.A. (2017). VAMP3/Syb and YKT6 are required for the fusion of constitutive secretory carriers with the plasma membrane. *PLOS Genetics*, 13(4):e1006698. doi:10.1371/journal.pgen.1006698.

Gossing, M., Chidambaram, S., and Fischer von Mollard, G. (2013). Importance of the N-Terminal Domain of the Qb-SNARE Vti1p for Different Membrane Transport Steps in the Yeast Endosomal System. *PLoS ONE*, 8(6):e66304. doi:10.1371/journal.pone.0066304.

Gould, S.B., Garg, S.G., and Martin, W.F. (2016). Bacterial Vesicle Secretion and the Evolutionary Origin of the Eukaryotic Endomembrane System. *Trends in Microbiology*, 24(7):525–534. doi:10.1016/j.tim.2016.03.005.

Gray, M.W. (1992). The Endosymbiont Hypothesis Revisited. In *International Review of Cytology*, volume 141, pages 233–357. Elsevier. ISBN 978-0-12-364544-9. doi:10.1016/S0074-7696(08)62068-9.

Gray, M.W. and Cedergren, R.J. (1984). On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Research*, 12(14).

Gray, M.W. and Doolittle, W.F. (1982). Has the Endosymbiont Hypothesis Been Proven? *Microbiological Reviews*, 46:1–42.

Green, R.E., Braun, E.L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M.W., St John, J.A., Capella-Gutiérrez, S., Castoe, T.A., Kern, C., Fujita, M.K., Opazo, J.C., Jurka, J., Kojima, K.K., Caballero, J., Hubley, R.M., Smit, A.F., Platt, R.N., Lavoie, C.A., Ramakodi, M.P., Finger, J.W., Suh, A., Isberg, S.R., Miles, L., Chong, A.Y., Jaratlerdsiri, W., Gongora, J., Moran, C., Iriarte, A., McCormack, J., Burgess, S.C., Edwards, S.V., Lyons, E., Williams, C., Breen, M., Howard, J.T., Gresham, C.R., Peterson, D.G., Schmitz, J., Pollock, D.D., Haussler, D., Triplett, E.W., Zhang, G., Irie, N., Jarvis, E.D., Brochu, C.A., Schmidt, C.J., McCarthy, F.M., Faircloth, B.C., Hoffmann, F.G., Glenn, T.C., Gabaldón, T., Paten, B., and Ray, D.A. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science (80-.).*, 346(6215). doi:10.1126/science.1254449.

Griffiths, G., Hoflack, B., Simons, K., Mellman, I., and Kornfeld, S. (1988). The mannose 6-phosphate receptor and the biogenesis of lysosomes. *Cell*, 52(3):329–341. doi:10.1016/S0092-8674(88)80026-6.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., and Dubchak, I. (2012). The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research*, 40(Database issue):D26–32. doi:10.1093/nar/gkr947.

Grosshans, B.L., Ortiz, D., and Novick, P. (2006). Rabs and their effectors: Achieving specificity in membrane traffic. *Proceedings of the National Academy of Sciences*, 103(32):11821–11827. doi:10.1073/pnas.0601617103.

Guan, R., Zhao, Y., Zhang, H., Fan, G., Liu, X., Zhou, W., Shi, C., Wang, J., Liu, W., Liang, X., Fu, Y., Ma, K., Zhao, L., Zhang, F., Lu, Z., Lee, S.M.Y., Xu, X., Wang, J., Yang, H., Fu, C., Ge, S., and Chen, W. (2016). Draft genome of the living fossil Ginkgo biloba. *GigaScience*, 5(1). doi:10.1186/s13742-016-0154-1.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321. doi:10.1093/sysbio/syq010.

Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704. doi:10.1080/10635150390235520.

Guo, X., Mattera, R., Ren, X., Chen, Y., Retamal, C., González, A., and Bonifacino, J.S. (2013). The Adaptor Protein-1 μ1B Subunit Expands the Repertoire of Basolateral Sorting Signal Recognition in Epithelial Cells. *Developmental Cell*, 27(3):353–366. doi:10.1016/j.devcel.2013.10.006.

Guy, L. and Ettema, T.J. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology*, 19(12):580–587. doi:10.1016/j.tim.2011.09.002.

Haft, C.R., Sierra, M.d.l.L., Bafford, R., Lesniak, M.A., Barr, V.A., and Taylor, S.I. (2000). Human Orthologs of Yeast Vacuolar Protein Sorting Proteins Vps26, 29, and 35: Assembly into Multimeric Complexes. *Molecular Biology of the Cell*, 11(12):4105–4116. doi:10.1091/mbc.11.12.4105.

Hampl, V., Čepička, I., and Eliáš, M. (2019). Was the Mitochondrion Necessary to Start Eukaryogenesis? *Trends in Microbiology*, 27(2):96–104. doi:10.1016/j.tim.2018.10.005.

Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G.B., and Roger, A.J. (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proceedings of the National Academy of Sciences*, 106(10):3859–3864. doi:10.1073/pnas.0807880106.

Hanson, P.I., Roth, R., Morisaki, H., Jahn, R., and Heuser, J.E. (1997). Structure and Conformational Changes in NSF and Its Membrane Receptor Complexes Visualized by Quick-Freeze/Deep-Etch Electron Microscopy. *Cell*, 90(3):523–535. doi:10.1016/S0092-8674(00)80512-7.

Hasegawa, M. and Kishino, H. (1994). Accuracies of the Simple Methods for Estimating the Bootstrap Probability of a Maximum-Likelihood Tree. *Molecular Biology and Evolution*, 11:142–145. doi:10.1093/oxfordjournals.molbev.a040097.

Hatsuzawa, K., Hirose, H., Tani, K., Yamamoto, A., Scheller, R.H., and Tagaya, M. (2000). Syntaxin 18, a SNAP Receptor That Functions in the Endoplasmic Reticulum, Intermediate Compartment, and *cis* -Golgi Vesicle Trafficking. *Journal of Biological Chemistry*, 275(18):13713–13720. doi:10.1074/jbc.275.18.13713.

Hatsuzawa, K., Lang, T., Fasshauer, D., Bruns, D., and Jahn, R. (2003). The R-SNARE Motif of Tomosyn Forms SNARE Core Complexes with Syntaxin 1 and SNAP-25 and Down-regulates Exocytosis. *Journal of Biological Chemistry*, 278(33):31159–31166. doi:10.1074/jbc.M305500200.

Hawes, C. (2004). Cell biology of the plant Golgi apparatus. *New Phytologist*, 165(1):29–44. doi:10.1111/j.1469-8137.2004.01218.x.

Hay, J.C., Klumperman, J., Oorschot, V., Steegmaier, M., Kuo, C.S., and Scheller, R.H. (1998). Localization, Dynamics, and Protein Interactions Reveal Distinct Roles for ER and Golgi SNAREs. *The Journal of Cell Biology*, 141(7):1489–1502. doi:10.1083/jcb.141.7.1489.

He, D., Fiz-Palacios, O., Fu, C.J., Fehling, J., Tsai, C.C., and Baldauf, S.L. (2014). An Alternative Root for the Eukaryote Tree of Life. *Current Biology*, 24(4):465–470. doi:10.1016/j.cub.2014.01.036.

Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., and Keeling, P.J. (2017). Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Current Biology*, 27(13):2043–2050.e6. doi:10.1016/j.cub.2017.06.006.

Heimerl, T., Flechsler, J., Pickl, C., Heinz, V., Salecker, B., Zweck, J., Wanner, G., Geimer, S., Samson, R.Y., Bell, S.D., Huber, H., Wirth, R., Wurch, L., Podar, M., and Rachel, R. (2017). A Complex Endomembrane System in the Archaeon *Ignicoccus hospitalis* Tapped by Nanoarchaeum equitans. *Frontiers in Microbiology*, 8. doi:10.3389/fmicb.2017.01072.

Heinz, E., Williams, T.A., Nakjang, S., Noël, C.J., Swan, D.C., Goldberg, A.V., Harris, S.R., Weinmaier, T., Markert, S., Becher, D., Bernhardt, J., Dagan, T., Hacker, C., Lucocq, J.M., Schweder, T., Rattei, T., Hall, N., Hirt, R.P., and Embley, T.M. (2012). The genome of the obligate intracellular parasite *Trachipleistophora hominis*: New insights into microsporidian genome dynamics and reductive evolution. *PLoS pathogens*, 8(10):e1002979. doi:10.1371/journal.ppat.1002979.

Helenius, A., Mellman, I., Wall, D., and Hubbard, A. (1983). Endosomes. *Trends in Biochemical Sciences*, 8(7):245–250.

Hennies, H.C., Kornak, U., Zhang, H., Egerer, J., Zhang, X., Seifert, W., Kühnisch, J., Budde, B., Nätebus, M., Brancati, F., Wilcox, W.R., Müller, D., Kaplan, P.B., Rajab, A., Zampino, G., Fodale, V., Dallapiccola, B., Newman, W., Metcalfe, K., Clayton-Smith, J., Tassabehji, M., Steinmann, B., a Barr, F., Nürnberg, P., Wieacker, P., and Mundlos, S. (2008). Gerodermia osteodysplastica is caused by mutations in SCYL1BP1, a Rab-6 interacting golgin. *Nat. Genet.*, 40(12):1410–2. doi:10.1038/ng.252.

Hill, E., van der Kaay, J., Downes, C.P., and Smythe, E. (2001). The Role of Dynamin and Its Binding Partners in Coated Pit Invagination and Scission. *The Journal of Cell Biology*, 152(2):309–324. doi:10.1083/jcb.152.2.309.

Hine, R. (2019). *A Dictionary of Biology*. Oxford University Press, 8 edition. ISBN 978-0-19-882148-9.

Hirst, J., Barlow, L.D., Francisco, G.C., Sahlender, D.A., Seaman, M.N.J., Dacks, J.B., and Robinson, M.S. (2011). The Fifth Adaptor Protein Complex. *PLoS Biology*, 9(10):e1001170. doi:10.1371/journal.pbio.1001170.

Hirst, J., Borner, G.H.H., Edgar, J., Hein, M.Y., Mann, M., Buchholz, F., Antrobus, R., and Robinson, M.S. (2013a). Interaction between AP-5 and the hereditary spastic paraplegia proteins SPG11 and SPG15. *Molecular Biology of the Cell*, 24(16):2558–2569. doi:10.1091/mbc.E13-03-0170.

Hirst, J., Bright, N.A., Rous, B., and Robinson, M.S. (1999). Characterization of a Fourth Adaptor-related Protein Complex. *Molecular Biology of the Cell*, 10(8):2787–2802. doi:10.1091/mbc.10.8.2787.

Hirst, J., Edgar, J.R., Esteves, T., Darios, F., Madeo, M., Chang, J., Roda, R.H., Dürr, A., Anheim, M.,

Gellera, C., Li, J., Züchner, S., Mariotti, C., Stevanin, G., Blackstone, C., Kruer, M.C., and Robinson, M.S. (2015). Loss of AP-5 results in accumulation of aberrant endolysosomes: Defining a new type of lysosomal storage disease. *Hum. Mol. Genet.*, 24(17):4984–4996. doi:10.1093/hmg/ddv220.

Hirst, J., Irving, C., and Borner, G.H. (2013b). Adaptor Protein Complexes AP-4 and AP-5: New Players in Endosomal Trafficking and Progressive Spastic Paraplegia. *Traffic*, 14(2):153–164. doi:10.1111/tra.12028.

Hirst, J., Itzhak, D.N., Antrobus, R., Borner, G.H.H., and Robinson, M.S. (2018). Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLOS Biology*, 16(1):e2004411. doi:10.1371/journal.pbio.2004411.

Hirst, J., Schlacht, A., Norcott, J.P., Traynor, D., Bloomfield, G., Antrobus, R., Kay, R.R., Dacks, J.B., and Robinson, M.S. (2014). Characterization of TSET, an ancient and widespread membrane trafficking complex. *eLife*, 3:e02866. doi:10.7554/eLife.02866.

Ho, H.H., He, C.Y., de Graffenried, C.L., Murrells, L.J., and Warren, G. (2006). Ordered assembly of the duplicating Golgi in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U. S. A.*, 103(20):7676–7681. doi:10.1073/pnas.0602595103.

Holt, M., Varoqueaux, F., Wiederhold, K., Takamori, S., Urlaub, H., Fasshauer, D., and Jahn, R. (2006). Identification of SNAP-47, a Novel Qbc-SNARE with Ubiquitous Expression. *Journal of Biological Chemistry*, 281(25):17076–17083. doi:10.1074/jbc.M513838200.

Holthuis, J.C. (1998). Two syntaxin homologues in the TGN/endosomal system of yeast. *The EMBO Journal*, 17(1):113–126. doi:10.1093/emboj/17.1.113.

Hong, W. and Lev, S. (2014). Tethering the assembly of SNARE complexes. *Trends in Cell Biology*, 24(1):35–43. doi:10.1016/j.tcb.2013.09.006.

Hoogenraad, C.C., Wulf, P., Schiefermeier, N., Stepanova, T., Galjart, N., Small, J.V., Grosveld, F., De Zeeuw, C.I., and Akhmanova, A. (2003). Bicaudal D induces selective dynein-mediated microtubule minus end-directed transport. *EMBO J.*, 22(22):6004–6015. doi:10.1093/emboj/cdg592.

Horwitz, M.A. (1984). *Legionella Pneumophila* inhibits acidification of its phagosome in human monocytes. *The Journal of Cell Biology*, 99(6):1936–1943. doi:10.1083/jcb.99.6.1936.

Hu, F., Shi, X., Li, B., Huang, X., Morelli, X., and Shi, N. (2015a). Structural basis for the interaction between the Golgi reassembly-stacking protein GRASP65 and the Golgi matrix protein GM130. *J. Biol. Chem.*, 290(44):26373–26382. doi:10.1074/jbc.M115.657940.

Hu, S., Parker, J., and Wright, J.T. (2015b). Towards Unraveling the Human Tooth Transcriptome: The Dentome. *PLoS One*, 10(4):e0124801. doi:10.1371/journal.pone.0124801.

Huelsenbeck, J.P. (1997). Phylogenetic Methods Come of Age: Testing Hypotheses in an Evolutionary Context. *Science*, 276(5310):227–232. doi:10.1126/science.276.5310.227.

Huerta-Cepas, J., Dopazo, J., and Gabaldón, T. (2010). ETE: A python Environment for Tree Exploration. *BMC Bioinformatics*, 11(1):24. doi:10.1186/1471-2105-11-24.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6):1635–1638. doi:10.1093/molbev/msw046.

Hug, L.A., Stechmann, A., and Roger, A.J. (2010). Phylogenetic Distributions and Histories of Proteins Involved in Anaerobic Pyruvate Metabolism in Eukaryotes. *Molecular Biology and Evolution*, 27(2):311–324. doi:10.1093/molbev/msp237.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3):90–95. doi:10.1109/MCSE.2007.55.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, a.F., Selengut, J.D., Sigrist, C.J.A., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H., and Yeats,

C. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Res.*, 37(Database):D211–D215. doi:10.1093/nar/gkn785.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H., and Takai, K. (2019). Isolation of an archaeon at the prokaryote-eukaryote interface. Preprint, Microbiology. doi:10.1101/726976.

Itakura, E., Kishi-Itakura, C., and Mizushima, N. (2012). The Hairpin-type Tail-Anchored SNARE Syntaxin 17 Targets to Autophagosomes for Fusion with Endosomes/Lysosomes. *Cell*, 151(6):1256–1269. doi:10.1016/j.cell.2012.11.001.

Jackson, A.J., Clucas, C., Mamczur, N.J., Ferguson, D.J., and Meissner, M. (2013). *Toxoplasma Gondii* Syntaxin 6 is required for vesicular transport between endosomal-like compartments and the Golgi complex. *Traffic*, 14(11):1166–81. doi:10.1111/tra.12102.

Jahn, R. and Niemann, H. (1994). Molecular Mechanisms of Clostridial Neurotoxins. *Annals of the New York Academy of Sciences*, 733(1 Molecular and):245–255. doi:10.1111/j.1749-6632.1994.tb17274.x.

Jahn, R. and Scheller, R.H. (2006). SNAREs — engines for membrane fusion. *Nature Reviews Molecular Cell Biology*, 7(9):631–643. doi:10.1038/nrm2002.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schüssler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lücking, R., Büdel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443(7113):818–822. doi:10.1038/nature05110.

James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J.E. (2013). Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. *Current biology: CB*, 23(16):1548–1553. doi:10.1016/j.cub.2013.06.057.

Janouškovec, J., Tikhonenkov, D.V., Mikhailov, K.V., Simdyanov, T.G., Aleoshin, V.V., Mylnikov, A.P., and Keeling, P.J. (2013). Colponemids represent multiple ancient alveolate lineages. *Curr. Biol.*, 23(24):2546–2552. doi:10.1016/j.cub.2013.10.062.

Jékely, G. (2003). Small GTPases and the evolution of the eukaryotic cell: Hypothesis. *BioEssays*, 25(11):1129–1138. doi:10.1002/bies.10353.

Jékely, G. (2007). Evolution of eukaryotic endomembranes: A critical evaluation of different scenarios. In G. Jékely, editor, *Eukaryotic Membranes and Cytoskeleton: Origins and Evolution*, number v. 607 in Advances in Experimental Medicine and Biology. Springer Science+Business Media ; Landes Bioscience, New York, N.Y. : Austin, Tex. ISBN 978-0-387-74020-1. OCLC: ocn157002733.

Jékely, G. (2008). Evolution of the Golgi complex. In *The Golgi Apparatus*, pages 675–691. Springer-Verlag/Wien. ISBN 978-3-211-76309-4.

Jiao, Y., Li, J., Tang, H., and Paterson, A.H. (2014). Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots. *The Plant Cell*, 26(7):2792–2802. doi:10.1105/tpc.114.127597.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., and dePamphilis, C.W. (2011). Ancestral polyploidy in seed plants and angiosperms. *NATURE*, 473(7345):97–U113. doi:10.1038/nature09916.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder:

Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587–589. doi:10.1038/nmeth.4285.

Kanazawa, T., Era, A., Minamino, N., Shikano, Y., Fujimoto, M., Uemura, T., Nishihama, R., Yamato, K.T., Ishizaki, K., Nishiyama, T., Kohchi, T., Nakano, A., and Ueda, T. (2016). SNARE Molecules in *Marchantia polymorpha* : Unique and Conserved Features of the Membrane Fusion Machinery. *Plant and Cell Physiology*, 57(2):307–324. doi:10.1093/pcp/pcv076.

Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcântara, D.M.C., Roger, A.J., Shadwick, L.L., Smirnov, A., Kudryavtsev, A., Lahr, D.J.G., and Brown, M.W. (2017). Between a Pod and a Hard Test: The Deep Evolution of Amoebae. *Molecular Biology and Evolution*, 34(9):2258–2270. doi:10.1093/molbev/msx162.

Karnahl, M., Park, M., Mayer, U., Hiller, U., and Jürgens, G. (2017). ER assembly of SNARE complexes mediating formation of partitioning membrane in *Arabidopsis* cytokinesis. *Elife*, 6:1–10. doi:10.7554/eLife.25327.

Karnkowska, A., Treitli, S.C., Brzoň, O., Novák, L., Vacek, V., Soukal, P., Barlow, L.D., Herman, E.K., Pipaliya, S.V., Pánek, T., Žihala, D., Petrželková, R., Butenko, A., Eme, L., Stairs, C.W., Roger, A.J., Eliáš, M., Dacks, J.B., and Hampl, V. (2019). The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion. *Molecular Biology and Evolution*, page msz147. doi:10.1093/molbev/msz147.

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C.W., Roger, A.J., Eliáš, M., Dacks, J.B., Vlček, Č., and Hampl, V. (2016). A Eukaryote without a Mitochondrial Organelle. *Current Biology*, 26(10):1274–1284. doi:10.1016/j.cub.2016.03.053.

Karol, K.G. (2001). The Closest Living Relatives of Land Plants. *Science*, 294(5550):2351–2353. doi:10.1126/science.1065156.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivarès, C.P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, 414(6862):450–453. doi:10.1038/35106579.

Katz, L. and Brennwald, P. (2000). Testing the 3Q:1R "Rule": Mutational Analysis of the Ionic "Zero" Layer in the Yeast Exocytic SNARE Complex Reveals No Requirement for Arginine. *Molecular Biology of the Cell*, 11(11):3849–3858. doi:10.1091/mbc.11.11.3849.

Katz, L., Hanson, P.I., Heuser, J.E., and Brennwald, P. (1998). Genetic and morphological analyses reveal a critical interaction between the C-termini of two SNARE proteins and a parallel four helical arrangement for the exocytic SNARE complex. *The EMBO Journal*, 17(21):6200–6209. doi:10.1093/emboj/17.21.6200.

Katz, L.A., Grant, J.R., Parfrey, L.W., and Burleigh, J.G. (2012). Turning the Crown Upside Down: Gene Tree Parsimony Roots the Eukaryotic Tree of Life. *Systematic Biology*, 61(4):653–660. doi:10.1093/sysbio/sys026.

Kauko, A. and Lehto, K. (2018). Eukaryote specific folds: Part of the whole. *Proteins: Structure, Function, and Bioinformatics*, 86(8):868–881. doi:10.1002/prot.25517.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B.,

Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., and Worden, A.Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.*, 12(6). doi:10.1371/journal.pbio.1001889.

Keeling, P.J. and Fast, N.M. (2002). Microsporidia: Biology and evolution of highly reduced intracellular parasites. *Annual Review of Microbiology*, 56:93–116. doi:10.1146/annurev.micro.56.012302.160854.

Kelley, D.R. and Salzberg, S.L. (2010). Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome Biology*, 11(3):R28. doi:10.1186/gb-2010-11-3-r28.

Kenrick, P., Wellman, C.H., Schneider, H., and Edgecombe, G.D. (2012). A timeline for terrestrialization: Consequences for the carbon cycle in the Palaeozoic. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1588):519–536. doi:10.1098/rstb.2011.0271.

Khoriaty, R., Hesketh, G.G., Bernard, A., Weyand, A.C., Mellacheruvu, D., Zhu, G., Hoenerhoff, M.J., McGee, B., Everett, L., Adams, E.J., Zhang, B., Saunders, T.L., Nesvizhskii, A.I., Klionsky, D.J., Shavit, J.A., Gingras, A.C., and Ginsburg, D. (2018). Functions of the COPII gene paralogs SEC23A and SEC23B are interchangeable in vivo. *Proceedings of the National Academy of Sciences*, 115(33):E7748–E7757. doi:10.1073/pnas.1805784115.

Kibria, K.M.K., Ferdous, J., Sardar, R., Panda, A., Gupta, D., Mohmmed, A., and Malhotra, P. (2019). A genome-wide analysis of coatomer protein (COP) subunits of apicomplexan parasites and their evolutionary relationships. *BMC Genomics*, 20(1). doi:10.1186/s12864-019-5463-1.

Kienle, N., Kloepper, T.H., and Fasshauer, D. (2009). Differences in the SNARE evolution of fungi and metazoa. *Biochemical Society Transactions*, 37(4):787–791. doi:10.1042/BST0370787.

Kim, S.Y., Xu, Z.Y., Song, K., Kim, D.H., Kang, H., Reichardt, I., Sohn, E.J., Friml, J., Juergens, G., and Hwang, I. (2013). Adaptor Protein Complex 2-Mediated Endocytosis Is Crucial for Male Reproductive Organ Development in *Arabidopsis*. *The Plant Cell*, 25(8):2970–2985. doi:10.1105/tpc.113.114264.

Kinseth, M.A., Anjard, C., Fuller, D., Guizzunti, G., Loomis, W.F., and Malhotra, V. (2007). The Golgi-Associated Protein GRASP Is Required for Unconventional Protein Secretion during Development. *Cell*, 130(3):524–534. doi:10.1016/j.cell.2007.06.029.

Kirchhausen, T. (1990). Identification of a putative yeast homolog of the mammalian beta chains of the clathrin-associated protein complexes. *Molecular and Cellular Biology*, 10(11):6089–6090. doi:10.1128/mcb.10.11.6089.

Kishino, H., Miyata, T., and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2):151–160. doi:10.1007/BF02109483.

Kissmehl, R., Schilde, C., Wassmer, T., Danzer, C., Nuehse, K., Lutter, K., and Plattner, H. (2007). Molecular Identification of 26 Syntaxin Genes and their Assignment to the Different Trafficking Pathways in *Paramecium*. *Traffic*, 8(5):523–542. doi:10.1111/j.1600-0854.2007.00544.x.

Klinger, C.M., Ramirez-Macias, I., Herman, E.K., Turkewitz, A.P., Field, M.C., and Dacks, J.B. (2016a). Resolving the homology—function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology. *Molecular and Biochemical Parasitology*, 209(1-2):88–103. doi:10.1016/j.molbiopara.2016.07.003.

Klinger, C.M., Spang, A., Dacks, J.B., and Ettema, T.J. (2016b). Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Molecular Biology and Evolution*, 33(6):1528–1541. doi:10.1093/molbev/msw034.

Kloepper, T.H., Kienle, C.N., and Fasshauer, D. (2007). An Elaborate Classification of SNARE Proteins Sheds Light on the Conservation of the Eukaryotic Endomembrane System. *Molecular Biology of the Cell*, 18(9):3463–3471. doi:10.1091/mbc.e07-03-0193.

Kloepper, T.H., Kienle, C.N., and Fasshauer, D. (2008). SNAREing the Basis of Multicellularity: Consequences of Protein Family Expansion during Evolution. *Molecular Biology and Evolution*, 25(9):2055–2068. doi:10.1093/molbev/msn151.

Klute, M.J., Melancon, P., and Dacks, J.B. (2011). Evolution and Diversity of the Golgi. *Cold Spring Harb. Lab. Press*, 100:29–35. doi:10.1024/1661-8157/a000403.

Koch, M., Haubold, B., and Mitchell-Olds, T. (2001). Molecular systematics of the Brassicaceae: Evidence from coding plastidic matK and nuclear Chs sequences. *American Journal of Botany*, 88(3):534–544. doi:10.2307/2657117.

Kodani, A. and Sutterlin, C. (2008). The Golgi Protein GM130 Regulates Centrosome Morphology and Function. *Mol. Biol. Cell*, 19(February):745–753. doi:10.1091/mbc.E07.

Koike, S. and Jahn, R. (2019). SNAREs define targeting specificity of trafficking vesicles by combinatorial interaction with tethering factors. *Nature Communications*, 10(1):1608. doi:10.1038/s41467-019-09617-9.

Kondylis, V. and Rabouille, C. (2009). The Golgi apparatus: Lessons from *Drosophila*. *FEBS Lett.*, 583(23):3827–38. doi:10.1016/j.febslet.2009.09.048.

Koornneef, M. and Meinke, D. (2010). The development of *Arabidopsis* as a model plant. *The Plant Journal*, 61(6):909–921. doi:10.1111/j.1365-313X.2009.04086.x.

Koreishi, M., Gniadek, T.J., Yu, S., Masuda, J., Honjo, Y., and Satoh, A. (2013). The golgin tether giantin regulates the secretory pathway by controlling stack organization within Golgi apparatus. *PLoS One*, 8(3):e59821. doi:10.1371/journal.pone.0059821.

Koumandou, V.L., Dacks, J.B., Coulson, R.M., and Field, M.C. (2007). Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evolutionary Biology*, 7(1):29. doi:10.1186/1471-2148-7-29.

Koumandou, V.L., Klute, M.J., Herman, E.K., Nunez-Miguel, R., Dacks, J.B., and Field, M.C. (2011). Evolutionary reconstruction of the retromer complex and its function in *Trypanosoma brucei*. *J. Cell Sci.*, 124(Pt 9):1496–509. doi:10.1242/jcs.081596.

Koumandou, V.L., Wickstead, B., Ginger, M.L., van der Giezen, M., Dacks, J.B., and Field, M.C. (2013). Molecular paleontology and complexity in the last eukaryotic common ancestor. *Critical Reviews in Biochemistry and Molecular Biology*, 48(4):373–396. doi:10.3109/10409238.2013.821444.

Kraynack, B.A., Chan, A., Rosenthal, E., Essid, M., Umansky, B., Waters, M.G., and Schmitt, H.D. (2005). Dsl1p, Tip20p, and the Novel Dsl3(Sec39) Protein Are Required for the Stability of the Q/t-SNARE Complex at the Endoplasmic Reticulum in Yeast. *Molecular Biology of the Cell*, 16:15.

Kreykenbohm, V. (2002). The SNAREs vti1a and vti1b have distinct localization and SNARE complex partners. *European Journal of Cell Biology*, 81(5):273–280. doi:10.1078/0171-9335-00247.

Kühnle, J., Shillcock, J., Mouritsen, O.G., and Weiss, M. (2010). A modeling approach to the self-assembly of the Golgi apparatus. *Biophys. J.*, 98(12):2839–47. doi:10.1016/j.bpj.2010.03.035.

Kwon, C., Neu, C., Pajonk, S., Yun, H.S., Lipka, U., Humphry, M., Bau, S., Straus, M., Kwaaitaal, M., Rampelt, H., Kasmi, F.E., Jürgens, G., Parker, J., Panstruga, R., Lipka, V., and Schulze-Lefert, P. (2008). Co-option of a default secretory pathway for plant immune responses. *Nature*, 451(7180):835–840. doi:10.1038/nature06545.

Lam, S.K., Tse, Y.C., Robinson, D.G., and Jiang, L. (2007). Tracking down the elusive early endosome. *Trends in Plant Science*, 12(11):497–505. doi:10.1016/j.tplants.2007.09.001.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., and Huala, E. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40(D1):D1202–D1210. doi:10.1093/nar/gkr1090.

Lane, N. and Martin, W. (2010). The energetics of genome complexity. *Nature*, 467(7318):929–934. doi:10.1038/nature09486.

Larson, R.T., Dacks, J.B., and Barlow, L.D. (2019). Recent gene duplications dominate evolutionary dynamics of adaptor protein complex subunits in embryophytes. *Traffic*, page tra.12698. doi:10.1111/tra.12698.

Latijnhouwers, M., Gillespie, T., Boevink, P., Kriechbaumer, V., Hawes, C., and Carvalho, C.M. (2007). Localization and domain characterization of *Arabidopsis* golgin candidates. *Journal of Experimental Botany*, 58(15-16):4373–4386. doi:10.1093/jxb/erm304.

Latijnhouwers, M., Hawes, C., Carvalho, C., Oparka, K., Gillingham, A.K., and Boevink, P. (2005). An *Arabidopsis* GRIP domain protein locates to the trans-Golgi and binds the small GTPase ARL1: An *Arabidopsis* GRIP protein and ARL1 GTPase interact. *The Plant Journal*, 44(3):459–470. doi:10.1111/j.1365-313X.2005.02542.x.

Lauber, M.H., Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz, W., and Jürgens, G. (1997). The *Arabidopsis* KNOLLE Protein Is a Cytokinesis-specific Syntaxin. *The Journal of Cell Biology*, 139(6):1485–1493. doi:10.1083/jcb.139.6.1485.

Laufman, O., Hong, W., and Lev, S. (2013). The COG complex interacts with multiple Golgi SNAREs and enhances fusogenic assembly of SNARE complexes. *Journal of Cell Science*, 126(6):1506–1516. doi:10.1242/jcs.122101.

Lax, G., Eglit, Y., Eme, L., Bertrand, E.M., Roger, A.J., and Simpson, A.G.B. (2018). Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature*. doi:10.1038/s41586-018-0708-8.

Lazarow, P.B. and Fujiki, Y. (1985). Biogenesis of Peroxisomes. *Annual Review of Cell Biology*, 1:489–530.

Le, S.Q. and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7):1307–1320. doi:10.1093/molbev/msn067.

Le Borgne, R., Alconada, A., Bauer, U., and Hoflack, B. (1998). The Mammalian AP-3 Adaptor-like Complex Mediates the Intracellular Transport of Lysosomal Membrane Glycoproteins. *Journal of Biological Chemistry*, 273(45):29451–29461. doi:10.1074/jbc.273.45.29451.

Lee, I., Tiwari, N., Dunlop, M.H., Graham, M., Liu, X., and Rothman, J.E. (2014). Membrane adhesion dictates Golgi stacking and cisternal morphology. *Proc. Natl. Acad. Sci. U. S. A.*, 111(5):1849–54. doi:10.1073/pnas.1323895111.

Lee, L.J., Klute, M.J., Herman, E.K., Read, B., and Dacks, J.B. (2015a). Losses, Expansions, and Novel Subunit Discovery of Adaptor Protein Complexes in Haptophyte Algae. *Protist*, 166(5):585–597. doi:10.1016/j.protis.2015.07.004.

Lee, M., Ko, Y.J., Moon, Y., Han, M., Kim, H.W., Lee, S.H., Kang, K., and Jun, Y. (2015b). SNAREs support atlastin-mediated homotypic ER fusion in *Saccharomyces cerevisiae*. *The Journal of Cell Biology*, 210(3):451–470. doi:10.1083/jcb.201501043.

Lee, M.C.S., Orci, L., Hamamoto, S., Futai, E., Ravazzola, M., and Schekman, R. (2005). Sar1p N-Terminal Helix Initiates Membrane Curvature and Completes the Fission of a COPII Vesicle. *Cell*, 122(August 26):605–617.

Lee, M.H. and Hwang, I. (2014). Adaptor proteins in protein trafficking between endomembrane compartments in plants. *Journal of Plant Biology*, 57(5):265–273. doi:10.1007/s12374-014-0314-8.

Lencer, W.I., Hirst, T.R., and Holmes, R.K. (1999). Membrane traffic and the cellular uptake of cholera toxin. *Biochimica et Biophysica Acta*, 1450:177–190.

Lewis, M.J. (1997). A novel SNARE complex implicated in vesicle fusion with the endoplasmic reticulum. *The EMBO Journal*, 16(11):3017–3024. doi:10.1093/emboj/16.11.3017.

Lewis, M.J., Nichols, B.J., Prescianotto-Baschong, C., Riezman, H., and Pelham, H.R.B. (2000). Specific Retrieval of the Exocytic SNARE Snc1p from Early Yeast Endosomes. *Molecular Biology of the Cell*, 11(1):23–38. doi:10.1091/mbc.11.1.23.

Lewis, M.J. and Pelham, H.R.B. (2002). A New Yeast Endosomal SNARE Related to Mammalian Syntaxin 8. *Traffic*, 3(12):922–929. doi:10.1034/j.1600-0854.2002.31207.x.

Li, W., Puertollano, R., Bonifacino, J.S., Overbeek, P.A., and Everett, E.T. (2010). Disruption of the Murine *Ap2β1* Gene Causes Nonsyndromic Cleft Palate. *The Cleft Palate-Craniofacial Journal*, 47(6):566–573. doi: 10.1597/09-145.

Li, Y., Kelly, W.G., Logsdon, J.M., Schurko, A.M., Harfe, B.D., Hill-Harfe, K.L., and a Kahn, R. (2004). Functional genomic analysis of the ADP-ribosylation factor family of GTPases: Phylogeny among diverse eukaryotes and function in C. elegans. *FASEB J.*, 18(15):1834–1850. doi:10.1096/fj.04-2273com.

Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H., and Barker, M.S. (2015). Early genome duplications in conifers and other seed plants. *Science Advances*, 1(10):e1501084. doi:10.1126/sciadv.1501084.

Lin, R.C. and Scheller, R.H. (1997). Structural Organization of the Synaptic Exocytosis Core Complex. *Neuron*, 19(5):1087–1094. doi:10.1016/S0896-6273(00)80399-2.

Link, E., McMahon, H., Fischer von Mollard, G., Yamasaki, S., Niemann, H., Sudhof, T.C., and Jahn, R. (1993). Cleavage of Cellubrevin by Tetanus Toxin Does Not Affect Fusion of Early Endosomes. *The Journal of Biological Chemistry*, 268(25):18423–18426.

Liu, Y. and Barlowe, C. (2002). Analysis of Sec22p in Endoplasmic Reticulum/Golgi Transport Reveals Cellular Redundancy in SNARE Protein Function. *Molecular Biology of the Cell*, 13(9):3314–3324. doi:10.1091/mbc.e02-04-0204.

Loughran, G., Pinter, K., Newell, P.C., and Gross, J.D. (2000). Identification of STKA-dependent genes in *Dictyostelium discoideum*. *Differentiation*, 66(2-3):71–80. doi:10.1046/j.1432-0436.2000.660202.x.

Ludwig, T. (1991). Distribution of newly synthesized lysosomal enzymes in the endocytic pathway of normal rat kidney cells. *The Journal of Cell Biology*, 115(6):1561–1572. doi:10.1083/jcb.115.6.1561.

Lupashin, V.V., Pokrovskaya, I.D., McNew, J.A., and Waters, M.G. (1997). Characterization of a Novel Yeast SNARE Protein Implicated in Golgi Retrograde Traffic. *Molecular Biology of the Cell*, 8(12):2659–2676. doi:10.1091/mbc.8.12.2659.

Lürick, A., Gao, J., Kuhlee, A., Yavavli, E., Langemeyer, L., Perz, A., Raunser, S., and Ungermann, C. (2017). Multivalent Rab interactions determine tether-mediated membrane fusion. *Molecular Biology of the Cell*, 28(2):322–332. doi:10.1091/mbc.e16-11-0764.

Lynch, M., Field, M.C., Goodson, H.V., Malik, H.S., Pereira-Leal, J.B., Roos, D.S., Turkewitz, A.P., and Sazer, S. (2014). Evolutionary cell biology: Two origins, one objective. *Proceedings of the National Academy of Sciences*, 111(48):16990–16994. doi:10.1073/pnas.1415861111.

Mallard, F., Tang, B.L., Galli, T., Tenza, D., Saint-Pol, A., Yue, X., Antony, C., Hong, W., Goud, B., and Johannes, L. (2002). Early/recycling endosomes-to-TGN transport involves two SNARE complexes and a Rab6 isoform. *The Journal of Cell Biology*, 156(4):653–664. doi:10.1083/jcb.200110081.

Mani, S. and Thattai, M. (2016). Stacking the odds for Golgi cisternal maturation. *Elife*. doi:10.7554/eLife.16231.

Maniak, M. (2003). Fusion and Fission Events in the Endocytic Pathway of *Dictyostelium*: Fusion and Fission in Endocytosis. *Traffic*, 4(1):1–5. doi:10.1034/j.1600-0854.2003.40101.x.

Manna, P.T., Kelly, S., and Field, M.C. (2013). Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids. *Molecular Phylogenetics and Evolution*, 67(1):123–128. doi:10.1016/j.ympev.2013.01.002.

Markus Boehm and Juan S. Bonifacino (2001). Adaptins the final recount. *Molecular Biology of the Cell*, 12(October):2907–2920.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P., and Nellen, W. (2002). RNAi in *Dictyostelium:* The Role of RNA-directed RNA Polymerases and Double-stranded RNase. *Molecular Biology of the Cell*, 13(2):445–453. doi:10.1091/mbc.01-04-0211.

Marti, M. and Hehl, A.B. (2003). Encystation-specific vesicles in *Giardia*: A primordial Golgi or just another secretory compartment? *Trends Parasitol.*, 19(10):440–446. doi:10.1016/S1471-4922(03)00201-0.

Marti, M., Regös, A., Li, Y., Schraner, E.M., Wild, P., Müller, N., Knopf, L.G., and Hehl, A.B. (2003). An Ancestral Secretory Apparatus in the Protozoan Parasite *Giardia intestinalis*. *J. Biol. Chem.*, 278(27):24837–24848. doi:10.1074/jbc.M302082200.

Martijn, J., Vosseberg, J., Guy, L., Offre, P., and Ettema, T.J.G. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, 557(7703):101–105. doi:10.1038/s41586-018-0059-5.

Martin, W. and Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature*, 392(6671):37–41. doi:10.1038/32096.

Martin, W.F., Tielens, A.G.M., Mentel, M., Garg, S.G., and Gould, S.B. (2017). The Physiology of Phagocytosis in the Context of Mitochondrial Origin. *Microbiology and Molecular Biology Reviews*, 81(3):e00008–17, /mmbr/81/3/e00008–17.atom. doi:10.1128/MMBR.00008-17.

Martinez-Arca, S., Rudge, R., Vacca, M., Raposo, G., Camonis, J., Proux-Gillardeaux, V., Daviet, L., Formstecher, E., Hamburger, A., Filippini, F., D'Esposito, M., and Galli, T. (2003). A dual mechanism controlling the localization and function of exocytic v-SNAREs. *Proceedings of the National Academy of Sciences*, 100(15):9011–9016. doi:10.1073/pnas.1431910100.

Mast, F.D., Barlow, L.D., Rachubinski, R.A., and Dacks, J.B. (2014). Evolutionary mechanisms for establishing eukaryotic cellular complexity. *Trends in Cell Biology*, 24(7):435–442. doi:10.1016/j.tcb.2014.02.003.

Matsuura-Tokita, K., Takeuchi, M., Ichihara, A., Mikuriya, K., and Nakano, A. (2006). Live imaging of yeast Golgi cisternal maturation. *Nature*, 441(7096):1007–1010. doi:10.1038/nature04737.

Mattera, R., Park, S.Y., De Pace, R., Guardia, C.M., and Bonifacino, J.S. (2017). AP-4 mediates export of ATG9A from the *trans* -Golgi network to promote autophagosome formation. *Proceedings of the National Academy of Sciences*, 114(50):E10697–E10706. doi:10.1073/pnas.1717327114.

Mazur, D.R. (2010). *Combinatorics A Guided Tour*. The Mathematical Association of America (Incorporated). ISBN 978-1-61444-607-1.

McConville, M.J., Ilgoutz, S.C., Teasdale, R.D., Foth, B.J., Matthews, A., a Mullin, K., and a Gleeson, P. (2002). Targeting of the GRIP domain to the trans-Golgi network is conserved from protists to animals. *Eur. J. Cell Biol.*, 81(9):485–95.

McLelland, G.L., Lee, S.A., McBride, H.M., and Fon, E.A. (2016). Syntaxin-17 delivers PINK1/parkin-dependent mitochondrial vesicles to the endolysosomal system. *The Journal of Cell Biology*, 214(3):275–291. doi:10.1083/jcb.201603105.

McMahon, H.T., Ushkaryov, Y.A., Edelmann, L., Link, E., Binz, T., Niemann, H., Jahn, R., and Südhof, T.C. (1993). Cellubrevin is a ubiquitous tetanus-toxin substrate homologous to a putative synaptic vesicle fusion protein. *Nature*, 364(6435):346–349. doi:10.1038/364346a0.

McNew, J.A., Parlati, F., Fukuda, R., Johnston, R.J., Paz, K., Paumet, F., and Rothman, J.E. (2000a). Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature*, 407:7.

McNew, J.A., Søgaard, M., Lampen, N.M., Machida, S., Ye, R.R., Lacomis, L., Tempst, P., Rothman, J.E., and Söllner, T.H. (1997). Ykt6p, a Prenylated SNARE Essential for Endoplasmic Reticulum-Golgi Transport. *Journal of Biological Chemistry*, 272(28):17776–17783. doi:10.1074/jbc.272.28.17776.

McNew, J.A., Weber, T., Parlati, F., Johnston, R.J., Melia, T.J., Söllner, T.H., and Rothman, J.E. (2000b). Close Is Not Enough: Snare-Dependent Membrane Fusion Requires an Active Mechanism That Transduces Force to Membrane Anchors. *The Journal of Cell Biology*, 150(1):105–118. doi:10.1083/jcb.150.1.105.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pages 1–8. doi:10.1109/GCE.2010.5676129.

Minh, B.Q., Nguyen, M.A.T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.*, 30(5):1188–1195. doi:10.1093/molbev/mst024.

Mironov, A.A., Banin, V.V., Sesorova, I.S., Dolgikh, V.V., Luini, A., and Beznoussenk, G.V. (2007). Evolution of the Endoplasmic Reticulum and the Golgi Complex. In G. Jékely, editor, *Eukaryotic Membranes and Cytoskeleton: Origins and Evolution*, number v. 607 in Advances in Experimental Medicine and Biology, pages 61–72. Springer Science+Business Media ; Landes Bioscience, New York, N.Y. : Austin, Tex. ISBN 978-0-387-74020-1. OCLC: ocn157002733.

Misumi, Y., Sohda, M., Tashiro, A., Sato, H., and Ikehara, Y. (2001). An Essential Cytoplasmic Domain for the Golgi Localization of Coiled-coil Proteins with a COOH-terminal Membrane Anchor. *Journal of Biological Chemistry*, 276(9):6867–6873. doi:10.1074/jbc.M010121200.

Misura, K.M.S., Bock, J.B., Gonzalez, L.C., Scheller, R.H., and Weis, W.I. (2002). Three-dimensional structure of the amino-terminal domain of syntaxin 6, a SNAP-25 C homolog. *Proceedings of the National Academy of Sciences*, 99(14):9184–9189. doi:10.1073/pnas.132274599.

Monnat, J., Neuhaus, E.M., Pop, M.S., Ferrari, D.M., Kramer, B., and Soldati, T. (2000). Identification of a Novel Saturable Endoplasmic Reticulum Localization Mechanism Mediated by the C-Terminus of a *Dictyostelium* Protein Disulfide Isomerase. *Molecular Biology of the Cell*, 11(10):3469–3484. doi:10.1091/mbc.11.10.3469.

Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., and Soltis, D.E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences*, 107(10):4623–4628. doi:10.1073/pnas.0907801107.

Moreno-De-Luca, A., Helmers, S.L., Mao, H., Burns, T.G., Melton, A.M.A., Schmidt, K.R., Fernhoff, P.M., Ledbetter, D.H., and Martin, C.L. (2011). Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *Journal of Medical Genetics*, 48(2):141–144. doi:10.1136/jmg.2010.082263.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10):E2274–E2283. doi:10.1073/pnas.1719588115.

Morris, S.C. (1989). Burgess Shale Faunas and the Cambrian Explosion. *Science*, 246(4928):339–346.

Mowbrey, K. and Dacks, J.B. (2009). Evolution and diversity of the Golgi body. *FEBS Letters*, 583(23):3738–3745. doi:10.1016/j.febslet.2009.10.025.

Mukai, A., Ichiraku, A., and Horikawa, K. (2016). Reliable handling of highly A/T-rich genomic DNA for efficient generation of knockin strains of *Dictyostelium discoideum*. *BMC Biotechnology*, 16(1):37. doi:10.1186/s12896-016-0267-8.

Mullock, B.M., Smith, C.W., Ihrke, G., Bright, N.A., Lindsay, M., Parkinson, E.J., Brooks, D.A., Parton, R.G., James, D.E., Luzio, J.P., and Piper, R.C. (2000). Syntaxin 7 Is Localized to Late Endosome Compartments, Associates with Vamp 8, and Is Required for Late Endosome–Lysosome Fusion. *Molecular Biology of the Cell*, 11(9):3137–3153. doi:10.1091/mbc.11.9.3137.

Munoz-Gomez, S.A., Hess, S., Burger, G., Lang, B.F., Susko, E., Slamovits, C.H., and Roger, A.J. (2019). An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife*, 8:e42535. doi:DOI:https://doi.org/10.7554/eLife.42535.

Munoz-Gomez, S.A., Slamovits, C.H., Dacks, J.B., Baier, K.A., Spencer, K.D., and Wideman, J.G. (2014). Ancient Homology of the Mitochondrial Contact Site and Cristae Organizing System Points to an Endosymbiotic Origin of Mitochondrial Cristae. *Curr. Biol.*, pages 1489–1495. doi:10.1016/j.cub.2015.04.006.

Munoz-Gomez, S.A., Wideman, J.G., Roger, A.J., and Slamovits, C.H. (2017). The Origin of Mitochondrial Cristae from Alphaproteobacteria. *Molecular Biology and Evolution*, 34(4):943–956. doi:10.1093/molbev/msw298.

Munro, S. (2011). The Golgin Coiled-Coil Proteins of the Golgi Apparatus. *Cold Spring Harbor Perspectives in Biology*, 3(6):a005256–a005256. doi:10.1101/cshperspect.a005256.

Munro, S. (2013). Open questions: What is there left for cell biologists to do? *BMC Biology*, 11(1):16. doi:10.1186/1741-7007-11-16.

Munro, S. and Nichols, B.J. (1999). The grip domain - A novel golgi-targeting domain found in several coiled-coil proteins. *Curr. Biol.*, 9(7):377–380. doi:10.1016/S0960-9822(99)80166-3.

Murphy, R.F. (1991). Maturation models for endosome and lysosome biogenesis. *Trends in Cell Biology*, 1(4):77–82. doi:10.1016/0962-8924(91)90022-2.

Murungi, E., Barlow, L.D., Venkatesh, D., Adung'a, V.O., Dacks, J.B., Field, M.C., and Christoffels, A. (2014). A comparative analysis of trypanosomatid SNARE proteins. *Parasitology International*, 63(2):341–348. doi:10.1016/j.parint.2013.11.002.

Nevin, W.D. and Dacks, J.B. (2009). Repeated secondary loss of adaptin complex genes in the Apicomplexa. *Parasitology International*, 58(1):86–94. doi:10.1016/j.parint.2008.12.002.

Newman, A.P., Shim, J., and Ferro-Novick, S. (1990). BET1, BOSI, and SEC22 Are Members of a Group of Interacting Yeast Genes Required for Transport from the Endoplasmic Reticulum to the Golgi Complex. *MOL. CELL. BIOL.*, 10:10.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1):268–274. doi:10.1093/molbev/msu300.

Nichols, B.J., Ungermann, C., Pelham, H.R.B., Wickner, W.T., and Haas, A. (1997). Homotypic vacuolar fusion mediated by t- and v-SNAREs. *Nature*, 387(6629):199–202. doi:10.1038/387199a0.

Novick, P., Ferro, S., and Schekman, R. (1981). Order of events in the yeast secretory pathway. *Cell*, 25(2):461–469. doi:10.1016/0092-8674(81)90064-7.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rosselló, J.A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T.R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., and Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451):579–584. doi:10.1038/nature12211.

Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdy, I., Paces, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc. Natl. Acad. Sci. U. S. A.*, 110(18):7371–6. doi:10.1073/pnas.1219590110.

Oas, T.G. and Endow, S.A. (1994). Springs and hinges: Dynamic coiled-coils and discontinuities. *Trends in Biochemical Sciences*, 19(2):51–54.

Obado, S.O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N.E., Chait, B.T., Field, M.C., and Rout, M.P. (2016). Interactome Mapping Reveals the Evolutionary History of the Nuclear Pore Complex. *PLOS Biology*, 14(2):e1002365. doi:10.1371/journal.pbio.1002365.

Orci, L., Ravazzola, M., Volchuk, A., Engel, T., Gmachl, M., Amherdt, M., Perrelet, A., Sollner, T.H., and Rothman, J.E. (2000). Anterograde flow of cargo across the Golgi stack potentially mediated via bidirectional "percolating" COPI vesicles. *Proceedings of the National Academy of Sciences*, 97(19):10400–10405. doi:10.1073/pnas.190292497.

Osterrieder, A. (2012). Tales of tethers and tentacles: Golgins in plants. *Journal of Microscopy*, 247(1):68–77. doi:10.1111/j.1365-2818.2012.03620.x.

Paek, I., Orci, L., Ravazzola, M., Erdjument-Bromage, H., Amherdt, M., Tempst, P., Söllner, T.H., and Rothman, J.E. (1997). ERS-24, a Mammalian v-SNARE Implicated in Vesicle Traffic between the ER and the Golgi. *The Journal of Cell Biology*, 137(5):1017–1028. doi:10.1083/jcb.137.5.1017.

Panchy, N., Lehti-Shiu, M.D., and Shiu, S.H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, page pp.00523.2016. doi:10.1104/pp.16.00523.

Panic, B., Whyte, J.R.C., and Munro, S. (2003). The ARF-like GTPases Arl1p and Arl3p act in a pathway that interacts with vesicle-tethering factors at the Golgi apparatus. *Curr. Biol.*, 13(5):405–410. doi:10.1016/S0960-9822(03)00091-5.

Pardi, F. and Gascuel, O. (2016). Distance-based methods in phylogenetics. In R.M. Kliman, editor, *Encyclopedia of Evolutionary Biology*, pages 458–465. Elsevier, Academic Press, Amsterdam Boston Heidelberg, 1 edition. ISBN 978-0-12-800049-6.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33):13624–13629. doi:10.1073/pnas.1110633108.

Parikh, A., Miranda, E.R., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., Curk, T., Sucgang, R., Chen, R., Zupan, B., Loomis, W.F., Kuspa, A., and Shaulsky, G. (2010). Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biology*, 11(3):R35. doi:10.1186/gb-2010-11-3-r35.

Park, M., Song, K., Reichardt, I., Kim, H., Mayer, U., Stierhof, Y.D., Hwang, I., and Jurgens, G. (2013). Arabidopsis -adaptin subunit AP1M of adaptor protein complex 1 mediates late secretory and vacuolar traffic and is required for growth. *Proceedings of the National Academy of Sciences*, 110(25):10318–10323. doi:10.1073/pnas.1300460110.

Parlati, F., McNew, J.A., Fukuda, R., Miller, R., Söllner, T.H., and Rothman, J.E. (2000). Topological restriction of SNARE-dependent membrane fusion. *Nature*, 407(6801):194–198. doi:10.1038/35025076.

Parsons, H.T. and Lilley, K.S. (2017). Mass spectrometry approaches to study plant endomembrane trafficking. *Semin. Cell Dev. Biol.* doi:10.1016/j.semcdb.2017.10.014.

Paumet, F., Rahimian, V., and Rothman, J.E. (2004). The specificity of SNARE-dependent fusion is encoded in the SNARE motif. *Proceedings of the National Academy of Sciences*, 101(10):3376–3380. doi:10.1073/pnas.0400271101.

Pearse, B.M. (1976). Clathrin: A unique protein associated with intracellular transfer of membrane by coated vesicles. *Proceedings of the National Academy of Sciences*, 73(4):1255–1259. doi:10.1073/pnas.73.4.1255.

Peden, A.A., Oorschot, V., Hesser, B.A., Austin, C.D., Scheller, R.H., and Klumperman, J. (2004). Localization of the AP-3 adaptor complex defines a novel endosomal exit site for lysosomal membrane proteins. *The Journal of Cell Biology*, 164(7):1065–1076. doi:10.1083/jcb.200311064.

Peden, A.A., Park, G.Y., and Scheller, R.H. (2001). The Di-leucine Motif of Vesicle-associated Membrane Protein 4 Is Required for Its Localization and AP-1 Binding. *Journal of Biological Chemistry*, 276(52):49183–49187. doi:10.1074/jbc.M106646200.

Peden, A.A., Rudge, R.E., Lui, W.W.Y., and Robinson, M.S. (2002). Assembly and function of AP-3 complexes in cells expressing mutant subunits. *The Journal of Cell Biology*, 156(2):327–336. doi:10.1083/jcb.200107140.

Petrželková, R. and Eliáš, M. (2014). Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida. *Acta Societatis Botanicorum Poloniae*, 83(4):303–315. doi:10.5586/asbp.2014.052.

Philippe, H. and Germot, A. (2000). Phylogeny of Eukaryotes Based on Ribosomal RNA: Long-Branch Attraction and Models of Sequence Evolution. *Molecular Biology and Evolution*, 17(5):830–834. doi:10.1093/oxfordjournals.molbev.a026362.

Pieperhoff, M.S., Schmitt, M., Ferguson, D.J.P., and Meissner, M. (2013). The role of clathrin in post-Golgi trafficking in *Toxoplasma gondii*. *PloS One*, 8(10):e77620. doi:10.1371/journal.pone.0077620.

Pittis, A.A. and Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*, 531(7592):101–104. doi:10.1038/nature16941.

Plattner, H. (2010). How to Design a Highly Organized Cell: An Unexpectedly High Number of Widely Diversified SNARE Proteins Positioned at Strategic Sites in the Ciliate, *Paramecium tetraurelia*. *Protist*, 161(4):497–516. doi:10.1016/j.protis.2010.05.001.

Pobbati, A.V., Razeto, A., Böddener, M., Becker, S., and Fasshauer, D. (2004). Structural Basis for the Inhibitory Role of Tomosyn in Exocytosis. *Journal of Biological Chemistry*, 279(45):47192–47200. doi:10.1074/jbc.M408767200.

Ponting, C.P. (2001). Issues in predicting protein function from sequence. *Briefings in Bioinformatics*, 2(1):19–29. doi:10.1093/bib/2.1.19.

Posada, D. and Buckley, T.R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, 53(5):793–808. doi:10.1080/10635150490522304.

Prekeris, R., Klumperman, J., Chen, Y.A., and Scheller, R.H. (1998). Syntaxin 13 Mediates Cycling of Plasma Membrane Proteins via Tubulovesicular Recycling Endosomes. *The Journal of Cell Biology*, 143(4):957–971. doi:10.1083/jcb.143.4.957.

Prekeris, R., Yang, B., Oorschot, V., Klumperman, J., and Scheller, R.H. (1999). Differential Roles of Syntaxin 7 and Syntaxin 8 in Endosomal Trafficking. *Molecular Biology of the Cell*, 10(11):3891–3908. doi:10.1091/mbc.10.11.3891.

Prescianotto-Baschong, C. and Riezman, H. (2002). Ordering of Compartments in the Yeast Endocytic Pathway: Order of Yeast Endocytic Compartments. *Traffic*, 3(1):37–49. doi:10.1034/j.1600-0854.2002.30106.x.

Protopopov, V., Govindan, B., Novick, P., and Gerst, J.E. (1993). Homologs of the Synaptobrevin/VAMP Family of Synaptic Vesicle Proteins Function on the Late Secretory Pathway in *S. cerevisiae*. *Cell*, 74(September 10):855–861.

Pryor, P.R., Jackson, L., Gray, S.R., Edeling, M.A., Thompson, A., Sanderson, C.M., Evans, P.R., Owen, D.J., and Luzio, J.P. (2008). Molecular Basis for the Sorting of the SNARE VAMP7 into Endocytic Clathrin-Coated Vesicles by the ArfGAP Hrb. *Cell*, 134(5):817–827. doi:10.1016/j.cell.2008.07.023.

Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D., and Donoghue, P.C. (2018). The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. *Current Biology*, 28(5):733–745.e2. doi:10.1016/j.cub.2018.01.063.

Quiroga, R., Trenchi, A., Gonzalez Montoro, A., Valdez Taubas, J., and Maccioni, H.J.F. (2013). Short transmembrane domains with high-volume exoplasmic halves determine retention of Type II membrane proteins in the Golgi complex. *Journal of Cell Science*, 126(23):5344–5349. doi:10.1242/jcs.130658.

Rabinovich, E., Kerem, A., Frohlich, K.U., Diamant, N., and Bar-Nun, S. (2002). AAA-ATPase p97/Cdc48p, a Cytosolic Chaperone Required for Endoplasmic Reticulum-Associated Protein Degradation. *MOL. CELL. BIOL.*, 22:9.

Rad, M.R., Phan, H.L., Kirchrath, L., Tan, P.K., Kirchhausen, T., Hollenberg, C.P., and Payne, G.S. (1995). *Saccharomyces Cerevisiae* Apl2p, a homologue of the mammalian clathrin AP beta subunit, plays a role in clathrin-dependent Golgi functions. *Journal of Cell Science*, 108 (Pt 4):1605–1615.

Raingo, J., Khvotchev, M., Liu, P., Darios, F., Li, Y.C., Ramirez, D.M.O., Adachi, M., Lemieux, P., Toth, K., Davletov, B., and Kavalali, E.T. (2012). VAMP4 directs synaptic vesicles to a pool that selectively maintains asynchronous neurotransmission. *Nature Neuroscience*, 15(5):738–745. doi:10.1038/nn.3067.

Ramadas, R. and Thattai, M. (2013). New Organelles by Gene Duplication in a Biophysical Model of Eukaryote Endomembrane Evolution. *Biophysical Journal*, 104(11):2553–2563. doi:10.1016/j.bpj.2013.03.066.

Ramirez, I.B.R., de Graffenried, C.L., Ebersberger, I., Yelinek, J., He, C.Y., Price, A., and Warren, G. (2008). TbG63, a golgin involved in Golgi architecture in *Trypanosoma brucei*. *J. Cell Sci.*, 121:1538–1546. doi:10.1242/jcs.014324.

Ramirez, I.B.R. and Lowe, M. (2009). Golgins and GRASPs: Holding the Golgi together. *Semin. Cell Dev. Biol.*, 20(7):770–9. doi:10.1016/j.semcdb.2009.03.011.

Ramirez-Macias, I., Barlow, L.D., Anton, C., Spang, A., Roncero, C., and Dacks, J.B. (2018). Evolutionary cell biology traces the rise of the exomer complex in Fungi from an ancient eukaryotic component. *Scientific Reports*, 8(1):11154. doi:10.1038/s41598-018-29416-4.

Ranea, J.A.G., Yeats, C., Grant, A., and Orengo, C.A. (2007). Predicting Protein Function with Hierarchical Phylogenetic Profiles: The Gene3D Phylo-Tuner Method Applied to Eukaryotic Genomes. *PLoS Computational Biology*, 3(11):13.

Rayner, J.C. and Pelham, H.R. (1997). Transmembrane domain-dependent sorting of proteins to the ER and plasma membrane in yeast. *The EMBO Journal*, 16(8):1832–1841. doi:10.1093/emboj/16.8.1832.

Redman, C., Siekevitz, P., and Palade, G. (1966). Synthesis and Transfer of Amylase in Pigeon Pancreatic Microsomes. *The Journal of Biological Chemistry*, 241(5):1150–1158.

Reggiori, F., Black, M.W., and Pelham, H.R.B. (2000). Polar Transmembrane Domains Target Proteins to the Interior of the Yeast Vacuole. *Molecular Biology of the Cell*, 11(11):3737–3749. doi:10.1091/mbc.11.11.3737.

Rehling, P., Darsow, T., Katzmann, D.J., and Emr, S.D. (1999). Formation of AP-3 transport intermediates requires Vps41 function. *Nature Cell Biology*, 1(6):346–353. doi:10.1038/14037.

Renna, L., Hanton, S.L., Stefano, G., Bortolotti, L., Misra, V., and Brandizzi, F. (2005). Identification and characterization of AtCASP, a plant transmembrane Golgi matrix protein. *Plant Molecular Biology*, 58(1):109–122. doi:10.1007/s11103-005-4618-4.

Rensing, S.A. (2017). Why we need more non-seed plant models. *New Phytologist*, 216(2):355–360. doi:10.1111/nph.14464.

Rink, J., Ghigo, E., Kalaidzidis, Y., and Zerial, M. (2005). Rab Conversion as a Mechanism of Progression from Early to Late Endosomes. *Cell*, 122(5):735–749. doi:10.1016/j.cell.2005.06.043.

Rivero, S., Cardenas, J., Bornens, M., and Rios, R.M. (2009). Microtubule nucleation at the cis-side of the Golgi apparatus requires AKAP450 and GM130. *EMBO J.*, 28(8):1016–1028. doi:10.1038/emboj.2009.47.

Robinson, M.S. (2015). Forty Years of Clathrin-coated Vesicles: Forty Years of Clathrin-coated Vesicles. *Traffic*, 16(12):1210–1238. doi:10.1111/tra.12335.

Robinson, M.S., Sahlender, D.A., and Foster, S.D. (2010). Rapid inactivation of proteins by rapamycin-induced rerouting to mitochondria. *Developmental Cell*, 18(2):324–331. doi:10.1016/j.devcel.2009.12.015.

Roboti, P., Sato, K., and Lowe, M. (2015). The golgin GMAP-210 is required for efficient membrane trafficking in the early secretory pathway. *J. Cell Sci.*, 128(8):1595–606. doi:10.1242/jcs.166710.

Rodríguez-Ezpeleta, N. and Embley, T.M. (2012). The SAR11 Group of Alpha-Proteobacteria Is Not Related to the Origin of Mitochondria. *PLoS ONE*, 7(1):e30520. doi:10.1371/journal.pone.0030520.

Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current Biology*, 27(21):R1177–R1192. doi:10.1016/j.cub.2017.09.015.

Rohena, L., Neidich, J., Truitt Cho, M., Gonzalez, K.D., Tang, S., Devinsky, O., and Chung, W.K. (2013). Mutation in *SNAP25* as a novel genetic cause of epilepsy and intellectual disability. *Rare Diseases*, 1(1):e26314. doi:10.4161/rdis.26314.

Ronquist, F. and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574. doi:10.1093/bioinformatics/btg180.

Rossi, V., Banfield, D., Vacca, M., Dietrich, L., Ungermann, C., Desposito, M., Galli, T., and Filippini, F. (2004). Longins and their longin domains: Regulated SNAREs and multifunctional SNARE regulators. *Trends in Biochemical Sciences*, 29(12):682–688. doi:10.1016/j.tibs.2004.10.002.

Rothman, J.E. (1987). Protein sorting by selective retention in the endoplasmic reticulum and Golgi stack. *Cell*, 50(4):521–522. doi:10.1016/0092-8674(87)90024-9.

Rothman, J.E. (1994). Mechanisms of intracellular protein transport. *Nature*, 372(6501):55–63. doi:10.1038/372055a0.

Rothman, J.E. (2019). Jim's View: Is the Golgi stack a phase-separated liquid crystal? *FEBS Letters*, pages 1873–3468.13609. doi:10.1002/1873-3468.13609.

Rothman, J.E. and Fine, R.E. (1980). Coated vesicles transport newly synthesized membrane glycoproteins from endoplasmic reticulum to plasma membrane in two successive stages. *Proceedings of the National Academy of Sciences*, 77(2):780–784. doi:10.1073/pnas.77.2.780.

Rothman, J.E. and Warren, G. (1994). Implications of the SNARE hypothesis for intracellular membrane topology and dynamics. *Current Biology*, 4(3):220–233. doi:10.1016/S0960-9822(00)00051-8.

Sagan, L. (1967). On the Origin of Mitosing Cells. *Journal of Theoretical Biology*, 14:225–274.

Salman-Minkov, A., Sabath, N., and Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. *Nature Plants*, 2(8). doi:10.1038/nplants.2016.115.

Sanderfoot, A. (2007). Increases in the Number of SNARE Genes Parallels the Rise of Multicellularity among the Green Plants. *PLANT PHYSIOLOGY*, 144(1):6–17. doi:10.1104/pp.106.092973.

Sanderfoot, A.A., Assaad, F.F., and Raikhel, N.V. (2000). The *Arabidopsis* Genome. An Abundance of Soluble *N*-Ethylmaleimide-Sensitive Factor Adaptor Protein Receptors. *Plant Physiology*, 124(4):1558–1569. doi:10.1104/pp.124.4.1558.

Sanderfoot, A.A., Kovaleva, V., Bassham, D.C., and Raikhel, N.V. (2001). Interactions between Syntaxins Identify at Least Five SNARE Complexes within the Golgi/Prevacuolar System of the *Arabidopsis* Cell. *Molecular Biology of the Cell*, 12(12):3733–3743. doi:10.1091/mbc.12.12.3733.

Sato, M.H., Nakamura, N., Ohsumi, Y., Kouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., and Wada, Y. (1997). The *AtVAM3* Encodes a Syntaxin-related Molecule Implicated in the Vacuolar Assembly in *Arabidopsis thaliana*. *Journal of Biological Chemistry*, 272(39):24530–24535. doi:10.1074/jbc.272.39.24530.

Sato, T.K., Darsow, T., and Emr, S.D. (1998). Vam7p, a SNAP-25-Like Molecule, and Vam3p, a Syntaxin Homolog, Function Together in Yeast Vacuolar Protein Trafficking. *Molecular and Cellular Biology*, 18(9):5308–5319. doi:10.1128/MCB.18.9.5308.

Scales, S.J., Hesser, B.A., Masuda, E.S., and Scheller, R.H. (2002). Amisyn, a Novel Syntaxin-binding Protein That May Regulate SNARE Complex Assembly. *Journal of Biological Chemistry*, 277(31):28271–28279. doi:10.1074/jbc.M204929200.

Scheuring, D., Viotti, C., Krüger, F., Künzl, F., Sturm, S., Bubeck, J., Hillmer, S., Frigerio, L., Robinson, D.G., Pimpl, P., and Schumacher, K. (2011). Multivesicular Bodies Mature from the *Trans*-Golgi Network/Early Endosome in *Arabidopsis*. *The Plant Cell*, 23(9):3463–3481. doi:10.1105/tpc.111.086918.

Schilde, C., Lutter, K., Kissmehl, R., and Plattner, H. (2008). Molecular Identification of a SNAP-25-Like SNARE Protein in *Paramecium*. *Eukaryotic Cell*, 7(8):1387–1402. doi:10.1128/EC.00012-08.

Schilde, C., Schönemann, B., Sehring, I.M., and Plattner, H. (2010). Distinct Subcellular Localization of a Group of Synaptobrevin-Like SNAREs in *Paramecium tetraurelia* and Effects of Silencing SNARE-Specific Chaperone NSF. *Eukaryotic Cell*, 9(2):288–305. doi:10.1128/EC.00220-09.

Schilde, C., Wassmer, T., Mansfeld, J., Plattner, H., and KissmehI, R. (2006). A Multigene Family Encoding R-SNAREs in the Ciliate *Paramecium tetraurelia*: The R-SNAREs of *Paramecium*. *Traffic*, 7(4):440–455. doi:10.1111/j.1600-0854.2006.00397.x.

Schlacht, A. and Dacks, J.B. (2015). Unexpected ancient paralogues and an evolutionary model for the COPII coat complex. *Genome Biol. Evol.*, 7(4):1098–1109. doi:10.1093/gbe/evv045.

Schlacht, A., Herman, E.K., Klute, M.J., Field, M.C., and Dacks, J.B. (2014). Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System. *Cold Spring Harbor Perspectives in Biology*, 6(10):a016048–a016048. doi:10.1101/cshperspect.a016048.

Schlacht, A., Mowbrey, K., Elias, M., Kahn, R.A., and Dacks, J.B. (2013). Ancient Complexity, Opisthokont Plasticity, and Discovery of the 11th Subfamily of Arf GAP Proteins: Arf GAP Evolution. *Traffic*, 14(6):636–649. doi:10.1111/tra.12063.

Schledzewski, K., Brinkmann, H., and 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. *Journal of Molecular Evolution*, 48(6):770–778.

Schlüter, A., Fourcade, S., Ripp, R., Mandel, J.L., Poch, O., and Pujol, A. (2006). The Evolutionary Origin of Peroxisomes: An ER-Peroxisome Connection. *Molecular Biology and Evolution*, 23(4):838–845. doi:10.1093/molbev/msj103.

Schmid, S. (1988). Two distinct subpopulations of endosomes involved in membrane recycling and transport to lysosomes. *Cell*, 52(1):73–83. doi:10.1016/0092-8674(88)90532-6.

Schoberer, J. and Strasser, R. (2011). Sub-Compartmental Organization of Golgi-Resident N-Glycan Processing Enzymes in Plants. *Molecular Plant*, 4(2):220–228. doi:10.1093/mp/ssq082.

Schroeder, C.C. (1990). Vesicle transport along microtubular ribbons and isolation of cytoplasmic dynein from *Paramecium*. *The Journal of Cell Biology*, 111(6):2553–2562. doi:10.1083/jcb.111.6.2553.

Schroer, T.A. (1988). The role of kinesin and other soluble factors in organelle movement along microtubules. *The Journal of Cell Biology*, 107(5):1785–1792. doi:10.1083/jcb.107.5.1785.

Seals, D.F., Eitzen, G., Margolis, N., Wickner, W.T., and Price, A. (2000). A Ypt/Rab effector complex containing the Sec1 homolog Vps33p is required for homotypic vacuole fusion. *Proceedings of the National Academy of Sciences*, 97(17):9402–9407. doi:10.1073/pnas.97.17.9402.

Seaman, M.N., Michael McCaffery, J., and Emr, S.D. (1998). A Membrane Coat Complex Essential for Endosome-to-Golgi Retrograde Transport in Yeast. *The Journal of Cell Biology*, 142(3):665–681. doi:10.1083/jcb.142.3.665.

Seaman, M.N.J. (2004). Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *J. Cell Biol.*, 165(1):111–122. doi:10.1083/jcb.200312034.

Setty, S.R.G., Shin, M.E., Yoshino, A., Marks, M.S., and Burd, C.G. (2003). Golgi Recruitment of GRIP Domain Proteins by Arf-like GTPase 1 Is Regulated by Arf-like GTPase 3. *Current Biology*, 13(5):401–404. doi:10.1016/S0960-9822(03)00089-7.

Shen, J., Tareste, D.C., Paumet, F., Rothman, J.E., and Melia, T.J. (2007). Selective Activation of Cognate SNAREpins by Sec1/Munc18 Proteins. *Cell*, 128(1):183–195. doi:10.1016/j.cell.2006.12.016.

Shen, X.M., Selcen, D., Brengman, J., and Engel, A.G. (2014). Mutant SNAP25B causes myasthenia, cortical hyperexcitability, ataxia, and intellectual disability. *Neurology*, 83(24):2247–2255. doi:10.1212/WNL.0000000000001079.

Shi, X., Halder, P., Yavuz, H., Jahn, R., and Shuman, H.A. (2016). Direct targeting of membrane fusion by SNARE mimicry: Convergent evolution of *Legionella* effectors. *Proceedings of the National Academy of Sciences*, 113(31):8807–8812. doi:10.1073/pnas.1608755113.

Shimada, T., Kunieda, T., Sumi, S., Koumoto, Y., Tamura, K., Hatano, K., Ueda, H., and Hara-Nishimura, I. (2018). The AP-1 Complex is Required for Proper Mucilage Formation in Arabidopsis Seeds. *Plant and Cell Physiology*. doi:10.1093/pcp/pcy158.

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3):492–508. doi:10.1080/10635150290069913.

Shimodaira, H. and Hasegawa, M. (2001). CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics*, 17(12):1246–1247. doi:10.1093/bioinformatics/17.12.1246.

Shohdy, N., Efe, J.A., Emr, S.D., and Shuman, H.A. (2005). Pathogen effector protein screening in yeast identifies *Legionella* factors that interfere with membrane trafficking. *Proceedings of the National Academy of Sciences*, 102(13):4866–4871. doi:10.1073/pnas.0501315102.

Short, B., Preisinger, C., Korner, R., Kopajtich, R., Byron, O., and Barr, F.A. (2001). A GRASP55-rab2 effector complex linking Golgi structure to membrane traffic. *J. Cell Biol.*, 155(6):877–883. doi:10.1083/jcb.200108079.

Shorter, J., Beard, M.B., Seemann, J., Dirac-Svejstrup, A.B., and Warren, G. (2002). Sequential tethering of Golgins and catalysis of SNAREpin assembly by the vesicle-tethering protein p115. *The Journal of Cell Biology*, 157(1):45–62. doi:10.1083/jcb.200112127.

Shotland, Y., Krämer, H., and Groisman, E.A. (2003). The *Salmonella* SpiC protein targets the mammalian Hook3 protein function to alter cellular trafficking. *Molecular Microbiology*, 49(6):1565–1576. doi:10.1046/j.1365-2958.2003.03668.x.

Shu-Miaw Chaw, Chien-Chang Chang, Hsin-Liang Chen, and Wen-Hsiung Li (2004). Dating the Monocot-Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. *Journal of Molecular Evolution*, 58(4):424–441. doi:10.1007/s00239-003-2564-9.

Sicheritz-Ponten, T., Kurland, C.G., and Andersson, S.G.E. (1998). A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. *Biochimica et Biophysica Acta*, 1365:545–551.

Sievers, F., Dineen, D., Wilm, A., and Higgins, D.G. (2013). Making automated multiple alignments of very large numbers of protein sequences. *Bioinformatics*, 29(8):989–995. doi:10.1093/bioinformatics/btt093.

Simonsen, A., Gaullier, J.M., D'Arrigo, A., and Stenmark, H. (1999). The Rab5 Effector EEA1 Interacts Directly with Syntaxin-6. *Journal of Biological Chemistry*, 274(41):28857–28860. doi:10.1074/jbc.274.41.28857.

Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., and Stenmark, H. (1998). EEA1 links PI(3)K function to Rab5 regulation of endosome fusion. *Nature*, 394(6692):494–498. doi:10.1038/28879.

Simpson, A. (2004a). Early Evolution within Kinetoplastids (Euglenozoa), and the Late Emergence of Trypanosomatids. *Protist*, 155(4):407–422. doi:10.1078/1434461042650389.

Simpson, F., Bright, N.A., West, M.A., Newman, L.S., Darnell, R.B., and Robinson, M.S. (1996). A novel adaptor-related protein complex. *The Journal of Cell Biology*, 133(4):749–760. doi:10.1083/jcb.133.4.749.

Simpson, L.L. (2004b). Identification of the major steps in botulinum toxin action. *Annual Review of Pharmacology and Toxicology*, 44(1):167–193. doi:10.1146/annurev.pharmtox.44.101802.121554.

Singh, M.K. and Jürgens, G. (2018). Specificity of plant membrane trafficking – ARFs, regulators and coat proteins. *Seminars in Cell & Developmental Biology*, 80:85–93. doi:10.1016/j.semcdb.2017.10.005.

Słabicki, M., Theis, M., Krastev, D.B., Samsonov, S., Mundwiller, E., Junqueira, M., Paszkowski-Rogacz, M., Teyra, J., Heninger, A.K., Poser, I., Prieur, F., Truchetto, J., Confavreux, C., Marelli, C., Durr, A., Camdessanche, J.P., Brice, A., Shevchenko, A., Pisabarro, M.T., Stevanin, G., and Buchholz, F. (2010). A genome-scale DNA repair RNAi screen identifies SPG48 as a novel gene associated with hereditary spastic paraplegia. *PLoS biology*, 8(6):e1000408. doi:10.1371/journal.pbio.1000408.

Slamovits, C.H., Burri, L., and Keeling, P.J. (2006). Characterization of a divergent Sec61beta gene in microsporidia. *Journal of Molecular Biology*, 359(5):1196–1202. doi:10.1016/j.jmb.2006.04.028.

Sloutsky, R. and Naegle, K.M. (2016). High-Resolution Identification of Specificity Determining Positions in the LacI Protein Family Using Ensembles of Sub-Sampled Alignments. *PLOS ONE*, 11(9):e0162579. doi:10.1371/journal.pone.0162579.

Smits, P., Bolton, A.D., Funari, V., Hong, M., Boyden, E.D., Lu, L., Manning, D.K., Dwyer, N.D., Moran, J.L., Prysak, M., Merriman, B., Nelson, S.F., Bonafé,

L., Superti-Furga, A., Ikegawa, S., Krakow, D., Cohn, D.H., Kirchhausen, T., Warman, M.L., and Beier, D.R. (2010). Lethal skeletal dysplasia in mice and humans lacking the golgin GMAP-210. *N. Engl. J. Med.*, 362(3):206–16. doi:10.1056/NEJMoa0900158.

Sogin, M. (1997). History assignment: When was the mitochondrion founded? *Current Opinion in Genetics & Development*, 7(6):792–799.

Sohda, M., Misumi, Y., Yamamoto, a., Yano, a., Nakamura, N., and Ikehara, Y. (2001). Identification and characterization of a novel Golgi protein, GCP60, that interacts with the integral membrane protein giantin. *J. Biol. Chem.*, 276(48):45298–306. doi:10.1074/jbc.M108961200.

Söllner, T., Bennett, M.K., Whiteheart, S.W., Scheller, R.H., and Rothman, J.E. (1993a). A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell*, 75(3):409–418. doi:10.1016/0092-8674(93)90376-2.

Söllner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J.E. (1993b). SNAP receptors implicated in vesicle targeting and fusion. *Nature*, 362(25 March):318–324.

Sonnhammer, E.L. and Koonin, E.V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics*, 18(12):619–620. doi:10.1016/S0168-9525(02)02793-2.

Spang, A., Eme, L., Saw, J.H., Caceres, E.F., Zaremba-Niedzwiedzka, K., Lombard, J., Guy, L., and Ettema, T.J.G. (2017). Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLOS Genetics*.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(May 14):173–179.

Sprecher, E., Ishida-Yamamoto, A., Mizrahi-Koren, M., Rapaport, D., Goldsher, D., Indelman, M., Topaz, O., Chefetz, I., Keren, H., O'Brien, T.J., Bercovich, D., Shalev, S., Geiger, D., Bergman, R., Horowitz, M., and Mandel, H. (2005). A Mutation in SNAP29, Coding for a SNARE Protein Involved in Intracellular Trafficking, Causes a Novel Neurocutaneous Syndrome Characterized by Cerebral Dysgenesis, Neuropathy, Ichthyosis, and Palmoplantar Keratoderma. *The American Journal of Human Genetics*, 77(2):242–251. doi:10.1086/432556.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)*, 22(21):2688–2690. doi:10.1093/bioinformatics/btl446.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Stamnes, M.A. and Rothman, J.E. (1993). The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein. *Cell*, 73(5):999–1005. doi:10.1016/0092-8674(93)90277-W.

Steegmaier, M., Klumperman, J., Foletti, D.L., Yoo, J.S., and Scheller, R.H. (1999). Vesicle-associated Membrane Protein 4 is Implicated in *Trans*-Golgi Network Vesicle Trafficking. *Molecular Biology of the Cell*, 10(6):1957–1972. doi:10.1091/mbc.10.6.1957.

Stelly, N., Mauger, J.P., Claret, M., and Adoutte, A. (1991). Cortical alveoli of *Paramecium*: A vast submembranous calcium storage compartment. *The Journal of Cell Biology*, 113(1):103–112. doi:10.1083/jcb.113.1.103.

Steneck, R.S., Graham, M.H., Bourque, B.J., Corbett, D., Erlandson, J.M., Estes, J.A., and Tegner, M.J. (2002). Kelp forest ecosystems: Biodiversity, stability, resilience and future. *Environmental Conservation*, 29(4):436–459. doi:10.1017/S0376892902000322.

Stenmark, H. and Olkkonen, V.M. (2001). The Rab GTPase family. *Genome Biology*, 2(5):reviews3007.1–3007.7.

Stoorvogel, W., Strous, G.J., Geuze, H.J., Oorschot, V., and Schwartzt, A.L. (1991). Late endosomes derive from early endosomes by maturation. *Cell*, 65(3):417–427. doi:10.1016/0092-8674(91)90459-C.

Strimmer, K. and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1487):137–142. doi:10.1098/rspb.2001.1862.

Struck, N.S., de Souza Dias, S., Langer, C., Marti, M., Pearce, J.A., Cowman, A.F., and Gilberger, T.W. (2005). Re-defining the Golgi complex in *Plasmodium falciparum* using the novel Golgi marker PfGRASP. *J. Cell Sci.*, 118(Pt 23):5603–13. doi:10.1242/jcs.02673.

Subramaniam, V.N., Loh, E., Horstmann, H., Habermann, A., Xu, Y., Coe, J., Griffiths, G., and Hong, W. (2000). Localization of syntaxin 8 to the early endosome. *Journal of Cell Science*, 113:997–1008.

Subramaniam, V.N., Peter, F., Philp, R., Wong, S.H., and Hong, W. (1996). GS28, a 28-Kilodalton Golgi SNARE That Participates in ER-Golgi Transport. *Science*, 272(5265):1161–1163. doi:10.1126/science.272.5265.1161.

Sudhof, T.C., Baumert, M., Perin, M.S., and Jahn, R. (1989). A Synaptic Vesicle Membrane Protein Is Conserved from Mammals to Drosophila. *Neuron*, 2(May):1475–1481.

Sugano, S., Hayashi, N., Kawagoe, Y., Mochizuki, S., Inoue, H., Mori, M., Nishizawa, Y., Jiang, C.J., Matsui, M., and Takatsuji, H. (2016). Rice OsVAMP714, a membrane-trafficking protein localized to the chloroplast and vacuolar membrane, is involved in resistance to rice blast disease. *Plant Molecular Biology*, 91(1-2):81–95. doi:10.1007/s11103-016-0444-0.

Sugiura, A., McLelland, G.L., Fon, E.A., and McBride, H.M. (2014). A new pathway for mitochondrial quality control: Mitochondrial-derived vesicles. *The EMBO Journal*, 33(19):2142–2156. doi:10.15252/embj.201488104.

Surkont, J. and Pereira-Leal, J.B. (2015). Evolutionary Patterns in Coiled-Coils. *Genome Biology and Evolution*, 7(2):545–556. doi:10.1093/gbe/evv007.

Sutton, R.B., Fasshauer, D., Jahn, R., and Brunger, A.T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature*, 395(6700):347–353. doi:10.1038/26412.

Suwastika, I.N., Uemura, T., Shiina, T., H. Sato, M., and Takeyasu, K. (2008). SYP71, a Plant-specific Qc-SNARE Protein, Reveals Dual Localization to the Plasma Membrane and the Endoplasmic Reticulum in *Arabidopsis*. *Cell Structure and Function*, 33(2):185–192. doi:10.1247/csf.08024.

Synek, L., Schlager, N., Eliáš, M., Quentin, M., Hauser, M.T., and Žárský, V. (2006). AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. *The Plant Journal*, 48(1):54–72. doi:10.1111/j.1365-313X.2006.02854.x.

Tai, G., Lu, L., Wang, T.L., Tang, B.L., Goud, B., Johannes, L., and Hong, W. (2004). Participation of the Syntaxin 5/Ykt6/GS28/GS15 SNARE Complex in Transport from the Early/Recycling Endosome to the Trans-Golgi Network. *Molecular Biology of the Cell*, 15:12.

Takahashi, M., Shibata, H., Shimakawa, M., Miyamoto, M., Mukai, H., and Yoshitaka, O. (1999). Characterization of a novel giant scaffolding protein, CG-NAP, that anchors multiple signaling enzymes to centrosome and the Golgi apparatus. *J. Biol. Chem.*, 274(24):17267–17274. doi:10.1074/jbc.274.24.17267.

Takishita, K., Kolisko, M., Komatsuzaki, H., Yabuki, A., Inagaki, Y., Cepicka, I., Smejkalová, P., Silberman, J.D., Hashimoto, T., Roger, A.J., and Simpson, A.G. (2012). Multigene Phylogenies of Diverse *Carpediemonas*-like Organisms Identify the Closest Relatives of 'Amitochondriate' Diplomonads and Retortamonads. *Protist*, 163(3):344–355. doi:10.1016/j.protis.2011.12.007.

Takvorian, P.M., Buttle, K.F., Mankus, D., Mannella, C.A., Weiss, L.M., and Cali, A. (2013). The multilayered interlaced network (MIN) in the sporoplasm of the microsporidium *Anncaliia algerae* is derived from Golgi. *The Journal of Eukaryotic Microbiology*, 60(2):166–178. doi:10.1111/jeu.12019.

Takvorian, P.M. and Cali, A. (1994). Enzyme histochemical identification of the Golgi apparatus in the microsporidian, *Glugea stephani*. *The Journal of Eukaryotic Microbiology*, 41(5):63S–64S.

Tatusov, R.L. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36. doi: 10.1093/nar/28.1.33.

Teh, O.K., Shimono, Y., Shirakawa, M., Fukao, Y., Tamura, K., Shimada, T., and Hara-Nishimura, I. (2013). The AP-1 μ Adaptin is Required for KNOLLE Localization at the Cell Plate to Mediate Cytokinesis in *Arabidopsis*. *Plant and Cell Physiology*, 54(6):838–847. doi:10.1093/pcp/pct048.

Tekle, Y.I., Anderson, O.R., Katz, L.A., Maurer-Alcal, X.X., Romero, M.A.C., and Molestina, R. (2016). Phylogenomics of 'Discosea': A new molecular phylogenetic perspective on Amoebozoa with flat body forms. *Mol. Phylogenet. Evol.*, 99:144–154. doi: 10.1016/j.ympev.2016.03.029.

TerBush, D.R., Maurice, T., Roth, D., and Novick, P. (1996). The Exocyst is a multiprotein complex required for exocytosis in *Saccharomyces cerevisiae*. *The EMBO Journal*, 15(23):6483–6494. doi:10.1002/j.1460-2075.1996.tb01039.x.

Tochio, H. (2001). An Autoinhibitory Mechanism for Nonsyntaxin SNARE Proteins Revealed by the Structure of Ykt6p. *Science*, 293(5530):698–702. doi: 10.1126/science.1062950.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (1999). Evolution of protein function, from a structural perspective. *Current Opinion in Chemical Biology*, 3(5):548–556. doi:10.1016/S1367-5931(99)00007-1.

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27(10):527–533.

Toonen, R.F. and Verhage, M. (2003). Vesicle trafficking: Pleasure and pain from SM genes. *Trends in Cell Biology*, 13(4):177–186. doi:10.1016/S0962-8924(03)00031-X.

Trimble, W.S., Cowan, D.M., and Scheller, R.H. (1988). VAMP-1: A synaptic vesicle-associated integral membrane protein. *Proceedings of the National Academy of Sciences*, 85(12):4538–4542. doi:10.1073/pnas.85.12.4538.

Tromer, E.C., van Hooff, J.J.E., Kops, G.J.P.L., and Snel, B. (2019). Mosaic origin of the eukaryotic kinetochore. *Proceedings of the National Academy of Sciences*, page 201821945. doi:10.1073/pnas.1821945116.

Uemura, T., Morita, M.T., Ebine, K., Okatani, Y., Yano, D., Saito, C., Ueda, T., and Nakano, A. (2010). Vacuolar/pre-vacuolar compartment Qa-SNAREs VAM3/SYP22 and PEP12/SYP21 have interchangeable functions in *Arabidopsis*: Redundant functions of vacuolar SNAREs. *The Plant Journal*, 64(5):864–873. doi:10.1111/j.1365-313X.2010.04372.x.

Uemura, T., Sato, M.H., and Takeyasu, K. (2005). The longin domain regulates subcellular targeting of VAMP7 in *Arabidopsis thaliana*. *FEBS Letters*, 579(13):2842–2846. doi:10.1016/j.febslet.2005.04.022.

Uemura, T., Ueda, T., Ohniwa, R.L., Nakano, A., Takeyasu, K., and Sato, M.H. (2004). Systematic Analysis of SNARE Molecules in *Arabidopsis*: Dissection of the post-Golgi Network in Plant Cells. *Cell Structure and Function*, 29(2):49–65. doi:10.1247/csf.29.49.

Ungermann, C., Nichols, B.J., Pelham, H.R., and Wickner, W. (1998). A Vacuolar v–t-SNARE Complex, the Predominant Form In Vivo and on Isolated Vacuoles, Is Disassembled and Activated for Docking and Fusion. *The Journal of Cell Biology*, 140(1):61–69. doi: 10.1083/jcb.140.1.61.

van Dam, T.J.P., Townsend, M.J., Turk, M., Schlessinger, A., Sali, A., Field, M.C., and Huynen, M.A. (2013). Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proceedings of the National Academy of Sciences*, 110(17):6943–6948. doi: 10.1073/pnas.1221011110.

Van Der Giezen, M. (2009). Hydrogenosomes and Mitosomes: Conservation and Evolution of Functions. *Journal of Eukaryotic Microbiology*, 56(3):221–231. doi:10.1111/j.1550-7408.2009.00407.x.

Van Valkenburgh, H., Shern, J.F., Sharer, J.D., Zhu, X., and Kahn, R.A. (2001). ADP-ribosylation factors (ARFs) and ARF-like 1 (ARL1) have both specific and shared effectors. Characterizing ARL1-binding proteins. *J. Biol. Chem.*, 276(25):22826–22837. doi: 10.1074/jbc.M102359200.

Vávra, J. and Lukeš, J. (2013). Microsporidia and 'The Art of Living Together'. In *Advances in Parasitology*, volume 82, pages 253–319. Elsevier. ISBN 978-0-12-407706-5. doi:10.1016/B978-0-12-407706-5.00004-6.

Vedovato, M., Rossi, V., Dacks, J.B., and Filippini, F. (2009). Comparative analysis of plant genomes allows the definition of the "Phytolongins": A novel non-SNARE longin domain protein family. *BMC Genomics*, 10(1):510. doi:10.1186/1471-2164-10-510.

Veit, M., Söllner, T.H., and Rothman, J.E. (1996). Multiple palmitoylation of synaptotagmin and the t-SNARE SNAP-25. *FEBS Letters*, 385(1-2):119–123. doi:10.1016/0014-5793(96)00362-6.

Veltman, D.M., Akar, G., Bosgraaf, L., and Van Haastert, P.J. (2009). A new set of small, extrachromosomal expression vectors for *Dictyostelium discoideum*. *Plasmid*, 61(2):110–118. doi:10.1016/j.plasmid.2008.11.003.

Venkatesh, D., Boehm, C., Barlow, L.D., Nankissoor, N.N., O'Reilly, A., Kelly, S., Dacks, J.B., and Field, M.C. (2017). Evolution of the endomembrane systems of trypanosomatids – conservation and specialisation. *Journal of Cell Science*, 130(8):1421–1434. doi:10.1242/jcs.197640.

Viotti, C., Bubeck, J., Stierhof, Y.D., Krebs, M., Langhans, M., van den Berg, W., van Dongen, W., Richter, S., Geldner, N., Takano, J., Jürgens, G., de Vries, S.C., Robinson, D.G., and Schumacher, K. (2010). Endocytic and Secretory Traffic in *Arabidopsis* Merge in the Trans-Golgi Network/Early Endosome, an Independent and Highly Dynamic Organelle. *The Plant Cell*, 22(4):1344–1357. doi:10.1105/tpc.109.072637.

Volchuk, A., Ravazzola, M., Perrelet, A., Eng, W.S., Di Liberto, M., Varlamov, O., Fukasawa, M., Engel, T., Söllner, T.H., Rothman, J.E., and Orci, L. (2004). Countercurrent Distribution of Two Distinct SNARE Complexes Mediating Transport within the Golgi Stack. *Molecular Biology of the Cell*, 15(4):1506–1518. doi:10.1091/mbc.e03-08-0625.

von Mollard, G.F., Nothwehr, S.F., and Stevens, T.H. (1997). The Yeast v-SNARE Vti1p Mediates Two Vesicle Transport Pathways through Interactions with the t-SNAREs Sed5p and Pep12p. *The Journal of Cell Biology*, 137(7):1511–1524. doi:10.1083/jcb.137.7.1511.

Vossbrinck, C.R. and Debrunner-Vossbrinck, B.A. (2005). Molecular phylogeny of the Microsporidia: Ecological, ultrastructural and taxonomic considerations. *Folia Parasitologica*, 52(1-2):131–142; discussion 130.

Wacey, D., Kilburn, M.R., Saunders, M., Cliff, J., and Brasier, M.D. (2011). Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*, 4(10):698–702. doi:10.1038/ngeo1238.

Wada, Y. and Anraku, Y. (1992). Genes for Directing Vacuolar Morphogenesis in *Saccharomyces cerevisiae*: II. VAM7, A gene for regulating morphogenic assembly of the vacuoles. *The Journal of Biological Chemistry*, 267(September 15):18671–18675.

Wada, Y., Nakamura, N., Ohsumi, Y., and Hirata, A. (1997). Vam3p, a new member of syntaxin related protein, is required for vacuolar assembly in the yeast *Saccharomyces cerevisiae*. *Journal of Cell Science*, 110:1299–1306.

Walker, G., Dorrell, R.G., Schlacht, A., and Dacks, J.B. (2011). Eukaryotic systematics: A user's guide for cell biologists and parasitologists. *Parasitology*, 138(13):1638–1663. doi:10.1017/S0031182010001708.

Walker, G., Simpson, a.G.B., Edgcomb, V., Sogin, M.L., and Patterson, D.J. (2001). Ultrastructural identities of *Mastigamoeba punctachora*, *Mastigamoeba simplex* and *Mastigella commutans* and assessment of hypotheses of relatedness of the pelobionts (Protista). *Eur. J. Protistol.*, 37(1):25–49. doi:10.1078/0932-4739-00780.

Wang, H., Frelin, L., and Pevsner, J. (1997). Human syntaxin 7: A Pep12p/Vps6p homologue implicated in vesicle trafficking to lysosomes. *Gene*, 199(1-2):39–48. doi:10.1016/S0378-1119(97)00343-0.

Wang, Z. and Wu, M. (2014). Phylogenomic Reconstruction Indicates Mitochondrial Ancestor Was an Energy Parasite. *PLoS ONE*, 9(10):e110685. doi:10.1371/journal.pone.0110685.

Wang, Z. and Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *SCIENTIFIC REPORTS*, 5:7949.

Waterham, H., Titorenko, V., Swaving, G., Harder, W., and Veenhuis, M. (1993). Peroxisomes in the methylotrophic yeast *Hansenula polymorpha* do not necessarily derive from pre-existing organelles. *The EMBO Journal*, 12(12):4785–4794. doi:10.1002/j.1460-2075.1993.tb06167.x.

Watson, R.T. and Pessin, J.E. (2001). Transmembrane domain length determines intracellular membrane compartment localization of syntaxins 3, 4, and 5. *American Journal of Physiology-Cell Physiology*, 281(1):C215–C223. doi:10.1152/ajpcell.2001.281.1.C215.

Weimbs, T., Mostov, K., Hui Low, S., and Hofmann, K. (1998). A model for structural similarity between different SNARE complexes based on sequence relationships. *Trends in Cell Biology*, 8(7):260–262. doi:10.1016/S0962-8924(98)01285-9.

Wen, W., Chen, L., Wu, H., Sun, X., Zhang, M., and Banfield, D.K. (2006). Identification of the Yeast R-SNARE Nyv1p as a Novel Longin Domain-containing Protein. *Molecular Biology of the Cell*, 17(10):4282–4299. doi:10.1091/mbc.e06-02-0128.

Whelan, N.V., Kocot, K.M., Moroz, L.L., and Halanych, K.M. (2015). Error, signal, and the placement of Ctenophora sister to all other animals. *Proceedings of the National Academy of Sciences*, 112(18):5773–5778. doi:10.1073/pnas.1503453112.

White, J.F., Kingsley, K., Harper, C.J., Verma, S.K., Brindisi, L., Chen, Q., Chang, X., Micci, A., and Bergen, M. (2018). Reactive Oxygen Defense Against Cellular Endoparasites and the Origin of Eukaryotes. In *Transformative Paleobotany*, pages 439–460. Elsevier. ISBN 978-0-12-813012-4. doi:10.1016/B978-0-12-813012-4.00018-8.

WHO (2010). Working to overcome the global impact of neglected tropical diseases: First WHO report on neglected tropical diseases. Technical report, Department of Reproductive health and Research, World Health Organization, Geneva.

WHO (2018). World Health Organization World Malaria Report 2018. Technical report, World Health Organization.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.S., and Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, 111(45):E4859–E4868. doi:10.1073/pnas.1323926111.

Williams, A.L., Bielopolski, N., Meroz, D., Lam, A.D., Passmore, D.R., Ben-Tal, N., Ernst, S.A., Ashery, U., and Stuenkel, E.L. (2011). Structural and Functional Analysis of Tomosyn Identifies Domains Important in Exocytotic Regulation. *Journal of Biological Chemistry*, 286(16):14542–14553. doi:10.1074/jbc.M110.215624.

Williams, B.A.P., Hirt, R.P., Lucocq, J.M., and Embley, T.M. (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature*, 418(6900):865–869. doi:10.1038/nature00949.

Williams, T.A., Foster, P.G., Nye, T.M.W., Cox, C.J., and Embley, T.M. (2012). A congruent phylogenomic signal places eukaryotes within the Archaea. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):4870–4879. doi:10.1098/rspb.2012.1795.

Williams, T.A., Heaps, S.E., Cherlin, S., Nye, T.M.W., Boys, R.J., and Embley, T.M. (2015). New substitution models for rooting phylogenetic trees. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1678):20140336. doi:10.1098/rstb.2014.0336.

Witkos, T.M. and Lowe, M. (2016). The Golgin Family of Coiled-Coil Tethering Proteins. *Frontiers in Cell and Developmental Biology*, 3. doi:10.3389/fcell.2015.00086.

Wong, M., Gillingham, A.K., and Munro, S. (2017). The golgin coiled-coil proteins capture different types of transport carriers via distinct N-terminal motifs. *BMC Biology*, 15(1):3. doi:10.1186/s12915-016-0345-3.

Wong, M. and Munro, S. (2014). The specificity of vesicle traffic to the Golgi is encoded in the golgin coiled-coil proteins. *Science*, 346(6209):1256898–1256898. doi:10.1126/science.1256898.

Woo, Y.H., Ansari, H., Otto, T.D., Klinger, C.M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., Ali, S., Bernal, A., del Campo, J., Cihlář, J., Flegontov, P., Gornik, S.G., Hajdušková, E., Horák, A., Janouškovec, J., Katris, N.J., Mast, F.D., Miranda-Saavedra, D., Mourier, T., Naeem, R., Nair, M., Panigrahi, A.K., Rawlings, N.D., Padron-Regalado, E., Ramaprasad, A., Samad, N., Tomčala, A., Wilkes, J., Neafsey, D.E., Doerig, C., Bowler, C., Keeling, P.J., Roos, D.S., Dacks, J.B., Templeton, T.J., Waller, R.F., Lukeš, J., Oborník, M., and Pain, A. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife*, 4:e06974. doi:10.7554/eLife.06974.

Xu, Y., Martin, S., James, D.E., and Hong, W. (2002). GS15 Forms a SNARE Complex with Syntaxin 5, GS28, and Ykt6 and Is Implicated in Traffic in the Early Cisternae of the Golgi Apparatus. *Molecular Biology of the Cell*, 13(10):3493–3507. doi:10.1091/mbc.e02-01-0004.

Yadav, S., Puri, S., and Linstedt, A.D. (2009). A Primary Role for Golgi Positioning in Directed Secretion, Cell Polarity, and Wound Healing. *Mol. Biol. Cell*, 20(March):1728–1736. doi:10.1091/mbc.E08.

Yadav, S., Puthenveedu, M.A., and Linstedt, A.D. (2012). Golgin160 Recruits the Dynein Motor to Position the Golgi Apparatus. *Dev. Cell*, 23:153–165. doi:10.1016/j.devcel.2012.05.023.

Yamaguchi, T., Dulubova, I., Min, S.W., Chen, X., Rizo, J., and Südhof, T.C. (2002). Sly1 Binds to Golgi and ER Syntaxins via a Conserved N-Terminal Peptide Motif. *Developmental Cell*, 2(3):295–305. doi:10.1016/S1534-5807(02)00125-9.

Yamakawa, H., Seog, D.H., Yoda, K., Yamasaki, M., and Wakabayashi, T. (1996). Uso1 Protein Is a Dimer with Two Globular Heads and a Long Coiled-Coil Tail. *Journal of Structural Biology*, 116(3):356–365. doi:10.1006/jsbi.1996.0053.

Yamaoka, S., Shimono, Y., Shirakawa, M., Fukao, Y., Kawase, T., Hatsugai, N., Tamura, K., Shimada, T., and Hara-Nishimura, I. (2013). Identification and Dynamics of *Arabidopsis* Adaptor Protein-2 Complex and Its Involvement in Floral Organ Development. *The Plant Cell*, 25(8):2958–2969. doi:10.1105/tpc.113.114082.

Yang, B., Gonzalez, L., Prekeris, R., Steegmaier, M., Advani, R.J., and Scheller, R.H. (1999). SNARE Interactions Are Not Selective: IMPLICATIONS FOR MEMBRANE FUSION SPECIFICITY. *Journal of Biological Chemistry*, 274(9):5649–5653. doi:10.1074/jbc.274.9.5649.

Yang, H.J., Nakanishi, H., Liu, S., McNew, J.A., and Neiman, A.M. (2008). Binding interactions control SNARE specificity in vivo. *The Journal of Cell Biology*, 183(6):1089–1100. doi:10.1083/jcb.200809178.

Yeung, B.G., Phan, H.L., and Payne, G.S. (1999). Adaptor Complex-independent Clathrin Function in Yeast. *Molecular Biology of the Cell*, 10(November):3643–3659.

Yoshizawa, A.C., Kawashima, S., Okuda, S., Fujita, M., Itoh, M., Moriya, Y., Hattori, M., and Kanehisa, M. (2006). Extracting Sequence Motifs and the Phylogenetic Features of SNARE-Dependent Membrane Traffic: SNARE Sequence Motifs and Phylogenetic Features. *Traffic*, 7(8):1104–1118. doi:10.1111/j.1600-0854.2006.00451.x.

Yu, H., Rathore, S.S., Gulbranson, D.R., and Shen, J. (2014). The N- and C-terminal Domains of Tomosyn Play Distinct Roles in Soluble *N*-Ethylmaleimide-sensitive Factor Attachment Protein Receptor Binding and Fusion Regulation. *Journal of Biological Chemistry*, 289(37):25571–25580. doi:10.1074/jbc.M114.591487.

Yutin, N., Wolf, M.Y., Wolf, Y.I., and Koonin, E.V. (2009). The origins of phagocytosis and eukaryogenesis. *Biology Direct*, 4(1):9. doi:10.1186/1745-6150-4-9.

Zanchi, R., Howard, G., Bretscher, M.S., and Kay, R.R. (2010). The exocytic gene secA is required for *Dictyostelium* cell motility and osmoregulation. *J. Cell Sci.*, 123(Pt 19):3226–3234. doi:10.1242/jcs.072876.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., and Ettema, T.J.G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541(7637):353–358. doi:10.1038/nature21031.

Zeng, L., Zhang, N., Zhang, Q., Endress, P., Huang, J., and Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *The New Phytologist*, 214(3):1338–1354. doi:10.1111/nph.14503[doi].

Zeng, Q., Subramaniam, V.N., Wong, S.H., Tang, B.L., Parton, R.G., Rea, S., James, D.E., and Hong, W. (1998). A Novel Synaptobrevin/VAMP Homologous Protein (VAMP5) Is Increased during In Vitro Myogenesis and Present in the Plasma Membrane. *Molecular Biology of the Cell*, 9(9):2423–2437. doi:10.1091/mbc.9.9.2423.

Zhang, T. and Hong, W. (2001). Ykt6 Forms a SNARE Complex with Syntaxin 5, GS28, and Bet1 and Participates in a Late Stage in Endoplasmic Reticulum-Golgi Transport. *Journal of Biological Chemistry*, 276(29):27480–27487. doi:10.1074/jbc.M102786200.

Zhang, X. and Wang, Y. (2016). GRASPs in Golgi Structure and Function. *Front. Cell Dev. Biol.*, 3(January):1–8. doi:10.3389/fcell.2015.00084.

Zhang, X.M., Ellis, S., Sriratana, A., Mitchell, C.A., and Rowe, T. (2004). Sec15 Is an Effector for the Rab11 GTPase in Mammalian Cells. *Journal of Biological Chemistry*, 279(41):43027–43034. doi:10.1074/jbc.M402264200.

Zhao, J., Li, B., Huang, X., Morelli, X., and Shi, N. (2017). Structural basis for the interaction between Golgi reassembly-stacking protein GRASP55 and Golgin45. *J. Biol. Chem.*, 292(7):2956–2965. doi:10.1074/jbc.M116.765990.

Zheng, H., Bednarek, S.Y., Sanderfoot, A.A., Alonso, J., Ecker, J.R., and Raikhel, N.V. (2002). NPSN11 Is a Cell Plate-Associated SNARE Protein That Interacts with the Syntaxin KNOLLE. *Plant Physiology*, 129(2):530–539. doi:10.1104/pp.003970.

Zizioli, D., Meyer, C., Guhde, G., Saftig, P., and Schu, P. (1999). Early Embryonic Death of Mice Deficient in gamma-Adaptin. *The Journal of Biological Chemistry*, 274(9):5385–5390.

Zlatic, S.A., Grossniklaus, E.J., Ryder, P.V., Salazar, G., Mattheyses, A.L., Peden, A.A., and Faundez, V. (2013). Chemical-genetic disruption of clathrin function spares adaptor complex 3-dependent endosome vesicle biogenesis. *Molecular Biology of the Cell*, 24(15):2378–2388. doi:10.1091/mbc.E12-12-0860.

Zuckerkandl, E. and Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2):357–366. doi:10.1016/0022-5193(65)90083-4.

zur Stadt, U. (2005). Linkage of familial hemophagocytic lymphohistiocytosis (FHL) type-4 to chromosome 6q24 and identification of mutations in syntaxin 11. *Human Molecular Genetics*, 14(6):827–834. doi:10.1093/hmg/ddi076.

Zwiewka, M., Feraru, E., Möller, B., Hwang, I., Feraru, M.I., Kleine-Vehn, J., Weijers, D., and Friml, J. (2011). The AP-3 adaptor complex is required for vacuolar function in *Arabidopsis*. *Cell Research*, 21(12):1711–1722. doi:10.1038/cr.2011.99.

Appendix A

Online supplemental material

Supplemental material is available for Chapters 2 to 5 online at the following URL: `https://drive.google.com/drive/folders/1OZhkDUgxF1T3lpw0jWPVS0QNwxZw-92U?usp=sharing`